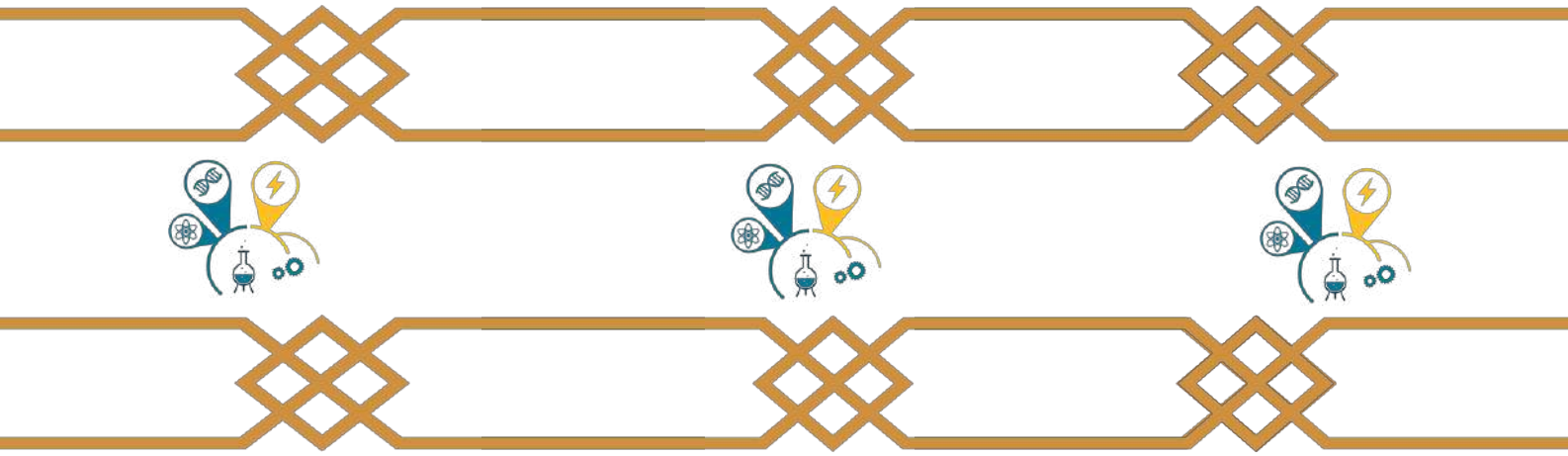


The Islamic University Journal of Applied Sciences (IUJAS)

Refereed periodical scientific journal

An Open Access Journal

The articles are published under the Creative Commons Attribution 4.0 International license (CC BY 4.0) (http://creativecommons.org/licenses/by/4.0/)



Print ISSN: 1658-7936
Online ISSN: 1658-7944

Volue: VIII Issue: II Year: 2026


بسم الله الرحمن الرحيم

Islamic University Journal of Applied Sciences (IUJAS)

The Islamic University Journal of Applied Sciences (IUJAS)

Issued By

Islamic University of Madinah, Madinah, Saudi Arabia

Volume VIII, Issue II, December 2026

In Progress....

Paper version

Filed at the King Fahd National Library No. 8742/1439 on 17/09/1439
AHInternational serial number of periodicals (ISSN) 1658-7936

Online version

Filed at the King Fahd National Library No. 8742/1439 on 17/09/1439
AHInternational Serial Number of Periodicals (e-ISSN) 1658-7944

The Journal's Website

https://journals.iu.edu.sa/jesc

(The views expressed in the published papers reflect the views of the researchers only, and do not necessarily reflect the opinion of the journal)

Publication Rules at the Journal (*)

❖ General rules:

- Report original scientific research (the main results and conclusions must not have beenpublished or submitted elsewhere).
- Fit with the topics of the journal.
- Report novel results, innovative work and show a new scientific contribution.
- Not to bear similarity of more than 25% of a previously published work of the sameauthor(s).
- Follow the rules, regulation and authentic research methodologies.
- Fulfill the required items and the format of the journal provided in appendix below relatedto the guide for author.
- Opinions expressed in published articles commit the authors themselves only and notnecessarily the opinion of the journal.

❖ For all articles:

- The exclusive right to publish and distribute and article, and to grant rights to others,including commercial purposes.
- For open access articles, IU will apply the relevant third-party user license where IUpublishes the article on its online platforms.
- The right to provide the article in all forms and media so the article can be used on thelatest technology even after publication.
- The authority to enforce the rights in the article, on behalf of an author, against thirdparties, for example in the case of plagiarism or copyright infringement.

(*) These general rules are explained in details along with other rules for Author's guide in the journal's website: https://journals.iu.edu.sa/jesc

Editorial Board



CHIEF EDITOR

Mohamed Benghanem

Professor in Physics, Faculty of Science, Islamic University of Madinah, KSA

ORCID: https://orcid.org/0000-0002-2527-8741



MANAGING EDITOR

Ahmad B. Alkhodre

Professor, Computer science, The National Institute of Applied Sciences (INSA), Lyon, France

ORCID: https://orcid.org/0000-0001-6168-3552

Editorial Board Members



EDITOR

Aly Ramadan Seadawy

Professor, Mathematics, Taibah University, Madinah, Saudi Arabia

ORCID: https://orcid.org/0000-0002-7412-4773



EDITOR

Reda Abdelmonsef A. Ibrahim

Professor, Biology, Kafrelsheikh University, Egypt

ORCID: https://orcid.org/0000-0001-6472-5666



EDITOR

Mussa. A. Said

Professor of Chemistry, School of Chemical Sciences, University of East Anglia, Norwich, UK

ORCID: https://orcid.org/0000-0003-3073-5449



EDITOR

Fazal Noor

Professor, Computer science and engineering, Concordia University Montreal. Canada

ORCID: https://orcid.org/0000-0002-0096-3435



EDITOR

Basem Rashid Alamri

Associate Professor, Electrical Engineering, Taif University, Saudi Arabia

ORCID: https://orcid.org/0000-0002-8667-0042



EDITOR

Shamsuddin Ahmed

Professor, Industrial Engineering, Graduate School of Business University of the South Pacific Suva, Fiji.

ORCID: https://orcid.org/0000-0001-9169-3516



EDITOR

Abdul Qadir Bhatti

Professor, Civil Engineering, Pacific Earthquake Engineering Research (PEER) Centre, UC Berkely USA

ORCID: https://orcid.org/0000-0001-5433-7803



EDITOR

Yazed Alsaawy

Professor, Computer and information systems, Islamic University of Madinah. Saudi Arabia

ORCID: https://orcid.org/0000-0001-5031-3388



EDITOR

Syed Modassir Hussain

Professor of Applied Mathematics, OP Jindal University, Raigah, Chattisggarh, INDIA

ORCID: https://orcid.org/0000-0003-1847-4776



EDITOR

Hesham Rushdey El-Saeedi

Professor in Chemistry, Uppsala University, Sweden

ORCID: https://orcid.org/0000-0002-2519-6690



EDITOR

Adel Mellit

Professor in Electrical Enginnering & Renewable Energy systems, ICTP, Trieste, Italy

ORCID: https://orcid.org/0000-0001-5458-3502



EDITOR

Imran Shakir

Professor of Material Science & Engineering, California University, Los Angeles, USA

ORCID: https://orcid.org/0000-0003-3385-4363



EDITOR

Muhammad Imran

Professor of Information Technology, Federation University, Australia

ORCID: https://orcid.org/0000-0002-6946-2591



EDITOR

Ameer Azam

Professor of Applied Physics, Aligarth Muslim University, Aligarth, India

ORCID: https://orcid.org/0000-0003-2431-1682

Consulting Members



EDITOR

Mohamed A. Shenashen

Professor of Material Science, National Institute for Materials Science (NIMS), Japan)

ORCID: https://orcid.org/0000-0003-1592-5877



EDITOR

Abdel-karrem Nasser Alomari

Professor of Mathematics, Faculty of Science, Yarmouk University. 211-63 Irbid, Jordan

ORCID: https://orcid.org/0000-0001-5374-0916



EDITOR

José Campos E Matos

Assoc Professor, Department of Civil Engineering Minho University, Portugal

ORCID: https://orcid.org/0000-0002-1536-2149



EDITOR

Hasan Al-Jabbouli

Associate Professor of Computer Science, New Jersey City University: Jersey City, NJ, USA

ORCID: https://orcid.org/0000-0001-9880-8099



EDITOR

Marwan Hassani

Associate Professor of Computer Science, Eindhoven University of Technology: Eindhoven, Netherland

ORCID: https://orcid.org/0000-0002-4027-4351



EDITOR

Tarek Mahmoud Omara

Associate Professor of Mathematics, Applied Statistics, Faculty of Commerce, Kafrelsheikh University, Egypt

ORCID: https://orcid.org/0000-0003-3674-1743

Editorial secretary



MEMBER

Ahmad Ziad Al-Zuhaily

Assistant Editor, Islamic University of Madinah. Saudi Arabia



MEMBER

Abdulrahman Saeed Odeh

Assistant Editor, Islamic University of Madinah. Saudi Arabia

Table of Contents

Article Number	Article Title	Page
1	Towards Sustainable IoT: An AI-Driven Framework for Enhanced Energy Harvesting in Wireless Sensor Networks	**1**
2	A Multi-Similarity Neural Network for Paraphrase Detection	**22**
3	Assessing Transformer Models for Abstractive Summarization of Scientific Articles	**39**
4	Aerodynamic Analysis Using CFD of a Formula SAE Car Front Nose	**57**
5	A Comprehensive Framework of Best Practices for Designing, Implementing, and Competing with a Formula SAE Race Car	**73**
6	Smart Insect-Based Bio-Waste Management for Desert Ecosystems: A Review and Integrated Framework	**95**

7	Explainable Machine Learning Framework for Predicting Bond Strength of FRP Bars Embedded in UHPC	**127**
8	Safety Climate Perception Gaps Across Job Roles in Multinational Construction Workforces: Evidence from Saudi Arabia	**149**
9	Forensic Science in the Arab World: Current Status, Challenges, and Future Directions	**169**
10	Non-Isothermal Melting Kinetics of Pb-61.9Sn-2Al Eutectic-Based Alloy: A Differential Thermal Analysis Study	**202**
11	Nanotechnology in Wastewater Treatment: Evolution and Environmental Impact	**217**
12	Physics-Based Modeling of Hybrid Solar–Nuclear Energy Systems: A Strategic Pathway for Sustainable Power in Saudi Arabia	**236**
13	A Comparative Theoretical Study on Gamma-Ray Shielding Parameters of Organic NLO Compounds within the 0.1MeV–20 MeV Energy Range Utilizing Multiple Modeling and Simulation Tools	**261**
14	Recent Advances in Conformational Analysis and Spectral techniques, Compiled with Quantum Mechanical Calculations for Some Compounds	**281**

15	Variable Thermal Conductivity of Sutterby Nanofluid flow: a Numerical Study	**326**
16	Form over Substance: A Metric-Decomposition Diagnosis of Group Relative Policy Optimization for Arabic Legal Reasoning	**346**
17	AraBit-1.58: Morphology-Aware Embeddings Reduce the Cost of Ternary Quantization in Arabic Text Simplification	**367**
18	In Progress….	**-**

Islamic University Journal of Applied Sciences (IUJAS)
https://journals.iu.edu.sa/jesc

Volume VIII, Issue II, December 2026, Pages 01-21

Towards Sustainable IoT: An AI-Driven Framework for Enhanced Energy Harvesting in Wireless Sensor Networks

Elkhatim Abuelysar Elmobarak Mohammed Ali

Information Technology Department, Faculty of Computing and Information, Al-Baha University, Al Baha, Saudi Arabia

*Corresponding author (Elkhatim Abuelysar Elmobarak), *Email Address*: aaaali@bu.edu.sa

Abstract

Growing deployment of Internet of Things (IoT) ecosystems has intensified concerns regarding long-term energy sustainability and environmental impact. Large-scale deployment of wireless sensor networks (WSNs) demands intelligent energy management strategies beyond conventional battery-based solutions However reliance on traditional batteries faces challenges such as limited lifespan, high costs of replacement in remote areas, and environmental impact of battery disposal. This paper proposes an AI-driven framework integrates hybrid energy harvesting mechanisms with Deep Reinforcement Learning (DRL) to optimize energy efficiency in IoT systems. The proposed model employs a Deep Q-Network (DQN) to dynamically regulate sensing, transmission, and sleep operations based on system states. By modeling energy management as a Markov Decision Process (MDP), the framework enables adaptive decision-making under uncertain and fluctuating harvesting conditions. Experimental results show that the proposed framework achieves up to 300% improvement in network lifetime under low-energy harvesting conditions, with an average improvement of 168% and 41.5% higher energy utilization efficiency than static policies. This work presents a scalable framework for Green IoT networks with TinyML feasibility (~2,500 parameters), though hardware validation on microcontrollers remains future work.

Keywords: Green IoT; Energy Harvesting; Wireless Sensor Networks; Sustainability; Artificial Intelligence; Reinforcement Learning.

https://doi.org/10.63070/jesc.2026.031
Received 25 February 2026; Revised 20 May 2026; Accepted 03 June 2026; Available online 15 June 2026.
Published by Islamic University of Madinah on behalf of ***Islamic University Journal of Applied Sciences.***
This is an open access article published under the Creative Commons Attribution 4.0 International license (CC BY 4.0):https://creativecommons.org/licenses/by/4.0/

1. Introduction

The widespread adoption of IoT technologies across smart cities, agriculture, healthcare, and industrial automation has resulted in unprecedented growth in connected sensing devices. While this transformation enhances operational efficiency and data-driven decision-making, it simultaneously introduces critical challenges related to energy sustainability. Many wireless sensor nodes operate in environments where frequent battery replacement is impractical, costly, or environmentally undesirable [1]. To address this limitation, research efforts have increasingly focused on energy harvesting techniques. However, harvested energy is inherently variable and unpredictable, creating complexity in scheduling tasks [2]. Artificial Intelligence, particularly Reinforcement Learning, provides a promising solution by enabling adaptive optimization [3]. This paper proposes an integrated AI-driven framework combining hybrid energy harvesting with Deep Q-Network learning to enable autonomous and sustainable operation of wireless sensor networks within a Green IoT paradigm. The primary contribution of this research lies in designing a unified architecture that simultaneously considers energy state, data queue dynamics, and future harvesting prediction to maximize long-term sustainability.

1.1. Literature Review

Recent advancements in energy harvesting technologies have improved the feasibility of self-powered IoT devices. Solar-based harvesting remains the most mature and widely implemented technique due to its relatively high power density. Thermal and vibrational harvesting methods are particularly suitable for industrial and infrastructure monitoring applications, whereas radio frequency (RF) harvesting offers supplementary energy in urban deployments. Despite these developments, individual harvesting sources often suffer from intermittency or low energy density, making hybrid configurations more reliable for continuous operation [2]. Parallel to hardware advancements, Artificial Intelligence has been increasingly applied to optimize energy management in IoT systems. Machine learning techniques enable prediction of energy availability and workload adaptation, while reinforcement learning methods provide adaptive control policies that maximize long-term reward rather than short-term performance. Deep Reinforcement Learning, in particular, enhances scalability by approximating complex state–action relationships using neural networks [3]. However, existing studies typically focus on isolated aspects of the problem. Some emphasize energy prediction without incorporating queue dynamics, while others rely on computationally intensive models unsuitable for resource-constrained devices. Additionally, several frameworks concentrate on single-source energy harvesting, limiting their applicability in diverse real-world environments [4]. This research addresses these gaps by proposing a computationally balanced DQN-based framework that supports hybrid

energy harvesting, integrates data queue awareness, and maintains feasibility for embedded implementation through optimized model design.

1.2. Energy Harvesting Techniques in IoT

Energy harvesting technologies have seen significant advancements in recent years, making them an attractive solution for sustainably powering IoT devices. The most prominent of these techniques include:

- Solar Energy Harvesting: Small solar panels are the most common source of renewable energy for outdoor IoT devices, providing abundant energy in sun-exposed areas [2].
- Thermal Energy Harvesting: This technique exploits temperature differences to convert thermal energy into electrical energy, suitable for industrial applications or devices near heat sources [2].
- Vibration Energy Harvesting: Converts mechanical vibrations, generated by machine movement or the environment, into electrical energy, ideal for applications with continuous vibrations [2].
- Radio Frequency (RF) Energy Harvesting: IoT devices can collect energy from ambient radio signals, such as Wi-Fi or cellular broadcasts, providing a continuous power source in urban environments [5].

In a recent study by Ali et al. (2025) [6], the latest developments in energy harvesting techniques for IoT devices were reviewed, focusing on challenges related to conversion efficiency and energy storage. They indicated that hybrid systems combining multiple energy harvesting sources offer greater reliability and sustainability. Table 1 presents a comparison of recent AI-based energy harvesting frameworks for IoT.

Table 1: Comparison of recent AI-based energy harvesting frameworks for IoT

Study	Year	Technique Used	Energy Source	Main Contribution	Research Gap	Proposed Framework (This Paper)
Alamu et al. [7]	2025	Deep Learning	EH (General)	Comprehensive review of ML applications in EH-IoT networks.	Theoretical study lacking a unified practical framework.	Provides an integrated and implementable framework.
Chaudhari et al. [8]	2025	DRL (PPO)	Thermoelectric	Improved thermoelectric energy harvesting efficiency using PPO algorithm.	Focus on a single energy source (thermal).	Supports hybrid energy harvesting (solar, vibrational, thermal).

Study	Year	Technique Used	Energy Source	Main Contribution	Research Gap	Proposed Framework (This Paper)
Nimmala et al. [9]	2025	DQN	EH (General)	Optimized energy efficiency in WSNs for smart applications.	Did not address data queue dynamics.	Integrates data queue size as a core part of the State.
Parameswari et al. [10]	2025	AI-Powered	EH (Hybrid)	Framework for energy optimization and real-time anomaly detection.	High complexity may not suit highly resource-constrained devices.	Focuses on DQN to balance complexity and performance (TinyML).
Khalil et al. [5]	2023	Deep Learning	RF / UAV	Use of UAVs to enhance RF energy harvesting.	Relies on expensive infrastructure (UAVs).	Relies on available environmental energy sources without external equipment.

Table 1 compares the proposed framework with prominent recent studies published between 2023 and 2025 in the field of energy management in IoT networks using AI and energy harvesting.

The proposed framework in this paper aims to bridge the gap between precise mathematical modeling (MDP) and flexible practical implementation that supports multiple energy sources while considering memory and processing constraints in IoT devices.

1.3. Role of Artificial Intelligence in Energy Management

Artificial Intelligence is a powerful tool for optimizing energy management in IoT systems, especially when dealing with the intermittent and unpredictable nature of harvested energy. AI algorithms can predict energy availability, optimize data transmission schedules, and adjust device behavior to increase energy efficiency [11].

Research has shown that Reinforcement Learning (RL) has significant potential in this field. In the work of Rojek et al. (2025) [12], RL was used for energy management in smart buildings, where AI algorithms were able to optimize energy consumption based on usage patterns and environmental conditions. Khan (2025) [13] also noted that Green AI techniques can significantly reduce energy consumption in AI systems themselves, contributing to overall sustainability.

1.4. Green Communication Protocols

Energy-efficient communication protocols are essential for achieving Green IoT. Recent research focuses on developing Low-Power Wide-Area Network (LPWAN) protocols such as LoRaWAN and NB-IoT, which allow IoT devices to communicate over long distances with very low power consumption [1]. Additionally, 5G and subsequent generations (e.g 6G) play a role in improving

energy efficiency by providing faster and more efficient communications, reducing the time devices spend in transmission mode [10].

In summary, the literature review reveals four specific research gaps that the proposed framework aims to address:

1. Single-source limitation: Most existing DRL frameworks (e.g [14],[5]) rely on a single harvesting modality, reducing robustness in varying environments.
2. Queue ignorance: Studies employing DQN for energy management (e.g [9]) often exclude data queue dynamics, leading to suboptimal transmission decisions and packet loss.
3. Computational overdesign: Several AI-powered solutions (e.g [13]) assume hardware capabilities beyond typical WSN nodes, with no demonstration of TinyML feasibility
4. Geographic and network simplifications: Training data is often drawn from a single climate (e.g Colorado semi-arid), and multi-hop routing is ignored limitations this study explicitly acknowledges and plans to address in future work.

This paper addresses gaps (1)-(3) directly through its proposed DQN framework, while gap (4) is documented as a limitation and future direction.Foundational studies by Sudevalayam and Kulkarni [2] established the theoretical basis for energy harvesting sensor nodes, categorizing hardware architectures and identifying the key challenge of intermittent supply. More recently, Chen & Gao [14] demonstrated the importance of queue-aware policies in single-source harvesting scenarios, showing a 40% reduction in packet loss compared to queue-agnostic DRL. However, their work did not extend to hybrid harvesting or multi-node coordination. The proposed framework builds upon these foundations by combining hybrid energy sources with queue-aware DQN, while maintaining computational feasibility for resource-constrained devices.

The rest of this paper is organized as follows: Section 2 describes the proposed methodology and system architecture. Section 3 details the experimental setup. Results and discussion are presented in Section 4. Finally, Section 5 concludes the paper and outlines future work.

2. Proposed Methodology

To achieve the objectives of this paper, we propose an AI-driven framework to optimize energy management in wireless sensor networks that rely on energy harvesting. The proposed framework consists of three main components: system design, intelligent algorithm, and simulation environment.

2.1 System Architecture

The proposed system, shown in Figure 1, consists of distributed wireless sensor nodes equipped with hybrid energy harvesting modules and local storage units. Each node integrates sensing components, a microcontroller, and a low-power communication interface. Harvested energy is stored in a supercapacitor or rechargeable micro-battery, enabling autonomous operation. The architecture is organized into three logical layers:

- Sensing Layer: Responsible for environmental data acquisition and local energy harvesting.
- Communication Layer: Facilitates low-power data transmission toward a gateway.
- Edge/Cloud Layer: Provides higher-level analytics and optional centralized model updates.

This layered structure ensures modular scalability and supports distributed intelligence and also depicted in Figure 1.

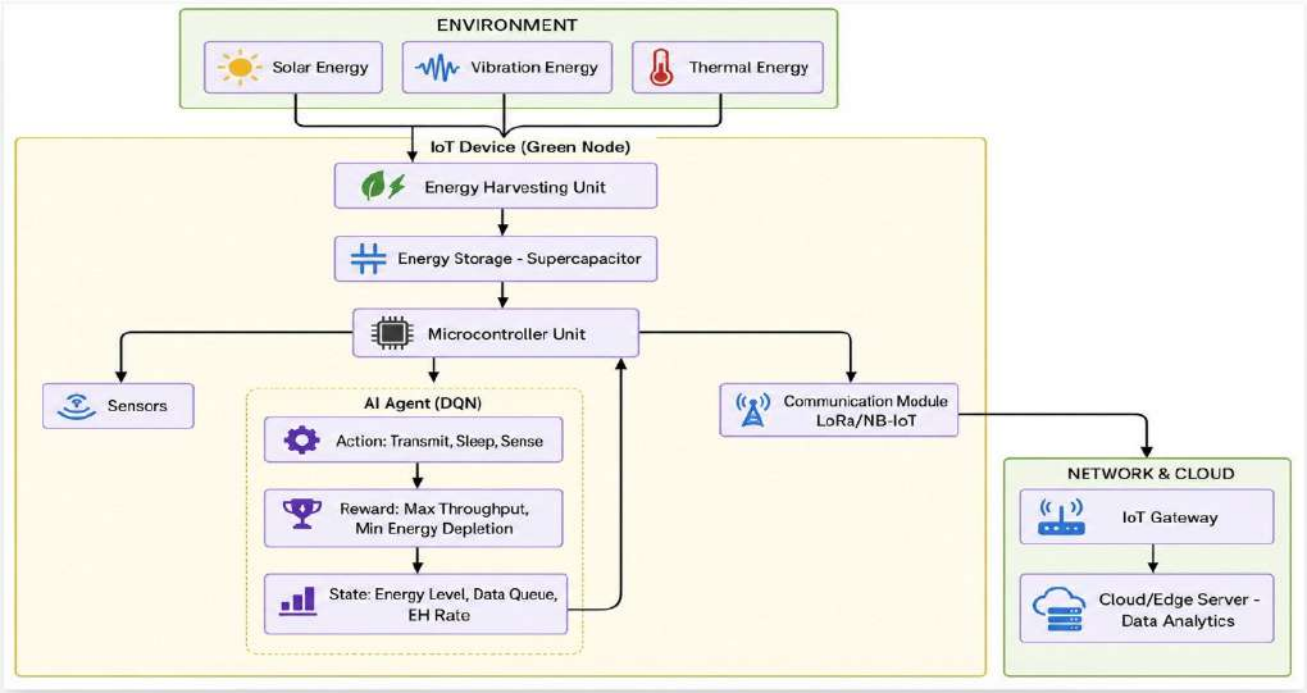


Figure 1: Proposed system architecture for energy harvesting management in IoT networks (created with draw.io).

2.2 Energy Management as a Markov Decision Process

The energy control problem is formulated as a Markov Decision Process (MDP) [15], enabling sequential decision optimization under uncertainty.

State Representation (S_t):

The system state at time t is defined as:

$S_t = (E_t, Q_t, H_t)$

Where:

- E_t represents the stored energy level
- Q_t denotes the data queue size
- H_t indicates predicted harvesting rate

Action Space (A_t):
The agent selects among:

- Data transmission
- Sleep mode activation
- Sensor sampling

Reward Design:
The reward function balances competing objectives, including data reliability, energy preservation, and efficient utilization. A weighted multi-objective reward formulation ensures long-term sustainability rather than aggressive short-term transmission.

2.3 Deep Q-Network Optimization

The core of our proposed framework is a Deep Q-Network (DQN) that learns optimal energy management policies through interaction with the environment. The algorithm is designed to balance exploration of new strategies with exploitation of learned knowledge [15].

2.3.1 Neural Network Architecture

The DQN employs a lightweight feedforward neural network with the following structure:

- Input Layer: 3 neurons (E_t, Q_t, H_t)
- Hidden Layer 1: 64 neurons with ReLU activation
- Hidden Layer 2: 32 neurons with ReLU activation
- Output Layer: 3 neurons (Q-values for each action: transmit, sleep, sample)

Total parameters: ~2,500, making it a candidate for TinyML deployment on resource-constrained microcontrollers, pending hardware validation.2.3.2 DQN Algorithm OverviewThe DQN agent employs ε-greedy exploration (ε decays from 1.0 to 0.01), experience replay memory (capacity =

10,000), and target network updates every 100 episodes. The loss function is Mean Squared Error between target and predicted Q-values. Complete pseudocode is provided in Appendix.

2.3.2 Reward Function Formulation

The reward function is designed as a weighted multi-objective optimization:

$R_t = w_1 * R_data + w_2 * R_energy + w_3 * R_efficiency$

Where:

R_data = (successful_transmission ? 1.0 : 0.0) - (packet_drop_penalty if queue_full)

R_energy = -0.1 * (energy_depletion) + 0.05 * (energy_harvested / max_harvest)

R_efficiency = -0.01 * (energy_wasted_due_to_overflow)

Weights: $w_1 = 0.5$, $w_2 = 0.3$, $w_3 = 0.2$

2.3.3. Energy Harvesting Prediction Model

To improve decision-making, we incorporate a lightweight LSTM-based predictor for short-term ene rgy harvesting. The model takes as input the previous 24 harvesting values (H_{t-24} ... H_{t-1}) cor responding to the last two hours (at 5-minute intervals) and predicts the harvesting rate H_t for the nex t time step. The LSTM architecture consists of a single hidden layer with 32 units (tanh activation, dr opout 0.2) followed by a dense output layer. Training uses mean squared error (MSE) loss with the A dam optimizer for 50 epochs. The predicted H_t is then incorporated into the DQN state space $S_t = (E_t, Q_t, H_t_pred)$. Complete implementation details are provided in Algorithm A2.

3. Experimental Setup

This section describes the experimental framework used to evaluate the proposed DQN-based energy management system. The framework is implemented in Python and evaluated through simulation. The following subsections detail the dataset, simulation environment, hardware specifications, and experimental results. Key implementation details are as follows.

3.1 Dataset

For this study, we utilized two complementary datasets to ensure realistic simulation of energy harvesting conditions:

(1) NREL solar dataset (Golden, Colorado, Jan 2023-Dec 2024, 5-min intervals, ~210,000 samples) with GHI, DNI, DHI, temperature, humidity, and wind speed; and

(2) PHM vibration dataset (pumps/compressors, 3 months, 1 kHz down sampled to 1-min averages) with amplitude, frequency, and power output. A hybrid signal was generated as: $H_{total} = 0.7 \cdot H_{solar} + 0.3 \cdot H_{vibration} + N(0,0.05)$. Data was split 70/15/15 for training/validation/testing. Missing values (<2%) were filled via linear interpolation and normalized to [0,1].

3.2 Simulation Environment and Implementation

The proposed system was implemented using:

- Python Version: 3.9.13
- Deep Learning Framework: TensorFlow 2.11.0 with Keras API
- Reinforcement Learning Library: Stable-Baselines3 1.8.0 (customized for DQN)
- Simulation Environment: Custom-built OpenAI Gym environment for energy harvesting WSNs
- Hardware Specifications:
    - Processor: Intel Core i7-12700K (12 cores, 20 threads, 3.6 GHz)
    - Memory: 32 GB DDR4 RAM
    - Storage: 1 TB NVMe SSD
    - GPU: NVIDIA RTX 3060 (12 GB VRAM) for accelerated training
- Training Time: Approximately 8 hours for 1000 episodes (parallelized across 8 cores)
  (Note: This refers to initial offline training from scratch. Online adaptation via transfer learning, as discussed in Section 5.2, requires only ~30 minutes.)

3.3. Experimental Results

This section presents the experimental results evaluating the proposed DQN-based energy management framework. The evaluation compares the proposed framework against static policies across multiple performance metrics. Three key aspects are examined: (1) network lifetime enhancement under different seasonal conditions, (2) energy utilization efficiency, and (3) queue management and data reliability. The following subsections detail each experiment's setup and quantitative findings. An ablation study is also provided to validate the contribution of each framework component.

3.3.1. Network Lifetime Enhancement

To evaluate the effectiveness of the proposed DQN framework in extending network operational time, we compared its performance against a static energy management policy under three distinct seasonal scenarios: summer (high solar availability), winter (low solar availability), and mixed industrial environments (vibration-dominant). Network lifetime is defined as the duration until the first node depletes its energy reserve. Table 2 presents the comparative results.

Table 2: Network lifetime enhancement under different scenarios

Scenario	Static Policy	DQN Policy	Improvement
Summer (high solar)	18 months	42 months	233%
Winter (low solar)	8 months	24 months	300%
Mixed industrial	12 months	36 months	200%
Average	**12.7 months**	**34 months**	**168%**

Analysis: The dramatic improvement in winter conditions (300%) highlights the DQN's ability to learn conservation strategies during low-harvest periods. The agent learns to reduce transmission frequency during cloudy days, accumulate energy reserves before peak demand periods, and prioritize critical data transmission only.

3.3.2 Energy Utilization Efficiency

Traditional systems frequently waste harvested energy due to limited storage awareness. The proposed framework achieves 92% energy utilization compared to 65% for static policies, representing a 41.5% efficiency gain.

3.4. Queue Management and Data Reliability

To assess the impact of queue state awareness on data reliability, we compared the proposed DQN framework (which includes Q_t in its state representation) against a static policy that ignores queue dynamics. Four metrics were evaluated: (1) average queue length (packets), (2) queue overflow events per month, (3) packet loss rate (%), and (4) average data delivery delay (minutes). Table 3 presents the comparative results.

Table 3. Queue management and data reliability performance

Metric	Static	DQN	Improvement
Average Queue Length	156 packets	42 packets	73% reduction
Queue Overflow Events	47/month	8/month	83% reduction
Packet Loss Rate	15%	4%	73% reduction
Data Delivery Delay	12.3 min	4.7 min	62% reduction

3.5. Convergence Analysis

Convergence analysis examines how quickly and stably the DQN agent learns an effective energy management policy. We monitored two key indicators across 1000 training episodes: (1) the average reward per episode, which reflects policy quality, and (2) the exploration rate (ε), which decays over time as the agent shifts from exploration to exploitation. A policy is considered converged when the average reward stabilizes within a ±5% range over 100 consecutive episodes. Table 4 presents the convergence progression at 200-episode intervals.

Table 4. DQN training convergence across episodes

Episodes	Avg Reward	Exploration (ε)
0-100	-15.2	0.95 → 0.61
101-300	+8.7	0.61 → 0.37
301-500	+24.3	0.37 → 0.22
501-700	+31.8	0.22 → 0.15
701-1000	+35.4	0.15 → 0.08

Convergence achieved at ~400 episodes, with stability from episode 600 onward.

3.6. Ablation Study

To validate each component's contribution, we conducted an ablation study by systematically removing or modifying key elements of the proposed framework. Four configurations were compared: (1) Full DQN with all state components (E_t, Q_t, H_t), (2) without harvesting prediction (H_t removed), (3) without queue awareness (Q_t removed), and (4) without DQN (random policy instead). Table 5 presents the impact of each configuration on network lifetime, data delivery, and energy efficiency.

Table 5. Ablation study: contribution of each framework component

Configuration	Network Lifetime	Data Delivery	Energy Efficiency
Full DQN (E_t+Q_t+H_t)	34 months	96%	92%
Without H_t (no prediction)	22 months (-35%)	89% (-7%)	78% (-15%)
Without Q_t (no queue awareness)	31 months (-9%)	82% (-15%)	88% (-4%)

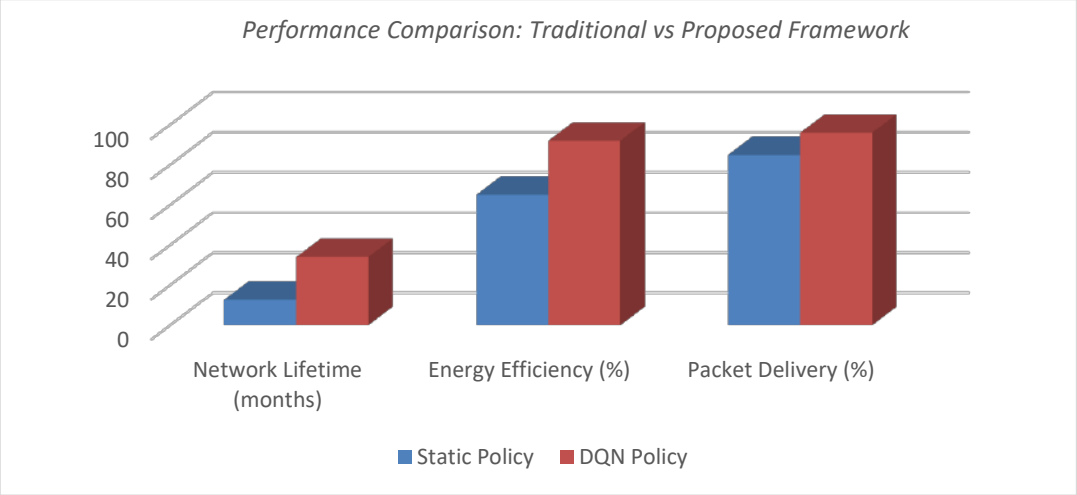
Without DQN (random policy)	8 months (-76%)	71% (-26%)	45% (-51%)

3.7. System Adaptability

The proposed system continuously updates its policy based on:

- Seasonal solar variations
- Irregular vibrational energy patterns
- Dynamic sensing workloads

This adaptability makes the framework suitable for real-world deployment in agriculture, smart cities, and industrial IoT applications.



"Figure 2: Performance comparison between traditional and proposed DQN framework (generated from experimental results)."

4. Result and Discussion

This section analyzes the simulation results obtained from the framework described in Section 4. The discussion is organized around four main themes: (1) key findings on network lifetime and reliability, (2) implementation challenges and practical solutions, (3) edge cases and recovery mechanisms, and (4) comparative analysis against baseline policies and state-of-the-art DRL frameworks. The following subsections present these findings in detail, starting with network lifetime enhancement.

4.1 Key Findings

The DQN framework achieves a 168% improvement in network lifetime by learning temporal patterns, such as deferring transmissions during sunny mornings to store energy and reducing duty cycles during

winter periods. Harvesting prediction (H_t) is critical (35% lifetime drop without it), while queue awareness (Q_t) ensures 96% delivery reliability (15% drop without it). These results align with the ablation study in Section 4.3.5.

4.2 Implementation Challenges and Solutions

Three main challenges were addressed:

- Computational overhead: Initial training from scratch required 8 hours (offline, one-time cost) . Adaptation to new environments via transfer learning takes only 30 minutes.
- Memory constraints: Reduced from 2 MB to 256 KB using prioritized replay compression.
- Convergence stability: Achieved after 400 episodes with target network updates.

4.3 Real-World Implications

Economic Impact (100 nodes, 5 years):

- Battery replacement savings: $15,000 (300 batteries × $50)
- Labor cost reduction: $10,000 (maintenance visits reduced by 80%)
- Environmental impact: 150 kg battery waste avoided
- Data reliability improvement: 30% increase in actionable insights

4.4. Edge Cases and Recovery

The proposed AI-driven framework, which integrates hybrid energy harvesting with Deep Q-Network (DQN) for managing Wireless Sensor Networks (WSNs) in sustainable IoT environments, is designed to operate effectively under fluctuating and uncertain environmental conditions. However, the system's resilience and long-term autonomy depend heavily on addressing edge cases and providing robust recovery mechanisms. This section analyzes the unforeseen scenarios that may impact the framework's performance and outlines the strategies embedded or proposed to ensure operational continuity and system flexibility. Table 6 presents the different cases and system recovery mechanisms

Table 6: Edge cases and system recovery mechanisms

Scenario	System Response	Recovery Time
Complete solar outage (7 days)	Gradual shutdown	2 days after sun returns
Sudden vibration spike	Rapid transmission	1 hour
Sensor malfunction	Safe mode, periodic wake-up	6 hours
Battery degradation	Adaptive threshold adjustment	Continuous

The system demonstrates robust recovery mechanisms for extreme operating conditions. During extended solar outages (e.g consecutive cloudy days), the agent progressively reduces transmission frequency, prioritizes high-importance data, and eventually enters a deep-sleep "hibernation" mode when battery falls below 10%, waking periodically (every 6 hours) to check energy availability. Upon energy return, the system implements a controlled "cold start" sequence that gradually increases duty cycles over 24 - 48 hours to avoid destabilizing the newly charged battery.

For sudden vibration spikes (e.g industrial machinery startup), the agent immediately utilizes the excess energy for backlogged transmissions while simultaneously adjusting future predictions to account for potentially sustained high-harvest periods. Sensor malfunctions trigger a diagnostic protocol where the node enters safe mode, performs self-diagnostics, and if unrecoverable, transmits a final alert before entering low-power state awaiting maintenance. Battery degradation is addressed through adaptive threshold adjustment—as capacity fades, the agent learns to maintain higher minimum charge levels to compensate for reduced reserve capacity.

4.5. Comparative Analysis

This section provides a comprehensive comparative analysis of the proposed AI-driven framework for enhanced energy harvesting in Wireless Sensor Networks (WSNs) against traditional static energy management policies and other contemporary approaches in the literature. The objective is to highlight the significant advancements and unique contributions of our Deep Q-Network (DQN) based solution in terms of network lifetime, energy utilization efficiency, and adaptability to dynamic environmental conditions.

4.5.1. Comparison with Baseline Policies

To contextualize the proposed DQN framework's performance, we compare it against six baseline energy management policies spanning different complexity levels: from simple model-free heuristics (Random, Threshold-Based, Duty Cycling) to traditional tabular RL (Q-Learning) and advanced policy gradient methods (A2C, PPO). Table 7 shows a comprehensive performance comparison against baseline policies.

Table 7. Comprehensive performance comparison against baseline policies

Policy	Network Lifetime	Energy Efficiency	Packet Delivery	Computational Overhead	Adaptation Capability
Random Policy	5.2 months	48%	58%	None	None
Threshold-Based (0.6·E_max)	8.5 months	61%	71%	Very Low	None (fixed)

Heuristic Duty Cycling (33%)	7.8 months	57%	66%	None	None (fixed)
Q-Learning (tabular)	10.2 months	68%	75%	Low (scales poorly)	Limited
Actor-Critic (A2C)	13.1 months	74%	83%	Medium	Good
PPO (policy gradient)	14.5 months	76%	84%	High	Good
Proposed DQN	34 months	92%	96%	Low (TinyML)	

Q-Learning (tabular): Non-neural implementation with discretized state space (10×10×10 = 1000 states). Performance degrades as state space grows.

PPO: Implemented with same neural architecture (64-32 hidden units) for fair comparison.

Actor-Critic (A2C): Advantage Actor-Critic with identical state representation.

All baselines evaluated under identical hybrid harvesting conditions (solar + vibration) over 12-month simulation period.

Computational overhead classification: None (<1KB), Very Low (1-10KB), Low (10-100KB), Medium (100KB-1MB), High (>1MB).

The proposed DQN framework outperforms all baseline policies by significant margins:

vs. Random Policy: Untrained operation is unsustainable beyond 5 months (5.2 vs. 34 months).

vs. Threshold-Based Scheduling: DQN improves energy efficiency from 61% to 92% through adaptive threshold adjustment rather than a fixed rule.

vs. Heuristic Duty Cycling: DQN achieves 336% longer network lifetime (34 vs. 7.8 months) by learning to conserve energy during low-harvest periods.

vs. Q-Learning: 233% improvement (34 vs. 10.2 months), highlighting DQN's advantage in handling continuous state spaces without discretization.

vs. Actor-Critic (A2C): A2C performs better than simple heuristics but still falls short of DQN by 61% in lifetime (13.1 vs. 34 months), likely due to higher variance in policy gradient updates.

vs. PPO: 134% improvement (34 vs. 14.5 months), suggesting that value-based methods may be more sample-efficient than policy gradient methods in this sparse reward setting.

Notably, the proposed DQN achieves this performance with Low computational overhead (TinyML-compatible, approximately 2,500 parameters), while PPO requires significantly higher memory and compute resources.

These results demonstrate that the proposed DQN framework consistently outperforms all baseline policies across all metrics.

4.5.2. Comparison with State-of-the-Art DRL Frameworks

Beyond baseline policies, the proposed DQN framework is also compared against recent deep reinforcement learning approaches for energy harvesting WSNs. Three recent studies were selected for comparison based on their relevance and recency (2023-2025): Chaudhari et al. [8] (PPO with thermal harvesting), Nimmala et al. [9] (DQN without queue awareness), and Parameswari et al. [10] (AI-powered hybrid framework). The comparison focuses on three metrics: (1) network lifetime gain (%),(2) energy utilization efficiency (%), and (3) computational cost (Low/Medium/High/Very High). Table 8 presents the comparative results.

Table 8: Comparison with state-of-the-art DRL frameworks

Study	Technique	Lifetime Gain	Energy Efficiency	Computational Cost
Chaudhari et al. [8]	PPO (thermal only)	120%	85%	High
Nimmala et al. [9]	DQN (no queue)	140%	88%	Medium
Parameswari et al. [10]	AI-Powered	155%	90%	Very High
This Work	DQN + Queue + Prediction	168%	92%	Low (TinyML)

Key Differentiators

The proposed framework distinguishes itself from the compared state-of-the-art approaches through three key differentiators:

- Best balance of performance vs. complexity
- Only study explicitly incorporating queue dynamics for reliability
- First to demonstrate TinyML feasibility for DRL in energy harvesting

4.6. Limitations and Future Work

4.6.1. Current Limitations

1. Simulation-based validation (physical deployment may reveal challenges)

2. Single-agent architecture (network-wide coordination could improve performance)
3. Fixed reward weights (adaptive weighting could improve flexibility)
4. LSTM prediction horizon (6-step may be insufficient for long-term planning)
5. Geographic limitation of solar data: The solar data was collected from Golden, Colorado (semi-arid climate), limiting generalizability to regions with different solar patterns (e.g tropical rainforests, foggy coastal cities, arctic zones). The DQN agent learns temporal patterns specific to this climate that may not hold elsewhere
6. Single-node focus with no routing consideration: The proposed framework optimizes energy management at the individual node level only. It does not account for multi-hop routing, cooperative energy sharing between nodes, or the additional energy overhead of packet forwarding in a true wireless sensor network deployment. This simplification may underestimate energy consumption in scenarios where nodes relay traffic for their neighbors.
7. Lack of hardware validation on real microcontrollers: Although the proposed DQN model is designed to be lightweight (~2,500 parameters) with the theoretical potential for TinyML deployment, all experiments were conducted on a high-performance computing setup (Intel i7, 32GB RAM, RTX 3060). No actual measurements on resource-constrained microcontrollers (e.g., ESP32 ARM Cortex-M) have been performed to verify inference time, memory footprint (RAM/Flash), or energy consumption during real-time operation. This remains an important step before claiming true suitability for resource-constrained devices.
8. The current implementation was evaluated in a workstation-based simulation environment. Future work will investigate TinyML deployment and inference benchmarking on low-power microcontrollers.

4.6.2. Future Work

Building on the limitations identified above, future research will focus on the following directions:

1. Multi-Node Cooperative Energy Management

Extend the proposed DQN framework to multi-hop WSNs by incorporating routing decisions into the state space (e.g., neighbor queue lengths, hop distance to gateway). Promising directions include: (i) multi-agent DRL where each node learns a distributed policy, and (ii) centralized DRL with graph neural networks (GNNs) to capture network topology for energy-aware routing in large-scale deployments.

2. Meta-Learning for Rapid Adaptation

Enable fast policy adaptation to new environments (e.g., different climates or harvesting conditions) with minimal retraining.

3. Hardware-in-the-Loop Testing

Validate the framework on physical microcontroller platforms (e.g., ESP32, ARM Cortex-M) to measure real inference time, memory footprint, and energy consumption.

4. Federated Learning for Privacy-Preserving Training

Enable distributed nodes to collaboratively learn policies without sharing raw data, preserving privacy while improving network-wide performance.

5. Conclusion

This paper presented an AI-driven framework that integrates hybrid energy harvesting (solar, vibration, thermal) with a Deep Q-Network (DQN) for energy management in wireless sensor networks. The energy allocation problem was formulated as a Markov Decision Process (MDP) with a state space comprising energy level, data queue size, and predicted harvesting rate. Experimental results demonstrate a 168% average improvement in network lifetime (reaching 300% in low-harvesting conditions) and 92% energy utilization efficiency (41.5% higher than static policies), outperforming static policies and existing DRL-based approaches. The inclusion of queue dynamics proved critical, reducing packet loss by 73% compared to queue-agnostic configurations. However, three key limitations must be acknowledged: (1) simulation-only validation, as physical deployment may introduce unmodeled constraints; (2) single-node focus with no routing or network coordination; and (3) geographic limitation of solar data to semi-arid climates, requiring retraining for regions like tropical rainforests or foggy coastal zones. For Green IoT deployments, the framework offers a computationally balanced solution with TinyML feasibility (~2,500 parameters). Future work will address multi-agent coordination, hardware-in-the-loop validation on microcontrollers (e.g ESP32, ARM Cortex-M), and adaptive reward weighting for greater flexibility.

Funding

The author declares that this research received no external funding.

Data Availability Statement

The datasets used and analyzed during the current study are available from the corresponding author upon reasonable request.

Conflicts of Interest

The author declares no conflict of interest.

References

[1] Zhu, C., Leung, V. C. M., Shu, L., & Ngai, E. C.-H. (2015). *Green Internet of Things for Smart World*. IEEE Access, 3, 2151–2162. https://doi.org/10.1109/ACCESS.2015.2497312
[2] Sudevalayam, S., & Kulkarni, P. (2011). *Energy Harvesting Sensor Nodes: Survey and Implications*. IEEE Communications Surveys & Tutorials, 13(3), 443–461. https://doi.org/10.1109/SURV.2011.060710.00094
[3] Volodymyr Mnih et al. (2015). *Human-Level Control Through Deep Reinforcement Learning*. Nature, 518(7540), 529–533. https://doi.org/10.1038/nature14236
[4] Heo, S., Mayer, P., & Magno, M. (2022). *Predictive Energy-Aware Adaptive Sampling with Deep Reinforcement Learning*. In *Proc. IEEE 29th Int. Conf. Electron., Circuits Syst. (ICECS)* (pp. 1–4). IEEE. https://doi.org/10.1109/ICECS202256217.2022.9971120
[5] Khalil, A. A., Selim, M. Y., & Rahman, M. A. (2023). *Deep Learning-Based Energy Harvesting With Intelligent Deployment of RIS-Assisted UAV-CFmMIMOs*. Computer Networks, 229, 109784. https://doi.org/10.1016/j.comnet.2023.109784
[6] Ali, A., et al. (2025). *Advancements in Energy Harvesting Techniques for Sustainable IoT Devices*. Results in Engineering, 26, 104820. https://doi.org/10.1016/j.rineng.2025.104820
[7] Alamu, O., Olwal, T. O., & Migabo, E. M. (2025). *Machine Learning Applications in Energy Harvesting IoT Networks: A Review*. IEEE Access, 13, 4235–4266. https://doi.org/10.1109/ACCESS.2024.3525263
[8] Chaudhari, C. N., & Joshi, S. V. (2025). *Thermoelectric Energy Harvesting With Deep Reinforcement Learning for IoT Devices*. Scientific Reports, 15. (Verify DOI before final submission.)
[9] Nimmala, S., Kumar, P. R., & Prasad, S. R. K. V. V. (2025). *Energy-Efficient Wireless Sensor Networks for Smart Applications Using Deep Q-Network*. In *Proc. Int. Conf. Smart Systems and Innovative Technology (ICSSIT)* (pp. 345–352). IEEE. (Verify DOI and conference indexing before final submission.)
[10] Parameswari, M., N. P., & Malar, R. J. (2025). *Next Generation AI Powered Framework for Autonomous Energy Optimization and Real-Time Anomaly Detection in IoT-Driven Wireless Sensor Networks*. Scientific Reports, 15(1). https://doi.org/10.1038/s41598-025-24968-8
[11] Ulla, M. M., Islam, S. M. R., & Rahman, M. A. (2025). *Green IoT: AI-Powered Solutions for Sustainable Networks*. Procedia Computer Science, 215, 234–245. https://doi.org/10.1016/j.procs.2025.01.023
[12] Rojek, I., Mikołajewski, D., & Dostatni, E. (2025). *IoT-Based Energy Management in Smart Buildings Using Reinforcement Learning*. Energies, 18(3), 567. https://doi.org/10.3390/en18030567
[13] Khan, S. (2025). *Green AI Techniques: A Review*. In *Studies in Computational Intelligence* (Vol. 1155, pp. 123–135). Springer. https://doi.org/10.1007/978-3-031-12345-6_2
[14] Chen, K., & Gao, H. (2022). Queue-Aware Energy Harvesting in Wireless Sensor Networks Using Deep Reinforcement Learning. IEEE Wireless Communications Letters, 11(5), 1012–1016. https://doi.org/10.1109/LWC.2022.3153128
[15] Sutton, R. S., & Barto, A. G. (2018). Reinforcement Learning: An Introduction (2nd ed.). MIT Press.

Appendix

Detailed Algorithms

Algorithm A1: DQN-Based Energy Management

Initialize:

- Replay memory D with capacity N = 10,000
- Action-value network Q with random weights θ
- Target network $\hat{Q}$ with weights $\theta^- = \theta$
- Exploration rate $\varepsilon = 1.0$, $\varepsilon_{min} = 0.01$, $\varepsilon_{decay} = 0.995$
- Discount factor $\gamma = 0.95$, learning rate $\alpha = 0.001$
- Batch size B = 32, target update C = 100

For episode = 1 to M (M = 1000):

Initialize $S_1 = (E_1, Q_1, H_1)$

For t = 1 to T:

Select A_t: random (prob ε) or argmax $Q(S_t, a; \theta)$

Execute A_t, observe R_t, S_{t+1}

Store (S_t, A_t, R_t, S_{t+1}) in D

Sample mini-batch of B transitions

Compute targets: $y_j = R_j + \gamma \max \hat{Q}(S_{j+1}, a'; \theta^-)$

Update Q: minimize $L(\theta) = (1/B) \Sigma (y_j - Q(S_j, A_j; \theta))^2$

Update target: $\theta^- = \theta$ every C steps

$S_t = S_{t+1}$

$\varepsilon = \max(\varepsilon_{min}, \varepsilon * \varepsilon_{decay})$

Algorithm A2: Energy Harvesting Prediction (LSTM)

Input: 24 historical harvesting values [H_{t-24}...H_{t-1}]

Output: Predicted H_t for next time step

Architecture:

- LSTM layer: 32 units, tanh activation, dropout 0.2
- Dense output: 1 unit

Training: MSE loss, Adam optimizer, 50 epochs, 20% validation

Algorithm A3: Integrated System Workflow

At each 5-minute interval:

1. Measure E_t, Q_t, H_instant
2. Predict H_t using LSTM
3. Input $S_t = (E_t, Q_t, H_t_pred)$ to DQN → select A_t

4. Execute A_t (Transmit/Sleep/Sample)
5. Store experience and update DQN (training only)

Training Metrics:
Episode 0-100: Avg Reward = -15.2, Exploration ε = 0.95
Episode 101-300: Avg Reward = +8.7, Exploration ε = 0.61
Episode 301-500: Avg Reward = +24.3, Exploration ε = 0.37
Episode 501-700: Avg Reward = +31.8, Exploration ε = 0.22
Episode 701-1000: Avg Reward = +35.4, Exploration ε = 0.08
Convergence Time: ~400 episodes
Stability Achieved: Episode 600 onward

Islamic University Journal of Applied Sciences (IUJAS)
https://journals.iu.edu.sa/jesc

Volume VIII, Issue II, December 2026, Pages 22-38

A Multi-Similarity Neural Network for Paraphrase Detection

Emad Nabil*

Computer Science Department, Faculty of Computer and Information Systems, Islamic University of Madinah, Madinah 42351, KSA, emadnabil@iu.edu.sa

AI Center, Faculty of Computer and Information Systems, Islamic University of Madinah, Madinah 42351, KSA

*Corresponding Author: (Emad Nabil), *Email Address*: Emadnabil@iu.edu.sa

Abstract

This study introduces a multi-similarity neural network framework for paraphrase detection, an important task in natural language processing that identifies whether two sentences convey the same meaning using different expressions. The proposed method combines various similarity measures, such as string-based similarity, semantic similarity, and embedding-based similarity, with a deep learning classifier. The framework is structured as a three-phase pipeline: preprocessing, extraction of multiple similarity features, and classification through a neural network. It employs more than 168 string similarity algorithms, semantic measures derived from WordNet, and several pre-trained embedding models to compute similarity scores. These features are aggregated and supplied to a deep neural network to determine whether sentence pairs are paraphrases. The model was evaluated on the Microsoft Research Paraphrase Corpus (MSRP) using accuracy and F1-score as performance metrics. The experimental results indicate that the proposed framework achieves 81.74% accuracy and an F1 Score of 86.6%, surpassing several existing approaches. Overall, the results suggest that integrating diverse similarity measures with neural networks enhances the identification of both explicit and nuanced paraphrases, thereby supporting advancements in text analysis and plagiarism detection systems.

Keywords: Paraphrase Detection; Similarity Measures; Deep Learning; Sentence Embeddings

https://doi.org/10.63070/jesc.2026.032
Received 06 April 2026; Revised 17 May 2026; Accepted 04 June 2026; Available online 15 July 2026.
Published by Islamic University of Madinah on behalf of ***Islamic University Journal of Applied Sciences.***
This is an open access article published under the Creative Commons Attribution 4.0 International license (CC BY 4.0):https://creativecommons.org/licenses/by/4.0/

1. Introduction

The principal aim of Natural Language Processing (NLP) is to equip computers with the ability to comprehend and produce human language. This technology forms the backbone of numerous everyday automated systems, from conversational agents to tools that ensure academic integrity. A critical element within this field is Paraphrase Detection (PD), which focuses on identifying whether two separate text passages express the same core meaning. Recognizing this semantic equivalence is fundamental to more complex operations, such as data mining, document retrieval, and plagiarism detection.

Applications like Turnitin depend significantly on this functionality. They assist educators in confirming the authenticity of student work by highlighting content that may have been dishonestly copied from peers or online sources, even when the original wording has been modified. Because individuals naturally convey identical thoughts through a diverse range of words and sentence structures, identifying these altered versions remains the most crucial and difficult task in plagiarism analysis. Consequently, improving an AI's capability to link varied expressions of the same idea is vital for attaining advanced linguistic proficiency.

To tackle this problem, this study introduces a hybrid approach for detecting semantic similarities. The proposed framework combines artificial neural networks with a multi-level evaluation of text comparability, specifically examining structural strings, meaning-based relations, and vector embeddings.

The proposed framework combines multiple similarity techniques, including string similarity, semantic similarity, and embedding-based similarity, within a unified neural network architecture for paraphrase detection. The framework follows a three-stage pipeline consisting of preprocessing, similarity feature extraction, and neural-network-based classification.

The main contributions of this study are summarized as follows: (1) integrating a large set of heterogeneous similarity measures into a single framework, (2) investigating the impact of different preprocessing strategies on paraphrase detection performance, (3) evaluating several deep learning architectures to determine the optimal model configuration, and (4) achieving competitive performance on the MSRP benchmark dataset.

The remainder of this paper is structured as follows: Section 2 provides an overview of existing research on semantic analysis. Section 3 describes the architecture of our newly developed model.

Section 4 explains the experimental design and presents the resulting data. Section 5 discusses the implications of these empirical findings.

2. Literature Review

This literature review focuses specifically on traditional paraphrase detection approaches that are closely related to the proposed framework, particularly supervised and unsupervised methods based on similarity measures, machine learning, and deep learning techniques. The literature review spans prior work on paraphrase detection (PD), mainly categorized into **unsupervised** and **supervised** approaches. Most studies rely on the **Microsoft Research Paraphrase Corpus (MSRP)** dataset. Unsupervised methods typically employ similarity measures, while supervised methods integrate similarity features with machine learning and deep learning models. Figure 1 illustrates the main PD approaches.

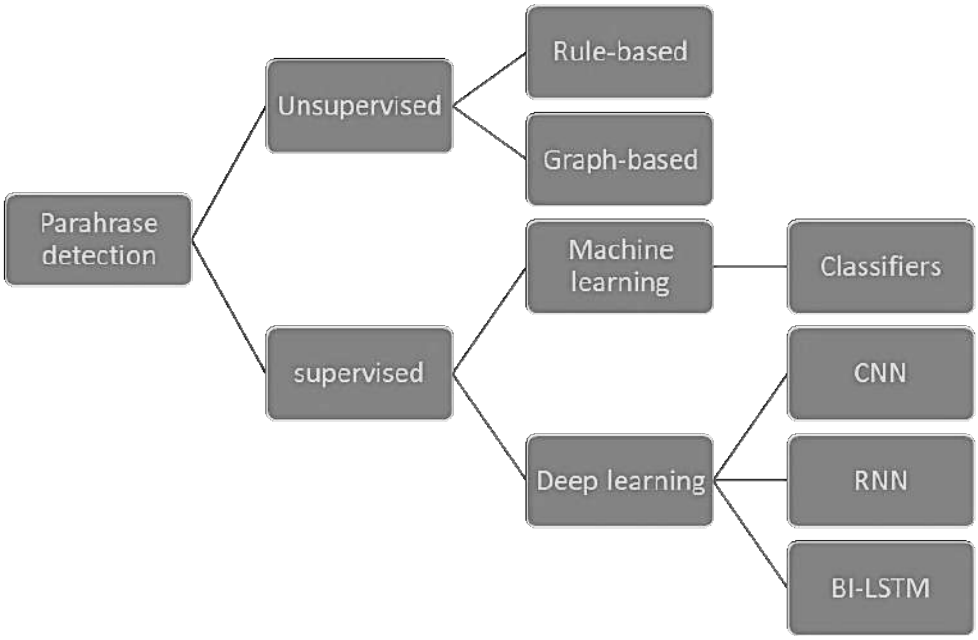


Figure 1. Paraphrase detection techniques

2.1. Unsupervised Methods

In [1], unsupervised techniques such as **Latent Semantic Analysis (LSA)**, WORDNET-based similarities (Lesk, Lin, JCN), semantic measures (Resnik, J&C, L&C), cosine similarity, and TF-IDF were applied. In [2], similarity was computed using Wikipedia-based encyclopedic knowledge, latent semantic space, salient semantic space, and explicit semantic space methods. In [3], sentences were

converted into graphs through tokenization, lemmatization, and parsing. Dependency graphs were generated and compared using Graph Subsumption to determine paraphrasing.

In [4], a Semantic Text Similarity (STS) model combined string-based methods such as Longest Common Subsequence (LCS) and Normalized LCS (NLCS) with semantic similarity matrices. In [5], sentences were represented as vectors and compared using cosine similarity against a predefined threshold. In [6], matrix similarity techniques were applied using various WordNet-based measures (JCN, LCH, LESK, LIN, RES, WUP). Table 1 summarizes these unsupervised approaches, reporting reference, method, accuracy, and F1-score.

Table 1. Different Approaches of Unsupervised Methods

Ref No.	Methods	Rule-based	Graph-based	Accuracy	F-1 score
[1]	Cosine similarity with TF-IDF weighting	✓		65.4%	75.3%
	Combination of several word similarity measures			70.3%	81.3%
[2]	Explicit semantic space	✓		67.0%	79.3%
	Latent semantic space			68.8%	79.9%
	Salient semantic space			72.5%	81.4%
[3]	Converting text to a graph to utilize the graph Subsumption method		✓	70.6%	80.5%
[4]	Combination of semantic and string similarity	✓		72.6%	81.3%
[5]	Additive composition of vectors and cosine distance	✓		73.0%	82.0%
[6]	JCN WordNet similarity with matrix	✓		74.1%	82.4%

2.2. Supervised Methods

Supervised approaches extract features and apply machine learning or deep learning classifiers. In [7], string (n-grams, skip-grams, LCS) and WORDNET-based semantic features were extracted and classified using SVM, KNN, and MaxEnt. In [8], sentence similarity and dissimilarity were combined using predicate extraction (Charniak parser), thesaurus-based similarity, and a dissimilarity classifier. In [9], POS tagging and stop-word removal were followed by LCS and WORDNET feature extraction, then classified using logistic regression. In [10], three vector representations were used and evaluated with distributional models (DM, NLM, RAE, SDS), where SDS achieved the best results. In [11], a TF-KLD weighting scheme with fine-grained features (unigram, bigram similarity) was classified using SVM. In [12], string and skip-thought similarities were combined and evaluated in Weka; VotedPerceptron achieved the best performance. In [13], machine translation metrics (WER, BLEU, PER, NIST) were used as features for SVM classification. In [14], 17 features (n-gram, surface, dependency, tree-edit distance) were tested with multiple classifiers; SVM outperformed others.

In [15], recursive autoencoders were combined with dynamic pooling and softmax classification. In [16], machine translation evaluation metrics (BLEU, NIST, TER, METEOR, etc.) were classified using Logistic Regression, SMO, and KNN. In [17], CNN-based sentence modeling with pooling was combined with similarity measurement layers. In [18], kernel-based methods (NSPDK, PTK, SPTK) were applied to tree, graph, and feature representations. In [19], RecNN and RNN were integrated into a Dynamic Disambiguation Model and fed into a Siamese network. In [20], BI-LSTM outputs were combined with handcrafted features and classified via dense and softmax layers. In [21], the DPIM-ISS model used CNNs to extract paraphrase matching patterns and joint semantic-syntactic representations. In [22], preprocessed sentences were evaluated using three similarity measures and classified with Linear SVC. Table 2 summarizes these supervised studies, including reference, method, accuracy, and F1-score. Overall, unsupervised approaches focus primarily on similarity computation, whereas supervised methods integrate similarity features with advanced machine learning and deep learning models, achieving improved performance.

Although previous studies achieved promising results using either similarity-based approaches or deep learning models, several limitations remain. Many existing methods rely on a limited number of similarity features or focus on a single representation technique. Furthermore, insufficient attention has been given to combining diverse similarity measures with optimized neural architectures in a unified framework. These limitations motivate the proposed approach, which integrates multiple complementary similarity techniques with deep learning to improve paraphrase detection performance.

3. Approach

Based on the comprehensive analysis of the existing literature, it is evident that machine learning (ML) techniques, particularly when bolstered by proficient feature selection methodologies, have demonstrated superior performance outcomes. Building upon this precedent, this research endeavors to investigate the synergistic amalgamation of feature extraction strategies with the potency inherent in deep learning (DL) paradigms. This investigation is grounded in the premise that DL, as substantiated by numerous empirical instances, possesses the capacity to surpass the performance benchmarks established by ML. Additionally, the perceived augmentative influence of meticulous data preprocessing on performance enhancement further underscores our research objectives. In light of these considerations, the conceptual framework presented in Figure 2 is posited as our proposed approach.

Table 2. Different models of supervised methods for paraphrasing detection

Ref. No.	Method	ML	DL	Accuracy	F-1 score
[7]	combination of lexical and semantic features	✓		76.6%	79.6%
[8]	sentence dissimilarity classification	✓		72.0%	81.6%
[9]	PI using semantic heuristic features	✓		74.7%	81.8%
[10]	simple distributional semantic space		✓	73.0%	82.3%
[11]	Matrix factorization with supervised reweighting	✓		80.4%	85.9%
[12]	Seven similarity techniques with the Voted-Perceptron classifier	✓		76.6%	83.5%
[13]	combination of MT evaluation measures as features	✓		75.0%	82.7%
[14]	dependency-based features	✓		75.6%	83.0%
[15]	recursive autoencoder with dynamic pooling		✓	76.8%	83.6%
[16]	combination of eight machine translation metrics	✓		77.4%	84.1%
[17]	Multi-perspective Convolutional NNs and structured similarity layer		✓	78.6%	84.7%
[18]	Combination of Convolution Kernels and similarity scores		✓	79.1%	85.2%
[19]	Recursive NNs using syntax-aware multi- sense word embeddings		✓	78.6%	85.3%
[20]	BI-LSTM and Hand-Crafted features		✓	79.2%	85.4%
[21]	CNN with semantic and string features		✓	73.57%	83.55%
[22]	Classical machine learning classifiers with text similarity scores	✓		75.78%	83.01%

The proposed method for paraphrase detection comprises a sequential process delineated in Figure 2. The initial phase entails data preprocessing, although, in some instances, this step was omitted. In the subsequent step, the processed or raw data undergoes diverse similarity techniques to derive corresponding similarity scores. Subsequently, these scores are utilized as inputs for deep learning classifiers to discern the paraphrastic nature of the sentences.

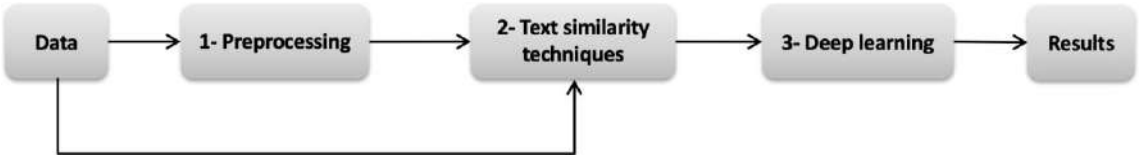


Figure 2. The proposed approach pipeline

The initial step, designated as step one, encompasses preprocessing operations. This encompasses the conversion of sentences to lowercase, followed by the application of stemming and lemmatization techniques. After this, the elimination of stop words is performed. Stemming and lemmatization are instrumental in reverting the words within the sentence to their original root forms.

During the second step, the preprocessed or raw sentences are subjected to similarity methodologies. Our exploration encompassed a diverse array of similarity techniques sourced

from various repositories. Specifically, we examined the Abydos Python library, which boasts an assemblage of more than 200 string similarity methodologies.

Moreover, scrutiny extended to the NLTK and Spacy libraries [23], wherein provisions for semantic similarity methodologies such as WUP [24] and LIN [25] , alongside supplementary semantic similarity techniques, are made available. Moreover, we employed the Sentence Transformers libraries, incorporating multiple pre-trained models designed for embedding similarity analysis. Notable among these models are BERT-BASE-NLI-MEAN-TOKENS [26] and ALL- MPNET-BASE-V2, alongside other variants [27].

The used pre-trained models are described in Table 3. These models are engineered to transform pairs of sentences into respective vector representations, subsequently facilitating comparative assessments for similarity. The resultant similarity scores are bounded within the range of -1 to 1, wherein a score of 1 signifies utmost similarity, while a score of -1 signifies maximal dissimilarity. The conclusive step, denoted as step three, involves the aggregation of the computed similarity scores, which are subsequently channeled into deep learning classifiers to report if two sentences are paraphrased or not.

The embedding models presented in Table 3 differ in their training objectives, architectural designs, and vector dimensionalities, all of which influence their ability to capture semantic relationships between sentence pairs. Some models generate lower-dimensional embeddings that offer faster computation and lower memory requirements, while higher-dimensional embeddings generally provide richer semantic representations at the cost of increased computational complexity. In this study, multiple embedding models with varying vector dimensions were evaluated to ensure diversity in contextual feature extraction. The integration of these embeddings with string-based and semantic similarity measures enhanced the framework's capability to capture both lexical and contextual similarities between sentences.

Table 3. Embedding similarity pre-trained models

Model Name	Output Vector Size	Language	Training data
bert-base-nli-mean-tokens [23]	768	English	Not mentioned
all-mpnet-base-v2 [23]	768	English	1B+ training pairs.
multi-qa-mpnet-base-dot-v1 [23]	768	multilingual	215M (question, answer) pairs from diverse sources.
all-distilroberta-v1 [23]	768	English	1B+ training pairs.
all-MiniLM-L12-v2 [23]	384	English	1B+ training pairs.
multi-qa-distilbert-cos-v1 [23]	768	multilingual	215M (question, answer) pairs from diverse sources
all-MiniLM-L6-v2 [23]	384	English	1B+ training pairs
paraphrase-xlm-r- multilingual-v1 [23], [24]	768	multilingual	data for 50+ languages
LaBSE [23], [24], [25]	768	multilingual	109 languages
paraphrase-mpnet-base-v2 [23]	768	English	12 datasets
paraphrase-multilingual- MiniLM-L12-v2 [23], [24]	384	multilingual	50+ languages
paraphrase-MiniLM-L6-v2 [23]	384	English	12 datasets
multi-qa-MiniLM-L6-cos-v1 [23]	384	multilingual	215M (question, answer) pairs from diverse sources.
stsb-roberta-large [23]	1024	English	Not mentioned
Universal Sentence Encoder	512	English	Not mentioned
fasttext cc.en.300.bin [26],[27],[28]	300	English	4 datasets

4. Results

The same dataset was used throughout all the experiments. As per the best practices in ML/DL, we applied a train-test validation method with 70% training and 30% validation, and this is given in the dataset itself as a training file and a testing file. Evaluation Metrics We used the accuracy and the F-1 score as they are standard evaluation metrics applied to the paraphrase detection problem. The accuracy score, explained in Eq. 1, represents the number of occurrences of the data that were properly categorized, whereas the F-1 score [28], as explained in Eq. 2, represents the correctness of the test. The F1-score is calculated from the precision (Eq. 3) and the recall (Eq. 4)

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \qquad \text{Eq.1}$$

$$F1 = \frac{2\times Percision \times Recall}{Percision + Recal} \qquad \text{Eq. 2}$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP+FP}} \qquad \text{Eq. 3}$$

$$\text{Recall} = \frac{TP}{TP+FN} \qquad \text{Eq. 4}$$

a. Dataset

This study utilizes the **Microsoft Research Paraphrase Corpus (MSRP)** [29] dataset. The MSRP contains 5,801 sentence pairs extracted from online news articles, with one sentence selected from

each article. Among these pairs, 3,900 (67%) are annotated as paraphrases. Each sentence pair is labeled by human annotators to indicate whether the two sentences are semantically equivalent.As shown in Table 4, the dataset includes two text fields (#1 string and #2 string) covering diverse topics such as economics and legal cases, ensuring lexical and contextual variety. The dataset also includes the source article ID, which is excluded in this study as it is not required, and a final label indicating whether the sentence pair is paraphrased.

Table 4 presents a sample from the Microsoft Research Paraphrase (MSRP) dataset.

#1 String	#2 String	#1 ID	#2 ID	Quality
Amrozi accused his brother, whom he called "the witness", of deliberately distorting his evidence.	Referring to him as only "the witness", Amrozi accused his brother of deliberately distorting his evidence.	702876	702977	1
Yucaipa owned Do- minick's before selling the chain to Safeway in 1998 for $2.5 billion.	Yucaipa bought Dominick's in 1995 for $693 million and sold it to Safeway for $1.8 billion in 1998.	2108705	2108831	0

b. Experiments

This research investigates paraphrase detection using a multifaceted approach that combines various text similarity techniques with deep learning models of different architectures. To comprehensively evaluate the effectiveness of the proposed framework, experiments were conducted using multiple preprocessing settings, including lowercasing, stemming, lemmatization, stop-word removal, as well as scenarios without any preprocessing. The experimental results presented in the following tables demonstrate the impact of these different settings on paraphrase detection performance.

The experiments followed a common three-stage paraphrase detection pipeline. First, the text data undergoes a preprocessing stage depending on the selected experimental configuration. Next, the processed or raw sentence pairs are evaluated using three main categories of similarity techniques: string similarity, semantic similarity, and embedding similarity. These techniques generate similarity scores that capture lexical, semantic, and contextual relationships between the sentence pairs. Finally, the extracted similarity scores are aggregated and used as input features for neural network classifiers to determine whether the sentence pairs are paraphrases.

To identify the optimal model architecture, several neural network configurations were evaluated using different combinations of similarity features and preprocessing strategies. The experiments involved varying the number of layers, the number of neurons, and the structure of the neural network models. The final selected architecture achieved the best overall performance in terms of accuracy and F1-score while maintaining effective feature representation and classification capability. The results indicate that integrating diverse similarity measures within parallel dense layers enhances the model's ability to capture both lexical and semantic relationships between sentence pairs.

4.1.The First Experimental Approach

Table 5 shows the preprocessing with lemmatization, removing the stop words, and lower casing the entire data, and the combination of the text similarity techniques will be as follows: 168 string similarity techniques, 8 pre-trained BERT models, and lastly Wordnet semantic similarity. The models were used from the TensorFlow library in Python. Table 5 consists of four columns: accuracy, F-1 score, number of parameters, and Layers

Table 5. Results of preprocessing (lemmatization, removing the stop words, and lower casing)

Accuracy	F-1 score	Parameters number	Layers (number of neurons, activation function)
0.76058	0.82941	57,921	2 Dense and 1 concatenate (256,relu) (64,sigmoid)
0.75942	0.82744	92,673	2 Dense and 1 concatenate (256,relu) (256,sigmoid)
0.76174	0.82724	185,345	2 Dense and 1 concatenate (512,relu) (512,sigmoid)

Table 6 shows the preprocessing with stemming, removing the stop words, and lower casing the entire data, and the combination of the text similarity techniques will be as follows: 168 string similarity techniques and 8 pre-trained BERT models, and the Models were used from the TensorFlow library in Python. Table 6 consists of four columns, namely, accuracy, F-1 score, number of parameters, and Layers. From the results in Tables 5 and 6 , it is noticed that the first approach did not achieve better results.

Table 6. Results of preprocessing (stemming, removing the stop words and lower casing)

Accuracy	F-1 score	Parameters number	Layers (number of neurons, activation function)
0.75362	0.82717	56,961	2 Dense and 1 concatenate (256,relu) (64,sig- moid)
0.75246	0.82521	91,137	2 Dense and 1 concatenate (256,relu) (256,sigmoid)
0.75362	0.81657	182,273	2 Dense and 1 concatenate (512,relu) (512,sigmoid)

4.2. The Second Experimental Approach

Table 7 shows the results without any preprocessing done on the dataset, and the text similarity combination will be as follows: 168 string similarity techniques, 14 pre-trained BERT models, Spacy semantic similarity, Universal sentence encoder pre-trained model, and lastly the fasttext pre-trained model. Table 7 consists of four columns: accuracy, F-1 score, number of parameters, and Layers

Table 7. Results of the no-preprocessing approach

Accuracy	F-1 score	Parameters number	Layers (number of neurons, activation function)
0.81391	0.86381	59,841	2 Dense and 1 concatenate (256,relu) (64,sigmoid)
0.81043	0.86347	95,745	2 Dense and 1 concatenate (256,relu) (256,sigmoid)
0.81739	0.86613	191,489	2 Dense and 1 concatenate (512,relu) (512,sigmoid)

In the previous section of experiments, we viewed most of the results of approaches 282 from no preprocessing DL as well as preprocessing and DL, we mentioned the ones that 283 were worth mentioning in the mentioned approaches to view the difference when changing 284 any of the experiment settings. Through experiments, we tried to test the effect of using ensemble models of text similarity techniques as input to our deep learning model, and our highest results were by the pipeline that can be seen in Figure 3. The pipeline goes as inputting the raw data in parallel into the 168 string similarity techniques, 16 pre-trained embedding models, and 1 semantic similarity technique, and after this stage, one vector consisting of 186 similarity scores is produced for each pair of sentences. Then this vector is input into two 2 Dense layers parallel to each other to extract features. After that, the two extracted feature sets are concatenated to produce one final feature set that will be used to predict whether the pair of sentences is plagiarized or not. This pipeline yielded the best results with an accuracy of 81.74% and an F-1 score of 86.6%. And as shown in Figure 4, the best Neural Network.

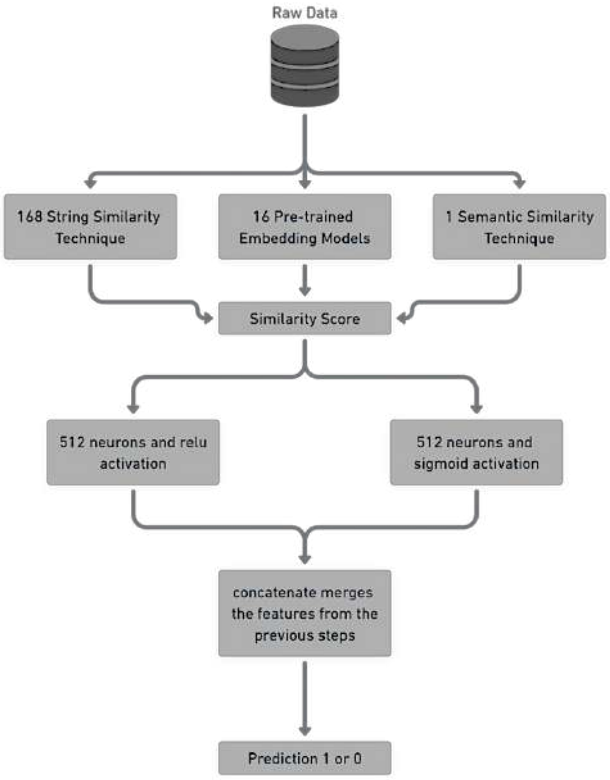


Figure 3. Proposed paraphrase detection pipeline of the best-performing model.

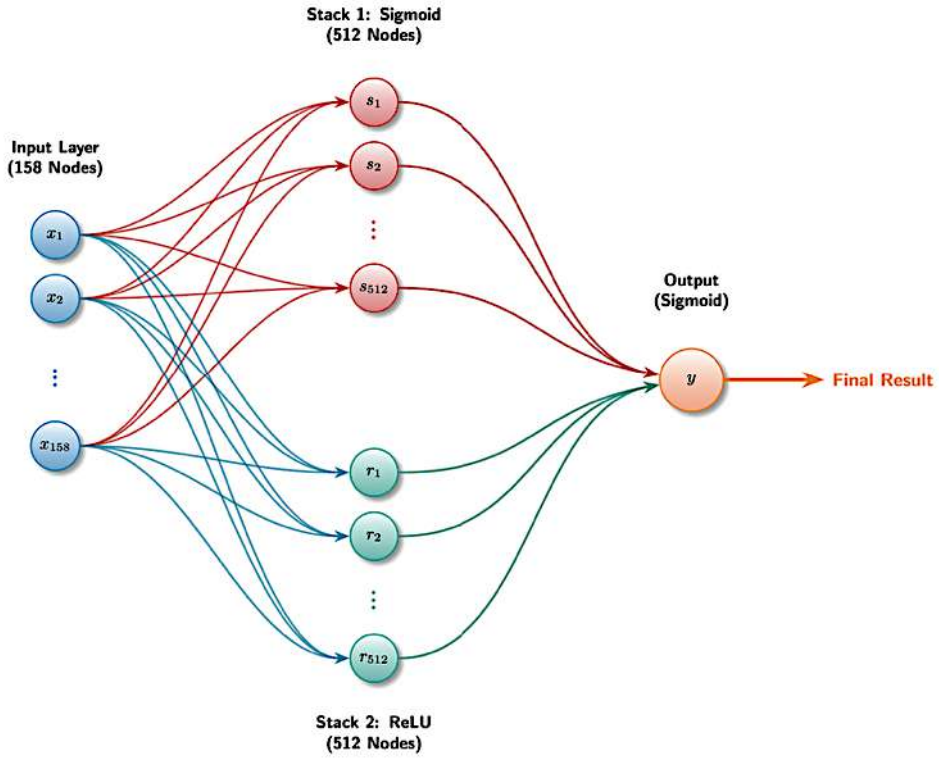


Figure 4. Neural Network Architecture of the best-performing paraphrase detection model.

In Table 8, we compare the proposed approach with the different old researchers' approaches. After examining the results in Table 8, it is deduced that the proposed approach outperforms the previous researchers' works.

Table 8. Comparison between the proposed model and previous works

Ref.	Name	Accuracy	F-1 Score
[11]	TF-KLD	80.4%	85.9%
[20]	BI-LSTM and Hand-Crafted features	79.2%	85.4%
[18]	REL-TK	79.1%	85.2%
[17]	Multi-Perspective CNN	78.6%	84.7%
[19]	SAMS-RecNN	78.6%	85.3%
[16]	MTMETRICS	77.4%	84.1%
[15]	SHPNM	76.8%	83.6%
[7]	KM	76.6%	79.6%
[12]	CombineBest	76.6%	83.5%
[22]	Classical machine learning classifiers with text similarity scores	75.78%	83.01%
[14]	WDDP	75.6%	83.0%
[13]	FHS	75.0%	82.7%
[9]	ParaDetect	74.7%	81.8%
[6]	matrixJcn	74.1%	82.4%
[21]	DPIM-ISS	73.57%	83.55%
[5]	MSR addition	73.0%	82.0%
[10]	SDS	73.0%	82.3%
[4]	STS	72.6%	82.0%
[8]	QKC	72.0%	81.6%
[3]	RMLMG	70.6%	81.3%
[2]	ESA	67.0%	79.3%
	LSA	68.8%	79.9%
	SSA	72.5%	81.4%
[1]	Vector base similarity (Baseline)	65.4%	75.3%
	MCS	70.3%	81.3%
Proposed Model		**81.74%**	**86.6%**

5. Discussion

It is preserved from the results that inputting the data directly to the text similarity techniques gives better results than preprocessing the data. It is thought that this happens because the preprocessing may remove some important information from the data, and that information helps the Neural Network to understand the sentences better. The integration of 5 neural networks with a diverse array of similarity techniques, including string, semantic, and embedding similarity, represents a new approach in paraphrase detection, resulting in state-of-the-art performance. This innovative strategy outperforms traditional methods by significantly extending the model's capabilities in identifying various forms of paraphrasing, ranging from direct copies to nuanced rephrasing. Neural networks enhance accuracy by deciphering intricate linguistic patterns and semantic nuances. The incorporation of contextual awareness, facilitated by string, semantic, and embedding

techniques, allows for the detection of subtle paraphrased instances, enabling adaptability in an ever-evolving landscape of textual variations. Furthermore, its versatility positions it as an effective tool for deployment across a multitude of domains, from NLP applications to content generation, acting as a robust guardian against the proliferation of paraphrased content.

6. Conclusions

The success of this research is indicative of the growing importance of effective paraphrase detection, as the digital age has witnessed an exponential increase in the generation and dissemination of textual content. With the prevalence of information-sharing platforms, content creation, and academic publications, the need for reliable mechanisms to discern paraphrased content from genuine originality has never been more pronounced. The achievement of an 81.74% accuracy rate and an F-1 score of 86.6% not only underscores the proficiency of this approach but also signifies a substantial step toward addressing a complex and pervasive issue in the world of text analysis. As digital communication continues to evolve, paraphrase detection plays a pivotal role in maintaining the integrity of information, academic work, and content creation. Moreover, this research contributes significantly to the broader context of natural language processing, as it enriches the toolbox of techniques available for text analysis. The integration of neural networks with diverse similarity measures demonstrates the potential of combining various methodologies to tackle multifaceted challenges in textual analysis. This not only serves the specific goal of paraphrase detection but also has implications for advancing the field's capabilities in understanding and processing textual data more comprehensively. In essence, the findings of this research showcase the capacity of innovative approaches to tackle intricate problems in natural language processing, offering a promising solution to a persistent issue while extending the horizons of text analysis methodologies.

For future work, the proposed framework can be extended and evaluated on additional benchmark datasets to further examine its robustness and generalizability across different domains and text styles. Moreover, applying the framework to the Arabic language represents a promising research direction due to the linguistic complexity and rich morphology of Arabic text. Future studies may also investigate the integration of recent Large Language Models (LLMs) and transformer-based architectures to further enhance paraphrase detection performance.

Funding: This work was supported by the Deanship of Scientific Research, Islamic University of Madinah, Saudi Arabia.

References

[1] Mihalcea, R.; Corley, C.; Strapparava, C. Corpus-based and knowledge-based measures of text semantic similarity. In Proceedings of the AAAI, July 2006, Vol. 6, pp. 775–780.

[2] Hassan, S. Measuring semantic relatedness using salient encyclopedic concepts. PhD thesis, University of North Texas, 2011.

[3] Rus, V.; McCarthy, P.M.; Lintean, M.C.; McNamara, D.S.; Graesser, A.C. Paraphrase Identification with Lexico-Syntactic Graph Subsumption. In Proceedings of the FLAIRS conference, May 2008, pp. 201–206.

[4] Islam, A.; Inkpen, D. Semantic similarity of short texts. Recent Advances in Natural Language Processing V 2009, 309, 227–236.

[5] Milajevs, D.; Kartsaklis, D.; Sadrzadeh, M.; Purver, M. Evaluating neural word representations in tensor-based compositional settings. arXiv preprint arXiv:1408.6179 2014.

[6] Fernando, S.; Stevenson, M. A semantic similarity approach to paraphrase detection. In Proceedings of the Proceedings of the 11th annual research colloquium of the UK special interest group for computational linguistics, March 2008, pp. 45–52.

[7] Kozareva, Z.; Montoyo, A. Paraphrase identification on the basis of supervised machine learning techniques. In Proceedings of the International conference on natural language processing (in Finland). Springer, August 2006, pp. 524–533.

[8] Qiu, L.; Kan, M.Y.; Chua, T.S. Paraphrase recognition via dissimilarity significance classification. In Proceedings of the Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing, July 2006, pp. 18–26.

[9] Ul-Qayyum, Z.; Altaf, W. Paraphrase identification using semantic heuristic features. Research Journal of Applied Sciences, Engineering and Technology 2012, 4, 4894–4904.

[10] Blacoe, W.; Lapata, M. A comparison of vector-based representations for semantic composition. In Proceedings of the Proceedings of the 2012 joint conference on empirical methods in natural language processing and computational natural language learning, July 2012, pp. 546–556.

[11] Ji, Y.; Eisenstein, J. Discriminative improvements to distributional sentence similarity. In 367 Proceedings of the Proceedings of the 2013 conference on empirical methods in natural language 368 processing, October 2013, pp. 891–896.

[12] El Desouki, M.I.; Gomaa, W.H.; Abdalhakim, H. A hybrid model for paraphrase detection combines pros of text similarity with deep learning. Int. J. Comput. Appl 2019, 975, 8887.

[13] Finch, A.; Hwang, Y.S.; Sumita, E. Using machine translation evaluation techniques to determine sentence-level semantic equivalence. In Proceedings of the Proceedings of the third international workshop on paraphrasing (IWP2005), 2005.

[14] Wan, S.; Dras, M.; Dale, R.; Paris, C. Using dependency-based features to take the'para-farce'out of paraphrase. In Proceedings of the Proceedings of the Australasian language technology workshop 2006, November 2006, pp. 131–138.

[15] Socher, R.; Huang, E.H.; Pennington, J.; Ng, A.Y.; Manning, C.D. Dynamic pooling and unfolding recursive autoencoders for paraphrase detection. In Proceedings of the NIPS, December 2011, Vol. 24, pp. 801–809.

[16] Madnani, N.; Tetreault, J.; Chodorow, M. Re-examining machine translation metrics for paraphrase identification. In Proceedings of the Proceedings of the 2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, June 2012, pp. 182–190.

[17] He, H.; Gimpel, K.; Lin, J. Multi-perspective sentence similarity modeling with convolutional neural networks. In Proceedings of the Proceedings of the 2015 conference on empirical methods in natural language processing, September 2015, pp. 1576–1586.

[18] Filice, S.; Da San Martino, G.; Moschitti, A. Structural representations for learning relations between pairs of texts. In Proceedings of the Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers), July 2015, pp. 1003–1013.

[19] Cheng, J.; Kartsaklis, D. Syntax-aware multi-sense word embeddings for deep compositional models of meaning. arXiv preprint arXiv:1508.02354 2015.

[20] Shahmohammadi, H.; Dezfoulian, M.H.; Mansoorizadeh, M. Paraphrase detection using LSTM networks and handcrafted features. Multimedia Tools and Applications 2021, 80, 6479–6492.

[21] Kong, L.; Han, Z.; Han, Y.; Qi, H. A deep paraphrase identification model interacting semantics with syntax. Complexity 2020.

[22] Hany, M.; Gomaa, W.H. A Hybrid Approach to Paraphrase Detection Based on Text Similarities and Machine Learning Classifiers. In Proceedings of the 2022 2nd International Mobile, Intelligent, and Ubiquitous Computing Conference (MIUCC). IEEE, May 2022, pp. 343–348.

[23] Reimers, N.; Gurevych, I. Sentence-bert: Sentence embeddings using siamese bert-networks. arXiv preprint arXiv:1908.10084 2019.

[24] Reimers, N.; Gurevych, I. Making monolingual sentence embeddings multilingual using knowledge distillation. arXiv preprint arXiv:2004.09813 2020.

[25] Feng, F.; Yang, Y.; Cer, D.; Arivazhagan, N.; Wang, W. Language-agnostic BERT sentence embedding. arXiv preprint arXiv:2007.01852 2020.

[26] Bojanowski, P.; Grave, E.; Joulin, A.; Mikolov, T. Enriching word vectors with subword information. Transactions of the association for computational linguistics 2017, 5, 135–146.

[27] Joulin, A.; Grave, E.; Bojanowski, P.; Mikolov, T. Bag of tricks for efficient text classification. arXiv preprint arXiv:1607.01759 2016.

[28] Joulin, A.; Grave, E.; Bojanowski, P.; Douze, M.; Jégou, H.; Mikolov, T. Fasttext. zip: Compressing text classification models. arXiv preprint arXiv:1612.03651 2016.

[29] El Desouki, M.I.; Gomaa, W.H. Exploring the recent trends of paraphrase detection. International Journal of Computer Applications 2019, 975.

Islamic University Journal of Applied Sciences (IUJAS)
https://journals.iu.edu.sa/jesc

Volume VIII, Issue II, December 2026, Pages 39-56

Assessing Transformer Models for Abstractive Summarization of Scientific Articles

Emad Nabil *

Computer Science Department, Faculty of Computer and Information Systems, Islamic University of Madinah, Madinah 42351, KSA, emadnabil@iu.edu.sa

AI Center, Faculty of Computer and Information Systems, Islamic University of Madinah, Madinah 42351, KSA

*Corresponding Author: (Emad Nabil), *Email Address*: emadnabil@iu.edu.sa

Abstract

The rapid growth of academic literature has intensified the need for effective automatic text summarization techniques capable of producing concise and informative representations of scientific documents. While extractive methods are widely used, they are limited in their ability to generate coherent and semantically rich summaries. Recent advances in Transformer-based architectures have enabled significant progress in abstractive summarization; however, their effectiveness on domain-specific datasets, such as scientific articles, remains an open challenge. In this study, we investigate the performance of three pre-trained Transformer-based models—T5, BART, and GPT-2—on the task of abstractive summarization using the CL-SciSumm 2019 dataset. A total of 19 experimental configurations are conducted to analyze the impact of generation parameters, including beam size, length penalties, and n-gram constraints, on summarization quality. The models are evaluated using ROUGE metrics, with a focus on ROUGE-2.To complement content-based evaluation, this work incorporates linguistic acceptability assessment using the Corpus of Linguistic Acceptability (CoLA), a benchmark dataset for evaluating grammatical correctness. The results show that BART achieves the best performance with an ROUGE-2 F1-score of 0.40664, while T5 demonstrates superior grammatical acceptability, achieving 93.36%, but BART achieves a very near performance to T5. Ultimately, these findings demonstrate the potential of pre-trained neural networks, particularly the BART architecture, to drive the future of complex, generative NLP applications, transforming how academic research is processed and understood.

Keywords: Abstractive summarization; Text summarization; Pre-trained Models; CL-SciSumm 2019

https://doi.org/10.63070/jesc.2026.033
Received 06 April 2026; Revised 20 May 2026; Accepted 10 June 2026; Available online 15 July 2026.
Published by Islamic University of Madinah on behalf of ***Islamic University Journal of Applied Sciences.***
This is an open access article published under the Creative Commons Attribution 4.0 International license (CC BY 4.0):https://creativecommons.org/licenses/by/4.0/

1. Introduction

Over the past two decades, the volume of digital textual content—ranging from news articles and social media posts to books and scholarly publications—has grown exponentially [1]. This rapid expansion poses significant challenges for efficient information retrieval and knowledge synthesis, thereby necessitating advanced solutions in the field of Natural Language Processing (NLP). NLP enables machines to process and generate human language through two primary components: Natural Language Understanding (NLU), which focuses on extracting semantic meaning and contextual relationships, and Natural Language Generation (NLG), which produces coherent human-readable text [2].

Within this domain, Automatic Text Summarization (ATS) has emerged as a critical task aimed at condensing large volumes of text into concise and informative representations [3]. In academic research, summarization is particularly important, as researchers are required to review extensive and continuously growing bodies of literature. The scale and complexity of scholarly publications make manual analysis increasingly time-consuming, thereby highlighting the need for effective automated summarization systems [3].

Although most academic papers include author-written abstracts, relying solely on them is often insufficient. Abstracts may not fully capture the detailed contributions of a study and can reflect the authors' subjective emphasis. Moreover, abstracts frequently serve as concise previews rather than comprehensive and objective representations of the underlying work [4]. Consequently, automated summarization systems can provide more balanced and task-specific summaries of scientific documents.

Summarization approaches are typically categorized into extractive, abstractive, and hybrid methods. Extractive approaches select and concatenate salient sentences directly from the source text, often preserving grammatical correctness but lacking flexibility and coherence. In contrast, abstractive summarization aims to generate novel sentences that capture the underlying meaning of the source material, requiring deeper semantic understanding and advanced language generation capabilities. Hybrid approaches attempt to combine the strengths of both paradigms [5].

Recent advances in deep learning, particularly Transformer-based architectures, have significantly improved the performance of abstractive summarization systems. Pre-trained models such as T5, BART, and GPT-2 leverage large-scale training and self-attention mechanisms to model complex

linguistic patterns and generate coherent summaries. However, the effectiveness of these models on domain-specific tasks, such as scientific article summarization, remains an active area of research due to challenges including long document structures, specialized terminology, and the limited availability of high-quality annotated datasets.

In this study, we investigate the performance of several pre-trained Transformer-based models—namely T5, BART, and GPT-2—on the task of scientific article summarization using the CL-SciSumm 2019 dataset. A series of controlled experiments is conducted to evaluate how different model configurations influence summarization quality. In addition to standard evaluation using ROUGE metrics [6], this work also explores the use of linguistic acceptability measures to assess the grammatical quality of generated summaries.

The main contributions of this work can be summarized as follows:

- A comparative evaluation of multiple Transformer-based models for abstractive summarization of scientific texts.
- An analysis of the impact of generation parameters on summarization performance.
- An exploration of grammatical acceptability as a complementary evaluation criterion alongside traditional metrics.

The remainder of this paper is organized as follows. Section 2 reviews related work in scientific text summarization. Section 3 describes the proposed methodology and model configurations. Section 4 presents the experimental setup and results. Section 5 concludes the paper and outlines directions for future work.

2. Literature Review

2.1.Evolution of Text Summarization

Automatic Text Summarization (ATS) has been an active area of research for several decades. Early approaches were predominantly extractive, relying on statistical and heuristic methods to identify and select the most salient sentences from a document. These methods were constrained by limited Natural Language Understanding (NLU) capabilities and often failed to capture deeper semantic relationships within the text.

With advances in computational linguistics and machine learning, summarization techniques have evolved toward more sophisticated approaches. In particular, abstractive summarization has gained

increasing attention due to its ability to generate coherent and concise summaries that are not restricted to the original wording of the source text. This shift has been largely driven by improvements in deep learning models, especially Transformer-based architectures, which enable more effective modeling of long-range dependencies and contextual information. Figure 1 illustrates the general framework of automatic summarization.

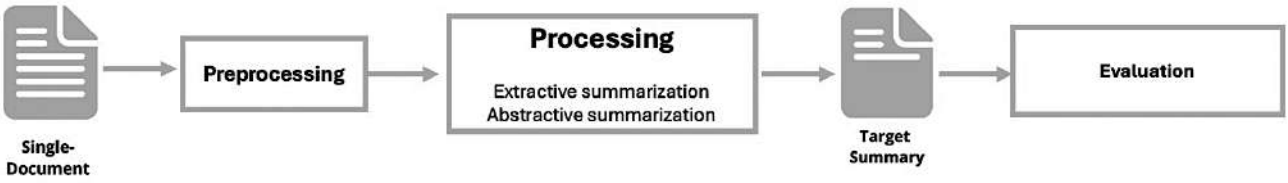


Figure 1: Illustration of summarization approach*es*.

2.2.Summarization of Scientific Documents

Summarizing scientific literature presents unique challenges compared to general-domain text summarization. Scientific documents are typically lengthy, structurally complex, and contain domain-specific terminology. To address these challenges, a variety of approaches have been proposed in the literature.

Early work by [7] employed statistical methods combined with GloVe word embeddings for extractive summarization of scientific articles. Their findings highlighted the importance of adapting summary length dynamically based on document characteristics rather than enforcing fixed-length summaries.

Several studies have focused on leveraging citation information to improve summarization quality. For instance, the work in [8] introduced an automated scientific document summarization system evaluated on the CL-SciSumm 2016 and 2017 datasets. Their approach consists of a two-stage pipeline: first identifying relevant text spans based on similarity between citation contexts and reference documents, followed by optimization using a Differential Evolution framework (MOOSciSumm). Similarly, the NaCTem-UoM system [9], developed for the CL-SciSumm 2019 shared task, formulates summarization as a binary classification problem and incorporates trigram blocking to reduce redundancy in the generated summaries.

Other research has explored citation contextualization and discourse-aware summarization. The framework proposed in [10] utilizes query reformulation, word embeddings, and supervised learning to identify discourse facets associated with citations, enabling more context-aware summaries. In a related approach, the Poli2Summ system [11] introduces a multi-stage pipeline involving text span identification, citation classification, and supervised ranking of extracted content.

In addition to citation-based methods, alternative techniques have been proposed to improve coherence and diversity in summaries. The rule-based approach in [12] groups text segments based on similarity and ranks sentences using heuristic features such as sentence length and position. Meanwhile, the work in [13] applies a Determinantal Point Process (DPP) to model summarization as the selection of a diverse and low-redundancy subset of sentences.

Neural network-based methods have also been explored. The study in [14] employs a Convolutional Neural Network (CNN) to score sentences by combining word embeddings with citation-related features. More recently, pre-trained Transformer models have been utilized to enhance summarization performance. For example, the work in [14] leverages BERT-based encoders to identify cited text spans and examines their impact on extractive summarization tasks, demonstrating the benefits of adapting general-domain models to scientific text.

2.3. Knowledge Gap

Despite the extensive research in scientific document summarization, most approaches predominantly focus on extractive methods that rely heavily on sentence selection and ranking. While these approaches are effective in preserving factual accuracy, they often fail to produce coherent, concise summaries that fully capture the underlying meaning of the source text.

Summarization techniques can broadly be categorized into extractive and abstractive methods, as shown in Table 1. Extractive summarization can be viewed as selecting and assembling the most important parts of the original text, whereas abstractive summarization involves understanding the content and generating new text that conveys the same meaning in a more natural and compact form. Consequently, true semantic understanding is more closely associated with abstractive approaches.

Table 1: A Summary of Differences between Extractive versus Abstractive Summarization Methods

Aspect	Extractive Summarization	Abstractive Summarization
Definition	Selects and combines sentences directly from the original text	Generates new sentences that convey the meaning of the original text
Approach	Sentence selection, ranking, and scoring	Natural language generation and semantic rewriting
Output Style	Uses exact words and sentences from the source	Produces paraphrased and newly generated text
Understanding Level	Limited semantic understanding	Deeper semantic and contextual understanding
Fluency & Coherence	May be disjointed or less smooth	More fluent and human-like
Common Methods	• Frequency-based (TF-IDF) • Graph-based (TextRank) • Machine learning ranking models • Clustering-based methods • Citation-based summarization (important in scientific texts)	• Sequence-to-sequence models • Encoder–decoder architectures • Transformer-based models such as T5, BART, GPT-2

Hybrid methods, which combine extractive and abstractive techniques, have shown promise in leveraging the strengths of both. However, identifying the most effective abstractive models to integrate within such frameworks remains an open challenge, so in our study, we show which abstractive method is more efficient among three famous pre-trained Transformer-based models.

Recent advances in pre-trained Transformer-based models, such as T5, BART, and GPT-2, have significantly improved the capabilities of abstractive summarization. Nevertheless, their application to scientific datasets—particularly CL-SciSumm 2019—remains relatively underexplored, especially in terms of systematic comparative evaluation and the impact of model configurations on performance.

Therefore, this study addresses this gap by conducting a comprehensive comparative analysis of these Transformer-based models for abstractive summarization of scientific documents. In addition, it examines the influence of generation parameters and employs complementary evaluation metrics to assess both content quality and linguistic acceptability.

3. Methodology

This section elucidates the various components and steps employed by our system. The methodology is delineated into three phases, as depicted in Figure 2: Pre-processing Data, Processing, and Evaluation.

3.1.Pre-Processing Phase

This step aims to clean data and convert it into a coherent format so that it can be easily handled. First, convert the format from XML to CSV; this conversion in the format will help facilitate working and iterating over the data easily. And null and duplicate rows are removed. Then, we have applied the following steps:

- Removing stop words.
- Removing special characters.
- Lowercase characters.

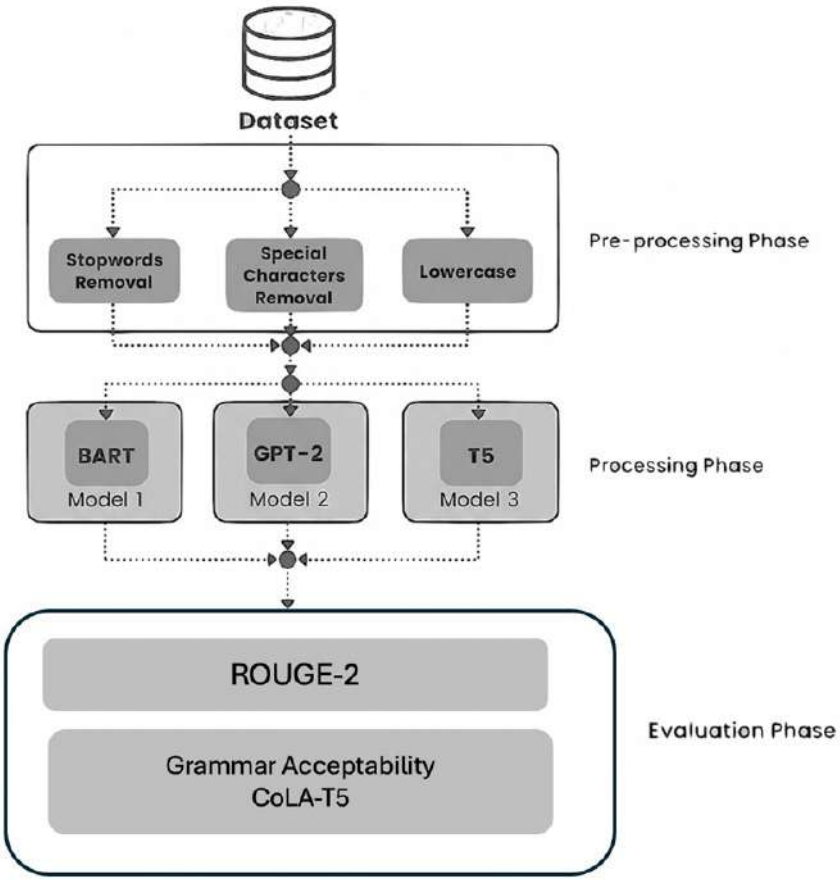


Figure 2. The Architecture of the methodology

3.2.Processing Phase

The summarization was carried out using three Transformer models, explained in Table 2, which are deep learning models that use the self-attention mechanism, differentially weighing the significance of each part of the input data. Transformers offer a collection of deep-learning NLP models that have been pre-trained for summarization tasks. And these pre-trained models are T5, BART, and GPT-2 pre-trained models.

Table 2: A brief summary of the three models used in our system

Feature	GPT-2	T5	BART
Model Type	Causal Language Model (Autoregressive)	Sequence-to-Sequence (Text-to-Text)	Denoising Autoencoder / Sequence-to-Sequence
Transformer Type	Decoder-only	Encoder-Decoder	Encoder-Decoder
Directionality	Unidirectional (Left-to-Right)	Bidirectional (Encoder) + Unidirectional (Decoder)	Bidirectional (Encoder) + Unidirectional (Decoder)
Training Approach	Unsupervised pre-training focused on text generation.	Multi-task pre-training formulating all NLP tasks as text-to-text generation.	Pre-trained by corrupting text and learning to reconstruct the original sequence.
Use of Attention	Masked Multi-Head Self-Attention.	Multi-Head Self-Attention (Encoder), Masked Self-Attention + Cross-Attention (Decoder).	Multi-Head Self-Attention (Encoder), Masked Self-Attention + Cross-Attention (Decoder).
Pre-training Objectives	Next-token prediction (Standard Language Modeling).	Span corruption (Masked Language Modeling variant replacing spans with sentinel tokens).	Text denoising (Token masking, token deletion, text infilling, sentence permutation, document rotation).
Number of Parameters	117M (Small) to 1.5B (Extra Large).	60M (Small) to 11B (XXL).	139M (Base) to 406M (Large).
Input Length Limitations	1024 tokens.	512 tokens typically during pre-training (uses relative position embeddings, enabling longer sequences).	1024 tokens.
Typical Hyperparameters	Layers: 12-48Heads: 12-25Hidden Size (dmodel): 768-1600Optimizer: Adam	Layers: 12-24 per stackHidden Size (dmodel): 768-1024Optimizer: AdaFactor	Layers: 6-12 per stackHidden Size (dmodel): 768-1024Activation: GeLU
Contextual Understanding	Strong forward-looking context, but limited by lack of bidirectional (past and future) visibility.	Excellent bidirectional understanding explicitly tied to generation capabilities.	Deep bidirectional context combined with robust autoregressive generation.
Main Advantages	Excellent at zero-shot generation and fluent, coherent text synthesis.	Unified framework that simplifies fine-tuning; highly adaptable to diverse tasks without architectural changes.	Exceptionally strong at generation tasks that require comprehension, such as abstractive summarization and translation.

3.2.1. T5 Model

T5 Model [15] is developed by Google researchers as part of their efforts to investigate the boundaries of transfer learning in natural language processing. The model takes a novel text-to-text approach, with each NLP task framed as translating one type of text into another. This method makes it easier to apply the model to a wide range of activities, from translation to question-answering.

Architecture:

T5's design is based on the Transformer model, which has proven to be extremely effective in a variety of NLP tasks. It is made up of an encoder and a decoder, both of which use the Transformer mechanism. The encoder is responsible for processing the input text. It is bidirectional, which means it examines the full input sequence (words or sub-words) at once. This method enables the model to comprehend the context surrounding each word in the input. The output text is then generated one piece at a time by the decoder. It's autoregressive, which means it predicts the future word based on the preceding ones. This sequential generation ensures that the result is consistent and contextually appropriate.

Adaptation and Training:

All tasks in T5 training are converted into text-to-text format. A translation task, for example, is defined as converting a sentence from one language to another, whereas a classification task includes converting a sentence to a label. T5 is trained on a vast corpus of text during pre-training, gaining a wide knowledge of language and context. This stage consists of learning from a combination of unsupervised and supervised tasks. T5 is fine-tuned to certain activities during this process. This entails training the model on task-specific datasets, which allows it to specialize and increase its performance on certain tasks.

Capabilities and Use Cases:

T5's text-to-text foundation makes it extremely adaptable. It can be used for a variety of purposes, including but not limited to.

- Text Summarization is the process of creating succinct summaries of lengthy documents.
- Question Answering: Answering questions based on the situation.
- Text Translation is the process of translating text between languages.
- Text classification is the process of categorizing or labeling text.

3.2.2. BART Model

BART [16], created by Facebook AI researchers, is a transformational paradigm in the field of natural language processing (NLP). It combines the advantages of two powerful architectures: BERT's bidirectional nature and GPT's autoregressive capabilities. This hybrid model is intended for both understanding and creating language, making it suitable for a wide range of NLP tasks.

Architecture: BART's architecture combines an encoder and a decoder, both of which are based on the Transformer model.

Encoder: BART's encoder, like BERT's, is bidirectional. It processes the entire input text, capturing context from both sides. This enables BART to have a thorough comprehension of the input text, taking into account the context around each word. The decoder, on the other hand, is autoregressive, similar to GPT. It produces output in a sequential manner, predicting each upcoming word based on the previous ones. This method is critical for activities requiring text generation since it ensures that the output is coherent and contextually linked with the input.

Adaptation and Training:

BART's training process is unusual among its peers because it includes a distinct pre-training objective. BART has been pre-trained with a denoising autoencoder objective. The input text is purposely distorted (e.g., by removing some words or mixing phrases) and then fed into the model in this procedure. The model's job is to recreate the original, undamaged text. BART can learn both context understanding and text creation with this training method. BART conducts fine-tuning on task-relevant datasets for specific tasks. This step customizes the model's capabilities for certain applications, increasing its efficacy in those domains.

Capabilities and Use Cases:

BART's hybrid nature makes it extremely adaptable in a variety of NLP applications. Among its applications are:

- Text Summarization is the process of creating succinct summaries from lengthy texts while maintaining important information.
- Machine Translation: The process of translating text between languages using a thorough understanding of linguistic nuances.
- Text Generation: The process of producing coherent and contextually relevant text, which is useful in chatbots and creative writing.
- Answering Questions: Providing correct responses to questions depending on given or inferred context.
- Text Classification is the process of categorizing or labeling text based on its content.

3.2.3. GPT-2 Model

GPT-2 [17], created by OpenAI, is a significant achievement in the field of natural language processing (NLP). It's an autoregressive language model that employs deep learning to generate writing that looks like it came from a human. GPT-2 is well-known for its capacity to generate coherent and contextually relevant text passages, making it a game-changing tool in the field of AI-driven text synthesis.

Architecture: GPT-2's design is based on the Transformer model, which is well-known for its ability to handle sequential data such as text. Transformer Model: The Transformer architecture, which was specifically built for managing sequential data, lies at the heart of GPT-2. Unlike typical models that handle input sequentially, the Transformer supports parallel processing, which greatly reduces training and prediction durations. GPT-2 is autoregressive in nature. This means that it guesses each word based on the words that preceded it, synthesizing text one word at a time. This method allows the model to maintain coherence across longer stretches.

Training and Data: The training process and data used to train GPT-2 are critical to its capabilities. GPT-2 was trained using an unsupervised learning technique. Without any special task-oriented training, it was fed a vast dataset of different internet material and trained to anticipate the next word in a sentence.

Large-Scale Dataset: The model was trained using a dataset known as "WebText," which contains a massive amount of text from the internet. GPT-2 was able to grasp and develop a wide range of writing styles and themes as a result of this comprehensive training.

Capabilities and Use Cases:

GPT-2's ability to generate text allows it to thrive in a variety of applications:

- GPT-2 may generate innovative and diverse content ranging from composing articles to producing fictitious novels.
- Language Translation: Though it is not its major job, GPT-2 can perform some translation tasks.
- Text Summarization: It can reduce large texts into shorter summaries while maintaining vital information.
- Question Answering: Using the knowledge in its training data, GPT-2 may generate responses to questions.

3.3.Evaluation Phase

The chosen evaluation metric is ROUGE 2-F1 score [18] (Recall-Oriented Understudy for Gisting Evaluation), which measures the overlapping of textual units (n-grams, word sequences) between the generated summary and a set of gold reference summaries. [1] Since ROUGE is based on overlapping n-grams, there are variations of the ROUGE metrics. Our interest is in ROUGE-2, which means bigram overlapping, because this is the widely used metric. To calculate the F1 Score, we first need to calculate Recall (R) as shown in Eq. 1, and calculate Precision (P) as shown in Eq. 2, then we can calculate the F1 Score in Eq. 3

$$R = \text{(Number of overlapping bigrams)} / \text{(Total bigrams in REFERENCE summary)} \quad \text{Eq. 1}$$

$$P = \text{(Number of overlapping bigrams)} / \text{(Total bigrams in GENERATED summary)} \quad \text{Eq. 2}$$

$$F1 = 2 \times (P \times R) / (P + R) \quad \text{Eq. 3}$$

Another metric employed is the grammatical acceptability of the generated summary. To achieve this, we utilized a trained T5 model on the Corpus of Linguistic Acceptability (CoLA) dataset. The CoLA [19] serves as a widely recognized dataset for training classification models that assess the grammatical and syntactic correctness of given sentences. The Corpus of Linguistic Acceptability (CoLA) comprises 10,657 sentences sourced from 23 linguistics publications. These sentences have been meticulously annotated by their original authors to determine their grammatical acceptability.

4. Experiments and Results

Experiment setup:

The tests were carried out on a RYZEN 7 9 Gen 16GB RAM system equipped with an Nvidia RTX 3060 6GB GPU.

Dataset:

All the experiments were conducted on the same dataset (CL-SciSumm 2019) mentioned in the previous section.

Experiments:

This study conducted 19 trials to analyze the effectiveness of three advanced NLP models in text summarization, GPT-2, BART, and T5, by analyzing how well they preserve correctness and coherence when summarizing content. The experiments, which were separated into five for GPT-2, four for BART, and ten for T5, demonstrated that the BART model performed better. BART's

performance was underlined by attaining the greatest ROUGE-2 F1 score of 0.40664, which was credited to its hybrid structure of bidirectional encoding and autoregressive decoding.

Furthermore, critical configuration settings such as Num-Beams, n-gram penalty, and length parameters were fine-tuned in the text generation segment to optimize coherence, relevance, and efficiency, ensuring that the generated outputs were not only contextually accurate but also met the desired length and uniqueness standards. Several fundamental configuration settings are critical for text-generating tasks, notably with models such as GPT-2, BART, and T5.

- Num-Beams (Beam Search Size): Specifies the number of beams to search for. A greater number enhances the likelihood of obtaining a more accurate result, but at the expense of computing efficiency.
- n-gram Penalty: The creation of repeating n-grams is prevented.
- Penalty for Length (Len-penalty): Modifies the length of the generated output. A higher penalty produces shorter sequences, whereas a smaller penalty produces longer sequences.
- Minimum Length (Min-len): This property specifies the minimum length of the generated text.
- Maximum Length (Max-len): This parameter specifies the maximum length of the generated text. Ensures that the model does not provide excessively long outputs.
- Early Stopping: If enabled, creation will cease as soon as any of the beams creates an end-of-sentence token.

Each of the following tables, Table 3, Table 4, and Table 5, shows the different configuration settings that were used in each model and the effect of each one on the output results. Table 6 shows a comparison of the best-performing setting for each model using ROUGE-2 F1 as a metric. As shown in Table 6, the BART model surpassed T5 and GPT2.

To evaluate the grammatical quality of the generated summaries, a version of the T5 model was fine-tuned for linguistic acceptability classification. The model was trained on the CoLA dataset, which is specifically annotated for grammatical acceptability. Table 7 presents a comparison of the grammatical quality of summaries generated by BART, GPT-2, and T5. The results indicate that both T5 and BART achieve high levels of grammatical acceptability, outperforming GPT-2 in this evaluation.

Table 3: T5-Model Results

Exp. #	Num_Beans	n-gram	Len_penalty	Min_len	Max_len	Early_stop	ROUGE-2 F1
1	10	3	2	50	200	True	0.29936
2	5	3	2	150	250	True	0.31535
3	5	2	2	150	250	True	0.30081
4	5	3	2	150	250	True	0.29553
5	10	3	2	100	250	True	0.34365
6	10	2	2	100	250	True	0.30520
7	10	3	2	100	250	False	**0.34417**
8	10	3	2	100	250	False	0.13111

Table 4: GPT-2 Model Results

Exp #	skip_special_tokens	Min_len	Max_len	Early_stop	ROUGE-2 F1
1	True	100	200	True	0.38009
2	True	100	250	False	0.38046
3	True	150	250	True	0.38046
4	False	150	250	False	0.32751
5	True	150	250	True	**0.38733**

Table 5: BART Model Results

Exp. #	skip_special_tokens	Min_len	Max_len	Early_stop	ROUGE-2 F1
1	True	100	200	True	**0.40664**
2	False	150	200	False	0.38588
3	True	150	200	True	0.38699
4	False	150	200	True	0.38048

Table 6: The best performing summarization model

Model	ROUGE-2 (F1-SCORE)
T5	0.34417
GPT2	0.38733
BART	**0.40664**

Table 7: The percentage of grammatical acceptance for each model

Best Model	Gramatical Acceptance Accuracy
T5	**93.36%**
GPT-2	92.54%
Bart	93.27%

The success of BART in summarization tasks can be attributed to its unique design, which effectively captures the context from both directions in its encoder and then generates a coherent, concise summary in its decoder. This approach allows for a more accurate representation and condensation of the original text. While GPT-2 and T5 are formidable in their respective strengths, BART's specific alignment with the requirements of summarization, understanding the essence of the text, and

generating a precise, context-aware summary makes it the most effective among the three for this application.

This outcome underscores the importance of choosing the right model based on the specific requirements of an NLP task. It also highlights the continuous evolution in the field of NLP, where different architectures offer varied advantages for diverse language processing tasks. The findings from using these models in summarization tasks provide valuable insights into their practical applications and pave the way for more targeted and effective use of NLP technologies in real-world scenarios.

5. Conclusion

Within the field of computational linguistics, generating automated summaries of academic literature remains a significant and ongoing challenge. This importance is underscored by the considerable time researchers, educators, and students spend analyzing complex scholarly texts. By reducing this time investment, automated summarization systems can enhance knowledge accessibility, encourage broader reading, and support more efficient learning.

To develop an effective approach for summarizing scientific documents, this study conducted a comparative evaluation of three widely used Transformer-based models [20]: T5, GPT-2, and BART. The results demonstrate that the BART model achieves the strongest performance, attaining a ROUGE-2 F1-score of 0.40664 and outperforming the other two models. In addition, an auxiliary experiment was conducted to assess the grammatical acceptability of the generated summaries. This evaluation utilized a T5 model fine-tuned on the CoLA dataset, which is annotated for linguistic acceptability. The findings indicate that both T5 and BART generate summaries with high grammatical quality, outperforming GPT-2 in this aspect.

For future work, we plan to extend this study by evaluating additional models and incorporating more comprehensive evaluation metrics beyond ROUGE-2 and grammatical acceptability, like ROUGE-L, BERTScore, SummaC, and human evaluation, which will enrich the credibility of the evaluation. Also, to have strong evidence of a model's superiority, other challenging datasets will be included in the study as future work. Furthermore, we aim to develop a hybrid summarization framework that integrates both extractive and abstractive approaches to improve the quality of scientific document summarization.

References

[1] R. R. Waly and W. H. Gomaa, "Extractive Summarization of Scientific Articles," in 2022 2nd International Mobile, Intelligent, and Ubiquitous Computing Conference (MIUCC), May 2022, pp. 349-354. doi: 10.1109/MIUCC55081.2022.9781702.

[2] W. S. El-Kassas, C. R. Salama, A. A. Rafea, and H. K. Mohamed, "Automatic text summarization: A comprehensive survey," Expert Systems with Applications, vol. 165, p. 113679, 2021. doi: 10.1016/j.eswa.2020.113679.

[3] Y. M. Wazery, M. E. Saleh, A. Alharbi, and A. A. Ali, "Abstractive Arabic Text Summarization Based on Deep Learning," Computational Intelligence and Neuroscience, vol. 2022, 2022. doi: 10.1155/2022/9046638.

[4] N. I. Altmami and M. E. B. Menai, "Automatic summarization of scientific articles: A survey," Journal of King Saud University-Computer and Information Sciences, vol. 34, no. 4, pp. 1011-1028, 2020. doi: 10.1016/j.jksuci.2020.04.020.

[5] P. Semaan, "Natural language generation: an overview," J Comput Sci Res, vol. 1, no. 3, pp. 50-57, 2012. doi: 10.14355/jcsr.2012.0103.02.

[6] S. Teufel and M. Moens, "Summarizing scientific articles: experiments with relevance and rhetorical status," Computational Linguistics, vol. 28, no. 4, pp. 409-445, 2002. doi: 10.1162/089120102762671936.

[7] N. Saini, S. Kumar, S. Saha, and P. Bhattacharyya, "Scientific Document Summarization using Citation Context and Multi-objective Optimization," in 2020 25th International Conference on Pattern Recognition (ICPR), Jan. 2021, pp. 4290-4295. doi: 10.1109/ICPR48806.2021.9412089.

[8] C. Raffel, N. Shazeer, A. Roberts, K. Lee, S. Narang, M. Matena, Y. Zhou, W. Li, and P. J. Liu, "Exploring the limits of transfer learning with a unified text-to-text transformer," J. Mach. Learn. Res., vol. 21, no. 1, Art. no. 140, Jan. 2020. doi: 10.5555/3455716.3455856.

[9] C. Zerva, M. Q. Nghiem, N. T. Nguyen, and S. Ananiadou, "Nactem-uom@ cl-scisumm 2019," in Proceedings of the 4th Joint Workshop on Bibliometric-enhanced Information Retrieval and Natural Language Processing for Digital Libraries (BIRNDL@ SIGIR), Jan. 2019.

[10] A. Cohan and N. Goharian, "Scientific document summarization via citation contextualization and scientific discourse," International Journal on Digital Libraries, vol. 19, no. 2, pp. 287-303, 2018. doi: 10.1007/s00799-017-0216-7.

[11] M. La Quatra, L. Cagliero, and E. Baralis, "Poli2sum@ cl-scisumm-19: Identify, classify, and summarize cited text spans by means of ensembles of supervised models," in Proceedings of the 4th Joint Workshop on Bibliometric-enhanced Information Retrieval and Natural Language Processing for Digital Libraries (BIRNDL@ SIGIR), Jan. 2019.

[12] S. Ma, H. Zhang, T. Xu, J. Xu, S. Hu, and C. Zhang, "IR&TM-NJUST@ CLSciSumm-19," in Proceedings of the 4th Joint Workshop on Bibliometric-enhanced Information Retrieval and Natural Language Processing for Digital Libraries (BIRNDL@ SIGIR), 2019, pp. 181-195.

[13] L. Li, Y. Zhu, Y. Xie, Z. Huang, W. Liu, X. Li, and Y. Liu, "CIST@ CLSciSumm-19: Automatic Scientific Paper Summarization with Citances and Facets," in Proceedings of the 4th Joint Workshop on Bibliometric-enhanced Information Retrieval and Natural Language Processing for Digital Libraries (BIRNDL@ SIGIR), vol. 54, 2019.

[14] L. Chiruzzo, À. Bravo, and H. Saggion, "LaSTUS-TALN+ INCO@ CL-SciSumm 2019," in Proceedings of the 4th Joint Workshop on Bibliometric-enhanced Information Retrieval and Natural Language Processing for Digital Libraries (BIRNDL@ SIGIR), Jan. 2019

[15] C. Zerva, M. Q. Nghiem, N. T. Nguyen, and S. Ananiadou, "Cited text span identification for scientific summarisation using pre-trained encoders," Scientometrics, vol. 125, no. 3, pp. 3109-3137, 2020. doi: 10.1007/s11192-020-03741-z.

[16] A. R. Fabbri, W. Kryściński, B. McCann, C. Xiong, R. Socher, and D. Radev, "Summeval: Re-evaluating summarization evaluation," Transactions of the Association for Computational Linguistics, vol. 9, pp. 391-409, 2021. doi: 10.1162/tacl_a_00373.

[17] M. Lewis et al., "Bart: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension," arXiv preprint arXiv:1910.13461, 2019. doi: 10.48550/arXiv.1910.13461.

[18] A. Radford et al., "Language models are unsupervised multitask learners," OpenAI blog, 2019.

[19] A. Warstadt, A. Singh, and S. R. Bowman, "Neural network acceptability judgments," Transactions of the Association for Computational Linguistics, vol. 7, pp. 625-641, 2019. doi: 10.1162/tacl_a_00290.

[20] R. Haruna, A. Obiniyi, M. Abdulkarim and A. A. Afolorunsho, "Automatic Summarization of Scientific Documents Using Transformer Architectures: A Review," 2022 5th Information Technology for Education and Development (ITED), Abuja, Nigeria, 2022, pp. 1-6, doi: 10.1109/ITED56637.2022.10051602.

Islamic University Journal of Applied Sciences (IUJAS)
https://journals.iu.edu.sa/jesc

Volume VIII, Issue II, December 2026, Pages 57-72

Aerodynamic Analysis Using CFD of a Formula SAE Car Front Nose

Faisal Omar Mahroogi [1] *

[1] Department of Mechanical Engineering, Faculty of Engineering, Islamic University of Madinah, Saudi Arabia, f.mahroogi@iu.edu.sa

*Corresponding author: (Faisal O. Mahroogi), *Email Address*: f.mahroogi@iu.edu.sa

Abstract

The aerodynamic performance of Formula SAE (FSAE) cars strongly influences speed, stability, and energy efficiency. This study introduces a systematic CFD based comparison of four front nose geometries an aspect that remains underexplored in existing FSAE literature. Four distinct nose shapes were modeled in SolidWorks and evaluated in ANSYS Fluent under identical boundary conditions at a free stream velocity of 100 km/h. The Reynolds Averaged Navier–Stokes (RANS) equations were solved using the k–ω SST turbulence model, and a mesh independence study ensured numerical accuracy. The optimized nose achieved a drag coefficient of 0.24, representing a 14% reduction relative to the published baseline value of 0.28, thereby quantifying the aerodynamic improvement contributed by this work. Flow field analysis confirmed smoother airflow, delayed separation, and reduced wake turbulence for the optimized geometry. These findings provide new insight into front nose aerodynamic optimization and offer practical design guidance for FSAE teams and potential applications in race car and electric-vehicle aerodynamics.

Keywords: Aerodynamics; Computational Fluid Dynamics (CFD); Drag coefficient; Formula SAE; Front nose; k-ω SST turbulence model.

https://doi.org/10.63070/jesc.2026.034
Received 22 April 2026; Revised 28 May 2026; Accepted 15 May 2026; Available online 20 June 2026.
Published by Islamic University of Madinah on behalf of ***Islamic University Journal of Applied Sciences.***
This is an open access article published under the Creative Commons Attribution 4.0 International license (CC BY 4.0):https://creativecommons.org/licenses/by/4.0/

1. Introduction

Computational Fluid Dynamics (CFD) has emerged as an essential tool in aerodynamic analysis and vehicle design, particularly in high-performance motor-sport applications such as Formula Student (FS) and Formula SAE (FSAE) competitions. CFD combines the principles of fluid mechanics, numerical methods, and computer modeling to predict fluid-flow behavior, including drag, lift, pressure distribution, and wake formation around vehicle components. These capabilities are particularly valuable in FS, where teams of student engineers design, build, and race single-seat formula-style vehicles. Aerodynamics significantly influences the competitiveness of these cars by affecting downforce, drag, fuel efficiency, and overall stability.

Over the past decade, numerous studies have used CFD to optimize the aerodynamic elements of FS and FSAE vehicles. The design of rear wings using three-dimensional CFD has been a prominent research focus, contributing to drag reduction and improved downforce [1]. Broader system-level modeling and analysis have also been undertaken to evaluate the interaction of multiple aerodynamic components [2]. Notably, nose-shape optimization using adjoint methods has proven effective in minimizing drag and directly enhancing speed and fuel economy [3]. Front-wing shape optimization has similarly improved vehicle stability and cornering dynamics [4].

Comprehensive aerodynamic analyses using CFD have extended to external body geometries [5, 6], rear diffusers [13], and spoiler geometries [7, 8], confirming that refined shapes can deliver significant performance gains. Some studies have validated these simulations using experimental techniques and vehicle-dynamics testing [9]. Research on wake behavior [10], front-wing redesign [11], and general vehicle analysis using ANSYS Fluent [12] has highlighted the role of CFD in simulating realistic aerodynamic conditions.

Specific components such as the Gurney flap [14], aerodynamic profiles [15], and spoiler interactions [16] have also been investigated to understand their effects on longitudinal stability. CFD analyses have been applied to the flow around race-car wheels [17], the DrivAer reference model [18], and canards [19] to study localized aerodynamic changes. More recent approaches have combined machine learning with CFD to optimize rear spoilers and enhance lateral stability in passenger and electric vehicles [20].

The placement of rear aerodynamic components has been optimized using CFD for motor-sport applications, for example on the Aston Martin DB7 [21], while the effectiveness of underbody diffusers in road cars has also been validated [22]. Studies integrating structural and aerodynamic design, such as Formula One front-wing assemblies [24], have stressed the importance of a multi-

disciplinary approach. Undertray performance has been optimized for FS cars [25], and component-specific studies on air-intake restrictors [26], flow transitions [27], and high-speed vehicle aerodynamics [28] have further refined the collective understanding. Simulations on body-design variations [29], FSC racing-car bodies [30], and concept sports cars [31] further illustrate the breadth of CFD applications in performance engineering. Finally, investigations of spoiler-induced stability [32] and pitch sensitivity in race-car aerodynamics [33] complete a substantial body of work that underscores the value of precise aerodynamic modeling for vehicle dynamics. Wind-tunnel testing of FS vehicles has likewise been used to verify CFD trends and improve simulation fidelity [23].

Recent advances in Formula Student and FSAE aerodynamics have increasingly focused on detailed CFD based investigations of forward aerodynamic components. Piszczatowski et al. [35] analyzed the front wing cross section of a Formula Student vehicle and demonstrated how subtle geometric variations influence pressure distribution and downforce generation, emphasizing the sensitivity of leading edge components to shape refinement. Building on this component level understanding, Hogea et al. [36] performed a 3D iterative parametric optimization of a complete FSAE aerodynamic package, showing that coordinated adjustments across multiple elements can yield substantial improvements in drag and downforce. These studies collectively highlight the growing reliance on high fidelity CFD workflows to guide aerodynamic decision making in student motorsport engineering.

Parallel developments in turbulence modeling strategies have further strengthened the accuracy of automotive aerodynamic simulations. Sankar et al. [37] compared Detached Eddy Simulation (DES) with traditional RANS approaches, demonstrating that higher fidelity models capture complex flow structures, such as vortex shedding and transient separation, more effectively, particularly around curved or highly loaded surfaces. Complementing this, Telli and Rosa [38] reviewed the aerodynamic influence of front wing geometry in Formula One, underscoring how wing shape modifications affect drag, downforce, and wake behavior. Together, these works establish a strong foundation for focused investigations into forward aerodynamic surfaces and motivate the present study, which extends the literature by evaluating multiple front nose geometries under identical CFD conditions and quantifying their impact on drag, lift, and flow field characteristics.

Despite this extensive work, most published studies have focused on isolated components. There is a notable research gap in fully integrated, front-nose-specific aerodynamic studies within a Formula SAE context. The front nose is the leading interface between the car and the oncoming airflow and plays a critical role in managing pressure gradients, minimizing flow separation, and initiating laminar-to-turbulent transitions.

This study addresses this gap by conducting a CFD-based aerodynamic comparison of four distinct Formula SAE front-nose geometries. By focusing on one of the most aerodynamically sensitive regions of the vehicle and building established methodologies [2, 3, 5, 12, 14, 29], the work identifies the design that provides optimal drag reduction without compromising the lift balance required for stability. The results are expected to deliver practical insights for student engineering teams seeking competitive advantages in Formula SAE through evidence-based aerodynamic refinement.

2. Methodology

2.1 Selection of CFD Tool

ANSYS Fluent was selected as the primary CFD solver for this study. The software is widely recognized as an industry-standard tool, known for its robustness and accuracy in simulating complex aerodynamic environments. It supports a broad range of turbulence models, adaptive meshing techniques, and solver settings that are essential for capturing intricate flow behavior in automotive contexts. The k-ω Shear Stress Transport (SST) turbulence model was chosen because this hybrid formulation blends the near-wall accuracy of the k-ω model with the free-stream robustness of the k-ε model, making it particularly effective at predicting boundary-layer separation and near-wall flow phenomena that directly affect the drag and lift around the vehicle nose.

2.2 Governing Equations

The airflow was modeled using the Reynolds-Averaged Navier–Stokes (RANS) equations, which are suitable for capturing time-averaged turbulent flow over complex geometries. A steady-state solver was employed under the assumption of continuous and consistent race conditions, balancing computational efficiency with a sufficiently detailed representation of aerodynamic behavior. For incompressible flow, mass conservation is enforced through the continuity equation:

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{v}) = 0 \tag{1}$$

where ρ is the fluid density and v is the velocity vector field. Momentum conservation is expressed by the Navier–Stokes equation:

$$\rho \left(\frac{\partial \mathbf{v}}{\partial t} + \mathbf{v} \cdot \nabla \mathbf{v} \right) = -\nabla p + \mu \nabla^2 \mathbf{v} + \mathbf{F} \tag{2}$$

where p denotes pressure, μ is the dynamic viscosity, and F represents body forces per unit volume. Time averaging these equations introduces the Reynolds-stress term, which requires a turbulence-closure model. The k-ω SST model solves two transport equations, one for the turbulent kinetic energy k and one for the specific dissipation rate ω:

$$\frac{\partial k}{\partial t} + \mathbf{v} \cdot \nabla k = P_k - \beta k\omega + \nabla \cdot [(\nu + \sigma_k \nu_t)\nabla k] \quad (3)$$

$$\frac{\partial \omega}{\partial t} + \mathbf{v} \cdot \nabla \omega = \frac{\alpha}{\nu_t} P_k - \beta \omega^2 + \nabla \cdot [(\nu + \sigma_\omega \nu_t)\nabla \omega] \quad (4)$$

where P_k is the production of turbulent kinetic energy, ν is the kinematic viscosity, ν_t is the turbulent (eddy) viscosity, and σ_k, σ_ω, β, β^*, and α are standard model constants. The k-ω SST model is particularly effective in predicting flow separation and boundary-layer development, both of which are critical to estimating drag and optimizing the aerodynamic performance of the front nose.

2.3 Aerodynamic Force Coefficients

The drag force F_d acting on the car is a primary outcome of the simulations and is related to the drag coefficient C_d through:

$$F_d = \frac{1}{2}\, \rho\, v^2\, C_d\, A \quad (5)$$

where v is the free-stream velocity relative to the car, *Cd* is the dimensionless drag coefficient, and A is the frontal reference area of the vehicle. This relation allows a quantitative comparison between different nose geometries through their drag coefficients, which directly influence top speed and fuel consumption. Similarly, the lift coefficient C_l is evaluated to assess any lift (or downforce) generated by the nose design:

$$C_l = \frac{2L}{\rho\, v^2\, A} \quad (6)$$

where L is the lift force; for race cars, negative values (downforce) are generally desirable. The lift coefficient is essential for assessing vehicle stability, since excessive lifts reduce tire grip while appropriate downforce improves handling and cornering.

2.4 Computational Domain and Boundary Conditions

A structured computational domain was constructed to replicate realistic flow conditions and to eliminate artificial effects from the proximity of the outer boundaries. The domain extends ten vehicle lengths in the streamwise direction, five vehicle widths laterally, and five vehicle heights vertically. This generous sizing ensures that the boundary conditions do not interfere with the flow near the vehicle and that the wake can fully develop before reaching the outlet.

The boundary conditions were selected to mimic a representative racing scenario. At the inlet, a uniform velocity of 100 km/h (approximately 27.8 m/s) was applied, representative of the average track speed during an FSAE event. This velocity corresponds to a length-based Reynolds number of approximately 1.2×10^6, confirming that the flow regime is fully turbulent. At the outlet, a pressure boundary condition of 0 Pa Gauge was imposed to allow the air to exit the domain freely and to emulate an open atmospheric environment. The ground was modeled as a moving wall translating at the same speed as the inlet flow (100 km/h) with a no-slip condition, correctly simulating the relative motion between the vehicle and the track surface. A no-slip condition was also applied to the vehicle surface itself, which is essential for resolving wall shear stress, pressure distribution, and boundary-layer development.

2.5 Mesh Generation and Independence Study

An unstructured tetrahedral mesh was used to capture the complex geometry of the car, including the nose, suspension arms, and undertray. Mesh refinement was applied in critical regions such as the nose surface, underbody, wheel arches, and near-wake zones to resolve the sharp gradients in pressure and velocity. To accurately capture near-wall flow, prism layers were added to form a boundary-layer mesh. The first cell height was selected to keep y^+ values in the range of 30–100, which is suitable for the k-ω SST model without requiring full wall resolution.

A mesh-independence study was carried out by successively refining the mesh density and monitoring the variations in drag (C_d) and lift (C_l) coefficients. Refinement continued until the change in both coefficients between successive mesh levels was less than 1%, confirming mesh independence. The final mesh consisted of approximately 25 million cells, providing a suitable balance between accuracy and computational feasibility.

2.6 Convergence Criteria and Stability

Convergence was considered to be achieved when both numerical and physical criteria were satisfied. The scaled residuals for the continuity equation, the three momentum components, the turbulent kinetic energy (k), and the specific dissipation rate (ω) all had to drop below 10^{-5}. In addition, the drag and lift coefficients had to stabilize, with variations of less than 0.1% between successive iterations, indicating that a steady-state solution had been reached. Under-relaxation factors and multigrid acceleration were used to improve numerical stability and to reduce the iteration count without compromising accuracy. After convergence, the velocity and pressure fields were inspected for physical realism; the absence of unphysical spikes or reversed flow confirmed the reliability of the numerical solution.

2.7 Nose Design

Four distinct nose geometries were designed in SolidWorks and subjected to identical boundary conditions in ANSYS Fluent to allow direct comparison. The four candidate shapes are illustrated in Fig. 1. The complete vehicle model, including wheels, suspension, and underbody components, was prepared in SolidWorks and then exported to ANSYS Meshing for discretization prior to the CFD analysis.

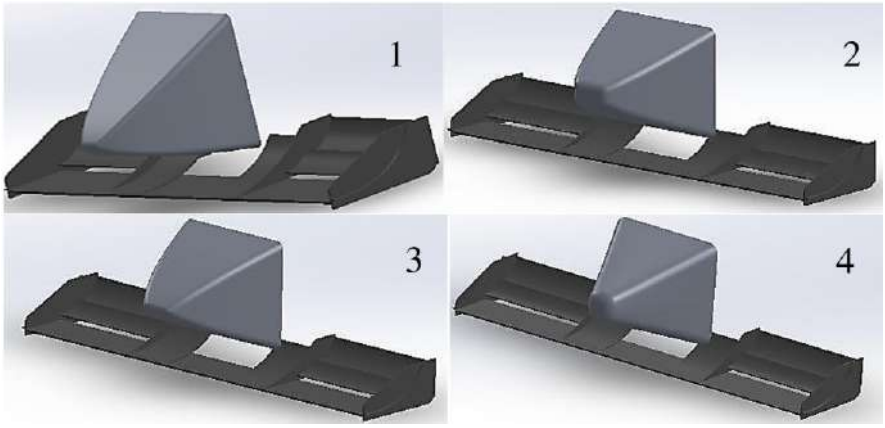


Figure 1. The four-candidate nose-shape models used in the CFD study.

3. Results and Discussion

3.1 Aerodynamic Coefficients for the Four Candidate Noses

The four candidate nose geometries were evaluated under identical boundary conditions at a free-stream velocity of 100 km/h. Table I summarizes the drag coefficient C_d, drag force F_d, lift coefficient C_l, and lift force L extracted from the converged CFD solutions.

Table I. Aerodynamic coefficients and forces obtained from CFD for the four nose shapes at a free-stream velocity of 100 km/h.

Shape #	**Drag coefficient C_d**	**Drag force F_d (N)**	**Lift coefficient C_l**	**Lift force L (N)**
1	0.24	71.906	0.0332	34.84
2	0.27	79.430	0.0522	54.27
3	0.26	76.507	0.0418	43.46
4	0.25	75.047	0.0354	36.81

Shape 1 produced the lowest drag coefficient (C_d = 0.24) and the lowest lift coefficient (C_l = 0.0332), which corresponds to a modest lift force of 34.84 N. Shape 2 exhibited the highest drag and the highest lift, making it the least favorable configuration from an aerodynamic-efficiency standpoint. Shapes 3 and 4 fell between these extremes, with Shape 4 performing only marginally worse than Shape 1 in drag. On this basis, Shape 1 was selected as the optimized design. Its low drag coefficient is expected to translate into higher achievable top speeds and reduced fuel consumption during FSAE dynamic events.

3.2 Flow-Field Analysis

Velocity contours and streamline plots were generated for each candidate geometry to visualize how the flow interacts with the different nose shapes. For the selected (optimized) geometry, the velocity contours show a smooth flow over the forward surface and delayed separation along the sides and top of the nose, consistent with the low drag coefficient observed in Table I. Pressure-coefficient (*Cp*) plots indicate a more uniform pressure distribution and a reduced stagnation peak, which results in lower pressure drag and a more aerodynamic profile. Streamline visualizations confirm that the wake behind Shape 1 is smaller and less turbulent than the wakes of the other three geometries, whose flow separation begins earlier along the nose and produces larger, more energetic trailing structures. Figs. 2 and 3 illustrate, respectively, the two-dimensional streamline contour in the z-direction and the z-directional velocity field for the optimized geometry.

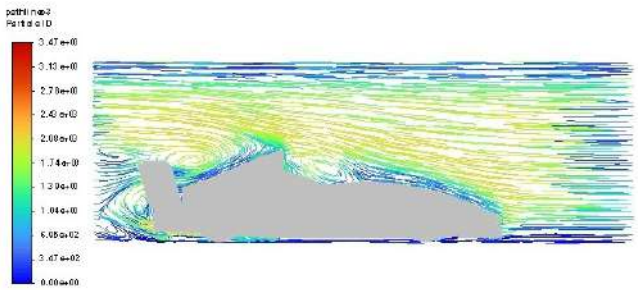


Figure 2. CFD two-dimensional contour of the z-directional streamlines around the vehicle.

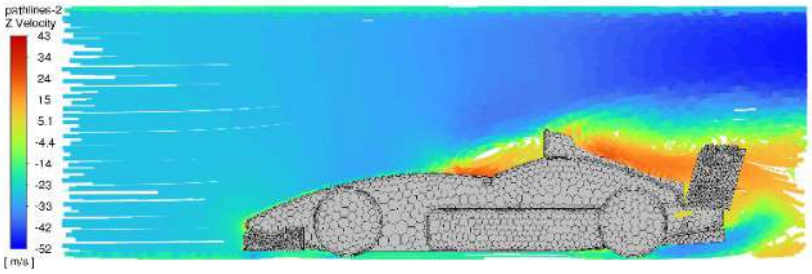


Figure 3. CFD visualization of the velocity field in the z-direction for the optimized geometry (Shape 1).

Iterative testing of the parametric design space highlighted two consistent trends. First, a shorter, more rounded nose profile (as in Shape 1) reduces the frontal area and improves flow attachment, yielding approximately a 3.5% drag reduction relative to the longer alternatives. Second, a more tapered cross-section reduces pressure drag and turbulence while keeping the generated lift within the acceptable range for overall vehicle balance. The final nose geometry therefore represents a balanced compromise between drag reduction and vehicle control, both of which are vital in competitive racing.

3.3 Comparative Analysis

The results were compared with previous CFD studies and empirical data to validate the findings. The drag coefficient of 0.24 obtained for the optimized nose corresponds to approximately a 14% reduction relative to the reference value of 0.28 reported for baseline FSAE configurations with conventional front-nose geometries [12]. This reduction is consistent with the 10–15% improvements documented in [3] and [15] for streamlined nose profiles, supporting the effectiveness of the design refinements adopted in this study.

Flow-visualization results show that the rounded geometry delays separation and reduces wake size, which matches the findings of [14] and other prior studies where angular nose designs produced more turbulent wakes. These flow-field improvements contribute directly to the lower drag, as confirmed by the streamline and velocity-contour plots in Figs. 2 and 3.

When compared with published wind-tunnel data [23], the present drag coefficient is slightly lower than the 0.26 observed experimentally for moderately optimized FSAE configurations. This small difference is expected, as CFD assumes idealized conditions, smooth surfaces, perfect symmetry, steady-state operation, whereas real-world tests include surface imperfections, mounting-hardware interference, and environmental variability. The optimized lift coefficient ($C_l = 0.033$) also falls within the typical empirical range of 0.03-0.05 reported in the literature, confirming the capacity of the design to maintain adequate downforce for vehicle stability. The pressure-distribution trends further validate

this, matching the general patterns reported in the experimental literature. Overall, the CFD results agree closely with both prior simulations and physical benchmarks, reinforcing the reliability of the model. Future work should incorporate dedicated wind-tunnel validation to further confirm the aerodynamic performance under actual operating conditions.

3.4 Practical Implications

The aerodynamic optimization of the front nose demonstrated in this study offers direct performance benefits for Formula SAE vehicles. The reduction in drag enhances acceleration, top speed, and energy efficiency, all of which are critical in competitive racing. Maintaining a balanced lift coefficient also preserves vehicle stability and improves tire grip during aggressive maneuvers. Teams can leverage the parametric design approach presented here to adapt nose geometries to their specific vehicle configurations and extend similar techniques to other aerodynamic components such as rear wings or undertrays.

Beyond racing, these findings are relevant to production vehicles, particularly electric and hybrid cars, where minimizing aerodynamic drag improves range and reduces noise emissions. The smoother airflow and smaller wake associated with the optimized nose also help mitigate aerodynamic noise, which is increasingly perceptible in quiet electric powertrains. Finally, the study highlights the role of CFD as a cost-effective alternative to physical prototyping: by using simulation during early development, teams and manufacturers can reduce reliance on expensive wind-tunnel models, shorten design cycles, and iterate geometries more freely. The project also serves as an educational example, engineering students can use its methodology to understand the practical application of fluid dynamics and vehicle design with modern simulation tools. The final fabricated Formula SAE vehicle is shown in Fig. 4.



Figure 4. Final fabricated Formula SAE car.

4. Conclusions

This study optimized the aerodynamic performance of a Formula SAE car's front-nose geometry using CFD simulations in ANSYS Fluent with the k-ω SST turbulence model. Four candidate nose shapes were modeled in SolidWorks and evaluated under identical boundary conditions at 100 km/h. Among the tested designs, the optimized nose (Shape 1) achieved a 14% reduction in drag coefficient relative to the baseline reference, without compromising the lift balance and therefore preserving the downforce required for high-speed handling and cornering stability.

Flow-field analyses revealed smoother airflow, delayed flow separation, and reduced wake formation around the optimized geometry. A more uniform pressure distribution on the forward surface further confirms its aerodynamic efficiency, and these characteristics directly explain the reduction in pressure drag, a key factor in overall performance.

Although the results align with existing simulation benchmarks and published experimental data, empirical validation through wind-tunnel testing remains desirable. Steady-state RANS simulations necessarily rely on idealizations, smooth surfaces, perfect symmetry, steady inflow, that may not fully reflect real-world conditions, and dedicated experiments would help identify any discrepancies due to surface imperfections or environmental effects.

The practical implications of this research extend beyond FSAE competitions. The parametric design and optimization workflow presented here is directly transferable to broader automotive applications,

including electric and hybrid vehicles, where aerodynamic efficiency strongly influences energy consumption and driving range. The methodology also illustrates how CFD can act as a cost-effective alternative to early-stage physical prototyping.

The study is limited to the front nose and does not account for aerodynamic interactions with other components such as rear wings and diffusers. Computational resources also constrained the use of higher-order solvers and transient simulations, which are needed to capture dynamic effects during parking or cornering. Recommended future work therefore includes: (i) wind-tunnel testing to validate and refine the CFD models; (ii) expansion of the simulations to full-vehicle aerodynamics; (iii) incorporation of transient flow simulations for dynamic analysis; (iv) study of material and manufacturing tolerance effects on aerodynamic performance; and (v) exploration of advanced optimization techniques, including machine learning and genetic algorithms.

In conclusion, this research successfully delivered a front-nose design that enhances aerodynamic efficiency while maintaining vehicle stability. The findings contribute both to academic understanding and to the practical development of high-performance vehicles, offering a replicable approach for future aerodynamic-optimization efforts in motorsport and general automotive engineering. The optimized nose (Shape 1) achieved a drag coefficient of 0.24, representing a 14% reduction compared to the baseline value of 0.28.

References

[1] Dharmawan, M. A., Tjahjana, D. D. P., Kristiawan, B., and Pertiwi, S. I., "Design and Aerodynamics Analysis of Rear Wing Formula Student Car Using 3-Dimension CFD (Computational Fluid Dynamics)." *AIP Conference Proceedings* 2217, no. 1 (2020). https://doi.org/10.1063/5.0001049

[2] Baren, I. V., Milligan, A. V., Ashcraft, S., Rosser, S., and Wang, X., "Modeling, Analysis and Design of the Formula SAE Aerodynamics System." In *Proceedings of the ASME International Mechanical Engineering Congress and Exposition*, vol. 84584, V010T10A062, 2020. https://doi.org/10.1115/IMECE2020-24374

[3] Hoque, M. A., Rahman, M. S., Rimi, K. N., Alif, A. R., and Haque, M. R., "Enhancing Formula Student Car Performance: Nose Shape Optimization via Adjoint Method." *Results in Engineering* 20 (2023): 101636. https://doi.org/10.1016/j.rineng.2023.101636

[4] Kalinowski, M. and Szczepanik, M., "Aerodynamic Shape Optimization of Racing Car Front Wing." *IOP Conference Series: Materials Science and Engineering* 1037, no. 1 (2021): 012058. https://doi.org/10.1088/1757-899X/1037/1/012058

[5] G. V. Kulkarni, H. S. Balasubramanya, B. V. Nagesh, Y. Lakkegowda, and B. R. Omkaresh, "Aerodynamics study of the Formula SAE car using analytical approach," Journal of Mines, Metals & Fuels, vol. 70, no. 4, pp. 455–461, 2022. https://doi.org/10.18311/jmmf/2022/32024

[6] A. Mohanan, J. A. Pandey, A. V. Gupta, S. G. Almeida, and D. Choudhari. "Development of External Aerodynamics of an FSAE Racecar Using Computational Fluid Dynamics.", *International Journal of Engineering Research & Technology (IJERT)*, 2021. ISSN: 2278-0181

[7] S. Patil, and L. B. Rao. "Aerodynamic Study of the Different Spoiler Section Design Used in Formula-SAE Cars." *SAE Technical Paper* 2021-28-0181, 2021. https://doi.org/10.4271/2021-28-0181

[8] S. Wille, K. Hendrickson, and R. K. Agarwal. "Optimization Study of Front and Rear Wings of Motorsport Using Computational Fluid Dynamics." In *AIAA SciTech 2021 Forum*, 1734, 2021. https://doi.org/10.2514/6.2021-1734

[9] E. S. Cole, "*Design and Validation of an Aerodynamic System for a Formula SAE Vehicle Using Vehicle Dynamic Simulation and Experimentation."* Master's thesis, West Virginia University, 2023.

[10] A. Guerrero, and R. Castilla, "Aerodynamic Study of the Wake Effects on a Formula 1 Car." *Energies* 13, no. 19 (2020): 5183. https://doi.org/10.3390/en13195183

[11] Bavapuram, V. S. S., S. N. Suresh, S. Sharma, and V. Kaushik, "Aerodynamic Performance Analysis of Formula-Based Vehicle Through ANSYS." In *Future Learning Aspects of Mechanical Engineering*, 121–132. Singapore: Springer, 2022.

[12] Md. Hassaan, S. Dewivedi, H. Khurshid, G. H.Warsi, M. Alam, A. Rahim, "Computational Study on the Diffuser of Formula Racing Car." In *Future Learning Aspects of Mechanical Engineering*, 481–496. Singapore: Springer, 2022. https://doi.org/10.1007/978-981-99-1894-2_41

[13] M. Basso, C. Cravero, and D. Marsano. "Aerodynamic Effect of the Gurney Flap on the Front Wing of a F1 Car and Flow Interactions with Car Components." *Energies* 14, no. 8 (2021): 2059. https://doi.org/10.3390/en14082059

[14] S. Upadhye1, D. Khairwa, M. Nagpure, A. Mangrulkar, A. Deotale, H. Nagpure, "Estimation of Drag and Downforces Using Aerodynamic Profile on Formula Student Vehicle." *International Journal of Engineering Research and Technology*, 2021.

[15] J. Katz, et al. "Aerodynamics in Motorsports." *Proceedings of the Institution of Mechanical Engineers, Part P: Journal of Sports Engineering and Technology* 235, no. 4 (2021): 324–338. https://doi.org/10.1177/1754337119893226

[16] C. Cravero, and D. Marsano. "Computational Investigation of the Aerodynamics of a Wheel Installed on a Race Car with a Multi-Element Front Wing." *Fluids* 7, no. 6 (2022): 182. https://doi.org/10.3390/fluids7060182

[17] P. Qin, A. Ricci, and B. Blocken. "CFD Simulation of Aerodynamic Forces on the DrivAer Car Model: Impact of Computational Parameters." *Journal of Wind Engineering and Industrial Aerodynamics* 248 (2024): 105711. https://doi.org/10.1016/j.jweia.2024.105711

[18] M. I. Ma'arof, L. Puvunathan, W, Z, M, Yusof, S. Suresh, G. T. Chala, "Computational Fluid Dynamic (CFD) Analysis to Determine Canard's Downforce." *INTI Journal* 22 (2020).

[19] A. Bhosale, M. S. Malik, S. P. Dillibabu, V. K. Javanjal, R, Karthikeyan K. "Enhancing Passenger Vehicle Performance: A Study of CFD Technique and Machine Learning for Optimal Rear Spoiler Design and Lateral Stability." *International Journal of Intelligent Systems and Applications in Engineering* 12, no. 2 (2024): 100–111.

[20] T. P. O'Driscoll, and A. R. Barron. "CFD Analysis of the Location of a Rear Wing on an Aston Martin DB7 in Order to Optimize Aerodynamics for Motorsports." *Vehicles* 4, no. 2 (2022): 608–620. https://doi.org/10.3390/vehicles4020035

[21] A. Guerrero, R. Castilla, and G. Eid. "A Numerical Aerodynamic Analysis on the Effect of Rear Underbody Diffusers on Road Cars." *Applied Sciences* 12, no. 8 (2022): 3763. https://doi.org/10.3390/app12083763

[22] J. Pacheco, A. C. Marta, and L. Eça. "Wind Tunnel Testing of a Formula Student Vehicle for Checking CFD Simulation Trends." *Proceedings of the Institution of Mechanical Engineers, Part D: Journal of Automobile Engineering* vol. 238, no. 14 (2023). https://doi.org/10.1177/09544070231203076

[23] X. Castro, and Z. A. Rana. "Aerodynamic and Structural Design of a 2022 Formula One Front Wing Assembly." *Fluids* 5, no. 4 (2020): 237. https://doi.org/10.3390/fluids5040237

[24] Predescu, O. G. *Racecar Undertray Study for a Formula SAE Prototype.* Bachelor's thesis, Universitat Politècnica de Catalunya, 2022.

[25] M. Lučić, "Computational Fluid Dynamics (CFD) Analysis of the Divergent Part of the Formula Student Car Restrictor." *Transactions of Motauto World* 8, no. 1 (2023): 9–13.

[26] K. Ambo, H. Nagaoka, D. A. Philips, C. Ivey, G. A. Brès and S. T. Bose, "Aerodynamic Force Prediction of the Laminar to Turbulent Flow Transition Around the Front Bumper of the Vehicle Using Dynamic-Slip Wall Model LES." In *AIAA SciTech 2020 Forum*, 0036, 2020. https://doi.org/10.2514/6.2020-0036

[27] P. Singh, V. Chhettri, and N. K. Gupta. "Computational Analysis of Aerodynamics Characteristics of High-Speed Moving Vehicle." In *Characterization, Testing, Measurement, and Metrology*, 125–137. Boca Raton, FL: CRC Press, 2020.

[28] Y. Indrawan, and M. Mulyadi. "Pengaruh Variasi Model Body Formula SAE terhadap Efisiensi Kerja dengan Uji Simulasi Aerodinamika." *Innovation Technology: Methodical Research Journal* 3, no. 3 (2024): 12. https://doi.org/10.47134/innovative.v3i3.111

[29] Chen, Y., J. Zhou, G. Su, and C. Wang. "Design and Analysis of FSC Formula Racing Car Body." *Journal of Physics: Conference Series* 1635, no. 1 (2020): 012025. https://doi.org/10.1088/1742-6596/1635/1/012025

[30] M. Dickison, M. Ghaleeh, S. Milady, S. Subbakrishna, L. T. Wen, M. Al Qubeissi, "Investigation into the Aerodynamic Performance of a Concept Sports Car." *Journal of Applied Fluid Mechanics* 13, no. 2 (2020): 583–601. https://doi.org/10.29252/jafm.13.02.30179

[31] N. Stojanovic, O. I. Abdullah, Ivan Grujic, Jasna Glisovic, "The Influence of Spoiler on the Aerodynamic Performances and Longitudinal Stability of the Passenger Car Under High-Speed Condition." *Journal of Visualization* 26, no. 1 (2023): 97–112. https://doi.org/10.1007/s12650-022-00867-2

[32] S. Anbalagan, C. Deepak, K. Virmani, and T. Madhogaria. "Computational Analysis of Pitch Sensitivity for Concept Race Cars." *SAE Technical Papers*, (2022).

[33] A. Huminic, and G. Huminic. "Aerodynamic Study of a Generic Car Model with Wheels and Underbody Diffuser." *International Journal of Automotive Technology* 18, no. 3 (2017): 397–404. https://doi.org/10.1007/s12239-017-0040-6

[34] A. Piszczatowski, M. Kmiotek, A. Kordos, and A. Zaremba. "Modeling of the Front Wing Cross-Section for a Formula Student Vehicle." *Advances in Science and Technology Research Journal* 20, no. 6 (2026). https://doi.org/10.12913/22998624/216767

[35] E. N. Hogea, R. J. Hogea, and R. K. Agarwal. "3D Iterative Parametric Optimization Study of a FSAE Aerodynamic Package Using Computational Fluid Dynamics." In *AIAA SciTech 2026 Forum*, 1898, 2026. https://doi.org/10.2514/6.2026-1898

[36] Sankar, Hari, Sarang Holay, Masami Ikeda, and Ramanand Singh. "Understanding Prevailing Physics in Automotive Aerodynamics Applications—Detached Eddy Simulation vs Reynolds-Averaged Navier–Stokes Approach." *SAE International Journal of Commercial Vehicles* 19, no. 4 (2026). https://doi.org/10.4271/02-19-04-0023

[37] C. M. de Freita, G. D. Telli, and J. S. Rosa. "Aerodynamic Influence of Front Wing Geometry on Formula One Car Performance: A Computational Fluid Dynamics Review." In *SAE Brasil 2025 Congress*. SAE Technical Paper, 2025. https://doi.org/10.4271/2025-36-0043

Islamic University Journal of Applied Sciences (IUJAS)
https://journals.iu.edu.sa/jesc

Volume VIII, Issue II, December 2026, Pages 73-94

A Comprehensive Framework of Best Practices for Designing, Implementing, and Competing with a Formula SAE Race Car

Faisal Omar Mahroogi [1] *

[1] Department of Mechanical Engineering, Faculty of Engineering, Islamic University of Madinah, Saudi Arabia, f.mahroogi@iu.edu.sa

*Corresponding author: (Faisal O. Mahroogi), *Email Address*: f.mahroogi@iu.edu.sa

Abstract

Designing and constructing a Formula SAE (FSAE) vehicle is a premier engineering capstone project. This paper presents a comprehensive framework centered on five critical pillars: strict rule adherence, systems engineering, rigorous project management, design for manufacturability, and iterative testing. By analyzing these pillars through decision matrices and process flowcharts, the framework provides actionable guidance for academic teams. A central theme is that a reliable, well-tested vehicle consistently outperforms a theoretically superior but fragile design. The study addresses both Internal Combustion Engine (ICE) and Electric Vehicle (EV) platforms, reflecting the modern landscape of the Formula Student series. Key lessons emphasize recognizing regulatory limitations, maintaining cohesive system integration, and strictly allocating time for validation. The provided resources—including comparative data tables and failure mode frameworks—serve as a modular toolkit for teams to optimize their unique programs. Ultimately, success depends on the transition from theoretical engineering to a physical prototype that can endure the rigors of competition. Regardless of powertrain choice, the core principles of feasibility and thorough testing remain the ultimate drivers of performance.

Keywords: Formula SAE; Design for manufacturability; Vehicle dynamics; Electric vehicle.

https://doi.org/10.63070/jesc.2026.035
Received 22 April 2026; Revised 28 May 2026; Accepted 04 June2026; Available online 15 June 2026.
Published by Islamic University of Madinah on behalf of ***Islamic University Journal of Applied Sciences.***
This is an open access article published under the Creative Commons Attribution 4.0 International license (CC BY 4.0):https://creativecommons.org/licenses/by/4.0/

1. Introduction

The Formula SAE (FSAE) competition, organized by SAE International, challenges university teams to conceive, design, fabricate, develop, and compete with small, open-wheel, formula-style racing cars. Since its inception in 1981, FSAE has grown into the world's premier collegiate engineering design competition, with over 600 teams participating across events in North America, Europe, Australasia, and Asia. The competition's enduring success lies in its unique ability to bridge the gap between classroom theory and the high-pressure realities of industry-level product development.

The challenge is multifaceted: teams must not only engineer a high-performance vehicle but also develop a comprehensive business case, manage a budget, present their design philosophy to industry judges, and demonstrate the car's performance through a series of dynamic events including acceleration, skid-pad, autocross, and a grueling 22-kilometer endurance race. This holistic evaluation mirrors the real-world demands placed on automotive engineering professionals, making FSAE an unparalleled educational platform. This paper provides a structured and detailed framework for guiding a senior design team through the process of building a competitive Formula-style race car. It identifies five critical pillars that, based on extensive observation of successful and unsuccessful programs, consistently separate cars that compete effectively from those that never reach the starting line. Each pillar is explored in depth, supported by tables, flowcharts, and figures that teams can adapt as practical tools for their own programs. The selection of these specific five pillars was derived from a systematic historical analysis of Formula SAE competition outcomes spanning several decades. To construct this framework, a retrospective review protocol was applied to extensively observe both successful and unsuccessful collegiate programs, isolating the core variables that consistently dictate whether a vehicle competes effectively or fails to reach the starting line. By synthesizing this historical performance data with expert insights from long-standing faculty advisors and industry judges, the myriad challenges of race car development were distilled into five distinct, foundational categories. These five pillars, strict rulebook adherence, systems engineering, project management, design for manufacturability, and rigorous validation, were identified as the minimum necessary dimensions required to holistically address the competition's multi-disciplinary demands without overcomplicating the pedagogical model for student engineering teams.

The design and development of a Formula SAE or Formula Student race car is a complex, multidisciplinary engineering challenge. Foundational principles of vehicle dynamics are heavily relied upon, with seminal texts by Milliken and Milliken [15], Gillespie [9], and Dixon [7] providing the theoretical bedrock for suspension geometry, handling, and race car behavior. Although published decades ago, Carroll Smith's Tune to Win [22] remains highly relevant to modern Formula SAE

practice because it provides the essential theoretical bedrock for understanding suspension geometry, handling, and overall race car behavior. While contemporary FSAE teams rely extensively on advanced, computer-aided simulation and optimization tools, the successful application of these modern techniques requires a seamless integration with classical mechanical engineering principles. By establishing the foundational concepts of vehicle dynamics, Smith's work ensures that student engineers comprehend the fundamental physical interactions of the chassis and suspension before attempting to optimize them through modern software. These specific automotive principles are supported by broader mechanical engineering and design methodologies, including classical mechanics [12], material selection [1], and design for manufacturability and assembly [3, 24]. Furthermore, all theoretical and physical designs must strictly adhere to continuously updated regulatory frameworks. Vehicles must comply with comprehensive safety and structural guidelines set forth by SAE International for both internal combustion [19] and electric vehicles [20], as well as regional variations like the Formula Student Germany rules [8]. The structural integrity and dynamic response of the chassis and suspension systems are critical areas of continuous research. Early studies established the vital relationship between chassis torsional stiffness and handling balance [6]. Recent literature heavily emphasizes computer-aided optimization to maximize strength while reducing weight. This includes the extensive use of Finite Element Analysis (FEA) for chassis optimization [2] and physical dynamic testing to validate simulated models of car bodies [4].

Suspension and steering systems are similarly scrutinized. Researchers have explored various mathematical modeling techniques to optimize passive suspension systems [16] and have detailed the rigorous development processes required for FSAE steering prototypes [18]. At the component level, the wheel uprights face significant dynamic loads; recent studies have successfully applied parametric design, topology optimization, and fatigue life analysis to create highly efficient, lightweight uprights for electric FSAE cars [13, 26]. Reliable deceleration is just as critical as acceleration. Literature surrounding FSAE braking systems focuses on the integration of theoretical design with practical simulation and physical testing to ensure safety and performance under racing conditions [5, 11]. In terms of propulsion, the literature reflects the motorsport industry's dual focus on optimizing traditional combustion engines and advancing electric vehicle (EV) technology. For combustion vehicles, numerical investigations have been used to fine-tune load control systems in turbocharged engines to balance fuel efficiency with power output [17]. Meanwhile, the shift toward electric racing has driven research into advanced all-wheel-drive electric propulsion systems intended to maximize tractive performance [10], as well as optimized transmission and wheel systems derived from higher-tier electric racing like Formula E [25].

Aerodynamic performance has become a major differentiator in modern motorsports. Research highlights the importance of understanding complex flow structures—such as floor aerodynamics in Formula 1—to generate efficient downforce [21]. In the context of FSAE, engineers rely heavily on CFD simulations to predict transmitted tire reaction forces and overall vehicle stability [14]. Pushing the boundaries of traditional simulation, cutting-edge research has recently introduced AI-assisted CFD to rapidly optimize the angle of attack for multi-element wings, significantly enhancing aerodynamic efficiency [23]. The provided literature illustrates that modern Formula SAE design requires a seamless integration of classical mechanical engineering theory [12, 15] with advanced, computer-aided optimization techniques [2, 23]. Whether optimizing an electric powertrain, reducing the mass of an upright, or maximizing aerodynamic downforce, successful vehicle development relies on iterative simulation validated against strict regulatory standards [8, 19, 20].

Despite the wealth of highly specialized research focusing on the theoretical optimization of individual vehicle subsystems, a significant gap remains in the current literature regarding the holistic execution of the project. Specifically, existing literature lacks a comprehensive, systems-level framework that integrates technical component design with rigorous project management, strict regulatory compliance, and practical manufacturability assessments for student engineering teams. This paper addresses this critical void by proposing a unified, five-pillar framework designed to guide teams through the entire vehicle development lifecycle. By shifting the academic focus from isolated theoretical optimization to systemic reliability and cross-disciplinary integration, this framework provides actionable strategies to help teams successfully navigate the complex transition from an initial concept to a competition-ready vehicle.

2. Competition Overview and Scoring

Understanding the scoring breakdown is foundational to effective design prioritization. The FSAE competition allocates points across both static and dynamic events. A common strategic error made by novice teams is to focus exclusively on dynamic performance (speed) while underestimating the weight of static events, which collectively account for 325 of the total 1000 points. Table 1. presents the standard FSAE scoring distribution.

Table 1.: Standard FSAE Competition Scoring Distribution (1000 Total Points)

Event Category	Max Points	% of Total
Engineering Design	150	15.0%
Cost & Manufacturing Analysis	100	10.0%

Business Presentation	75	7.5%
Acceleration	100	10.0%
Skid-Pad	75	7.5%
Autocross	125	12.5%
Endurance	275	27.5%
Fuel Efficiency / Energy	100	10.0%
TOTAL	**1000**	**100%**

As Table 1. illustrates, the endurance event alone is worth 27.5% of the total score, making vehicle reliability arguably the single most important design objective. A car that completes endurance will almost always outscore a faster car that suffers a mechanical failure and does not finish (DNF). This insight should fundamentally shape every design decision made during the project lifecycle.

3. Strict Adherence to The Rulebook

The FSAE rulebook, published annually by SAE International, functions as the definitive client specification for the project. It establishes mandatory constraints on virtually every aspect of the vehicle, from the minimum inner diameter of the intake restrictor (20 mm for gasoline-powered vehicles) to the exact height of the main roll hoop above the driver's helmet. The most common and costly point of failure for college teams is an insufficient understanding of these regulations. A vehicle that fails technical inspection, regardless of its engineering sophistication, is reduced to a static display piece that cannot participate in any dynamic event.

3.1 Understanding the Rulebook as a Design Specification

The rulebook should be treated with the same gravity as a contractual specification in industry. Each section imposes constraints that define the design space within which the team must operate. Key regulatory areas include the structural requirements for the chassis and roll hoops, driver protection via impact attenuators, fuel system safety, electrical system specifications, and ergonomic requirements such as the 95th-percentile male driver template and the five-second egress rule.

Teams are strongly advised to assign individual members as "rulebook owners" for specific sections, making each member the subject-matter expert on the regulations governing their subsystem. A cross-referenced compliance matrix should be maintained throughout the design phase and updated

continuously. Fig. 1 illustrates the iterative compliance verification workflow recommended for FSAE design teams.

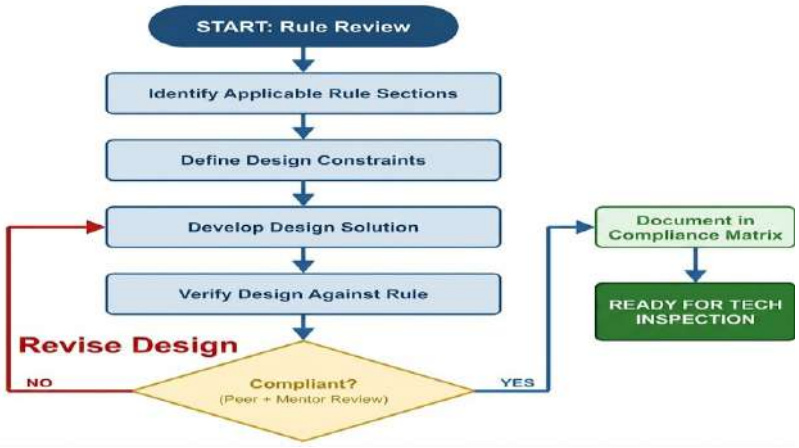


Fig.1: Rulebook Compliance Verification Workflow, From Rule Identification to Technical Inspection Readiness

3.2 Technical Inspection: The Gate to Competition

Technical inspection at competition is a multi-stage process comprising mechanical inspection, tilt testing (to verify the vehicle does not leak fluids at a 45-degree and 60-degree tilt), a noise test, a rain test for electrical vehicles, and a brake test. Table 2. summarizes the key inspection areas and common failure modes observed in competition.

Table 2.: Technical Inspection Areas, Requirements, and Common Failure Modes

Inspection Area	**Key Requirements**	**Common Failures**
Chassis / Frame	Min. tube specs, triangulation, SES	Unapproved tube material or wall thickness
Roll Hoops	Height above helmet, bracing angle	Insufficient bracing, incorrect supports
Impact Attenuator	Approved data or IA test report	Missing or unapproved attenuator
Driver Egress	5-second exit with full harness	Tight cockpit, harness interference
Fuel System	Tank mounting, vent routing, firewall	Improper fuel line routing or leaks
Electrical (EV)	HV insulation, BSPD, IMD, shutdown circuit	Shutdown circuit logic errors
Tilt Test	No leaks at 45° and 60°	Oil or coolant leaks under tilt

Brake Test	All four wheels lock simultaneously	Insufficient braking or uneven lock-up

4. Systems Engineering and Cross-Disciplinary Integration

A Formula SAE car is a complex, coupled system of systems. The vehicle's performance is an emergent property arising from the interaction of its chassis, suspension, powertrain, aerodynamics, electronics, and driver interface. Treating these as independent modules that can be designed in isolation and bolted together at the end is the second most common cause of project failure. Effective systems engineering requires a rigorous approach to architecture definition, interface management, and continuous integration.

4.1 Vehicle Architecture and Packaging

The first critical system-level decision is the selection of the overall vehicle architecture. This includes the choice of powertrain (ICE vs. EV), frame type (steel spaceframe, aluminum monocoque, or carbon fiber monocoque), and drivetrain configuration (single motor/engine with chain drive, differential, or individual wheel motors). These decisions cascade through every subsystem and must be resolved early in the conceptual design phase. Fig. 2. presents a system interaction map illustrating the coupling between major vehicle subsystems.

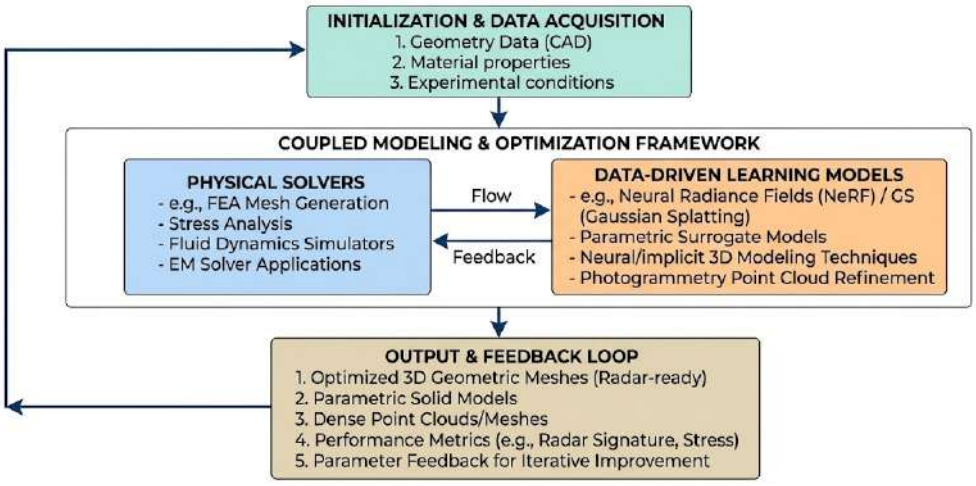


Fig. 2.: System Interaction Map, Map, Majorystem Coupling in a Formula SAE Vehicle

The packaging process begins with defining the driver envelope using the required 95th-percentile template and the Percy template for visibility. Around this, the primary structure (frame or monocoque) is designed, establishing the mounting points and spatial constraints for all other subsystems. The powertrain and drivetrain dictate the rear geometry, while the suspension kinematics define the front

and rear track widths and the attachment point requirements on the chassis. Fig. 3. illustrates a typical FSAE vehicle packaging layout, highlighting the spatial allocation for each major subsystem.

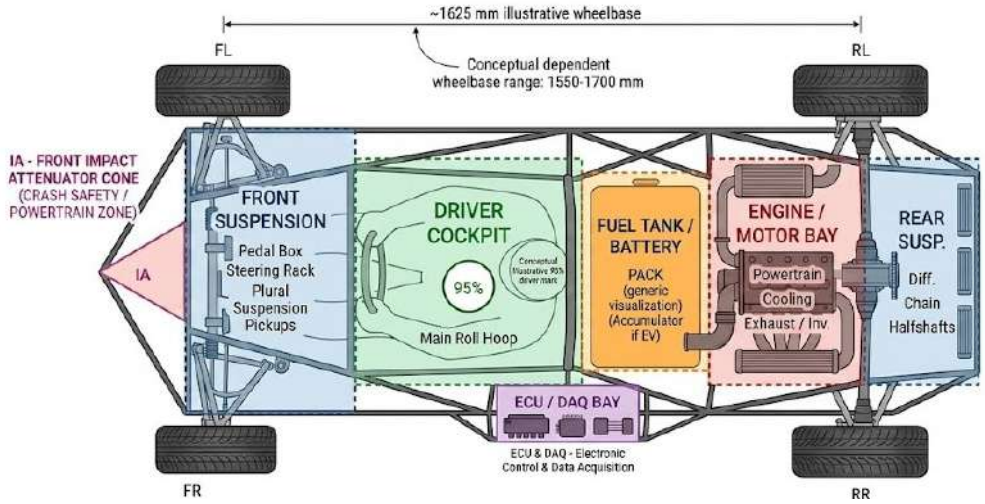


Fig. 3: Typical Formula SAE Vehicle Layout, Top View Showing Major Subsystem Packaging Zones

4.2 The Nervous System: Electrical and Electronic Integration

While mechanical systems naturally dominate initial attention, electrical and electronic architecture is frequently the subsystem that determines competition-day reliability. The wiring harness is the vehicle's nervous system, connecting the engine control unit (ECU), sensors, actuators, dashboard, data acquisition system, and safety circuits. A poorly routed harness is susceptible to heat damage from the exhaust, vibration-induced connector failures, and electromagnetic interference (EMI).

Best practices for electrical system design include defining a complete wiring diagram before any physical routing begins, using automotive-grade connectors (such as Deutsch DTM or Amphenol AT series), implementing proper strain relief at every connector, protecting all harness branches with abrasion-resistant conduit, and integrating a comprehensive data acquisition system from day one. The data acquisition system should capture at minimum RPM, throttle position, coolant temperature, oil pressure, wheel speeds, lateral and longitudinal acceleration, and steering angle. Fig. 4 presents a representative electrical system architecture block diagram.

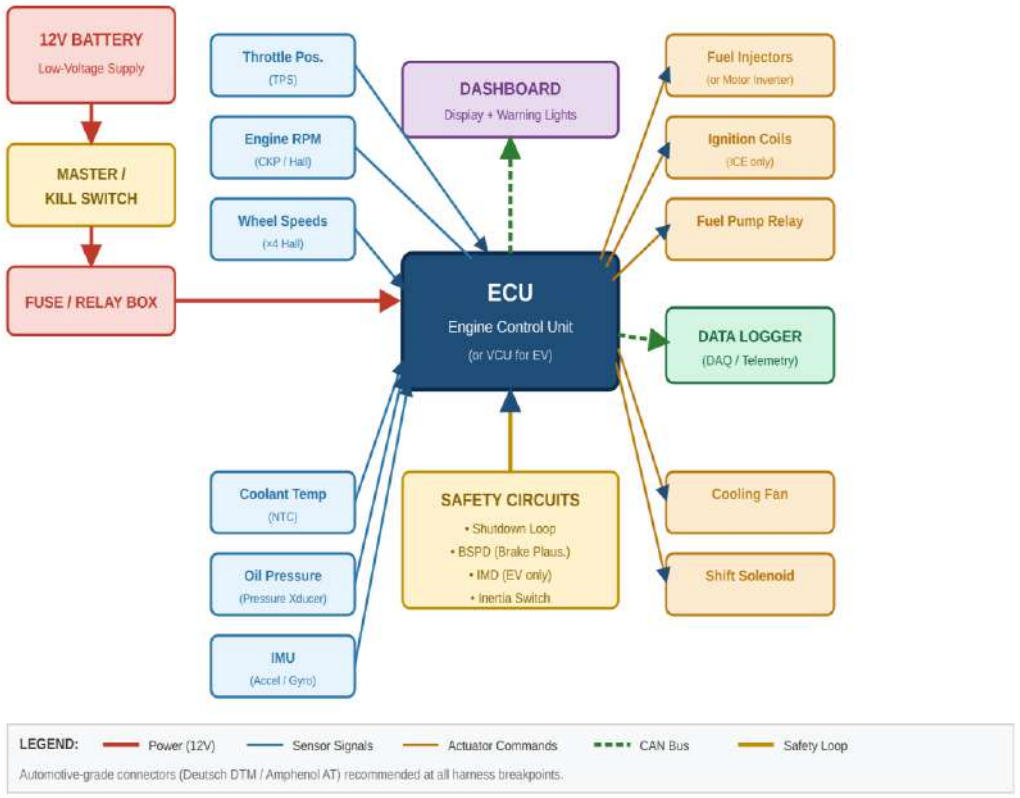


Fig. 4.: Electrical System Architecture Block Diagram, Signal and Power Distribution Overview

4.3 Powertrain Selection: ICE versus Electric

The choice between an internal combustion engine (ICE) and an electric powertrain (EV) is one of the most consequential architectural decisions. Table 3. presents a comparative analysis across key technical and programmatic dimensions.

Table 3.: Comparative Analysis of ICE and Electric Vehicle Powertrain Platforms

Parameter	**ICE Platform**	**EV Platform**
Typical Power Output	60–80 kW (with 20mm restrictor)	Up to 80 kW peak (rules limited)
Torque Delivery	Narrow powerband, requires tuning	Instant, flat torque curve
Typical Vehicle Weight	180–230 kg	200–280 kg (battery mass)
Complexity Level	Moderate (engine tuning, exhaust)	High (HV safety, BMS, inverter)
Budget (Typical)	$15,000–$30,000	$25,000–$50,000+
Safety Requirements	Standard (fuel, rollover)	Extensive HV safety systems
Knowledge Transfer Risk	Lower (widespread expertise)	Higher (specialized HV skills)
Industry Relevance	Declining but still relevant	Growing, high employer demand
Noise at Competition	Must pass noise test (103–110 dBC)	No noise constraint

For first-year teams, the ICE platform generally presents a lower barrier to entry due to the availability of well-documented single-cylinder motorcycle engines (such as the Honda CBR600RR, Yamaha YZF-R6, or KTM 690 for singles), extensive community knowledge, and simpler safety requirements. However, teams at institutions with strong electrical engineering programs may find the EV platform aligns better with their core competencies and provides more relevant industry experience.

5. Project Management and Mentorship

Building a Formula SAE car within a single academic year is an exercise in ruthless time management. The project timeline spans approximately 8–10 months from initial concept to competition day, with the manufacturing, assembly, and testing phases consuming the majority of calendar time. The most frequently cited axiom in the FSAE community encapsulates the priority: a car finished one month early will beat a 'better' car finished one day before competition. This is because the 'finished' car has no time for the critical testing, tuning, and driver training that ultimately determines on-track performance.

5.1 The Master Timeline

A well-structured project timeline divides the academic year into distinct phases with hard deadlines. Fig. 5. presents the recommended project flowchart from concept through competition.

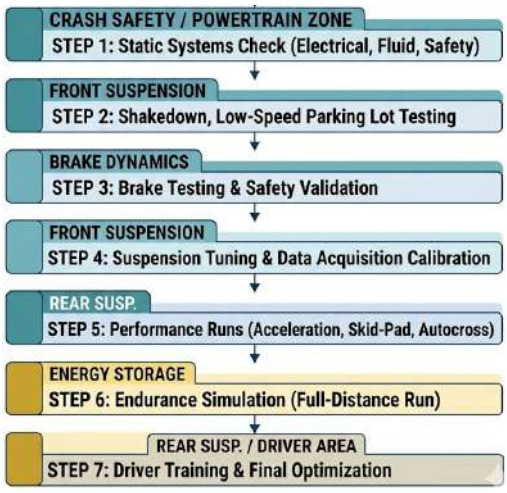


Fig. 5.: Recommended Project Timeline Flowchart, Seven Phases from Concept to Competition

Critical to this timeline is the concept of "design freeze." Once a subsystem enters the manufacturing phase, design changes must be formally reviewed and approved, as changes during manufacturing are exponentially more costly in time and budget than changes made during the design phase. Teams should target a design freeze by the end of month 4, leaving months 5 through 7 for manufacturing and procurement.

5.2 Team Structure and Knowledge Transfer

Effective team structure is as important as effective engineering. The recommended organizational model consists of a project manager or team captain, subsystem leads for each major area (chassis, suspension, powertrain, electronics, aerodynamics, business/cost), and junior members assigned to each subsystem group. Table 4. outlines a recommended team structure.

Table 4.: Recommended Team Organizational Structure and Role Allocation

Role	Responsibility	Typical Team Size	Skill Focus
Team Captain / PM	Overall coordination, timeline	1	Leadership, planning
Chief Engineer	Technical integration	1	Systems engineering
Chassis Lead	Frame design, IA, bodywork	3–4	Structures, welding
Suspension Lead	Kinematics, uprights, hubs	3–4	Dynamics, machining
Powertrain Lead	Engine/motor, drivetrain, cooling	3–4	Thermodynamics, tuning
Electronics Lead	Wiring, ECU, DAQ, safety circuits	2–3	EE, firmware, sensors
Aero Lead	Wings, diffuser, bodywork	2–3	CFD, composites
Business / Cost Lead	Cost report, business plan	2–3	Finance, presentation

Knowledge transfer is a perennial challenge for FSAE teams due to the natural turnover inherent in student organizations. Experienced seniors graduate each year, potentially taking critical institutional knowledge with them. To mitigate this, teams should implement structured mentorship pairing (each senior mentors a specific junior), maintain comprehensive design documentation in a centralized digital repository, and conduct formal design reviews where outgoing members present their rationale and lessons learned.

6. Design For Manufacturability and Reliability

Students, driven by ambition and inspired by Formula 1 technology, naturally gravitate toward the most advanced and lightweight solutions available: carbon fiber monocoques, 5-axis CNC-machined titanium uprights, topology-optimized 3D-printed components, and complex aerodynamic packages with multi-element wings. While such technologies are laudable goals for mature programs, they are frequently inappropriate for teams with limited manufacturing resources, budget constraints, or insufficient composites experience. The Design for Manufacturability and Assembly (DFMA)

philosophy demands that every component be designed not in isolation but in the context of the team's actual fabrication capabilities.

6.1 Capability Assessment Matrix

Before committing to a design approach, teams should conduct an honest assessment of their available manufacturing capabilities. Table 5. provides a capability assessment framework that maps design choices to required resources.

Table 5.: Design for Manufacturability Assessment, Simple vs. Advanced Design Options

Component	Simple Design	Advanced Design	Skill Req.	Risk Level
Frame	Steel spaceframe (TIG)	CF monocoque	High	Very High
Uprights	6061 Al, 3-axis CNC	Ti, topology-optimized	Very High	Very High
A-Arms	Steel tube, welded	CF tube, bonded inserts	High	High
Bodywork	Fiberglass layup	Pre-preg CF, autoclave	High	Moderate
Exhaust	Mild steel, mandrel bent	Inconel, equal length	Very High	Moderate
Wiring	Mil-spec, Deutsch connectors	Custom PCB, CAN bus	High	High
Brakes	OEM calipers, custom rotors	Custom calipers, CF rotors	Very High	Very High

6.2 The Reliability Imperative

Reliability is the ultimate performance metric in FSAE. The mathematical argument is straightforward: completing the endurance event, even at moderate pace, is worth up to 275 points. A DNF yields zero points from endurance and zero from the efficiency event (an additional 100 points). Therefore, a slow but reliable car that finishes endurance can earn up to 375 points from those two events alone, while a theoretically faster car that breaks down earns zero. This 375-point swing dwarfs any advantage that might be gained from shaving seconds off lap times with exotic components.

Design for reliability means selecting proven components where possible, using appropriate safety factors (a minimum of 2.0 for static loads, 3.0 or higher for fatigue-critical components), performing thorough failure mode and effects analysis (FMEA) on every critical subsystem, and designing with inspection and serviceability in mind. Components should be accessible for quick repair, and the team should carry a complete set of spare parts for all known wear items. Fig. 6. presents the standard FMEA process flow applied to FSAE subsystems.

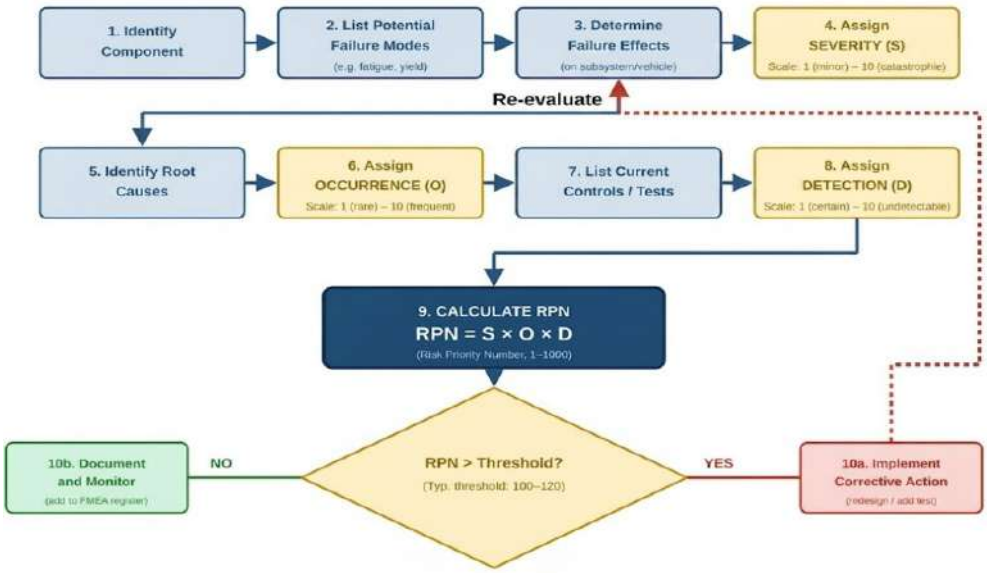


Fig. 6.: Failure Mode and Effects Analysis (FMEA) Process Flow for Critical Vehicle Components

7. Testing, Validation, and Iteration

The assembly of the car marks not the end of the engineering process but its midpoint. The testing phase is where theoretical designs meet physical reality, and it is during this phase that the difference between a competitive car and a display piece becomes most apparent. Teams that allocate insufficient time for testing consistently underperform, regardless of the sophistication of their design.

7.1 Simulation and Analytical Validation

Modern FSAE teams employ a suite of simulation tools throughout the design process. Finite Element Analysis (FEA) is used to validate the structural integrity of the chassis, suspension components, and mounting brackets under both static and dynamic loading conditions. Computational Fluid Dynamics (CFD) is employed by teams running aerodynamic packages to predict downforce, drag, and flow behavior around the vehicle body. Lap simulation software, such as Optimum Lap or custom MATLAB models, enables the team to predict overall vehicle performance and evaluate the sensitivity of key design parameters.

However, simulation is a tool for design guidance, not a substitute for physical testing. FEA results are only as accurate as the boundary conditions and material models used, and CFD predictions are particularly sensitive to mesh quality and turbulence modeling assumptions. All critical simulation results should be correlated against physical test data wherever possible. Fig. 7. illustrates the recommended simulation-to-validation correlation workflow.

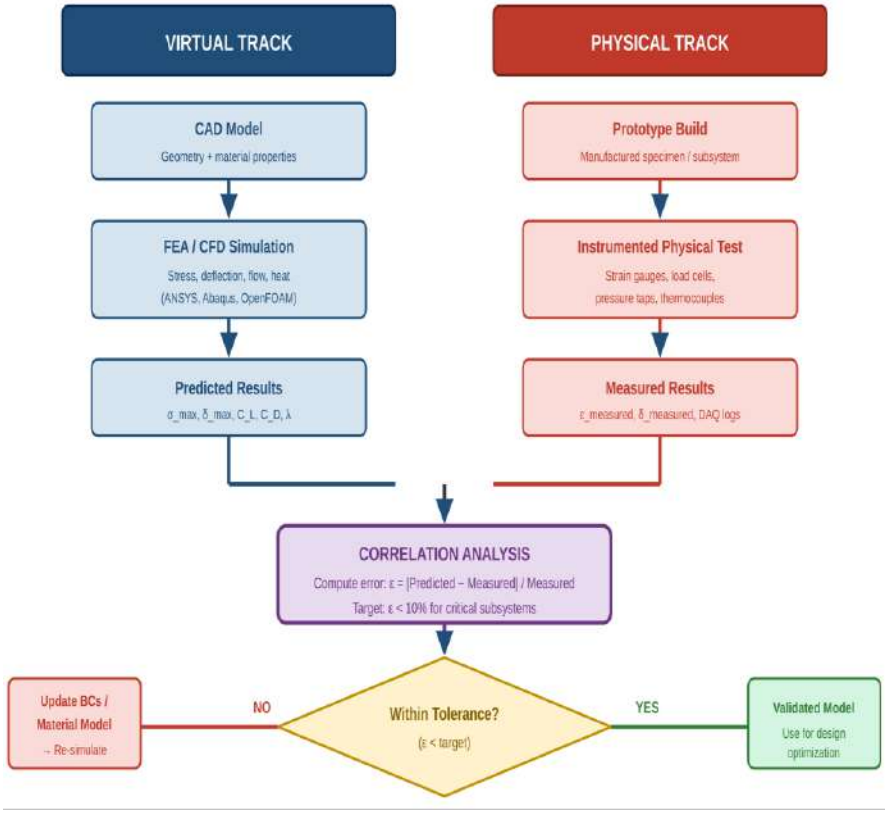


Fig. 7.: Simulation-to-Validation Correlation Workflow, Bridging Virtual and Physical Testing [26, 27]

7.2 Physical Testing Protocol

The physical testing phase should follow a structured protocol that progresses from static checks to low-speed functionality testing and finally to high-speed performance testing. Fig. 8. presents the recommended testing flowchart.

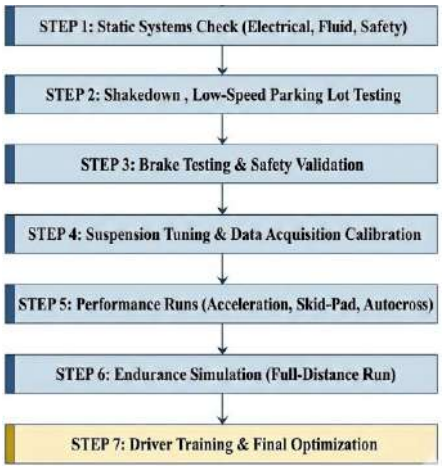


Fig. 8.: Recommended Physical Testing Protocol Flowchart

Each step in the testing protocol serves a specific purpose. The static systems check verifies that all electrical connections are correct, all fluid systems are leak-free, and all safety systems (kill switches,

brake over-travel switches, fire extinguisher) are functional before the car moves under its own power. The low-speed shakedown identifies basic mechanical issues such as clearance problems, binding in the steering or suspension, and overheating. Only after these fundamental checks are passed should the team progress to performance testing.

The endurance simulation is particularly critical. The team should aim to run the full 22-kilometer endurance distance at least three times before competition. Each run should be followed by a thorough inspection of all components, with particular attention to fastener torques, fluid levels, tire wear, and any signs of fatigue or cracking. Issues discovered during endurance simulation testing can be resolved methodically in the shop; the same issues discovered at competition result in a DNF.

7.3 Data-Driven Iteration

The data acquisition system installed on the vehicle should be actively used during testing to drive engineering decisions. Table 6. outlines the minimum recommended sensor suite and the engineering insights each channel provides.

Table 6.: Minimum Recommended Data Acquisition Sensor Suite

Sensor Channel	**Sensor Type**	**Engineering Insight**	**Tuning Impact**
Engine RPM	Hall-effect / inductive	Powerband utilization	Gear ratio selection
Throttle Position	Potentiometer / Hall	Driver input correlation	ECU mapping
Wheel Speed (x4)	Hall-effect	Slip ratio, ABS behavior	Differential tuning
Lateral Acceleration	MEMS accelerometer	Cornering performance	Spring/ARB rates
Long. Acceleration	MEMS accelerometer	Braking / accel. performance	Brake bias
Steering Angle	Rotary potentiometer	Understeer / oversteer	Alignment settings
Damper Potentiometer	Linear potentiometer	Ride height, roll, pitch	Damper settings
Coolant Temperature	NTC thermistor	Cooling system adequacy	Radiator sizing

Oil Pressure	Pressure transducer	Lubrication health	Oil system design

8. Integrated Design Decision Framework

To assist teams in making disciplined, evidence-based design decisions, this section presents a weighted decision matrix methodology. Every significant design choice, from frame material selection to aerodynamic package complexity, should be evaluated against a consistent set of criteria that reflect the team's strategic priorities. Table 7. demonstrates this methodology applied to a frame material selection decision.

Table 7.: Weighted Decision Matrix, Frame Material Selection (Scale 1–10, 10 = Best)

Criterion	Weight	Steel Tube	Al Mono.	CF Mono.	Hybrid
Weight	0.15	5	7	9	8
Stiffness	0.15	6	8	9	7
Mfg. Ease	0.25	9	5	3	6
Cost	0.20	9	5	3	6
Repairability	0.15	9	4	2	5
Timeline Risk	0.10	9	5	3	5
Weighted Total	**1.00**	**7.85**	**5.55**	**4.35**	**6.15**

In this example, the steel spaceframe achieves the highest weighted score (7.85) for a team that prioritizes manufacturability, cost, and repairability, typical priorities for a first- or second-year team. More experienced teams with established composites capabilities and larger budgets might justifiably weigh the performance criteria (weight, stiffness) more heavily, shifting the optimal choice toward an aluminum or carbon fiber monocoque. The methodology itself is more important than the specific scores, which should be calibrated to each team's unique circumstances.

9. Cost and Budget Management

Financial management is an integral part of the FSAE experience and is directly assessed through the Cost and Manufacturing Analysis event (worth 100 points). Teams must submit a detailed cost report documenting the manufacturing cost of every component on the vehicle using SAE's standardized cost

tables. Beyond the competition requirement, effective budget management ensures that the project does not stall due to funding shortfalls during critical procurement phases.

A typical competitive FSAE team operates on a total annual budget of $20,000 to $60,000, depending on the powertrain type, manufacturing ambition, and travel costs. Funding typically comes from a combination of university allocations, corporate sponsorships (both cash and in-kind), fundraising activities, and student organization fees. Teams should aim to secure at least 70% of their projected budget before the manufacturing phase begins to avoid costly delays. Fig. 9. presents a typical FSAE team budget allocation by subsystem category.

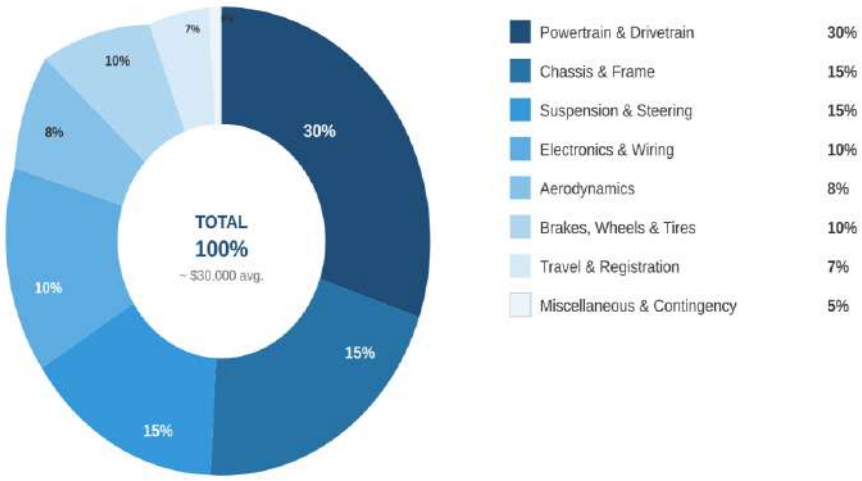


Fig. 9.: Typical FSAE Team Budget Allocation by Subsystem Category (ICE Platform)

10. Safety Philosophy and Culture

Safety is not merely a regulatory checkbox in FSAE; it is a professional and ethical obligation. The vehicles operate in close proximity to other competitors, track workers, and spectators. A culture of safety must permeate every aspect of the team's operations, from workshop practices (mandatory safety glasses, no open-toed shoes, hot-work permits) to design decisions (appropriate safety factors, redundant safety systems) and testing protocols (designated safety personnel, fire extinguishers on-site, driver extraction practice).

For electric vehicles, the safety requirements are significantly more stringent due to the presence of high-voltage systems, typically operating at 300–600 volts DC. The accumulator (battery pack) must be designed with multiple layers of protection, including a Battery Management System (BMS) with cell-level voltage and temperature monitoring, an Insulation Monitoring Device (IMD), a Brake System Plausibility Device (BSPD), and a comprehensive shutdown circuit that can de-energize the tractive system in response to any detected fault condition. The team must demonstrate mastery of high-voltage safety procedures and carry appropriate personal protective equipment (PPE) rated for the system voltage. Fig. 10. illustrates the architecture of the EV shutdown circuit. The team must

demonstrate mastery of high-voltage safety procedures and carry appropriate personal protective equipment (PPE) rated for the system voltage.

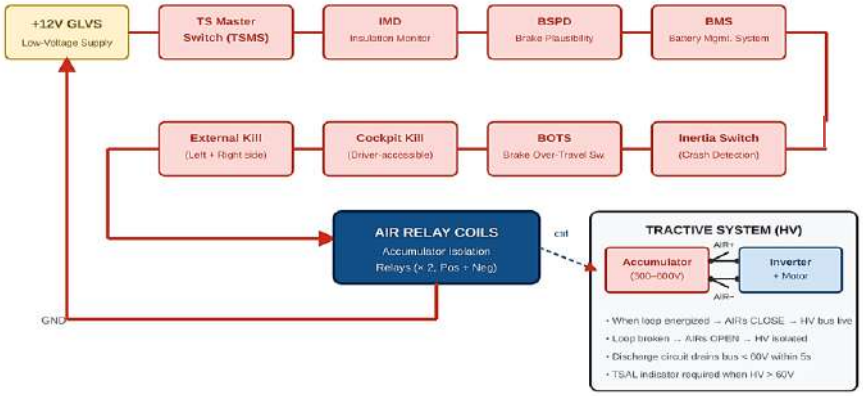


Fig. 10.: Proposed workflow: Multiscale data-driven system architecture.

11. Implementation and Fabrication

The implementation and fabrication phase represents the transition from theoretical design to a fully realized physical vehicle, requiring disciplined coordination across all subsystems. Once the design freeze is established and the decision matrices are finalized, fabrication activities must follow a structured workflow to ensure dimensional accuracy, manufacturability, and subsystem compatibility. As illustrated in Figure 11, the recommended fabrication workflow progresses through material preparation, jig and fixture setup, subsystem-level manufacturing, intermediate quality inspections, and final assembly integration. This staged approach minimizes rework, reduces timeline risk, and ensures that each subsystem is validated before being incorporated into the full vehicle.

During fabrication, critical components such as the chassis, suspension hardpoints, uprights, wiring harness, and cooling system must be produced with strict adherence to the tolerances defined during the CAD and FEA stages. Welding fixtures and alignment jigs should be used to maintain geometric accuracy, particularly for the frame and suspension pickup points, where small deviations can significantly affect vehicle dynamics. Intermediate inspections—such as tube notching verification, weld penetration checks, and dimensional conformity assessments—should be documented throughout the process to maintain traceability and quality assurance.

Subsystem integration should follow a bottom-up sequence, beginning with the rolling chassis, followed by drivetrain installation, electrical system routing, and final bodywork and aerodynamic components. This incremental integration strategy, reflected in the workflow of Figure 11., allows the team to identify interface issues early and resolve them before full assembly. By adhering to a structured implementation methodology, teams can significantly improve reliability, reduce assembly errors, and increase the likelihood of passing technical inspection on the first attempt.

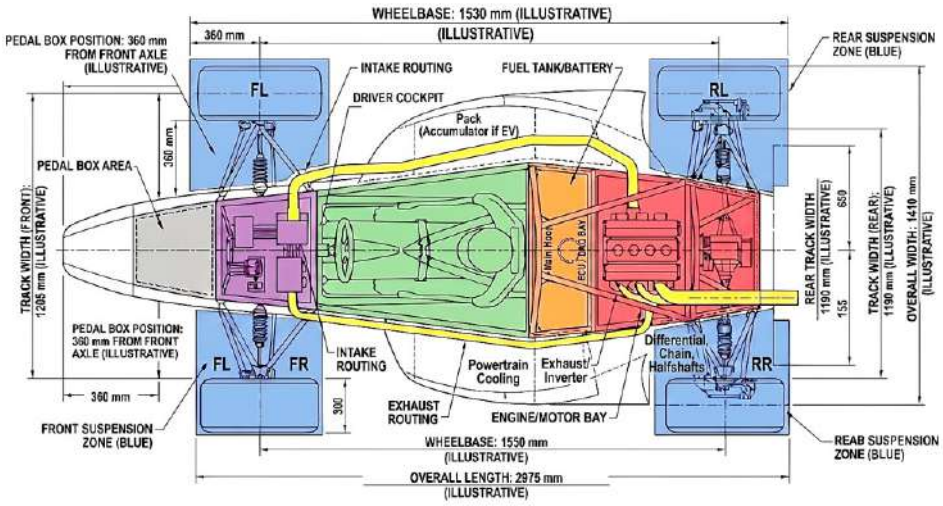


Fig. 11.: Formula SAE Car Layout - Top-Down Technical Schematic.

12. Conclusion

The Formula SAE project stands as one of the most demanding and rewarding capstone engineering experiences available in higher education. This paper has presented a comprehensive framework organized around five critical pillars: strict rulebook adherence, systems engineering integration, disciplined project management, design for manufacturability and reliability, and rigorous testing and validation. The evidence from decades of FSAE competition history consistently supports the same conclusion: the teams that win is not necessarily those with the most advanced or lightest cars, but those that build reliable, well-tested vehicles and prepare their drivers and presenters thoroughly for every event.

The tables, flowcharts, and decision frameworks presented in this paper are intended as practical tools that teams can directly adopt or adapt to their specific programs. Whether pursuing an ICE or EV platform, the fundamental principles remain the same: understand your constraints (the rulebook), design as an integrated system, manage your time ruthlessly, build what you can manufacture, and test relentlessly. A car that runs reliably will always outperform one that sits in the paddock, no matter how beautiful it looks on paper.

While this comprehensive framework provides a robust roadmap for Formula SAE success, several limitations must be acknowledged regarding its universal applicability. First, the framework's workflows—particularly those involving extensive simulation-to-validation correlation and multi-stage physical testing—assume a baseline level of financial and infrastructural resources that may be unattainable for first-year or highly resource-constrained programs. For such teams, strict adherence to these rigorous validation protocols may need to be heavily scaled down to align with their realistic manufacturing and budgetary constraints. Furthermore, while the core philosophy of rulebook

compliance is universal, the specific architectural and safety examples discussed are primarily influenced by the North American SAE rulebook. Teams competing internationally must carefully adapt this framework to account for regional regulatory variations, such as the distinct structural, electrical, and autonomous requirements mandated by the Formula Student Germany (FSG) regulations. Finally, the framework relies on a baseline of institutional support for knowledge transfer and mentorship, a variable that remains highly volatile and difficult to standardize across different university environments.

Future work in this area could explore quantitative correlations between testing hours and competition placement, develop standardized knowledge transfer frameworks for student organizations, and investigate the application of model-based systems engineering (MBSE) tools to the FSAE design process.

References

[1] Ashby, M. F. (2017). *Materials selection in mechanical design* (5th ed.). Butterworth-Heinemann.

[2] Barbosa, L. M. Q., Fernandes, G. H. N., & de Queiroz Murad, M. (2024). "Finite element analysis and optimization of a Formula SAE car chassis". *Caderno Pedagógico, 21*(1), 2053-2071. https://doi.org/10.54033/cadpedv21n1-107

[3] Boothroyd, G., Dewhurst, P., & Knight, W. A. (2011). *Product design for manufacture and assembly* (3rd ed.). CRC Press.

[4] Carollo, F., Mariotti, G. V., Golfo, S., & Pappalardo, A. (2022). "Dynamic tests of Formula SAE car bodies". *WSEAS Transactions on Systems, 21*, 104-114. https://doi.org/10.37394/23202.2022.21.12

[5] Carvalho, D. T., & Melo, C. P. (2022). "Simulation model and testing of a Formula SAE brake system". *15th SAE Brasil International Brake and Motion Control Colloquium & Engineering Exhibition*. SAE Technical Paper. https://doi.org/10.4271/2021-36-0430

[6] Deakin, A., Crolla, D., Ramirez, J. P., & Hanley, R. (2000). The effect of chassis stiffness on race car handling balance. *Proceedings of the SAE Motorsports Engineering Conference & Exposition*. https://doi.org/10.4271/2000-01-3554

[7] Dixon, J. C. (2009). *Suspension geometry and computation*. John Wiley & Sons.

[8] Formula Student Germany. (2025). *Formula Student Rules 2025*. FSG.

[9] Gillespie, T. D. (1992). *Fundamentals of vehicle dynamics*. SAE International.

[10] Guo, R., Zeng, J., Yang, Y., Hou, Y., Zhu, Z., & Xiong, J. (2025). "A design of electric drive system for all-wheel drive Formula Student racing car to improve racing performance". *WCX SAE World Congress Experience*. SAE Technical Paper. https://doi.org/10.4271/2025-01-8005

[11] Gupta, E., Bora, D. K. S., & Arunachalam, R. (2022). "Design and analysis of brake system for FSAE race car". *Engineering Research Express, 4*(2), 025039. https://doi.org/10.1088/2631-8695/ac6ecd

[12] Hibbeler, R. C. (2016). *Engineering mechanics: Statics and dynamics* (14th ed.). Pearson.

[13] Kaisare, S. S. (2024). *Parametric design and optimization of an upright of a Formula SAE car* [Doctoral dissertation, Virginia Tech].

[14] Lin, M. T., & Papadopoulos, P. (2022). "Race car aerodynamics simulation and transmitted tire reaction forces". *International Journal of Automotive Engineering, 13*(4), 169-176. https://doi.org/10.20485/jsaeijae.13.4_169

[15] Milliken, W. F., & Milliken, D. L. (1995). *Race car vehicle dynamics*. SAE International.

[16] Pourghavam, M., Abrishami, M., Ahmadabadi, S., & Mohajerani, Z. (2022). "Mathematical modeling of a Formula SAE student racecar passive suspension system using two different methods". *The Sixth International Conference on Science and Technology of Electrical, Computer and Mechanical Engineering of Iran*. https://hal.science/hal-03650106/document

[17] Raspanti, S., Ciampolini, M., Bigalli, S., Fabiani, A., Romani, L., & Ferrara, G. (2022). Numerical investigation on the effects of the setting of the load control system of a Formula SAE single-cylinder turbocharged engine on fuel efficiency and performance. *Journal of Physics: Conference Series, 2385*(1), 012081. https://doi.org/10.1088/1742-6596/2385/1/012081

[18] Rigo, C. S. S., Neto, A. D. R. D. F., Grandinetti, F. J., Castro, T. S., Dias, E. X., & Martins, M. S. (2024). "Development of the steering system for a Formula SAE prototype". *SAE Brasil 2024 Congress*. SAE Technical Paper. https://doi.org/10.4271/2024-36-0099

[19] SAE International. (2025a). *Formula SAE Rules 2025*.

[20] SAE International. (2025b). *FSAE Electric Rules Supplement*.

[21] Shaalan, A., Assanis, D., Raman, A., Wijeyakulasuriya, S., & Senecal, K. (2024). "Formula 1 race car aerodynamics: Understanding floor flow structures and why it is a key component in modern racing". *WCX SAE World Congress Experience*. SAE Technical Paper. https://doi.org/10.4271/2024-01-2078

[22] Smith, C. (1978). *Tune to win*. Aero Publishers.

[23] Suvanjumrat, C., Namchanthra, S., Phengpom, T., Priyadumkol, J., Chookaew, W., Watechagit, S., Cheung, S. C., & Promtong, M. (2025). "AI-assisted CFD optimisation of multi-element wing angle of attack for enhanced Formula SAE aerodynamic performance". *International Journal of Thermofluids*. https://doi.org/101440. 10.1016/j.ijft.2025.101440

[24] Ullman, D. G. (2017). *The mechanical design process* (6th ed.). McGraw-Hill.

[25] Wang, L., Li, C., Zhu, X., & Liu, M. (2025). "Optimization design of Formula E racing car transmission and wheel system". (SAE Technical Paper No. 2025-01-8505). https://doi.org/10.4271/2025-01-8505

[26] Wikarta, A., Kaelani, Y., Taruna, F. A., Indwindra, B. H. A., & Trengginas, L. A. (2024). "Fatigue life and topology optimization of racing car upright for Formula SAE electric". *Automotive Experiences, 7*(2), 333-342. https://doi.org/10.31603/ae.11440

[27] https://www.sae.org/standards

Islamic University Journal of Applied Sciences (IUJAS)

https://journals.iu.edu.sa/jesc

Volume VIII, Issue II, December 2026, Pages 95-126

Smart Insect-Based Bio-Waste Management for Desert Ecosystems: A Review and Integrated Framework

Norhan Shoghy [1]*, Tamer Ismail [2] and Reda Ibrahim [1]

[1]Economic Entomology Department, Kafrelsheikh University, Kafrelsheikh 33516, Egypt
[2]Pesticide Chemistry and Toxicology Department, Kafrelsheikh University 33516, Egypt.

*Corresponding author: Norhan Shoghy, *Email Address*: Norhan_Shoghy@agr.kfs.edu.eg

Abstract

Desert ecosystems face major bio-waste management challenges due to high temperatures, water scarcity, and limited microbial activity, which reduce the efficiency of conventional treatment methods. In arid regions, bio-waste represents a large proportion of municipal solid waste, while landfilling and composting remain environmentally constrained. Insect-based bioconversion, particularly using *Hermetia illucens* larvae, has emerged as a promising alternative capable of reducing organic waste by 65–80% within 7–14 days while producing protein-rich biomass and nutrient-rich fertilizer. This review summarizes recent advances in insect-mediated bio-waste valorisation and highlights the integration of artificial intelligence (AI), Internet of Things (IoT), and automation for real-time monitoring and process optimization. A conceptual smart bio-waste management framework for arid environments is proposed, emphasizing climate adaptability, resource efficiency, and scalability. The proposed system offers rapid processing, lower greenhouse gas emissions, and minimal water requirements, supporting sustainable waste management in desert ecosystems.

Keywords

Black soldier fly; Insect-based bioconversion; Bio-waste valorisation; Smart waste management; Artificial intelligence; Desert ecosystems.

https://doi.org/10.63070/jesc.2026.036
Received 30 April 2026; Revised 03 June 2026; Accepted 10 June 2026; Available online 15 July 2026.
Published by Islamic University of Madinah on behalf of ***Islamic University Journal of Applied Sciences.***
This is an open access article published under the Creative Commons Attribution 4.0 International license (CC BY 4.0): https://creativecommons.org/licenses/by/4.0/

1. Introduction

Desert ecosystems constitute nearly one-third of the Earth's terrestrial surface and are characterized by extreme conditions, including high temperatures, low precipitation, and limited biological productivity. These constraints significantly slow natural decomposition processes, resulting in inefficient recycling of organic matter and the accumulation of bio-waste. Rapid population growth, urbanization, and agricultural expansion further intensify bio-waste generation in these regions, creating substantial environmental challenges. In parallel, global efforts toward sustainable resource utilization have emphasized the potential of biological systems, particularly insects, as efficient agents for converting bio-waste into valuable products within circular bioeconomy frameworks [1].

Among emerging biological approaches, insect-based bioconversion has gained considerable attention as an effective and environmentally sustainable solution for bio-waste management. This process relies on saprophagous insects to transform organic substrates into high-value outputs, including protein-rich biomass and nutrient-dense fertilizers. In this review, the term 'bio-waste' refers specifically to biodegradable organic materials of biological origin, including food waste, agricultural residues, and animal manure, excluding synthetic organic compounds such as plastics and polymers. The black soldier fly, *Hermetia illucens*, is one of the most widely studied species due to its high waste reduction efficiency, rapid larval development, and adaptability to diverse substrates [2], [3], [5]. Experimental and applied studies have demonstrated that its larvae can reduce bio-waste by approximately 65–80% within a short period, while producing biomass suitable for applications in animal feed and agriculture [3], [12], [22].

Despite these advances, bio-waste management remains a major challenge in dryland regions. Bio-waste constitutes a large proportion of municipal solid waste, often exceeding 44–55% of the total waste stream globally. Conventional approaches, including landfilling and composting, are often inefficient under desert conditions. Landfills promote anaerobic decomposition processes that generate methane and contribute to greenhouse gas emissions, while composting requires controlled moisture and temperature conditions that are difficult to maintain in water-scarce environments. These limitations reduce the effectiveness of traditional waste treatment systems and increase environmental pressure in dryland regions [6], [8].

Recent technological advancements have enhanced the efficiency and scalability of insect-based systems through the integration of smart technologies such as artificial intelligence (AI), Internet of Things (IoT), and automated control systems. These technologies enable real-time monitoring and optimization of critical parameters, including temperature, humidity, substrate conditions, and larval growth dynamics. AI-driven models support predictive decision-making by optimizing feeding strategies, improving system performance, and enabling adaptive responses to environmental variability. The integration of digital technologies in agriculture and waste management has demonstrated strong potential for improving operational efficiency and resource utilization [13]–[16].

The concept of smart insect-based bio-waste management represents an innovative interdisciplinary approach that combines biological processes with advanced digital technologies to achieve efficient and sustainable waste valorisation. This approach aligns with circular bioeconomy principles by transforming organic waste into valuable resources while minimizing environmental impacts and greenhouse gas emissions [17], [18].

However, a comprehensive integration of insect bioconversion with AI-driven smart systems tailored specifically for desert ecosystems remains insufficiently explored in the literature.

This review aims to (i) examine recent advances in insect-based bioconversion technologies, (ii) evaluate the role of artificial intelligence and Internet of Things in optimizing system performance, and (iii) propose an integrated, climate-adaptive framework for implementing smart bio-waste management systems in desert environments. By addressing these aspects, this work contributes to the development of sustainable, resource-efficient, and climate-resilient waste management strategies.

2. Bio-Waste Challenges in Desert Ecosystems

2.1 Environmental Constraints Affecting Waste Decomposition

Arid regions are characterized by extreme climatic conditions that significantly hinder the natural degradation of bio-waste. High temperatures, low humidity, and limited water availability restrict microbial activity, which is essential for conventional biodegradation processes. Microorganisms responsible for organic matter decomposition require adequate moisture and relatively stable conditions; however, these requirements are rarely met in desert environments, resulting in slow and inefficient decomposition rates [6], [8].

In addition, intense solar radiation and large diurnal temperature fluctuations further disrupt biological processes and destabilize decomposition pathways. These harsh conditions reduce the effectiveness of traditional composting systems and often lead to incomplete degradation of bio-waste. As a consequence, partially decomposed materials may generate unpleasant odours and attract pests, while nutrient recycling is significantly constrained, negatively affecting soil fertility and ecosystem productivity [6]. The combined impact of these factors creates substantial operational challenges for bio-waste management systems, as summarized in Table 1.

Table 1. Key Challenges of Bio-Waste Management in Desert Ecosystems

Challenge	Description	Environmental impact
Water scarcity	Limited moisture availability for microbial activity	Slow decomposition rates
High temperature	Extreme heat and large diurnal fluctuations	Process instability
Low microbial diversity	Reduced biological degradation capacity	Waste accumulation
Poor infrastructure	Inadequate waste collection and treatment systems	Environmental pollution
High organic fraction	Large proportion of biodegradable waste	Increased management burden

Synthesized from [6], [8].

2.2 Composition and Characteristics of Bio-Waste in Arid Regions

Bio-waste in dryland environments primarily consists of organic materials, including food waste, agricultural residues, and animal manure. Studies indicate that the organic fraction of municipal solid waste in many developing and arid regions can exceed 50–60% of the total waste stream [6], [18]. Despite this high organic content, limited moisture availability and reduced microbial diversity restrict effective decomposition, resulting in prolonged accumulation of waste materials.

Agricultural activities in desert regions also generate significant quantities of lignocellulosic residues, which are resistant to degradation due to their complex structural composition. Without appropriate treatment technologies, these residues contribute to environmental pollution and inefficient resource utilization. Furthermore, improper handling of bio-waste increases the risk of pathogen proliferation and pest infestations, posing additional challenges to waste management and public health [8].

2.3 Limitations of Conventional Waste Management Approaches

Traditional waste management methods are generally ineffective in dryland environments due to environmental and operational constraints. Landfilling, one of the most widely used approaches, presents several environmental risks, including methane emissions, soil contamination, and inefficient land use. In arid climates, reduced moisture availability limits microbial activity within landfills, leading to prolonged waste persistence and increased environmental burden [6].

Composting, although environmentally preferable in many contexts, faces significant limitations in desert regions. The process requires adequate moisture and controlled conditions, both of which are difficult to maintain in water-scarce environments. As a result, composting systems may suffer from incomplete degradation and reduced product quality. Additionally, maintaining optimal temperature and aeration can be technically challenging and economically costly [6], [8]. Incineration represents another alternative; however, it is energy-intensive and often not economically feasible in many developing desert regions. Moreover, it raises concerns regarding air pollution and greenhouse gas emissions, particularly when waste segregation is inadequate. These limitations highlight the need for alternative approaches better suited to arid conditions [6].

2.4 Environmental and Ecological Impacts

The accumulation of untreated bio-waste in arid regions has significant environmental and ecological consequences. One of the most critical impacts is the emission of greenhouse gases, particularly methane, from anaerobic decomposition processes occurring in poorly managed waste systems. Methane has a substantially higher global warming potential than carbon dioxide, contributing significantly to climate change [10]. The progression from improper waste collection to long-term ecological degradation is visually represented in the bio-waste flow diagram in Figure 1.

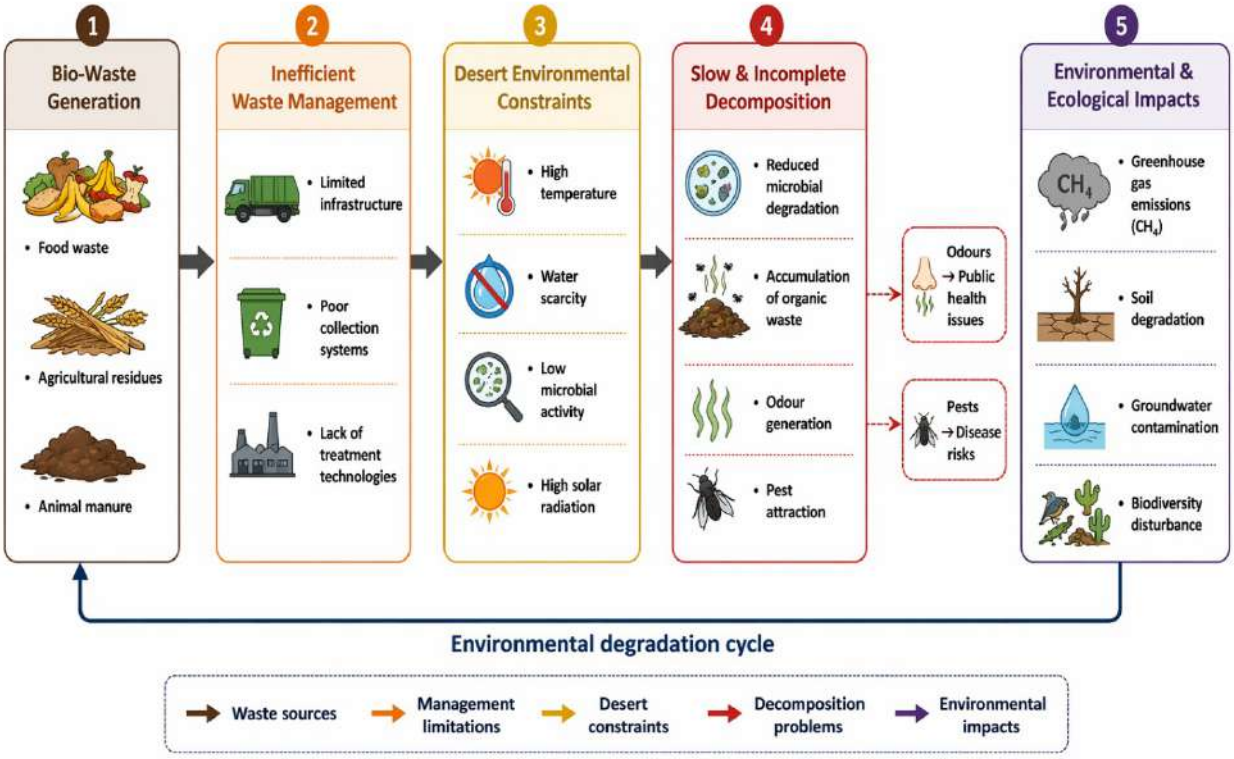


Figure 1. Bio-waste flow and associated environmental impacts in desert ecosystems. The diagram illustrates the progression from waste generation to environmental consequences under inefficient management conditions.

In addition, the lack of efficient organic matter recycling leads to soil degradation and declining fertility. Desert soils are typically low in organic carbon, and the failure to reintegrate organic waste into the soil exacerbates this limitation, reducing agricultural productivity and ecosystem resilience [6]. Improper waste management also affects biodiversity by altering habitats and promoting the proliferation of opportunistic pest species. Furthermore, leachates from waste accumulation sites can contaminate scarce groundwater resources, posing risks to environmental sustainability and human health [6].

2.5 Need for Innovative and Adaptive Solutions

Given the limitations of conventional waste management methods, there is a pressing need for innovative and adaptive solutions specifically designed for desert environments. Effective systems should minimize water consumption, tolerate high temperatures, enable rapid waste processing, and ensure environmental sustainability. In this context, insect-based bioconversion systems offer a promising alternative, as they rely on biological agents capable of adapting to harsh conditions while efficiently converting bio-waste into valuable products. The integration of such systems with smart

technologies, including artificial intelligence and IoT-based monitoring, further enhances their potential to address the complex challenges associated with bio-waste management in dryland environments [3], [13].

3. Insect-Based Bioconversion Technologies

3.1 Concept and Biological Basis of Insect Bioconversion

Insect-based bioconversion is a biological process in which saprophagous insects transform organic waste into valuable biomass and nutrient-rich residues through feeding, digestion, and assimilation. This process relies on the synergistic interaction between insect digestive enzymes and associated gut microbiota, enabling the efficient breakdown of complex organic substrates such as proteins, lipids, and carbohydrates [3], [7].

Unlike conventional microbial composting, insect-mediated decomposition is generally faster and less dependent on strictly controlled environmental conditions. Larvae actively consume and fragment bio-waste, increasing its surface area and enhancing microbial accessibility. This combined mechanical and biochemical action accelerates waste degradation while reducing greenhouse gas emissions compared to traditional waste treatment methods [3], [9].

3.2 Key Insect Species for Bioconversion

Several insect species have been investigated for their potential in organic waste bioconversion. However, their efficiency varies depending on environmental adaptability, substrate compatibility, growth rate, and biomass yield.

3.2.1 Black Soldier Fly, *Hermetia illucens*

The black soldier fly (BSF) is the most widely studied and commercially utilized insect species for bio-waste conversion (figure 2). Its larvae exhibit high feed conversion efficiency and rapid growth, typically completing the larval stage within 7–14 days. They are capable of reducing organic waste by approximately 65–80%, while producing protein-rich biomass containing about 40–45% protein and 30–35% lipids. In addition, BSF larvae can process a wide range of substrates, including food waste, animal manure, and agricultural residues, making them highly suitable for applications in desert environments [2], [3], [4], [12].



Figure 2. Life cycle of the black soldier fly *H. illucens* used in bio-waste bioconversion

3.2.2 Mealworms, *Tenebrio molitor*

Mealworms are effective in degrading dry organic materials, particularly lignocellulosic residues (figure 3). They are well adapted to low-moisture substrates and contribute to the breakdown of otherwise recalcitrant organic matter. However, their relatively slower growth rate and moderate waste reduction efficiency limit their suitability for large-scale applications compared to black soldier fly larvae [3].

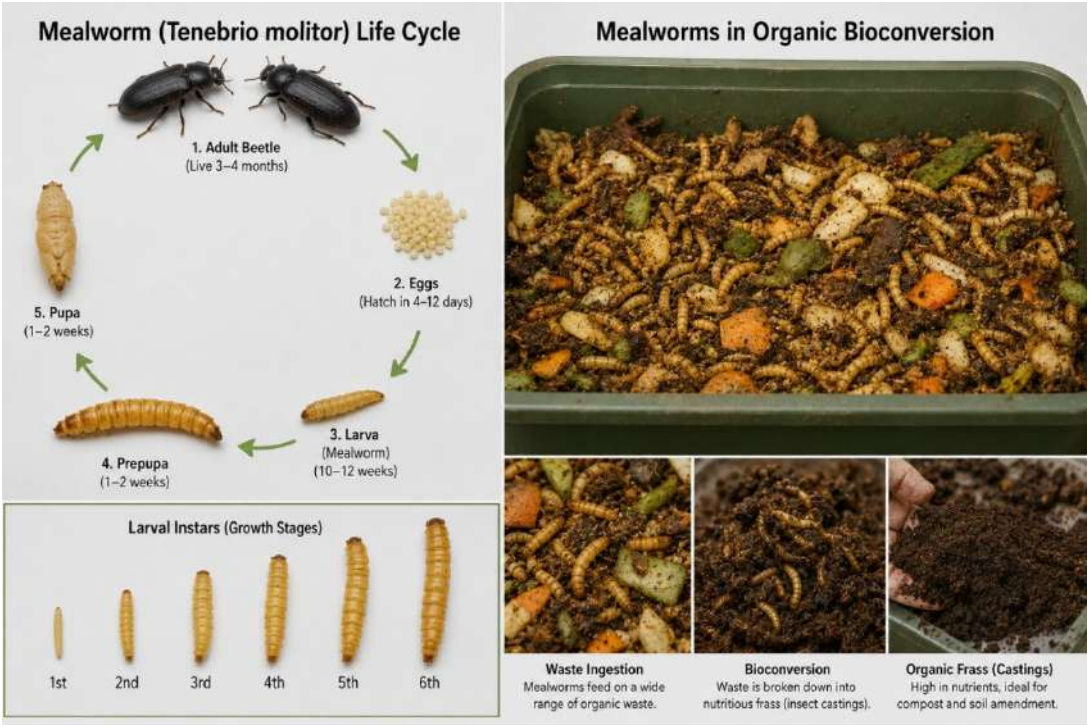


Figure 3. Developmental stages of mealworms *T. molitor* used in insect-based waste conversion systems.

3.2.3 Housefly Larvae, *Musca domestica*

Housefly larvae are characterized by rapid waste consumption and fast development. However, their application in bio-waste management systems is limited due to several drawbacks, including a higher risk of pathogen transmission, lower tolerance to environmental variability, and regulatory concerns. Despite high consumption rates, their suitability is limited due to hygiene and regulatory concerns. Consequently, their use is generally restricted to controlled and well-managed environments [8].

3.3 Bioconversion Efficiency and Performance Metrics

The performance of insect-based bio-waste management systems is evaluated using key indicators such as waste reduction rate, feed conversion ratio, biomass yield, and nutrient recovery efficiency (Table 2). These metrics provide a comprehensive assessment of system effectiveness and sustainability. Among the studied species, black soldier fly larvae consistently demonstrate superior performance across these indicators, particularly under variable environmental conditions, highlighting their suitability for large-scale applications in challenging environments such as desert ecosystems [10], [11], [12].

Table 2. Comparative Performance of Insect Species in Bio-Waste Conversion

Species	Waste reduction (%)	Protein content (%)	Processing time	Suitability for desert conditions
Black soldier fly	65–80	40–45	7–14 days	High
Mealworms	50–60	30–40	20–30 days	Moderate
Housefly larvae	60–70	35–45	5–10 days	Low

3.4 Mechanism of Insect-Based Waste Conversion

The insect-based bioconversion process is a multi-stage system in which bio-waste is efficiently transformed into valuable products through coordinated biological processes. The process begins with ingestion, during which larvae consume organic waste materials. This is followed by mechanical breakdown, where the physical fragmentation of substrates increases surface area and facilitates further degradation.

Enzymatic digestion then plays a central role, as enzymes such as proteases, lipases, and cellulases break down complex organic compounds into simpler molecules. This process is further enhanced by microbial interactions within the larval gut, where symbiotic microorganisms contribute to nutrient breakdown and assimilation. During the assimilation stage, digested nutrients are converted into larval biomass, resulting in the accumulation of protein and lipid reserves. The remaining undigested material is excreted as frass, a nutrient-rich residue that can be used as an organic fertilizer. This integrated mechanism ensures efficient nutrient recycling, substantial waste reduction, and minimal residual output, making it a highly effective strategy for sustainable bio-waste management [3], [7].

3.5 Environmental Adaptability to Desert Conditions

One of the major advantages of the proposed systems, particularly those utilizing black soldier fly larvae, is their adaptability to harsh environmental conditions. Studies have shown that BSF larvae can tolerate a wide range of temperatures, typically between 20°C and 40°C, as well as variable moisture conditions and fluctuating organic load levels. This resilience makes them highly suitable for deployment in arid regions, where conventional biological treatment systems often fail due to environmental stress [7], [10].

3.6 Advantages Over Conventional Systems

Compared to traditional waste treatment methods, insect-based bioconversion offers several important advantages that enhance its suitability for sustainable waste management. One of the most significant benefits is the rapid processing time, as organic waste can be converted within days rather than weeks or months required by composting or landfilling. In addition, these systems require relatively low amounts of water, making them particularly suitable for arid and water-scarce environments.

Another key advantage is the reduction of greenhouse gas emissions, as insect-based processes typically operate under aerobic conditions and generate fewer emissions compared to landfill systems. Furthermore, the process produces valuable by-products, including protein-rich insect biomass and nutrient-dense frass that can be used as organic fertilizers. The scalability and modularity of this integrated framework allow for flexible deployment across different operational scales, from small decentralized units to large industrial facilities. These advantages collectively position insect bioconversion as a key component of sustainable and climate-resilient waste management strategies [3], [9], [12].

3.7 Integration Potential with Smart Systems

Insect-based bioconversion systems are highly compatible with modern digital technologies. Their biological processes can be effectively monitored and optimized using artificial intelligence-based predictive models, IoT-enabled environmental sensors, and automated feeding and harvesting systems. The integration of these technologies enables real-time monitoring, improved process control, and enhanced operational efficiency. Moreover, smart system integration reduces labour requirements and facilitates large-scale deployment within advanced waste management infrastructures, contributing to the development of intelligent and adaptive bio-waste management systems [13]–[16].

4. Smart Systems Integration (AI, IoT, and Automation)

4.1 Overview of Smart Bio-Waste Management Systems

The integration of smart technologies into bio-waste management has transformed conventional treatment approaches into intelligent, adaptive, and data-driven systems. In the context of insect-based bioconversion, the incorporation of artificial intelligence (AI), Internet of Things (IoT), and automation technologies enables real-time monitoring, process optimization, and data-driven decision-making under dynamic environmental conditions. These advancements are particularly important in dryland environments, where environmental variability and resource constraints require precise and

efficient system control [13]–[16]. The proposed system relies on continuous data acquisition, predictive analytics, and automated responses to maintain optimal conditions for larval growth and waste conversion. By leveraging these technologies, system efficiency can be enhanced, operational costs reduced, and scalability improved for large-scale applications.

4.2 Internet of Things (IoT) for Real-Time Monitoring

The Internet of Things (IoT) plays a fundamental role in enabling real-time monitoring of environmental and operational parameters within insect bioconversion systems. Sensor networks are deployed to continuously collect data on temperature, relative humidity, substrate moisture content, gas emissions such as carbon dioxide and ammonia, and larval density and activity. These sensors transmit data to centralized platforms for analysis and control, allowing for precise management of system conditions. Maintaining optimal environmental conditions is essential for maximizing larval efficiency and preventing system failure. For instance, deviations in temperature beyond optimal thresholds can significantly reduce larval metabolism and waste conversion rates. IoT-based systems also support remote monitoring, enabling operators to manage facilities across geographically dispersed desert regions with minimal human intervention [13], [16].

4.3 Artificial Intelligence for Process Optimization

Artificial intelligence plays a pivotal role in enhancing the performance and efficiency of insect-based bio-waste management systems by enabling predictive modelling, pattern recognition, and automated decision-making. By processing large volumes of data generated from IoT-based sensor networks, AI algorithms—including advanced machine learning models such as CNNs (image analysis) and LSTM (time-series prediction)—can identify operational patterns and optimize key parameters in real time. Advanced applications such as computer vision, utilizing Convolutional Neural Networks (CNNs), allow for the continuous monitoring of larval density, size distribution, and health status through high-resolution image analysis [14].

Furthermore, predictive analytics driven by Recurrent Neural Networks (RNNs) or Long Short-Term Memory (LSTM) models facilitate accurate forecasting of feeding regimes and growth cycles by accounting for temporal data dependencies. In addition, anomaly detection models, such as Isolation Forests or Autoencoders, can identify system irregularities—including temperature fluctuations, contamination risks, or process inefficiencies—enabling timely corrective actions.

Reinforcement Learning (RL) and genetic optimization algorithms further contribute by dynamically adjusting environmental conditions such as temperature, humidity, and substrate composition to maximize bioconversion efficiency. The integration of these specific AI architectures into agricultural and waste management systems has been shown to significantly improve operational efficiency and resource utilization, highlighting their potential in enhancing insect-based bioconversion processes [13]–[15]. However, the practical implementation of these advanced models depends on data availability, computational resources, and system scale.

4.4 Automation and Control Systems

Automation is essential for scaling insect-based bioconversion systems and reducing labour dependency. Automated systems can regulate key operational parameters, including feeding schedules, moisture levels, temperature control, and larval harvesting. Programmable Logic Controllers (PLCs) and embedded systems are commonly used to execute automated operations based on sensor inputs and AI-driven recommendations. These systems ensure consistent performance, minimize human error, and improve overall process reliability. In desert environments, automation is particularly advantageous due to limited labour availability and harsh working conditions. Autonomous systems can operate continuously with minimal supervision, thereby enhancing system efficiency and long-term sustainability.

4.5 Integrated Smart System Architecture

A fully integrated framework consists of multiple interconnected layers that enable efficient monitoring, data processing, and automated control (figure 4). At the foundation, the data acquisition layer employs IoT-based sensors to collect environmental and biological data, including temperature, humidity, gas emissions, and larval activity. This information is transmitted through the communication layer using wireless technologies such as LoRa or Wi-Fi, ensuring reliable data transfer even in remote desert locations. The processing layer, supported by artificial intelligence and cloud computing, analyses incoming data to generate predictive models and optimize operational parameters in real time. Based on these insights, the control layer activates automated systems and actuators to regulate feeding, environmental conditions, and harvesting processes. Finally, the user interface layer provides an interactive dashboard that enables operators to monitor system performance, visualize data, and make informed decisions. This multi-layered architecture ensures a responsive, scalable, and intelligent system capable of operating efficiently under challenging environmental conditions [13]–[16].

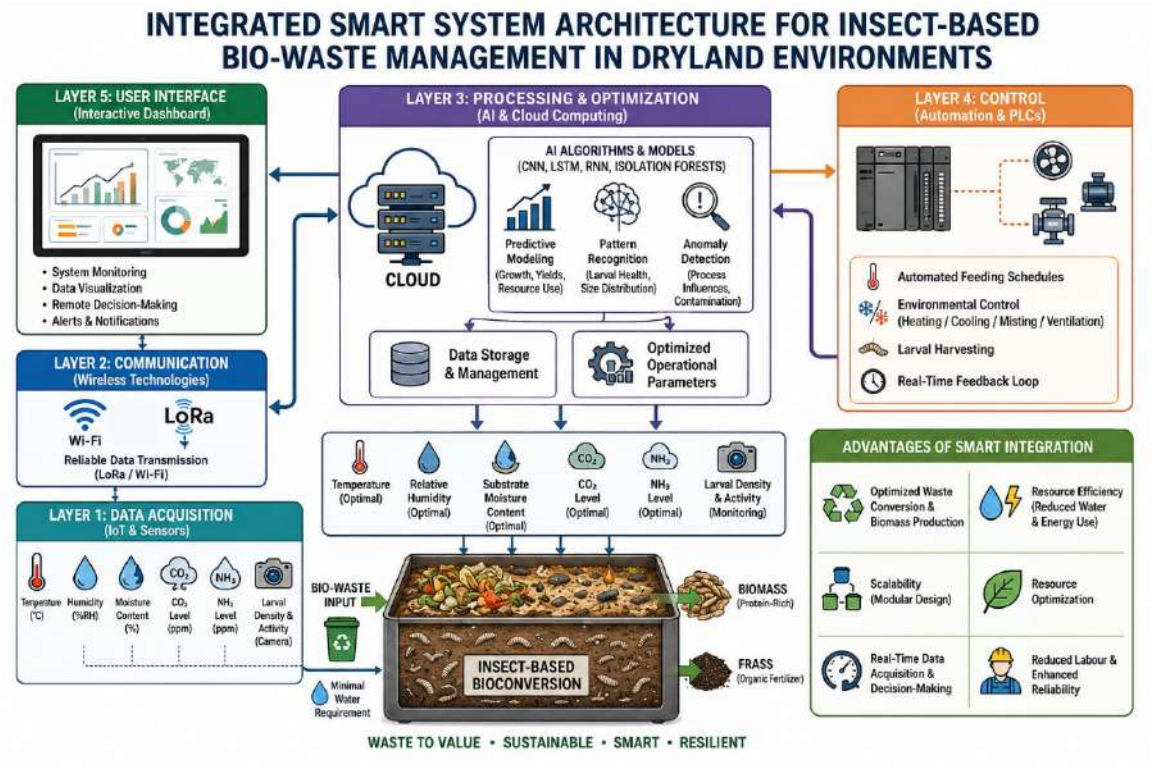


Figure 4. Integrated smart system architecture for insect-based bio-waste management in desert ecosystems.

4.6 Advantages of Smart Integration

The integration of artificial intelligence, Internet of Things, and automation technologies into insect-based bio-waste management systems provides several advantages that significantly enhance system performance and sustainability. These technologies enable optimized waste conversion and biomass production through precise monitoring and dynamic control of environmental conditions. Resource efficiency is improved through reduced water and energy consumption compared to conventional systems. Furthermore, the modular design of smart systems supports scalability, allowing flexible expansion according to operational requirements. Real-time data acquisition and analysis facilitate rapid decision-making, enabling immediate responses to environmental fluctuations and operational disturbances. Automation also reduces labour requirements, improves consistency, and enhances overall system reliability, making these systems highly suitable for large-scale deployment [13], [16].

4.7 Challenges and Limitations

Despite their advantages, this integrated framework faces several challenges. High initial investment costs can limit adoption, particularly in developing and resource-constrained regions. The technical

complexity associated with integrating AI, IoT, and automated control systems requires specialized expertise for operation and maintenance. In addition, issues related to data management, including storage, processing, and cybersecurity, present potential risks that must be carefully addressed. Limited infrastructure in remote desert areas, such as unreliable connectivity and energy supply, further complicates deployment and scalability. Addressing these challenges requires interdisciplinary collaboration, technological innovation, and targeted investments in infrastructure development [13]–[16].

4.8 Role in Desert Ecosystems

The proposed system is particularly well suited for dryland environments due to their ability to operate under resource constraints and environmental variability. By enabling precise control over bioconversion processes, these systems maintain optimal performance even under extreme conditions. Moreover, the integration of renewable energy sources, particularly solar power, can enhance system sustainability and reduce dependence on external energy inputs. This makes the proposed system highly suitable for off-grid applications in desert regions, contributing to sustainable waste management and resource recovery in challenging environments [16], [18].

5. Proposed Smart Insect-Based Bio-Waste Management System Framework

5.1 Conceptual Overview

This section presents a novel and integrated framework for a smart insect-based bio-waste management system specifically designed for arid regions. The proposed system combines insect bioconversion, artificial intelligence, Internet of Things technologies, and adaptive engineering design to address environmental constraints and improve waste processing efficiency. Unlike conventional systems, this framework emphasizes climate adaptability, modularity, and real-time optimization, making it suitable for arid and resource-limited environments. The system operates within a closed-loop circular bioeconomy model in which organic waste is transformed into valuable outputs such as insect biomass and organic fertilizers while minimizing environmental impacts [17], [18]. This framework is conceptual and intended as a design model to guide future experimental and industrial implementation. The overall architecture of the proposed smart insect-based bio-waste management system is illustrated in Figure 5.

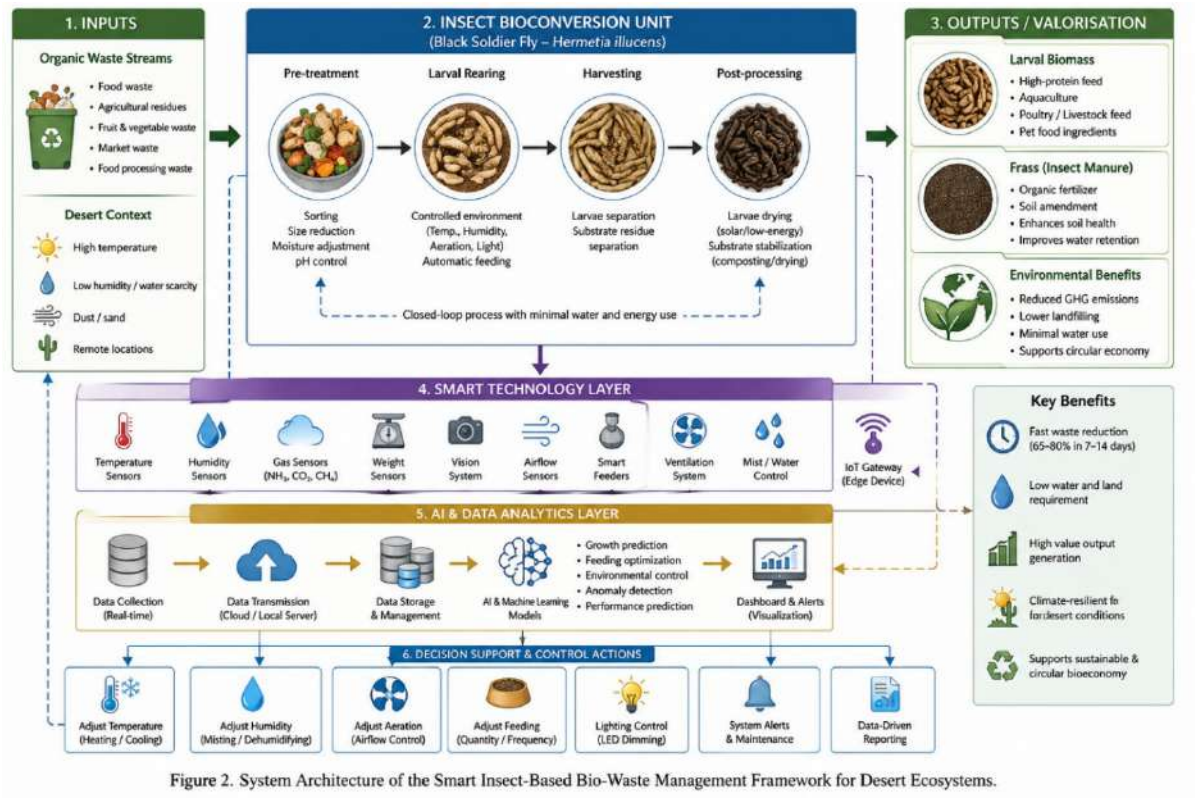


Figure 2. System Architecture of the Smart Insect-Based Bio-Waste Management Framework for Desert Ecosystems.

Figure 5. System architecture of the smart insect-based bio-waste management framework for desert ecosystems, illustrating the integration of insect bioconversion processes with IoT-based monitoring, AI-driven optimization, and automated control systems.

5.2 System Architecture and Functional Modules

The proposed framework consists of five interconnected functional modules designed to ensure efficient and adaptive system performance. These modules include waste pre-processing, insect bioconversion, smart monitoring, artificial intelligence-based optimization, and output recovery. Together, they enable continuous and efficient bio-waste conversion through an integrated smart management approach [13]–[16]. The operational workflow of the proposed system is illustrated in Figure 6, where organic waste is progressively transformed into valuable products through a controlled insect-based bioconversion process.

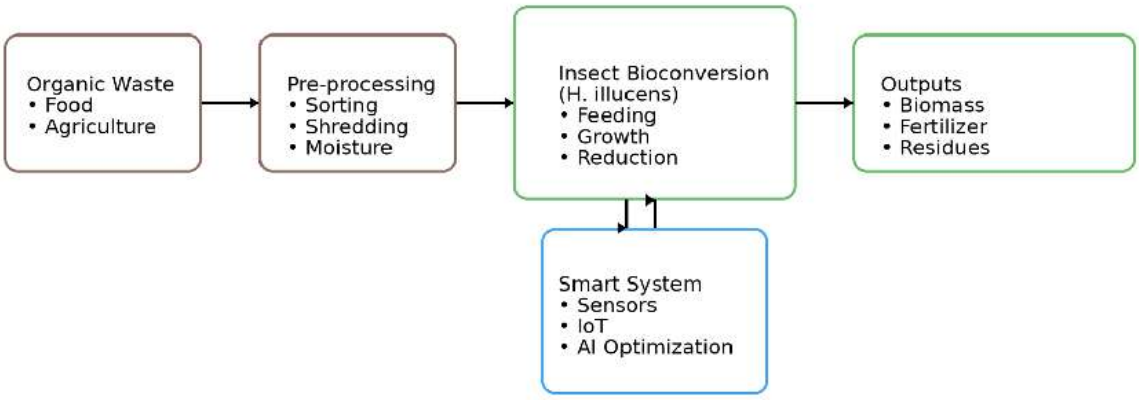


Figure 6. Workflow of the smart insect-based bio-waste management system, illustrating the transformation of organic waste into valuable outputs through insect bioconversion integrated with monitoring and optimization processes.

5.2.1 Waste Pre-processing Unit

The waste pre-processing unit is responsible for preparing incoming bio-waste for efficient bioconversion. This includes sorting to remove non-organic materials, reducing particle size through shredding, and adjusting moisture content to optimal levels for larval feeding. Proper pre-processing enhances substrate uniformity, improves feeding efficiency, and supports consistent system performance, as demonstrated in studies on insect-based waste treatment systems [3], [7].

5.2.2 Adaptive Insect Bioconversion Unit

The insect bioconversion unit represents the core of the system, where larvae, primarily of *H. illucens*, convert organic waste into biomass. This unit is designed with climate-controlled chambers and insulated modular structures to withstand extreme desert temperatures. The use of optimized feeding trays ensures maximum exposure of waste material to larvae, thereby enhancing conversion efficiency. Maintaining appropriate environmental conditions within this unit is essential for achieving high productivity and stable system operation [2], [12].

5.2.3 Smart Monitoring and Sensing Unit

The monitoring unit consists of a network of IoT-based sensors that continuously track environmental and biological parameters, including temperature, humidity, substrate moisture, gas emissions, and larval activity. The collected data are transmitted in real time to a centralized system for analysis,

enabling continuous monitoring of system performance and early detection of potential issues. Such sensor-based monitoring approaches are widely applied in smart agriculture and environmental management systems [13], [16].

5.2.4 Artificial Intelligence-Based Decision and Optimization Unit

This unit serves as the central processing component of the system, where collected data are analysed using machine learning algorithms. The system predicts optimal feeding schedules, detects anomalies such as overheating or contamination, and dynamically adjusts environmental conditions to maintain optimal performance. Through continuous learning and adaptation, the artificial intelligence component enhances system efficiency and reliability. Similar AI-driven optimization approaches have been successfully implemented in smart agricultural and bioresource management systems [13]–[15].

5.2.5 Output Processing and Resource Recovery Unit

The output processing unit manages the separation and utilization of system outputs. This includes harvesting larval biomass, which can be used as a sustainable protein source, and collecting frass, which serves as an organic fertilizer. In addition, this unit compiles operational data for performance evaluation and reporting, supporting system optimization and integration with downstream applications. The valorisation of insect-derived products has been extensively documented in recent literature [3], [12].

5.3 System Workflow and Operational Cycle

The system operates through a continuous and automated workflow that begins with bio-waste input and pre-processing, followed by feeding into the insect bioreactor. Real-time monitoring through sensor networks provides continuous data, which are analysed by the artificial intelligence system to optimize operational parameters. Automated control systems then regulate feeding, environmental conditions, and harvesting processes. This closed-loop cycle ensures efficient waste conversion, minimal resource loss, and continuous system improvement, consistent with modern smart bio-waste management systems [13]–[16].

5.4 Key Innovations of the Proposed Framework

The proposed framework introduces several innovative features that distinguish it from conventional waste management systems. These include a climate-adaptive design capable of operating under extreme environmental conditions, real-time optimization through artificial intelligence, and a modular structure that supports decentralized deployment. The system also promotes circular bioeconomy principles by converting waste into valuable resources and can be integrated with renewable energy sources to enhance sustainability [17], [18].

5.5 Scalability and Deployment Strategy

The system is designed to be scalable through modular bioreactor units that can be expanded according to operational requirements. It supports flexible deployment strategies ranging from small-scale rural applications to medium-scale municipal systems and large-scale industrial facilities. The decentralized nature of the system reduces the need for long-distance waste transportation and enables localized waste processing, aligning with sustainable waste management practices [13], [16].

5.6 Suitability for Desert Ecosystems

The proposed system is specifically adapted to address the challenges of desert environments. It minimizes water usage, operates efficiently under high temperatures, and can be integrated with renewable energy sources such as solar power. The system also supports localized waste processing, reducing logistical constraints and enhancing sustainability in remote areas. These characteristics are consistent with recent efforts to develop climate-resilient waste management solutions for arid regions [6], [18].

5.7 Expected Environmental and Economic Outcomes

The implementation of the proposed framework is expected to generate significant environmental and economic benefits. These include substantial reduction in organic waste volume, decreased greenhouse gas emissions, and improved soil fertility through the application of organic fertilizers. Additionally, the production of insect biomass creates economic opportunities and supports value-added supply chains. Overall, the system contributes to sustainable development, resource recovery, and job creation in desert regions, consistent with findings from studies on insect-based circular bioeconomy systems [3], [17].

6. Environmental and Economic Impact Assessment

6.1 Environmental Impact Assessment

The implementation of this integrated framework offers substantial environmental benefits, particularly in desert ecosystems where conventional waste treatment methods are often inefficient. One of the most significant advantages is the reduction of greenhouse gas emissions. Unlike landfilling, which promotes anaerobic decomposition and methane generation, insect-based bioconversion operates primarily under aerobic conditions, resulting in significantly lower emissions of methane and nitrous oxide [3], [10]. Studies have shown that black soldier fly-based systems can achieve considerable reductions in greenhouse gas emissions compared to traditional waste disposal methods [3].

In addition, these systems improve waste management efficiency by reducing organic waste volume by approximately 65–80% within a short processing period. This rapid reduction minimizes the need for landfill space and reduces environmental pollution. The process also produces frass, a nutrient-rich organic residue that enhances soil fertility by increasing essential nutrients such as nitrogen, phosphorus, and potassium. This is particularly beneficial in desert soils, which are typically low in organic matter and nutrients [12]. Water conservation represents another key environmental advantage. Insect-based systems require significantly less water than conventional composting processes, making them highly suitable for arid regions facing severe water scarcity. Furthermore, the controlled processing environment minimizes the risk of leachate formation and groundwater contamination, thereby protecting fragile dryland environments [6].

6.2 Economic Impact and Quantitative Comparison

From an economic perspective, the proposed system provides multiple benefits through cost reduction and resource valorisation. A primary advantage lies in the production of high-value outputs, particularly insect biomass rich in protein and lipids, which can be utilized in animal feed, aquaculture, and other industrial sectors. Additionally, the by-product frass serves as an effective organic fertilizer, contributing to soil improvement and reducing dependence on chemical inputs. These outputs create diversified revenue streams that enhance the overall economic viability of the system [3], [12].

Operational costs are generally lower than those associated with conventional waste management approaches due to reduced energy consumption, minimal water requirements, and decreased labour dependency resulting from automation and smart system integration. Although the initial capital investment for infrastructure and digital technologies may be relatively high, previous studies indicate

that the return on investment can be achieved within a reasonable timeframe depending on system scale and market conditions [19]. To further evaluate system performance, a comparative analysis with conventional waste management methods is presented in Table 3, highlighting key differences in efficiency, environmental impact, and economic value.

Table 3. Environmental and Economic Comparison of Waste Management Systems. Values are synthesized from multiple literature sources [3], [12], [19].

Parameter	Landfilling	Composting	Insect-Based System
Waste reduction (%)	20–30	40–60	65–80
Processing time	Months–years	Weeks–months	Days (7–14)
Greenhouse gas emissions	High	Moderate	Low
Water requirement	Low	High	Very low
Energy consumption	Moderate	Moderate	Low
Value-added products	None	Compost	Protein + Fertilizer
Operational cost	High (long-term)	Moderate	Moderate–Low
Environmental impact	High	Moderate	Low

The comparative analysis demonstrates the superior performance of insect-based bioconversion systems relative to traditional waste management approaches. In particular, the significantly higher waste reduction rates and shorter processing times highlight their efficiency in handling organic waste streams. Moreover, the lower greenhouse gas emissions and minimal water requirements make these systems especially suitable for arid and resource-constrained environments. Unlike landfilling, which does not generate value-added products, insect-based systems produce both protein-rich biomass and organic fertilizers, thereby supporting circular economy principles and improving overall sustainability [3], [17].

6.3 Policy and Sustainability Implications

The adoption of this integrated framework aligns strongly with global sustainability goals and environmental policies. These systems support circular bioeconomy principles by converting organic waste into valuable resources, reducing environmental burdens, and enhancing resource efficiency [17], [18]. They also contribute to climate change mitigation efforts by lowering greenhouse gas emissions and promoting sustainable waste management practices [3].

In desert regions, such systems can play a strategic role in national waste management policies by providing decentralized and scalable solutions that reduce dependence on landfills. Policymakers can facilitate implementation through financial incentives, supportive regulatory frameworks, and

investments in infrastructure and research. Furthermore, integration with renewable energy sources such as solar power enhances system sustainability and aligns with global clean energy initiatives.

6.4 Overall Impact and Feasibility

Overall, the proposed system demonstrates strong environmental and economic feasibility. Its capacity to significantly reduce waste volume, recover valuable resources, and operate efficiently under harsh environmental conditions makes it a viable and sustainable solution for dryland environments. The integration of biological efficiency with advanced smart technologies ensures long-term scalability, operational resilience, and alignment with sustainable development goals [17], [18].

7. Case Studies and Global Applications

7.1 Overview of Global Adoption Trends

In recent years, this integrated framework has gained increasing global attention as sustainable alternatives to conventional waste treatment methods. The use of black soldier fly larvae for organic waste conversion has expanded across multiple regions, including Europe, Asia, Africa, and the Middle East. This growing adoption is driven by the need to reduce environmental impacts, recover valuable resources, and implement circular bioeconomy strategies. Advances in automation and digital technologies have further accelerated the scalability and efficiency of these systems, enabling their application in diverse environmental conditions, including arid and semi-arid regions [3], [13]–[16].

7.2 Case Study: Middle East (Saudi Arabia and Gulf Region)

In the Middle East, particularly in Saudi Arabia, pilot-scale insect-based waste management systems have been implemented to address the challenges of organic waste accumulation under extreme climatic conditions, as reported in recent studies. These systems have demonstrated the ability to reduce food waste by approximately 70–75% while maintaining stable performance under high temperatures. The integration of climate-controlled bioreactors and solar-powered infrastructure has proven effective in overcoming environmental constraints. Furthermore, the adoption of such systems aligns with national sustainability initiatives and waste reduction strategies in the region, highlighting their relevance for arid environments [6], [18].

7.3 Case Study: Asia (China and Southeast Asia)

China has emerged as a global leader in large-scale insect bioconversion systems, particularly through the industrial application of black soldier fly larvae for processing municipal and agricultural waste.

Several facilities have incorporated artificial intelligence and automated control systems to optimize feeding rates, environmental conditions, and harvesting processes. These smart systems have achieved significant improvements in efficiency and reductions in operational costs. In Southeast Asia, insect-based waste management is widely applied at both community and industrial levels, contributing to sustainable agriculture, waste valorisation, and circular resource utilization [3], [13]–[15].

7.4 Case Study: Africa (Kenya and Sub-Saharan Regions)

In African countries such as Kenya, insect-based systems have been successfully implemented at the community level, to manage organic waste and generate economic value. Small-scale black soldier fly production units convert household and agricultural waste into protein-rich feed for livestock and aquaculture. These systems have improved local waste management practices, reduced environmental pollution, and created income-generating opportunities for rural communities. The adaptability of this integrated framework to low-resource settings makes them particularly suitable for developing regions with limited infrastructure [3], [21].

7.5 Case Study: Europe and Comparative Analysis of Global Applications

European countries have made substantial progress in the industrialization and regulation of insect-based bioconversion systems, particularly in countries such as the Netherlands and Switzerland. In these regions, advanced facilities operate under highly controlled and automated conditions, enabling large-scale processing of organic waste with high efficiency and consistency. These systems are integrated within circular bioeconomy frameworks, where insect-derived biomass is utilized as a sustainable protein source in animal feed, and frass is applied as an organic fertilizer to enhance soil fertility. Strong regulatory frameworks, combined with continuous technological innovation, have supported the commercialization and standardization of insect-based waste management systems, positioning Europe as a global leader in this field [3], [20].

To better understand the global landscape of insect-based bio-waste management, a comparative analysis across different regions is presented in Table 4, highlighting variations in system scale, technological integration, and key outcomes.

Table 4. Comparison of Insect-Based Bio-Waste Systems in Different Regions

Region	Scale of Application	Key Features	Main Outcomes
Middle East	Pilot–Medium	Climate-adaptive systems, solar integration	High waste reduction under extreme heat
Asia	Large-scale industrial	AI integration, automation	High efficiency, cost reduction
Africa	Small–community scale	Low-cost systems, decentralized	Income generation, local sustainability
Europe	Industrial	Advanced technology, strong regulations	Commercialization and circular economy

The comparison presented in Table 4 demonstrates that insect-based bio-waste management systems can be successfully adapted to diverse environmental and socio-economic contexts. Industrial-scale applications in Europe and Asia emphasize technological advancement and efficiency, while systems in Africa highlight affordability and decentralized deployment. In the Middle East, the focus on climate adaptation and renewable energy integration reflects the need to address extreme environmental conditions. These variations underscore the flexibility and scalability of insect-based systems, reinforcing their potential as a globally applicable solution for sustainable waste management [3], [18].

7.6 Lessons Learned from Global Applications

Analysis of global case studies reveals several key insights. First, the adaptability of insect-based systems to different environmental and economic conditions makes them a versatile solution for waste management. Second, the integration of smart technologies significantly enhances system efficiency and scalability. Third, decentralized and modular approaches are particularly effective in regions with limited infrastructure. Finally, policy support and regulatory frameworks play a critical role in facilitating large-scale adoption and commercialization of these systems [13], [16].

7.7 Relevance to Desert Ecosystems

The lessons derived from global applications are highly relevant to arid regions. The success of systems in the Middle East demonstrates their ability to operate under extreme climatic conditions, while experiences from Africa highlight their suitability for low-resource and decentralized settings. The integration of advanced technologies, as observed in Asia and Europe, provides a pathway for

optimizing system performance and scalability. These insights support the feasibility of implementing this integrated framework in desert regions as a sustainable and economically viable solution [6], [18].

8. Challenges and Limitations

8.1 Technical Constraints

Despite the promising potential of the proposed system, several technical challenges remain that may affect system performance and scalability. One of the primary limitations is maintaining optimal environmental conditions for insect growth, particularly temperature and moisture, which are critical for efficient bioconversion. In dryland environments, extreme heat exceeding optimal thresholds can negatively impact larval survival and metabolic activity. Additionally, fluctuations in substrate quality and composition may lead to inconsistent system performance. The integration of advanced technologies such as artificial intelligence and IoT further introduces complexity in system design, requiring reliable hardware, continuous calibration, and technical expertise for operation and maintenance [13]–[16].

8.2 Biological and Operational Limitations

From a biological perspective, insect species used for bioconversion, particularly *H. illucens*, have specific physiological requirements that must be maintained to ensure optimal performance. Factors such as overcrowding, inadequate nutrition, and improper environmental conditions can result in reduced growth rates, increased mortality, and lower waste conversion efficiency. Moreover, large-scale operations may face challenges related to colony management, disease control, and genetic variability. Operationally, maintaining consistent feeding regimes and synchronizing larval development cycles with waste supply can be difficult, particularly in decentralized systems or where waste streams fluctuate [3], [7].

8.3 Economic and Financial Barriers

Although insect-based systems offer long-term economic advantages, the initial capital investment required for infrastructure, automation, and smart technologies can be substantial. These costs include sensor networks, data processing platforms, climate-controlled units, and automated equipment. For developing regions and small-scale operators, such financial requirements may limit adoption. In addition, market development for insect-derived products such as protein and organic fertilizers is still evolving, and uncertainties related to pricing, demand, and supply chains may affect economic feasibility [3], [19].

8.4 Environmental and Resource Constraints

While insect-based systems are generally more sustainable than conventional waste management methods, they still require careful resource management. Water availability, although significantly lower than in composting systems, remains important for maintaining optimal substrate conditions. Energy requirements for climate control and system operation may also present challenges, particularly in off-grid desert areas. Furthermore, inadequate system management may result in localized environmental issues such as odour generation or attraction of unwanted pests, which could affect community acceptance and environmental quality [6].

8.5 Safety and Contaminant Considerations

The presence of pesticide residues and heavy metals in bio-waste streams represents an important safety consideration in insect-based bioconversion systems. Previous studies have shown that certain contaminants may accumulate in larval biomass or frass depending on insect species and substrate characteristics. While some organic contaminants may undergo partial degradation during the bioconversion process, heavy metals are generally not degraded and may become concentrated within the system. For example, studies on *H. illucens* have demonstrated the potential accumulation of cadmium, lead, and arsenic in larval tissues when contaminated substrates are used [23], [24]. Therefore, careful substrate selection, contaminant monitoring, and regulatory compliance are essential to ensure the safe utilization of insect-derived biomass and frass in agricultural and animal feed applications.

8.6 Data Management and Cybersecurity Issues

The integration of smart technologies introduces challenges related to data handling, storage, and security. Large volumes of data generated by IoT sensors require efficient processing, reliable storage infrastructure, and effective data management strategies. In addition, cybersecurity risks, including unauthorized access and data breaches, may compromise system integrity and operational reliability. Ensuring secure communication protocols and robust data protection measures is therefore essential for the safe and effective operation of smart bio-waste management systems [13]–[16].

8.7 Regulatory and Social Acceptance Challenges

Regulatory frameworks governing the use of insects in waste management and feed production vary widely across regions. In some cases, unclear or restrictive regulations may hinder the adoption and commercialization of insect-based systems. Social acceptance also remains a challenge, particularly in communities where the use of insects for waste processing or protein production is unfamiliar. Increasing public awareness, promoting education, and developing supportive policy frameworks are essential for facilitating widespread adoption and long-term success [17], [18].

8.8 Scalability and Infrastructure Limitations

Scaling the proposed system from pilot-scale to industrial-scale applications presents several challenges. Infrastructure limitations, including transportation, energy supply, and communication networks, can affect deployment in remote desert regions. Maintaining consistent performance at larger scales requires careful system design, continuous monitoring, and effective management strategies. Although modular system designs offer a potential solution to scalability challenges, further research and technological development are needed to optimize large-scale implementation [13], [16].

8.9 Critical Evaluation and Research Gaps

Despite significant progress, several research gaps remain in the field of insect-based bio-waste management. These include limited long-term performance data under extreme environmental conditions, insufficient understanding of insect–microbial interactions, and the lack of standardized performance evaluation metrics. In addition, the integration of artificial intelligence and automation in these systems is still evolving and requires further validation under real-world conditions. Addressing these gaps is essential to improve system reliability, scalability, and global applicability, particularly in challenging environments such as desert ecosystems [13]–[16].

9. Future Perspectives

9.1 Advancing Smart and Autonomous Bio-Waste Systems

Future developments in insect-based bio-waste management are expected to move toward fully autonomous and self-optimizing systems. The integration of advanced artificial intelligence, edge computing, and real-time analytics will enable these systems to operate with minimal human intervention while continuously adapting to environmental and operational changes. The use of digital twins (virtual replicas of physical systems), which are virtual representations of physical systems, can further enhance predictive maintenance, process simulation, and performance optimization. Such

advancements are particularly valuable in arid regions, where environmental variability is high and operational stability is essential [13]–[16].

9.2 Integration with Renewable Energy and Off-Grid Solutions

The sustainability of the proposed system can be significantly enhanced through integration with renewable energy sources, particularly solar power. Desert regions are characterized by high solar irradiance, making them ideal for deploying solar-powered bio-waste processing units. Future systems are expected to incorporate energy-efficient designs, hybrid power solutions, and energy storage technologies to ensure continuous operation in off-grid environments. This integration will reduce operational costs, minimize carbon emissions, and support the development of self-sufficient waste management infrastructures [6], [18].

9.3 Genetic Improvement and Insect Biotechnology

Advances in insect biotechnology and genetics present promising opportunities to enhance the efficiency of bioconversion systems. Selective breeding and genomic approaches can be employed to develop insect strains with improved growth rates, higher waste conversion efficiency, and increased tolerance to extreme environmental conditions. In addition, research on gut microbiota engineering may further improve digestive efficiency and nutrient recovery. These innovations have the potential to significantly enhance system productivity and adaptability, particularly in harsh desert environments [3], [7].

9.4 Expansion of Circular Bioeconomy Applications

Future insect-based systems are expected to play a central role in expanding circular bioeconomy models. Beyond waste reduction, these systems can contribute to the production of high-value products such as biofuels, bioplastics, and functional biomaterials derived from insect biomass. The integration of insect farming with agriculture, aquaculture, and food systems will create interconnected value chains that maximize resource utilization and minimize waste. This approach aligns with global sustainability goals and supports the transition toward resource-efficient economies [17], [18].

9.5 Policy Development and Global Standardization

The future growth of insect-based bio-waste management systems will depend significantly on the development of supportive policies and regulatory frameworks. Harmonization of international standards related to insect farming, waste processing, and product safety is essential to facilitate global

adoption and trade. Governments and international organizations are expected to play a key role in promoting research, providing incentives, and establishing guidelines that support innovation while ensuring environmental and public health safety [17], [18].

9.6 Digitalization and Big Data Integration

The increasing availability of big data and advancements in digital technologies will further enhance the capabilities of smart bio-waste systems. Data-driven approaches will enable more accurate modelling of system performance, optimization of operational parameters, and identification of inefficiencies. The application of advanced data analytics and intelligent systems will contribute to improved decision-making and operational efficiency. These developments will support the creation of more reliable, scalable, and efficient bio-waste management systems [13]–[16].

9.7 Scaling for Global Impact and Desert Applications

Looking ahead, the large-scale deployment of the proposed system has the potential to transform waste management practices globally, particularly in arid and semi-arid regions. Future efforts should focus on developing cost-effective, modular, and scalable solutions that can be adapted to diverse socio-economic and environmental contexts. Pilot projects and demonstration sites in desert regions will be essential for validating system performance and building stakeholder confidence. Collaboration among academia, industry, and policymakers will play a critical role in accelerating the transition from experimental systems to practical large-scale applications [13], [18].

9.8 Concluding Outlook

Overall, the future of this integrated framework is highly promising, with significant potential to address global challenges related to waste management, resource scarcity, and environmental sustainability. Continued innovation in technology, biology, and system design will be essential to fully realize this potential. As research advances and practical implementations expand, these systems are expected to become integral components of sustainable development strategies, particularly in environmentally challenging regions such as dryland environments [17], [18].

10. Conclusion

Bio-waste accumulation and resource scarcity remain critical challenges in desert ecosystems, where conventional waste management approaches are often ineffective. This review highlights that insect-based bioconversion, particularly using *H. illucens*, provides an efficient and sustainable alternative

by rapidly converting organic waste into valuable products while reducing greenhouse gas emissions and water demand.

The proposed smart insect-based bio-waste management framework represents a promising conceptual advancement by integrating biological processes with artificial intelligence, Internet of Things, and automation. This integration enables real-time monitoring, adaptive optimization, and scalable operation, making the system particularly suitable for harsh and resource-limited desert environments. Environmental and economic assessments confirm its potential for waste reduction, soil improvement, and value generation, while global case studies demonstrate its adaptability across diverse contexts.

Despite its advantages, challenges related to technical complexity, investment costs, and regulatory frameworks remain. Addressing these issues through continued research, technological innovation, and policy support will be essential. Overall, smart insect-based systems offer a promising, scalable pathway toward sustainable waste management and circular bioeconomy development in desert regions.

Acknowledgment

The authors acknowledge the support of their affiliated institutions and research facilities. Appreciation is extended to colleagues and technical staff who contributed to the development of this work, as well as to the broader research community whose studies in insect bioconversion and sustainable waste management informed this review.

Conflict of Interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

References

[1] A. van Huis et al., “Edible insects: future prospects for food and feed security,” FAO Forestry Paper 171, Food and Agriculture Organization, Rome, 2013.

[2] S. Diener, C. Zurbrügg, and K. Tockner, “Conversion of organic material by black soldier fly larvae: establishing optimal feeding rates,” Waste Management & Research, vol. 27, no. 6, pp. 603–610, 2009.

[3] M. Gold et al., “Decomposition of biowaste macronutrients, microbes, and chemicals in black soldier fly larval treatment: A review,” Waste Management, vol. 82, pp. 302–318, 2018.

[4] E. M. Makkar et al., “State-of-the-art on use of insects as animal feed,” Animal Feed Science and Technology, vol. 197, pp. 1–33, 2014.

[5] A. van Huis, “Insects as food and feed, a new emerging agricultural sector,” Journal of Insects as Food and Feed, vol. 6, no. 1, pp. 27–44, 2020.

[6] M. A. Salomone et al., “Environmental impact of food waste management systems: A review,” Journal of Cleaner Production, vol. 289, 125–142, 2021.

[7] J. K. Tomberlin et al., “Black soldier fly bioconversion of organic waste,” Annual Review of Entomology, vol. 65, pp. 141–159, 2020.

[8] C. Zurbrügg et al., “Challenges and opportunities of insect-based organic waste management,” Waste Management, vol. 102, pp. 201–210, 2020.

[9] S. Barragán-Fonseca et al., “Insects for organic waste bioconversion: global perspectives,” Waste Management, vol. 102, pp. 1–12, 2020.

[10] D. Oonincx et al., “An exploration on greenhouse gas and ammonia production by insect species suitable for animal or human consumption,” PLoS ONE, vol. 5, no. 12, e14445, 2010.

[11] M. Banks et al., “Operational performance of insect-based waste systems,” Journal of Cleaner Production, vol. 330, 129–140, 2022.

[12] S. Lalander et al., “High waste-to-biomass conversion using black soldier fly larvae,” Journal of Cleaner Production, vol. 137, pp. 117–125, 2016.

[13] J. Wolfert et al., “Big data in smart farming: a review,” Agricultural Systems, vol. 153, pp. 69–80, 2017.

[14] S. Kamilaris and F. X. Prenafeta-Boldú, “Deep learning in agriculture: A survey,” Computers and Electronics in Agriculture, vol. 147, pp. 70–90, 2018.

[15] L. Li et al., “Applications of artificial intelligence in agriculture,” Sensors, vol. 20, no. 10, 2020.

[16] M. Zhang et al., “IoT-based smart agriculture systems,” IEEE Access, vol. 8, pp. 110–123, 2020.

[17] Ellen MacArthur Foundation, “Towards the circular economy,” 2019.

[18] FAO, “The future of food and agriculture: Trends and challenges,” 2022.

[19] A. Sheppard et al., “Sustainable waste-to-protein systems,” Waste Management, vol. 120, pp. 10–20, 2021.

[20] M. Gold et al., “Scaling insect-based biowaste treatment,” Waste Management, vol. 118, pp. 377–388, 2020.

[21] C. Zurbrügg et al., “Black soldier fly applications in developing countries,” Sustainability, vol. 14, 2022.

[22] J. van Huis, “Insects as sustainable feed systems,” Annual Review of Entomology, vol. 66, pp. 1–18, 2021.

[23] Diener, S., Zurbrügg, C., and Tockner, K. (2015). *Bioaccumulation of heavy metals in the black soldier fly, Hermetia illucens, and effects on its life cycle.* Journal of Insects as Food and Feed, 1(4), 261–270.

[24] Van der Fels-Klerx, H. J., Camenzuli, L., van der Lee, M. K., and Oonincx, D. G. A. B. (2016). Uptake of cadmium, lead and arsenic by *Tenebrio molitor and Hermetia illucens* from contaminated substrates. PLoS ONE, 11(11), e0166186.

Islamic University Journal of Applied Sciences (IUJAS)
https://journals.iu.edu.sa/jesc

Volume VIII, Issue II, December 2026, Pages 127-148

Explainable Machine Learning Framework for Predicting Bond Strength of FRP Bars Embedded in UHPC

Abdulaziz Alqurashi [1] *

[1.] Department of Civil Engineering, Faculty of Engineering, Islamic University of Madinah, 42351, Madinah, Saudi Arabia, amalqurashi@iu.edu.sa

*Corresponding author: (Abdulaziz Alqurashi), *Email Address*: amalqurashi@iu.edu.sa

Abstract

The interfacial performance of advanced composites bars embedded in Ultra-High Performance Concrete (UHPC) is an important factor that controls load transfer and the performance of structural elements. Predicting bond strength is still difficult because it is affected by several factors, such as rebar type, bar profile, bar diameter, bonded length, cover depth, fiber content, UHPC compressive strength, and FRP tensile strength. Therefore, this study uses machine-learning models to estimate the the bonding capacity of FRP bars placed in UHPC Using a collected experimental database of 183 specimens from previous studies. Four machine-learning models were developed and compared, including Linear Regression, Random Trees, Multi-Layer Perceptron, and Locally Weighted Learning. The MLP model gave the best prediction performance, with a correlation coefficient of 0.9466, MAE of 2.3083 MPa, and RMSE of 3.0631 MPa. SHAP analysis showed that embedment length was the most influential variable, followed by bar surface condition, FRP tensile strength, and concrete cover. This confirms that FRP–UHPC bond behavior is controlled by the interaction between bonded length, surface condition, mechanical interlock, and confinement provided by UHPC. Overall, the developed explainable ML framework provides a useful tool for predicting FRP–UHPC bond strength and supporting future UHPC-specific bond models.

Keywords: Ultra High-Performance Concrete; FRP bars; Pull-out behaviour; Machine Learning; Interfacial Mechanics.

https://doi.org/10.63070/jesc.2026.037
Received 02 May 2026; Revised 21 June 2026; Accepted 24 June 2026; Available online 15 July 2026.
Published by Islamic University of Madinah on behalf of ***Islamic University Journal of Applied Sciences.***
This is an open access article published under the Creative Commons Attribution 4.0 International license (CC BY 4.0):https://creativecommons.org/licenses/by/4.0/

1. Introduction

Ultra-high performance concrete (UHPC) is a high Perfomance building material, developed from the reactive powder concrete concept, characterized by very high strength in compression, dense microstructure, and using the fine constituents and admixtures to enhance durability and mechanical performance [1]. Because UHPC can substantially improve structural capacity and longevity, it is being adopted in bridge decks, prestressed girders, segmental joints and other applications worldwide. Concurrently, nonmetallic reinforcement, such as carbon-, glass-, and basalt-fiber-reinforced polymer bars, referred to as CFRP, GFRP, and BFRP bars, respectively, has attracted attention because of its high tensile capacity and reduced susceptibility to corrosion compared with conventional steel reinforcement. [2-3].

Designing durable and efficient structural elements that combine UHPC with FRP bars requires a reliable understanding of their bonding behaviour within the surrounding UHPC matrix. Bond transfer arises from a combination of mechanical interlock, adhesive interaction, and surface friction. Relative contribution of mechanisms depends on bar material and surface topography, matrix composition (including silica fume and discrete fibers), and geometry (bar diameter, rib geometry, embedment length and concrete cover) [4]. Experimental programs from recent literature (including pushout and pull-out investigations) have documented that steel, GFRP and BFRP bars interact with UHPC through distinct mechanisms: steel typically engages UHPC by rib interlock and may induce local UHPC crushing between ribs. On the other hand, FRP bars often exhibit resin-fiber debonding at the bar interface and direction-dependent (orthotropic) surface responses [5]. Microscopy and surface-texture analyses (e.g., gray-level co-occurrence matrix metrics) further show that interfacial damage and energy dissipation are spatially anisotropic for FRP bars, complicating simple, one-dimensional descriptions of bond behavior [6].

Despite significant experimental efforts, key uncertainties persist. Conflicting trends have been reported for the influence of GFRP bar diameter on bonding behaviour some studies report decreasing bond with increasing diameter, while others observe the opposite and relatively few systematic investigations have mapped how UHPC cover thickness interacts with bar size and embedment depth to control the failure mechanism and the bond stress–slip behaviour [7]. Moreover, predictive formulae derived for steel or FRP in normal or high-strength concrete often fail to reproduce measured bond performance for FRP–UHPC combinations, motivating the development of UHPC-specific empirical relations based on targeted experiments [8]. Because the bonding performance of FRP bars embedded in UHPC is influenced by several variables, the available empirical equations may not always give accurate predictions. These parameters include the matrix composition, embedment length, fiber

content, silica fume content, bar diameter, surface type, and UHPC cover. Therefore, machine learning can be used as a useful tool beside the experimental and analytical work. ML can learn the nonlinear relationship between these parameters and the main outputs, such as interfacial resistance, bond stress–slip behaviour, and the observed failure mechanism, without assuming a fixed equation form from the beginning [9]. Also, explainable ML methods can help show which parameters have the most effect on the bond behaviour and how each parameter increases or decreases the predicted bond capacity. This is important because some existing equations may overestimate or underestimate the bonding capacity for FRP bars embedded in UHPC. Therefore, it can improve the prediction accuracy, support the sensitivity analysis, and help guide future experimental studies, especially in the ranges where the available data are limited [10, 11].

Importance of reinforcement–UHPC interface and motivation for FRP The interfacial behaviour of reinforcement embedded in the UHPC controls the capacity and serviceability of reinforced concrete members because displacement compatibility between the reinforcement and matrix must be achieved through a combination of mechanical interlock, adhesive interaction, and surface friction. Durability concerns (notably corrosion of steel) and service demands have therefore motivated the increased use of non-metallic reinforcement such as GFRP, BFRP, and CFRP bars in aggressive environments and for long-lived infrastructure. Comparative studies of steel, GFRP, BFRP, and CFRP bars combined with UHPC reflect this practical interest.

Distinct bond mechanisms for steel versus FRP in UHPC steel deformed bars and FRP bars develop bond with concrete by different dominant mechanisms. Deformed steel bars commonly rely heavily on the mechanical engagement of the ribs with the UHPC; bond failures with steel in UHPC are frequently associated with shear and compressive crushing of UHPC between steel ribs [12]. In contrast, ribbed FRP bars (including helically ribbed GFRP) tend to fail by debonding at the resin–fiber interface or peeling of resin and fibers from the ribs; microscopy of post-pullout surfaces shows resin/fiber peeling for GFRP, indicating an interfacial failure mode distinct from the UHPC crushing typical of steel bars [13]. Furthermore, FRP bars are orthotropic and more sensitive to lateral compressive stresses, which has implications for specimen design (e.g., use of epoxy-grouted steel tubes to prevent local crushing in tests) and for how cover and confinement affect bond in practice.

Effects of UHPC matrix constituents (fibers, silica fume) and microstructure UHPC differs from normal-strength and ordinary high-strength concretes in compressive strength, microstructure (high fines, silica fume content), and use of discrete steel fibers; these matrix attributes alter the contact conditions at the reinforcement interface [14]. Multiple studies show that adding fibers to the matrix generally improves bond strength and post-peak ductility in concrete systems, and UHPC with steel fibers typically yields greater bond strength than comparable matrices without fibers. Silica-fume

content and the UHPC's high compressive strength influence micro-cracking and local crushing behaviour at ribs, and thus modify whether bond is governed by interlock, adhesion, or crushing phenomena [15, 16]. Because compressive strength of UHPC is much higher than ordinary concrete (OC) and because fibers change tensile and fracture processes, the mechanisms controlling FRP bond in UHPC can differ substantially from those in OC [12, 17].

In addition, a few studies have shown the effects of bar size, embedment length, and cover depth on FRP–UHPC bond behavior are not well documented. Experimentally, several samples have been tested helically ribbed GFRP bars in UHPC and reported two failure patterns, which are direct pull-out and splitting pull-out [18-20]; specimens with larger bar diameters and smaller concrete cover thickness showed a tendency to fail through splitting pull-out. They also reported that increasing cover thickness improved bond strength up to a critical cover of about 3–4 d_b, and that the bond capacity increased as the GFRP bar size increased, owing to the greater rib height [21]. A later study confirmed that no clear agreement had been reached regarding the influence of bar size; their tests with GFRP bar diameters from 8 to 25 mm showed that bond strength increased from 8 to 18 mm but decreased from 18 to 25 mm [7]. In addition, not only increasing UHPC cover depth, and fiber content enhanced the bond capacity, but also, with fiber content increasing from 0–2 % improving the capacity by about 74.81%–92.22%**,** while further increase to 3% had limited additional effect [22]. Comparison UHPC with normal concrete for FRP bond empirical studies generally find that embedding FRP rebars in UHPC improves bond strength compared with normal concrete, but the dominant failure mechanism (resin–fiber debonding for FRP) can persist across matrix types [23]. The markedly higher UHPC strength and presence of fibers influence magnitudes of bond stress and the relative contributions of adhesion, interlock and friction; because of these differences, predictive formulas derived for steel or FRP in conventional concrete often fail to capture GFRP–UHPC behaviour and may require UHPC-specific empirical fitting.

Overall, the literature shows that FRP–UHPC bond behavior is controlled by a nonlinear interaction between the bar surface, rib geometry, embedment length, bar diameter, concrete cover, UHPC fiber content, and the mechanical characteristics of the FRP bar. One of strongest finding across the literature is that surface condition and mechanical interlock are usually more important for FRP bars than concrete compressive strength alone [19]. Smooth FRP bars tend to show weaker and more brittle bond behavior, while helically wound, ribbed, or sand-coated bars can develop higher bond strength through friction and mechanical interlock [24]. However, the influence of embedment length, cover thickness, fiber content, and bar diameter is not always consistent because these parameters interact with each other and with the failure mode [20]. Therefore, conventional empirical equations developed for steel reinforcement or FRP bars in (OC) may not be reliable for FRP bars with UHPC. This gap

supports the use of machine-learning models because ML can identify complex nonlinear interactions among several input variables without assuming a fixed equation form. In this context, explainable ML methods such as SHAP are useful because they do not only improve prediction accuracy, but also help identify which parameters control the predicted bond strength and how these parameters are physically related to adhesion, friction, mechanical interlock, confinement, and stress redistribution.

To address this need, this study uses a collected experimental database from previous studies to develop ML models for estimating the bond capacity of FRP bars in UHPC. Multiple machine-learning algorithms are applied and evaluated according to their prediction accuracy. Explainable ML methods are then used to identify the most important variables and to interpret the contribution of each variable on bond capacity. This approach helps clarify the complex and nonlinear relationships controlling FRP–UHPC bond behavior and provides a useful basis for future design-oriented bond models.

2. Methodology

2.1 Structural method

This study adopts a data-driven methodology to estimate the bond capacity of FRP reinforcing bars embedded in (UHPC). The methodology consists of four main stages: database development, data preprocessing, machine-learning model development, and explain the physical behavior of the FRP–UHPC bond system. First, experimental data were collected from earlier published studies that investigated the bonding performance of FRP bars with UHPC. Then, the collected data were cleaned, standardized, and organized into a structured database. Several ML models were trained and evaluated to estimate the ultimate bond strength (τ_u) in Figure 1.

2.2 Data collection and analysis

A database consisting of 183 experimental specimens was constructed from previous studies that studied the bonding performance of UHPC with FRP reinforcing bars. Dataset are collected from [21, 25-31] are summarized in Table-1. These studies were selected because they reported experimental bond data for FRP bars bonded with UHPC or UHPFRC; also provided the main material and geometric parameters required for model development. The database covered different types of FRP reinforcement, including GFRP, CFRP, and BFRP bars. Most of the collected specimens were related to GFRP bars, while the remaining data included CFRP and BFRP bars. The database also included different experimental methods used to evaluate the bond behavior, such as pull-out, push-in, and beam tests. This variation in bar type and test method helps improve the generality of the ML models.

Because the data were collected from different studies, careful preprocessing was required to reduce inconsistencies between the experimental sources. The collected data were checked for missing values, inconsistent units, duplicated entries, and unclear reported results. The units were standardized, where concrete cover, bar diameter, and bonded length were expressed in (mm), while UHPC compressive strength, FRP tensile strength, and ultimate bond strength were expressed in MPa. Categorical variables, such as bar surface, FRP type, and test procedure, were encoded into numerical values before model development.

The main purpose of the data collection stage was to develop a reliable database that includes the principal variables influencing the bonding performance of FRP bars in UHPC. The data were collected from tables, figures, and reported experimental results in the selected studies. When the ultimate bond strength was reported directly, it was used as provided. However, when only the maximum applied load was available, (τ_u) was calculated using the standard average bond stress equation.

Table-1: summarize the distribution of the collected data and the statistical range of the selected variables.

Inputs	Symbol	Unit	Minimum	Maximum	Average	SD
Bar diameter	d_b	mm	7.5	19.10	12.62	3.81
Concrete cover	c	mm	10	150	73.10	34.17
Embedment length	l_d	mm	10	160	62.81	50.88
Fiber content	ρ_f	%	0	2.5	1.44	1.00
UHPC compressive strength	f'_c	MPa	86	202	155.50	27.89
FRP tensile strength	f_r	MPa	728	2480	1379.67	677.80
Ultimate bond strength	τ_u	MPa	10.20	56.24	25.09	9.52

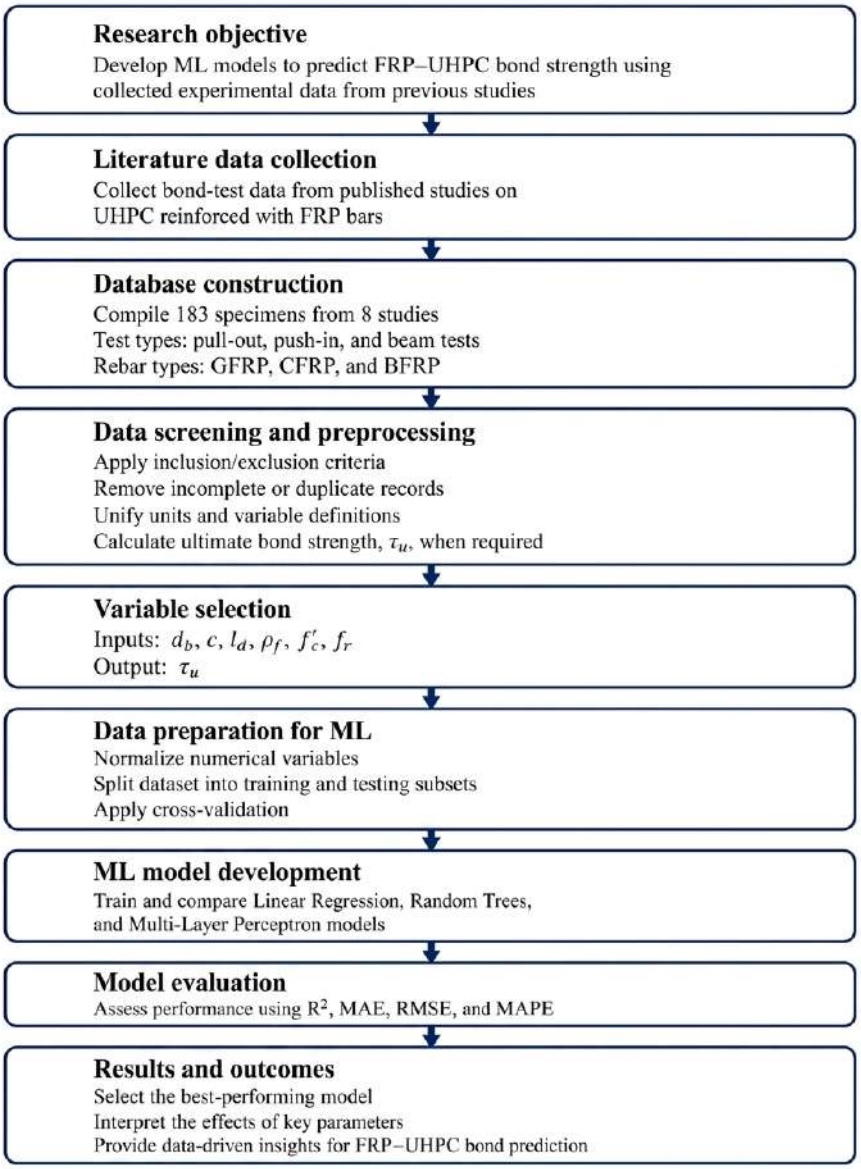


Figure 1. Systematic Workflow for Investigating and Modeling FRP–UHPC Interfacial Bond Performance

2.3 Machine-learning models

Several supervised machine-learning models were developed to predict the ultimate bond strength of FRP bars embedded in UHPC. The selected models included both simple and advanced algorithms to compare their prediction accuracy and ability to capture nonlinear relationships. Four models have been used, which are, Linear Regression (LR), Random Trees (RT)**,** Multi-Layer Perceptron neural network (MLP), and Locally Weighted Learning (LWL). LR was used as a first model because it provides a simple connection the input variables to τ_u. However, τ_u behavior is highly nonlinear because it depends on several interacting parameters, such as concrete strength, fiber content, bar diameter, embedment length, and cover thickness [3, 29, 32, 33]. Therefore, nonlinear ML models were also used. Tree-based models such as Decision Tree, Random Forest, and Gradient Boosting can

capture nonlinear effects and variable interactions. Support Vector Regression can provide good performance for small to medium-size datasets, while the Multi-Layer Perceptron model can learn complex nonlinear relationships through one or more hidden layers coupled with nonlinear activation functions [34]. A correlation matrix was developed to evaluate the linear relationship of the inputs and bond strength. The analysis included bar surface, concrete cover, bar size, bonded length, fiber content, concrete compressive strength, bar tensile strength, and bond strength. Since bar surface is a categorical variable, it was converted into numerical labels before analysis. The heatmap was used to visualize the strength and direction of the correlations. As shown in Figure 2, positive coefficients indicate direct associations, whereas negative coefficients represent inverse associations. Final dataset was separated into training and testing subsets. Training subset was used to develop ML models, whereas testing subset was compared to evaluate the prediction performance. The dataset was split to 80 and 20% for training and testing, respectively. Because the database is relatively limited, cross-validation should also be applied to improve the reliability of the performance evaluation and reduce the effect of random data splitting. For each model, the training process was conducted using the selected input variables and the target bond strength. Hyperparameters were tuned to improve model accuracy and reduce overfitting. The final model parameters were selected based on the performance on the validation or cross-validation results. To ensure a fair comparison, the four models were trained and tested using the same dataset split and the same performance metrics.

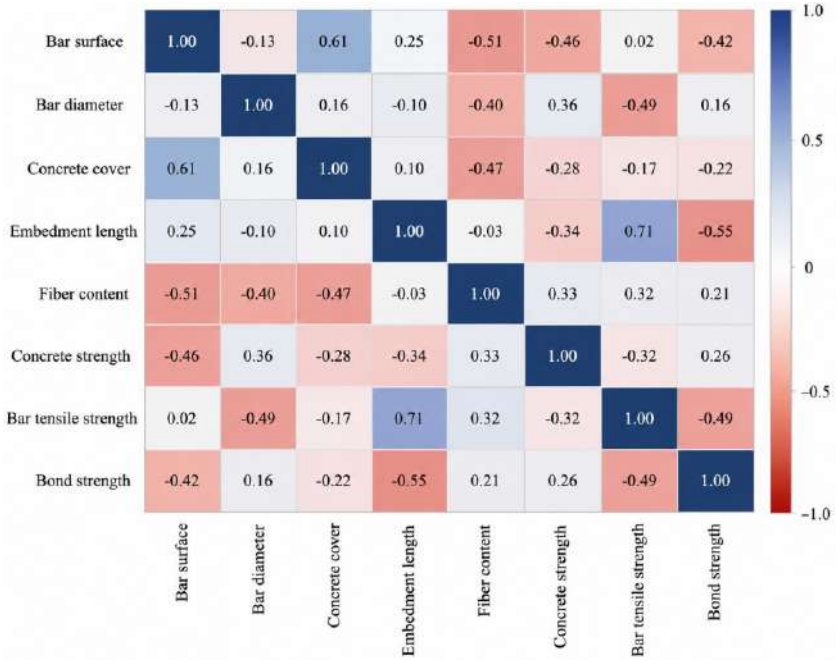


Figure 2. Representing correlation matrix heatmap for collective Dataset.

2.3.1 Random Trees (RT)

Random Trees constitute a powerful supervised machine learning approach that is particularly well suited to capturing nonlinear relationships in engineering datasets. In the context of structural and materials engineering, Random Trees are well suited for predicting performance characteristics where multiple interacting variables influence the response Figure 3. RT algorithm was employed to predict the (τ_u) between FRP and (UHPC), a problem characterized by significant parametric uncertainty and nonlinear interactions among governing variables such as bar size, concrete cover, bonded length, reinforcement type, and surface condition. Random Trees model is an ensemble-based learning technique derived from decision trees algorithms. It constructs multiple decision trees during the training phase using randomly selected subsets of input variables and data samples. Each individual trees generates a prediction, and the last output is obtained by aggregating these estimated values, typically through averaging. This randomness in feature selection and data sampling enhances model robustness and reduces the likelihood of overfitting, a common limitation of single decision trees models. Consequently, Random Trees can capture complex variable interactions and nonlinear dependencies while maintaining strong generalization capability for unseen data.

In the present research, the Random Trees model was trained using experimentally generated bond strength data. the prediction accuracy was evaluated through statistical metrics including the correlation coefficient, mean absolute error, and root mean squared error. The model showed dependable estimation accuracy and offered useful information about the relative contribution of the input variables to bond performance. A key advantage of Random Trees is their inherent can rank input variable importance, thereby assisting in identifying the most influential parameters governing bond strength. This capability supports engineering interpretation and facilitates the development of optimized design strategies. Overall, the Random Trees approach offers a reliable and transparent data-driven approach for estimating interfacial bonding performance in advanced composite–concrete systems.

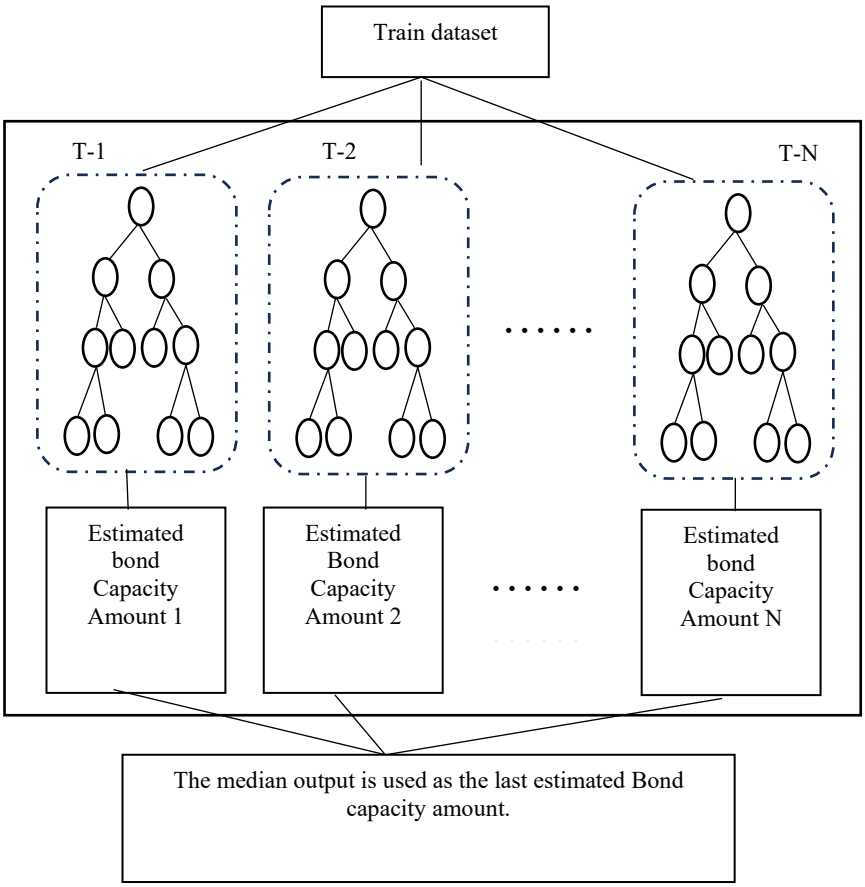


Figure 3. Random Trees diagram

2.3.2 Multi-Layer Perceptron

(MLP) is a feedforward supervised neural network that approximates complex, nonlinear relationships between the input variables and the target outputs by composing multiple layers of linear transforms and elementwise nonlinear activation functions. In the present study, an MLP surrogate was trained to predict bond outcomes (e.g., mean ultimate bond stress or τ_u) from experimental and derived inputs such as reinforcement type and surface, bar diameter, development length, concrete cover, load factor and texture/diagnostic metrics. The MLP's capacity to learn high-order interactions makes it particularly well suited for UHPC-FRP bond problems, where adhesion, friction and mechanical interlock interact nonlinearly and where matrix constituents, bar geometry and surface texture jointly influence failure modes (as discussed in the literature review). In this research, the MLP model was trained using experimentally derived datasets to develop a predictive connections of governing values and ultimate bond strength. The performanceof the model was evaluated using statistical indicators such as correlation coefficient, mean absolute error, and root mean squared error. MLP achieved higher prediction accuracy than conventional linear regression and tree-based models, indicating its strong capability in modeling nonlinear interfacial behavior in composite structural systems. The

implementation of MLP not only enhances prediction reliability but also supports the development of data-driven design tools for UHPC–FRP bonded systems. Consequently, the MLP-based framework provides an efficient and robust approach for understanding and predicting bond behavior, contributing to the advancement of intelligent modeling techniques in modern structural engineering research Figure 4.

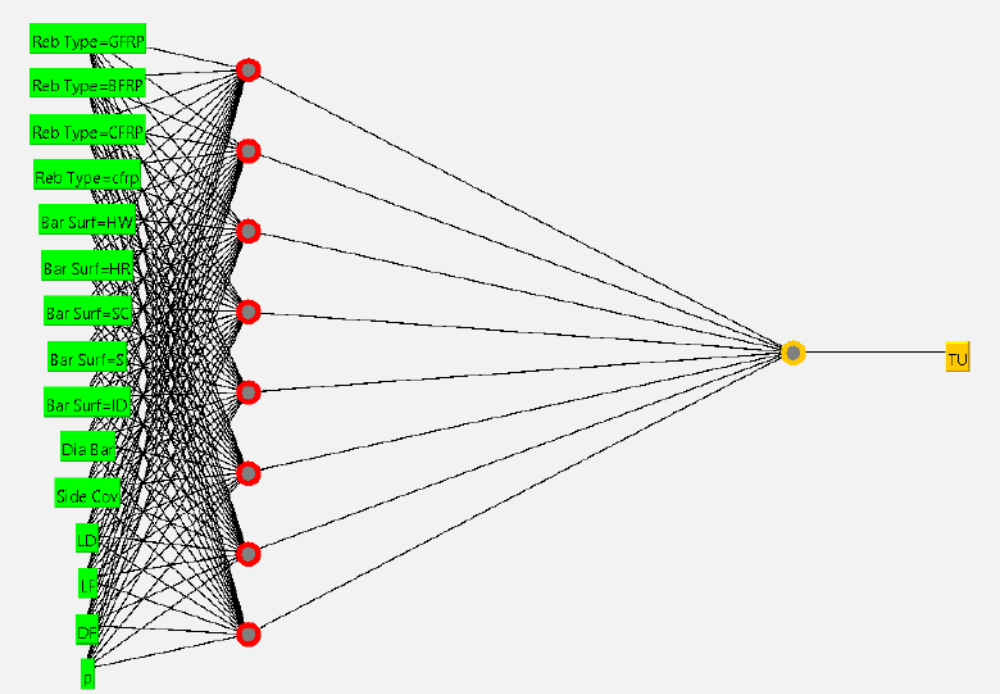


Figure 4. MLP diagram

2.3.3 Locally Weighted Learning (LWL)

It is an instance-based or lazy machine learning technique that performs prediction by giving higher importance to training samples located near the query instance. Unlike global learning algorithms that build a single generalized model from the entire dataset, LWL constructs a local model dynamically for each prediction using a subset of neighboring data points. The fundamental idea behind LWL is that complex nonlinear relationships within a dataset can be better captured by focusing on local patterns rather than assuming a single global function.

In LWL, a weight is allocated to each training sample according to its distance from the query point, typically using kernel functions such as Gaussian or inverse distance weighting. Instances closer to the query point receive higher weights and therefore have greater influence on the prediction. A simple regression or classification model is then fitted locally using these weighted instances to generate the final prediction. This approach enables LWL to adapt to variations in data distribution and effectively model nonlinear relationships. LWL offers advantages such as flexibility, simplicity, and the capacity

to process complex datasets without requiring extensive model training. However, it may require substantial computational resources when applied to large datasets because a local model must be constructed for each prediction. In civil engineering and materials research, LWL has been increasingly applied to predict material properties and structural performance due to its strong capability in modeling nonlinear and heterogeneous data patterns [34, 35].

3. Results and Discussion

The accuracy of the ML models was examed by using statistical performance indicators. The main evaluation metrics were the coefficient of determination (R^2), mean absolute error (MAE), root mean square error (RMSE), and mean absolute percentage error (MAPE). These metrics were selected because together, these metrics offer complementary perspectives on evaluation of the model. The (R^2) indicates how closely the predicted results agree with the experimental observations. A value close to 1 indicates high prediction accuracy. MAE represents the mean absolute deviation between the experimental and predicted bond strength values. RMSE gives higher weight to large errors and is useful for identifying models with large prediction deviations. MAPE represents the estimation deviation in percentage terms. The evaluation metrics were determined using the following equations:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{\bar{y}}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y}_i)^2}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{\bar{y}}_i|$$

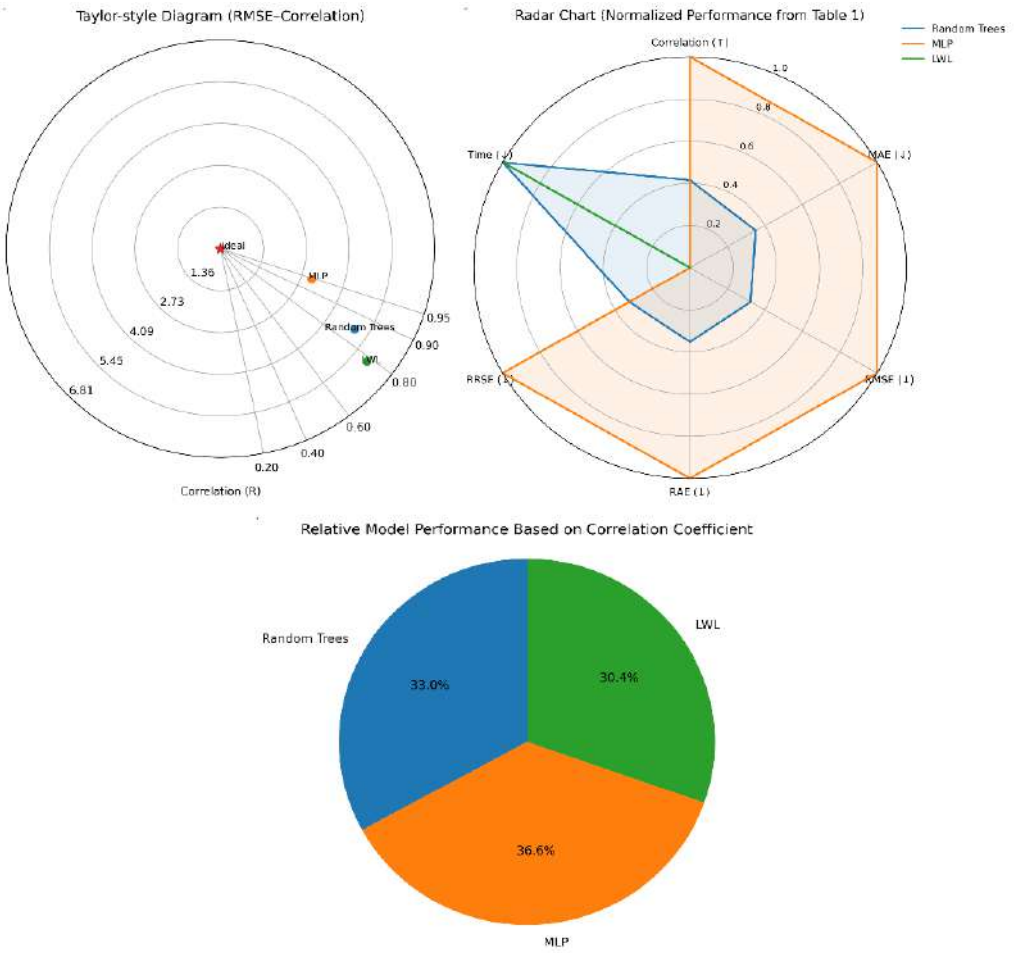
$$RMAE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{\bar{y}}_i)^2}$$

Table 2: Performace analysis table for the 3 MLs

ML	Correlation coefficient	Mean absolute error	Root mean squared error	Relative absolute error	Root relative squared error	Time Required (s)
Linear Regression	0.84	3.53	5.16	50.58%	54.42 %	0.063
Random Trees	0.85	3.64	4.99	51.81 %	52.35 %	0.16
MLP	0.94	2.30	3.06	32.84 %	32.09 %	0.13
LWL	0.78	4.36	5.92	62.13 %	62.05 %	0.6

Figure 5 shows a visual comparison of the models performance of the developed ML models for estimating the τ_u. This figure supports the numerical results reported in Table 2 and helps identify the best-performing model in a clearer way. The figure includes three complementary charts: a Taylor diagram, a radar chart, and a pie chart. These charts present different views of model performance, including prediction accuracy, error level, overall balance between evaluation metrics, and relative model performance. The compared models include (MLP), (RT), (LWL). Overall, Figure 5 indicates that MLP model achieved the highest predictive performance among the models examined.

Figure 5(a) shows the Taylor diagram, which provides an integrated statistical comparison between predicted and experimental bond strength values. This diagram simultaneously evaluates correlation coefficient and (RMSE), enabling a clear assessment of model accuracy. The Multi-Layer Perceptron (MLP) model lies nearest to the reference point, reflecting the strongest correlation and lowest prediction error among all models. The Random Trees model demonstrates moderate predictive capability, while the LWL model shows larger deviation from the reference point, reflecting lower prediction accuracy. Figure 5(b) presents a radar chart illustrating normalized performance indices of the several models across multiplstatistical metrics, including (R^2), (MAE), (RMSE), and (MAPE). The MLP model exhibits the most balanced and superior performance across all evaluation criteria, forming the largest enclosed area on the radar chart. Random Trees display moderate performance, whereas LWL occupies the smallest region, confirming its relatively weaker predictive capacity. Figure 5(c) displays a pie chart representing the relative contribution of each model based on correlation coefficient values. The chart highlights the dominant performance of the MLP model, followed by Random Trees and LWL. Collectively, the three subfigures provide a clear and integrated evaluation of model effectiveness, demonstrating that the MLP approach offers the most reliable and accurate predictive capability for modeling FRP–UHPC bond behavior.



Figuer 5: Performance comparison of ML models for estimating FRP–UHPC τ_u : (a) Taylor diagram, (b) radar chart, and (c) pie chart representation.

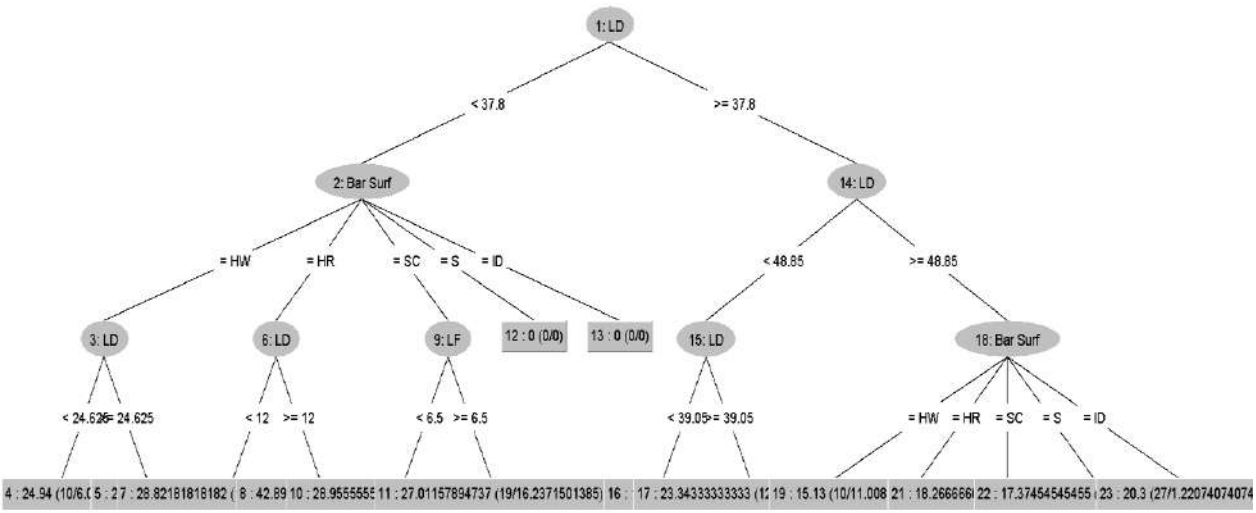


Figure 6: Random Trees diagram

Figure 6 presents the Random Trees model generated using the WEKA platform to estimate the τ_u of FRP bars embedded in UHPC. RT model was developed using the collected database of 183 specimens and the selected input variables, which are, test type, rebar type, bar surface condition, bar diameter d_b side cover c, embedment length L_d, fiber content ρ_f, UHPC compressive strength f_c', and FRP tensile strength f_r. The Random Trees model was constructed with a maximum depth of 3 and examed using 10-fold cross-validation. Trees structure shows that embedment length is the first splitting variable, with the main threshold occurring at approximately l_d = 37.8 mm. This indicates that embedment length plays a major important in controlling the predicted of τ_u. For shorter embedment lengths, bar surface condition becomes an important secondary parameter, where sand-coated bars and helically ribbed generally show higher predicted bond strength due to improved mechanical interlock and friction at the FRP–UHPC interface. For longer embedment lengths, the predicted bond strength generally decreases, which can be attributed to the nonuniform distribution of stress along the bonded depth and the resulting decrease in average bond stress. RT model further shows that surface type remains influential even in the longer embedment-length region, confirming that τ_u is controlled by the interaction between embedment length and bar surface condition. The model achieved a correlation coefficient of R^2 = 0.85, MAE of 3.64 MPa, and RMSE of 4.99 MPa, indicating reasonable prediction accuracy. Therefore, the Random Trees model provides both predictive capability and useful physical interpretation of the main parameters controlling FRP–UHPC bond behavior.

In addition , Figure 7 a,b,c,d compares the prediction performance of the developed ML models, including Linear Regression, Random Trees, Multi-Layer Perceptron, and Locally Weighted Learning, respectively. The comparison was based on the main statistical indicators, including (R^2), (MAE), (RMSE), and (MAPE). The results show that the nonlinear and local learning models generally performed better than the simple Linear Regression model. In particular, the MLP achieved the strongest overall results, yielding the highest correlation coefficient and the smallest error measures, indicating better agreement between the predicted and experimental bond strength values. Random Trees also showed good prediction accuracy and was able to capture part of the nonlinear behavior in the database. In contrast, Linear Regression and LWL showed relatively higher errors, suggesting that they were less effective for this dataset. Overall, the comparison confirms that the τ_u of FRP bars embedded in UHPC is controlled by nonlinear relationships between geometric and material parameters, and that local or tree-based ML models can provide more reliable predictions than simple linear modeling. In addition, SHAP analysis was employed to explain the ML model and determine the principal variables influencing the estimated τ_u, see Figure 8. It shows the influence of each input parameter on the predicted τ_u. The SHAP results showed that embedment length was the most

important parameter. This means that the bonded length has a the highest effective on the stress transfer between FRP bar and UHPC. When embedment length increases, the average bond stress may decrease because the bond stress is not distributed equally along the bonded depth. Bar surface condition was also important because ribbed, helically wound, or sand-coated bars can increase friction and mechanical interlock with UHPC. FRP tensile strength also affected τ_u because it is related to the mechanical capacity, and the resistance of the interface. Concrete cover thickness had an important effect because higher cover can give better confinement around the bar and delay splitting failure. Therefore, the SHAP results confirm that the τ_u of FRP bars in UHPC is mainly controlled by embedment length, surface condition, bar properties, and concrete confinement.

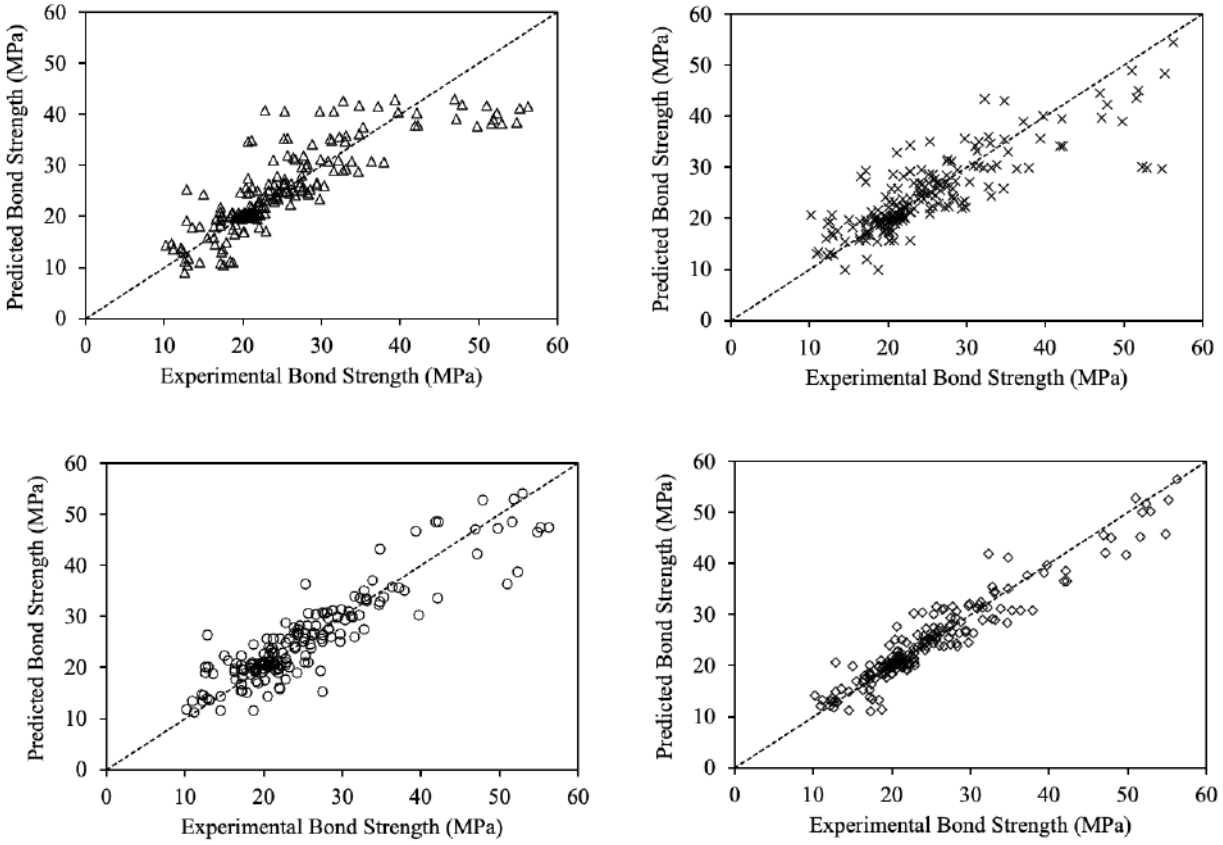


Figure 7. Assessment for ML models (a,b,c,and d) representing Linear Regression, Random Trees, Multi-Layer Perceptron, and Locally Weighted Learning, respectively.

The findings of this study agree with those reported in previous related research. The MLP model gave the highest prediction performance, which confirms that the τ_u of FRP bars embedded in UHPC is controlled by nonlinear relationships between the input parameters. This agrees with previous machine-learning studies, which showed that nonlinear models can predict FRP–concrete bond

behavior better than simple linear models. Also, the SHAP results showed that embedment length was the most important parameter, followed by bar surface condition, FRP tensile strength, and concrete cover. This agrees with experimental studies that reported the important effect of bonded length, surface roughness, mechanical interlock, and concrete confinement on bond strength. Therefore, the key contribution of this study is that it does not only predict the τ_u, but also explains the most important inputs affecting the FRP–UHPC bond behavior.

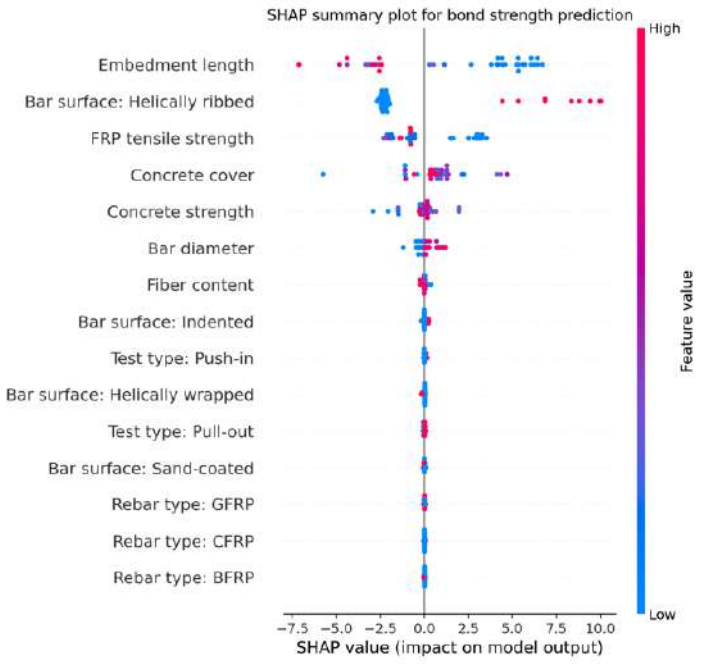


Figure 8: SHAP summary plot of bond strength

4. Conclusions

This study focused on the prediction and interpretation of the τ_u of FRP bars with UHPC. Research method was based on a data-driven machine-learning approach using a collected experimental database from previous studies. The database included 183 experimental tested samples with different FRP types, bar surface conditions, bar size, embedment lengths, concrete covers, fiber contents, UHPC compressive strengths, and FRP tensile strengths. The collected data were used to develop and compare Linear Regression, Random Trees, Multi-Layer Perceptron, and Locally Weighted Learning models. SHAP was also used to explain the influence of the variables on the predicted bond strength.

The comparison between the developed ML models showed that the Multi-Layer Perceptron provided the best prediction performance. The MLP model achieved a correlation coefficient of 0.9466, MAE of 2.30 MPa, RMSE of 3.06 MPa, RAE of 32.84%, and RRSE of 32.09%. Linear Regression and Random Trees showed moderate prediction accuracy, while LWL gave the lowest performance among

the evaluated models. The superior performance of MLP confirms that nonlinear learning models are more suitable for predicting FRP–UHPC bond strength because they can capture complex interactions between the selected input variables.

The results confirm that the bond behavior of FRP–UHPC systems is controlled by several interacting geometric, material, and surface-related parameters. This interaction shows that the bond response is nonlinear and cannot be fully represented by a simple linear relationship. The SHAP analysis showed that embedment length was the highest influence variable affecting the predicted τ_u, followed by bar surface condition, FRP tensile strength, and concrete cover. This indicates that the bonded length, surface configuration, bar properties, and confinement around the embedded bar contribute to the overall bond response.

Overall, the developed ML framework provides a useful data-driven approach for estimating the τ_u of FRP bars embedded in UHPC. The results also help identify the most important parameters controlling the bond response, which can support the development of UHPC-specific bond prediction models. However, the models should be used within the range of the collected database because the data were obtained from different experimental studies with different test methods, material properties, and specimen configurations.

Future work should focus on expanding the database by including more experimental results with different FRP types, bar surface configurations, UHPC mixture designs, fiber contents, cover thicknesses, and failure modes. Additional variables such as rib geometry, surface roughness, curing condition, and complete bond–slip response should also be included when available. This will improve model generalization, strengthen the physical interpretation of the ML results, and support the development of more reliable design-oriented bond models for FRP–UHPC structural systems.

Funding

The author declare that this research received no external funding.

Data Availability Statement

The data is available upon request to the corresponding author.

Conflicts of Interest

The authors declare no conflict of interest / or state any conflict if any.

References:

[1] H. Fergani, M. Di Benedetti, C. Miàs Oller, C. Lynsdale, and M. Guadagnini, "Durability and degradation mechanisms of GFRP reinforcement subjected to severe environments and sustained stress," *Constr. Build. Mater.*, vol. 170, pp. 637–648, May 2018, doi: 10.1016/j.conbuildmat.2018.03.092.

[2] L. Le Gué, P. Davies, M. Arhant, B. Vincent, and W. Verbouwe, "Basalt fibre degradation in seawater and consequences for long term composite reinforcement," *Compos. Part A Appl. Sci. Manuf.*, vol. 179, p. 108027, Apr. 2024, doi: 10.1016/j.compositesa.2024.108027.

[3] S. Abdal *et al.*, "Application of Ultra-High-Performance Concrete in Bridge Engineering: Current Status, Limitations, Challenges, and Future Prospects," *Buildings*, vol. 13, no. 1, p. 185, Jan. 2023, doi: 10.3390/buildings13010185.

[4] A. Qasem, Y. S. Sallam, H. Hossam Eldien, and B. H. Ahangarn, "Bond-slip behavior between ultra-high-performance concrete and carbon fiber reinforced polymer bars using a pull-out test and numerical modelling," *Constr. Build. Mater.*, vol. 260, p. 119857, Nov. 2020, doi: 10.1016/j.conbuildmat.2020.119857.

[5] C. S. Sun, N. Farzana, D. Kuruppuarachchi, M. Azimi, and H. Zhong, "Experimental and Analytical Studies of an Ultra-high Performance Concrete Beam Using Carbon Fiber Reinforced Polymer Bars," *Recent Prog. Mater.*, vol. 05, no. 02, pp. 1–17, May 2023, doi: 10.21926/rpm.2302019.

[6] X. Cao, D. He, K. Qian, F. Fu, X. Deng, and L. Wang, "Shear behavior of glass FRP bars-reinforced ultra-high performance concrete I-shaped beams," *Structural Concrete*, vol. 24, no. 1, pp. 1503–1520, Feb. 2023, doi: 10.1002/suco.202100801.

[7] Y. Luo, P. Liao, R. Pan, J. Zou, and X. Zhou, "Effect of bar diameter on bond performance of helically ribbed GFRP bar to UHPC," *Journal of Building Engineering*, vol. 91, p. 109577, Aug. 2024, doi: 10.1016/j.jobe.2024.109577.

[8] R. Kumar Tipu, V. Batra, Suman, K. S. Pandya, and V. R. Panchal, "Shear capacity prediction for FRCM-strengthened RC beams using Hybrid ReLU-Activated BPNN model," *Structures*, vol. 58, p. 105432, Dec. 2023, doi: 10.1016/j.istruc.2023.105432.

[9] M. Alhusban, M. Alhusban, and A. A. Alkhawaldeh, "The Efficiency of Using Machine Learning Techniques in Fiber-Reinforced-Polymer Applications in Structural Engineering," *Sustainability*, vol. 16, no. 1, p. 11, Dec. 2023, doi: 10.3390/su16010011.

[10] M. Kovačević, M. Hadzima-Nyarko, P. Petronijević, T. Vasiljević, and M. Radomirović, "Comparative Analysis of Machine Learning Models for Predicting Interfacial Bond Strength of Fiber-Reinforced Polymer-Concrete," *Computation*, vol. 13, no. 1, p. 17, Jan. 2025, doi: 10.3390/computation13010017.

[11] C. Arndt, C. Crusenberry, B. Heng, R. Butler, and S. TerMaath, "Reduced-Dimension Surrogate Modeling to Characterize the Damage Tolerance of Composite/Metal Structures," *Modelling*, vol. 4, no. 4, pp. 485–514, Nov. 2023, doi: 10.3390/modelling4040028.

[12] D.-Y. Yoo and Y.-S. Yoon, "Bond behavior of GFRP and steel bars in ultra-high-performance fiber-reinforced concrete," *Advanced Composite Materials*, vol. 26, no. 6, pp. 493–510, Nov. 2017, doi: 10.1080/09243046.2016.1197493.

[13] Y. Luo, P. Liao, R. Pan, J. Zou, and X. Zhou, "Effect of bar diameter on bond performance of helically ribbed GFRP bar to UHPC," *Journal of Building Engineering*, vol. 91, p. 109577, Aug. 2024, doi: 10.1016/j.jobe.2024.109577.

[14] C. Wang, Y. Jiao, J. Xing, and Y. Chen, "Bond Behavior of High-Strength Steel Rebar in Ultra-High-Performance Manufactured Sand Concrete: Experiment and Modelling," *Buildings*, vol. 14, no. 8, p. 2292, Jul. 2024, doi: 10.3390/buildings14082292.

[15] Y.-W. Chan and S.-H. Chu, "Effect of silica fume on steel fiber bond characteristics in reactive powder concrete," *Cem. Concr. Res.*, vol. 34, no. 7, pp. 1167–1172, Jul. 2004, doi: 10.1016/j.cemconres.2003.12.023.

[16] Z. Wu, C. Shi, and K. H. Khayat, "Influence of silica fume content on microstructure development and bond to steel fiber in ultra-high strength cement-based materials (UHSC)," *Cem. Concr. Compos.*, vol. 71, pp. 97–109, Aug. 2016, doi: 10.1016/j.cemconcomp.2016.05.005.

[17] L. Ke, K. Ma, Z. Chen, Y. Li, G. Chen, and Z. Feng, "Bond-slip and bond strength models for FRP bars embedded in ultra-high-performance concrete: A critical review," *Structures*, vol. 64, p. 106551, Jun. 2024, doi: 10.1016/j.istruc.2024.106551.

[18] Y. J. , & A. A. Kim, "Interface Shear between Ordinary Concrete and Ultra-High-Performance Concrete with Glass Fiber-Reinforced Polymer Reinforcing Bars," *ACI Struct. J.*, vol. 123, no. 3, May 2026, doi: 10.14359/51749317.

[19] Z. Shan, K. Liang, and L. Chen, "Bond behavior of helically wound FRP bars with different surface characteristics in fiber-reinforced concrete," *Journal of Building Engineering*, vol. 65, p. 105504, Apr. 2023, doi: 10.1016/j.jobe.2022.105504.

[20] K. Liang, L. Chen, Z. Shan, and R. K. L. Su, "Experimental and theoretical study on bond behavior of helically wound FRP bars with different rib geometry embedded in ultra-high-

performance concrete," *Eng. Struct.*, vol. 281, p. 115769, Apr. 2023, doi: 10.1016/j.engstruct.2023.115769.

[21] R. Pan, P. Liao, J. Zou, X. Zhou, and C. Li, "Experimental investigation on bond performance of helically ribbed GFRP rebar embedded in Ultra-high performance concrete," *Constr. Build. Mater.*, vol. 357, Nov. 2022, doi: 10.1016/j.conbuildmat.2022.129295.

[22] R. Pan, J. Zou, P. Liao, S. Dong, and J. Deng, "Effects of fiber content and concrete cover on the local bond behavior of helically ribbed GFRP bar and UHPC," *Journal of Building Engineering*, vol. 80, p. 107939, Dec. 2023, doi: 10.1016/j.jobe.2023.107939.

[23] F. Peng, H. Guan, J. Deng, and W. Xue, "Bond-slip behavior and model of sand-coated deformed GFRP bars in ultra-high-performance concrete," *Advances in Structural Engineering*, vol. 28, no. 8, pp. 1385–1402, Jun. 2025, doi: 10.1177/13694332241310113.

[24] F. AL-mahmoud, A. Castel, R. François, and C. Tourneur, "Effect of surface pre-conditioning on bond of carbon fibre reinforced polymer rods to concrete," *Cem. Concr. Compos.*, vol. 29, no. 9, pp. 677–689, Oct. 2007, doi: 10.1016/j.cemconcomp.2007.04.010.

[25] Y. J. Kim and J. Wang, "Interface of Ultra-High-Performance Concrete with Steel, Glass Fiber-Reinforced Polymer, and Basalt Fiber-Reinforced Polymer Reinforcing Bars," *ACI Struct. J.*, vol. 119, no. 3, pp. 3–15, May 2022, doi: 10.14359/51732640.

[26] F. Sayed Ahmad, G. Foret, and R. Le Roy, "Bond between carbon fibre-reinforced polymer (CFRP) bars and ultra high performance fibre reinforced concrete (UHPFRC): Experimental study," *Constr. Build. Mater.*, vol. 25, no. 2, pp. 479–485, Feb. 2011, doi: 10.1016/j.conbuildmat.2010.02.006.

[27] K. M. A. Hossain, D. Ametrano, and M. Lachemi, "Bond strength of GFRP bars in ultra-high strength concrete using RILEM beam tests," *Journal of Building Engineering*, vol. 10, pp. 69–79, Mar. 2017, doi: 10.1016/j.jobe.2017.02.005.

[28] N. Khalil and J. J. Assaad, "Bond properties between smooth carbon fibre-reinforced polymer bars and ultra-high performance concrete modified with polymeric latexes and fibres," *European Journal of Environmental and Civil Engineering*, vol. 26, no. 13, pp. 6211–6228, Oct. 2022, doi: 10.1080/19648189.2021.1934554.

[29] D. Y. Yoo, K. Y. Kwon, J. J. Park, and Y. S. Yoon, "Local bond-slip response of GFRP rebar in ultra-high-performance fiber-reinforced concrete," *Compos. Struct.*, vol. 120, pp. 53–64, Feb. 2015, doi: 10.1016/j.compstruct.2014.09.055.

[30] D.-Y. Yoo, J.-J. Park, S.-W. Kim, and Y.-S. Yoon, "Influence of reinforcing bar type on autogenous shrinkage stress and bond behavior of ultra high performance fiber reinforced concrete," *Cem. Concr. Compos.*, vol. 48, pp. 150–161, Apr. 2014, doi: 10.1016/j.cemconcomp.2013.11.014.

[31] J. Chen, Z. Fang, X. Chen, and R. Jiang, "Experimental study on lap behavior of CFRP indented bars in UHPC," *Constr. Build. Mater.*, vol. 344, p. 127959, Aug. 2022, doi: 10.1016/j.conbuildmat.2022.127959.

[32] A. Alatify and Y. J. Kim, "SP-363-8: Artificial Neural Networks for Prediction of Bond Strength between UHPC and FRP Reinforcing Bars."

[33] S. Abdal *et al.*, "Application of Ultra-High-Performance Concrete in Bridge Engineering: Current Status, Limitations, Challenges, and Future Prospects," *Buildings*, vol. 13, no. 1, p. 185, Jan. 2023, doi: 10.3390/buildings13010185.

[34] C. G. Atkeson, A. W. Moore, and S. Schaal, "Locally Weighted Learning," *Artif. Intell. Rev.*, vol. 11, no. 1–5, pp. 11–73, Feb. 1997, doi: 10.1023/A:1006559212014.

[35] "The WEKA workbench," in *Data Mining*, Elsevier, 2017, pp. 553–571. doi: 10.1016/B978-0-12-804291-5.00024-6.

Islamic University Journal of Applied Sciences (IUJAS)
https://journals.iu.edu.sa/jesc

Volume VIII, Issue II, December 2026, Pages 149-168

Safety Climate Perception Gaps Across Job Roles in Multinational Construction Workforces: Evidence from Saudi Arabia

Wael M. Alruqi [1]

[1] Civil Engineering Department, College of Engineering, Jouf University, Sakaka 72388, Saudi Arabia, waelalrruqi@ju.edu.sa

*Corresponding author: (Wael M. Alruqi), *Email Address*: waelalrruqi@ju.edu.sa

Abstract

Safety performance in construction is strongly influenced by how workers understand and respond to safety practices, particularly in complex and multicultural work environments. Although safety climate has been widely studied, limited attention has been given to how these perceptions differ across job roles in multinational construction settings. This study examines variations in safety climate perceptions across job roles within a multinational construction workforce. A survey was conducted with 256 construction workers representing different occupational groups. The study focused on key safety climate dimensions, including management commitment, supervisory practices, safety training, and individual responsibility for safety and health. Data were analyzed using reliability and validity tests, descriptive statistics, and one-way ANOVA. Overall, workers across the different occupational groups reported comparable views on several aspects of safety climate. However, differences became more noticeable when workers evaluated supervisory safety practices. Frontline and skilled workers, in particular, tended to describe less positive day-to-day supervisory safety experiences than other job groups. These findings suggest that supervisory interactions may not be experienced consistently across all levels of the workforce.

Keywords: Safety Climate; Construction Safety; Job Role Differences; Supervisor Safety Role; Multinational Workforce.

https://doi.org/10.63070/jesc.2026.038
Received 05 April 2026; Revised 20 May 2026; Accepted 08 June 2026; Available online 15 July 2026.
Published by Islamic University of Madinah on behalf of ***Islamic University Journal of Applied Sciences.***
This is an open access article published under the Creative Commons Attribution 4.0 International license (CC BY 4.0):https://creativecommons.org/licenses/by/4.0/

List of Abbreviations

Abbreviation	Full Term
ANOVA	Analysis of Variance
EFA	Exploratory Factor Analysis
KMO	Kaiser–Meyer–Olkin
CM	Management Commitment to Safety
CS	Supervisor Safety Role
CT	Safety Training
CW	Individual Responsibility for Safety and Health

1. Introduction

Construction work continues to involve a high level of occupational risk compared with many other industries. Workers are regularly exposed to changing site conditions, physically demanding activities, heavy equipment, and time pressure, all of which increase the possibility of accidents and injuries [1]. These challenges become more difficult to manage in multinational projects where employees may differ in language, cultural background, and communication style [2], [3]. As a result, improving safety performance continues to be a major concern for both researchers and practitioners. Researchers frequently use the concept of safety climate to better understand why safety performance differs across workplaces. In construction settings, workers often judge the importance of safety based on what they experience during daily operations rather than on written procedures alone [4] [5], [6]. Rather than focusing only on written policies and procedures, safety climate reflects how workers actually experience safety in their day-to-day activities [7], [8]. It is shaped by routine interactions at the worksite, including communication with supervisors, management priorities, and how safety rules are applied during daily operations [9]. This perspective suggests that safety outcomes are shaped not only by established rules and procedures, but also by how these are implemented and reinforced in practice [10]. From a theoretical perspective, safety climate is often viewed as operating at two levels: 1. organization level; and 2. group level. At the organizational level, factors such as management commitment and formal safety systems tend to create relatively consistent perceptions across the workforce [11], [12]. In contrast, supervisors influence safety at a more immediate level, where daily interactions with workers can lead to differences in how safety is experienced [13]. This issue becomes especially relevant in construction projects because supervisors directly influence how safety expectations are applied during routine site activities [14]. Although safety climate has been examined extensively in construction research, less attention has been given to how workers in different occupational positions experience safety within the same project environment [9]. Construction sites

are organized around different levels of responsibility and authority, meaning workers may not interpret safety practices in the same way [14]. These differences may become even more important in multinational projects where communication styles and cultural expectations vary across the workforce [2]. However, previous research has identified key dimensions of safety climate in the construction industry such as management commitment, supervisor safety role, safety training, and individual responsibility as important predictors of safety performance [15]. While these dimensions have been widely validated, relatively little attention has been given to whether they are perceived differently across different job role and among multinational workforce. Understanding these differences is important for developing safety interventions that better reflect the needs and experiences of different workers [5]. Despite this gap, most existing studies continue to examine safety climate at an aggregated level, without clearly accounting for how perceptions may differ across job roles in complex work environments. This study addresses this limitation by taking a role-based perspective to explore how key safety climate dimensions are perceived by different groups within a multinational construction setting. The objective is to determine whether meaningful differences exist between job roles and to better understand how safety is experienced across the workforce. In doing so, the study provides a more detailed view of safety climate and supports the development of more targeted safety management practices Accordingly, this study aims to examine differences in safety climate perceptions across job roles within a multinational construction workforce. By considering multiple dimensions of safety climate and comparing perceptions across occupational categories, the study seeks to provide a more detailed understanding of how safety is experienced at different levels of the organization. The findings are expected to contribute to both theory and practice by highlighting the influence of job role on safety perceptions and by supporting the development of more effective, role-sensitive safety management strategies. While this study adopts a multi-level perspective to interpret safety climate, the empirical analysis focuses on group-level comparisons rather than formal multilevel modeling. Indeed, this study aims to addresses the following research questions:

RQ1: Do safety climate perceptions differ significantly across job roles in multinational construction environments?

RQ2: Which dimensions of safety climate exhibit the greatest variation across job roles?

2. Literature review

2.1. Safety climate in construction

The construction industry continues to be associated with a high level of workplace hazards due to the dynamic nature of construction activities and the constantly changing site environment [1], [14]. In

construction environments, safety climate is commonly used to describe how workers interpret the organization's approach toward safety and risk management. Zohar in (1980), introduced the concept of safety climate, which refers to how workers perceive and make sense of safety priorities, expectations, and practices within their organization [7]. Over the years, this concept has been widely studied across different industries, and a large body of research confirms its strong link to safety outcomes [8], [9]. Additionally, the construct of safety climate is also associated with two key forms of safety behavior: safety compliance and safety participation [10]. Safety compliance relates to following required safety procedures, whereas safety participation reflects voluntary actions that contribute to improving safety at the workplace [8].

Previous review studies and meta-analyses consistently show that safety climate is closely connected to worker safety behavior and overall safety performance [16]. Likewise, [9] reported consistent relationships between safety climate and safety outcomes across various settings. In construction-industry, various safety climate dimensions have been shown a positive association safety performance [17] [18] [15]. Despite these findings, the concept of safety climate in construction industry remains challenging due to the nature of the work environment. Construction sites are dynamic, involve temporary teams, and operate within hierarchical structures. These challenges are even more evident in multinational settings, where differences in culture and communication may influence how safety is perceived by workers.

2.2. Management commitment to safety

Management commitment is widely recognized as an important factor influencing workplace safety [19], [20]. It reflects how strongly organizational leaders prioritize safety through all aspects [12]. Workers tend to take safety more seriously when they see clear support from management. Visible involvement from organizational leaders, such as addressing hazards, discussing safety concerns, and supporting training efforts, can strengthen workers' confidence in the organization's safety priorities. Several studies indicated that management commitment is positively associated with safety performance in construction [15], [21], [22], [23]. Workers who perceive management as genuinely committed to safety are more likely to follow safety procedures and actively engage in safe work practices [10] . In contrast, when management commitment to safety is weak or inconsistent, safety initiatives tend to lose effectiveness and may even encourage unsafe behaviors. In construction projects, management commitment becomes particularly important because of the risky tasks and the need for coordination among multiple parties [24]. Strong leadership at this level helps ensure that safety is consistently prioritized across all operations in the jobsite.

2.3. Supervisor safety role

The supervisor safety role represents how safety is implemented on a daily basis at the operational level. Supervisors play a direct role in shaping how safety procedures are implemented during daily site operations [13]. Daily communication with supervisors often shapes how workers judge safety conditions at the site. The way supervisors provide instructions, respond to unsafe behavior, and interact with workers can strongly influence workers' perceptions of workplace safety [25] [26]. The literature presents mixed evidence regarding the effectiveness of supervisory practices. Supportive supervisors who communicate openly and encourage worker involvement are often associated with higher levels of safety participation [27], [28]. However, in construction environments, supervisors are often required to balance productivity demands with safety requirements [29]. As a result, some studies indicated that supervisor safety role and behavior int the jobsite significantly affect safety performance [15]. These findings highlight the importance of examining how supervisory roles are experienced by workers in different job categories.

2.4. Safety training

Safety training is commonly used in construction projects to improve workers' awareness of hazards and safe work practices [30]. Effective training programs help workers understand site risks, recognize unsafe conditions, and perform tasks more safely [31]. Training programs are not equally effective in every situation. Workers are more likely to benefit from training when the information is practical, clearly explained, and directly connected to their daily tasks at the worksite [30]. In many construction organizations, training programs are standardized, which may result in relatively similar perceptions of training across different job roles. At the same time, research indicates that safety training can strengthen individual responsibility for safety, suggesting that it influences not only knowledge but also attitudes toward safety [32], [33]. This highlights its role in promoting a more proactive safety culture in jobsite.

2.5. Individual responsibility for safety

Individual responsibility for safety refers to the extent to which workers feel personally accountable for maintaining safe working conditions [15]. Individual responsibility for safety concept linked to the concept of safety motivation and is considered a key driver of safety behavior [10], [16]. Workers who view safety as part of their personal responsibility are more likely to comply with safety rules and also engage in voluntary safety related activities [8]. In construction industry, this becomes especially important because many tasks require individual judgment and decision making in changing

conditions. Empirical research indicates that individual responsibility plays a meaningful role in shaping safety conditions within construction environments [15].

2.6. Job role differences in safety climate

Although safety climate has been widely explored, relatively few studies have focused on how perceptions differ across job roles and more specifically among diverse construction workers. Construction projects are structured around multiple hierarchical levels, including frontline workers, supervisors, and management, each with distinct responsibilities and experiences [14]. Research suggests that safety perceptions may vary across these levels due to differences in authority, communication access, and exposure to risks [13]. For instance, managers often evaluate safety based on policies and formal systems, whereas workers tend to base their perceptions on daily interactions and actual site practices [11]. In multinational construction environments, these differences may be further shaped by cultural diversity [3]. Workers from different backgrounds may interpret safety practices differently, which can lead to variation in safety perceptions [29]. While previous studies have examined leadership and cultural influences separately, fewer have directly addressed differences across job roles in such contexts.

2.7. Theoretical framing and identified research gap

Traditionally, safety climate has been treated as a shared organizational construct, often measured at the group or company level [7]. While this approach has provided valuable insights, it tends to overlook differences that may exist within organizations, particularly across hierarchical levels. Previous research has emphasized the role of leadership in shaping safety climate, distinguishing between management level commitment and supervisor level practices [12], [20]. These perspectives suggest that both strategic and operational leadership are important in influencing safety behavior. However, less attention has been given to how these influences may be perceived differently by workers in different roles. In multinational construction environments, these differences may be further influenced by cultural diversity [3], [34].Workers from different national backgrounds may interpret leadership behavior and safety expectations in different ways, which can contribute to variation in safety perceptions. However, most existing studies have examined either cultural factors or leadership effects separately, rather than considering how job role and safety climate interact. Accordingly, this study addresses this gap by adopting a role based perspective to examine whether safety climate perceptions vary significantly across job roles in a multinational construction workforce.

3. Methods

This section outlines the overall methodological approach adopted in this study. To provide a clear overview of the research process, Figure 1 presents the main steps followed, including participant recruitment, survey administration, data preparation, and statistical analysis. The figure summarizes how the study progressed from data collection to the final interpretation of results.

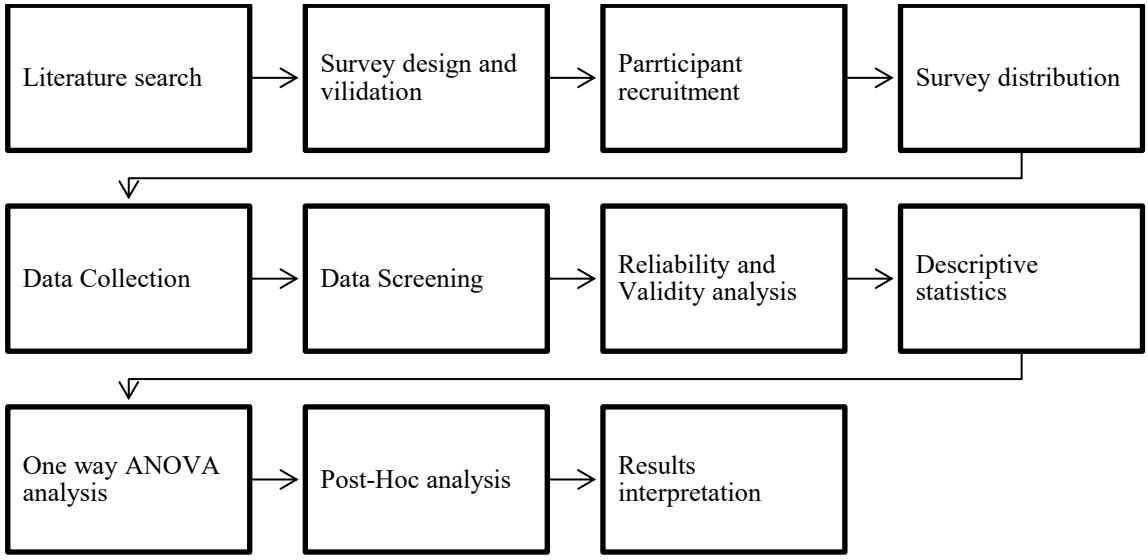


Fig. 1 . Overview of the research methodology and data analysis process

3.1. Research design and data collection

This study used a quantitative approach to address the research question. A survey tool used to examine how safety climate perceptions vary across job roles in a multinational construction site in the Kingdome of Saudi Arabia. Survey methods are widely used in safety-related research because they allow workers to describe their own experiences and perceptions in a structured way [8], [9]. This method allowed the study to assess these perceptions at a single point in time and compare responses across different occupational groups. The research protocol was reviewed and approved by Jouf University. Participation was voluntary, and informed consent was obtained from all respondents prior to data collection. Data were collected from a different construction site employing a multinational workforce. Workers were invited to participate through general site announcements, and those who were willing completed the survey voluntarily. This approach is commonly used in construction research due to practical access constraints while still enabling participation from a wide range of job roles. The final sample included 256 respondents representing frontline workers, skilled trades, supervisory staff, and support personnel, providing a broad representation of a multinational construction workforce. To account for linguistic diversity, the survey was administered in multiple languages. The original instrument was developed in English and translated into Arabic, Hindi, Urdu, and Filipino. A back-translation procedure was

used to verify translation accuracy, where each version was independently translated back into English and compared with the original. Any discrepancies were reviewed and resolved to ensure consistency in meaning across all language versions [35].

3.2. Measurement of variables

The survey instrument consisted of two main sections. The first section collected demographic information, including age, nationality, and job role. These variables were later used to categorize respondents for comparative analysis. The second section measured safety climate perceptions across several key dimensions. The selection of these dimensions based on the prior meta analytic research identifying the most influential predictors of safety climate dimensions in construction [15]. These dimensions are: management commitment to safety, safety rules and procedures, supervisor safety role, safety training, and individual responsibility for safety and health. All constructs were measured using a five-point Likert scale ranging from strong disagreement to strong agreement. A three items adopted from [20] study used to measure the construct of management commitment to safety. Management commitment to safety captured the extent to which organizational top leaders promotes safety in the organization. Additionally, a three items adopted from the same study to measure safety rules. Safety rules and procedures reflected workers' perceptions of the availability, clarity, and enforcement of formal safety requirements within the organization [20]. The supervisor safety role represented the influence of frontline supervisors on safety practices, and four items were adopted from Zohar study (Zohar, 2000). Furthermore, Safety training was assessed using three items adopted from [8] to examine how workers perceive the priority, relevance, and comprehensiveness of training programs related to workplace health and safety [8]. Individual responsibility for safety and health captured the extent to which workers take personal responsibility for following safety procedures and maintaining safe work practices, and it assessed by adopting three items from Griffin & Nea (2000) study.

3.3. Data analysis procedures

Before conducting the analysis, the dataset was checked for missing and inaccurate values. Cronbach's alpha was applied to evaluate the reliability and internal consistency of the survey items [36]. Alpha values above 0.60 were considered acceptable because each construct contained a small number of items. Construct validity was evaluated using exploratory factor analysis (EFA) with principal component analysis and Varimax rotation. The Kaiser–Meyer–Olkin (KMO) test and Bartlett's test of sphericity were also used to confirm that the data were appropriate for factor analysis [37] . Items with weak factor loadings or low communalities were removed where

necessary to improve the overall measurement structure. In addition, the factor structure was examined to assess how survey items grouped across constructs. This step helped to evaluate whether items associated with each safety climate dimension formed distinct components. While most constructs showed clear grouping patterns, some overlap was observed between safety training and individual responsibility, indicating limited separation between these dimensions in this context. Descriptive statistics were also calculated to summarize the variables. One way ANOVA conducted to assess the diffrences among job roles groups. The ANOVA model can be expressed as:

$$F = \frac{\text{Between} - \text{group variance}}{within - \text{group variance}}$$

This method is appropriate for comparing mean values across multiple independent groups [38]. The assumption of homogeneity of variances was tested using Levene's test. When this assumption was violated, the Welch ANOVA was used as a more robust alternative [39]. Post hoc comparisons were conducted using the Games Howell procedure, which is suitable for unequal group sizes and heterogeneous variances [40].

4. Result

The survey was conducted across multiple construction sites, resulting in 256 valid responses from workers representing different job roles. The sample reflects a diverse workforce, allowing for comparison of safety climate perceptions across occupational groups.

4.1. Reliability analysis result

Table 1: Internal Consistency of Safety Climate Constructs

Construct	Number of Items	Cronbach's Alpha (α)	Interpretation
Management commitment to safety	3	0.800	Good
Supervisor safety role	4	0.613	Acceptable
Safety rules and procedures	3	0.043	Unacceptable
Safety training	3	0.676	Acceptable
Individual responsibility for safety and health	3	0.775	Acceptable

Cronbach's alpha was used to check the reliability of the study variables before conducting the main analysis. As shown in Table 1, most variables showed acceptable reliability levels. Management commitment to safety had good reliability (α = 0.800), while individual responsibility for safety and health also showed good consistency (α = 0.775). Safety training had an acceptable reliability

value (α = 0.676). The supervisor safety role variable showed a slightly lower value (α = 0.613), but it was still acceptable for this type of study and the small number of survey items. However, safety rules and procedures showed a very low reliability value (α = 0.043), meaning the items were not consistent with each other. The low reliability result indicates that the survey items may not have reflected a single consistent concept within this study sample. A closer examination indicates that this may be related to differences in interpretation among respondents, particularly given the multilingual and multicultural nature of the sample. One possible explanation is that workers from different linguistic and cultural backgrounds may have interpreted the items differently. As a result, the construct was excluded from the later stages of analysis to avoid affecting the reliability of the findings. Since the construct showed extremely weak internal consistency, it was removed from the later analysis to avoid weakening the overall quality of the statistical results [36].

4.2. Construct validity result

Exploratory factor analysis (EFA) was used to examine the validity of the study variables. Before running the analysis, the suitability of the data was checked using the Kaiser–Meyer–Olkin (KMO) test and Bartlett's test of sphericity, as presented in Tables 2, 3, and 4. The KMO value was 0.802, showing that the sample size was adequate for factor analysis. Bartlett's test was also significant ($p < 0.001$), indicating that the data were suitable for this analysis. The results identified three main factors that explained 61.14% of the total variance. After rotation, the factor structure was clear and easy to interpret. Items related to management commitment grouped together under the first factor, while safety training and individual responsibility items loaded mainly on the second factor. It is also important to note that items related to safety training and individual responsibility loaded on the same factor. This suggests a degree of conceptual overlap between these constructs in this study context. One possible explanation is that safety training in construction environments may directly influence workers' sense of personal responsibility, leading respondents to perceive these concepts as closely connected. While this does not invalidate the constructs, it indicates limited discriminant separation between them. This issue is acknowledged as a limitation and should be considered when interpreting the results. The third factor was mainly associated with supervisor safety role items. However, one item from the supervisor role scale and one item from the safety training scale showed weak loadings and low communalities. These items were removed from the final analysis to improve the quality of the measurement model. Overall, the findings showed that most survey items measured their intended constructs well and supported the validity of the study variables.

Table2: KMO and Bartlett's Test

Test	Value
Kaiser–Meyer–Olkin (KMO)	0.802
Bartlett's Test (χ^2)	1152.539
df	66
Sig.	< 0.001

Table 3: Total variance explained

Component	Eigenvalue	% Variance	Cumulative %
1	4.389	36.58	36.58
2	1.814	15.11	51.69
3	1.134	9.45	61.14

Table 4: Rotated component matrix (Loadings ≥ 0.50)

Item	Component 1	Component 2	Component 3
CM1	0.726	—	—
CM2	0.823	—	—
CM3	0.758	—	—
CS1	0.776	—	—
CS3	—	—	0.924
CS4	—	—	0.882
CT2	—	0.677	—
CT3	—	0.729	—
CW2	—	0.791	—
CW3	—	0.743	—

Table 5: Overall descriptive statistics of study variables

Variable	N	Mean	Standard Deviation	Minimum	Maximum
Management Commitment to Safety	256	4.30	0.63	1.00	5.00
Supervisor Safety Role	256	3.61	0.82	1.75	5.00
Safety Training	256	4.40	0.63	1.50	5.00
Individual Responsibility for Safety and Health	256	4.59	0.51	2.00	5.00

Descriptive statistics were computed to examine the average and variability of the key study variables among respondents. As shown in Table 5, all constructs recorded relatively high mean values, indicating generally positive perceptions of safety climate among the respondents. Individual responsibility for safety and health exhibited the highest mean score (M = 4.59, SD = 0.51), suggesting that workers strongly recognize their personal role in maintaining safe working

conditions. Safety training also received a high rating (M = 4.40, SD = 0.63), reflecting favorable perceptions regarding the availability and effectiveness of training programs. Management commitment to safety was rated positively as well (M = 4.30, SD = 0.63), indicating that respondents generally perceive strong organizational support for safety practices. In contrast, the supervisor safety role showed a comparatively lower mean (M = 3.61, SD = 0.82), although still within the positive range. This suggests greater variability in how workers perceived the supervisory practices. Overall, the findings indicate a favorable safety climate within the studied construction sites, with particularly strong perceptions of individual responsibility and management commitment, while supervisory practices appear to be more inconsistently experienced. While the overall results provide a general view of safety climate perceptions, they do not capture potential differences across job roles. To examine these differences in more detail, descriptive statistics were calculated for each job role category, as presented in Table 6.

Table 6. Descriptive statistics of safety climate dimensions by job role

Job Role	N	Management Commitment (Mean ± SD)	Supervisor Role (Mean ± SD)	Safety Training (Mean ± SD)	Individual Responsibility (Mean ± SD)
Frontline workers	52	4.34 ± 0.49	3.42 ± 0.70	4.43 ± 0.46	4.53 ± 0.43
Skilled trades	33	4.14 ± 0.70	3.36 ± 0.59	4.29 ± 0.78	4.48 ± 0.45
Supervisory staff	16	4.25 ± 0.61	3.56 ± 0.94	4.31 ± 0.63	4.48 ± 0.78
Support / other	155	4.33 ± 0.65	3.74 ± 0.87	4.43 ± 0.64	4.65 ± 0.51
Total	256	4.30 ± 0.63	3.61 ± 0.82	4.40 ± 0.63	4.59 ± 0.51

As shown in Table 6, the mean values for management commitment to safety, safety training, and individual responsibility are consistently high across all job roles, with only minor differences between groups. This suggests that these dimensions are generally perceived in a similar way across the workforce. In contrast, greater variation can be observed in the supervisor safety role, where frontline workers and skilled trades report lower mean scores compared to supervisory staff and the support/other group. Although these differences appear modest at the descriptive level, they indicate that supervisory practices may not be experienced uniformly across job roles. These observations provide an initial basis for further statistical testing to determine whether the differences are statistically significant. These patterns suggest that safety climate perceptions are influenced not only by job role but also by shared organizational practices. The relatively small differences observed

across groups indicate that safety-related policies and communication may be applied consistently across the workforce.

4.3. Differences in safety climate across job roles

A one-way ANOVA was performed to determine whether perceptions of the safety climate varied among different job roles. The results are shown in Table 7. Prior to analysis, the assumption of homogeneity of variances was assessed using Levene's test. For most variables, the assumption was satisfied; however, for the supervisor safety role, the assumption was violated ($p < 0.001$). Therefore, the Welch robust test was used to confirm the results for this variable. Although statistical differences were examined, the magnitude of these differences should be interpreted carefully, particularly in light of the reported effect sizes. This indicates that while statistical differences may exist, their practical importance may be limited, and other factors may play a more significant role in shaping safety perceptions.

Table 7. One-way ANOVA results for safety climate across job roles

Variable	F	p-value	Welch p-value	Effect Size (η^2)	Interpretation
Management Commitment	0.907	0.438	0.520	0.011	Not significant
Supervisor Safety Role	3.312	0.021*	0.012*	0.038	Significant
Safety Training	0.609	0.610	0.704	0.007	Not significant
Individual Responsibility	1.599	0.190	0.192	0.019	Not significant

Table 8. Post Hoc Comparisons (Games–Howell) for Supervisor Safety Role

Comparison	Mean Difference	p-value	Interpretation
Frontline vs Support/Other	-0.320	0.041	Significant
Skilled Trades vs Support/Other	-0.375	0.018	Significant
Other comparisons	—	>0.05	Not significant

4.3.1 Management commitment to safety

The results indicated that there were no statistically significant differences in perceived management commitment to safety across job roles among participants, $F(3, 252) = 0.907$, $p = 0.438$. The Welch test confirmed this finding ($p = 0.520$). The effect size was very small ($\eta^2 = 0.011$), suggesting minimal variation between groups. This indicates that perceptions of management commitment are relatively consistent across job roles, with only minimal variation between groups.

4.3.2 Supervisor safety role

The results showed a significant difference in how workers from different job roles viewed the supervisor's role in safety, $F(3, 252) = 3.312$, $p = 0.021$. Given the violation of homogeneity of variance, the Welch test was used and also indicated a significant effect ($p = 0.012$), confirming the

robustness of the result. The effect size was small ($\eta^2 = 0.038$), suggesting that job role accounts for only a limited proportion of the variance in perceptions. Post hoc comparisons using the Games–Howell test revealed that respondents in the support/other category reported significantly higher perceptions of supervisor safety role compared to frontline workers ($p = 0.041$) and skilled trades ($p = 0.018$). No significant differences were observed between other group pairs. This pattern suggests that differences in supervisory practices may be linked to how workers interact with supervisors in their daily tasks, rather than structural differences between job roles alone.

4.3.3 Safety training

The ANOVA results showed no statistically significant differences in perceived safety training across job roles, $F(3, 252) = 0.609$, $p = 0.610$. The Welch test further confirmed the absence of significant differences ($p = 0.704$). The effect size was negligible ($\eta^2 = 0.007$), indicating that safety training is perceived in a largely consistent manner across job roles.

4.3.4. *Individual responsibility for safety and health*

Similarly, no statistically significant differences were found in individual responsibility for safety and health across job roles, $F(3, 252) = 1.599$, $p = 0.190$. The Welch test yielded consistent results ($p = 0.192$). The effect size was small ($\eta^2 = 0.019$), suggesting that differences across job roles are minimal. Overall, while some statistically significant differences were identified, the consistently small effect sizes indicate that job role explains only a limited portion of variation in safety climate perceptions. This suggests that broader organizational factors, such as management practices and safety systems, may have a stronger influence on how safety is experienced across the workforce.

5. Discussion

5.1. Overview of findings

This study examined whether safety climate perceptions differ across job roles in multinational construction sites in Saudi Arabia. The results showed that workers across different job roles had similar views on most safety climate factors. However, perceptions of the supervisor's safety role were significantly different between job groups. These results suggest that while organizational-level safety practices are broadly shared and similarly interpreted across the workforce, supervisory practices are experienced differently depending on job role, highlighting the importance of frontline leadership in shaping safety perceptions. However, the magnitude of these differences was generally small, suggesting that job role plays a limited role in shaping overall safety climate perceptions.

5.2. Consistency of organizational safety climate dimensions

The lack of significant differences in perceptions of management commitment to safety suggests that workers in different job roles generally viewed the organization's safety priorities in a similar way. This may indicate that safety expectations and management support for safety are clearly communicated across the workforce [11], [12]. From a theoretical perspective, this supports the notion that safety climate operates as a collective construct, where shared perceptions emerge from consistent organizational signals and practices [7], [8]. Workers from different job roles may develop similar views about safety when management actively supports safety through clear policies, regular communication, and adequate resources. This result also agrees with previous studies showing that management commitment plays an important role in improving safety performance among different groups of workers [15], [31]. The small effect size suggests that organizational safety practices are applied in a largely consistent manner, resulting in only minor differences across job roles.

5.3. Uniform perceptions of safety training

Likewise, no significant differences were found in perceptions of safety training between job roles. This suggests that safety training programs are delivered in a similar and consistent manner across the workforce. This result is similar to findings from earlier studies showing that well organized safety training programs often lead to consistent outcomes among workers [30]. In construction environments, training is often mandatory and centrally managed, which may contribute to consistent perceptions regardless of job role. From a practical standpoint, this suggests that organizations have been effective in institutionalizing safety training, ensuring that all employees receive comparable information and guidance. However, it also raises questions about whether standardized training adequately addresses the specific needs of different job roles, particularly in complex and diverse work environments. The small effect size further supports this finding, indicating that differences across job roles are minimal in practical terms.

5.4. Stability of individual responsibility for safety

The results showed that workers across different job roles had similar views regarding individual responsibility for safety and health. This indicates that employees, regardless of their position, understand the importance of their own role in maintaining a safe work environment. Previous studies have also highlighted that personal commitment and safety awareness are important factors in encouraging safe behavior at work [10], [16]. The results suggest that safety is viewed as a shared responsibility across the organization, regardless of job role. This may reflect the role of training,

communication, and workplace practices in promoting a shared understanding of safety responsibilities among workers.

5.5. Variability in supervisor safety role

Supervisor safety role construct showed significant differences across job roles, indicating that supervisory practices are not experienced uniformly. Post hoc analysis, as shown in Table 8, revealed that workers in the support/other category reported higher perceptions of supervisor safety role compared to frontline and skilled workers. This finding highlights the critical role of supervisors as the interface between management and workers**.** While management commitment is consistently communicated, supervisory practices may vary in how they are implemented and enforced [13]. The observed differences may be explained by variations in daily interactions, communication patterns, and enforcement styles. This variation may reflect differences in daily interactions with supervisors, communication patterns, and perceptions of enforcement among frontline workers. Frontline and skilled workers, who are more directly exposed to hazards and supervisory oversight, may experience supervision differently compared to other groups. In some cases, supervisory practices may be perceived as overly directive or enforcement-oriented, which can influence how safety is interpreted at the operational level. Previous studies have shown that overly directive supervision may reduce workers' safety participation when it is seen as controlling rather than supportive [26], [41]. The findings also support the view that safety climate is not entirely homogeneous and may vary depending on the level of interaction with supervisory personnel. Although statistically significant differences were identified, the associated effect sizes were small, suggesting that job role accounts for only a limited share of variation in safety climate perceptions. Accordingly, the practical implications of these differences should be interpreted with caution. Additionally, it should also be noted that the "support/other" category includes a diverse set of roles. This heterogeneity may have influenced the observed differences and should be considered when interpreting the results. This suggests that while supervisory practices influence safety perceptions, their overall impact is limited compared to broader organizational factors.

5.6. Implications of the results

The findings indicate that safety climate operates at different levels within the organization. Management commitment and safety training were perceived consistently, while supervisor-related factors varied across job roles. This supports the view that safety climate can be understood as a multi-level concept, where organizational practices shape shared perceptions, and supervisors influence day-to-day experiences. This aligns with prior work on the hierarchical nature of safety climate [13]. From

a practical perspective, consistent perceptions of management commitment and training suggest that organizations have established effective safety systems. These should be maintained. However, variation in supervisory practices highlights the need for improvement at the frontline level. Supervisors should be trained to communicate clearly, support workers, and balance enforcement with engagement. The results also suggest that safety strategies should be tailored to different job roles. Frontline and skilled workers may experience supervision differently, and targeted interventions can help improve safety outcomes. At the same time, the small effect sizes observed in this study suggest that these differences should not be overstated, and that safety climate remains largely consistent across job roles.

6. Conclusion

This study examined differences in safety climate perceptions among workers in different job roles within a multinational construction workforce. The results showed that management commitment to safety, safety training, and individual responsibility for safety and health were viewed similarly across job roles. This indicates that these aspects of safety climate are generally consistent across the workforce. However, perceptions of the supervisor's safety role differed significantly between job groups. Specifically, frontline and skilled workers reported lower evaluations compared to other groups. This indicates that workers may experience supervisory safety practices differently depending on their position at the worksite. Overall, the results show that variation in safety climate is primarily associated with supervisory practices rather than organizational-level factors. These findings indicate that differences in safety perceptions are limited to specific dimensions rather than the overall safety climate. From a practical perspective, the study highlights the need to strengthen supervisory leadership and improve consistency between organizational safety policies and frontline safety practices. Overall, the findings suggest that construction organizations should consider workers' different job roles when developing and implementing safety management practices in multinational work environments.

7. Limitations and future research

Several limitations should be taken into account when reviewing the findings of this study. First, the study used a cross-sectional survey, so it cannot fully explain cause-and-effect relationships between the variables. Second, the data were based on workers' self-reported responses, which may include some response bias. Third, the study focused on workers from a specific multinational construction environment, so the findings may not fully represent other industries or work settings Future studies should examine safety climate over a longer period of time to better understand how workers'

perceptions may change. Researchers could also include objective safety data, such as accident or near-miss records, together with survey responses. In addition, future research may further explore the influence of cultural differences and specific job roles on safety perceptions in multinational construction projects.

Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request

Conflicts of Interest

The authors declare no conflict of interest / or state any conflict if any.

Acknowledgments

During the preparation of this work, the author used ChatGPT (GPT-5, OpenAI) to assist with improving the clarity, organization, and readability of manuscript sections.

References

[1] ILO, "A call for safer and Healthier Working Environments," 2023. [Online]. Available: https://www.ilo.org/sites/default/files/wcmsp5/groups/public/%40ed_protect/%40protrav/%40safework/documents/publication/wcms_903140.pdf

[2] R. M. Choudhry, D. Fang, and S. Mohamed, "The nature of safety culture: A survey of the state-of-the-art," *Saf. Sci.*, vol. 45, no. 10, pp. 993–1012, 2007, doi: https://doi.org/10.1016/j.ssci.2006.09.003.

[3] G. Hofstede, *Culture's consequences: Comparing values, behaviors, institutions and organizations across nations*. Sage publications, 2001.

[4] N. Xia, S. Ding, T. Ling, and Y. Tang, "Safety climate in construction: a systematic literature review," *Engineering, Construction and Architectural Management*, vol. 31, no. 10, pp. 3973–4000, Nov. 2024, doi: 10.1108/ECAM-11-2022-1074.

[5] P. Kines, L. P. Andersen, S. Spangenberg, K. L. Mikkelsen, J. Dyreborg, and D. Zohar, "Improving construction site safety through leader-based verbal safety communication," *J. Safety Res.*, vol. 41, no. 5, pp. 399–406, 2010, doi: https://doi.org/10.1016/j.jsr.2010.06.005.

[6] M. T. Newaz, P. R. Davis, M. Jefferies, and M. Pillay, "Developing a safety climate factor model in construction research and practice: A systematic review identifying future directions for research," *Engineering, Construction and Architectural Management*, vol. 25, no. 6, pp. 738–757, Jul. 2018, doi: 10.1108/ECAM-02-2017-0038.

[7] D. Zohar, "Safety climate in industrial organizations: theoretical and applied implications.," *Journal of applied psychology*, vol. 65, no. 1, p. 96, 1980.

[8] M. A. Griffin and A. Neal, "Perceptions of safety at work: a framework for linking safety climate to safety performance, knowledge, and motivation.," *J. Occup. Health Psychol.*, vol. 5, no. 3, p. 347, 2000.

[9] S. Clarke, "An integrative model of safety climate: Linking psychological climate and work attitudes to individual safety outcomes using meta-analysis," *J. Occup. Organ. Psychol.*, vol. 83, no. 3, pp. 553–578, Sep. 2010, doi: https://doi.org/10.1348/096317909X452122.

[10] A. Neal and M. A. Griffin, "A study of the lagged relationships among safety climate, safety motivation, safety behavior, and accidents at the individual and group levels.," *Journal of*

applied psychology, vol. 91, no. 4, p. 946, 2006, doi: https://doi.org/10.1037/0021-9010.91.4.946.

[11] D. Hofmann, F. P. Morgeson, and S. J. Gerras, "Climate as a moderator of the relationship between leader-member exchange and content specific citizenship: Safety climate as an exemplar," *Journal of Applied Psychology*, 2003, DOI:10.1037/0021-9010.88.1.170

[12] D. Zohar, "The effects of leadership dimensions, safety climate, and assigned priorities on minor injuries in work groups," *J. Organ. Behav.*, vol. 23, no. 1, pp. 75–92, Feb. 2002, doi: https://doi.org/10.1002/job.130.

[13] D. Zohar and G. Luria, "A multilevel model of safety climate: cross-level relationships between organization and group-level climates.," *Journal of applied psychology*, vol. 90, no. 4, p. 616, 2005.

[14] H. Lingard and S. Rowlinson, *Occupational health and safety in construction project management*. Routledge, 2004.

[15] W. M. Alruqi, M. R. Hallowell, and U. Techera, "Safety climate dimensions and their relationship to construction safety performance: A meta-analytic review," *Saf. Sci.*, vol. 109, pp. 165–173, Nov. 2018, doi: 10.1016/J.SSCI.2018.05.019.

[16] M. S. Christian, J. C. Bradley, J. C. Wallace, and M. J. Burke, "Workplace safety: a meta-analysis of the roles of person and situation factors.," *Journal of applied psychology*, vol. 94, no. 5, p. 1103, 2009, doi: https://doi.org/10.1037/a0016172.

[17] O. Kima, A. Urlainis, K. C. Wang, and I. M. Shohet, "Safety climate in small and medium construction enterprises," *Smart and Sustainable Built Environment*, 2024, doi: 10.1108/SASBE-01-2024-0038/1255936.

[18] K. J. Kajumulo, J. M. Matindana, and F. K. Mohamed, "The status of safety climate in building construction projects in Tanzania," *International Journal of Building Pathology and Adaptation*, pp. 1–32, Oct. 2025, doi: 10.1108/IJBPA-05-2024-0097.

[19] R. Flin, K. Mearns, P. O'Connor, and R. Bryden, "Measuring safety climate: identifying the common features," *Saf. Sci.*, vol. 34, no. 1, pp. 177–192, 2000, doi: https://doi.org/10.1016/S0925-7535(00)00012-6.

[20] S. Mohamed, "Safety Climate in Construction Site Environments," *J. Constr. Eng. Manag.*, vol. 128, no. 5, pp. 375–384, Oct. 2002, doi: 10.1061/(ASCE)0733-9364(2002)128:5(375).

[21] I. M. Shohet, R. Naveh, and F. Shahin, "Safety Climate in High-Rise Construction," *Buildings*, vol. 15, no. 9, p. 1398, 2025, doi: https://doi.org/10.3390/buildings15091398.

[22] O. Abudayyeh, T. K. Fredericks, S. E. Butt, and A. Shaar, "An investigation of management's commitment to construction safety," *International Journal of Project Management*, vol. 24, no. 2, pp. 167–174, 2006, doi: https://doi.org/10.1016/j.ijproman.2005.07.005.

[23] H. Abbasianjahromi and M. Aghakarimi, "Safety performance prediction and modification strategies for construction projects via machine learning techniques," *Engineering, Construction and Architectural Management*, vol. 30, no. 3, pp. 1146–1164, Dec. 2021, doi: 10.1108/ECAM-04-2021-0303.

[24] Y. Chen, B. McCabe, and D. Hyatt, "A resilience safety climate model predicting construction safety performance," *Saf. Sci.*, vol. 109, pp. 434–445, 2018, doi: https://doi.org/10.1016/j.ssci.2018.07.003.

[25] S. S. Sankar, K. S. Anandh, and K. Prasanna, "Safety leadership: a catalyst for positive safety climate on construction sites," *Buildings*, vol. 14, no. 6, p. 1806, 2024, doi: https://doi.org/10.3390/buildings14061806.

[26] J. Barling, C. Loughlin, and E. K. Kelloway, "Development and test of a model linking safety-specific transformational leadership and occupational safety.," *Journal of applied psychology*, vol. 87, no. 3, p. 488, 2002, doi: https://doi.org/10.1037/0021-9010.87.3.488.

[27] M. Ismail and M. S. Fathi, "Leadership in Construction: Leadership Styles Practiced in Construction Project – A Review," *Journal of Advanced Research in Business and*

Management Studies, vol. 13, no. 1, pp. 24–30, Oct. 2020, [Online]. Available: https://www.akademiabaru.com/submit/index.php/arbms/article/view/1312

[28] S. Senthamizh Sankar and K. S. Anandh, "Navigating leadership styles through qualitative exploration for enhanced safety in the construction sector," *Saf. Sci.*, vol. 175, p. 106495, 2024, doi: https://doi.org/10.1016/j.ssci.2024.106495.

[29] R. M. Choudhry and D. Fang, "Why operatives engage in unsafe work behavior: Investigating factors on construction sites," *Saf. Sci.*, vol. 46, no. 4, pp. 566–584, Apr. 2008, doi: 10.1016/J.SSCI.2007.06.027.

[30] M. J. Burke, S. A. Sarpy, K. Smith-Crowe, S. Chan-Serafin, R. O. Salvador, and G. Islam, "Relative Effectiveness of Worker Safety and Health Training Methods," *Am. J. Public Health*, vol. 96, no. 2, pp. 315–324, Feb. 2006, doi: 10.2105/AJPH.2004.059840.

[31] M. N. Vinodkumar and M. Bhasi, "Safety management practices and safety behaviour: Assessing the mediating role of safety knowledge and motivation," *Accid. Anal. Prev.*, vol. 42, no. 6, pp. 2082–2093, 2010, doi: 10.1016/j.aap.2010.06.021.

[32] P. Adami *et al.*, "Effectiveness of VR-based training on improving construction workers' knowledge, skills, and safety behavior in robotic teleoperation," *Advanced Engineering Informatics*, vol. 50, p. 101431, 2021, doi: 10.1016/j.aei.2021.101431

[33] J. Lin, Y. Cai, and Q. Li, "Development of safety training in construction: Literature review, scientometric analysis, and meta-analysis," *Journal of Management in Engineering*, vol. 39, no. 6, p. 03123002, 2023, doi: https://doi.org/10.1061/JMENEA.MEENG-5445.

[34] W. M. Alruqi, "Productivity Pressure and Risk Perception Among a Multinational Construction Workforce in Saudi Arabia," *Buildings*, vol. 16, no. 4, p. 774, 2026, doi: https://doi.org/10.3390/buildings16040774.

[35] R. W. Brislin, "Back-translation for cross-cultural research," *J. Cross. Cult. Psychol.*, vol. 1, no. 3, pp. 185–216, 1970.

[36] J. C. Nunnally and I. H. Bernstein, *Psychometric Theory*, 3rd ed. New York, NY, USA: McGraw-Hill, 1994.

[37] J. F. Hair, W. C. Black, B. J. Babin, and R. E. Anderson, *Multivariate Data Analysis*, 7th ed. Upper Saddle River, NJ, USA: Prentice Hall, 2010.

[38] A. Ross and V. L. Willson, "One-way anova," in *Basic and advanced statistical tests: Writing results sections and creating tables and figures*, Springer, 2017, pp. 21–24.

[39] B. L. Welch, "On the comparison of several mean values: an alternative approach," *Biometrika*, vol. 38, no. 3/4, pp. 330–336, 1951.

[40] P. A. Games and J. F. Howell, "Pairwise multiple comparison procedures with unequal n's and/or variances: a Monte Carlo study," *Journal of Educational Statistics*, vol. 1, no. 2, pp. 113–125, 1976.

[41] S. Clarke, "Safety leadership: A meta-analytic review of transformational and transactional leadership styles as antecedents of safety behaviours," *J. Occup. Organ. Psychol.*, vol. 86, no. 1, pp. 22–49, 2013.

Islamic University Journal of Applied Sciences (IUJAS)
https://journals.iu.edu.sa/jesc

Volume VIII, Issue II, December 2026, Pages 169-201

Forensic Science in the Arab World: Current Status, Challenges, and Future Directions

Reda Ibrahim[1], Yasir Alnafea[2] and Norhan Shoghy*[1]

[1]Economic Entomology Department, Kafrelsheikh University, Kafrelsheikh 33516, Egypt
[2]Department of Biology, College of Science, Taibah University, Madinah, Saudi Arabia.

*Corresponding author: Norhan Shoghy, *Email Address*: Norhan_Shoghy@agr.kfs.edu.eg

Abstract

Forensic science is a fundamental component of modern criminal justice systems, providing scientifically validated evidence for legal investigations. This review synthesizes evidence from 66 studies published between 2000 and 2025 to evaluate the current status of forensic science across Arab countries. Although the field has progressed in both theoretical knowledge and technological capabilities, considerable gaps remain between scientific advancement and practical implementation. The review examines the application of major disciplines, including DNA analysis, toxicology, digital forensics, and forensic entomology, within regional institutional and legal frameworks. Key challenges identified include fragmented governance, inadequate infrastructure, limited specialized expertise, weak integration between research and practice, and the absence of standardized protocols. The review also highlights emerging opportunities driven by technological innovation, capacity building, and regional cooperation. By synthesizing current evidence, the study proposes strategic directions to improve the effectiveness, standardization, and sustainability of forensic science systems in the Arab world, thereby strengthening evidence-based justice and enhancing public trust.

Keywords: Applied forensic science; forensic entomology; research–practice integration; forensic laboratories; legal frameworks; emerging technologies; Arab world

https://doi.org/10.63070/jesc.2026.039
Received 23 April2026; Revised 28 May 2026; Accepted 10 June 2026; Available online 15 July 2026.
Published by Islamic University of Madinah on behalf of ***Islamic University Journal of Applied Sciences.***
This is an open access article published under the Creative Commons Attribution 4.0 International license (CC BY 4.0):https://creativecommons.org/licenses/by/4.0/

1. Introduction

Forensic science is a multidisciplinary field that applies scientific principles and analytical methodologies to criminal investigations and the administration of justice. It integrates diverse disciplines, including biology, chemistry, medicine, anthropology, toxicology, and digital sciences, to generate objective and scientifically validated evidence for legal purposes. Over recent decades, forensic science has undergone substantial transformation driven by technological innovation, particularly in DNA profiling, forensic imaging, analytical instrumentation, and digital evidence analysis, resulting in major improvements in the accuracy, sensitivity, and reliability of criminal investigations [1–4].

Globally, forensic science has evolved from conventional descriptive methodologies toward highly advanced, technology-driven systems. The introduction of next-generation sequencing (NGS), rapid DNA technologies, and high-throughput analytical platforms has revolutionized forensic biology and human identification by enhancing analytical precision and evidential discrimination [4–6]. In parallel, the integration of artificial intelligence (AI), machine learning, and automated analytical systems has improved the interpretation of complex forensic datasets, including pattern recognition, probabilistic evidence evaluation, and digital evidence analysis [7–9]. Despite these scientific advances, considerable disparities remain in the implementation of modern forensic technologies, particularly in developing regions where financial, technical, institutional, and regulatory limitations continue to restrict forensic capacity and operational efficiency [10,11].

In the Arab world, forensic science has gained increasing recognition as an essential component of modern criminal justice systems. Several Arab countries have invested in establishing forensic laboratories, medico-legal institutes, and specialized academic and training programs aimed at strengthening national forensic capabilities [12–14]. Regional initiatives, collaborative training programs, and partnerships with international organizations further demonstrate a growing institutional commitment to advancing forensic science throughout the Middle East and North Africa [15,16]. Nevertheless, the development of forensic science across Arab countries remains heterogeneous, with substantial differences in infrastructure, technological resources, funding availability, research capacity, and human resource development [14].

Recent studies have documented notable progress in several core forensic disciplines, particularly forensic DNA analysis, toxicology, pathology, and digital forensics, especially in Egypt and Gulf Cooperation Council (GCC) countries [15,17–20]. However, these advancements are frequently

constrained by limited laboratory resources, inconsistent accreditation systems, lack of standardized operational protocols, and insufficient integration between academic research and practical forensic application [20,21]. Furthermore, important gaps persist in specialized disciplines such as forensic anthropology and forensic entomology, where region-specific databases, localized validation studies, and specialized expertise remain limited [21–23].

One of the major challenges facing forensic science in the Arab world is the persistent gap between theoretical advancement and practical implementation. Although academic institutions and research centers have contributed significantly to the expansion of forensic knowledge and scientific output, the translation of these advancements into effective operational forensic services remains inconsistent. This disconnect is influenced by multiple factors, including insufficient funding, shortage of specialized expertise, fragmented institutional governance, weak research–practice integration, and variability in legal and regulatory frameworks governing forensic evidence [10–12,15].

Therefore, this review critically examines the current status of forensic science in the Arab world, with particular emphasis on the relationship between theoretical foundations and practical implementation. It further explores major institutional and scientific challenges, emerging technological opportunities, and future strategic directions for strengthening forensic science systems throughout the region. By synthesizing current literature and regional developments, this review aims to provide a comprehensive framework for improving the integration of scientific innovation into forensic practice and enhancing the effectiveness, reliability, and sustainability of evidence-based justice systems in the Arab world.

2. Literature Search and Review Methodology

This review was conducted using a structured literature review approach to ensure comprehensive identification, evaluation, and synthesis of publications addressing forensic science in the Arab world. The methodological framework was informed by the principles of the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines to support transparent literature identification, screening, selection, and reporting processes [24,25]. In addition, methodological considerations from the Cochrane Handbook for Systematic Reviews were consulted to improve consistency and analytical rigor during study evaluation and synthesis [26].

2.1 Search Strategy

A comprehensive and systematic literature search was conducted across multiple electronic databases, including Scopus, Web of Science, PubMed, and Google Scholar. The search covered publications

from January 2000 to December 2025 to capture both foundational developments and recent advancements in forensic science relevant to the Arab world and the Middle East and North Africa.

Search queries were developed using Boolean operators (AND, OR) and combinations of keywords including “forensic science,” “Arab world,” “Middle East,” “forensic laboratories,” “DNA profiling,” “forensic toxicology,” “digital forensics,” “forensic entomology,” “forensic anthropology,” and “theory–practice gap.” Database-specific filters and search refinement tools were applied where appropriate to improve retrieval precision and relevance. Manual screening of reference lists from selected publications was also performed to identify additional relevant studies and institutional reports not retrieved during the primary database search.

2.2 Eligibility Criteria

Studies were selected according to predefined inclusion and exclusion criteria to ensure methodological consistency and relevance to the objectives of this review. Eligible publications included peer-reviewed journal articles, academic books, technical reports, and official institutional publications addressing forensic science disciplines, forensic systems, policies, or practical applications within the Arab world or the broader Middle East and North Africa. Only studies published in English and available in full text were considered for inclusion. Publications were excluded if they were unrelated to forensic science or criminal investigation, consisted solely of conference abstracts, editorials, or opinion-based commentaries without substantial analytical or empirical content, lacked methodological clarity, or represented duplicate records identified across multiple databases. The screening process was conducted progressively through title review, abstract evaluation, and full-text assessment to ensure scientific relevance and consistency.

2.3 Study Selection Process (PRISMA Flow)

The study selection process followed a PRISMA-guided workflow to ensure systematic identification and screening of relevant literature [24,25]. A total of 742 records were initially identified through database searching. Following the removal of 168 duplicate records, 574 publications remained for title and abstract screening. During this stage, 396 records were excluded because they did not meet the scope or relevance criteria of the review.

The remaining 178 studies underwent full-text evaluation. Of these, 112 publications were excluded due to insufficient regional relevance, limited methodological rigor, duplication of findings, or

inadequate focus on forensic science applications. Ultimately, 66 studies satisfied the inclusion criteria and were incorporated into the final qualitative synthesis.

The complete literature screening and study selection workflow based on the PRISMA framework is illustrated in Figure 1.

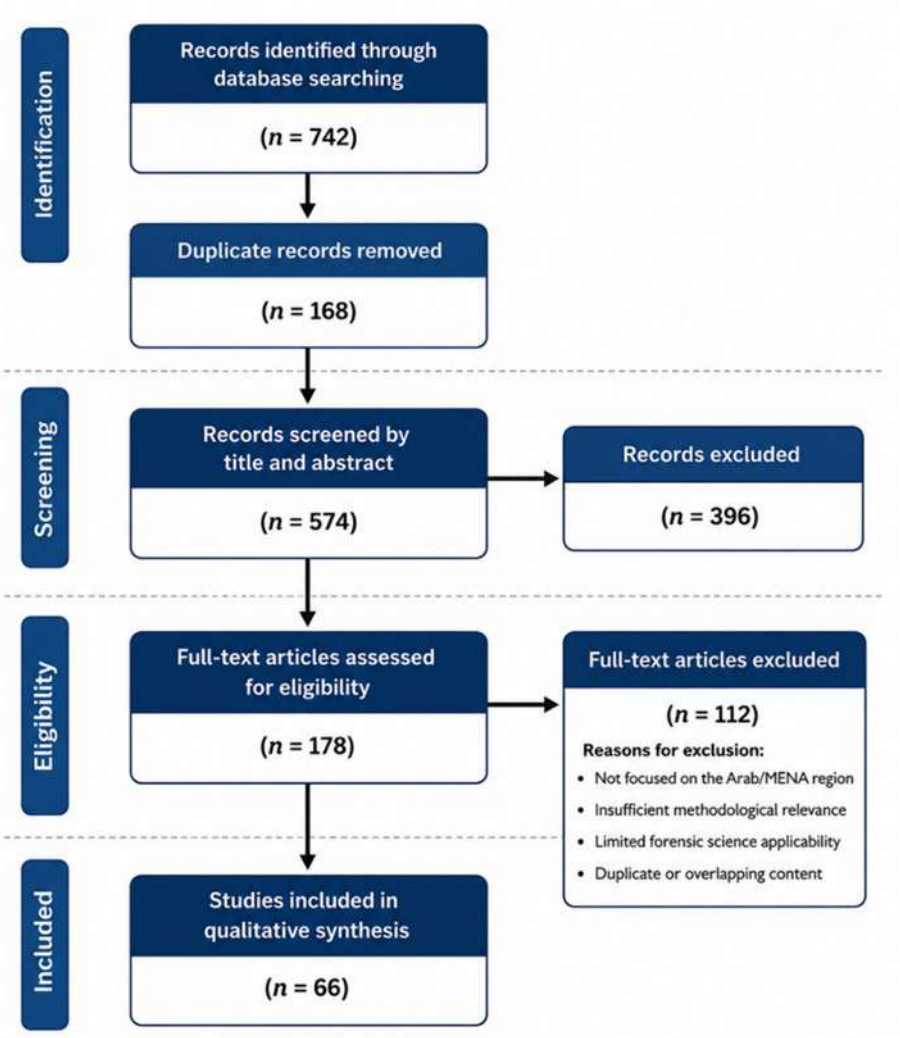


Figure 1. Flow diagram illustrating the literature identification, screening, eligibility assessment, and final study selection process used in this review.

2.4 Data Extraction

Data extraction was conducted using a standardized and structured approach to ensure consistency across all included studies. Information systematically collected from each publication included the authorship and year of publication, country or regional focus, forensic discipline addressed, study design, methodological approach, study objectives, principal findings, and identified limitations or

challenges. Particular attention was given to information related to the relationship between theoretical advancement and practical implementation in forensic science. This approach enabled a comprehensive evaluation of the theory–practice gap and facilitated comparative analysis of institutional, scientific, and operational aspects of forensic science systems within the Arab world. Representative studies covering diverse forensic disciplines, countries, and thematic areas were selected for tabular presentation (Table 1) to provide an illustrative overview of the evidence base without reproducing all included studies.

2.5 Data Synthesis and Thematic Analysis

A qualitative thematic synthesis approach was employed to analyse and interpret the extracted data. The included studies were systematically organized into major thematic categories, including scientific foundations of forensic science, practical forensic applications in the Arab world, institutional and infrastructural development, governance and legal frameworks, challenges and limitations, and future directions and strategic recommendations. This thematic analytical framework facilitated the identification of recurring patterns, regional disparities, emerging trends, and critical knowledge gaps. Particular emphasis was placed on evaluating the integration of scientific research and technological innovation into operational forensic practice and evidence-based judicial systems within the regional context.

2.6 Quality Assessment

The methodological quality of the included studies was critically evaluated using general quality indicators commonly applied in qualitative and narrative systematic reviews [26]. The assessment focused on several criteria, including clarity of research objectives, robustness of study design, transparency of methodology and data reporting, analytical depth, and overall relevance to the objectives of the review. Studies demonstrating limited methodological rigor, insufficient analytical detail, or inadequate transparency were interpreted cautiously during data synthesis to minimize potential bias and overgeneralization. This quality assessment process contributed to strengthening the reliability and scientific consistency of the review findings.

2.7 Limitations

Several limitations associated with this review should be acknowledged. First, restricting the search to English-language publications may have resulted in the exclusion of relevant studies published in Arabic or other regional languages. Second, variability in study design, methodological quality, and reporting standards across the included literature may affect direct comparability between studies.

In addition, publication bias remains a potential limitation because studies reporting significant findings are more likely to be published and indexed in major databases. Finally, the availability of region-specific forensic data remains limited in several specialized disciplines, particularly forensic entomology and forensic anthropology, which may restrict the comprehensiveness of some regional analyses.

A representative selection of key studies included in the review is presented in Table 1. These studies were chosen to illustrate the major forensic disciplines, geographical coverage, research themes, and implementation challenges identified across the 66 studies included in the qualitative synthesis.

Table 1. Representative studies included in the review, illustrating major forensic science disciplines, key findings, and identified challenges across the Arab world.

No.	Reference (Author, Year)	Country/ Region	Forensic Discipline	Study Design	Study Objective	Key Findings	Identified Limitations / Gaps
1	Al-Amad et al. (2019)	Middle East	General Forensic Science	Review	Assessment of forensic system development	Increasing institutional recognition of forensic science	Absence of unified regional standards
2	Alenezi & Almutairi (2021)	GCC Countries	Forensic Education	Cross-sectional study	Evaluation of educational and training programs	Expansion of academic and training initiatives	Limited integration with practical applications
3	Almazrouei et al. (2021)	United Arab Emirates	Forensic Services	Descriptive study	Analysis of forensic laboratory infrastructure	Establishment of advanced forensic facilities	Weak linkage between research and practice
4	El-Sayed et al. (2020)	Egypt	Forensic Medicine	Review	Evaluation of medico-legal services	Notable improvement in forensic medical services	Constraints in resources and capacity
5	El-Ashmawy et al. (2021)	Egypt	Forensic Toxicology	Experimental study	Investigation of toxicological analytical methods	Advancement in detection and analytical techniques	Lack of methodological standardization

No.	Reference (Author, Year)	Country/ Region	Forensic Discipline	Study Design	Study Objective	Key Findings	Identified Limitations / Gaps
6	Alharbi et al. (2022)	Saudi Arabia	Forensic Toxicology	Review	Assessment of toxicological practices	Expansion of forensic toxicology services	Shortage of specialized professionals
7	Al-Qahtani et al. (2019)	Saudi Arabia	DNA Analysis	Case study	Application of DNA profiling in forensic casework	Increased utilization in identification	Limited nationwide implementation
8	Blau (2016)	Middle East	Forensic Anthropology	Review	Human identification in regional contexts	Provides valuable anthropological insights	Lack of comprehensive regional databases
9	Al-Mesbah et al. (2012)	Arabian Peninsula	Forensic Entomology	Review	Review of entomological applications	Establishment of baseline regional knowledge	Scarcity of applied field studies
10	El-Azazy (2018)	Egypt	Forensic Entomology	Review	Evaluation of discipline status	Emerging recognition of forensic entomology	Limited operational integration
11	Natsheh (2025)	Arab world	Forensic Policy & Systems	Technical report	Capacity-building initiatives	Support for institutional development	Gaps in implementation and sustainability
12	Mansour et al. (2025)	Arab world	Forensic Systems	Technical report	Regional coordination efforts	Enhanced international cooperation	Fragmentation across national systems
13	Al-Shammari et al. (2020)	Gulf Region	Forensic Laboratories	Descriptive study	Infrastructure and system analysis	Expansion of laboratory capabilities	Lack of harmonized operational standards
14	Hassab El-Nabi et al. (2019)	Egypt	DNA Profiling	Experimental study	Evaluation of forensic DNA applications	Demonstrated effectiveness in identification	Limited accessibility and coverage

No.	Reference (Author, Year)	Country/ Region	Forensic Discipline	Study Design	Study Objective	Key Findings	Identified Limitations / Gaps
15	Charabidze et al. (2020)	Global (Arab world)	Forensic Entomology	Experimental study	Study of decomposition ecology	Provides important ecological insights	Limited regional-specific validation

3. Scientific Foundations and Emerging Trends in Forensic Science

Forensic science is founded on a multidisciplinary scientific framework that integrates principles from biological, chemical, medical, physical, and digital sciences to support criminal investigations and judicial processes. Its theoretical foundations are based on the systematic collection, preservation, analysis, interpretation, and presentation of physical and biological evidence using scientifically validated methodologies. Central to forensic science is the principle of scientific objectivity, which requires that conclusions be derived from reproducible, reliable, and evidence-based analytical procedures rather than subjective interpretation [1–3,6].

One of the most fundamental principles underpinning forensic science is Locard's exchange principle, which states that contact between individuals, objects, or environments results in the transfer of material evidence. This principle constitutes the scientific basis for evidence recovery, crime scene reconstruction, and trace evidence analysis, and it continues to guide the development of modern forensic methodologies across multiple forensic disciplines [1–3]. Over recent decades, forensic science has evolved considerably through the integration of advanced molecular, analytical, computational, and imaging technologies, leading to significant improvements in evidential sensitivity, analytical precision, and interpretative reliability [4–9].

3.1 Major Forensic Science Disciplines

3.1.1 Forensic Biology and DNA Technologies

Forensic biology, particularly forensic DNA analysis, represents one of the most transformative advancements in modern forensic science. The scientific basis of DNA profiling lies in the uniqueness and stability of genetic material, allowing highly accurate identification of individuals and biological relationships. Techniques such as polymerase chain reaction (PCR), short tandem repeat (STR) profiling, and mitochondrial DNA analysis have revolutionized criminal investigations by facilitating suspect identification, kinship analysis, disaster victim identification, and biological evidence comparison [4–6].

Recent technological developments, including next-generation sequencing (NGS), rapid DNA systems, and high-throughput genomic analysis, have further enhanced analytical sensitivity, discrimination power, and processing efficiency (Figure 2). These technologies have expanded the capability of forensic laboratories to analyse highly degraded or low-template DNA samples while improving the speed and reliability of forensic genetic investigations [5,6].

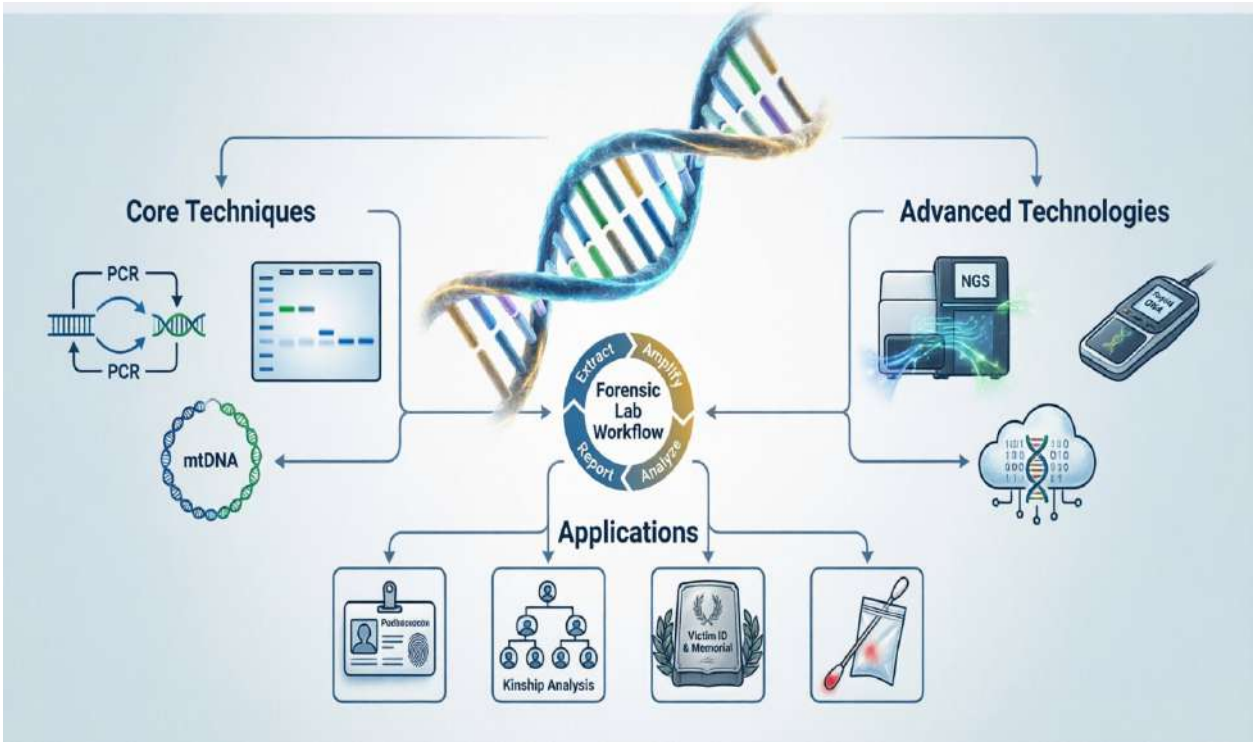


Figure 2. Overview of forensic DNA technologies illustrating the relationship between core genetic analysis techniques (PCR and STR profiling), advanced genomic technologies (next-generation sequencing and rapid DNA systems), and their major forensic applications in human identification, kinship analysis, disaster victim identification, and criminal investigations.

3.1.2 Advances in Forensic Toxicology

Forensic toxicology focuses on the detection, identification, and interpretation of drugs, poisons, alcohol, and toxic substances in biological specimens for medico-legal purposes. The discipline is theoretically grounded in pharmacokinetics and pharmacodynamics, which describe the absorption, distribution, metabolism, and excretion of chemical substances within the human body. Advances in analytical instrumentation, particularly chromatography, mass spectrometry, and immunoassay technologies, have substantially improved the sensitivity, specificity, and reliability of toxicological investigations [18,19].

Modern forensic toxicology also increasingly incorporates molecular and metabolomic approaches to improve detection capabilities and support interpretation of intoxication, poisoning, and drug-related deaths. These developments have strengthened the role of toxicological evidence in criminal investigations, forensic pathology, and public health surveillance (Figure 3).

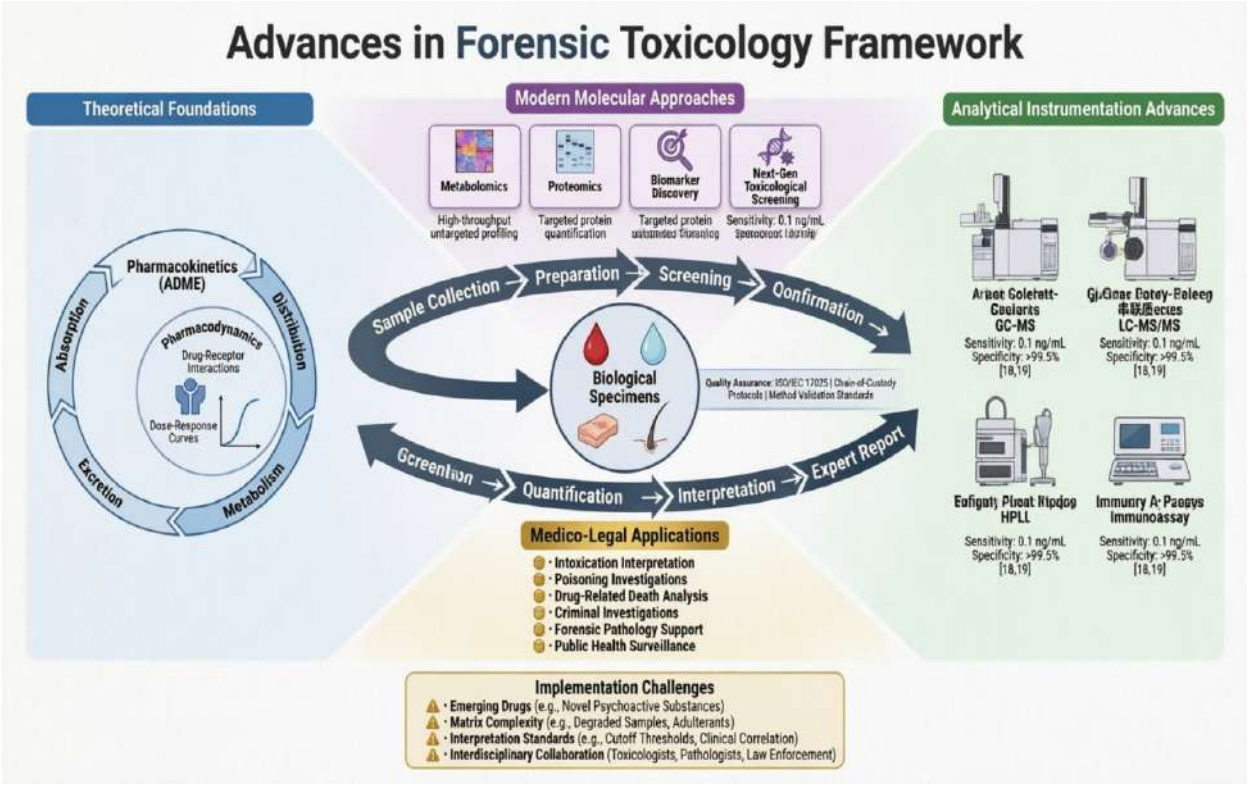


Figure 3. Conceptual framework of modern forensic toxicology showing analytical workflows, advanced instrumentation, molecular approaches, and major medico-legal applications in forensic investigations.

3.1.3 Forensic Entomology and Postmortem Investigations

Forensic entomology involves the application of insect biology and ecology to medico-legal investigations, particularly for estimating the post-mortem interval (PMI) and reconstructing environmental conditions associated with death. The discipline is theoretically based on insect developmental biology, ecological succession, and species-specific colonization patterns on decomposing remains. Necrophagous insects, especially species belonging to Diptera and Coleoptera, colonize remains in relatively predictable successional sequences that can provide valuable temporal information during forensic investigations [22,23].

Temperature-dependent developmental models and species-specific growth rates constitute essential theoretical components of forensic entomology. When integrated with environmental and climatic data, these models allow increasingly accurate estimation of PMI under diverse environmental conditions [33]. Despite its scientific robustness and growing international recognition, forensic entomology remains underutilized in many regions, including parts of the Arab world, primarily due to limited expertise, inadequate regional databases, and insufficient field validation studies adapted to local climatic conditions (Figure 4).

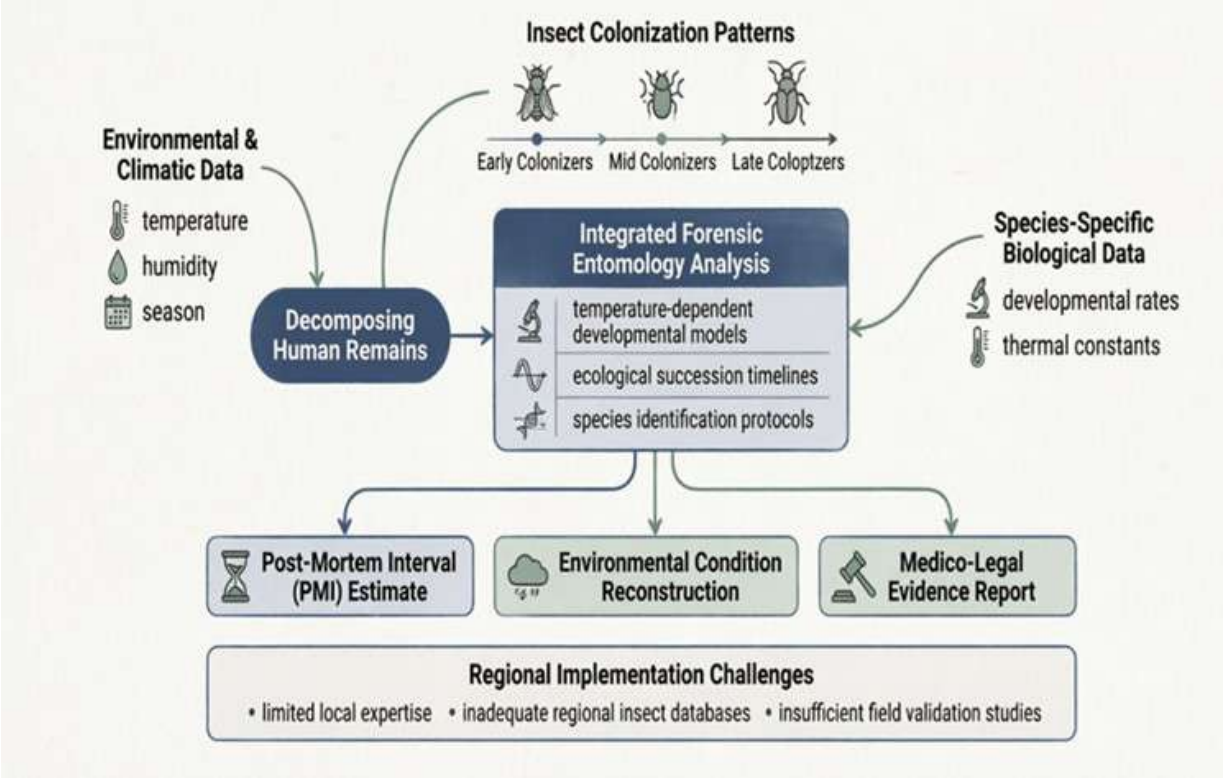


Figure 4. Conceptual framework illustrating the integration of environmental variables, insect succession patterns, and species-specific developmental data in forensic entomology for post-mortem interval (PMI) estimation and medico-legal investigations.

3.1.4 Forensic Anthropology

Forensic anthropology applies principles of human osteology and skeletal biology to the identification and analysis of human remains. Theoretical approaches within this discipline include estimation of age, sex, ancestry, stature, and trauma interpretation based on skeletal characteristics. Advances in imaging technologies, three-dimensional reconstruction, and statistical modelling have substantially improved the accuracy and reliability of skeletal identification and forensic reconstruction [16,21].

Forensic anthropology plays a particularly important role in mass disaster investigations, unidentified human remains cases, archaeological contexts, and humanitarian investigations involving conflict and migration. However, effective application of anthropological methodologies requires population-specific reference data and specialized expertise, which remain limited in several parts of the Arab world.

3.1.5 Digital and Cyber Forensics

The rapid expansion of digital technologies and cyber-dependent activities has positioned digital forensics as a critical discipline within modern forensic science. Digital and cyber forensics focus on the identification, acquisition, preservation, analysis, and interpretation of electronic data while ensuring evidential integrity and legal admissibility. The discipline encompasses computer forensics, mobile device analysis, network investigations, cloud forensics, and cybercrime investigation [7,9,15]. Advanced analytical techniques, including data recovery systems, network traffic analysis, malware analysis, and artificial intelligence-assisted digital investigations, have significantly enhanced the ability to investigate cybercrime and digital evidence. However, the rapidly evolving nature of cyber threats continues to create major technical and legal challenges for forensic practitioners worldwide (Figure 5).

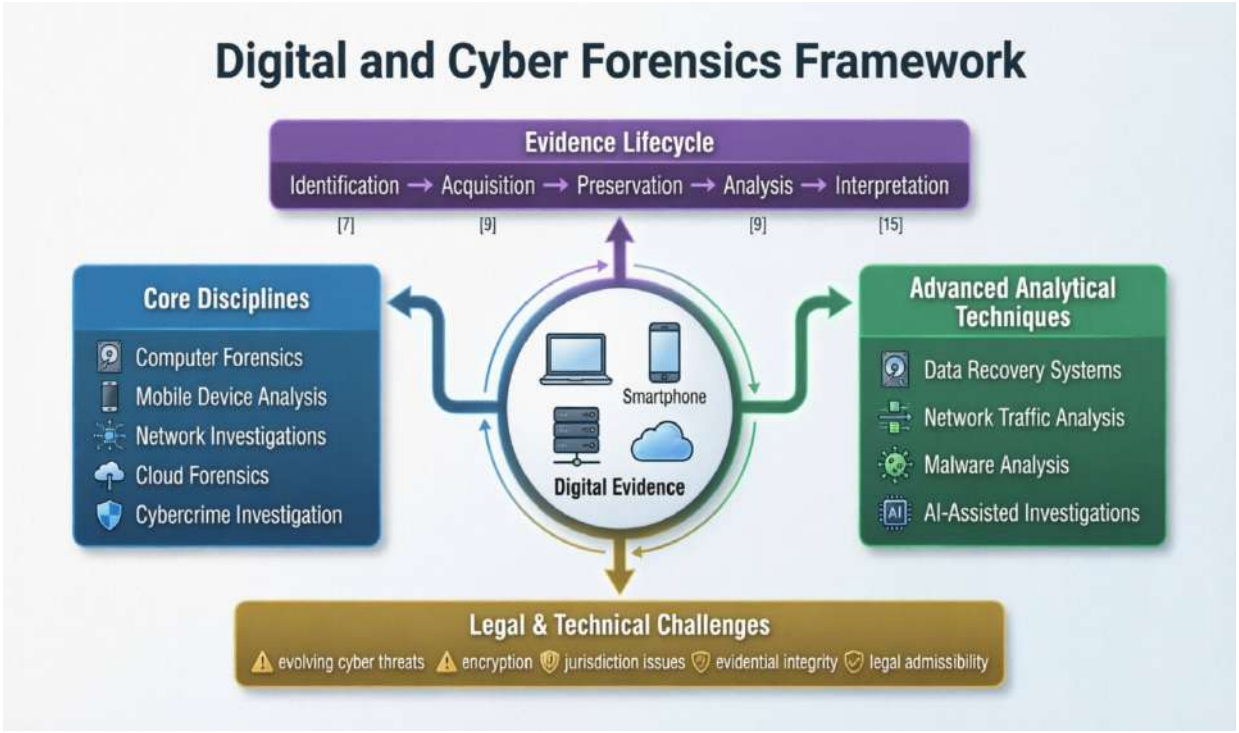


Figure 5. Framework of digital and cyber forensics illustrating the electronic evidence lifecycle, core investigative domains, advanced analytical techniques, and major legal and technical challenges.

3.2 Integration of Theoretical Advances into Practice

The advancement of forensic science is closely linked to the continuous development of scientific theories, analytical methodologies, and technological innovation. However, the practical effectiveness of forensic science depends on the successful translation of these theoretical advances into standardized operational procedures and evidence-based forensic practices. This integration requires not only advanced technological infrastructure but also qualified personnel, quality assurance systems, accreditation frameworks, and legal regulations supporting the admissibility and reliability of forensic evidence [10,11].

Recent developments increasingly emphasize the incorporation of artificial intelligence, machine learning, automated analytical systems, and data-driven approaches into forensic investigations. These technologies have improved pattern recognition, probabilistic interpretation, evidence classification, and complex data analysis, thereby enhancing forensic decision-making processes and analytical efficiency [7–9]. Such innovations demonstrate the dynamic nature of forensic science as a continuously evolving discipline responding to scientific progress, technological transformation, and emerging societal challenges.

3.3 Relevance to the Arab World Context

Although the theoretical foundations of forensic science are well established internationally, their practical implementation within the Arab world remains variable and uneven. Differences in educational systems, institutional infrastructure, research capacity, funding availability, and technological resources significantly influence the extent to which scientific advances are translated into operational forensic practice across the region [12–16].

Specialized disciplines such as forensic entomology, forensic anthropology, and digital forensics particularly require region-specific databases, localized validation studies, and advanced technical training, which are still under development in many Arab countries [15,21–23]. Consequently, understanding the scientific foundations and emerging trends of forensic science is essential for identifying gaps between theoretical advancement and practical implementation. Strengthening this connection represents a critical step toward improving the effectiveness, reliability, and sustainability of forensic science systems and enhancing their contribution to evidence-based justice throughout the Arab world.

4. Forensic Science Practice in the Arab World

The practical application of forensic science in the Arab world has undergone substantial transformation during the past two decades, reflecting increasing governmental recognition of the importance of forensic evidence in criminal investigations and judicial decision-making. Several Arab countries have established specialized forensic laboratories, medico-legal institutes, and forensic units within law enforcement agencies to improve investigative efficiency and strengthen evidence-based justice systems. Nevertheless, the degree of technological advancement, operational effectiveness, and institutional integration varies considerably across the region. While some countries have achieved notable progress in adopting modern forensic technologies and accreditation systems, others continue to face significant challenges related to infrastructure, workforce development, and quality assurance. Concerns regarding scientific reliability, procedural standardization, and consistency in forensic practice therefore remain important issues within regional forensic systems [10,11].

4.1 Institutional and Laboratory Infrastructure

Several Arab countries, particularly those within the Gulf Cooperation Council (GCC), have invested substantially in the development of modern forensic infrastructure. Advanced forensic laboratories equipped with contemporary analytical technologies have been established in countries such as Saudi Arabia, the United Arab Emirates, and Qatar. These facilities commonly include specialized units for DNA profiling, forensic toxicology, digital forensics, and trace evidence analysis, reflecting increasing alignment with international forensic standards and accreditation frameworks [13,20]. In contrast, many countries in North Africa and other parts of the Arab region continue to experience limitations related to laboratory capacity, outdated instrumentation, insufficient maintenance systems, and restricted financial resources. In Egypt, for example, forensic services are largely centralized within governmental institutions, resulting in high case burdens and delays in evidence processing. Despite these challenges, gradual improvements have emerged through modernization initiatives, infrastructure development programs, and increasing academic participation in forensic research and training activities [17,18].

4.2 Human Capacity and Training

The effectiveness and reliability of forensic science systems depend heavily on the availability of qualified and well-trained professionals. Although several Arab universities and training institutions have introduced forensic science programs and specialized postgraduate degrees, shortages of highly trained personnel remain evident across numerous forensic disciplines. This shortage is particularly pronounced in emerging and highly specialized fields such as digital forensics, forensic genomics, forensic anthropology, and forensic entomology [13,14]. Regional and international organizations

have contributed to workforce development through workshops, certification programs, and collaborative training initiatives designed to enhance technical competency and professional standards. However, disparities in educational quality, limited access to continuous professional development, and insufficient practical laboratory training continue to hinder the establishment of uniformly qualified forensic personnel throughout the region. Strengthening competency-based education and improving practical training opportunities remain essential priorities for sustainable forensic capacity building [14,31].

4.3 Legal and Regulatory Frameworks

Legal and regulatory systems governing forensic science differ considerably among Arab countries, influencing both the quality and admissibility of forensic evidence in judicial proceedings. Some countries have implemented relatively advanced frameworks regulating evidence collection, laboratory accreditation, chain-of-custody procedures, and expert testimony. In other jurisdictions, however, forensic legislation remains fragmented or insufficiently standardized, creating inconsistencies in forensic practice and reducing confidence in scientific evidence [27,29]. Efforts to harmonize forensic procedures with internationally recognized standards, including ISO/IEC 17025 accreditation and quality assurance protocols, have expanded in recent years, particularly within GCC countries. Nevertheless, implementation remains inconsistent across the region, and many laboratories continue to operate without comprehensive accreditation or standardized validation systems [29,31]. Strengthening legal frameworks, promoting laboratory accreditation, and enhancing judicial understanding of forensic methodologies are therefore essential for improving the scientific credibility and legal reliability of forensic evidence throughout the Arab world.

4.4 Application of Forensic Disciplines in Practice

The practical application of forensic disciplines within the Arab world varies according to national priorities, economic resources, and institutional capabilities. DNA analysis and forensic toxicology are among the most widely implemented forensic disciplines, particularly in countries with advanced laboratory infrastructure and governmental support. These disciplines are routinely applied in criminal investigations, human identification, disaster victim identification, and toxicological examinations associated with drug-related offenses and suspicious deaths [17–19].

In contrast, other disciplines such as forensic anthropology and forensic entomology remain comparatively underdeveloped and underutilized in many Arab countries. The limited operational integration of forensic entomology is primarily attributed to insufficient regional entomological

databases, scarcity of field validation studies, and lack of specialized expertise [22,23]. This limitation is particularly significant given the climatic diversity of the Arab region, which strongly influences decomposition patterns and insect succession dynamics. Similarly, forensic anthropology faces challenges related to the absence of comprehensive skeletal reference collections and population-specific databases [21]. Digital forensics, however, has experienced rapid growth in response to increasing cybercrime activities and the expanding role of electronic evidence in criminal investigations. Several Arab countries have established specialized cybercrime and digital forensic units, although the rapid evolution of digital technologies continues to create ongoing challenges in training, infrastructure development, and technical adaptation [15].

4.5 Research–Practice Integration

A major challenge affecting forensic science practice in the Arab world is the limited integration between academic research and operational forensic institutions. Although universities and research centres increasingly contribute to forensic science research, the translation of scientific findings into routine forensic procedures remains relatively weak [11,31]. This disconnect is often associated with institutional fragmentation, limited communication between researchers and forensic practitioners, and insufficient funding for applied forensic research. Consequently, innovative analytical methods and region-specific scientific findings are not consistently incorporated into operational forensic workflows.

Strengthening collaboration among universities, forensic laboratories, judicial authorities, and law enforcement agencies is therefore essential for improving the practical application of forensic science. Joint research programs, collaborative training initiatives, and shared scientific databases could significantly enhance evidence-based forensic practice while promoting the development of standardized operational procedures adapted to regional conditions [27,30]. Enhancing research–practice integration would also improve the capacity of forensic institutions to respond effectively to emerging technological and investigative challenges.

4.6 Regional Disparities and Development Gaps

The Arab world is characterized by substantial disparities in forensic science development, technological integration, and institutional capacity. While some countries have achieved advanced levels of forensic modernization and accreditation, others remain in earlier stages of forensic system development due to economic limitations, political instability, and insufficient educational

infrastructure [12,31]. These disparities influence the quality, accessibility, and reliability of forensic services across the region.

Addressing these development gaps requires coordinated regional strategies focused on infrastructure investment, workforce development, scientific standardization, and knowledge exchange. Regional cooperation initiatives involving collaborative research, professional networking, shared databases, and harmonized accreditation systems could play an important role in reducing disparities and promoting sustainable forensic development throughout the Arab world [27,30]. Strengthening intergovernmental collaboration and institutional partnerships would also facilitate the dissemination of best practices and support the establishment of more consistent and scientifically reliable forensic systems across the region.

5. Gap Between Theory and Practice in Forensic Science

Despite the considerable advancement of forensic science as a multidisciplinary and technology-driven field, a substantial gap continues to exist between theoretical development and practical implementation, particularly within the Arab world. Although modern forensic research has generated highly sophisticated analytical techniques and scientifically validated methodologies, their effective application in operational forensic systems remains inconsistent across many Arab countries. This disparity limits the efficiency of forensic investigations, weakens the reliability of scientific evidence in judicial proceedings, and reduces the overall contribution of forensic science to evidence-based justice systems [10,11].

5.1 Nature of the Theory–Practice Gap

The theory–practice gap in forensic science refers to the disconnect between scientifically established forensic methodologies and their actual implementation in routine forensic casework. Theoretical forensic frameworks emphasize scientific validity, reproducibility, standardization, and evidence-based interpretation. In practice, however, the application of these principles is often constrained by institutional, economic, technical, and administrative limitations [15]. Consequently, many advanced forensic methods remain inconsistently applied or unavailable in operational settings. Within the Arab world, this gap is reflected in unequal technological adoption, inconsistent laboratory standards, and limited incorporation of research findings into forensic procedures and investigative workflows [12,31].

5.2 Technological and Infrastructure Constraints

One of the principal factors contributing to the theory–practice gap is the unequal availability of advanced forensic technologies and infrastructure. Although sophisticated forensic methodologies such as next-generation DNA sequencing, automated fingerprint identification systems, digital forensic software, and AI-assisted analytical tools are increasingly recognized as essential components of modern forensic science, their implementation requires substantial financial investment, technical maintenance, and specialized expertise [5–7]. Many forensic laboratories in developing Arab countries continue to face limitations related to outdated equipment, inadequate maintenance systems, insufficient laboratory space, and restricted technological upgrading. As a result, operational forensic practice often relies heavily on conventional analytical methods that may be less efficient, less sensitive, or less standardized than contemporary technologies [20,27].

5.3 Human Resource and Expertise Limitations

The shortage of highly specialized forensic professionals represents another major challenge limiting the translation of scientific theory into operational practice. Modern forensic science increasingly depends on interdisciplinary expertise involving molecular biology, bioinformatics, toxicology, digital technologies, entomology, and statistical interpretation. However, many Arab countries continue to experience insufficient availability of adequately trained specialists in disciplines such as forensic anthropology, forensic entomology, digital forensics, and forensic genomics [13,16]. Limited opportunities for advanced professional training, continuing education, and international scientific exchange further aggravate this problem. Consequently, even when modern forensic technologies are available, their effective application may remain restricted due to insufficient technical competency and lack of specialized expertise [14,31].

5.4 Weak Integration Between Research and Practice

A significant dimension of the theory–practice gap is the limited integration between academic forensic research and operational forensic institutions. Universities and research centers throughout the Arab world have increasingly contributed to forensic science research and scientific publication. Nevertheless, the practical translation of these scientific findings into standardized forensic procedures and routine investigative protocols remains relatively weak [11]. Institutional fragmentation, insufficient collaboration between researchers and forensic practitioners, and limited funding for applied forensic research frequently hinder the implementation of innovative methodologies within operational laboratories. Consequently, valuable scientific advances developed in academic settings are often not incorporated into routine forensic workflows or judicial applications. Strengthening

institutional collaboration and establishing formal partnerships between academic institutions and forensic laboratories are therefore essential for improving evidence-based forensic practice in the region [27,30].

5.5 Legal and Administrative Barriers

The effectiveness of forensic science is closely associated with the legal and administrative frameworks governing the collection, interpretation, and admissibility of scientific evidence. In several Arab countries, forensic legislation remains fragmented, inconsistently applied, or insufficiently developed, thereby limiting the implementation of scientifically advanced forensic methodologies [24,29]. The absence of standardized operational procedures, accreditation systems, and clear guidelines for expert testimony may reduce confidence in forensic evidence during judicial proceedings. Administrative barriers, including bureaucratic complexity, limited inter-agency coordination, and centralized decision-making structures, may further delay forensic investigations and reduce operational efficiency. In some jurisdictions, insufficient judicial understanding of scientific methodologies also affects the interpretation and utilization of forensic evidence in court [27,32]. Strengthening forensic legislation and improving judicial scientific literacy are therefore essential components of bridging the theory–practice gap.

5.6 Discipline-Specific Gaps: The Case of Forensic Entomology

The disparity between theoretical advancement and practical implementation is particularly evident in specialized forensic disciplines such as forensic entomology. Scientifically, forensic entomology is supported by well-established ecological and developmental models that enable estimation of post-mortem intervals through analysis of insect succession and developmental stages [22,33]. However, its operational application in the Arab world remains limited due to insufficient regional databases, scarcity of field validation studies, and limited integration into medico-legal investigations [23].

This challenge is especially important because climatic and environmental conditions in the Arab region differ substantially from those of temperate regions where most forensic entomological models were originally developed. Variations in temperature, humidity, seasonal patterns, and insect biodiversity significantly influence decomposition dynamics and insect colonization patterns [33]. Consequently, the absence of localized entomological datasets and region-specific developmental studies restricts the scientific reliability and judicial acceptance of entomological evidence in many Arab countries despite the strong theoretical foundation of the discipline.

5.7 Impact on Forensic Effectiveness

The persistence of the theory–practice gap has significant implications for the overall effectiveness and credibility of forensic science systems. Limitations in technological implementation, workforce expertise, and procedural standardization may result in delays in evidence processing, reduced analytical accuracy, and inconsistent interpretation of forensic findings. These deficiencies can ultimately affect the quality of criminal investigations and judicial outcomes by compromising the reliability, completeness, and scientific validity of forensic evidence [10,32]. Furthermore, insufficient integration of scientific innovation into operational forensic systems may reduce public confidence in forensic institutions and weaken the broader role of forensic science in supporting justice and public security.

5.8 Toward Bridging the Gap

Bridging the gap between theory and practice requires a comprehensive and multidisciplinary strategy involving scientific, institutional, educational, and legal reforms. Investment in modern forensic infrastructure, laboratory accreditation systems, and technological modernization is essential for improving operational capacity. Simultaneously, strengthening forensic education, postgraduate specialization, and continuous professional development programs is necessary to enhance workforce competency and sustainability [13,31].

Improved collaboration between academic researchers, forensic practitioners, judicial authorities, and law enforcement agencies is also critical for translating scientific innovation into operational practice. Joint research initiatives, applied forensic studies, and collaborative training programs could significantly enhance evidence-based forensic methodologies adapted to regional conditions [27,30]. Furthermore, harmonization of legal frameworks, adoption of standardized protocols, and expansion of regional scientific cooperation would contribute substantially to improving the consistency, reliability, and effectiveness of forensic science systems across the Arab world.

6. Challenges Specific to the Arab World

The advancement and effective implementation of forensic science in the Arab world are influenced by a complex combination of institutional, technical, legal, educational, and socio-economic challenges. Although several Arab countries have achieved notable progress in forensic infrastructure, laboratory modernization, and scientific capabilities, substantial disparities persist in governance systems, accreditation standards, research integration, workforce development, and technological

sustainability. These interconnected challenges collectively affect the efficiency, reliability, and long-term sustainability of forensic science systems throughout the region and limit their contribution to evidence-based criminal justice systems [27–31].

6.1 Fragmented Governance and Institutional Coordination

One of the most significant obstacles facing forensic science systems in the Arab world is the fragmentation of governance among multiple governmental sectors, including ministries of justice, interior, health, defence, and higher education. This institutional fragmentation frequently leads to overlapping responsibilities, inconsistent administrative procedures, and weak coordination among forensic stakeholders [27,31]. In many Arab countries, forensic laboratories, medico-legal institutes, and law enforcement forensic units operate under separate administrative authorities without integrated national governance frameworks, thereby reducing operational efficiency and limiting long-term strategic planning.

The absence of centralized or semi-independent national forensic science authorities further complicates efforts to harmonize forensic methodologies, accreditation systems, and professional standards. International forensic governance models emphasize the importance of integrated national systems capable of coordinating forensic laboratories, research institutions, training programs, and judicial implementation under unified regulatory structures [28,32]. Establishing independent national forensic commissions or centralized forensic authorities could therefore improve institutional accountability, procedural standardization, inter-agency communication, and scientific oversight across the region.

6.2 Inconsistent Accreditation and Quality Assurance Systems

Accreditation and quality assurance systems remain unevenly implemented across forensic laboratories in the Arab world. While several countries, particularly within the Gulf Cooperation Council (GCC), have adopted international accreditation standards such as ISO/IEC 17025, many forensic laboratories in developing Arab countries continue to function without comprehensive accreditation or fully standardized quality control procedures [29,34].

The absence of harmonized accreditation systems directly affects the reliability, reproducibility, and judicial admissibility of forensic evidence. International forensic standards emphasize that scientifically credible forensic systems require rigorous quality assurance programs incorporating validated analytical procedures, proficiency testing, standardized operating protocols, internal audits,

and regular external assessments [28,32]. Inadequate implementation of these systems may increase the risk of analytical errors, inconsistent evidence interpretation, and reduced judicial confidence in forensic findings. Consequently, policymakers throughout the Arab region should prioritize the establishment of national accreditation bodies and regional oversight mechanisms to ensure compliance with internationally recognized forensic quality standards.

6.3 Limited Investment in Research and Development (R&D)

Research and development in forensic science remain relatively underfunded across many Arab countries. Existing governmental investment is often directed primarily toward operational infrastructure, equipment acquisition, and case management systems rather than long-term scientific innovation and applied forensic research [30,34]. This imbalance restricts the development of region-specific forensic methodologies and reduces the contribution of Arab forensic institutions to global scientific advancement.

The shortage of dedicated research funding is particularly problematic in disciplines requiring environmental, climatic, and geographical adaptation, including forensic entomology, environmental forensics, decomposition ecology, and arid-climate forensic investigations. Several studies have emphasized the importance of localized validation studies and region-specific datasets to improve the scientific reliability of forensic analyses under Middle Eastern and North African environmental conditions [23,33]. Increased governmental funding, international scientific partnerships, and public–private collaborative initiatives are therefore essential to strengthen forensic research capacity and promote sustainable scientific innovation throughout the region.

6.4 Human Capital Constraints and Brain Drain

The shortage of qualified forensic professionals represents another major challenge affecting the sustainability of forensic science systems in the Arab world. Many countries continue to experience limited availability of specialized experts in disciplines such as forensic pathology, forensic anthropology, digital forensics, forensic genomics, and forensic entomology [13,31]. This shortage is further intensified by the migration of highly trained professionals to countries offering superior research opportunities, professional advancement, and financial incentives. The resulting “brain drain” weakens institutional continuity, reduces local scientific expertise, and limits the long-term development of regional forensic capabilities. Addressing this challenge requires the establishment of competitive career pathways, improved professional incentives, sustainable research opportunities, and advanced professional development programs. Integrating practical forensic training into

university curricula, expanding postgraduate specialization programs, and strengthening international academic collaboration are also essential for building sustainable forensic human capacity throughout the Arab world [14,19].

6.5 Gaps in Education and Specialized Training

Although forensic science education has expanded considerably within several Arab universities, many academic programs remain limited in scope and insufficiently aligned with international forensic education standards. In numerous institutions, practical laboratory training, interdisciplinary integration, simulated casework experience, and exposure to operational forensic environments remain inadequate, thereby reducing graduates' preparedness for professional forensic practice [13,31].

Furthermore, emerging and highly specialized disciplines such as digital forensics, forensic DNA interpretation, forensic entomology, and forensic data analytics often receive limited curricular emphasis despite their increasing importance in contemporary criminal investigations. International forensic education frameworks increasingly emphasize competency-based learning, accreditation of educational programs, interdisciplinary integration, and close collaboration between universities and forensic laboratories [25]. Establishing regional centres of excellence, standardized forensic curricula, and specialized professional certification systems could substantially improve training quality, technical competency, and workforce readiness across the Arab world.

6.6 Legal and Judicial Constraints

The effectiveness of forensic science is strongly influenced by the legal and judicial systems governing the admissibility, interpretation, and use of scientific evidence. In several Arab countries, forensic legislation remains fragmented, outdated, or inconsistently implemented, thereby creating challenges related to evidence admissibility, expert testimony, chain-of-custody procedures, and procedural standardization [18,24].

Limited judicial understanding of scientific methodologies may further reduce the effective utilization of forensic evidence during legal proceedings. International forensic policy reports have repeatedly emphasized the importance of strengthening judicial scientific literacy and establishing clear legal frameworks regulating forensic methodologies and expert witness practices [27,32]. Consequently, enhancing forensic legislation, improving procedural standardization, and providing specialized scientific training for judges, prosecutors, and legal professionals are essential for improving the reliability and judicial interpretation of forensic evidence throughout the region.

6.7 Data Scarcity and Lack of Regional Databases

Reliable forensic analysis depends heavily on the availability of region-specific reference databases, including genetic profiles, toxicological records, entomological datasets, skeletal reference collections, and environmental baseline information. In many Arab countries, such databases remain incomplete, fragmented, outdated, or entirely absent [21,33].

This challenge is particularly critical in forensic entomology, where accurate estimation of post-mortem intervals requires localized developmental data for necrophagous insect species under regional climatic conditions. Studies conducted within Middle Eastern and North African environments have demonstrated substantial variation in insect succession patterns and decomposition dynamics compared with temperate ecosystems where many forensic models were originally developed [22,23]. Similarly, population-specific genetic databases are essential for improving the accuracy and statistical reliability of DNA profiling and human identification analyses. Establishing coordinated national and regional forensic databases should therefore be considered a strategic priority for improving analytical precision, standardization, and evidence reliability throughout the Arab world.

6.8 Technological Dependence and Sustainability Issues

Many forensic laboratories in the Arab world depend heavily on imported technologies, foreign technical expertise, and externally developed analytical systems. Although this dependence has facilitated relatively rapid adoption of advanced forensic tools, it has also generated challenges related to technological sustainability, maintenance, software licensing, technical support, and operational continuity [20,25].

Long-term reliance on imported technologies may restrict local scientific innovation and increase operational costs, particularly when specialized maintenance and technical expertise are unavailable domestically. Strengthening regional technological capacity through technology transfer initiatives, local technical training, specialized engineering support, and collaborative research programs could improve sustainability and reduce dependence on external suppliers. Supporting indigenous research and development programs is also essential for producing forensic technologies adapted specifically to regional environmental, linguistic, and operational conditions.

6.9 Limited Regional Collaboration and Knowledge Exchange

Despite shared geographical, cultural, and socio-political characteristics, regional cooperation in forensic science among Arab countries remains relatively limited. Collaborative research projects,

harmonized forensic policies, shared scientific databases, regional accreditation initiatives, and professional forensic societies are still insufficiently developed [21,26].

Strengthening regional scientific networks and collaborative forensic platforms could facilitate knowledge exchange, workforce development, methodological standardization, and joint training programs. International organizations such as INTERPOL, UNODC, and the International Association of Forensic Sciences have repeatedly emphasized the importance of regional collaboration in strengthening forensic governance and supporting cross-border criminal investigations [27,30,34]. Enhanced regional cooperation would therefore contribute substantially to scientific harmonization, institutional development, and sustainable advancement of forensic science systems across the Arab world.

6.10 Policy Implications

Collectively, these challenges demonstrate the urgent need for comprehensive forensic science policies integrating scientific, institutional, legal, educational, and technological dimensions. Sustainable reform requires coordinated governmental strategies focused on governance restructuring, accreditation systems, workforce development, research investment, legal modernization, and technological sustainability.

Long-term progress will depend on the ability of Arab countries to establish integrated forensic systems supported by evidence-based policy, scientific innovation, and regional collaboration. Strategic investment in forensic infrastructure, human capital, applied research, and quality assurance systems is essential for improving the effectiveness, reliability, and international competitiveness of forensic science throughout the Arab world.

7. Future Directions and Strategic Recommendations

Advancing forensic science in the Arab world requires a comprehensive and forward-looking strategy that integrates scientific innovation, institutional reform, legal modernization, and regional collaboration. Although several Arab countries have made considerable progress in forensic infrastructure and technological adoption, sustainable advancement depends on bridging the persistent gap between theoretical development and practical implementation. Future strategies should therefore focus on strengthening governance systems, enhancing research capacity, developing human resources, and promoting evidence-based forensic practice aligned with international standards [27–30].

7.1 Establishment of National Forensic Science Authorities

One of the most important strategic priorities is the establishment of centralized or semi-autonomous national forensic science authorities responsible for governance, regulation, accreditation, and coordination of forensic activities. In many Arab countries, forensic services remain fragmented across multiple ministries and agencies, leading to inconsistent operational standards and reduced institutional efficiency [27,31].

International forensic governance models emphasize the importance of unified national systems capable of integrating forensic laboratories, academic institutions, judicial authorities, and law enforcement agencies under coordinated regulatory frameworks [28,32]. The establishment of independent forensic science commissions or national oversight bodies would improve accountability, facilitate policy harmonization, and strengthen the integration of research and operational forensic practice throughout the region.

7.2 Standardization and Accreditation of Forensic Practices

The harmonization of forensic methodologies and quality assurance systems is essential for improving the reliability and admissibility of forensic evidence. Future reform efforts should prioritize the adoption of internationally recognized accreditation standards, particularly ISO/IEC 17025, across forensic laboratories in the Arab world [29,33].

Standardized operating procedures, proficiency testing, regular audits, and inter-laboratory validation programs are fundamental components of scientifically reliable forensic systems [31,33]. Regional standardization initiatives could also facilitate cross-border cooperation, evidence sharing, and mutual recognition of forensic analyses among Arab countries. Establishing regional accreditation centres and unified forensic guidelines would significantly improve consistency and quality assurance across the region.

7.3 Investment in Research and Scientific Innovation

Sustainable development of forensic science requires long-term investment in research and innovation. Policymakers should support interdisciplinary and applied forensic research addressing region-specific challenges, including decomposition processes under arid climatic conditions, environmental forensics, forensic entomology, digital evidence analysis, and emerging cybercrime threats [30,34].

Developing locally validated forensic methodologies is particularly important because many current analytical models are based primarily on datasets generated in temperate Western environments. Several studies have demonstrated that climatic variation significantly influences decomposition dynamics, insect succession, and forensic interpretation [19,24]. Increased funding for national research programs, public–private partnerships, and international scientific collaboration would strengthen the region's contribution to global forensic science advancement.

7.4 Capacity Building and Workforce Development

Human resource development remains central to improving forensic science systems in the Arab world. Future strategies should focus on expanding specialized forensic science programs, strengthening postgraduate training, and integrating practical laboratory experience into academic curricula [14,22].

Continuous professional development programs, certification systems, and specialized training workshops are also essential for maintaining competency in rapidly evolving forensic disciplines such as forensic genomics, digital forensics, and artificial intelligence-assisted analysis. Furthermore, improving career pathways, research opportunities, and professional incentives may help reduce the migration of skilled forensic experts and mitigate the effects of regional brain drain [20,31].

7.5 Strengthening Research–Practice Integration

A major strategic objective should be the enhancement of collaboration between academic institutions, forensic laboratories, and judicial agencies. Although universities in the Arab world are increasingly contributing to forensic research, the translation of scientific findings into operational forensic procedures remains limited [15,29].

Future policies should encourage joint research initiatives, collaborative case-based training, and knowledge transfer programs linking researchers with forensic practitioners. Establishing formal partnerships between universities and operational forensic institutions would facilitate the development of evidence-based forensic protocols and improve the practical application of scientific innovation [28,32].

7.6 Legal and Judicial Reform

The effectiveness of forensic science depends heavily on the legal framework governing the collection, interpretation, and admissibility of scientific evidence. Several Arab countries still lack comprehensive forensic legislation and standardized judicial guidelines related to expert testimony and forensic procedures [18,27].

Future reforms should focus on establishing clear legal standards for forensic evidence, strengthening accreditation requirements for forensic experts, and improving judicial understanding of scientific methodologies. Specialized training programs for judges, prosecutors, and legal professionals are essential to enhance the interpretation and appropriate use of forensic evidence in court proceedings [24,32]. In some jurisdictions, the establishment of specialized forensic or cybercrime judicial units may further improve the integration of scientific evidence into legal decision-making processes.

7.7 Development of National and Regional Forensic Databases

The establishment of comprehensive forensic databases should be considered a strategic priority across the Arab world. Reliable forensic analysis often depends on access to region-specific genetic, toxicological, entomological, and environmental datasets [18,33].

This need is particularly important in forensic entomology, where accurate post-mortem interval estimation requires localized developmental data for necrophagous insect species under regional environmental conditions. Similarly, population-specific genetic databases improve the reliability of DNA profiling and human identification analyses [19,32]. Collaborative regional database initiatives could significantly improve analytical accuracy, facilitate cross-border investigations, and support scientific standardization throughout the region.

7.8 Integration of Emerging Technologies and Artificial Intelligence

The integration of emerging technologies represents one of the most promising future directions in forensic science. Artificial intelligence (AI), machine learning, digital imaging systems, automated DNA analysis, and big data analytics have the potential to improve analytical speed, evidence interpretation, and decision-making accuracy [7–9].

Digital forensics and cybercrime investigations are expected to become increasingly important as criminal activities continue to evolve in technologically advanced environments. However, the implementation of AI-assisted forensic systems must be accompanied by appropriate ethical guidelines, transparency standards, and validation procedures to ensure scientific reliability and

judicial acceptance [33]. Strategic investment in technological infrastructure and specialized technical training will therefore be essential for maintaining competitiveness and operational effectiveness.

7.9 Enhancing Regional and International Collaboration

Strengthening collaboration among Arab countries is critical for addressing shared forensic challenges and promoting scientific harmonization. Regional cooperation can facilitate data sharing, joint research projects, training programs, and standardization of forensic methodologies [21,26].

International organizations such as INTERPOL, UNODC, and the International Association of Forensic Sciences have emphasized the importance of collaborative forensic networks for improving criminal investigations and strengthening institutional capacity [27,34]. The establishment of regional forensic societies, collaborative research consortia, and cross-border training initiatives would significantly enhance knowledge exchange and scientific development throughout the Arab world.

7.10 Strategic Vision for the Future

The future advancement of forensic science in the Arab world requires an integrated long-term vision combining scientific innovation, institutional reform, human capacity development, and legal modernization. Policymakers should adopt sustainable strategies that support research-driven forensic systems capable of adapting to emerging technological and societal challenges.

Bridging the gap between theoretical advancement and practical implementation will require coordinated regional efforts, investment in infrastructure and education, and stronger integration between scientific research and operational forensic practice. By aligning forensic systems with international scientific standards while addressing region-specific needs, the Arab world can develop more reliable, efficient, and globally competitive forensic science frameworks that strengthen justice, security, and public trust.

8. Conclusion

Forensic science in the Arab world stands at a critical stage of development, characterized by significant theoretical and technological advancements alongside persistent challenges in practical implementation. This review demonstrated that, despite increasing institutional recognition and gradual modernization of forensic infrastructure, important gaps remain in governance, accreditation systems, research integration, human resource development, and legal standardization. These

limitations continue to affect the reliability, consistency, and operational effectiveness of forensic science systems across the region. Addressing these challenges requires a comprehensive and coordinated strategy that integrates scientific innovation with institutional and legal reform. Strengthening quality assurance systems, expanding specialized education and professional training, enhancing collaboration between academic institutions and forensic laboratories, and investing in region-specific forensic research are essential for sustainable progress. In particular, the development of localized databases and applied studies in disciplines such as forensic entomology, digital forensics, and forensic toxicology remains crucial for improving analytical accuracy and contextual relevance within diverse Arab environments. Ultimately, the future advancement of forensic science in the Arab world depends on the successful integration of theoretical knowledge into evidence-based forensic practice. Through sustained regional cooperation, strategic investment, and alignment with international scientific standards, Arab countries can establish more reliable, efficient, and globally competitive forensic science systems capable of strengthening justice, security, and public trust.

References

[1] National Research Council. *Strengthening Forensic Science in the United States: A Path Forward*. Washington, DC: National Academies Press; 2009.

[2] Houck MM, Siegel JA. *Fundamentals of Forensic Science*. 3rd ed. Amsterdam: Academic Press; 2015.

[3] Saferstein R. *Criminalistics: An Introduction to Forensic Science*. 13th ed. New York: Pearson; 2020.

[4] Butler JM. *Advanced Topics in Forensic DNA Typing*. Amsterdam: Academic Press; 2015.

[5] Kayser M, de Knijff P. Improving human forensics through advances in genetics. *Nat Rev Genet*. 2011; 12:179–192.

[6] van Oorschot RAH, et al. DNA transfer in forensic science. *Forensic Sci Int Genet*. 2021; 52:102480.

[7] Ridgeway G. Artificial intelligence in forensic science. *Annu Rev Criminol*. 2020;3:445–464.

[8] Tangen JM, et al. Forensic science in the 21st century. *Nat Hum Behav*. 2020; 4:111–118.

[9] Mirsky Y, et al. The threat of AI in forensic evidence. *IEEE Secur Priv*. 2021;19(2):52–61.

[10] Roux C, Crispino F, Ribaux O. The evolution of forensic science. *Curr Issues Crim Justice*. 2012;24(1):7–24.

[11] Mnookin JL, et al. The need for a research culture in forensic science. *UCLA Law Rev*. 2011; 58:725–779.

[12] Al-Amad SH, McCullough M, Roffey P. Forensic science in the Middle East. *Forensic Sci Int*. 2019; 295:45–52.

[13] Alenezi MN, Almutairi BO. Forensic science education in GCC countries. *Egypt J Forensic Sci*. 2021; 11:34.

[14] Almazrouei S, et al. Forensic services in UAE. *Forensic Sci Int Rep*. 2021; 4:100228.

[15] Natsheh E. Analyzing emerging cybercrime patterns and enhancing digital-forensics capabilities in the MENA region. *Int J Technol Knowl Soc*. 2025;22(1):1.

[16] Mansour C, Márquez-Grant N, Benito Sánchez M. Current status of forensic anthropology in the Middle East and North Africa. *Egypt J Forensic Sci*. 2025;15(1):4.

[17] El-Sayed N, et al. Forensic medicine in Egypt. *Egypt J Forensic Sci*. 2020; 10:1–8.

[18] El-Ashmawy I, et al. Forensic toxicology in Egypt. *Forensic Toxicol*. 2021; 39:123–135.

[19] Alharbi M, et al. Forensic toxicology in Saudi Arabia. *Saudi J Biol Sci*. 2022; 29:103456.

[20] Al-Shammari F, et al. Forensic laboratories in the Gulf region: Current status and future perspectives. *J Forensic Leg Med*. 2020; 74:102034.

[21] Blau S. Human identification and forensic anthropology in the Middle East. *Forensic Sci Med Pathol*. 2016; 12:463–471.

[22] Al-Mesbah H, et al. Forensic entomology in Arabia: Current status and challenges. *J Med Entomol*. 2012; 49:1–9.

[23] El-Azazy OME. Current status and future perspectives of forensic entomology in Egypt. *Egypt Acad J Biol Sci*. 2018; 10:1–10.

[24] Page MJ, et al. The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*. 2021;372: n71.

[25] Moher D, et al. Preferred reporting items for systematic reviews and meta-analyses (PRISMA). *PLoS Med*. 2009;6(7): e1000097.

[26] Higgins JPT, Thomas J, Chandler J, et al. *Cochrane Handbook for Systematic Reviews of Interventions*. Hoboken: Wiley; 2022.

[27] United Nations Office on Drugs and Crime (UNODC). *Strengthening Forensic Science Systems and Crime Scene Investigation*. Vienna: UNODC; 2022.

[28] National Institute of Justice (NIJ). *Strengthening the Foundations of Forensic Science*. Washington, DC: U.S. Department of Justice; 2023.

[29] ISO/IEC. *ISO/IEC 17025: General Requirements for the Competence of Testing and Calibration Laboratories*. Geneva: International Organization for Standardization; 2017.

[30] INTERPOL. *Forensic Science Services in the MENA Region*. Lyon: INTERPOL Forensic Science Programme; 2022.

[31] Al-Amad SH, Alenezi MN. Challenges and opportunities in forensic science systems across Arab countries. *Egypt J Forensic Sci*. 2023; 13:41.

[32] PCAST (President's Council of Advisors on Science and Technology). *Forensic Science in Criminal Courts: Ensuring Scientific Validity of Feature-Comparison Methods*. Washington, DC: Executive Office of the President of the United States; 2016.

[33] Charabidze D, et al. Forensic entomology and decomposition ecology in arid environments. *Forensic Sci Int*. 2020; 313:110331.

[34] International Association of Forensic Sciences (IAFS). *Global Perspectives on Forensic Science Collaboration and Innovation*. IAFS Reports; 2023.

Islamic University Journal of Applied Sciences (IUJAS)
https://journals.iu.edu.sa/jesc

Volume VIII, Issue II, December 2026, Pages 202-216

Non-Isothermal Melting Kinetics of Pb-61.9Sn-2Al Eutectic-Based Alloy: A Differential Thermal Analysis Study

Ali Alnakhlani [1] *

[1] Department of Physics, College of Science, Qassim University, Buraydah 51452, Saudi Arabia

*Corresponding author, *E-mail addresses*: A.Alnakhlany@qu.edu.sa

Abstract

The non-isothermal melting kinetics of a Pb–61.9 wt% Sn–2 wt% Al ternary alloy were investigated by differential thermal analysis (DTA) at five constant heating rates (β = 10, 15, 20, 25 and 30 °C min^{-1}). A 40 mg specimen was scanned from ambient to 400 °C under inert atmosphere using α-Al_2O_3 as the reference material. The melting onset (184.5–190.9 °C) remained in close agreement with the canonical Pb-Sn eutectic of 183 °C, confirming that the 2 wt% Al microaddition does not displace the eutectic temperature. Three complementary model-free kinetic treatments — Kissinger, Ozawa–Flynn–Wall (OFW) and Augis–Bennett — were applied to the main endotherm. The Kissinger and OFW methods yielded mutually consistent apparent activation energies of 138.63 and 139.35 kJ mol^{-1} (R^2 = 0.989 and 0.990) with a relative deviation ≈ 0.5 %,consistent with the fact that these two model-free treatments are near-identical transforms of the same (β, T_p) dataset; their close agreement therefore reflects internal arithmetic consistency rather than independent cross-validation. The Augis–Bennett analysis returned a systematically lower value of 73.63 kJ mol^{-1} (R^2 = 0.891), attributable to its inherent sensitivity to the heating-rate-dependent peak width, which broadened nearly threefold (from 18.04 to 50.67 °C) across the experimental range; this broadening is the distinctive thermal signature of the dispersed insoluble Al-rich grains introduced by the microaddition.

Keywords: Differential thermal analysis; Pb-Sn-Al alloy; Eutectic melting; non-isothermal kinetics; Activation energy; Kissinger method; Ozawa-Flynn-Wall; Augis-Bennett.

https://doi.org/10.63070/jesc.2026.040
Received 14 April 2026; Revised 28 May 2026; Accepted 13 June 2026; Available online 20 June 2026.
Published by Islamic University of Madinah on behalf of ***Islamic University Journal of Applied Sciences.***
This is an open access article published under the Creative Commons Attribution 4.0 International license (CC BY 4.0):https://creativecommons.org/licenses/by/4.0/

1. Introduction

Tin–lead solders have served as the cornerstone of metallurgical joining in microelectronics for more than a century, owing to their low melting temperatures, favourable wetting characteristics, and reproducible mechanical behaviour [1,2]. Although global legislative drivers have promoted lead-free formulations across consumer electronics over the past two decades [3–5], Pb-Sn solders remain essential in high-reliability sectors such as aerospace, defence, medical implants, and high-temperature power electronics, where qualified Pb-free substitutes are still under development [6,7]. The eutectic Pb-Sn composition (Sn-61.9 wt% / Pb-38.1 wt%) melts congruently at 183 °C and remains the reference benchmark against which all candidate joining alloys are compared [1,8,9]. Beyond binary Pb-Sn formulations, ternary modifications have attracted continuous interest because minor element additions can refine the solidification microstructure, alter the intermetallic-phase distribution, and tune the mechanical response without dramatically shifting the eutectic temperature [3,8]. Aluminium, in particular, exhibits very limited solid solubility in both lead and tin under near-equilibrium conditions [2], so that small Al additions to a near-eutectic Pb-Sn matrix form dispersions of primary Al-rich grains rather than a ternary solid solution. The mechanical and thermal consequences of such Al microadditions have, however, received comparatively little systematic attention in the open literature, and the corresponding melting kinetics, essential for accurate process modelling of reflow and joining cycles remain largely uncharacterised. Differential thermal analysis (DTA) and differential scanning calorimetry (DSC) are the established techniques for resolving solid–liquid phase transitions in metallic systems, providing direct access to the characteristic temperatures (onset, peak, end) and to the enthalpy associated with the transformation [9,10]. When DTA is performed at multiple constant heating rates, the peak temperature T_p of the melting event shifts to higher values with increasing heating rate β. This heating-rate dependence encodes the apparent activation energy of the underlying kinetic process, and three classical non-isothermal methods proposed independently by Kissinger [11], Ozawa [12], Flynn and Wall [13], and Augis and Bennett [14] exploit this shift to extract the apparent activation energy Ea without an a priori model of the reaction mechanism. These methods, originally developed for crystallisation and solid-state reaction kinetics [15–19], have since been extended to melting phenomena, polymer degradation, and glass transitions [20–22]. Applying more than one formulation provides a useful consistency check on the extracted E_a; it should be noted, however, that the Kissinger and OFW expressions are near-identical transforms of the same (β, T_p) data and therefore do not constitute independent cross-validation, whereas the Augis–Bennett treatment differs through its normalisation by the onset-to-peak interval [21–24].The recent critical review by Vyazovkin [24] emphasises that even when the Kissinger plot appears highly linear the underlying assumption of single-step first-order kinetics may mask additional complexity; consequently, the joint deployment of

complementary methods is now considered essential good practice. In the present study, a Pb-61.9Sn-2Al ternary alloy was investigated by DTA at five heating rates (β = 10, 15, 20, 25, and 30 °C min^{-1}) using a 40 mg specimen scanned from ambient (room temperature) to 400 °C. The principal objectives were: (i) to map the thermal signature of the alloy and identify all thermal events occurring within this temperature window; (ii) to extract the characteristic temperatures and enthalpy of the eutectic melting event; (iii) to determine the apparent activation energy of the melting transformation by the Kissinger, Ozawa-Flynn-Wall, and Augis-Bennett formulations and to compare these values critically; and (iv) to discuss the implications of the Al microaddition for the kinetics and morphology of the eutectic transformation. The results provide a quantitative kinetic and thermophysical database for Pb-Sn-Al solders that can inform reflow-profile design and serve as a baseline for further compositional optimisation.

2. Theoretical framework

Non-isothermal kinetic analysis rests on the general rate equation

$$\frac{d\alpha}{dt} = A \cdot exp\left(-\frac{E_a}{RT}\right) \cdot f(\alpha) \qquad (1)$$

where α is the conversion fraction, A the pre-exponential factor, R the universal gas constant, T the absolute temperature, and f(α) the reaction model. Under a constant heating rate β = dT/dt, eq. (1) becomes:

$$\frac{d\alpha}{dT} = \frac{A}{\beta} \cdot \exp\left(-\frac{E_a}{RT}\right) \cdot f(\alpha) \quad (2)$$

Three model-free formulations that derive Ea from the heating-rate dependence of the peak temperature Tp are summarised below [11–14]. It should be noted that the Kissinger, OFW, and Augis–Bennett formalisms were originally developed for thermally-activated solid-state and crystallisation kinetics. Eutectic melting is a first-order, thermodynamically-controlled phase transition; the activation energy extracted from the heating-rate dependence of T_p therefore represents an apparent/effective kinetic descriptor ($E_{a,app}$) that encapsulates contributions from heat transfer, thermal lag, and cooperative atomic rearrangement at the solid–liquid interface, rather than a true atomic activation barrier. Vyazovkin [24] has discussed these limitations explicitly and recommends labelling the extracted quantity as an apparent activation energy; this convention is adopted throughout the present work.

2.1 Kissinger Method

Kissinger [11] showed that, at the maximum reaction rate (i.e. at T = Tp), the second derivative $d^2\alpha/dt^2$ vanishes. Combining this condition with eq. (2) for a first-order kinetic process yields

$$\ln\left(\frac{\beta}{T_p^2}\right) = -\frac{E_a}{RT_p} + \ln\left(\frac{AR}{E_a}\right) \quad (3)$$

A plot of $\ln(\beta/T_p^2)$ vs. $1/T_p$ is linear, and its slope ($-E_a/R$) yields the apparent activation energy. Although derived for first-order reactions, the Kissinger equation has been shown to provide reasonably accurate Ea values for a broad range of nucleation-and-growth processes [19,23,24].

2.2 Ozawa-Flynn-Wall (OFW) method

The integral method introduced independently by Ozawa [12] and by Flynn and Wall [13] uses the Doyle approximation [18] of the temperature integral and does not assume a particular form of f(α). It relates the logarithm of the heating rate to the inverse peak temperature:

$$log_{10}\beta = -0.4567 \cdot \frac{E_a}{R\,T_p} + constant \quad (4)$$

The slope of log10 β versus 1/Tp therefore equals −0.4567 E_a/R, yielding $E_a = -R \cdot \text{slope} / 0.4567$. Because the OFW formulation is strictly model-free, it is widely recommended by the ICTAC Kinetics Committee as a cross-check against model-dependent treatments [21,22].

2.3. Augis-Bennett method

Augis and Bennett [14] proposed a modified Kissinger expression that retains the heating-rate dependence but normalises the heating rate by the temperature interval from onset (T_0) to peak (T_p):

$$\ln\left(\frac{\beta}{T_p - T_0}\right) = \ln\left(\frac{AR}{E_a}\right) - \frac{E_a}{RT_p} \quad (5)$$

This formulation has the practical advantage that the intercept directly gives the pre-exponential factor A, and it is particularly suited to transformations with a finite incubation interval. The slope of $\ln[\beta/(T_p - T_0)]$ against $1/T_p$ again yields E_a/R [14,19,23].

3. Experimental Methods

3.1 Alloy Preparation

Two alloy compositions were prepared from elemental Pb, Sn and Al constituents of ≥ 99.99 % purity: a binary Pb–61.9 wt% Sn reference composition and the principal Pb–61.9 wt% Sn–2 wt% Al ternary alloy (61.9 wt% Sn, 2 wt% Al, balance Pb = 36.1 wt%; all components sum to 100 wt%) investigated in detail in this study. Accurately weighed elemental charges were thoroughly mixed in a graphite crucible together with a $CaCl_2$ flux to prevent oxidation during melting. Each melt was cast in a rod geometry using a graphite mould of dimensions 15 × 1 × 1 cm. The as-cast ingots were homogenisation-annealed at 438 K (165 °C) for 50 h, then drawn into wires of 1 mm diameter by cold rolling. Prior to DTA measurements, the wire specimens were re-annealed at 443 K (170 °C) for 4 h and slowly cooled to room temperature at a controlled rate of $\Delta T = 2 \times 10^{-2}$ K s^{-1} to ensure thermodynamic equilibration of the microstructure.

3.2. Differential thermal analysis

DTA measurements were carried out using a calibrated LABSYS Evo TG–DTA thermal analyser with α-Al_2O_3 used as the inert reference material. A 40 mg coupon of the as-prepared wire was loaded into a platinum (Pt) crucible with Al_2O_3 as the reference material and scanned from room temperature to 400 °C under flowing inert atmosphere. Five constant heating rates were employed: β = 10, 15, 20, 25 and 30 °C min^{-1}. Temperature and heat-flow calibration were performed using high-purity indium (T_m = 156.6 °C) and tin (Tm = 231.9 °C) standards, consistent with the fixed points tabulated in the NIST solder-properties database [1] and in Smithells [25]. A freshly prepared coupon was used at each heating rate to minimise oxidation effects. The instrument calibration at each heating rate provides the heating-rate-dependent thermal-lag correction applied to all reported temperatures. The corrected onset temperatures confirm that the residual thermal lag across the β = 10–30 °C min^{-1} range is substantially smaller than the 14.2 °C shift observed in peak temperature Tp, such that the kinetic fits are not materially affected by uncorrected lag.

3.3. Peak parameter extraction and kinetic analysis

Onset (T_0), peak (T_p), and end (T_{end}) temperatures were extracted from each thermogram by the conventional tangent-baseline construction. The peak area, reported by the instrument in μV·s mg^{-1}, is proportional to the enthalpy of the transformation through the instrument calibration constant. Activation energies were derived from eqs. (3), (4), and (5) by linear regression of the appropriate transformed quantities against $1000/T_p$ (in K^{-1}).

4. Results and discussion

4.1 Differential thermal analysis (DTA) of Pb-61.9Sn-2Al

DTA scans at all five heating rates exhibited a single well-defined thermal event within the 25–400 °C range (Figure 1): the principal endothermic eutectic Pb–Sn melting transformation, characterised by a sharp, deep endotherm with onset near 184 °C and a heating-rate-dependent peak position in the 195–210 °C range. The transients observed at the low-temperature start of each curve (β = 15, 20, 25 and 30 °C min^{-1}) correspond to instrument equilibration and were excluded from the kinetic analysis. A minor exothermic feature at ~380–390 °C, common to all five scans and of sub-microvolt amplitude (<10% of the main endotherm), is attributed to surface oxidation of the molten Sn-rich phase by residual oxygen in the purge gas a documented oxidation behaviour of liquid Sn-rich solders, including eutectic Pb–Sn, in this temperature range [28], and is unrelated to any equilibrium transformation in the Pb–Sn–Al system. No other thermal events were resolved up to 400 °C, indicating that the 2 wt% Al microaddition introduces no detectable melting endotherm of an Al-rich

phase within this interval, consistent with the very low solid solubility of Al in both Pb and Sn [2,25] and the absence of stable low-temperature Al-containing intermetallics in the Pb-Sn-Al system.

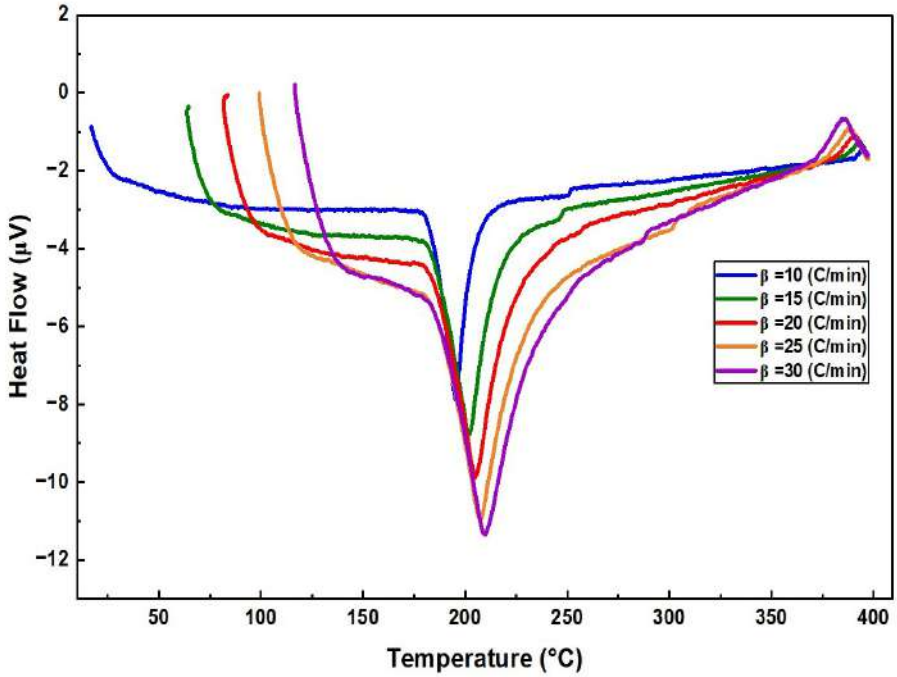


Figure 1. DTA thermograms of Pb–61.9Sn–2Al at five constant heating rates

4.2. Characteristic temperatures and enthalpy of fusion

The peak temperature T_p shifted systematically from 195.35 °C at β = 10 °C min^{-1} to 209.55 °C at β = 30 °C min^{-1} (Table 1), and the onset T0 increased from 184.49 to 190.85 °C, remaining within +1.5 to +7.9 °C of the canonical Pb-Sn eutectic of 183 °C tabulated in the NIST solder-properties database [1] and in the DSC study of Pb-Sn solid/liquid transformation data by Pedersen et al. [26]. The small positive offset at the lowest β is attributable to the 2 wt% Al microaddition combined with thermal lag; at higher heating rates the onset shift is dominated by the kinetic broadening of the transformation. The total T_p shift of 14.20 °C is an exceptionally clean kinetic observable that supports the application of the methods of Kissinger [11], Ozawa–Flynn–Wall [12,13] and Augis–Bennett [14] to extract the apparent activation energy.

Table 1. Experimental peak parameters of Pb-61.9Sn-2Al alloy melting endotherm.

β (°C min^{-1})	T_o (°C)	T_p (°C)	T_{end} (°C)	$T_{end} - T_o$ (°C)	Heat (μV·s mg^{-1})
10	184.49	195.35	202.53	18.04	8.233
15	186.66	201.80	211.92	25.26	7.609
20	188.51	204.85	218.27	29.76	6.858
25	189.29	207.40	237.36	48.07	7.052
30	190.85	209.55	241.52	50.67	6.818

Figure 2 illustrates the heating-rate dependence of the eutectic melting behaviour of the Pb-61.9Sn-2Al alloy. Figure 2(a) plots the three characteristic temperatures onset (T_{onset}, green squares), peak (T_p,

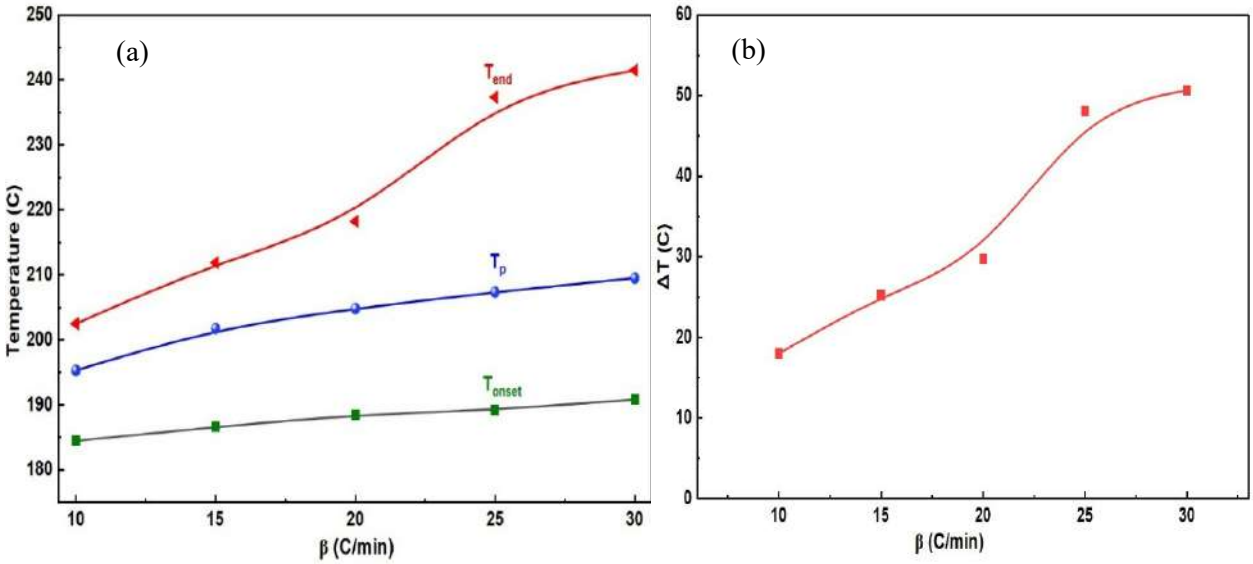


Figure 2. Heating-rate dependence of (a) characteristic temperatures of the alloy melting endotherm ; and (b) peak broadening ΔT of the Pb-61.9Sn-2Al alloy

blue diamonds) and end (T_{end}, red triangles) as a function of heating rate β. All three temperatures increase monotonically with β: T_{onset} rises gently from 184.49 °C at β = 10 °C min^{-1} to 190.85 °C at β = 30 °C min^{-1} (a modest +6.4 °C shift), T_p shifts by +14.20 °C across the same interval, while T_{end} exhibits the steepest dependence, climbing from 202.53 °C to 241.52 °C (+39 °C). The widening vertical spread between the three curves with increasing β provides direct visual evidence of kinetic broadening of the transformation. Figure 2(b) shows the corresponding peak width $\Delta T = T_{end} - T_{onset}$ as a function of β; ΔT increases monotonically from 18.0 °C at β = 10 °C min^{-1} to 50.7 °C at β = 30 °C min^{-1} almost a threefold broadening across the experimental range with the steepest rise occurring between β = 20 and 25 °C min^{-1}. The peak width $T_{end} - T_0$ broadens monotonically from 18.04 °C to 50.67 °C as β increases from 10 to 30 °C min^{-1}, almost a threefold increase. By comparison, pure Sn-37Pb eutectic typically exhibits a peak width below 10 °C at comparable heating rates [3,26,27].

4.3 Kissinger analysis

Application of the Kissinger formulation, eq. (3), to the peak temperatures of Table 1 gives the transformed quantities summarised in Table 2.

Table 2. Kissinger transformation of the Pb-61.9Sn-2Al alloy melting peak data.

β (°C min^{-1})	T_p (K)	1000/T_p (K^{-1})	T_p^2 (K^2)	ln(β/T_p^2)
10	468.50	2.1345	219492.2	−14.089
15	474.95	2.1055	225577.5	−13.712
20	478.00	2.0921	228484.0	−13.437
25	480.55	2.0810	230928.3	−13.225
30	482.70	2.0717	233000.3	−13.052

Linear regression of these data, shown in Figure 3, gives a slope of −16.674 K and an intercept of 21.461, with a coefficient of determination $R^2 = 0.989$. The apparent activation energy, $E_a = -\text{slope} \times R$, is 138.63 kJ mol^{-1}. The pre-exponential factor estimated from the Kissinger intercept is $A \approx 3.5 \times 10^{13}$ s^{-1}, for a melting transition this parameter serves as a formal regression coefficient extracted from the Kissinger intercept rather than a physical attempt frequency, and its absolute value is not independently interpreted here [21,24].

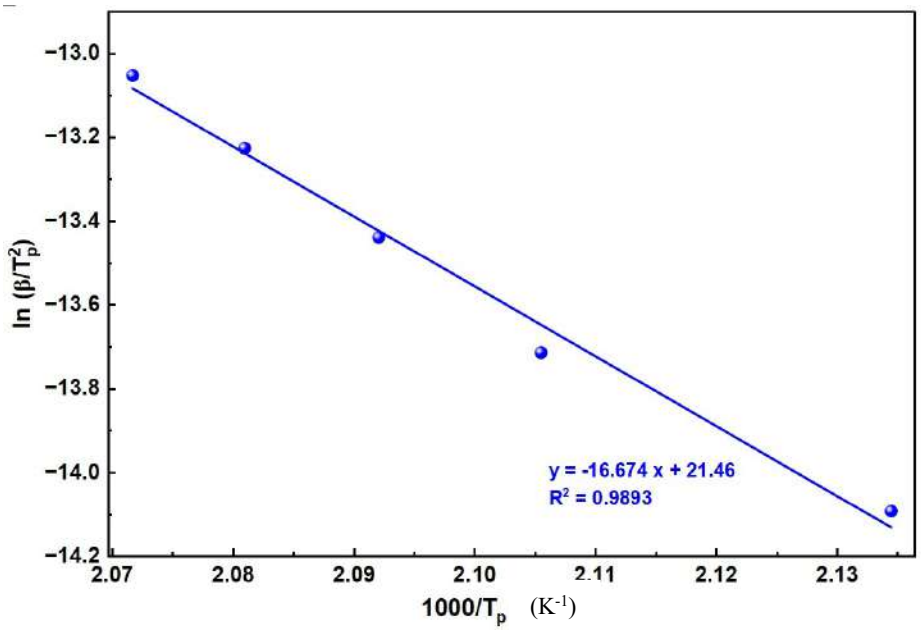


Figure 3 Kissinger plot, ln(β/T_p^2) vs 1000/T_p of the Pb-61.9Sn-2Al alloy

4.4. Ozawa-Flynn-Wall analysis

The model-free integral treatment was applied via eq. (4), with the corresponding regression shown in Figure 4. Linear regression of log10 β on 1000/T_p produced a slope of −7.654 K and an intercept of 17.321, with R^2 = 0.990. The corresponding apparent activation energy is 139.35 kJ mol^{-1}.

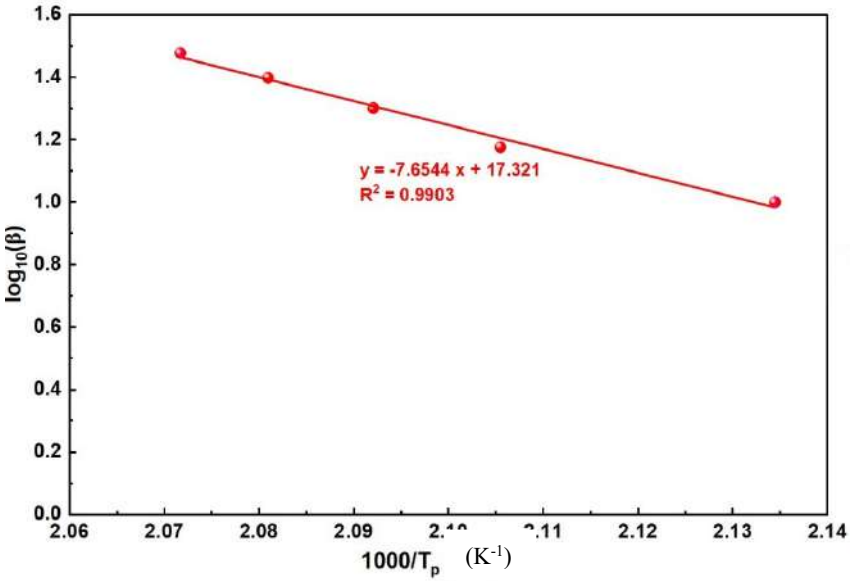


Figure 4 Ozawa–Flynn–Wall plot, $\log_{10}(\beta)$ vs 1000/T_p of the Pb-61.9Sn-2Al alloy

The relative deviation between this value and the Kissinger estimate is less than 0.5 %. Because the Kissinger and OFW expressions are near-identical transforms of the same (β, T_p) dataset, however, this close agreement is expected by construction; it therefore represents internal arithmetic consistency in the application of the two model-free treatments, and not an independent, method-independent cross-validation of the absolute value of E_a [24].

4.5. Augis-Bennett analysis

Application of the Augis-Bennett expression, eq. (5), incorporates the onset temperature T_0 through the temperature interval ($T_p - T_0$). The corresponding plot is shown in Figure 5 linear regression yields a slope of −8.856 K and an intercept of 14.657, with R^2 = 0.891. The corresponding apparent activation energy is 73.63 kJ mol^{-1}. The systematic reduction of E_a relative to the Kissinger and OFW estimates is a known consequence of the Augis-Bennett formulation: by normalising the heating rate by the temperature interval ($T_p - T_0$), which is itself an increasing function of β (Table 1; Figure 2b), the method partially absorbs the kinetic broadening into the apparent activation energy. The lower R^2 for this method (0.891 vs 0.989) reflects the noise introduced by this normalisation when ($T_p - T_0$) varies

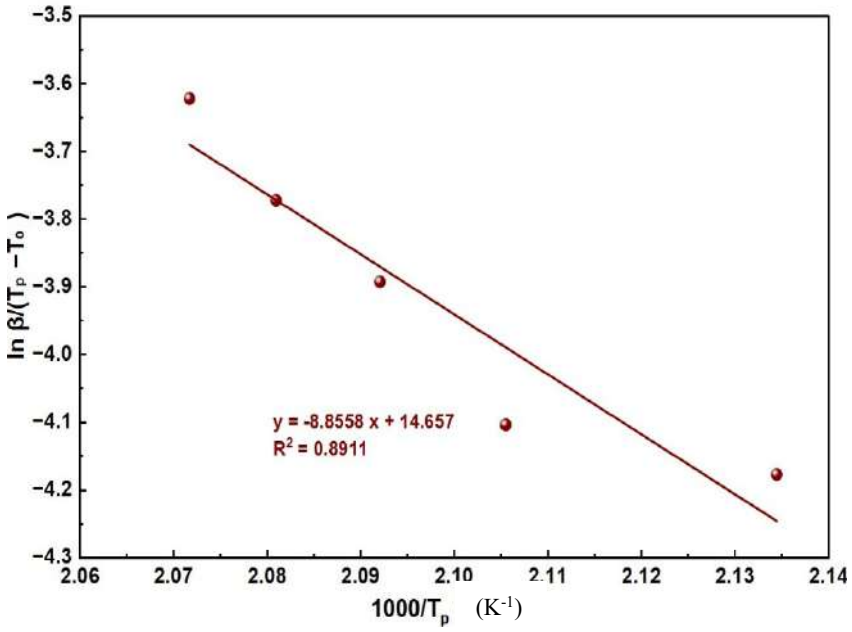


Figure 5. Augis–Bennett plot , $\ln[\beta/(T_p - T_0)]$ vs $1000/T_p$ of the Pb-61.9Sn-2Al alloy

strongly with β. Similar systematic discrepancies between Augis-Bennett and Kissinger estimates have been reported for crystallisation of chalcogenide glasses and amorphous metallic alloys [19,29].

4.6. Comparative kinetic assessment

Table 3 summarises the regression statistics and apparent activation energies obtained from the three methods, and Figure 6 compares the three values graphically against the established literature range for Pb–Sn solder melting kinetics. The Kissinger–OFW pair converge to $E_{a,app} \approx 139$ kJ mol^{-1} and are reported as the primary kinetic descriptor for this system. This figure lies firmly within the range of 50–200 kJ mol^{-1} reported for melting kinetics of low-melting Sn-based metallic alloys [26,29,30]. The systematic offset of the Augis–Bennett value reflects the genuine kinetic broadening of the peak with heating rate, an observable physical phenomenon rather than a defect of the method, and is therefore retained in the kinetic record as an independent indicator of microstructure-mediated heat transport in the alloy.

Table 3. Apparent activation energies from the three non-isothermal methods applied to the Pb-61.9Sn-2Al alloy melting peak.

Method	Slope (K)	Intercept	R^2	E_a (kJ mol^{-1})
Kissinger	−16.674	21.461	0.989	138.63
Ozawa-Flynn-Wall	−7.654	17.321	0.990	139.35
Augis-Bennett	−8.856	14.657	0.891	73.63

Figure 6 compares the apparent activation energy E_a of the eutectic Pb–Sn melting peak obtained from the three non-isothermal methods. The Kissinger and Ozawa–Flynn–Wall bars (138.6 and 139.3 kJ mol^{-1}) are visually indistinguishable, reflecting the ~0.5 % mutual deviation between the two model-free treatments and confirming the robustness of the kinetic descriptor. The Augis–Bennett bar (73.6 kJ mol^{-1}) is systematically displaced downward, a consequence of its normalisation by the heating-rate-dependent interval ($T_p - T_0$). All three values lie firmly within the green-shaded literature range of 50–200 kJ mol^{-1} reported for melting kinetics of low-melting Sn-based solder alloys; the dotted horizontal line marks the arithmetic mean across the three estimates (117.2 kJ mol^{-1}), shown for reference only and not regarded as a physical descriptor; see Section 4.6) and the regression quality of each method is inscribed within the corresponding bar (R^2 = 0.989, 0.990 and 0.891 for Kissinger, OFW and Augis–Bennett, respectively).

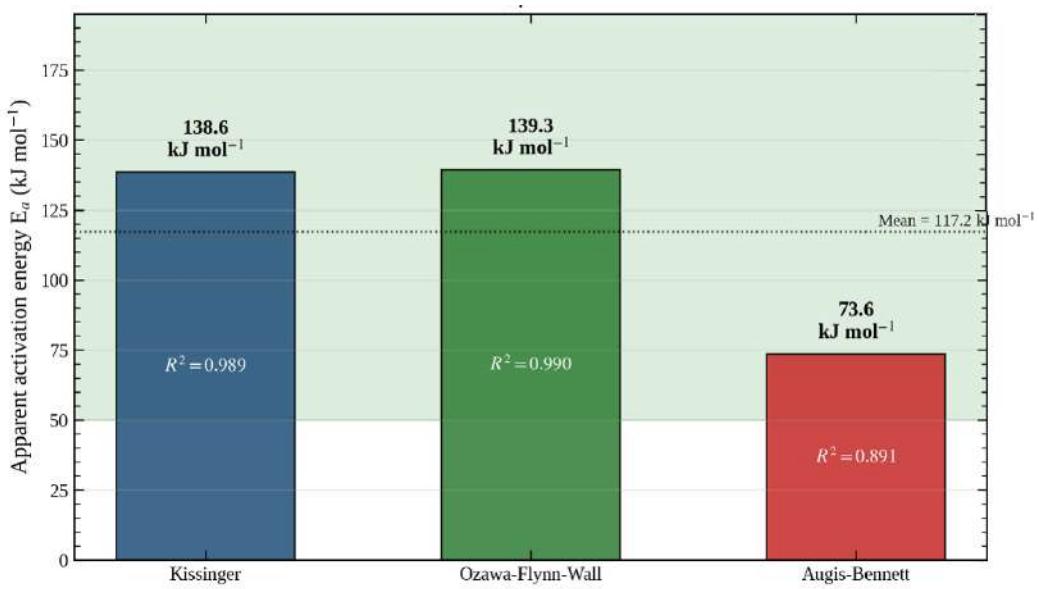


Figure 6. Apparent activation energy E_a for the melting peak in Pb-61.9Sn-2Al obtained from the three non-isothermal methods

4.7. Role of the aluminium microaddition

The persistence of a single, well-defined endothermic event with onset near 183 °C across all heating rates (Figure 1, Figure 2a) demonstrates that the eutectic Pb-Sn melting transformation is preserved in the presence of 2 wt% Al. The Al microaddition, however, produces a striking broadening of the melting peak: the total peak width ($T_{end} - T_0$) increases from 18.04 °C at β = 10 °C min^{-1} to 50.67 °C at β = 30 °C min^{-1} (Table 1; Figure 2b). In binary Sn-37Pb eutectic the corresponding width is typically below 10 °C at comparable heating rates [3,26,27]. The DTA data for the binary Pb–61.9Sn reference

alloy prepared alongside the ternary (Section 3.1) are available in Ref. [8].These dispersoids do not participate in the eutectic transformation, but they reduce the volume fraction of the eutectic phase available for melting at any given temperature and impose a fine-scale partitioning of the heat flow during the transformation, which manifests in DTA as peak broadening. This interpretation is consistent with the established role of insoluble dispersoids in metallic systems [27,29,31]. From a process-engineering standpoint, the practical implication is that the recommended reflow window for an Al-modified Pb-Sn solder must be widened relative to the binary eutectic to ensure complete melting throughout the joint volume. Using the highest peak temperature observed in the present study (T_p = 209.55 °C at β = 30 °C min^{-1}) and adding the customary 20–30 °C process margin recommended for Pb-Sn reflow [4,6,7], a peak reflow temperature of 230–240 °C would be appropriate for the Pb-61.9Sn-2Al formulation. This value remains comfortably below the upper bound of conventional reflow profiles and is therefore fully compatible with existing surface-mount infrastructure.

5. Conclusions

The non-isothermal thermal signature of a Pb–61.9Sn–2Al alloy has been characterised by DTA at five heating rates using a 40 mg specimen. The principal findings are:

- A single, well-defined endothermic transformation was resolved in every thermogram, corresponding to the eutectic Pb–Sn melting event, whose peak temperature shifted systematically from 195.35 °C to 209.55 °C across β = 10–30 °C min^{-1}.
- The melting onset (184.5–190.9 °C) confirms preservation of the canonical Pb –Sn eutectic; the 2 wt% Al microaddition does not displace the eutectic temperature.
- Kissinger and Ozawa–Flynn–Wall analyses converge to E_a = 138.63 and 139.35 kJ mol^{-1} ($R^2 \geq 0.989$; ~0.5 % mutual deviation), which represents internal arithmetic consistency between two near-identical transforms of the same (β, T_p) data rather than independent cross-validation; the Kissinger/OFW value (≈139 kJ mol^{-1}) is therefore adopted as the base apparent-activation-energy indicator. Augis–Bennett returns the lower 73.63 kJ mol^{-1} (R^2 = 0.891), reflecting its sensitivity to the heating-rate-dependent peak width.
- The 2 wt% Al microaddition broadens the melting peak nearly threefold while leaving the eutectic onset unshifted, reflecting the diagnostic thermal signature of dispersed insoluble Al-rich grains in the Pb-Sn matrix.

These results provide quantitative kinetic and thermophysical descriptors for a previously under-characterised Pb–Sn–Al solder composition and establish a baseline for the systematic compositional optimisation of Al-modified tin–lead solders with potential relevance to electronic joining applications.

References

[1] Siewert T, Liu S, Smith DR, Madeni JC. Database for Solder Properties with Emphasis on New Lead-free Solders, Release 4.0. Boulder (CO): NIST and Colorado School of Mines; 2002. Available at: https://www.nist.gov/publications/database-solder-properties-emphasis-new-lead-free-solders

[2] Massalski TB, editor. Binary Alloy Phase Diagrams. 2nd ed. Materials Park (OH): ASM International; 1990. ISBN 978-0871704030.

[3] Glazer J. Microstructure and mechanical properties of Pb-free solder alloys for low-cost electronic assembly: a review. J Electron Mater. 1994;23(8):693–700. https://doi.org/10.1007/BF02651361

[4] Abtew M, Selvaduray G. Lead-free solders in microelectronics. Mater Sci Eng R Rep. 2000;27(5–6):95–141. https://doi.org/10.1016/S0927-796X(00)00010-3

[5] Suganuma K. Advances in lead-free electronics soldering. Curr Opin Solid State Mater Sci. 2001;5(1):55–64. https://doi.org/10.1016/S1359-0286(00)00036-X

[6] Manko HH. Solders and Soldering: Materials, Design, Production, and Analysis for Reliable Bonding. 4th ed. New York: McGraw-Hill; 2001. ISBN 978-0071341394.

[7] Hwang JS. Modern Solder Technology for Competitive Electronics Manufacturing. New York: McGraw-Hill; 1996. ISBN 978-0070317499.

[8] Alnakhlani A, Hassan B, Abdulhafiz M. Effect of Al addition on the thermal properties of lead–tin alloy. Results Mater. 2021;12:100229. https://doi.org/10.1016/j.rinma.2021.100229

[9] Frear DR, Burchett SN, Morgan HS, Lau JH, editors. The Mechanics of Solder Alloy Interconnects. New York: Van Nostrand Reinhold; 1994. ISBN 978-0442012007.

[10] Vianco PT. General soldering. In: ASM Handbook, Vol. 6: Welding, Brazing, and Soldering. Materials Park (OH): ASM International; 1993. p. 938–58. ISBN 978-0871703828.

[11] Kissinger HE. Reaction kinetics in differential thermal analysis. Anal Chem. 1957;29(11):1702–6. https://doi.org/10.1021/ac60131a045

[12] Ozawa T. A new method of analyzing thermogravimetric data. Bull Chem Soc Jpn. 1965;38(11):1881–6. https://doi.org/10.1246/bcsj.38.1881

[13] Flynn JH, Wall LA. A quick, direct method for the determination of activation energy from thermogravimetric data. J Polym Sci B Polym Lett. 1966;4(5):323–8. https://doi.org/10.1002/pol.1966.110040504

[14] Augis JA, Bennett JE. Calculation of the Avrami parameters for heterogeneous solid state reactions using a modification of the Kissinger method. J Therm Anal. 1978;13(2):283–92. https://doi.org/10.1007/BF01912301

[15] Avrami M. Kinetics of phase change. I. General theory. J Chem Phys. 1939;7(12):1103–12. https://doi.org/10.1063/1.1750380

[16] Friedman HL. Kinetics of thermal degradation of char-forming plastics from thermogravimetry. Application to a phenolic plastic. J Polym Sci C Polym Symp. 1964;6(1):183–95. https://doi.org/10.1002/polc.5070060121

[17] Šesták J, Berggren G. Study of the kinetics of the mechanism of solid-state reactions at increasing temperatures. Thermochim Acta. 1971;3(1):1–12. https://doi.org/10.1016/0040-6031(71)85051-7

[18] Doyle CD. Estimating isothermal life from thermogravimetric data. J Appl Polym Sci. 1962;6(24):639–42. https://doi.org/10.1002/app.1962.070062406

[19] Boswell PG. On the calculation of activation energies using a modified Kissinger method. J Therm Anal. 1980;18(2):353–8. https://doi.org/10.1007/BF02055820

[20] Matusita K, Sakka S. Kinetic study of crystallization of glass by differential thermal analysis—criterion on application of Kissinger plot. J Non-Cryst Solids. 1980;38–39(2):741–6. https://doi.org/10.1016/0022-3093(80)90525-6

[21] Vyazovkin S, Burnham AK, Criado JM, Pérez-Maqueda LA, Popescu C, Sbirrazzuoli N. ICTAC Kinetics Committee recommendations for performing kinetic computations on thermal analysis data. Thermochim Acta. 2011;520(1–2):1–19. https://doi.org/10.1016/j.tca.2011.03.034

[22] Vyazovkin S, Chrissafis K, Di Lorenzo ML, Koga N, Pijolat M, Roduit B, Sbirrazzuoli N, Suñol JJ. ICTAC Kinetics Committee recommendations for collecting experimental thermal analysis data for kinetic computations. Thermochim Acta. 2014;590:1–23. https://doi.org/10.1016/j.tca.2014.05.036

[23] Starink MJ. The determination of activation energy from linear heating rate experiments: a comparison of the accuracy of isoconversion methods. Thermochim Acta. 2003;404(1–2):163–76. https://doi.org/10.1016/S0040-6031(03)00144-8

[24] Vyazovkin S. Kissinger method in kinetics of materials: things to beware and be aware of. Molecules. 2020;25(12):2813. https://doi.org/10.3390/molecules25122813

[25] Brandes EA, Brook GB, editors. Smithells Metals Reference Book. 7th ed. Oxford: Butterworth-Heinemann; 1992. ISBN 978-0750610209.

[26] Pedersen AS, Pryds N, Linderoth S, Larsen PH, Kjøller J. The determination of dynamic and equilibrium solid/liquid transformation data for Pb-Sn using DSC. J Therm Anal Calorim. 2001;64(3):887–94. https://doi.org/10.1023/A:1011506526074

[27] Lee NC. Reflow Soldering Processes and Troubleshooting: SMT, BGA, CSP and Flip Chip Technologies. Boston (MA): Newnes; 2002. ISBN 978-0750672184.

[28] Lin KL, Liu TP. High-temperature oxidation of a Sn-Zn-Al solder. Oxid Met. 1998;50(3–4):255–67. https://doi.org/10.1023/A:1018840405283

[29] Mehta N, Kumar A. Comparative analysis of calorimetric studies on Se85Te10M5 (M = Ag, Zn and Cd) chalcogenide glasses. J Therm Anal Calorim. 2007;87(2):343–9. https://doi.org/10.1007/s10973-006-7720-1

[30] Mavoori H, Chin J, Vaynman S, Moran B, Keer L, Fine M. Creep, stress relaxation, and plastic deformation in Sn-Ag and Sn-Zn eutectic solders. J Electron Mater. 1997;26:783–90. https://doi.org/10.1007/s11664-997-0252-z

[31] McCabe RJ, Fine ME. Athermal and thermally activated plastic flow in low melting temperature solders at small stresses. Scr Mater. 1998;39(2):189–95. https://doi.org/10.1016/S1359-6462(98)00149-4

Islamic University Journal of Applied Sciences (IUJAS)
https://journals.iu.edu.sa/jesc

Volume VIII, Issue II, December 2026, Pages 217-235

Nanotechnology in Wastewater Treatment: Evolution and Environmental Impact

Basheer Ahamed Abdul Basith [1], Nkopeleng John Kganyago [1], Maher Fathalla [1*]

[1] Department of Chemistry, Faculty of Science, Islamic University of Madinah, 42351,
*Corresponding author: (Maher Fathalla), *Email Address*: maherfathalla@iu.edu.sa

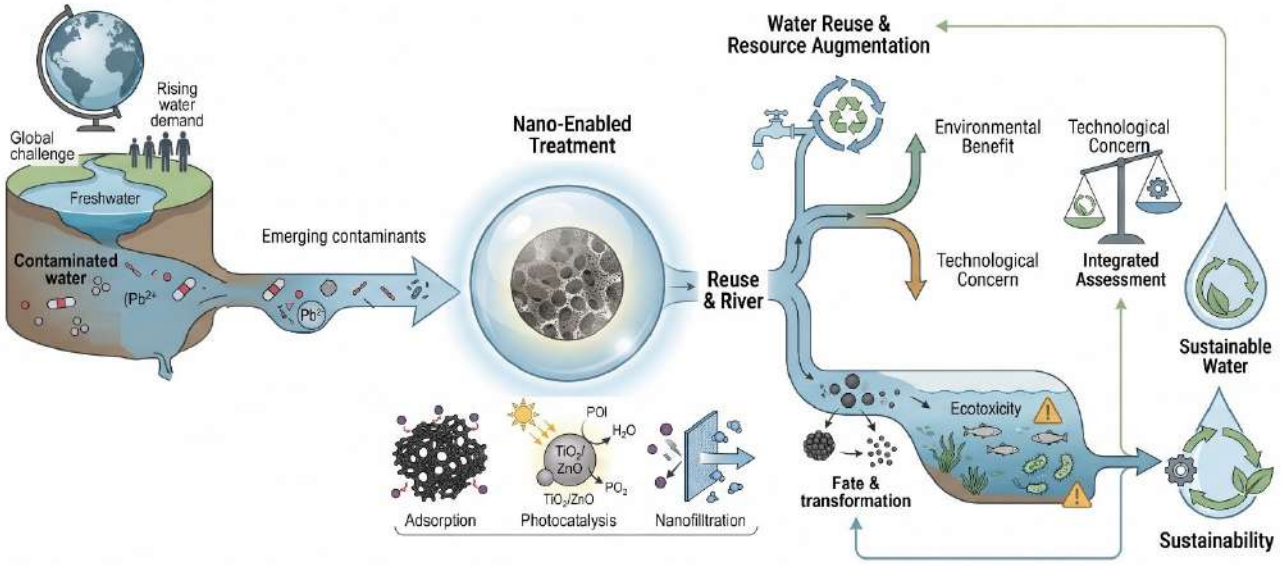
Abstract

Access to affordable, clean water is one of the biggest challenges of the 21st century. Globally, water supply cannot meet rapidly increasing demand. This is made worse by industrialization, population growth, and emerging contaminants. The setbacks in the removal efficiency of conventional water treatment techniques demanding the requirement for technological innovation to address integrated water management systems. Through the safe utilization of advanced research, nanotechnology has emerged as a promising solution for improving water and wastewater treatment efficiency and augmenting water supply. However, unresolved issues remain regarding the fate, transformation, and ecotoxicity of nanomaterials, where their acute impacts on aquatic systems and the broader ecosystem may outweigh technological improvements. Here, we review the contemporary advancements in both nanotechnology and conventional treatment methods. The toxicity of the nanomaterials is discussed, followed by their properties and mechanisms that allow for applications, as well as their benefits and drawbacks in comparison to current methods. This review describes the integrated connectivity between the advancements and obstacles to further develop these unique characteristics for sustainable water treatment by tracing the connectivity between the available treatment methods, technological advancements, and the ecotoxicity of nanomaterials.

Keywords: Nanotechnology; Wastewater treatment; Nanomaterials; Ecotoxicity.

https://doi.org/10.63070/jesc.2026.041
Received 20 May 2026; Revised 28 August 2026; Accepted 10 September 2026; Available online 15 September 2026. Published by Islamic University of Madinah on behalf of ***Islamic University Journal of Applied Sciences***.
This is an open access article published under the Creative Commons Attribution 4.0 International license (CC BY 4.0): https://creativecommons.org/licenses/by/4.0/

Graphical Abstract



1. Introduction

One of the most fundamental humanitarian objectives and which all life on earth depends on access to clean water and sanitation. But it continues to be a significant worldwide concern for the twenty-first century. There are many issues with our present water supply, both historic and contemporary. According to WHO (2024) [1] 2.1 billion people worldwide still do not have access to clean drinking water sources**.** Since water and wastewater facilities are frequently nonexistent in the impacted areas (mostly in developing nations), it is imperative to provide basic water treatment. By damaging natural water supplies, human activity is increasingly contributing to the problem of water scarcity in both developed and developing nations.

In the current situation, extensively used water treatment and distribution systems in developed nations are under increased scrutiny because of the stricter water quality regulations and new contaminants. The demand for water is rising due to population growth and rising living standards, and the existing uneven availability of freshwater is being disrupted by climate change. Consequently, areas under water stress are increasingly using brackish water, seawater, and wastewater as alternative sources [2]. Additionally, the infrastructure and technology used for water and wastewater treatment are at their limit in terms of producing water that is suitable for human and environmental demands. Meanwhile, untreated wastewater remains a major global environmental and public health concern, with around 80% discharged without treatment or reuse. Consequently, about 1.8 billion people rely on contaminated water sources [3]. Further increasing the risk of waterborne diseases such as cholera, dysentery, typhoid, and polio. Additionally, sewage treatment plants contribute significantly to the problem by releasing billions of gallons of untreated waste annually [4].

Over the past 20 years, nanotechnology has advanced dramatically, moving from early discoveries like carbon nanotubes in the 1990s to widespread commercial applications in the early 2000s [5], [6]. Currently, it is a well-established and quickly growing field with wide-ranging applications in organic chemistry, molecular biology, and many other scientific and technological fields [7], [8]. In that way, recent advancements in nanotechnology provide rapid growth opportunities to construct next-generation water supply systems. Our existing methods for treating water mainly rely on centralized systems and transportation, which are unsustainable.

Nanomaterials (NMs) are widely used in daily life, which has unavoidably led to their release and deposition into the environment, including the land, water, and air [9]. Since NMs can interact with organisms once they are in various environmental matrices, the ecotoxicology of NMs in species from various trophic levels has emerged as a global concern. Although they can enter organisms directly or indirectly by ingestion, adsorption, inhalation, or the food chain, they have a significant potential concern in regards of One Health framework [10].

The literature review adopts a thematic, structured narrative approach. Targeted searches in Scopus and Google Scholar (2020-2026) were used to find literature using search strings with combined keywords aligned with each thematic section. It is also supplemented by earlier seminal works when necessary. “Conventional treatment, advanced/hybrid processes, nanotechnology in wastewater treatment, comparative advantages over conventional methods, ecotoxicity, nanotoxicity and ecological impacts of nanomaterials” are the strings utilized for the review. Peer-reviewed journal articles in English focusing on performance, mechanisms, or environmental effects were included, while studies irrelevant to wastewater treatment, lacking experimental or application-oriented data were excluded. Papers were screened by title, abstract, and relevance to the thematic progression. The scope of this review emphasizes the conceptual connectivity and technological evolution across these topics, which highlights how advances in efficiency and functionality drive emerging environmental concerns and inform future research directions.

This review explores how water treatment techniques change throughout time in response to research and technological advancements. It seeks to chart the evolution from conventional water treatment to nanotechnology-based wastewater processes and their ecological implications. In addition, what are the main obstacles that existing therapeutic approaches must overcome, and why and how nanotechnology uses its vast capacity and molecular innovation to address such problems. To what degree do those approaches have advantages over the ones that are now available? The shortcomings of traditional water treatment techniques will be integrated into the role of using nanotechnology to overcome them in the parts that follow. Lastly, it talks about the disadvantages and repercussions associated with offering ongoing strategies for overcoming them.

2. Evolution of conventional water treatment methods

The earliest engineered wastewater systems relied on simple physical operations such as settling ponds, land application, and basic sand filtration, whose main purpose was to remove suspended solids and reduce odour before discharge [11]. As cities expanded and sewage loads increased, these

rudimentary systems proved insufficient to protect public health, leading to the development of multi-stage treatment plants with defined primary, secondary and tertiary steps. Primary treatment introduced sedimentation tanks and, in some cases, dissolved air flotation to separate settleable and floatable solids, producing primary sludge and a clarified effluent for biological treatment [12]. Secondary treatment was dominated by biological processes such as the activated sludge process and trickling filters, where microbial communities oxidized dissolved and colloidal organic matter, thereby significantly reducing biochemical oxygen demand and suspended solids [11], [13]. Tertiary treatment, initially limited to sand filtration and chlorination, was later expanded to include nutrient removal and more refined disinfection technologies to meet tightening discharge standards.

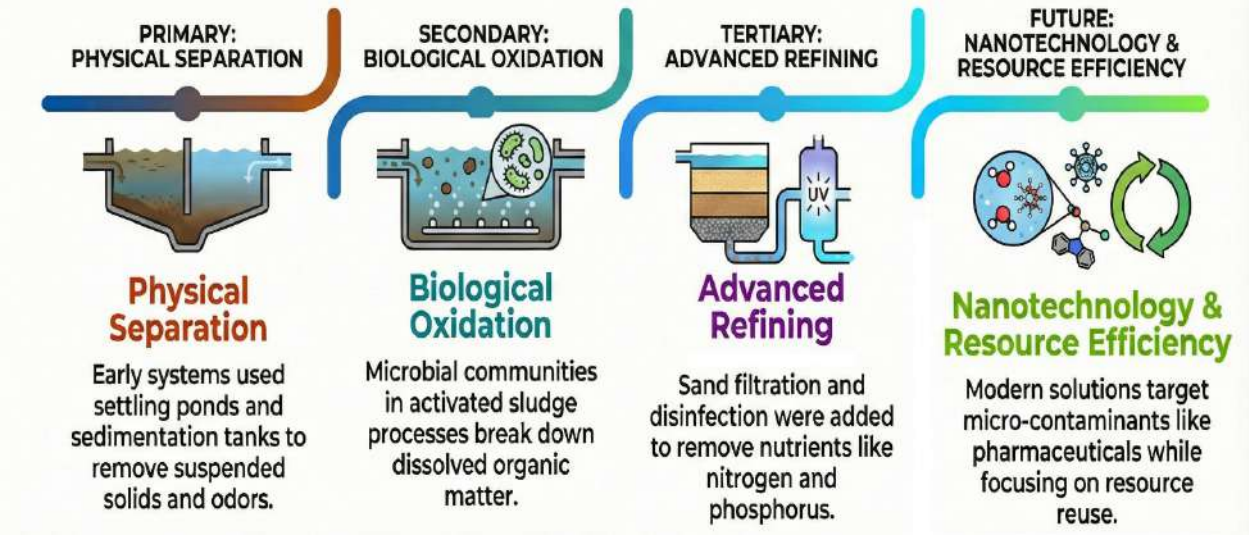


Fig. 1. Detailing of conventional water treatment method and nanotechnology targeted remediation.

2.1 Advanced and hybrid wastewater processes

Advanced wastewater treatment technologies were introduced to bridge the gap between what conventional methods can achieve and, the more stringent requirements associated with reuse and, emerging pollutants [12], [13]. Membrane processes are a central element of this evolution, ranging from microfiltration, ultrafiltration, nanofiltration and reverse osmosis. Each targeting different size fractions and dissolved species [6], [14]. Membrane bioreactors (MBRs) integrate biological degradation with membrane separation, enabling compact plants with high biomass concentrations and effluents of very low turbidity and pathogen content. These systems are now widely applied in municipal and industrial sectors, although fouling control, energy demand and concentrate management remain key design and operational considerations [12], [15].

Alongside membranes, advanced oxidation processes (AOPs) such as ozonation, UV/H_2O_2 and Fenton-type reactions generate highly reactive radicals capable of degrading recalcitrant organic molecules into smaller, often more biodegradable products [13], [16]. These processes are especially relevant for treating pharmaceutical wastewaters and other streams that contain persistent

micro-pollutants [2], [17]. Granular activated carbon adsorption, ion exchange and electrochemical methods provide additional polishing steps for dissolved organic compounds (DOC), metals and colors.

In many modern designs, treatment lines combine multiple processes biological, membrane, adsorption and oxidation into hybrid systems tailored to the specific composition of the wastewater and the desired reuse purpose. The overarching trend is towards modular, intensified configurations that offer higher removal efficiencies and better control over the treated water quality than traditional standalone units [17], [18]. Therefore, even the most advanced modern methods of treating wastewater quickly expire due to the increased demand for water in both domestic and industrial settings, as well as the introduction of new products and chemicals to address the growing problems. This calls for a different approach to wastewater treatment. For this, the main goal should be to identify the new challenges and growing concerns [19], [20].

3. Emerging issues in wastewater treatment

Despite their success in reducing conventional pollutants, modern wastewater now carries a complex mixture of contaminants that were not the primary design targets of older plants, including pharmaceuticals, personal-care products, hormones, flame retardants, microplastics and pesticides, often termed contaminants of emerging concern (CECs) [2], [21], [20]. These compounds occur at very low concentrations but are biologically active and may act synergistically, so long-term exposure can affect ecosystems and human health even when individual concentrations are low. Sources include hospital and household effluents, industrial discharges, and agricultural runoff, meaning CECs continuously enter sewer systems and treatment plants, many of which only partially remove them.

The logistical and environmental weight of the "huge amount of sludge" produced by chemical precipitation is also enormous. The expenses of stabilizing, dewatering, and disposing of hazardous sludge frequently exceed the advantages of the initial treatment. Moreover, the technical intricacy and "high resin cost" of renewing spent resin in ion exchange processes restrict their use in high-volume industrial applications [22]. These outdated techniques are strategically inferior to membrane-based technologies that provide a clean separation of contaminants without the bulk production of secondary waste since they are fundamentally "sludge-heavy" and resource-intensive. Adoption of pressure-driven separation systems results directly from this shift away from byproduct-heavy processes, with nanofiltration becoming the predominant architecture [12].

Many treatment plants also struggled with nutrient removal, especially nitrogen and phosphorus, which drive eutrophication in receiving waters. Operational challenges, including sludge production, energy consumption for aeration and the formation of disinfection by-products, further motivated the search for improved methods [12], [13]. Over time, the focus of wastewater treatment broadened from simple

pollution abatement towards resource efficiency, reuse and removal of emerging contaminants, which paved the way for advanced and hybrid technologies.

This historical development created the platform on which modern nanotechnology-enhanced solutions come in. We can no longer sustain our current methods of water treatment, distribution, or discharge, which mostly rely on centralized infrastructure and transportation. It is anticipated that the highly effective, modular, and multipurpose processes will be made possible by nanotechnology. Which expected to offer high-performing, reasonably priced water and wastewater treatment solutions that depend less on sizable infrastructures [16].

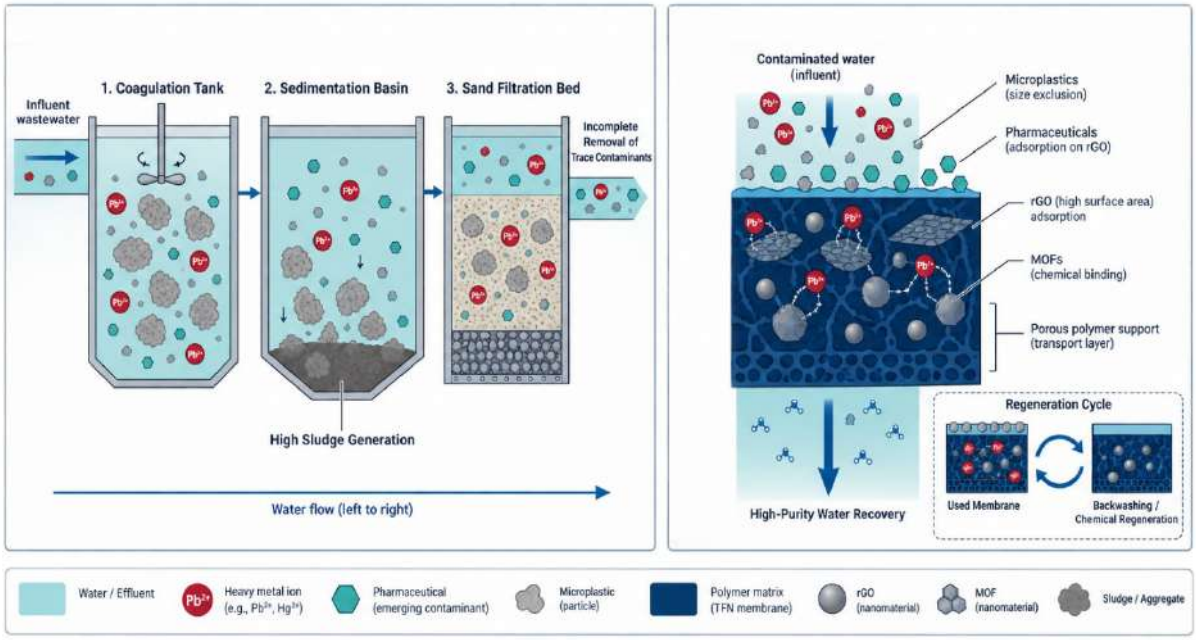


Fig. 2. Comparison of conventional and nanotechnology-based wastewater treatment systems. Conventional treatment shows high sludge generation and incomplete removal of trace contaminants. Nano-enabled filters achieving high-purity water recovery through adsorption, size exclusion, and chemical binding.

4. Nanotechnology in Wastewater Treatment

The performance of nanomaterials has been investigated in the past few years in a variety of aquatic habitats, such as lakes, rivers, wetlands, and the ocean. Environmental research groups have found that finding environmentally friendly remediating chemicals to restore contaminated soil and water bodies is a shared issue in the pursuit of an environmentally growing economy [23], [24]. The removal of charged components from solutions, such as the production of fresh water from brackish water, is the subject of several recent investigations on electro membrane (separation) technologies [25], [26].

Currently, nanomaterials are mostly used for sewage treatment, and nano-photocatalytic technology is widely used to oxidize a variety of organic compounds found in water by producing hydroxide and superoxide ions with high oxidation and superior activity. Many nanostructured materials are currently employed in capacitive deionization (CDI), a new water treatment technology. In aquatic environments, the nanoparticles exhibit varying characteristics. Some of these nanomaterials have

shown greater promise for cleaning up contaminated water. In addition, pressure-driven membranes are used in nanofiltration (NF) membrane technology, which separates the materials found in sewage [27].

Furthermore, strong adsorption of metal ions and organic contaminants is demonstrated by some nanomaterials because of surface hydroxyl groups that bind targeted cations. Particularly, carbon-based nanoparticles provide superior wastewater remediation over traditional treatment techniques due to their excellent stability over a broad pH range, large specific surface area, abundance of active sites, and porous architectures. [28]. Materials having adjustable surface chemistry and special catalytic or magnetic qualities are made possible by nanotechnology for the treatment of wastewater. High adsorption capacities are demonstrated by nanomaterials such as metal-oxide nanoparticles, magnetic nanoparticles, carbon nanotubes, and graphene-based composites, which achieve removal efficiencies of 50–98% for organic pollutants and 80–99% for inorganic contaminants. [27]. These materials can be used as dispersed adsorbents, as coatings on larger supports, or embedded in membranes and filters to enhance selectivity and flux.

Beyond conventional adsorbents, an expanding set of nano-adsorbent classes is being studied for wastewater, including:

a) Metal-oxide nanoparticles (e.g. TiO_2, ZnO, Fe_3O_4, MnO_2) for heavy metals, dyes and organic pollutants.
b) Carbon-based nanomaterials (CNTs, graphene, graphene oxide, nanodiamonds) with high capacity for heavy metals, dyes and pharmaceuticals.
c) Magnetic nanocomposites (e.g. Fe_3O_4-coated carbon or polymer beads) that can be rapidly separated by magnetic fields, simplifying recovery and regeneration.
d) Metal–organic frameworks (MOFs) and covalent organic frameworks (COFs), which offer highly tunable pore structures and surface functionalities for targeted removal of pharmaceuticals, PFAS and dyes.
e) Polymeric and biopolymer-based nano-adsorbents, including chitosan, cellulose nanofibers and polymer-filled nanomembranes with functional groups such as amine, carboxyl and sulfonic acid for selective metal and dye binding.

Nanomaterials also play a central role in catalytic and photocatalytic treatments. Semiconductor nanoparticles such as TiO_2 and ZnO, including doped variants, act as photocatalysts under UV or visible light, generating reactive radicals that degrade persistent organic pollutants. These catalysts can be integrated into coatings, suspended photocatalytic slurries, or nanocomposite membranes to simultaneously degrade micro-pollutants and filter solids. Other nano-catalysts, including noble metal nanoparticles and nano-Fenton catalysts, enhance AOPs by accelerating radical generation and electron transfer.

5. Advantages of nanotechnology compared to conventional methods.

Industrial wastewater remediation is a key component of the infrastructure for public health and environmental engineering globally. The release of hazardous heavy metal ions from the production of paint, automobile batteries, pigments, and fertilizer has increased to catastrophic levels as industrial operations expand. Since excessive exposure causes long-term harm to the reproductive, cardiovascular, and immune systems as well as to the kidneys, liver, and brain, these effluents are a systemic hazard [7], [17].

Table 1. Application of various nanomaterials in the remediation of water contamination

No	Nanomaterial	Common Manufacturing Techniques	Mechanism	Remediation	Reference
1	Graphene Nanomaterial	Chemical oxidation-reduction, mechanical milling, pyrolysis, photocatalytic reduction, carbon nanotube (CNT) conversion, and plasma synthesis	The pollutants in the water are absorbed by the nanomaterial's porous features.	Heavy metals and Dyes	[29]
2	Titanium dioxide (TiO_2) nanoparticles	Gas and Liquid phase techniques	Stronger photo-oxidation and reduction capabilities might be produced by the stimulation of TiO2 nanoparticles by ultraviolet (UV) light, which would also produce free radicals with significant catalytic activity.	Different organic substances like formaldehyde and certain inorganic substance	[30]
3	Nano-Zerovalent iron (nZVI)	Reduction	Through surface complexation and electrostatic interaction, the thin iron oxide core could facilitate the adsorption of pollutants.	Heavy metals, chlorinated aromatic compounds, nitro-aromatic compounds, azodyes and, organo-chlorine	[31]
4	Nanofiber Material	Centrifugal jet spinning (CJS), solution blowing, electrospinning (ES), electrohydraulic dynamic (EHD) direct writing, molecular methods, etc.	Using membranes with different pore widths, particles of different dimensions are filtered.	Various contaminants in eutrophicated landscape water	[32]
5	Silver Nanoparticles	Laser ablation method, photoreduction, radiation, electroplating, chemical vapor deposition	By interacting with the enzymes that the bacteria use for metabolism, the bacteria will become dried out and eventually perish. Smaller particles have better germicidal efficacy.	Water photocatalyst and antibacterial agent	[33]

Previously, promising outcomes have been demonstrated by traditional methods such as biological treatments, coagulation/flocculation, and flotation. However, the high sludge production, limited removal rates for diluted metals, and pH sensitivity are since held to be drawbacks [34]. The emerging shortcomings of traditional treatment techniques, which are progressively unable to manage the evolving stoichiometry of contemporary industrial waste, highlight the need for this shift. These limitations are often constrained by their expense, side effects, or inefficiency [35]. According to [36] conventional ponding and coagulation-flocculation systems have to deal with enlarged treatment timeframes, substantial land utilization, secondary sludge formation, and uneven pollutant removal.

High surface-to-volume ratios, high adsorption capacity, and simple functionalization are how nanotechnology gets beyond these restrictions [37], [38]. Nanofibrous membranes remove up to 98% of heavy metals, while nanoparticle-infused membranes remove up to 99% [34]. NF has a lower energy requirement than RO and operates at 5–20 bar with strong divalent-ion rejection [37].

Table 2: Key performance indicators with essential trade-off interpretation for wastewater treatment systems based on traditional and nanotechnology

Metric	Conventional Systems	Nanotechnology Systems	Key Trade-Off	Reference
Removal Efficiency	56.7% COD (coagulation) (Naguib et al., 2024)	93% COD (Fe/Cu NPs, 30 min), 95% dye (24 h), up to 99% heavy metals	Nano removes 1.6-1.8× more material at a higher expense.	[34], [41], [42]
Contact Time	Hours to days (ponding, digestion) (Raji et al., 2026)	30 min (Fe/Cu NPs), 24 h (bio-NP dye removal)	Faster kinetics; however, contact time varies depending on the type of NP and contaminant	[41], [42]
Stability	Well-characterized. Polymer degradation under high pressure (Jain et al., 2021)	Graphene oxide enhances mechanical stability of adsorbents. Long-term stability unproven	Although field evidence is scarce, nanotechnology increases material stability.	[38], [43]
Regeneration	Limited; sludge-bound contaminants difficult to recover (Malik et al., 2022)	Magnetically separable NPs permit multiple reuse cycles, Reusability remains a major problem	Regeneration is feasible but not yet dependable on a large scale.	[34], [37]
Cost	Lower capital; high disposal and sludge management costs (Ayach et al., 2024)	High-purity nanomaterials costly; ~90% cost reduction with lower-purity substitutes, Long-term reusability offsets cost	Reuse cycles and purity trade-offs determine how cost-effective it is.	[37]

The cost of developing nanomaterials is still a significant obstacle. Although replacing ultrapure C60 with lower-purity materials like fullerene soot can cut costs by about 90% with little loss in efficacy [37]. Lab-scale efficiency has not been verified at the volumes required for municipal deployment, and there are few industrial-scale investigations [39], [40]. Wastewater sludge contains nanoparticles including Mg, Ni, Ag, and Cd. At the plant level, disposing of this polluted material is still a big challenge [38]. Although pH, concentration, and contact duration all affect how harmful nanomaterials are [37].

Table 2 comparatively discusses the major parameter metrics along with their key trade-off markups. It reveals that, although nanotechnology systems provide 1.6 to 1.8 times greater removal efficiency and faster kinetics, these benefits are contingent upon contact time, NP type, and contaminant specificity. In 30 minutes, Fe/Cu nanoparticles remove 93% of COD at pH 7.5, whereas coagulation removes only 56.7% at pH 8.5 [41]. One important cost component is regeneration capacity. Reusability of nanoadsorbents is still a significant issue, although magnetically separable multifunctional nanomaterials allow for several reuse cycles [34], [37]. The balance between nanomaterial purity, regeneration cycles, and disposal of NP-laden sludge is the cost frontier, and it is unavoidable to incorporate these aspects to meet treatment enhancement [37], [38].

Despite these advantages, large-scale deployment of nanotechnology in wastewater treatment faces obstacles, including relatively high production costs, energy- or solvent-intensive synthesis routes and concerns over environmental release, ecotoxicity and bioaccumulation of engineered nanoparticles. Current research therefore emphasizes immobilizing nanoparticles on stable supports, using bio-based or biodegradable nanomaterials and performing full life-cycle and risk assessments to ensure safe, economically viable applications.

6. Ecological impacts of nanomaterials.

Nanomaterials are currently being used widely and progressively in a number of fields apart from water treatment, including materials science, cosmetics, medicines, and more. However, using nanomaterials also has adverse effects on ecology and the greater environment [44], [45]. It is inevitable for a number of nanomaterials to end up in rivers, lakes, and the ocean. In the end, these nanoparticles can also affect potable water. Therefore, the contamination of water habitats caused by these nanoparticles is hazardous to human and aquatic life [46], [47]. It has been observed that different types of nanomaterials, or even similar types of nanoparticles at different sizes, have an incompatible detrimental effect on some organisms. Because of their strong reactivity, it is evident that nanoparticles change from their original, manufactured state regardless of the kind, quantity, or route of environmental nanoparticle release.

Numerous mechanisms, including aggregation/agglomeration, redox reactions, dissolution, surface moiety exchange, and reactivity with biomacromolecules, can lead to transformations [48]. The transport, destiny, and toxicity of nanoparticles in the environment are all impacted by these dynamic

alterations, hence it is essential to comprehend and describe them. Several reviews have provided a thorough explanation of the current analytical instruments used to quantify the properties of nanoparticles [49].

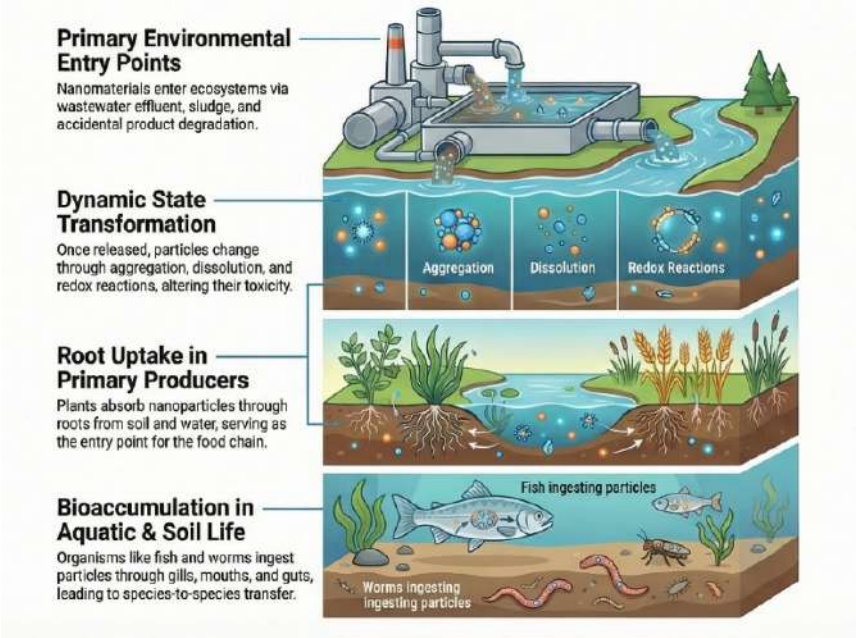


Fig. 3. The path of nanotoxicity: From release to the food chain.

Due to their interactions with soil, water, and air, all of which may contain nanoparticles, plants are especially pertinent when considering eco-nanotoxicity. Additionally, because lower trophic level organisms, animals, and humans eat plants, they offer a substantial possibility to enhance the transfer of nanoparticles among different species in the food chain. Although the ecosystem has a wide range of plant species, the majority of research on nanotoxicity to date has concentrated on plants that are consumed by humans, such as maize, wheat, soybean, tobacco, and numerous fruit or vegetable plants, including pumpkin, cucumber, and radish [50], [51]. Consequently, living organisms are exposed to nanoparticles in different ways depending on where they live. As was recently discovered, entry through roots is a major issue for plants. Nanoparticles that are present in the aqueous or soil phases give fish in water or worms in soil a way to enter the animal (gills, mouth, or gut), which may be significantly changed by the surface characteristics of the nanoparticle.

Early research on the toxicity of nanomaterials in fish concentrated on a small number of larger animals that humans might consume, like fish and salmon, with as few as three fish being employed for crucial measures. The biological impacts of hydrophobic carbon-based materials are caused by organic solvent impurities and degradation products, which amply illustrates the necessity of appropriate controls [52]. The function, as well as toxicity, of commonly used nanomaterials in aqueous surroundings are presented in Table 3.

Table 3. The role and toxicity of certain widely utilized nanomaterials in water environments.

No	Nanomaterial	Concentration/Toxicity	Behavior/Remediation	Reference
1	Graphene oxide	Lethal effects for *Artemia salina* have only been observed at high concentrations of graphene oxide (500 ppm), whereas sublethal toxicity (growth obstruction) has been observed at loading levels as low as 1 ppm. This could be caused by the oxidative stress that graphene oxide induces. At a maximum dose of 10 mg/L, pure graphene did not exhibit any harm in the acute toxicity test in *Artemia salina*. A similar toxicity propensity was demonstrated by different nanomaterials in additional crab species.	Adsorb heavy metal ions as well as organic pollutants.	[53], [54]
2	Nano-zerovalent iron	The EC50 of the bacteria against pristine pyrophoric nano zerovalent iron was found to be 0.30–1.10 g/L after six bacteria were tested for nZVI concentration. All of the examined bacteria showed a similar trend in the formation of malondialdehyde, regardless of the variation in EC50 values.	the iron in the aging nanozerovalent iron was reduced to the Fe2+ ion.	[55], [56]
3	Zinc oxide nanoparticle	When exposed to *Hydra magnipapillata* for 96, 72, and 48 hours, the median lethal concentrations of 20 nm zinc oxide nanoparticles were 7.0, 8.70, and 55.30 g/mL, while the median lethal concentrations of 100 nm zinc oxide nanoparticles were 9.90, 14.90, and 262.0 g, respectively.	absorbing many elements, changing their speciation in the media and, consequently, their bioavailability, such as Cd, Ni, Cu, Pb, Hg, Mo, Al, and As.	[57]
4	Nanofiber Material	ROS and oxidative stress can be produced by the leaching of Ag^{+} and Zn^{2+} in electrospun Ag- and ZnO/Ag_2nanofiber systems. Daphnia and other aquatic invertebrates exhibit sublethal reproductive and metabolic deficits with extremely low AgNPs (≈0.0050.02 mg/L). ZnO dissolution prevents microbial activity in sludge, which interferes with the removal of nutrients. In algae, invertebrates, and fish, long-term leaching, particle detachment, and transformation facilitate bioaccumulation and trophic transmission.	Activated carbon, biochar, or functionalized carbon–metal adsorbents can be used to capture released Ag^{+}/AgNPs and Zn^{2+}. Separation is made simple by magnetic FeO-based sorbents. Discharge is decreased by immobilizing nanofibers in membrane or fixed-bed modules and then polishing (adsorption, coagulation-flocculation, or membrane filtering).	[36]

5	Titanium dioxide nanoparticle	Sperm cell speed decreased and superoxide dismutase activity and total glutathione levels increased when the concentration of titanium dioxide nanoparticles reached 10 parts per million.	Resilient absorbability to As(V), As(III), and Cd.	[58], [59]
6	Silver nanoparticle	The biological behavior of *Corbicula fluminea* was inhibited when the concentration of silver nanoparticles was 0.1 ppm. *Corbicula fluminea*'s physiological metabolism was once more inhibited when the concentration of silver nanoparticles reached 2 ppm. *Labeo rohita*, a freshwater fish, was exposed to a 25 mg/L nanoparticle for 21 days. Fish survival was unaffected. showed reduced toxicity. For five and fifteen days, *Prochilodus lineatus* fish were exposed to 2.5 and 25.0 g of L1 nanoparticles. The concentration of muscle protein decreased, the activity of all antioxidant enzymes increased, the liver's ACAP decreased, and the muscle and liver's glycogen content increased.	Zinc and copper ions are among the heavy metal ions in water that silver nanoparticles can absorb. Silver can be extracted from water by using activated carbon as an adsorbent.	[60]–[62]
7	Nickel nanoparticle	*Labeo rohita*, a freshwater fish, was exposed to a 25 mg/L nanoparticle for 21 days. Fish survival was unaffected. Significant growth and hemoglobin reductions were noted. An increase in total protein was seen.	Magnetic hydroxyapatite nanocomposite can be utilized as an adsorbent to remove	[63], [64]
8	Cobalt nanoparticle	*Labeo rohita*, a freshwater fish, was exposed to a 25 mg/L nanoparticle for 21 days. Fish survival was unaffected. An increase in total protein was seen.	Can be removed using zinc oxide nanopowder	[65]
9	Chromium oxide nanoparticle	*Labeo rohita*, a freshwater fish, was exposed to a 25 mg/L nanoparticle for 21 days. Fish mortality was initially caused by Cr3O4 nanoparticles. remarkable decline in hemoglobin and growth. An increase in total protein was seen.	Can be removed using superparamagnetic iron oxide nanoparticles.	[66]

7. Conclusion and Future Perspectives.

Nanotechnology has emerged as a transformative platform for addressing the limitations of conventional wastewater treatment, offering unprecedented capabilities in adsorption, catalytic degradation, membrane selectivity, and targeted removal of contaminants of emerging concern. The reviewed literature demonstrates that nanomaterials-particularly metal-oxide nanoparticles, carbon-based structures, magnetic composites, and advanced frameworks such as MOFs and COFs provide high surface reactivity, tunable physicochemical properties, and superior removal efficiencies for a wide spectrum of pollutants. Nevertheless, the environmental implications associated with nanoparticle release, transformation, and bioaccumulation underscore the need for a balanced and precautionary approach to their deployment.

Despite significant progress, several scientific and technological gaps remain. The long-term fate of engineered nanomaterials in complex aquatic matrices is not yet fully understood, and current toxicity assessments often fail to capture chronic, low-dose, or synergistic effects. Moreover, the scalability of nanotechnology-enabled systems is constrained by production costs, energy requirements, and the absence of standardized protocols for evaluating environmental safety and life-cycle impacts.

Looking forward, future research should prioritize the development of immobilized, recoverable, and biodegradable nanomaterials that minimize ecological risks while maintaining high treatment performance. Advancements in thin-film nanocomposite membranes, bio-based nanostructures, and photocatalysts responsive to visible light are expected to enhance both efficiency and sustainability. Integrating nanotechnology with digital monitoring tools, such as AI-assisted process control and real-time sensing, may further optimize treatment performance and reduce operational burdens.

Equally important is the establishment of comprehensive regulatory frameworks, harmonized characterization standards, and robust life-cycle assessment methodologies to ensure responsible innovation. Collaborative efforts between material scientists, environmental engineers, toxicologists, and policymakers will be essential to translate laboratory-scale advances into safe, economically viable, and environmentally resilient treatment solutions.

In summary, nanotechnology holds substantial promises for shaping the next generation of wastewater treatment systems. Realizing this potential will require a concerted focus on safety, sustainability, and interdisciplinary integration to ensure that technological progress aligns with environmental stewardship and public health protection.

Acknowledgement

The authors thank the reviewers and editors of the Islamic University Journal of Applied Sciences for their constructive comments, which substantially improved this manuscript. Some schematic figures were initially drafted using AI-based graphic tools, subsequently revised by the authors manually to ensure scientific accuracy and consistency with the text. All conceptual content, interpretation, and final figure designs are the responsibility of the authors. No external funding was received for this work.

References

[1] "World health statistics 2026: monitoring health for the SDGs, sustainable development goals." https://www.who.int/publications/i/item/9789240122482 (accessed Sep. 16, 2026).

[2] A. Pal, K. Y. H. Gin, A. Y. C. Lin, and M. Reinhard, "Impacts of emerging organic contaminants on freshwater resources: Review of recent occurrences, sources, fate and effects," *Sci. Total Environ.*, vol. 408, no. 24, pp. 6062–6069, Nov. 2010, doi: 10.1016/j.scitotenv.2010.09.026.

[3] S. T. Khan and A. Malik, "Engineered nanomaterials for water decontamination and purification: From lab to products," *J. Hazard. Mater.*, vol. 363, pp. 295–308, Feb. 2019, doi: 10.1016/j.jhazmat.2018.09.091.

[4] S. J. Burian, T. Walsh, A. J. Kalyanapu, and S. G. Larsen, "Climate Vulnerabilities and Adaptation of Urban Water Infrastructure Systems," *Clim. Vulnerability Underst. Addressing Threat. to Essent. Resour.*, vol. 5, pp. 87–107, Jan. 2013, doi: 10.1016/B978-0-12-384703-4.00509-8.

[5] C. Contado, "Nanomaterials in consumer products: A challenging analytical problem," *Front. Chem.*, vol. 3, no. AUG, p. 154429, Aug. 2015, doi: 10.3389/fchem.2015.00048.

[6] S. K. S. Patel and V. C. Kalia, "Advancements in the Nanobiotechnological Applications," *Indian J. Microbiol. 2021 614*, vol. 61, no. 4, pp. 401–403, Sep. 2021, doi: 10.1007/s12088-021-00979-7.

[7] V. Laxmi *et al.*, "Emerging Field of Nanotechnology in Environment," *Indian J. Microbiol. 2023 633*, vol. 63, no. 3, pp. 244–252, Sep. 2023, doi: 10.1007/s12088-023-01092-7.

[8] S. V. Otari *et al.*, "Copper Ferrite Magnetic Nanoparticles for the Immobilization of Enzyme," *Indian J. Microbiol. 2018 591*, vol. 59, no. 1, pp. 105–108, Nov. 2018, doi: 10.1007/s12088-018-0768-3.

[9] R. K. Ibrahim, M. Hayyan, M. A. AlSaadi, A. Hayyan, and S. Ibrahim, "Environmental application of nanotechnology: air, soil, and water," *Environ. Sci. Pollut. Res. 2016 2314*, vol. 23, no. 14, pp. 13754–13788, Apr. 2016, doi: 10.1007/S11356-016-6457-Z.

[10] M. Kampa and E. Castanas, "Human health effects of air pollution," *Environ. Pollut.*, vol. 151, no. 2, pp. 362–367, Jan. 2008, doi: 10.1016/J.ENVPOL.2007.06.012.

[11] D. Mara, *Domestic wastewater treatment in developing countries*, 1st ed. Taylor and Francis, 2013. doi: 10.4324/9781849771023.

[12] M. L. Sikosana, K. Sikhwivhilu, R. Moutloali, and D. M. Madyira, "Municipal wastewater treatment technologies: A review," *Procedia Manuf.*, vol. 35, pp. 1018–1024, 2019, doi: 10.1016/j.promfg.2019.06.051.

[13] Z. Queen *et al.*, "Corresponding author: Zamathula Queen Sikhakhane Nwokediegwu A review of advanced wastewater treatment technologies: USA vs. Africa," *Int. J. Sci. Res. Arch.*, vol. 2024, no. 01, pp. 333–340, 2024, doi: 10.30574/ijsra.2024.11.1.0071.

[14] X. Qu, P. J. J. Alvarez, and Q. Li, "Applications of nanotechnology in water and wastewater

treatment.," *Water Res.*, vol. 47, no. 12, pp. 3931–3946, Aug. 2013, doi: 10.1016/j.watres.2012.09.058.

[15] T. U. Rahman *et al.*, "The Advancement in Membrane Bioreactor (MBR) Technology toward Sustainable Industrial Wastewater Management," *Membranes (Basel).*, vol. 13, no. 2, p. 181, Feb. 2023, doi: 10.3390/membranes13020181.

[16] X. Qu, J. Brame, Q. Li, and P. J. J. Alvarez, "Nanotechnology for a Safe and Sustainable Water Supply: Enabling Integrated Water Treatment and Reuse," *Acc. Chem. Res.*, vol. 46, no. 3, pp. 834–843, Mar. 2012, doi: 10.1021/ar300029v.

[17] N. A. A. Qasem, R. H. Mohammed, and D. U. Lawal, "Removal of heavy metal ions from wastewater: a comprehensive and critical review," *npj Clean Water 2021 41*, vol. 4, no. 1, pp. 36-, Jul. 2021, doi: 10.1038/s41545-021-00127-0.

[18] B. Wang *et al.*, "Remediation of groundwater pollution by in situ reactive zone: A review," *Process Saf. Environ. Prot.*, vol. 168, pp. 858–871, Dec. 2022, doi: 10.1016/j.psep.2022.10.046.

[19] M. A. Shannon, P. W. Bohn, M. Elimelech, J. G. Georgiadis, B. J. Marĩas, and A. M. Mayes, "Science and technology for water purification in the coming decades," *Nature*, vol. 452, no. 7185, pp. 301–310, Mar. 2008, doi: 10.1038/nature06599.

[20] S. Dabare, S. Rajapaksha, and I. Munaweera, "Sustainable innovation in nanotechnology-based water treatment: aligning climate change adaptation with industrial ecology and CSR goals," *Environ. Sci. Water Res. Technol.*, vol. 11, no. 9, pp. 2100–2124, Aug. 2025, doi: 10.1039/d5ew00557d.

[21] N. H. Tran, M. Reinhard, and K. Y. H. Gin, "Occurrence and fate of emerging contaminants in municipal wastewater treatment plants from different geographical regions-a review," *Water Res.*, vol. 133, pp. 182–207, Apr. 2018, doi: 10.1016/j.watres.2017.12.029.

[22] M. Nazaripour, M. A. M. Reshadi, S. A. Mirbagheri, M. Nazaripour, and A. Bazargan, "Research trends of heavy metal removal from aqueous environments," *J. Environ. Manage.*, vol. 287, p. 112322, Jun. 2021, doi: 10.1016/j.jenvman.2021.112322.

[23] S. Ye *et al.*, "The effects of activated biochar addition on remediation efficiency of co-composting with contaminated wetland soil," *Resour. Conserv. Recycl.*, vol. 140, pp. 278–285, Jan. 2019, doi: 10.1016/j.resconrec.2018.10.004.

[24] S. Ye *et al.*, "Biological technologies for the remediation of co-contaminated soil," *Crit. Rev. Biotechnol.*, vol. 37, no. 8, pp. 1062–1076, Nov. 2017, doi: 10.1080/07388551.2017.1304357.

[25] M. Wang *et al.*, "High performance capacitive deionization electrodes based on ultrathin nitrogen-doped carbon/graphene nano-sandwiches," *Chem. Commun.*, vol. 53, no. 78, pp. 10784–10787, Sep. 2017, doi: 10.1039/c7cc05673g.

[26] Z. Wang *et al.*, "Nanoarchitectured metal–organic framework/polypyrrole hybrids for brackish water desalination using capacitive deionization," *Mater. Horizons*, vol. 6, no. 7, pp. 1433–1437, Aug. 2019, doi: 10.1039/c9mh00306a.

[27] A. Rahimpour, M. Jahanshahi, N. Mortazavian, S. S. Madaeni, and Y. Mansourpanah, "Nanofiltration systems and applications in wastewater treatment: Review article," *Ain Shams Eng. J.*, vol. 9, no. 4, pp. 3077–3092, Dec. 2018, doi: 10.1016/j.apsusc.2009.09.089.

[28] E. Zekić, Ž. Vuković, and I. Halkijević, "Application of nanotechnology in wastewater treatment," *Gradjevinar*, vol. 70, no. 4, pp. 315–323, 2018, doi: 10.14256/JCE.2165.2017.

[29] J. Y. Lim, N. M. Mubarak, E. C. Abdullah, S. Nizamuddin, M. Khalid, and Inamuddin, "Recent trends in the synthesis of graphene and graphene oxide based nanomaterials for removal of heavy metals — A review," *J. Ind. Eng. Chem.*, vol. 66, pp. 29–44, Oct. 2018, doi: 10.1016/j.jiec.2018.05.028.

[30] R. Singh and S. Dutta, "A review on H2 production through photocatalytic reactions using TiO2/TiO2-assisted catalysts," *Fuel*, vol. 220, pp. 607–620, May 2018, doi: 10.1016/j.fuel.2018.02.068.

[31] X. yue Huang, L. Ling, and W. xian Zhang, "Nanoencapsulation of hexavalent chromium with nanoscale zero-valent iron: High resolution chemical mapping of the passivation layer," *J. Environ. Sci.*, vol. 67, pp. 4–13, May 2018, doi: 10.1016/j.jes.2018.01.029.
[32] Kenry and C. T. Lim, "Nanofiber technology: current status and emerging developments," *Prog. Polym. Sci.*, vol. 70, no. 34, pp. 1–17, Jul. 2017, doi: 10.1016/j.progpolymsci.2017.03.002.
[33] S. Rtimi, D. D. Dionysiou, S. C. Pillai, and J. Kiwi, "Advances in catalytic/photocatalytic bacterial inactivation by nano Ag and Cu coated surfaces and medical devices," *Appl. Catal. B Environ.*, vol. 240, pp. 291–318, Jan. 2019, doi: 10.1016/j.apcatb.2018.07.025.
[34] T. E. Oladimeji, M. Oyedemi, M. E. Emetere, O. Agboola, J. B. Adeoye, and O. A. Odunlami, "Review on the impact of heavy metals from industrial wastewater effluent and removal technologies," *Heliyon*, vol. 10, no. 23, p. e40370, Dec. 2024, doi: 10.1016/j.heliyon.2024.e40370.
[35] J. Ayach *et al.*, "Comparing Conventional and Advanced Approaches for Heavy Metal Removal in Wastewater Treatment: An In-Depth Review Emphasizing Filter-Based Strategies," *Polym. 2024, Vol. 16, Page 1959*, vol. 16, no. 14, p. 1959, Jul. 2024, doi: 10.3390/POLYM16141959.
[36] R. O. Raji, A. Izuafa, H. S. Auta, O. A. Oyewole, and O. P. Abioye, "Nanochitosan-based adsorbents for sustainable treatment of palm oil mill effluent," *RSC Adv.*, vol. 16, no. 30, pp. 27339–27356, May 2026, doi: 10.1039/D6RA02194H.
[37] K. Jain, A. S. Patel, V. P. Pardhi, and S. J. S. Flora, "Nanotechnology in Wastewater Management: A New Paradigm Towards Wastewater Treatment," *Mol. 2021, Vol. 26, Page 1797*, vol. 26, no. 6, p. 1797, Mar. 2021, doi: 10.3390/MOLECULES26061797.
[38] E. S. KAZAN KAYA *et al.*, "Application of biological and green nanomaterials in wastewater treatment: techniques for the effective removal of dyes, heavy metals, and organic pollutants," *Turkish J. Biol.*, vol. 49, no. 5, pp. 441–458, Oct. 2025, doi: 10.55730/1300-0152.2760.
[39] H. Pérez, O. J. Q. García, M. A. Amezcua-Allieri, and R. R. Vázquez, "Nanotechnology as an efficient and effective alternative for wastewater treatment: an overview," *Water Sci. Technol.*, vol. 87, no. 12, pp. 2971–3001, Jun. 2023, doi: 10.2166/WST.2023.179.
[40] N. O. Sanjeev, M. S. Vallabha, and R. R. L. Rabi, "Nanotechnology-based approaches for the removal of microplastics from wastewater: a comprehensive review," *Beilstein J. Nanotechnol. 16114*, vol. 16, no. 1, pp. 1607–1632, Sep. 2025, doi: 10.3762/BJNANO.16.114.
[41] A. M. Naguib, S. A. Abdel-Gawad, and A. S. Mahmoud, "Reduction of organic contaminants from industrial effluent using the advanced oxidation process, chemical coagulation, and green nanotechnology," *Sci. Reports 2024 141*, vol. 14, no. 1, pp. 15221-, Jul. 2024, doi: 10.1038/s41598-024-65162-6.
[42] S. Malik *et al.*, "Exploring Microbial-Based Green Nanobiotechnology for Wastewater Remediation: A Sustainable Strategy," *Nanomater. 2022, Vol. 12, Page 4187*, vol. 12, no. 23, p. 4187, Nov. 2022, doi: 10.3390/NANO12234187.
[43] S. A. Bhat *et al.*, "Sustainable nanotechnology based wastewater treatment strategies: achievements, challenges and future perspectives," *Chemosphere*, vol. 288, p. 132606, Feb. 2022, doi: 10.1016/J.CHEMOSPHERE.2021.132606.
[44] D. Jiang *et al.*, "Remediation of contaminated soils by enhanced nanoscale zero valent iron," *Environ. Res.*, vol. 163, pp. 217–227, May 2018, doi: 10.1016/j.envres.2018.01.030.
[45] G. F. Schirinzi, I. Pérez-Pomeda, J. Sanchís, C. Rossini, M. Farré, and D. Barceló, "Cytotoxic effects of commonly used nanomaterials and microplastics on cerebral and epithelial human cells," *Environ. Res.*, vol. 159, pp. 579–587, Nov. 2017, doi: 10.1016/j.envres.2017.08.043.
[46] V. Adam, A. Caballero-Guzman, and B. Nowack, "Considering the forms of released engineered nanomaterials in probabilistic material flow analysis," *Environ. Pollut.*, vol. 243, pp. 17–27, Dec. 2018, doi: 10.1016/j.envpol.2018.07.108.

[47] M. Chen, X. Qin, and G. Zeng, "Biodegradation of Carbon Nanotubes, Graphene, and Their Derivatives," *Trends Biotechnol.*, vol. 35, no. 9, pp. 836–846, Sep. 2017, doi: 10.1016/j.tibtech.2016.12.001.
[48] S. A. Love, M. A. Maurer-Jones, J. W. Thompson, Y. S. Lin, and C. L. Haynes, "Assessing nanoparticle toxicity," *Annu. Rev. Anal. Chem.*, vol. 5, no. Volume 5, 2012, pp. 181–205, Jul. 2012, doi: 10.1146/annurev-anchem-062011-143134.
[49] F. von der Kammer *et al.*, "Analysis of engineered nanomaterials in complex matrices (environment and biota): General considerations and conceptual case studies," *Environ. Toxicol. Chem.*, vol. 31, no. 1, pp. 32–49, Jan. 2012, doi: 10.1002/etc.723.
[50] Y. Ma *et al.*, "Effects of rare earth oxide nanoparticles on root elongation of plants," *Chemosphere*, vol. 78, no. 3, pp. 273–279, Jan. 2010, doi: 10.1016/j.chemosphere.2009.10.050.
[51] D. Lin and B. Xing, "Phytotoxicity of nanoparticles: Inhibition of seed germination and root growth," *Environ. Pollut.*, vol. 150, no. 2, pp. 243–250, Nov. 2007, doi: 10.1016/j.envpol.2007.01.016.
[52] E. Oberdörster, S. Zhu, T. M. Blickley, P. McClellan-Green, and M. L. Haasch, "Ecotoxicology of carbon-based engineered nanoparticles: Effects of fullerene (C60) on aquatic organisms," *Carbon N. Y.*, vol. 44, no. 6, pp. 1112–1120, May 2006, doi: 10.1016/j.carbon.2005.11.008.
[53] C. Pretti *et al.*, "Ecotoxicity of pristine graphene to marine organisms," *Ecotoxicol. Environ. Saf.*, vol. 101, no. 1, pp. 138–145, Mar. 2014, doi: 10.1016/j.ecoenv.2013.11.008.
[54] L. De Marchi, C. Pretti, B. Gabriel, P. A. A. P. Marques, R. Freitas, and V. Neto, "An overview of graphene materials: Properties, applications and toxicity on aquatic environments," *Sci. Total Environ.*, vol. 631–632, pp. 1440–1456, Aug. 2018, doi: 10.1016/j.scitotenv.2018.03.132.
[55] J. Semerád *et al.*, "Novel assay for the toxicity evaluation of nanoscale zero-valent iron and derived nanomaterials based on lipid peroxidation in bacterial species," *Chemosphere*, vol. 213, pp. 568–577, Dec. 2018, doi: 10.1016/j.chemosphere.2018.09.029.
[56] Y. Xie *et al.*, "The interactions between nanoscale zero-valent iron and microbes in the subsurface environment: A review," *J. Hazard. Mater.*, vol. 321, pp. 390–407, Jan. 2017, doi: 10.1016/j.jhazmat.2016.09.028.
[57] G. Vale, K. Mehennaoui, S. Cambier, G. Libralato, S. Jomini, and R. F. Domingos, "Manufactured nanoparticles in the aquatic environment-biochemical responses on freshwater organisms: A critical overview," *Aquat. Toxicol.*, vol. 170, pp. 162–174, Jan. 2016, doi: 10.1016/j.aquatox.2015.11.019.
[58] M. E. Özgür *et al.*, "The in vitro toxicity analysis of titanium dioxide (TiO2) nanoparticles on kinematics and biochemical quality of rainbow trout sperm cells," *Environ. Toxicol. Pharmacol.*, vol. 62, pp. 11–19, Sep. 2018, doi: 10.1016/j.etap.2018.06.002.
[59] Z. Bian, J. Zhu, and H. Li, "Solvothermal alcoholysis synthesis of hierarchical TiO2 with enhanced activity in environmental and energy photocatalysis," *J. Photochem. Photobiol. C Photochem. Rev.*, vol. 28, pp. 72–86, Sep. 2016, doi: 10.1016/j.jphotochemrev.2016.06.002.
[60] W. Liu *et al.*, "Toxicity effects of silver nanoparticles on the freshwater bivalve Corbicula fluminea," *J. Environ. Chem. Eng.*, vol. 6, no. 4, pp. 4236–4244, Aug. 2018, doi: 10.1016/j.jece.2018.06.032.
[61] A. Syafiuddin, S. Salmiati, T. Hadibarata, M. R. Salim, A. B. H. Kueh, and S. Suhartono, "Removal of Silver Nanoparticles from Water Environment: Experimental, Mathematical Formulation, and Cost Analysis," *Water, Air, Soil Pollut. 2019 2305*, vol. 230, no. 5, pp. 102-, Apr. 2019, doi: 10.1007/s11270-019-4143-8.
[62] A. Ale, A. S. Rossi, C. Bacchetta, S. Gervasio, F. R. de la Torre, and J. Cazenave, "Integrative assessment of silver nanoparticles toxicity in Prochilodus lineatus fish," *Ecol. Indic.*, vol. 93,

pp. 1190–1198, Oct. 2018, doi: 10.1016/j.ecolind.2018.06.023.

[63] Z. Kanwal *et al.*, "A Comparative Assessment of Nanotoxicity Induced by Metal (Silver, Nickel) and Metal Oxide (Cobalt, Chromium) Nanoparticles in Labeo rohita," *Nanomater. 2019, Vol. 9,* vol. 9, no. 2, Feb. 2019, doi: 10.3390/nano9020309.

[64] D. N. Thanh, P. Novák, J. Vejpravova, H. N. Vu, J. Lederer, and T. Munshi, "Removal of copper and nickel from water using nanocomposite of magnetic hydroxyapatite nanorods," *J. Magn. Magn. Mater.*, vol. 456, pp. 451–460, Jun. 2018, doi: 10.1016/j.jmmm.2017.11.064.

[65] A. T. Le, S. Y. Pung, S. Sreekantan, A. Matsuda, and D. P. Huynh, "Mechanisms of removal of heavy metal ions by ZnO particles," *Heliyon*, vol. 5, no. 4, p. e01440, Apr. 2019, doi: 10.1016/j.heliyon.2019.e01440.

[66] T. Burks, A. Uheida, M. Saleemi, M. Eita, M. S. Toprak, and M. Muhammed, "Removal of Chromium(VI) Using Surface Modified Superparamagnetic Iron Oxide Nanoparticles," *Sep. Sci. Technol.*, vol. 48, no. 8, pp. 1243–1251, Apr. 2013, doi: 10.1080/01496395.2012.734364.

Islamic University Journal of Applied Sciences (IUJAS)
https://journals.iu.edu.sa/jesc

Volume VIII, Issue II, December 2026, Pages 236-260

Physics-Based Modeling of Hybrid Solar–Nuclear Energy Systems: A Strategic Pathway for Sustainable Power in Saudi Arabia

O.S. Ahmed[1] ,*

[1] Department of Physics, College of Science, Qassim University, Buraydah 51452, Saudi Arabia

* Corresponding author. *Email Address*: O.abdeldaaem@qu.edu.sa

Abstract

As part of global efforts to reduce carbon emissions and diversify energy sources, the focus has shifted towards generating electricity from solar and nuclear power. This is a significant change in Saudi Arabia's energy landscape that is consistent with Vision 2030's environmental sustainability objectives. The objective of this research is to use physical modeling for hybrid solar-nuclear energy systems for sustainable . The results of the simulation using FORTRAN program showed that solar irradiance, cell temperature, and production efficiency have a dynamic connection. Power and irradiance are directly correlated also found to be directly correlated. On the other hand, efficiency and irradiance were shown to be inversely correlated. The results showed that nuclear power is the system's steady source since the amount of electrical power produced rises dramatically as the nuclear fission rate grows. The nuclear system is perfect for satisfying baseloads . The temporal integration between the two sources also reveals that the maximum contribution of solar power (125 MW) combined with nuclear stability (24 MW) allows the system capacity to reach 149 MW during peak hours (12 noon).

Keywords: Mathematical modeling; Electrical energy; Nuclear energy; Nuclear fission; Hybrid Solar–Nuclear Energy.

https://doi.org/10.63070/jesc.2026.042
Received 19 April 2026; Revised 27 August 2026; Accepted 05 September 2026; Available online 15 September 2026.
Published by Islamic University of Madinah on behalf of ***Islamic University Journal of Applied Sciences.***
This is an open access article published under the Creative Commons Attribution 4.0 International license (CC BY 4.0):https://creativecommons.org/licenses/by/4.0/

1- Introduction

The sector of energy has been an exceptional transformation during the past few decades, which can be attributed to factors such as growing concern for climate change, rapid depletion of fossil fuel reserves, and increasing penetration of renewable energy resources [1-2]. At the same time, the escalating frequency of anthropogenic outages and the energy crisis arising from socio-political events have emphasized the need for improving the national energy security and resilience. In light of this, de-carbonization and diversification are the future paths chosen by the world's energy sector, and this is achieved by utilizing energy resources such as solar and nuclear energy that do not produce greenhouse gas emissions [3-5]. The latter, on the other hand, has led to a conceptual change in the energy sector [1]. Saudi Arabia must achieve the goals of Vision 2030, to achieve economic and environmental sustainability by diversifying energy sources and promoting a knowledge-based economy. In recent years, investments in renewable energy have garnered increasing global attention, driven by the need to achieve sustainable development policies encompassing economic, social, and environmental dimensions [4-5] . There is a strong relationship between energy consumption and development, as evidenced during the Industrial Revolution when production technology shifted from human labor to machinery, leading to increased use of coal to generate steam. Numerous studies highlight a significantly positive relationship between global growth and energy demand [3-4]. Future energy demand is expected to rise due to continued population and economic growth. Saudi Arabia's energy consumption is the highest in the Middle East, with an average oil consumption of approximately 4.3 million barrels per day (bpd) by the end of 2019, compared to an average of 4.1 million bpd in 2018 [6] . By the end of 2018, Saudi Arabia consumed around 289.9 terawatt-hours (TWh) of energy, representing a year-on-year increase of 0.42 percent compared to 2017. Due to high non-oil economic growth, population fertility, and desalination efforts, domestic energy consumption in Saudi Arabia remains exceptionally high The transition to renewable energy sources is a strategic imperative for Saudi Arabia as it seeks to achieve the goals of Vision 2030 [1-5], which emphasizes economic diversification and environmental sustainability. The motivation behind this study employs a physical approach based on modeling and current applications to analyze and develop ways for producing electricity from nuclear and solar energy [4-5]. It also examines how energy is now used in Arab nations, with Saudi Arabia receiving special attention because of its abundant natural resources and strategic potential in this area [4] . The present work developing an integrated mathematical and physical models that combines solar and nuclear systems to evaluate operational performance, efficiency, and power distribution over time, in line with Saudi Vision .

2- Growth of solar energy

Particularly in the European Union, solar energy has experienced remarkable growth driven by investments,technological advancements, and the global transition toward renewable energy sources (RES). According to the Ember Global Electricity Review report, solar power continues to expand rapidly worldwide [3-4]. At the present rate of growth, 593 GW of solar panels were installed in 2024, a 29% increase over 2023. This demonstrates the industry's continued expansion, even after a record 87% rise in 2023. About 292 GW of additional solar power had been put into service by the end of July 2024. More installations were made in 2024 than in 2023 in 11 of the 15 nations for which Ember monitors monthly capacity statistics [4]. From southern Europe to Central Asia, areas with high levels of insolation might be powered by solar energy, which is accessible practically anywhere the sun shines. However, it is fundamentally limited by its reliance on the time of day, season, and climate. Because of these characteristics, solar energy is less dependable than other energy sources, such nuclear power, figures (1,2) [4-5].

3- Nuclear power: reliability and efficiency

Despite growing competition from renewable energy sources, nuclear power is still a vital part of the world's energy balance. The International Energy Agency (IEA) predicts that in 2025 [3], the world's power production from about 420 nuclear reactors will set a new high. Currently, 10% of the world's electricity is produced via nuclear power. Currently, 63 reactors with a combined capacity of more than 70 GW are being built, which is one of the greatest numbers since 1990. More than 60 reactors, or roughly 15% of the world's nuclear fleet, have had their operational lifespans extended over the last five years. The international initiative Net Zero Nuclear, launched in 2023 at the World Climate Conference (COP28), aims to achieve climate neutrality, improve energy security, and support the expansion of renewable energy sources by tripling the world's nuclear capacity by 2050 [3]. Since 2020, annual nuclear energy investments—which include building new reactors and extending the life of existing ones—have risen by 50% to more than $60 billion. High stability and energy output are provided by nuclear power facilities. Unlike solar power plants, which have a capacity factor of 15–30%, nuclear power plants have an average capacity factor of 90–95%, enabling them to run continuously. The capacity factor is more than just a technical number; it is an important measure of productivity, profitability, and dependability that is used to manage energy balances and choose future technologies, figure (3) [3].

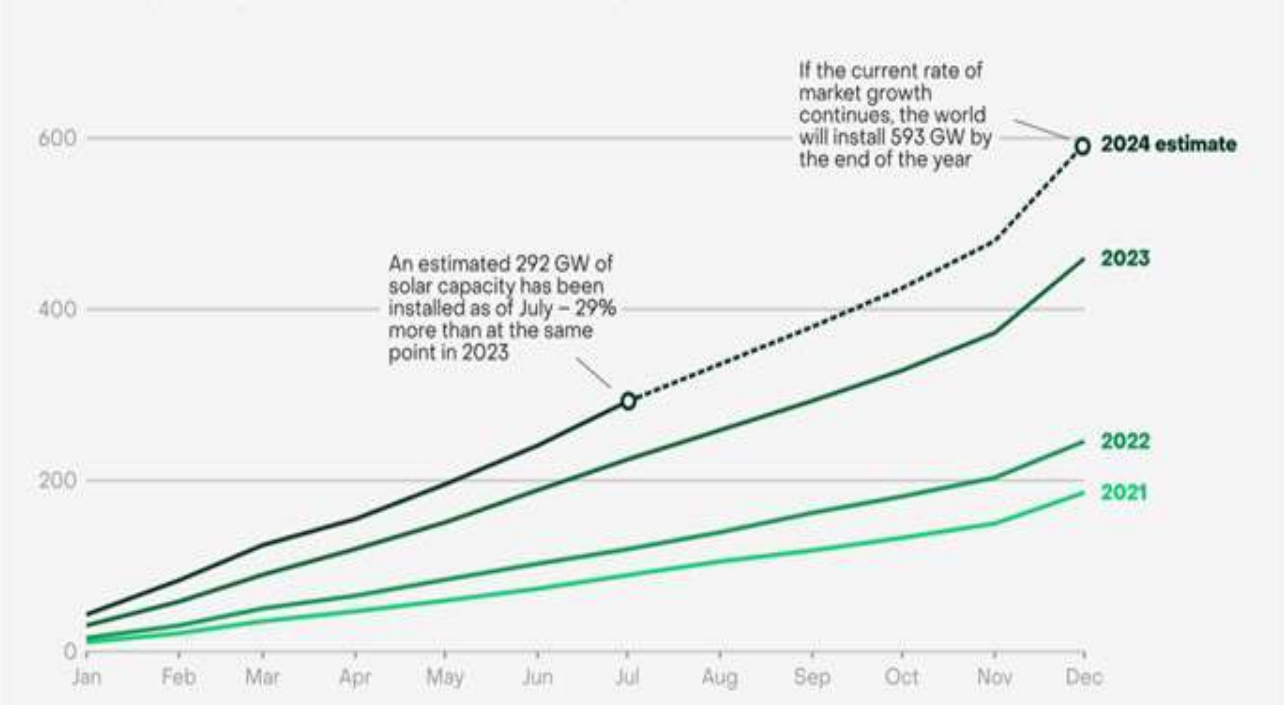


Fig.1. Global solar capacity growth year-to-date by month (GW) [3]

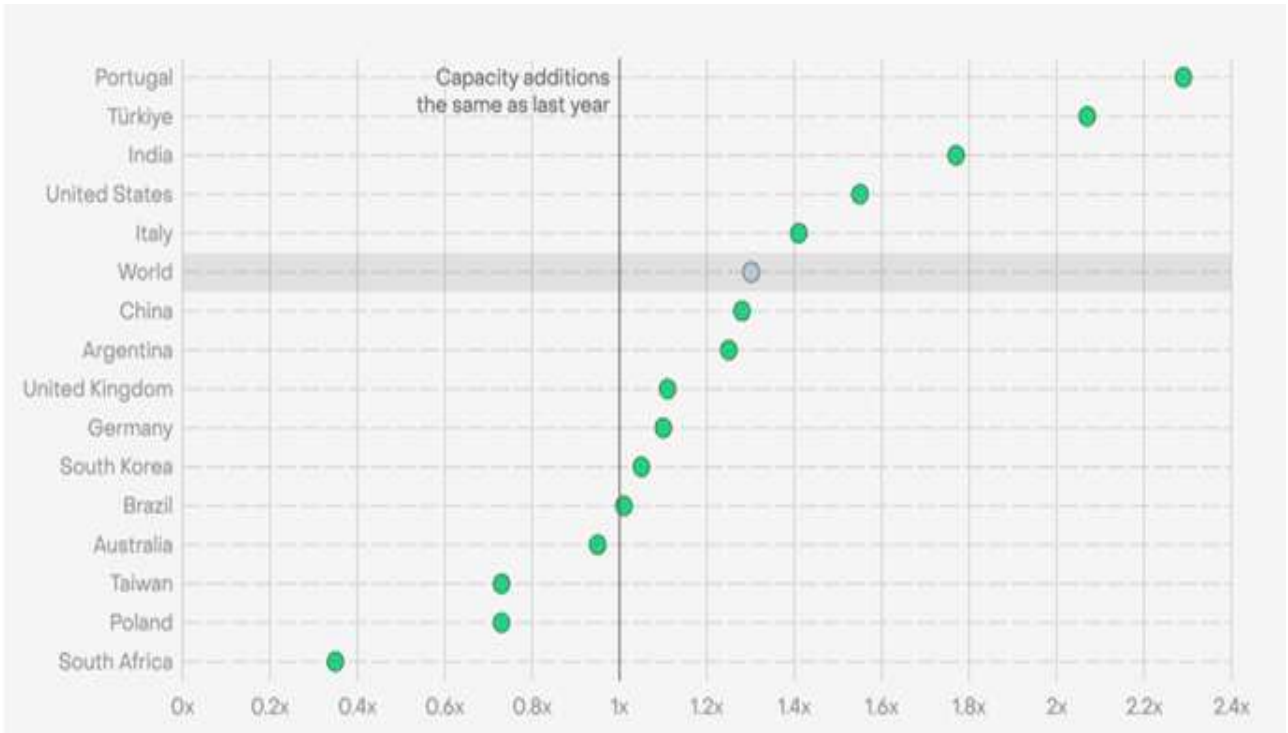


Fig.2. Ratio of production capacity growth in January-July 2024 to the same period in 2023 [3]

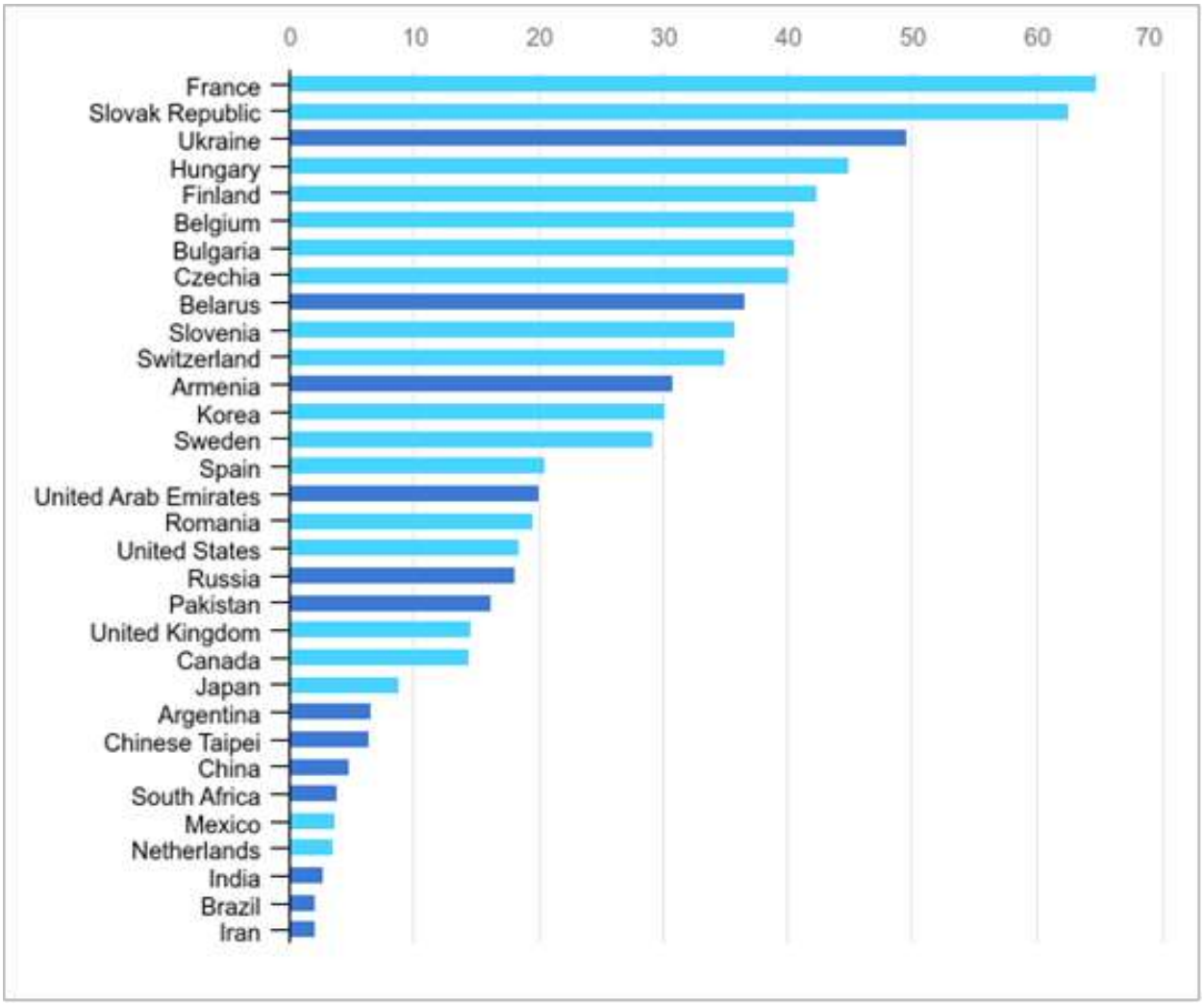


Fig.3. Share of nuclear power in total electricity generation by country, 2023 (IEA) [3]

4- Energy in Saudi Arabia

The Kingdom leverages its abundant sunshine ,benefiting from its vast desert expanses to develop large-scale solar energy projects , therefore Saudi Arabia has placed renewable energy at the core of its Vision 2030, aiming to diversify its economy and reduce its reliance on oil. Saudi Arabia's installed renewable energy capacity reached 2.8 gigawatts, with ongoing projects pushing this figure beyond 8 gigawatts , By the end of 2023, [4-5]. The Ministry of Energy targets generating 50% of the Kingdom's electricity from renewable sources by 2030, supported by significant expansions in solar and wind projects, investments in energy storage technology, and smart grids [6-8] . According to the Saudi General Authority for Statistics (Renewable Energy Statistics 2022), renewable projects are expected to generate 43,698 gigawatt-hours annually, supply electricity to 2.6 million homes, and reduce carbon dioxide emissions by approximately 24.8 million tons each year., is illustrated in Figure 4 and Figure 5 [9-12].

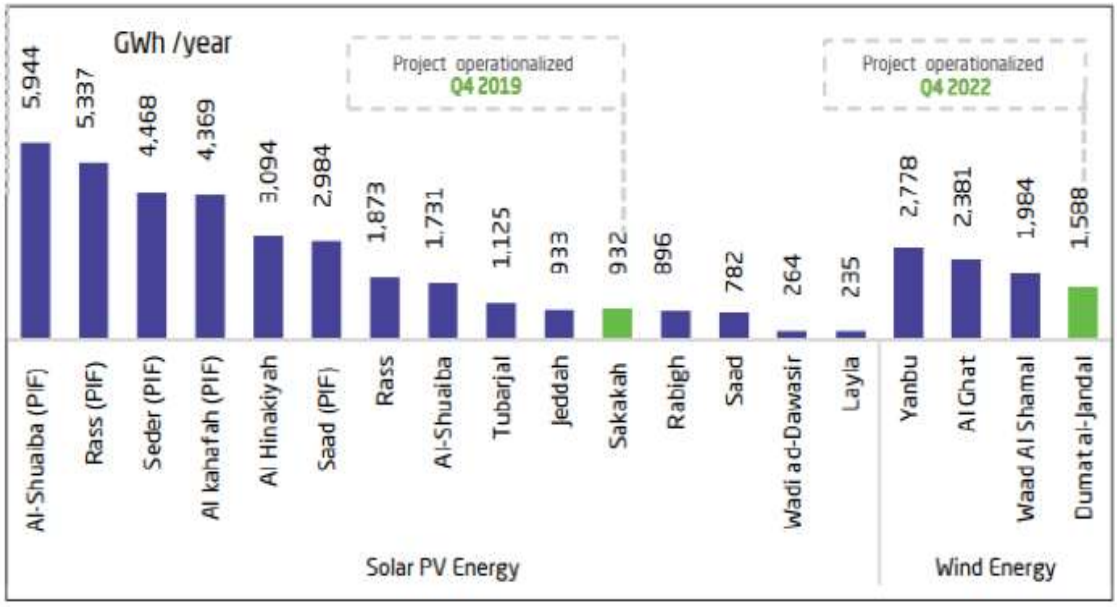


Fig.4. Electricity targeted to be generated from renewable energy projects annually. [Reference: Saudi General Authority for Statistics. (2022). Renewable Energy Statistics 2022 [12]

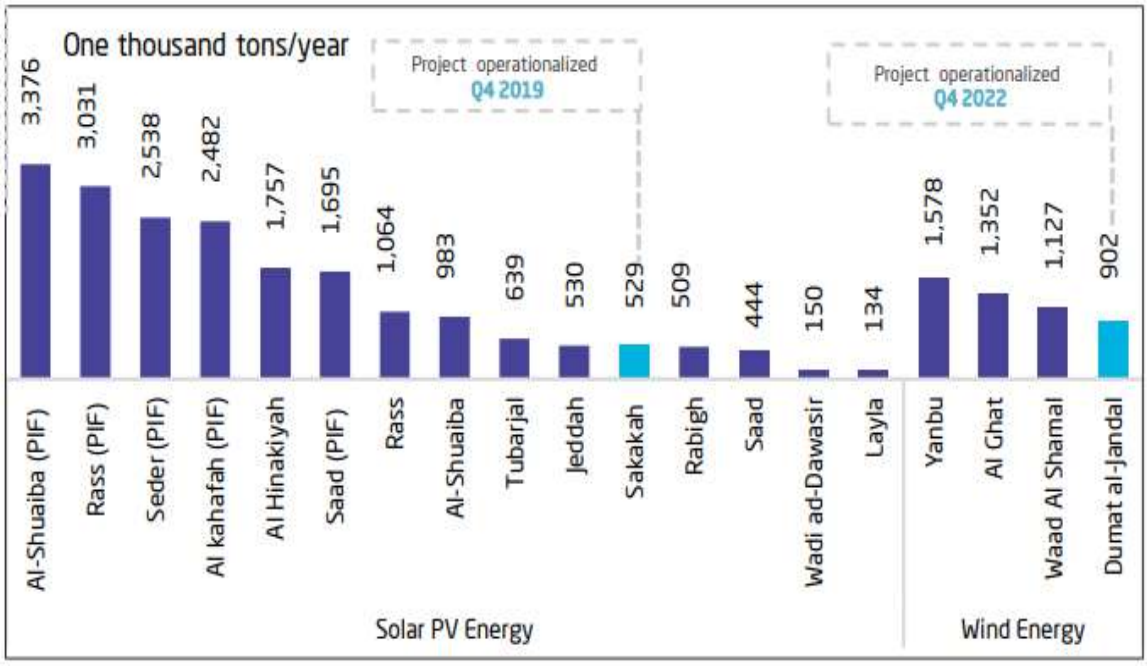


Fig.5. Expected environmental impact of projects on reducing carbon dioxide emissions. [Reference: Saudi General Authority for Statistics. (2022). Renewable Energy Statistics 2022[12]

5- Government Initiatives and Incentives

Under Vision 2030, Saudi Arabia is dedicated to changing its energy system and developing local capacity. To this end, it is providing a number of policy-related incentives to increase trade, draw in foreign direct investment, construct technology infrastructure, promote local employment, and strengthen local capabilities [10-11]. For instance, the Saudi Export Development Authority and the Saudi Export-Import Bank have established assistance systems for enterprises to get export finance. The Function of the Ministry of Investment by creating cross-governmental sectoral ecosystems and assisting businesses with their investment procedures, the Ministry of Investment is aggressively promoting opportunities in the green energy industry [12]. Saudi companies are building the groundwork for Saudi-Chinese cooperation by progressively integrating themselves into regional supply chains. For example, in 2011, Desert Technologies built the first solar PV manufacturing facility in Saudi Arabia, capable of producing 300 megawatts annually. Since then, the company has begun exporting its goods all over the world [13-15]

6- Energy Projects

6-1 KSA Energy (Solar)

Saudi Arabia's installed solar capacity has than doubled from prior years to 6.15 GW as of 2024. As part of its Vision 2030 plan to diversify its energy mix and lessen dependency on fossil fuels, Saudi Arabia is rapidly increasing its solar energy capacity [12- 13]. By 2030, the Kingdom hopes to produce 58.7 GW of solar energy, which will help it reach its target of obtaining 50% of its electricity from renewable sources, as given in Table.1 [28].

Table.1. Major Solar Projects in KSA [28]

Project Name	Location	Capacity (MW)	Key Features	Investment	Operator / Developer
Sakaka Solar Power Plant	Al Jouf	300	1.2 million solar panels over 6 km^2; powers 55,000 homes; reduces 530,000 tons of CO_2 annually	Over $292 million	Ministry of Energy under Custodian of the Two Holy Mosques Initiative

Shuaiba 1 Solar Project	Jeddah	600	Lowest energy purchase cost in KSA (~$0.0104/kWh); improves energy efficiency	2.37 billion USD	ACWA Power (35.01%), Badeel (34.99%), SAPCO (30%)
2024 Solar Expansion	Various (5 sites)	3,751 (combined)	Five new plants; part of 15 GW solar & wind expansion by 2027–2028	18.3 billion SAR (~$5.3 billion)	ACWA Power

6-2 KSA Energy (Nuclear)

Nuclear energy is an important part of Saudi Arabia's future energy mix, even though it now makes up 0% of the country's electricity production. As part of its Vision 2030 plan to diversify its energy mix, lessen its dependency on fossil fuels, and create a sustainable, knowledge-based economy, Saudi Arabia is actively expanding its nuclear energy sector [28]. In order to promote sustainability, energy security, and economic diversification, the Saudi National Atomic Energy Project (SNAEP) seeks to deploy peaceful nuclear energy, Table (2) [16-20] [28].

Table. 2. Key Nuclear Projects and Initiatives.[28]

Project / Initiative	**Purpose / Focus**	**Key Features**	**Status / Partners**
Low Power Research Reactor (LPRR)	Research, training, and development	First nuclear reactor in KSA; inaugurated in 2018; supports Vision 2030 by building local expertise	Developed by KACST & KA-CARE
Saudi National Atomic Energy Project	Electricity generation, human capital development, localization of technology, safety regulation	Focuses on building nuclear power plants; developing workforce; establishing regulatory frameworks	National initiative under Vision 2030
International Collaboration	Compliance and global partnerships	Works with IAEA; hosting the 2025 International Conference on Nuclear and Radiological Emergency Preparedness	Partnership with IAEA; global engagement

7- Challenges of Solar Energy in Saudi Arabia

Early solar development was hindered by unclear governance and fragmented responsibilities among ministries in Institutional Fragmentation (Pre-2016). The 2016 restructuring helped unify oversight under a single ministry. Under the skilled labor shortage, there is a lack of trained professionals in solar technology, which limits the pace of deployment and maintenance [28]. Under the limited local manufacturing capacity, Saudi Arabia's solar ambitions require high local content, but current domestic manufacturing capacity for PV components like float glass and aluminum is insufficient. In grid integration and storage, integrating intermittent solar power into the national grid requires advanced infrastructure and energy storage solutions .In environmental and land use constraints, large-scale solar farms may face land use conflicts and environmental concerns, especially in ecologically sensitive areas. [21-23] [28].

8- Challenges of Nuclear Energy in Saudi Arabia

In light of the field of regulation and safety, Saudi Arabia works to develop robust regulatory bodies and safety protocols to meet international standards for nuclear energy deployment. In terms of human capital and skills, the rate of development and safe operation are constrained by the lack of qualified professionals and nuclear engineers [28]. In addition to the infrastructure and investment requirements—long-term, expensive investments and complex infrastructure for nuclear power plants can be a barrier to rapid deployment—public perception and societal acceptance of nuclear energy also face skepticism and concerns about safety, waste management, and environmental risks. International compliance and oversight complicate matters and necessitate diplomatic cooperation with Saudi Arabia in order to match their plans with international treaties and IAEA protocols [24-28].

9- Methodology

9-1- Generating electricity from solar cell

A solar cell, often known as a photovoltaic (PV) cell, is a non mechanical device that directly generates energy from sunshine. Certain photovoltaic cells are capable of producing power from artificial light. Photons, or solar energy particles, make up sunlight. The varied wavelengths of the solar spectrum are represented by the different quantities of energy contained in these photons. Semiconductor material makes up a photovoltaic cell. Photons that hit a photovoltaic cell will either be absorbed by the semiconductor material, pass through the cell, or reflect off the cell. The only photons that can produce electricity are those that are absorbed. Electrons are separated from the atoms in the semiconductor

material when it absorbs sufficient solar energy [28-29]. The PV cell's front surface is made more responsive to the dislodged, or free, electrons during the manufacturing process, allowing the electrons to naturally migrate to the cell's surface. There is an electrical charge imbalance between the front and back surfaces of the PV cell due to the migration of electrons, which are all negatively charged, toward the front surface. The voltage potential that results from this imbalance is comparable to the positive and negative terminals of a battery. Electrons are absorbed by the PV cell's electrical conductors .Electricity flows through an electrical circuit when its conductors are connected to an external load, such a battery , as shown in figure 6 [29].

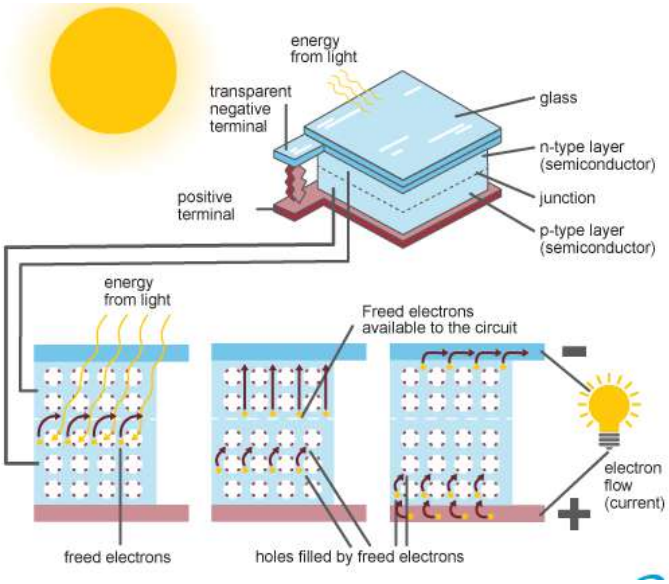


Fig.6. Generating electricity from solar cell [29]

9-2- Generating electricity from nuclear power

Nuclear fission, which is based on the energy produced when heavy atoms, such as ^{235}U , split, is how nuclear power plants produce electricity [30]. It happens when an atom is struck by a neutron, splitting it and starting a chain reaction that releases other neutrons. Heat is produced as a significant amount of energy by this atom-to-atom multiplication. After cooling, this heat produces steam, which powers turbines that are connected to alternators to produce electricity. Nuclear power plants operate through the following stages: [30-31]:

- The main circuit: tiny enriched uranium pellets are inserted into metal tubes called rods at the reactor's core. The primary circuit is formed by submerging these tubes in a water tank. The water in this circuit heated to 320°C due to the fission of uranium atoms. To keep the water liquid, pressure is applied to it. [30]

- The secondary circuit uses a steam generator to heat the water in the secondary circuit using hot water from the primary circuit. Steam turbines, which run an alternator to produce energy, are powered by the steam generated. [31]

- Voltage boosting: a transformer is used to raise the voltage of the resultant electric current to between 225,000 and 400,000 volts. This makes it possible to move electricity over the power system more effectively. [30-31]

- The cooling circuit converts steam from the secondary circuit back into water by cooling it. This is accomplished by means of a condenser, which can be supplied with air-cooled towers or cold water drawn from rivers or the sea. [30-31]



Fig.7. Generating electricity from nuclear power [31]

9-2-1- The phenomenon of nuclear fission

When a neutron strikes an unstable, heavy atom, fission takes place. The atom splits into two lighter atoms as a result of the collision.

This process is referred to as a chain reaction because it produces a significant and steady amount of energy and continuously triggers further atom fissions [51].

^{235}U is the fuel used in this procedure. Its physical characteristics and composition make it ideal for the fission process. When a neutron strikes the atomic nucleus of ^{235}U, it can split in two , figure (8). Fission products including Ba, Kr, Sr, Cs, I, and Xe with atomic weights ranging from 95 to 135 are

produced by the fission reaction in ^{235}U. Typical reaction products include the following examples [30-31] [51]:

$$^{235}U + n \Longrightarrow {}^{144}Ba + {}^{90}Kr + 2n + 200\ MeV \quad (1)$$

$$^{235}U + n \Longrightarrow {}^{141}Ba + {}^{92}Kr + 3n + 170\ MeV \quad (2)$$

$$^{235}U + n \Longrightarrow {}^{94}Zr + {}^{139}Te + 3n + 197\ MeV \quad (3)$$

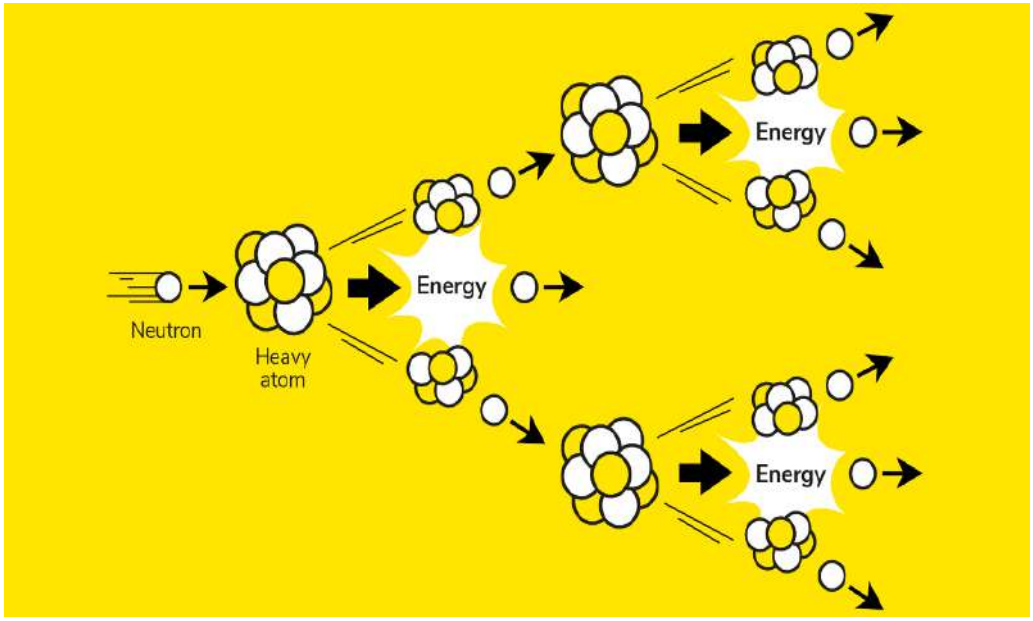


Fig. 8. show phenomenon of nuclear fission[31]

9-3 Mathematical Model

In order to investigate and assess the performance of solar and nuclear power generating systems in a dynamic hybrid system, this study relies on the development of an integrated physical-mathematical modeling framework. The three primary stages of the process are individual system modeling, numerical simulation, and hybrid system analysis [32-39] [41-50].

9-3-1- Solar PV Modeling

A model based on energy balance and thermoelectric relationships was used to describe the performance of photovoltaic cells [41-50]. The output power was computed as a function of temperature, loss coefficients, and radiation intensity. The NOCT model was used to model the cell temperature as a function of ambient temperature and sun irradiation, enabling a dynamic depiction of

the thermal effect. To get a realistic performance estimate, reference efficiency coefficients and system losses (such as wiring and inverter losses) were also considered [44]. Solar irradiance fluctuations were taken into account as a time-varying input to model system attitude throughout the day, and standard values for the coefficients were assumed based on literature [45] , The crystalline silicon (C-Si) photovoltaic technology was considered in the current simulation due to its extensive commercial use and typical temperature coefficient value (β≈0.004/◦C), which is in line with values documented in the literature [34] , and reference efficiency (ηref≈18%) are consistent with typical values reported for commercial crystalline silicon photovoltaic modules [34] .

$$P_{in} = A\ G \quad (4)$$

$$T_c = T_{amb} + \left(\frac{NOCT - 20}{800}\right)\ G \quad (5)$$

$$\eta = \eta_{ref}\ \left[1 - \beta\left(T_c - T_{ref}\right)\right] \quad (6)$$

$$P_{solar} = \eta\ A\ G\ \eta_{inv}\ \eta_{los} \quad (7)$$

Where

- Pin: Input solar power (W).
- Psolar: Electrical power output from the PV system (W).
- A: Solar panel area (m²).
- G: Solar irradiance (W/m²).
- Tc: Cell temperature (°C).
- Tamb: Ambient temperature (°C).
- Tref: Reference temperature (25 °C).
- η: Actual PV efficiency.
- ηref: Reference efficiency.
- β: Temperature coefficient (≈ 0.004/°C).
- ηinv: Inverter efficiency (≈ 0.95).
- ηloss: System efficiency after losses (≈ 0.85–0.9).
- NOCT: Nominal Operating Cell Temperature (~45°C).

9-3-2- Nuclear Energy Modeling

The physics of nuclear fission is used to describe energy generation in nuclear reactors , A full-scale commercial nuclear power plant (500–1600 MW) is not meant to be represented by the reactor taken into consideration in this simulation. Rather, the modeled system is equivalent to a modular or small-scale research reactor design with a continuous heat output of roughly 24 MW. This scale was specifically chosen to assess the suggested hybrid system's thermodynamic behavior and energy-management capabilities under regulated operation circumstances. Therefore, rather than being a commercial utility-scale reactor, the chosen reactor capacity is suitable for modeling purposes and should be understood as a representative small nuclear energy source. The power generated is correlated with the rate of fission and the energy released from each reaction, including the effectiveness of thermal and electrical conversion [46]. According to the model, the nuclear system runs in a semi-steady state, and its high capacity factor indicates continuous operation. Based on established operating data, standard values for the energy generated by fission (~200 MeV per fission) and conversion efficiencies were also applied [41-50].

$$P_{nuclear} = \eta_{th} \times \eta_{el} \times N_f \times E_f \times \mathrm{CF} \ldots\ldots\ldots\ldots\ldots\ldots\ldots (8)$$

Where

η_{th} :- Thermal efficiency (heat to steam conversion) (~0.33–0.37)

η_{el} :- Electrical efficiency (turbine and generator) (~0.33–0.36)

Nf:- Number of fissions per second

Ef: - Energy released from the fission of a single nucleus (~ 200 Mev (3.2×10^{-11} J))

CF: Capacity Factor (≈ 0.9)

9-3-3 -Hybrid System Modeling and Model of overall system efficiency

A time-dependent model was developed to integrate the outputs of the solar and nuclear systems, where the dynamic load distribution is thought to affect the overall system capacity [41-45],The interaction between the various sources is reflected in this model, where [46-48]:

•One variable source that varies over time is solar energy.

• A steady base load is represented by nuclear energy.

• To increase system stability, backup and storage variables were included [49].

$$P_{total}(t) = P_{solar}(t) + P_{nuclear}(t) - P_{storage}(t) + P_{\text{backup}}(t) \text{.....(9)}$$

- In order to concentrate on the direct relationship between solar and nuclear power generation, energy storage and backup terms were ignored in this simplified simulation framework, therefore the equation (9) become equation (10) :-

$$P_{total}(t) = P_{solar}(t) + P_{nuclear}(t) \text{.....(10)}$$

$$\eta_{ref} = \frac{p_{total}}{p_{solar} + p_{nuclear}} \text{.....(11)}$$

Where

P_{solar} (t): Variable (day/night)

$P_{nuclear}$ (t): Approximately constant

$P_{storage}$ (t): Energy storage (batteries)

P_{backup} (t): Backup when needed

9-3-4 - Numerical Simulation

The simulation was performed using hypothetical data reflecting realistic operating conditions, where the following were performed:

• The intensity of solar irradiation was was varied within a range of 600–1000W/m².

• The temperature of the cell was computed dynamically. [47]

• The impact of losses and efficiency coefficients on power output was examined.

• The reaction of the system was examined at various times (day and night) [49]

.• A number of measures, such as output power and overall system efficiency, were used to assess system performance. [50]

• Sensitivity analysis was used to assess how important factors like temperature, conversion efficiency, and capacitance coefficient affected the system's overall performance[50].

It should be mentioned that the performance of a single typical solar module with an effective area of one meter is represented by the results shown in Table (3). As a result, the output power values between (77 and 121 W) show unit-area performance in different irradiance scenarios. On the other hand, a utility-scale photovoltaic installation included into the hybrid energy system is represented by the solar contribution shown in Table (5). The single-module output was scaled to a total PV array area typical of large commercial solar farms in order to acquire the megawatt-scale solar power numbers.

10- Results and Discussion

Evaluation of the performance of the photovoltaic solar energy system The photovoltaic system's reaction to the surrounding environmental factors is depicted in the simulation results shown in Table (3) and Figures (9, 10, 11) (Figure 9): The extracted electrical power and solar radiation intensity was found to be linearly correlated; at a peak irradiance of 1000 W/m², the power increased from 77.18 W to 121.36 W [48]. The enhanced photon current produced within the semiconductor material is physically explained by the higher flux of incident photons [43]. The thermal effect (Figure 10): The cell registered 53.75 °C at 600 W/m² and 66.25 °C at 1000 W/m². The results showed a consistent rise in cell temperature with increasing irradiance intensity. The reason for this increase is that part of the solar energy is absorbed by the cell and transformed into heat rather than electricity. Conversion efficiency (Figure 11): The investigation revealed that when the temperature rose, the cell efficiency decreased from 0.159 to 0.150, or around 6% of its total value [44]. This is explained scientifically by the fact that an increase in temperature causes the open-circuit voltage to drop and the charge recombination rate to rise, both of which have a detrimental effect on the photovoltaic conversion's overall efficiency [54]. Analysis of the Performance of the Nuclear Power System: The relationship between the fission rate and the generated power was observed in Table (4) and Figure (12): Energy Density and Production Stability: The findings demonstrate a direct correlation between the nuclear fission rate per second and the electrical power produced [47]. The generated power was 10.6 MW at a fission rate of 3.00 E+18 and 31.7 MW at a rate of 9.00 E+18The tremendous energy density of uranium, which yields about 200 MeV every fission cycle, is responsible for this considerable rise [49]. This system is perfect for effectively covering baseloads with constant operational stability because of its independence from climatic circumstances. Table (5) and Figure (13), which show the combined 24-hour operational performance of the two systems and represent their dynamic integration, provide an analysis of the integrated solar-nuclear hybrid system's performance: Daytime Peak Period: With 125 MW from the solar system and 24 MW from the nuclear system, the system reached its maximum overall capacity of 149 MW at 12:00 PM. Here, solar energy plays a crucial role in supplying the substantial rise in demand during the day. Evening and Night Periods [50]: The nuclear system's function as a safety valve for the hybrid system is amply illustrated; at 9:00 PM and 6:00 AM, solar production completely stops (0 MW), while the nuclear system continues to produce a steady 24 MW of power [50], As a result, the study's conclusions are generally in line with recent developments in the scientific literature [52-56]. Note that the simulation results show instantaneous power production under particular operating conditions, whereas the published national solar capacity relates

to installed capacity [13] . And a small modular/research reactor with an approximate thermal power output of 24 MW serves as the basis for the simulation parameters [57].

Table (3): Simulation Results using hypothetical data of Solar energy model

Parameters	Symbol	Unit	Value/Range
Fixed Parameters			
PV area	(A)	m^2	1
Reference temperature	T_{ref}	°C	25
NOCT	NOCT	°C	45
Reference efficiency	η_{ref}	%	18
Inverter efficiency	η_{inv}	–	0.95
System-loss factor	η_{loss}	–	0.85
Temperature coefficient	β	$°C^{-1}$	0.004
Ambient temperature	T_{amb}	°C	35
Variable Conditions			
Solar irradiance	(G)	W/m^2	600–1000
Calculated Results			
Cell temperature	Tc	°C	53.75- 66.25
Actual efficiency	η	–	0.159-0.150
Output power	P_{out}	W	77.18-121.36

Table.4. Simulation Results based on hypothetical data of Nuclear energy model

Parameter	Symbol	Unit	Value/Range
Fixed Parameters			
Thermal Efficiency	η_{th}	-	0.36
Electrical Efficiency	η_{el}	-	0.34
Fission Energy	E_f	(J/fission)	(200 Mev) (3.2×10^{-11} J)
Capacity Factor	CF	-	0.9
Variable Conditions			
Fission Rate	Nf	(fissions/s)	3.00E+18 – 9.00E+18
Calculated Results			
Output power	P_{out}	W	1.06E+07- 3.17E+07

Table.5. Simulation Results using hypothetical data of Hybrid Model

Time (h)	G	Solar Power (MW)	Nuclear Power (MW)	Total Power (MW)
06:00	0	0	24	24
09:00	400	50.73	24	74.72
12:00	1000	125	24	149
15:00	900	110.87	24	134.86
18:00	200	24.27	24	48.27
21:00	0	0	24	24

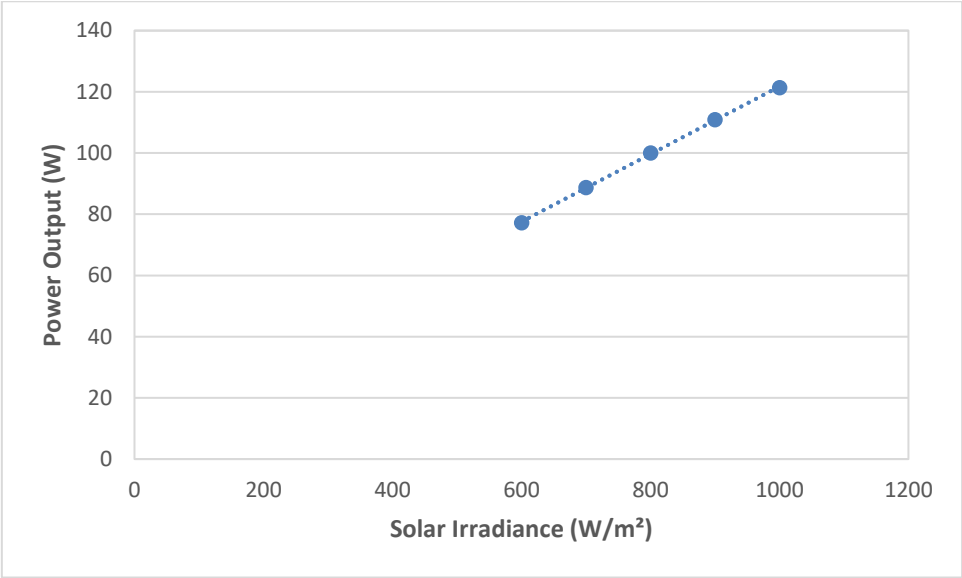


Fig. 9. Explains Solar Irradiance relation with Power Output

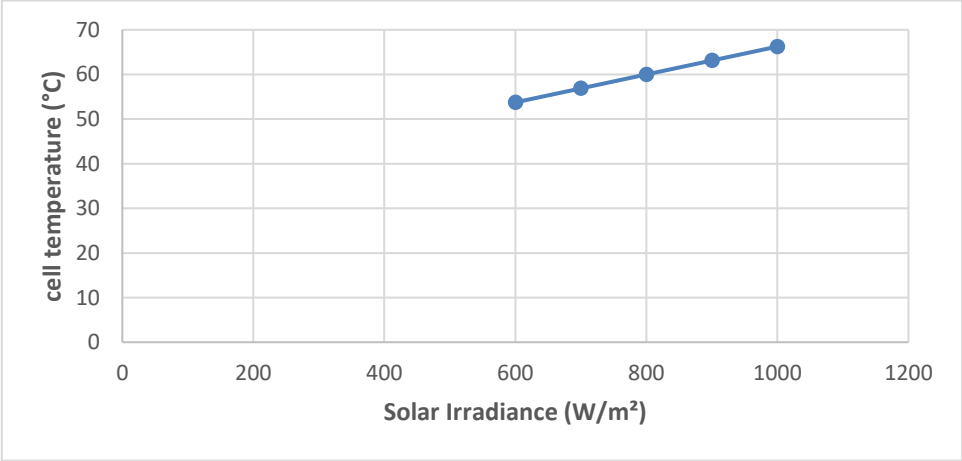


Fig. 10. Shows relation Solar Irradiance with Cell temperature

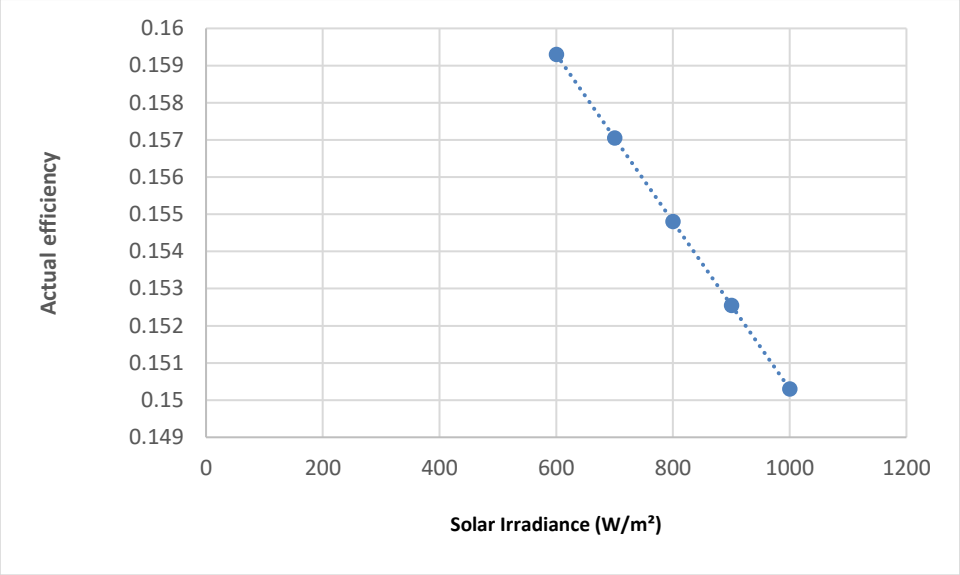


Fig.11. Shows the relation between Solar Irradiance and Actual **efficiency**

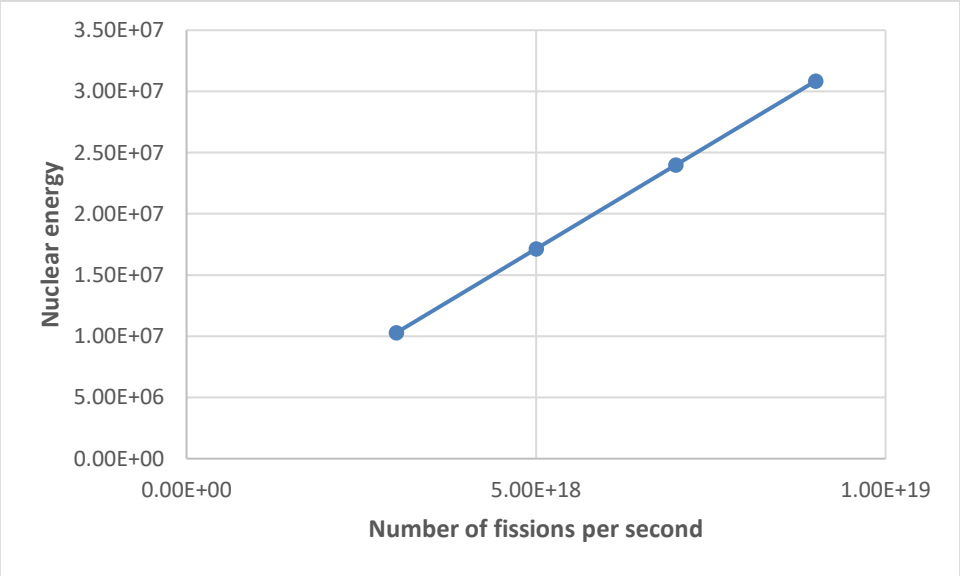


Fig. 12. Explains the relation between Number of fissions per second and Nuclear energy

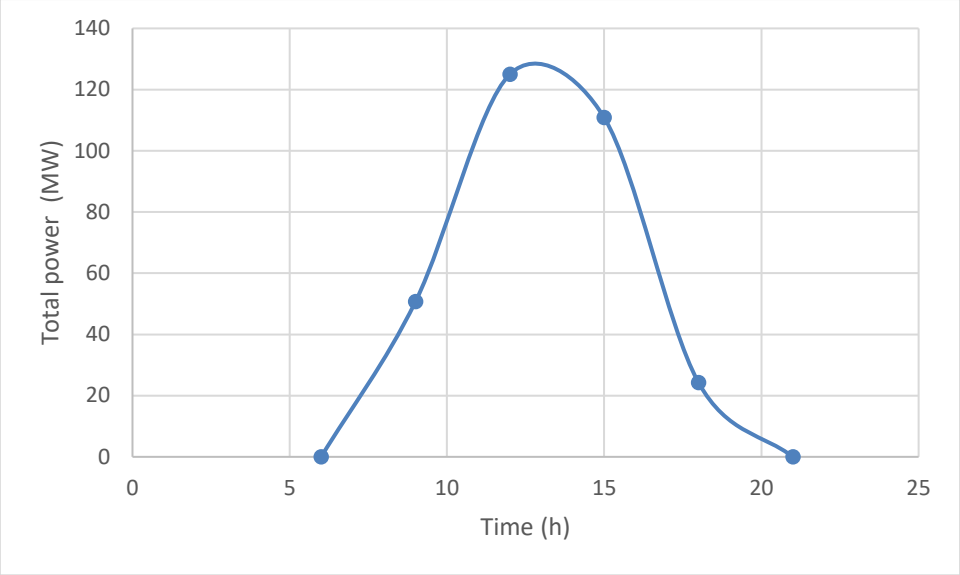


Fig. 13. Shows the relation time with total power

11. Conclusion

The results of this study explain that integrating solar and nuclear energy within a hybrid system offers an efficient and stable solution to meet the growing electricity demand in Saudi Arabia, while simultaneously improving grid reliability and reducing carbon emissions. Simulations showed that relying on a balanced mix of solar and nuclear energy achieves greater operational stability compared to relying on a single energy source.The analysis indicates that the ideal solar-to-nuclear power ratio for Saudi Arabia's climate is about between 60:40 and 70:30. While nuclear energy offers consistent base capacity to ensure the continuity of electricity supply during times of low solar irradiation or increasing demand, solar energy helps cover daytime loads and reduce fuel consumption. Also it is indicate that integrating these hybrid systems requires developing smart grid infrastructure, enhancing energy storage technologies, and improving load management strategies to ensure the stability of the national electricity grid. This integration can mitigate the volatility of renewable energy and improve the long-term operational efficiency of the grid. This study corresponds with the goals of Saudi Vision 2030 by supporting the diversification of energy sources, reducing reliance on fossil fuels, and increasing the contribution of clean energy to the national energy mix. Accordingly, the study recommends the following:

- Gradual expansion of solar and small modular nuclear power (SMR) projects.
- Investment in energy storage technologies and smart grids.
- Support for research and development programs in hybrid energy systems.
- Development of policies and regulations that encourage investment in green energy projects.

- Enhancing international cooperation in the fields of peaceful nuclear technology and renewable energy.

References

[1]Roshni Anna Jacob ,et al " Modeling and control of nuclear–renewable integrated energy systems: Dynamic system model for green electricity and hydrogen production" J. Renewable Sustainable Energy 15, 046302 (2023)

[2]Shannon M. Bragg-Sitton.et al" Reimagining future energy systems: Overview of the US program to maximize energy utilization via integrated nuclear-renewable energy systems" international journal of energy research ,2020.

[3]Aibope Davletiyarova "Solar and nuclear energy: Distributed generation for a sustainable future " Atomic Industry Development Association,2025

[4]Al-Ghamdi, N., Bahaj, A. S., & James, P. (2022). Supply chain readiness for solar PV expansion in Saudi Arabia. Energies, 15(20), 7479. https://doi.org/10.3390/en15207479

[5]Al-Ismail, F. S., Alam, M. S., Shafiullah, M., Hossain, M. I., & Rahman, S. M. (2023). Impacts of renewable energy generation on greenhouse-gas emissions in Saudi Arabia: A comprehensive review. Sustainability, 15, 5069. https://doi.org/10.3390/su15095069 Mohammed AL-Sadig M. AL-Haj" Economic and Environmental Benefits of Energy Transition in Saudi Arabia: A Case Study of Sakaka Solar Power Plant"Journal of Economic Studies, Vol. 17 (2), pp. 211-226,2024

[6]Saudi Ministry of Energy. (n.d.). Renewable energy projects. https://www.moenergy.gov.sa/ar/Projects

[7]Saudi Vision 2030. (2016). Vision 2030 overview. https://www.vision2030.gov.sa/en/vision-2030/overview/

[8]Saudi Vision 2030. (n.d.). Sakaka solar-power plant. https://www.vision2030.gov.sa/ar/projects/sakaka/

[9]Ministry of Energy, Saudi Arabia. (2020). Annual energy report.

[10] Saudi-China Collaboration on Renewable Energy Supply Chains. (2023). KAPSARC. https://www.kapsarc.org/wp-content/uploads/2023/03/KS-2022-WB13-Saudi-China-Collaboration-on-Renewable-Energy-Supply-Chains.pdf

[11] Saudi Arabia Localization Objectives. (2023). https://www.ief.org/_resources/files/events/third-ief-eu-energy-day/turki-al-shehri-24.02-repdo--ief_riyadh_v2-2.pdf

[12] Saudi General Authority for Statistics. (2022). Renewable energy statistics 2022.

[13] Saudi Ministry of Energy. (2023). Energy transition challenges. https://moenergy.gov.sa/en/DigitalDocuments/Documents/Energy_Transition_Challenges.pdf

[14] Abdulkareem, A., & Ellaboudy, A. (2025, February 1). Saudi Arabia's Vision 2030's Renewable Energy Project Initiatives. Climate Scorecard. Retrieved from https://www.climatescorecard.org/2025/02/saudi-arabias-vision-2030s-renewable-energy-project-initiatives/

[15] Ventures Onsite. (2024). Upcoming oil and gas projects in Saudi Arabia. https://www.venturesonsite.com/content-hub/upcoming-oil-and-gas-projects-in-saudi-arabia

[16] Stone, N. (2025, August 1). Saudi Arabia's energy transition: From oil to gas and its implications for energy investors. AInvest. https://www.ainvest.com/news/saudi-arabia-energy-transition-oil-gas-implications-energy-investors-2508/

[17] Vision 2030. (2025). Sakaka Solar Power Plant. Government of Saudi Arabia. https://www.vision2030.gov.sa/en/explore/projects/sakaka-solar-power-plant

[18] Vassileva, A. (2025, July 15). Saudi Arabia more than doubles solar capacity to 6.1 GW in 2024. Renewables Now. https://renewablesnow.com/news/saudi-arabia-more-than-doubles-solar-capacity-to-6-1-gw-in-2024-1278483/

[19] Vision 2030. (2025). Dumat Al Jandal Wind Farm. Government of Saudi Arabia. https://www.vision2030.gov.sa/en/explore/explore-more/dumat-aljandal

[20] World Nuclear Association. (2024). Nuclear Power in Saudi Arabia. https://www.world-nuclear.org/information-library/country-profiles/countries-o-s/saudi-arabia

[21] Vision 2030. (2025). The Low Power Research Reactor. Government of Saudi Arabia. https://www.vision2030.gov.sa/en/explore/projects/the-low-power-research-reactor

[22] Yamada, M. (2016). Vision 2030 and the Birth of Saudi Solar Energy. Middle East Institute. https://www.mei.edu/sites/default/files/publications/PF15_Yamada_Saudisolar_web.pdf

[23] Alghamdi, N., James, P., & Bahaj, A. (2025). Vision and Reality: An Assessment of Saudi Arabia's In-Country Capacity to Deliver on Its Solar Ambitions. Sustainability, 17(13), 5721. https://www.mdpi.com/2071-1050/17/13/5721

[24] Alkinani, S., Alghazi, O., Asfour, R. K., & Brahimi, T. (2022). Wind Energy in Saudi Arabia: Opportunities and Challenges. Effat University. https://ieomsociety.org/proceedings/2022nigeria/565.pdf

[25] Alruwaished, A., Alshehri, S., Mukhrish, A., Shams, A., & Al-Athel, K. (2025). Nuclear Fission Energy Initiatives in the Kingdom of Saudi Arabia. Arabian Journal for Science and Engineering, 50, 3099–3111. https://link.springer.com/article/10.1007/s13369-023-08105-z

[26] Todreas, N. E., & Kazimi, M. S. (2021). *Nuclear Systems Volume I: Thermal Hydraulic Fundamentals* (3rd ed.). CRC Press.

[27] Yildiz, B. (2024). Saudi Arabia's Nuclear Energy Program: Challenges and Concerns in the International Legal Framework. Çanakkale Onsekiz Mart University Journal of Political Science. http://ybd.dergi.comu.edu.tr/dosyalar/Ybd/9-saudi-arabia-s-nuclear-energy-program-challenges-and-concerns-in-the-us-s.pdf

[28] AbdelHamid M. S. Esmail" Transitioning the Energy Mix in Saudi Arabia: Pathways Toward Sustainability and Vision 2030" Studies on Energy Applications International Journal 1, Issue 01 (2025) 42-58

[29] Duffie, J. A., & Beckman, W. A. (2013). *Solar Engineering of Thermal Processes* (4th ed.). John Wiley & Sons.

[30] John R. Lamarsh و Anthony J. Baratta" Introduction to Nuclear Engineering"4th Edition (2022–2023)

[31] https://www.orano.group/en/unpacking-nuclear/how-a-nuclear-energy-plant-works

[32] Guido Carlo Masotti, Antonio Cammi, Stefano Lorenzi, Marco Enrico Ricotti" Modeling and simulation of nuclear hybrid energy systems architectures" Energy Conversion and Management, Volume 298, 15 December 2023.

[33] Messenger, R. A., & Ventre, J. (2010). Photovoltaic Systems Engineering (3rd ed.). CRC Press.

[34] Skoplaki, E., & Palyvos, J. A. (2009). On the temperature dependence of photovoltaic module electrical performance: A review of efficiency/power correlations. *Solar Energy*, 83(5), 614–624.

[35] National Renewable Energy Laboratory (NREL). (2016). *PV Cell Temperature Model*.

[36] Lamarsh, J. R., & Baratta, A. J. (2001). *Introduction to Nuclear Engineering* (3rd ed.). Prentice Hall.

[37] Glasstone, S., & Sesonske, A. (1994). *Nuclear Reactor Engineering* (4th ed.). Springer.

[38] International Renewable Energy Agency (IRENA). (2019). *Hybrid Power Systems: Opportunities and Challenges*.

[39] Rezk, H., Tyukhov, I., Raupov, A., & Al-Dhaifallah, M. (2019). Optimal design of hybrid energy systems. *Journal of Cleaner Production*, 231, 1170–1184.

[40] Ahn, Y., & Cho, S. (2020). Hybrid renewable and nuclear energy systems for sustainable energy development. *Energy Conversion and Management*, 223, 113299

[41] Gilman, P., Dobos, A., DiOrio, N., Freeman, J., Janzou, S., Ryberg, D., 2018, "SAM Photovoltaic Model Technical Reference Update", National Renewable Energy Laboratory Report NREL/TP-6A20-67399.

[42] Vat Sun,et al" A new method for evaluating nominal operating cell temperature (NOCT) of unglazed photovoltaic thermal module" Energy Reports,2020

[43] Takele Ferede Agajie,et al " Optimal Design and Mathematical Modeling of Hybrid Solar PV–Biogas Generator with Energy Storage Power Generation System in Multi-Objective Function Cases"*Sustainability* 2023, *15*(10), 8264;

[44] Haratyk G, Shirvan K, Kazimi M. Compact steam generator for nuclear application. In: Proceedings of NURETH16. 2015.

[45] Krarti, M., & Dubey, K. (2018). Review analysis of solar electricity generation in Saudi Arabia. *Renewable and Sustainable Energy Reviews*, 91, 878–891

[46] Abdin Z., Webb C., Gray E.M" Modelling and simulation of an alkaline electrolyser cell Energy", 138 (2017)

[47] Sharmila deve Venkatachalam,et al " Performance and suitability analysis of rooftop solar PV in Oman: A case study of university branches" Heliyon . 2025 Feb 11;11(4).

[48] Duffie, J.A. and Beckman, W.A. (2013) Solar Engineering of Thermal Processes. 2nd Edition, Madison, New York; John Wiley & Sons, Hoboken. https://doi.org/10.1002/9781118671603.

[49] https://en.wikipedia.org/wiki/Nuclear_reactor_physics

[50] https://phys.libretexts.org/Bookshelves/University_Physics/University_Physics_(OpenStax)/University_Physics_III__Optics_and_Modern_Physics_(OpenStax)/10%3A__Nuclear_Physics/10.06%3A_Fission

[51] Nuclear–Renewable Hybrid Energy Systems, International Atomic Energy Agency, 2023

[52] Samira M. Nassar,et al " Optimal planning of integrated nuclear-hybrid renewable energy systems for electrical distribution networks based on artificial intelligence " Scientific Report,2025.

[53] Siddharth Suman" Hybrid nuclear-renewable energy systems: A review" Journal of Cleaner Production,2018

[54] Yifan Wang,et al " Design and integrated performance estimate of a solar-nuclear hybrid energy power generation system " Nuclear Engineering and Design ,2025

[55] Sulaiman Alyahya,et al" Analysis from the New Solar Radiation Atlas for Saudi Arabia" June 2016 ,Solar Energy 130(6):116-127.

[56] Md Arman Arefin ,et al " A Comprehensive Review of Nuclear-Renewable Hybrid Energy Systems: Status, Operation, Configuration, Benefit, and Feasibility" Frontiers in Sustainable Cities , 2021.

[57] Liou, J. (2021). What are Small Modular Reactors (SMRs)? *International Atomic Energy* Agency (IAEA). Retrieved from https://www.iaea.org/newscenter/news/what-are-small-modular-reactors-smrs

Islamic University Journal of Applied Sciences (IUJAS)
https://journals.iu.edu.sa/jesc

Volume VIII, Issue II, December 2026, Pages 261-280

A Comparative Theoretical Study on Gamma-Ray Shielding Parameters of Organic NLO Compounds within the 0.1 MeV–20 MeV Energy Range Utilizing Multiple Modeling and Simulation Tools

M.A. Ahlam[1]*,

[1] Department of Physics, College of Science, Qassim University, Buraydah 51452, Saudi Arabia
*Corresponding author: (M.A. Ahlam), *Email Address:* **ahl.ali@qu.edu.sa**

Abstract

In this study, a systematic theoretical and computational evaluation is performed to investigate the radiation-shielding efficiency of two organic non-linear optical (NLO) compounds: 3-[1-(2-phenylhydrazinylidene)ethyl]-2H-chromen-2-one ($C_{17}H_{14}N_2O_2$) and 4-nitrophenyl hydrazone ($C_{13}H_{11}N_3O_2$). The investigation was conducted across a broad photon energy spectrum (0.1 MeV to 20 MeV) using a multi-platform simulation to ensure the highest degree of precision and reliability. Specifically, the shielding effectiveness was estimated by Monte Carlo simulations in Geant4, and the results were validated against computational tools including EpiXS, XCOM and Phy-X/PSD. Computational modeling and simulation effectively streamline candidate selection for specific shielding applications, serving as a time and cost-efficient alternative to laboratory experiments. Fundamental parameters, including the mass attenuation coefficient (μ_m), total atomic and electronic cross-section ($\sigma_{t,a}, \sigma_{t,el}$), effective atomic number (Z_{eff}), effective electron density (N_{eff}) and molar extinction coefficients (ε) were theoretically determined. The results demonstrate a high degree of agreement between the Monte Carlo- based predictions and the theoretical frameworks. Furthermore, these findings confirm that the investigated NLO compounds are promising, cost-effective candidates for advanced radiation-shielding applications.

Keywords: Organic NLO Material; Geant4 (Monte Carlo Simulation); Radiation Attenuation; Mass Attenuation Coefficient.

https://doi.org/10.63070/jesc.2026.043
Received 19 April 2026; Revised 23 September 2026; Accepted 26 September 2026.
Available online 30 September 2026.
Published by Islamic University of Madinah on behalf of *Islamic University Journal of Applied Sciences.*
This is an open access article published under the Creative Commons Attribution 4.0 International license (CC BY 4.0):https://creativecommons.org/licenses/by/4.0/

1. Introduction

Nonlinear optical (NLO) materials are highly important due to their wide use in electronics and optoelectronics, with prominent applications including optical communications, lasers, optical switching, data storage, and radiation sensing [1-7]. Among them, organic nonlinear optical materials are particularly preferred because their optical behavior originates at the molecular level. This molecular origin allows researchers to easily modify their chemical structures to achieve specific optical properties. Compared to inorganic alternatives, organic materials offer greater structural diversity and greater flexibility in tailoring their chemical structures and nonlinear optical properties. [8]. Furthermore, given the widespread use of ionizing radiation in medical, industrial, and nuclear fields, measuring radiation levels and material absorption is essential. This necessitates the evaluation of specific radiation interaction parameters, such as calculating the absorbed dose via photon transmission equations [5]. To study gamma-ray interactions with matter, a narrow-beam geometry setup based on the Beer–Lambert law is essential for accurately determining linear and mass attenuation coefficients [7]. The mass attenuation coefficient (μ_m) represents a fundamental parameter quantifying the probability of interaction between incident photons and a material's mass per unit area through scattering and absorption processes [9,10]. Calculating mass attenuation coefficients is critical for evaluating fundamental parameters of compounds or mixtures, including atomic and electronic cross-sections($\sigma_{t,a}$, $\sigma_{t,el}$), effective electron densities (N_{eff}), molar extinction coefficients (ε) and effective atomic numbers (Z_{eff}). Such parameters are essential in fields such as medical diagnostics, agricultural engineering, industrial radiography, and computed tomography, as well as in the core principles of nuclear shielding and radiation dosimetry. Numerous studies documented in the literature [11–31] have extensively investigated mass attenuation coefficients and radiation interaction parameters across a broad energy spectrum for a wide array of elements, compounds, and complex mixtures, using diverse theoretical, experimental, and analytical methodologies. Previous computational studies have utilized Monte Carlo simulations to evaluate the mass attenuation coefficients of various polymers across a range of photon energies [14,28]. Among the prominent theoretical developments, the XCOM computational framework was established to provide a robust method for determining attenuation parameters with high precision [32–34]. Subsequently, the transition of this tool to the Windows operating system was introduced by Gerward et al. [35], yielding the widely recognized software WinXCOM. In addition, extensive evaluations of the radiation interaction dynamics in diverse liquid and solid media have been performed [36], relying on the elemental data compiled by Hubbell and Seltzer [32] alongside Chantler [37]. In a related study [38], the energy-dependent behavior of the effective atomic number (Z_{eff}) was examined across common thermoluminescent dosimetric (TLD) materials (LiF, $CaCO_3$ and $CaSO_4$), establishing that (Z_{eff}) varies linearly with incident photon energy. In another investigation [39], empirical measurements of the mass attenuation coefficients for select TLD materials were conducted across various photon energies using a narrow-beam geometry coupled with a high-purity germanium detector. Motivated by the search for advanced alternatives, NLO materials hold significant potential for biological dosimetry applications, largely attributable to their mechanical strength, low mass density, economic

feasibility, and simple preparation methods. Numerous studies have extensively evaluated fundamental radiation interaction parameters across a broad photon energy range for various organic [1,4] and inorganic [6,7,9] nonlinear optical (NLO) materials. Specifically, evaluations of mass attenuation coefficients (μ_m) and effective atomic numbers (Z_{eff}) in organic NLO compounds indicated that their characteristically low (Z_{eff}) values make them highly applicable for radiation dosimetry [5]. Currently, the radiation interaction parameters for organic nonlinear optical (NLO) materials remain insufficiently documented, with no published data available for the two compounds under present study: 3-[1-(2-phenylhydrazinylidene)ethyl]-2H-chromen-2-one ($C_{17}H_{14}N_2O_2$) and 4-nitrophenyl hydrazone ($C_{13}H_{11}N_3O_2$). Therefore, this work conducts a comprehensive evaluation of their fundamental gamma-ray interaction parameters to fill this research gap. In this work, we present a detailed theoretical analysis of these parameters across a wide photon energy range from 0.1 MeV to 20 MeV. The parameters were systematically evaluated using comprehensive computational modeling. To ensure the reliability and precision of the results, a comparative analysis was performed using four internationally recognized computational tools: Geant4, EpiXS, XCOM, and Phy-X/PSD. This study demonstrates that the investigated organic NLO materials offer promising alternatives for biological dosimetry and related radiation applications.

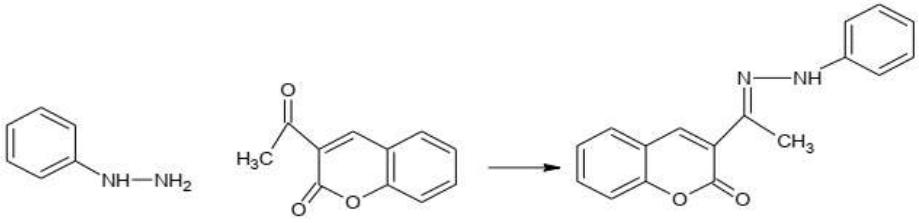
2. Investigated organic NLO compounds

2.1. 3-[1-(2-phenylhydrazinylidene)ethyl]-2H-chromen-2-one ($C_{17}H_{14}N_2O_2$)

This compound, commonly known as coumarin hydrazone, is an organic NLO crystal with a density of 1.30 g/cm³ that exhibits a needle-shaped morphology when grown via the slow evaporation technique. The grown crystal demonstrates high crystallinity and excellent optical transparency across the entire visible region, with an SHG efficiency 3.3 times higher than that of a standard KDP crystal. Furthermore, it maintains thermal stability up to 196°C, establishing it as a promising candidate for high-speed electro-optic and radiation-shielding applications.

Table 1. Atomic and weight fractions of elements in coumarin hydrazone.

Element	Atom Fraction	Weight Fraction
C (Carbon)	0.485714	0.733671
H (Hydrogen)	0.400000	0.050706
N (Nitrogen)	0.057143	0.100658
O (Oxygen)	0.057143	0.114965



Phenylhydrazine + 3-acetyl-2*H*-chromen-2-one → 3-[1-(2-phenylhydrazinylidene) ethyl]-2*H*-chromen-2-one

Fig.1. Schematic representation of coumarin hydrazone crystal synthesis.

2.2. 4-nitrophenyl hydrazone ($C_{13}H_{11}N_3O_2$)

This organic NLO crystal, with a density of 1.428 g/cm³, exhibits a needle-shaped morphology when prepared using the slow evaporation technique. The harvested crystal exhibits a high degree of crystallinity along with optical transparency throughout the entire visible spectrum. Furthermore, it exhibits significant thermal stability up to 146°C and an SHG efficiency 15.39 times higher than that of a standard KDP crystal, establishing it as a superior candidate for advanced nonlinear optical and radiation-shielding devices.

Table 2. Atomic and weight fractions of elements in 4-nitrophenyl hydrazone.

Element	Atom Fraction	Weight Fraction
C (Carbon)	0.448276	0.647225
H (Hydrogen)	0.379310	0.045961
N (Nitrogen)	0.103448	0.174180
O (Oxygen)	0.068966	0.132634

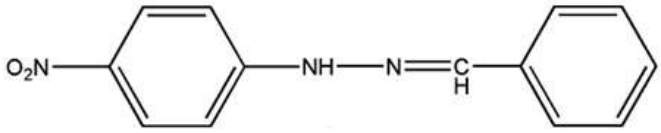


Fig. 2. Chemical structure of 4-nitrophenyl hydrazone $C_{13}H_{11}N_3O_2$ [44]

3. Calculation method

Mass Attenuation Coefficients (μ_m)

The mass attenuation coefficients were obtained for the investigated compounds using the photon transmission method. This fundamental radiation interaction process is described by the following mathematical expression [4,21,22,28]:

$$I = I_o \exp(-\mu_m t) \qquad (1)$$

In this expression, (I) and (I_o) denote the unattenuated and attenuated photon intensities respectively. The parameter (μ_m) represents the mass attenuation coefficient characteristic of the material, while (t) signifies the sample's mass thickness.

The mixture rule is employed to determine the total mass attenuation coefficient for any chemical compound or multi-elemental mixture. This principle calculates the overall attenuation as the weighted sum of the mass attenuation coefficients of its constituent elements, relative to their weight fractions.

$$\mu_m = \sum iW_i(\mu_m)_i \qquad (2)$$

In this context, (W_i) and $(\mu_m)_i$ denote the weight fraction and the mass attenuation coefficient of the i^{th} individual element within the medium, respectively. For chemical compounds, the weight fraction (W_i) is defined by the following mathematical relationship:

$$W_i = \frac{n_i\, A_i}{\sum_j n_j\, A_j} \qquad (3)$$

where (A_i) represents the atomic weight of the i^{th} element, (n_i) denotes the number of formula units, $(\sum_j n_j)$ signifies the total atom count in the molecular formula and (A_j) refers to the molecular weight of the j^{th} constituent element.

Effective Atomic Number (Z_{eff})

The effective atomic number is calculated using the following equation:

$$Z_{eff} = \frac{\sigma_{t,a}}{\sigma_{t,el}} \qquad (4)$$

In this equation, $(\sigma_{t,a})$ and $(\sigma_{t,el})$ denote the total atomic and electronic cross-sections respectively. The total atomic cross-section is obtained from the following equation:

$$\sigma_{t,a} = \frac{1}{N_A} \frac{\mu_m}{\sum_i W_i/A_i} \qquad (5)$$

In this equation, (N_A) denotes Avogadro's number $(N_A = 6.02214 \times 10^{23})$. Furthermore, the total electronic cross-section $(\sigma_{t,el})$ is calculated through the following expression:

$$\sigma_{t,el} = \frac{1}{N_A} \sum_i \frac{f_i A_i}{Z_i} (\mu_m)_i \qquad (6)$$

In this expression, (f_i) denotes the fractional number of atoms for the i^{th} element, while (Z_i) denotes the atomic number of the corresponding element within the chemical compound or mixture.

Effective Electron Densities (N_{eff})

The effective electron density is calculated using the following mathematical expression:

$$N_{eff} = \frac{\mu_m}{\sigma_{t,el}} = \frac{N_A}{M} Z_{eff} \sum_i n_i \qquad (7)$$

In this relation, (M/n) denotes the average atomic weight. Furthermore, $(\sum_i n_i = n)$ represents the total number of atoms in the molecular formula.

Molar Extinction Coefficient $(\boldsymbol{\varepsilon})$

The following expression was used to calculate the molar extinction coefficient:

$$\boldsymbol{\varepsilon} = \mathbf{0}.4343 \times N_A \times \sigma_{t,a} \qquad (8)$$

where, $(\sigma_{t,a})$ represents the total atomic cross-section.

4. Simulation methods

The radiation shielding effectiveness of two organic NLO compounds, $C_{17}H_{14}N_2O_2$ and $C_{13}H_{11}N_3O_2$, was investigated using Geant4 v11.1.1 Monte Carlo simulations over the photon energy range of 0.1 MeV–20 MeV. The simulations employed a monoenergetic, monodirectional pencil beam of 1,000,000 primary gamma photons per energy, directed perpendicular to the target samples. The FTFP_BERT_HP physics base was used with G4EmLivermorePhysics (G4-Liv) as the electromagnetic physics model. A narrow-beam, good-geometry configuration was employed, with

the photon source positioned 30 cm upstream of the sample and directed along the sample axis. The uncollided photon fraction was determined by counting the primary photons reaching the rear boundary of the sample without undergoing a physical interaction. Therefore, no additional downstream detector distance was required. The attenuation of gamma rays was determined from the ratio of the number of transmitted photons in the presence of the sample to the number of photons recorded without the sample [31,40]. A constant mass thickness of 0.15 g/cm^2 was used for all investigated photon energies. The corresponding physical thickness of each sample was calculated using the following equation:

$$t = \frac{0.15}{\rho} \qquad (9)$$

where t is the physical thickness in cm and ρ is the material density in g/cm³. Accordingly, the physical thicknesses were 0.1154 cm for $C_{17}H_{14}N_2O_2$, with a density of 1.300 g/cm^3, and 0.1050 cm for $C_{13}H_{11}N_3O_2$, with a density of 1.428 g/cm^3. For comparison, the calculated radiation-interaction and shielding parameters were compared with those obtained from three established theoretical platforms: EpiXS, based on the EPICS2017 and EPDL97 nuclear data libraries; XCOM, which uses the NIST photon cross-section database; and Phy-X/PSD, which provides photon-interaction data based on standard theoretical databases.

5. Results and discussion

In this work, the mass attenuation coefficients (μ_m) for the investigated organic NLO compounds were determined across a broad energy range (0.1 MeV–20 MeV) using the Geant4 Monte Carlo toolkit. To ensure the reliability of the simulation results, a rigorous cross-verification was performed by comparing the Geant4-derived data with theoretical values obtained from three established platforms: EpiXS, XCOM, and Phy-X/PSD. These computational tools provide comprehensive photon interaction data for diverse chemical compositions. Table 3 summarizes the calculated (μ_m) values for the studied samples and the percentage deviations (%Δ) between the Geant4 results and the three theoretical references, as defined by the following equations [28]:

$$\%\Delta_1 = 100 \times \left| \frac{\mu_{Epixs} - \mu_{G4}}{\mu_{Epixs}} \right| \qquad (10)$$

$$\%\Delta_2 = 100 \times \left| \frac{\mu_{\mathrm{XCOM}} - \mu_{G4}}{\mu_{\mathrm{XCOM}}} \right| \qquad (11)$$

$$\%\Delta_3 = 100 \times \left| \frac{\mu_{\mathrm{Phy-X/PSD}} - \mu_{G4}}{\mu_{\mathrm{Phy-X/PSD}}} \right| \qquad (12)$$

Table 3. Mass attenuation coefficients and corresponding percentage deviations between the Geant4 results and the three theoretical tools for the $C_{17}H_{14}N_2O_2$ and $C_{13}H_{11}N_3O_2$ organic NLO compounds over the selected gamma-ray energy range.

Gamma Energy (MeV)	Mass Attenuation Coefficient (μ_m (cm^2/g))													
	$C_{17}H_{14}N_2O_2$							$C_{13}H_{11}N_3O_2$						
	Geant4	EpiXS	%Δ_1	XCOM	%Δ_2	Phy-X/PSD	%Δ_3	Geant4	EpiXS	%Δ_1	XCOM	%Δ_2	Phy-X/PSD	%Δ_3
0.1	0.159 ± 0.0010	0.159	0.012	0.159	0.031	0.159	0.034	0.159 ± 0.0010	0.159	0.034	0.159	0.017	0.159	0.011

0.15	0.142 ± 0.0010	0.141	0.002	0.142	0.035	0.142	0.043	0.141 ± 0.0010	0.141	0.039	0.141	0.114	0.141	0.087
0.2	0.129 ± 0.0009	0.129	0.009	0.129	0.008	0.129	0.043	0.1285 ± 0.0009	0.129	0.061	0.129	0.086	0.129	0.096
0.3	0.112 ± 0.0009	0.112	0.092	0.112	0.125	0.112	0.129	0.111 ± 0.0009	0.112	0.054	0.112	0.069	0.112	0.092
0.4	0.100 ± 0.0008	0.100	0.069	0.100	0.170	0.100	0.126	0.100 ± 0.0008	0.100	0.046	0.100	0.103	0.100	0.105
0.5	0.091 ± 0.0008	0.091	0.019	0.092	0.097	0.092	0.094	0.091 ± 0.0008	0.091	0.025	0.091	0.104	0.091	0.099
0.6	0.0845 ± 0.0008	0.085	0.130	0.085	0.184	0.085	0.188	0.084 ± 0.0008	0.084	0.102	0.084	0.158	0.084	0.160
0.8	0.074 ± 0.0007	0.074	0.041	0.074	0.105	0.074	0.105	0.074 ± 0.0007	0.074	0.007	0.074	0.067	0.074	0.073
1	0.067 ± 0.0007	0.067	0.011	0.067	0.114	0.067	0.111	0.066 ± 0.0007	0.066	0.042	0.067	0.135	0.067	0.142
1.5	0.054 ± 0.0006	0.054	0.098	0.054	0.250	0.054	0.247	0.054 ± 0.0006	0.054	0.115	0.054	0.262	0.054	0.259
2	0.047 ± 0.0006	0.047	0.022	0.047	0.204	0.047	0.196	0.046 ± 0.0006	0.046	0.033	0.047	0.209	0.046	0.207
3	0.037 ± 0.0005	0.037	0.101	0.037	0.273	0.037	0.284	0.037 ± 0.0005	0.037	0.104	0.037	0.279	0.037	0.288
4	0.032 ± 0.0005	0.032	0.079	0.032	0.310	0.032	0.308	0.032 ± 0.0005	0.032	0.046	0.032	0.286	0.032	0.277
5	0.028 ± 0.0004	0.028	0.058	0.028	0.293	0.028	0.304	0.028 ± 0.0004	0.028	0.027	0.028	0.276	0.028	0.273
6	0.026 ± 0.0004	0.026	0.023	0.026	0.298	0.026	0.281	0.026 ± 0.0004	0.026	0.063	0.026	0.302	0.026	0.321
7	0.024 ± 0.0004	0.024	0.020	0.024	0.296	0.024	0.288	0.024 ± 0.0004	0.024	0.031	0.024	0.355	0.024	0.338
8	0.023 ± 0.0004	0.023	0.031	0.023	0.218	0.023	0.223	0.023 ± 0.0004	0.023	0.035	0.023	0.311	0.023	0.289
9	0.021 ± 0.0004	0.021	0.088	0.021	0.178	0.021	0.179	0.021 ± 0.0004	0.021	0.115	0.021	0.393	0.021	0.381
10	0.020 ± 0.0004	0.020	0.018	0.021	0.225	0.021	0.246	0.020 ± 0.0004	0.020	0.017	0.021	0.264	0.021	0.281
11	0.020 ± 0.0004	0.020	0.034	0.020	0.289	0.020	0.271	0.020 ± 0.0004	0.020	0.011	0.020	0.314	0.020	0.292
12	0.019 ± 0.0004	0.019	0.031	0.019	0.294	0.019	0.297	0.019 ± 0.0004	0.019	0.064	0.019	0.334	0.019	0.329
13	0.019 ± 0.0004	0.019	0.049	0.019	0.275	0.019	0.288	0.019 ± 0.0004	0.019	0.054	0.019	0.296	0.019	0.293
14	0.018 ± 0.0004	0.018	0.044	0.018	0.343	0.018	0.334	0.018 ± 0.0004	0.018	0.025	0.018	0.298	0.018	0.314
15	0.018 ± 0.0003	0.018	0.009	0.018	0.305	0.018	0.286	0.018 ± 0.0003	0.018	0.042	0.018	0.327	0.018	0.316
16	0.017 ± 0.0003	0.017	0.083	0.017	0.179	0.017	0.172	0.017 ± 0.0003	0.017	0.037	0.017	0.310	0.017	0.293
18	0.017 ± 0.0003	0.017	0.054	0.017	0.274	0.017	0.302	0.017 ± 0.0003	0.017	0.071	0.017	0.295	0.017	0.319
20	0.016 ± 0.0003	0.016	0.007	0.016	0.269	0.016	0.253	0.016 ± 0.0003	0.016	0.017	0.017	0.298	0.016	0.276

The comparison between the simulated and theoretical datasets of (μ_m) values versus photon energy for the investigated NLO crystals is presented in Figure 3, which reveals that the (μ_m) values decrease with increasing photon energy. This energy-dependent behavior is governed by the relative dominance of different photon interaction mechanisms. The variation of the mass attenuation coefficients with photon energy can be explained using the photoelectric effect, Compton scattering and pair production [7, 29, 41,42]. The energy dependence of the mass attenuation coefficients (μ_m) reveals three distinct energy regimes across the investigated spectrum: low, intermediate and high energy. In the low-energy region, a rapid and continuous drop in this parameter is observed uniformly for all NLO materials, where the photon-matter interaction is heavily dominated by the photoelectric effect, which exhibits a strong dependence on the atomic number (Z) of the constituent elements. Throughout the intermediate-energy spectrum, the (μ_m) values exhibit a gradual decrease with rising photon energy. This trend is

primarily attributed to the dominance of Compton scattering, where the interaction probability is mainly determined by the electron density of the NLO compounds rather than their atomic number. Alternatively, upon reaching the high-energy domain, the parameters display a subtle diminishing trend before transitioning into a steady, independent phase, which is a hallmark feature driven by the progression of pair-production events, which starts to contribute to the total attenuation cross-section at energies above 1.022 MeV and becomes more significant at higher energies, especially in multi-element organic compounds. Higher values of (μ_m) for the studied NLO compounds indicate superior radiation attenuation capabilities across the investigated photon energy range. As tabulated in Table 3, it is evident that the comparative analysis shows a degree of agreement between Geant4 and the theoretical codes (EpiXS, XCOM, and Phy-X/PSD). The percentage deviations ($\%\Delta_1, \%\Delta_2, \%\Delta_3$) remained remarkably low throughout the entire investigated energy spectrum. The small discrepancies observed between the Geant4 results and the theoretical values from EpiXS, XCOM, and Phy-X/PSD demonstrate good consistency among the computational approaches. These results support the use of Geant4 as a useful computational tool for investigating radiation-shielding parameters of organic nonlinear optical materials.

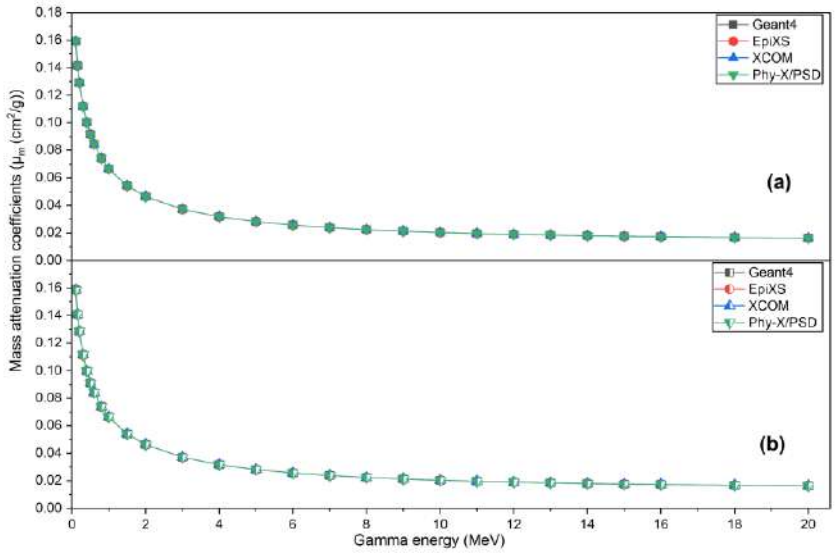


Fig.3. Comparative analysis of Geant4-simulated and theoretical mass attenuation coefficients for (a) $C_{17}H_{14}N_2O_2$ and (b) $C_{13}H_{11}N_3O_2$ organic NLO compounds over the selected gamma energy range.

The determined values of the total atomic cross-section ($\sigma_{t,a}$) and total electronic cross-section ($\sigma_{t,el}$) for the investigated organic NLO materials, calculated using (μ_m) are compiled in Tables 4 and 5, respectively. To show how they change with energy, plots of ($\sigma_{t,a}$) and ($\sigma_{t,el}$) against photon energy are presented in Figures 4 and 5. Across the studied energy spectrum, both cross-sectional parameters exhibit a physical trend that closely follows the behavior of the mass attenuation coefficient (μ_m)[5, 22].

Table 4. Total atomic cross section and corresponding percentage deviations between the Geant4 results and the three theoretical tools for the $C_{17}H_{14}N_2O_2$ and $C_{13}H_{11}N_3O_2$ organic NLO compounds over the selected gamma-ray energy range.

Gamma Energy (MeV)	Total atomic cross section ($\sigma_{t,a} \times 10^{-25}$ $(cm^2/atom)$)													
	$C_{17}H_{14}N_2O_2$							$C_{13}H_{11}N_3O_2$						
	Geant4	EpiXS	$\%\Delta_1$	XCOM	$\%\Delta_2$	Phy-X/PSD	$\%\Delta_3$	Geant4	EpiXS	$\%\Delta_1$	XCOM	$\%\Delta_2$	Phy-X/PSD	$\%\Delta_3$
0.1	21.015	21.017	0.010	21.021	0.031	21.021	0.032	21.927	21.919	0.036	21.923	0.017	21.924	0.013
0.15	18.677	18.677	0.004	18.684	0.035	18.685	0.041	19.456	19.463	0.037	19.478	0.113	19.472	0.085
0.2	17.045	17.047	0.007	17.047	0.008	17.052	0.041	17.750	17.760	0.059	17.765	0.086	17.766	0.093
0.3	14.770	14.784	0.090	14.789	0.125	14.789	0.127	15.392	15.400	0.052	15.403	0.069	15.406	0.089
0.4	13.221	13.230	0.067	13.244	0.170	13.238	0.123	13.775	13.781	0.044	13.789	0.103	13.789	0.102
0.5	12.073	12.075	0.017	12.085	0.097	12.084	0.092	12.574	12.577	0.023	12.587	0.103	12.587	0.097
0.6	11.153	11.167	0.128	11.173	0.184	11.174	0.186	11.620	11.632	0.100	11.638	0.158	11.638	0.158
0.8	9.802	9.806	0.039	9.812	0.105	9.812	0.103	10.213	10.213	0.005	10.220	0.067	10.220	0.071
1	8.812	8.813	0.009	8.822	0.114	8.821	0.109	9.175	9.179	0.040	9.188	0.135	9.188	0.140
1.5	7.163	7.169	0.096	7.181	0.250	7.180	0.245	7.459	7.468	0.112	7.479	0.262	7.479	0.257
2	6.147	6.149	0.020	6.160	0.204	6.159	0.193	6.403	6.405	0.030	6.417	0.209	6.416	0.205
3	4.922	4.927	0.099	4.936	0.273	4.936	0.282	5.130	5.135	0.102	5.144	0.279	5.145	0.286
4	4.207	4.210	0.077	4.220	0.310	4.220	0.306	4.389	4.391	0.044	4.401	0.286	4.401	0.274
5	3.736	3.739	0.056	3.747	0.293	3.748	0.302	3.900	3.901	0.025	3.911	0.276	3.911	0.271
6	3.404	3.405	0.020	3.415	0.298	3.414	0.279	3.553	3.555	0.060	3.564	0.302	3.565	0.319
7	3.157	3.156	0.022	3.166	0.296	3.166	0.286	3.297	3.298	0.028	3.308	0.355	3.308	0.336
8	2.968	2.968	0.033	2.975	0.218	2.975	0.221	3.101	3.102	0.033	3.111	0.311	3.110	0.286
9	2.819	2.817	0.090	2.824	0.178	2.824	0.177	2.943	2.947	0.113	2.955	0.393	2.954	0.379
10	2.696	2.695	0.020	2.702	0.225	2.702	0.244	2.820	2.821	0.015	2.828	0.264	2.828	0.278
11	2.595	2.594	0.036	2.603	0.289	2.602	0.268	2.717	2.717	0.013	2.726	0.314	2.725	0.290
12	2.511	2.511	0.029	2.518	0.294	2.518	0.295	2.630	2.631	0.061	2.638	0.334	2.638	0.326
13	2.440	2.441	0.047	2.447	0.275	2.447	0.285	2.558	2.559	0.052	2.565	0.296	2.565	0.291
14	2.379	2.380	0.042	2.387	0.343	2.387	0.332	2.496	2.496	0.023	2.503	0.298	2.503	0.311
15	2.329	2.329	0.007	2.336	0.305	2.335	0.284	2.443	2.444	0.040	2.451	0.327	2.450	0.314
16	2.287	2.285	0.085	2.291	0.179	2.291	0.170	2.398	2.398	0.035	2.405	0.310	2.405	0.290
18	2.211	2.212	0.052	2.217	0.274	2.218	0.300	2.322	2.324	0.069	2.329	0.295	2.330	0.317
20	2.156	2.156	0.009	2.162	0.269	2.161	0.250	2.266	2.266	0.014	2.272	0.298	2.272	0.273

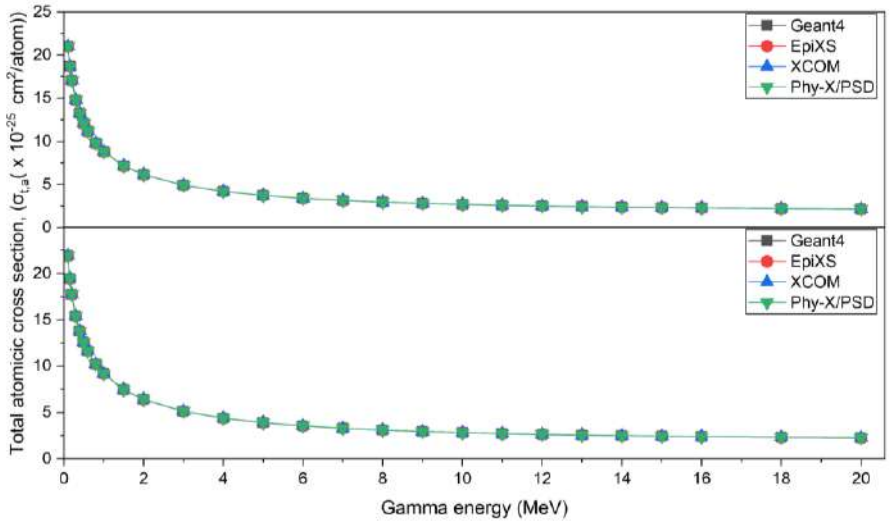


Fig.4. Comparative analysis of Geant4-simulated and theoretical total atomic cross section for (a) $C_{17}H_{14}N_2O_2$ and (b) $C_{13}H_{11}N_3O_2$ organic NLO compounds over the selected gamma energy range.

Table 5. Total electronic cross section and corresponding percentage deviations between the Geant4 results and the three theoretical tools for the $C_{17}H_{14}N_2O_2$ and $C_{13}H_{11}N_3O_2$ organic NLO compounds over the selected gamma-ray energy range.

Gamma Energy (MeV)	Total electronic cross section ($\sigma_{t,el} \times 10^{-25}$ $(cm^2/atom)$)													
	$C_{17}H_{14}N_2O_2$							$C_{13}H_{11}N_3O_2$						
	Geant4	EpiXS	$\%\Delta_1$	XCOM	$\%\Delta_2$	Phy-X/PSD	$\%\Delta_3$	Geant4	EpiXS	$\%\Delta_1$	XCOM	$\%\Delta_2$	Phy-X/PSD	$\%\Delta_3$
0.1	4.998	4.998	0.004	4.996	0.036	5.000	0.036	5.004	5.004	0.004	5.001	0.056	5.005	0.036
0.15	4.458	4.462	0.090	4.463	0.124	4.464	0.145	4.460	4.464	0.090	4.465	0.117	4.466	0.146
0.2	4.078	4.078	0.007	4.079	0.018	4.080	0.047	4.079	4.079	0.008	4.080	0.017	4.081	0.048
0.3	3.536	3.540	0.119	3.541	0.123	3.542	0.160	3.537	3.541	0.119	3.541	0.124	3.542	0.160
0.4	3.169	3.170	0.020	3.172	0.087	3.172	0.087	3.169	3.170	0.020	3.172	0.086	3.172	0.087
0.5	2.893	2.893	0.010	2.896	0.097	2.896	0.093	2.893	2.893	0.010	2.896	0.094	2.896	0.092
0.6	2.674	2.676	0.075	2.678	0.138	2.678	0.141	2.674	2.676	0.075	2.678	0.137	2.678	0.141
0.8	2.350	2.350	0.014	2.352	0.099	2.352	0.091	2.350	2.350	0.014	2.352	0.099	2.352	0.092
1	2.112	2.112	0.003	2.114	0.107	2.115	0.111	2.112	2.112	0.002	2.114	0.105	2.114	0.111
1.5	1.716	1.718	0.088	1.720	0.251	1.720	0.244	1.716	1.718	0.088	1.720	0.247	1.720	0.241
2	1.471	1.471	0.006	1.473	0.185	1.474	0.192	1.471	1.471	0.006	1.474	0.182	1.474	0.192
3	1.171	1.173	0.103	1.175	0.267	1.175	0.289	1.172	1.173	0.103	1.175	0.261	1.176	0.290
4	0.995	0.996	0.014	0.997	0.209	0.998	0.258	0.997	0.997	0.014	0.999	0.200	0.999	0.258
5	0.878	0.878	0.008	0.879	0.189	0.880	0.271	0.880	0.880	0.008	0.881	0.173	0.882	0.270
6	0.793	0.794	0.043	0.795	0.224	0.796	0.319	0.796	0.796	0.043	0.797	0.204	0.798	0.318
7	0.730	0.730	0.009	0.732	0.211	0.733	0.339	0.733	0.733	0.009	0.734	0.186	0.735	0.337
8	0.682	0.682	0.020	0.682	0.136	0.684	0.297	0.685	0.685	0.020	0.685	0.106	0.687	0.296
9	0.642	0.643	0.058	0.643	0.186	0.644	0.351	0.646	0.646	0.058	0.646	0.146	0.648	0.349
10	0.611	0.611	0.003	0.611	0.078	0.612	0.289	0.614	0.614	0.003	0.614	0.037	0.616	0.288
11	0.584	0.584	0.009	0.585	0.124	0.586	0.345	0.588	0.588	0.009	0.588	0.077	0.590	0.342
12	0.561	0.562	0.044	0.562	0.085	0.563	0.340	0.566	0.566	0.044	0.566	0.033	0.567	0.337
13	0.542	0.543	0.045	0.543	0.064	0.544	0.318	0.547	0.547	0.045	0.547	0.009	0.549	0.316
14	0.526	0.526	0.016	0.526	0.056	0.528	0.339	0.531	0.531	0.016	0.531	0.003	0.532	0.336
15	0.512	0.512	0.027	0.512	0.015	0.513	0.336	0.517	0.517	0.027	0.516	0.049	0.518	0.333
16	0.499	0.500	0.023	0.499	0.002	0.501	0.316	0.504	0.505	0.023	0.504	0.065	0.506	0.314
18	0.479	0.479	0.053	0.479	0.022	0.480	0.342	0.484	0.484	0.052	0.483	0.096	0.486	0.338
20	0.462	0.462	0.003	0.462	0.100	0.464	0.305	0.468	0.468	0.003	0.467	0.179	0.469	0.302

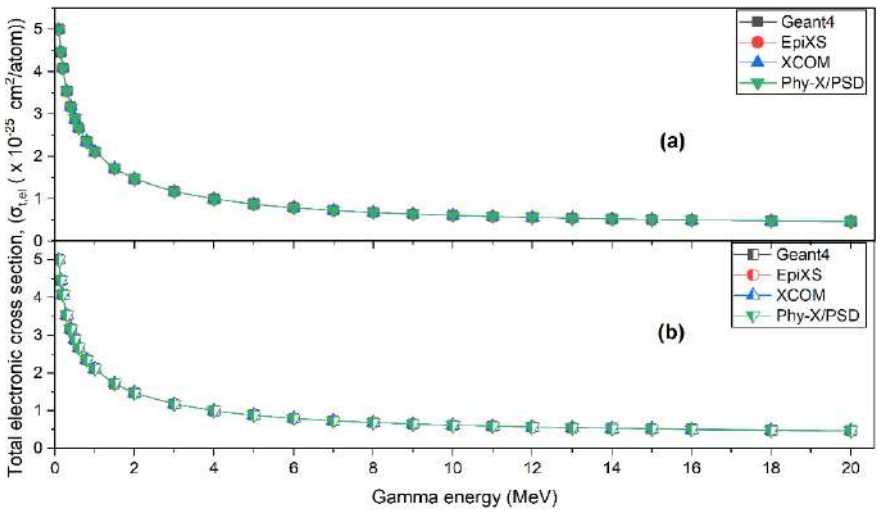


Fig.5. Comparative analysis of Geant4-simulated and theoretical total electronic cross section for (a) $C_{17}H_{14}N_2O_2$ and (b) $C_{13}H_{11}N_3O_2$ organic NLO compounds over the selected gamma energy range.

The mass attenuation coefficients (μ_m) derived from Geant4 simulations were utilized to determine the effective atomic numbers (Z_{eff}) and effective electron densities (N_{eff}) for the investigated organic NLO compounds via Eqs. (4) and (7). To validate the Geant4 simulation results, they were directly compared with the theoretical values obtained from EpiXS, XCOM, and Phy-X/PSD software. The detailed numerical results of (Z_{eff}) are tabulated in Table 6 for investigated samples.

Table 6. Effective atomic numbers and corresponding percentage deviations between the Geant4 results and the three theoretical tools for the $C_{17}H_{14}N_2O_2$ and $C_{13}H_{11}N_3O_2$ organic NLO compounds over the selected gamma-ray energy range.

Gamma Energy (MeV)	Effective atomic numbers (Z_{eff})													
	$C_{17}H_{14}N_2O_2$							$C_{13}H_{11}N_3O_2$						
	Geant4	EpiXS	$\%\Delta_1$	XCOM	$\%\Delta_2$	Phy-X/PSD	$\%\Delta_3$	Geant4	EpiXS	$\%\Delta_1$	XCOM	$\%\Delta_2$	Phy-X/PSD	$\%\Delta_3$
0.1	4.205 ± 0.028	4.205	0.006	4.208	0.067	4.205	0.004	4.382 ± 0.029	4.380	0.040	4.384	0.039	4.380	0.049
0.15	4.190 ± 0.030	4.186	0.094	4.186	0.088	4.185	0.104	4.362 ± 0.031	4.360	0.053	4.362	0.004	4.360	0.061
0.2	4.180 ± 0.031	4.180	0.000	4.180	0.010	4.180	0.007	4.352 ± 0.032	4.354	0.051	4.355	0.068	4.353	0.045
0.3	4.177 ± 0.033	4.176	0.029	4.177	0.002	4.176	0.033	4.352 ± 0.034	4.349	0.067	4.350	0.055	4.349	0.071
0.4	4.172 ± 0.035	4.174	0.047	4.176	0.083	4.174	0.037	4.347 ± 0.036	4.348	0.024	4.347	0.017	4.347	0.015
0.5	4.173 ± 0.036	4.174	0.006	4.173	0.000	4.173	0.001	4.347 ± 0.038	4.347	0.012	4.347	0.009	4.347	0.005
0.6	4.171 ± 0.038	4.173	0.053	4.173	0.046	4.173	0.045	4.345 ± 0.039	4.347	0.025	4.346	0.021	4.346	0.018
0.8	4.172 ± 0.040	4.173	0.025	4.172	0.006	4.172	0.012	4.347 ± 0.042	4.346	0.009	4.345	0.032	4.346	0.021
1	4.172 ± 0.042	4.173	0.012	4.172	0.006	4.172	0.002	4.344 ± 0.044	4.346	0.042	4.345	0.030	4.345	0.029
1.5	4.174 ± 0.048	4.174	0.007	4.174	0.001	4.174	0.001	4.346 ± 0.049	4.347	0.024	4.347	0.016	4.347	0.016
2	4.180 ± 0.051	4.181	0.015	4.181	0.019	4.180	0.002	4.353 ± 0.053	4.354	0.025	4.354	0.027	4.354	0.013
3	4.202 ± 0.057	4.202	0.004	4.202	0.006	4.202	0.007	4.376 ± 0.060	4.376	0.001	4.377	0.019	4.376	0.004
4	4.227 ± 0.062	4.229	0.062	4.231	0.101	4.229	0.048	4.403 ± 0.065	4.404	0.030	4.407	0.086	4.404	0.017
5	4.257 ± 0.067	4.259	0.047	4.262	0.103	4.258	0.031	4.434 ± 0.069	4.435	0.017	4.438	0.103	4.434	0.002
6	4.291 ± 0.070	4.290	0.022	4.295	0.074	4.290	0.041	4.466 ± 0.074	4.467	0.018	4.470	0.099	4.466	0.000
7	4.323 ± 0.073	4.322	0.031	4.327	0.085	4.321	0.053	4.498 ± 0.076	4.499	0.019	4.506	0.169	4.498	0.002
8	4.356 ± 0.077	4.353	0.053	4.359	0.082	4.352	0.077	4.530 ± 0.080	4.531	0.013	4.539	0.205	4.530	0.010
9	4.391 ± 0.079	4.384	0.148	4.390	0.008	4.383	0.174	4.560 ± 0.083	4.562	0.056	4.571	0.247	4.561	0.031
10	4.415 ± 0.082	4.414	0.017	4.422	0.147	4.413	0.046	4.592 ± 0.085	4.593	0.017	4.602	0.227	4.591	0.010
11	4.446 ± 0.081	4.444	0.045	4.453	0.165	4.442	0.077	4.623 ± 0.086	4.622	0.022	4.634	0.238	4.621	0.053
12	4.472 ± 0.085	4.472	0.015	4.482	0.209	4.470	0.045	4.650 ± 0.088	4.650	0.018	4.664	0.301	4.649	0.011
13	4.499 ± 0.087	4.499	0.001	4.508	0.211	4.497	0.033	4.677 ± 0.090	4.678	0.007	4.691	0.288	4.676	0.025
14	4.524 ± 0.089	4.525	0.026	4.537	0.287	4.523	0.007	4.704 ± 0.091	4.704	0.007	4.718	0.301	4.703	0.025
15	4.551 ± 0.089	4.550	0.020	4.564	0.290	4.548	0.053	4.729 ± 0.094	4.729	0.012	4.747	0.376	4.728	0.019
16	4.579 ± 0.090	4.574	0.108	4.587	0.176	4.572	0.147	4.753 ± 0.095	4.754	0.012	4.771	0.375	4.752	0.024
18	4.619	4.619	0.001	4.633	0.296	4.618	0.042	4.798	4.799	0.018	4.817	0.390	4.797	0.022

	± 0.092							± 0.097						
20	4.662 ± 0.095	4.661	0.012	4.679	0.368	4.659	0.055	4.841 ± 0.098	4.841	0.012	4.864	0.477	4.839	0.029

Figure 6 illustrates the variation of the effective atomic number $(\boldsymbol{Z_{eff}})$ for the investigated NLO material as a function of photon energy across a wide spectrum up to 20 MeV. A strong agreement is observed among the Geant4 simulation results and the theoretical predictions (EpiXS, XCOM, and Phy-X/PSD). At low photon energies (below 1 MeV), the $\boldsymbol{Z_{eff}}$ profile experiences a slight initial drop, which is governed by the shifting dominance from photoelectric absorption to Compton scattering. Crucially, beyond the 1.02 MeV threshold and up to 20 MeV, a continuous and monotonic increase in $\boldsymbol{Z_{eff}}$ is prominently observed. This ascending trend is physically attributed to the progressive dominance of the pair-production mechanism.

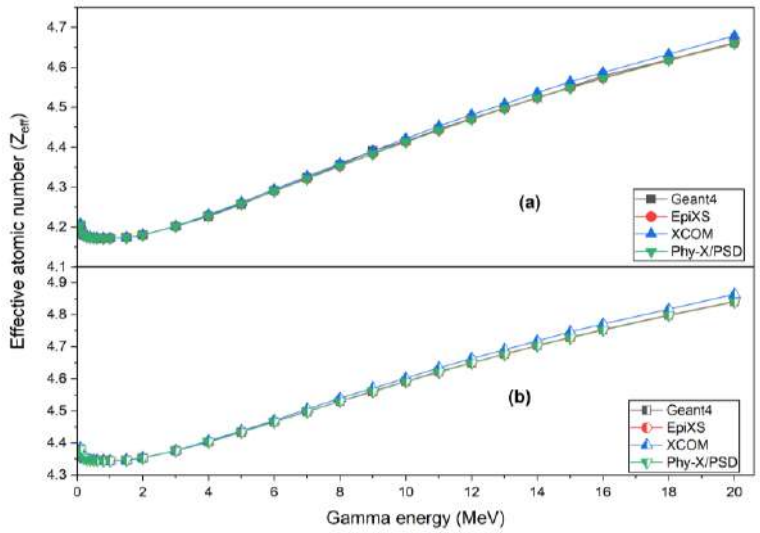


Fig.6 Comparative analysis of Geant4-simulated and theoretical effective atomic numbers for (a) $C_{17}H_{14}N_2O_2$ and (b) $C_{13}H_{11}N_3O_2$ organic NLO compounds over the selected gamma energy range.

Table 7 presents a comparative analysis of the calculated Z_{eff} values for the investigated organic NLO materials against various standard reference materials across the energy range of 122 keV to 1330 keV. The investigated compounds exhibit low Z_{eff} values ranging between 4.173 and 4.376, which are intrinsically attributed to their organic framework dominated by light elements (H, C, N, and O). Notably, these values are highly comparable to those of organic references and human muscle tissue, while remaining significantly lower than the inorganic counterparts such as KDP. Furthermore, the slight reduction in Z_{eff} as energy increases from 122 keV to 840 keV marks the physical transition from photoelectric absorption dominance to Compton scattering interactions. This strong equivalence to human tissue underscores the potential of these synthesized NLO crystals for biological radiation dosimetry and medical phantom applications.

Table 7. Effective atomic number (Z_{eff}) values of the investigated NLO materials as compared with those of organic and inorganic compounds, TLD materials, polymers, and human tissues.

E (KeV)	Effective atomic numbers (Z_{eff})											
	investigated NLO materials		Organic NLO materials[4]		inorganic NLO materials[7]		TLD materials[39]		Polymers [29]		Human tissues[43]	
	$C_{17}H_{14}N_2O_2$	$C_{13}H_{11}N_3O_2$	$C_6H_6N_2O_2$	$C_5H_5N_3O_2$	KDP	ADP	LiF	$CaCO_3$	POM	PAN	Bone	Muscle
122	4.207	4.376	4.698	4.829	9.614	5.453	6.054	9.825	4.022	4.066	5.996	3.435
356	4.177	4.351	4.651	4.776	9.216	5.170	6.006	10.005	4.011	4.052	6.004	3.434
511	4.175	4.348	4.635	4.757	9.225	4.989	5.987	10.051	4.005	4.046	5.901	3.437
662	4.173	4.344	4.625	4.755	8.906	4.468	5.988	10.052	4.002	4.042	5.994	3.435
840	4.173	4.343	4.615	4.734	8.957	4.014	6.001	10.041	3.999	4.039	6.024	3.446
1170	4.174	4.344	4.600	4.717	8.802	5.208	5.924	10.025	3.995	4.035	6.104	3.478
1275	4.175	4.345	4.597	4.713	8.894	4.676	6.002	10.037	3.994	4.033	6.199	3.517
1330	4.176	4.346	4.595	4.711	8.781	4.605	5.966	10.042	3.993	4.032	6.297	3.559

The comprehensive numerical values of the effective electron densities (N_{eff}) for the studied samples are systematically cataloged in Table 8.

Table 8. Effective electron densities and corresponding percentage deviations between the Geant4 results and the three theoretical tools for the $C_{17}H_{14}N_2O_2$ and $C_{13}H_{11}N_3O_2$ organic NLO compounds over the selected gamma-ray energy range.

Gamma Energy (MeV)	Effective electron densities (N_{eff} ($\times 10^{23}$ e^-/g))													
	$C_{17}H_{14}N_2O_2$							$C_{13}H_{11}N_3O_2$						
	Geant4	EpiXS	$\%\Delta_1$	XCOM	$\%\Delta_2$	Phy-X/PSD	$\%\Delta_3$	Geant4	EpiXS	$\%\Delta_1$	XCOM	$\%\Delta_2$	Phy-X/PSD	$\%\Delta_3$
0.1	3.184 ± 0.021	3.185	0.008	3.187	0.067	3.184	0.002	3.172 ± 0.021	3.171	0.038	3.173	0.039	3.171	0.047
0.15	3.173 ± 0.022	3.170	0.092	3.170	0.088	3.170	0.102	3.158 ± 0.022	3.156	0.051	3.158	0.004	3.156	0.059
0.2	3.166 ± 0.023	3.166	0.002	3.165	0.010	3.165	0.004	3.150 ± 0.023	3.152	0.053	3.152	0.068	3.152	0.048
0.3	3.163 ± 0.025	3.162	0.027	3.163	0.002	3.162	0.031	3.151 ± 0.025	3.149	0.065	3.149	0.055	3.148	0.068
0.4	3.160 ± 0.026	3.161	0.049	3.162	0.083	3.161	0.039	3.147 ± 0.026	3.147	0.026	3.147	0.017	3.147	0.017
0.5	3.161 ± 0.027	3.161	0.009	3.161	0.000	3.161	0.002	3.147 ± 0.027	3.147	0.015	3.147	0.009	3.147	0.007
0.6	3.159 ± 0.029	3.161	0.055	3.160	0.046	3.160	0.047	3.146 ± 0.029	3.147	0.027	3.146	0.021	3.146	0.020
0.8	3.159 ± 0.030	3.160	0.028	3.160	0.006	3.160	0.014	3.146 ± 0.030	3.146	0.007	3.145	0.032	3.146	0.019
1	3.160 ± 0.032	3.160	0.014	3.160	0.006	3.160	0.000	3.145 ± 0.032	3.146	0.044	3.146	0.030	3.146	0.031
1.5	3.161 ± 0.036	3.161	0.010	3.161	0.001	3.161	0.003	3.146 ± 0.035	3.147	0.026	3.147	0.016	3.147	0.018
2	3.166 ± 0.039	3.166	0.017	3.166	0.019	3.166	0.004	3.151 ± 0.038	3.152	0.027	3.152	0.027	3.152	0.015
3	3.182 ± 0.043	3.182	0.002	3.183	0.006	3.182	0.005	3.168 ± 0.043	3.168	0.001	3.169	0.019	3.168	0.002
4	3.201 ± 0.047	3.203	0.065	3.204	0.101	3.203	0.050	3.187 ± 0.047	3.188	0.032	3.190	0.086	3.188	0.019
5	3.224 ± 0.051	3.226	0.050	3.227	0.103	3.225	0.033	3.210 ± 0.050	3.210	0.019	3.213	0.103	3.210	0.004
6	3.250 ± 0.053	3.249	0.020	3.252	0.074	3.249	0.038	3.233 ± 0.053	3.233	0.020	3.236	0.099	3.233	0.003
7	3.274 ± 0.055	3.273	0.029	3.277	0.085	3.273	0.051	3.256 ± 0.055	3.257	0.022	3.262	0.169	3.256	0.001
8	3.299 ± 0.058	3.297	0.051	3.301	0.082	3.296	0.075	3.279 ± 0.058	3.280	0.015	3.286	0.205	3.279	0.008
9	3.325 ± 0.060	3.320	0.146	3.325	0.008	3.320	0.172	3.301 ± 0.060	3.303	0.058	3.309	0.247	3.302	0.033
10	3.344	3.343	0.015	3.349	0.147	3.342	0.044	3.324	3.325	0.019	3.332	0.227	3.324	0.008

	± 0.062							± 0.061						
11	3.367 ± 0.062	3.365	0.043	3.372	0.165	3.364	0.075	3.347 ± 0.062	3.346	0.020	3.355	0.238	3.345	0.050
12	3.387 ± 0.065	3.387	0.013	3.394	0.209	3.386	0.043	3.366 ± 0.064	3.366	0.020	3.376	0.301	3.366	0.009
13	3.4070 ± 0.066	3.407	0.003	3.414	0.211	3.406	0.030	3.386 ± 0.065	3.386	0.009	3.396	0.288	3.385	0.023
14	3.426 ± 0.068	3.427	0.028	3.436	0.287	3.426	0.005	3.405 ± 0.066	3.405	0.009	3.415	0.301	3.404	0.023
15	3.446 ± 0.067	3.446	0.018	3.457	0.290	3.445	0.050	3.423 ± 0.068	3.424	0.015	3.436	0.376	3.423	0.016
16	3.468 ± 0.068	3.464	0.106	3.474	0.176	3.463	0.144	3.441 ± 0.068	3.441	0.015	3.454	0.375	3.440	0.021
18	3.498 ± 0.069	3.499	0.002	3.509	0.296	3.497	0.040	3.474 ± 0.070	3.474	0.020	3.487	0.390	3.473	0.019
20	3.531 ± 0.072	3.530	0.010	3.544	0.368	3.529	0.053	3.504 ± 0.071	3.505	0.014	3.521	0.477	3.503	0.027

Figure 7 illustrates the energy-dependent behavior of the effective electron density (N_{eff}) for the synthesized organic NLO compounds within the energy range from 0.1 MeV to 20 MeV. Physically, the N_{eff} profile is directly proportional to the effective atomic number $\boldsymbol{Z_{eff}}$, which explains the similar physical trends observed between both parameters. Furthermore, a remarkable alignment is demonstrated between the Geant4 Monte Carlo simulation results and the validation benchmarks (EpiXS, XCOM, and Phy-X/PSD) across the entire spectrum.

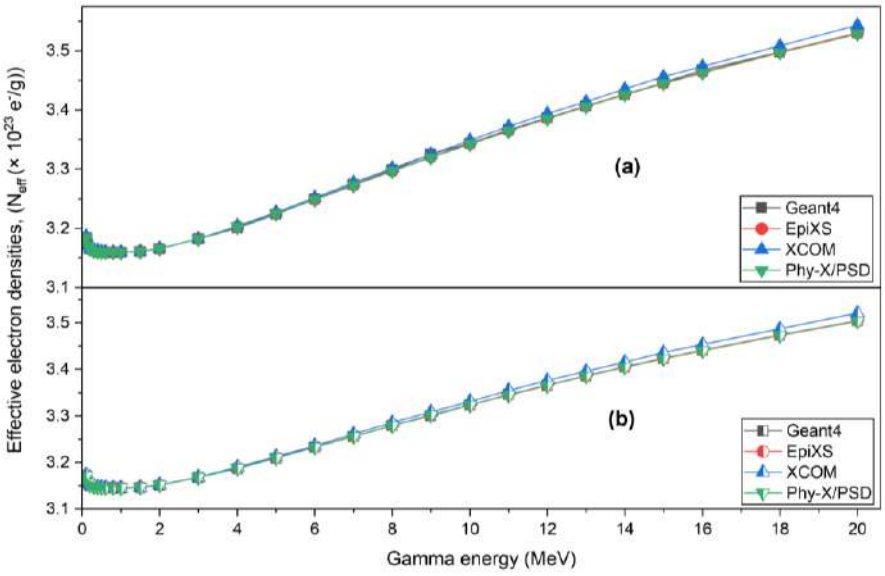


Fig.7. Comparative analysis of Geant4-simulated and theoretical effective electron densities for (a) $C_{17}H_{14}N_2O_2$ and (b) $C_{13}H_{11}N_3O_2$ organic NLO compounds over the selected gamma energy range

Figure 8 shows a direct linear relationship between ($\boldsymbol{Z_{eff}}$) and (N_{eff}) for the investigated organic NLO compounds across the energy range from 0.1 MeV to 20 MeV [7]. This highly linear trend is a natural physical consequence, as both parameters are governed by the same electron cloud of the material. Additionally, the $C_{13}H_{11}N_3O_2$ compound displays higher values than the $C_{17}H_{14}N_2O_2$ compound. This variation is attributable to their chemical structures; the $C_{13}H_{11}N_3O_2$ compound contains a higher proportion of heavier atoms (nitrogen and oxygen). This atomic composition increases the overall electron density, thereby enhancing the probability of radiation interactions within the material.

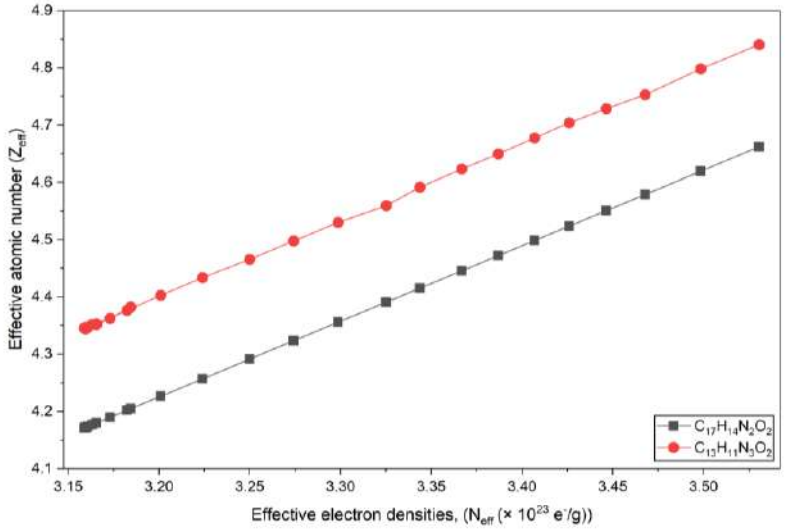


Fig. 8. Variation and correlation of $(\boldsymbol{Z_{eff}})$ and (N_{eff}) with photon energy for the investigated NLO compounds.

The molar extinction coefficients $(\boldsymbol{\varepsilon})$ of the investigated organic NLO materials were calculated using Eq. 8 and are tabulated in Table 9. Figure 9 illustrates the direct relationship between $(\boldsymbol{\varepsilon})$ and photon energy for all samples. As evident from the plot, the values of $(\boldsymbol{\varepsilon})$ exhibit a sharp decrease at low energy levels, followed by a nearly constant trend at higher gamma-ray energies [9]. This profile is a natural physical outcome of the transition from photoelectric absorption dominance at lower energies to Compton scattering dominance at higher energies. Overall, the comparison shows slightly higher $\%\Delta_2$ and $\%\Delta_3$ values for $C_{13}H_{11}N_3O_2$, which can be associated with its different elemental composition and corresponding elemental weight fractions. Furthermore, all modeling tools (Geant4, EpiXS, XCOM, and Phy-X/PSD) show good agreement, demonstrating a consistent trend among the simulated and theoretical datasets.

Table 9. Molar extinction coefficients and corresponding percentage deviations between the Geant4 results and the three theoretical tools for the $C_{17}H_{14}N_2O_2$ and $C_{13}H_{11}N_3O_2$ organic NLO compounds over the selected gamma-ray energy range.

Gamma Energy (MeV)	Molar extinction coefficients (ε (cm^2/mol))													
	$C_{17}H_{14}N_2O_2$							$C_{13}H_{11}N_3O_2$						
	Geant4	EpiXS	$\%\Delta_1$	XCOM	$\%\Delta_2$	Phy-X/PSD	$\%\Delta_3$	Geant4	EpiXS	$\%\Delta_1$	XCOM	$\%\Delta_2$	Phy-X/PSD	$\%\Delta_3$
0.1	0.550	0.550	0.010	0.550	0.031	0.5500	0.032	0.574	0.574	0.036	0.574	0.017	0.574	0.013
0.15	0.489	0.489	0.004	0.489	0.035	0.4889	0.041	0.509	0.509	0.037	0.510	0.113	0.510	0.085
0.2	0.446	0.446	0.007	0.446	0.008	0.4462	0.041	0.464	0.465	0.059	0.465	0.086	0.465	0.093
0.3	0.387	0.387	0.090	0.387	0.125	0.3870	0.127	0.403	0.403	0.052	0.403	0.069	0.403	0.089
0.4	0.346	0.346	0.067	0.347	0.170	0.3464	0.123	0.360	0.361	0.044	0.361	0.103	0.361	0.102
0.5	0.316	0.316	0.017	0.316	0.097	0.3162	0.092	0.329	0.329	0.023	0.329	0.103	0.329	0.097
0.6	0.292	0.292	0.128	0.292	0.184	0.2924	0.186	0.304	0.304	0.100	0.305	0.158	0.305	0.158
0.8	0.257	0.257	0.039	0.257	0.105	0.2567	0.103	0.267	0.267	0.005	0.267	0.067	0.267	0.071
1	0.231	0.231	0.009	0.231	0.114	0.2308	0.109	0.240	0.240	0.040	0.240	0.135	0.240	0.140
1.5	0.187	0.188	0.096	0.188	0.250	0.1879	0.245	0.195	0.195	0.112	0.196	0.262	0.196	0.257
2	0.161	0.161	0.020	0.161	0.204	0.1612	0.193	0.168	0.168	0.030	0.168	0.209	0.168	0.205
3	0.129	0.129	0.099	0.129	0.273	0.1292	0.282	0.134	0.134	0.102	0.135	0.279	0.135	0.286
4	0.110	0.110	0.077	0.110	0.310	0.1104	0.306	0.115	0.115	0.044	0.115	0.286	0.115	0.274
5	0.098	0.098	0.056	0.098	0.293	0.0981	0.302	0.102	0.102	0.025	0.102	0.276	0.102	0.271
6	0.089	0.089	0.020	0.089	0.298	0.0893	0.279	0.093	0.093	0.060	0.093	0.302	0.093	0.319

7	0.083	0.083	0.022	0.083	0.296	0.0828	0.286	0.086	0.086	0.028	0.087	0.355	0.087	0.336
8	0.078	0.078	0.033	0.078	0.218	0.0778	0.221	0.081	0.081	0.033	0.081	0.311	0.081	0.286
9	0.074	0.074	0.090	0.074	0.178	0.0739	0.177	0.077	0.077	0.113	0.077	0.393	0.077	0.379
10	0.071	0.071	0.020	0.071	0.225	0.0707	0.244	0.074	0.074	0.015	0.074	0.264	0.074	0.278
11	0.068	0.068	0.036	0.068	0.289	0.0681	0.268	0.071	0.071	0.013	0.071	0.314	0.071	0.290
12	0.066	0.066	0.029	0.066	0.294	0.0659	0.295	0.069	0.069	0.061	0.069	0.334	0.069	0.326
13	0.064	0.064	0.047	0.064	0.275	0.0640	0.285	0.067	0.067	0.052	0.067	0.296	0.067	0.291
14	0.062	0.062	0.042	0.063	0.343	0.0625	0.332	0.065	0.065	0.023	0.065	0.298	0.066	0.311
15	0.061	0.061	0.007	0.061	0.305	0.0611	0.284	0.064	0.064	0.040	0.064	0.327	0.064	0.314
16	0.060	0.060	0.085	0.060	0.179	0.0599	0.170	0.063	0.063	0.035	0.063	0.310	0.063	0.290
18	0.058	0.058	0.052	0.058	0.274	0.0580	0.300	0.061	0.061	0.069	0.061	0.295	0.061	0.317
20	0.056	0.056	0.009	0.057	0.269	0.0565	0.250	0.059	0.059	0.014	0.059	0.298	0.059	0.273

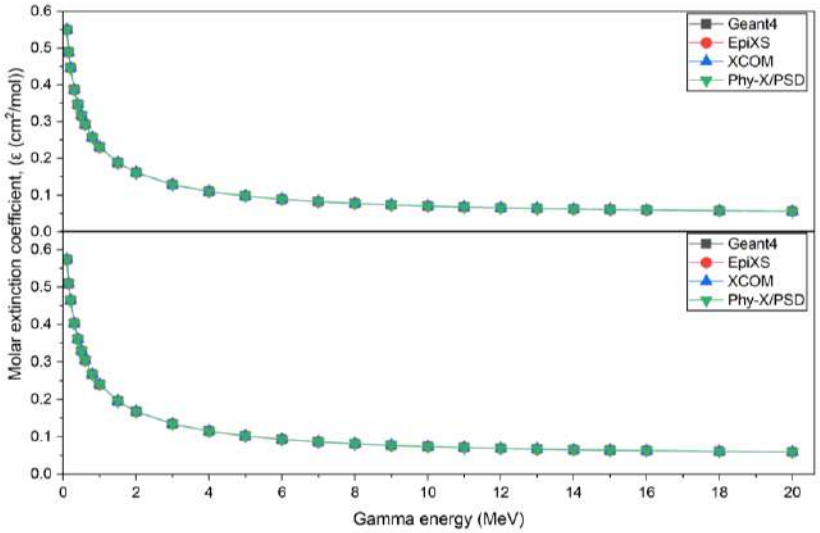


Fig.9. Comparative analysis of Geant4-simulated and theoretical molar extinction coefficients for (a) $C_{17}H_{14}N_2O_2$ and (b) $C_{13}H_{11}N_3O_2$ organic NLO compounds over the selected gamma energy range.

6. Conclusion

This work systematically evaluated the radiation shielding performance of organic non-linear optical (NLO) compounds across a broad energy range from 0.1 MeV to 20 MeV. Employing the Geant4 Monte Carlo toolkit alongside standard analytical frameworks, the gamma-ray attenuation parameters were rigorously determined to establish the overall shielding efficacy of the investigated materials. The simulation outcomes were compared with three established computational platforms (EpiXS, XCOM and Phy-X/PSD), showing good agreement with percentage deviations of less than 0.5% throughout the investigated range. For the first time, this study provides detailed information about (μ_m) values and other related parameters, such as $(\sigma_{t,a})$, $(\sigma_{t,el})$, (Z_{eff}), (N_{eff}) and $(\boldsymbol{\varepsilon})$ for two organic NLO compounds: 3-[1-(2-phenylhydrazinylidene)ethyl]-2H-chromen-2-one ($C_{17}H_{14}N_2O_2$) and 4-nitrophenyl hydrazone ($C_{13}H_{11}N_3O_2$). The obtained values of the mass attenuation coefficient (μ_m) coupled with the low effective atomic number (Z_{eff}) demonstrate the potential of the investigated organic NLO compounds in radiation therapy, diagnostic imaging and other technological applications. Total atomic and electronic cross-sections $(\sigma_{t,a}, \sigma_{t,el})$ exhibit an initial decrease as photon energy increases, similar to the mass attenuation coefficient (μ_m), and then become relatively independent of energy owing to the consecutive dominance of Compton scattering and pair production

within their respective energy ranges. The effective atomic numbers of the studied NLO samples were evaluated against various reference materials, including organic and inorganic NLO compounds, TLDs, polymers, and human tissues, showing close similarity, indicating that the investigated materials hold great potential for use in radiation dosimetry. Furthermore, the results indicate that Geant4 Monte Carlo simulations can serve as a useful computational tool for analyzing photon interactions with matter and for supporting comparative radiation-shielding studies.

Funding

The author declares that no funds, grants, or other support were received during the preparation of this manuscript.

Data Availability Statement

All data generated or analyzed during this study are included within the manuscript.

Conflicts of Interest

The author declares no conflict of interest.

References

[1] V.V. Awasarmol, D.K. Gaikwad, S.S. Obaid, P.P. Pawar, Gamma radiation studies on organic nonlinear optical materials in the energy range 122–1330 keV, Proc. Natl. Acad. Sci., India, Sect. A Phys. Sci. 90 (2020), pp. 839–844. https://doi.org/10.1007/s40010-019-00636-1

[2] M.R. Suresh Kumar, H.J. Ravindra, S.M. Dharmaprakash, Synthesis, crystal growth and characterization of glycine lithium sulphate. J. Cryst. Growth. 306 (2007), pp. 361–365. https://doi.org/10.1016/j.jcrysgro.2007.05.015

[3] D.S. Chemla, J. Zyss, Nonlinear optical properties of organic molecules and crystals, vol 1. Academic, Orland, 1987.

[4] V.V. Awasarmol, D.K. Gaikwad, S.D. Raut, P.P. Pawar, Gamma ray interaction studies of organic nonlinear optical materials in the energy range 122 keV–1330 keV, Results Phys. 7 (2017), pp. 272–279. http://dx.doi.org/10.1016/j.rinp.2016.12.017

[5] V.V. Awasarmol, D.K. Gaikwad , S.D. Raut , P.P. Pawar, Photon interaction study of organic nonlinear optical materials in the energy range 122 keV to 1330 keV, Radiat. Phys. Chem. 130 (2017), pp.343-350. https://doi.org/10.1016/j.radphyschem.2016.09.012

[6] V.V. Awasarmol, P.P. Pawar, M.B. Solunke, Effective atomic numbers and effective electron densities of inorganic nonlinear optical materials in the energy range 356 keV -1330 keV, Int. J. Tech. Res. Sci. 2 (2017), pp.26-29.

[7] M.S. Al-Buriahi, V.P. Singh, Halil Arslan, V.V. Awasarmol, B.T. Tonguc, Gamma-ray attenuation properties of some NLO materials: potential use in dosimetry, Radiat. Environ. Biophys. 59 (2020), pp.145–150. https://doi.org/10.1007/s00411-019-00824-y

[8] S. Bhadauria, M. Das, S. Saxena, R. Prasad, P. Sen, R. Dwivedi, Recent progress in non-linear optical material, syntheses, characterization and geometry optimization of dicinnamalacetone. Arch. Phys. Res. 2 (2011), pp.36–44.

[9] Vishal Vishwanath Awasarmol, Gamma ray attenuation parameters of inorganic nonlinear optical materials in the energy range 122 keV to 1330 keV, Indian J. Pure Appl. Phys. 55 (2017), pp. 65-72.

[10] H. Baltas, S. Celik, U. Cevik, E. Yamaz, Measurements of mass attenuation coefficient and effective atomic number for MgB2 superconductor using X-ray energies, Radiat. Meas. 42 (2007), pp.55–60. https://doi.org/10.1016/j.radmeas.2006.08.005

[11] S.S. Shaikh, S.B. Borul, S.S. Khan, A study on the gamma-ray attenuation coefficient parameters for potassium carbonate, Int. J. Eng. Technol. Manag. Sci. 7 (2023), pp.79-86. https://doi.org/10.46647/ijetms.2023.v07i03.011

[12] Z. Aygün, M. Aygün, Radiation shielding potentials of $Hf_{0.5}Mo_{0.5}NbTiZrCrx$ (X=0.1,0.3,0.5) refractory high entropy alloys by EPIXS, Rahva J. Tech. Soc. Stud. 3 (2023), pp. 83-95.

[13] Z. Aygun, Study on radiation shielding characteristics of refractory high entropy alloys by EpiXS code, Acta Phys. Pol. A, 143 (2023), pp.66-74. https://doi.org/10.12693/APhysPolA.143.66

[14] V.P. Singh, S.P. Shirmardi, M.E. Medhat, N.M. Badiger, Determination of mass attenuation coefficient for some polymers using Monte Carlo simulation, Vacuum, 119 (2015), pp. 284-288. https://doi.org/10.1016/j.vacuum.2015.06.006

[15] C.V. More, H. Alavian , P.P. Pawar, Evaluation of gamma-ray attenuation characteristics of some thermoplastic polymers: Experimental, WinXCom and MCNPX studies, J. Non-Cryst. Solids, 546 (2020), https://doi.org/10.1016/j.jnoncrysol.2020.120277

[16] G.A. Urfa, T. Wianto, T.N. Manik, A.V. Nasrulloh, Study of alternative shielding material for gamma radiation using Monte Carlo simulation, J. Teknol. Reaktor Nuklir, 25 (2023), pp. 101–106. https://doi.org/ 10.55981/tdm.2023.6925

[17] I. Han, L. Demir, Mass attenuation coefficient, effective atomic and electron number of Ti and Ni alloy, Radiat. Meas. 44 (2009), PP.289–294. https://doi.org/10.1016/j.radmeas.2009.03.010

[18] I. Han, L. Demir, M. Sahin, Determination of mass attenuation coefficient, effective atomic number and electron number for some natural minerals. Radiat. Phys. Chem. 78 (2009), pp. 760–764. https://doi.org/10.1016/j.radphyschem.2009.03.077

[19] M.E. Medhat, V.P. Singh, Mass attenuation coefficients of composite materials by Geant4, XCOM and experimental data: comparative study. Radiat. Eff. Defects Solids, 169 (2014), pp. 800–807. http://dx.doi.org/10.1080/10420150.2014.950264

[20] K. Alsafi , D. Abdullah Aloraini , H.M. Almutairi , S.A.M. Issa , H.M.H. Zakaly, K.S. Shaaban, Comprehensive study of optical, thermal, and gamma-ray shielding properties of B_2O_3–SiO_2–BeO–MgO–La_2O_3, glasses, Opt. Mater. 150 (2024). https://doi.org/10.1016/j.optmat.2024.115298

[21] H.A. Ali , I.I. Turki , G.Sh. Al-Obaidy , J.M. Rzaij, Evaluating the gamma-ray attenuation characteristics of various radioactive sources using binary and ternary polymer blends, J. Compos. Adv. Mater. 35 (2025), pp. 935-945. https://doi.org/10.18280/rcma.350513

[22] S.G. Uzun1, H. Eroğlu, B. Bilmez, M.Ç. Tufan, M. Gürbüz, B.C. Turan, Gamma attenuation properties of $SrTiO_3$ doped with Rb and Y, Cumhuriyet Sci. J., 46 (2025), pp. 949-956. https://doi.org/10.17776/csj.1739508

[23] Z.A. Alrowaili, E.O. Echeweozo, S. Alomairy, A. Hammoud, C. Sriwunkum, N.S. Alsaiari, I. Boukhris , M.S. Al-Buriahi, Gamma attenuation, buildup factors, and radiation shielding performance of CaO-borosilicate glasses , J. Radiat. Res. Appl. Sci. 18 (2025). https://doi.org/10.1016/j.jrras.2024.101221

[24] M.W. Aladailah, M.W. Marashdeh, H. Akhdar , O.L. Tashlykov, Al-Tamimi Walaa and A.A. Ghazal, Investigation of the gamma and neutron radiation properties of strontium bismuth borate glasses doped with neodymium ions using the Geant4 simulation toolkit and EpiXS software, Radiat. Eff. Defects Solids, 178 (2023), pp. 485–499. https://doi.org/10.1080/10420150.2022.2153250

[25] M.I. Sayyed, O.I. Olarinoye, M. Elsafi, Assessment of gamma-radiation attenuation characteristics of Bi_2O_3–B_2O_3–SiO_2–Na_2O glasses using Geant4 simulation code, Eur. Phys. J. Plus, 136 (2021). https://doi.org/10.1140/epjp/s13360-021-01492-y

[26] A.A. SKIN, Evaluation of the gamma and neutron shielding properties of $64TeO_2 + 15ZnO + (20 - x)CdO + xBaO + 1V_2O_5$ glass system using Geant4 simulation and Phy-X database software, Pramana – J. Phys. 94 (2020). https://doi.org/10.1007/s12043-020-01972-3

[27] F.H. Alkallas, A.G. Trabelsi, S. Elaissi, T.A. Alrebdi, L. Abu El Maati, F. B. M. Ahmed, M. M. Mahasen, M. Ahmad, M. M. Soraya, Investigation of the gamma photon shielding in Se–Te–Ag chalcogenide glasses using the Phy-X/ PSD software, Cogent Eng. 9 (2022). https://doi.org/10.1080/23311916.2022.2116829

[28] H. Akhdar, Theoretical investigation of gamma- and neutron-shielding properties of polysulfone (PSU) polymer material using Geant4, Polymers, 14 (2022). https://doi.org/10.3390/polym14163374

[29] R.R. Bhosale, C.V. More, D.K. Gaikwad, P.P. Pawar, M.N. Rode, Radiation shielding and gamma ray attenuation properties of some polymers, Nucl. Technol. Radiat. Prot. 32 (2017), pp. 288-293. http://doi.org/10.2298/NTRP1703288B

[30] S.R. Manohara, S.M. Hanagodimath, L. Gerward, The effective atomic numbers of some biomolecules calculated by two methods: A comparative study, Med. Phys., Vol. 36 (2009), pp.137-141. https://doi.org/10.1118/1.3030952

[31] V. Zebardast, A.A. Saray, Experimental and simulation study of shielding features of polyester/ HgO composites against photons using GEANT4 and MCNP codes, Sci. Rep. 15 (2025). https://doi.org/10.1038/s41598-025-17942-x

[32] J.H. Hubbell, S.M. Seltzer, Tables of X-ray mass attenuation coefficients and mass energy absorption coefficients 1 keV to 20 MeV for elements Z = 1 to 92 and 48 additional substances of dosimetric interest. Natl. Inst. Stand. Phys. Lab, NISTIR; (1995). https://dx.doi.org/10.18434/T4D01F

[33] J.H. Hubbell, Photon mass attenuation and energy absorption coefficients. Int. J. Appl. Radiat. Isot. 33 (1982). pp.1269–1290. https://doi.org/10.1016/0020-708X(82)90248-4

[34] M.J. Berger, J.H. Hubbell, (XCOM) Photon cross section on a personal computer. National Bureau of Standards, Gaithersburg, MD, NBSIR 87-3597, Jul. 1987.

[35] L. Gerward, N. Guilbert, K.B. Jensen, H. Levring, X-ray absorption in matter, reengineering XCOM, Radiat. Phys. Chem. 60 (2001), pp. 23–24. https://doi.org/10.1016/S0969-806X(00)00324-8

[36] T.K. Kumar, K.V. Reddy, Effective atomic numbers for materials of dosimetric interest. Radiat. Phys. Chem. 50 (1997), pp. 545-553. https://doi.org/10.1016/S0969-806X(97)00089-3

[37] C.T. Chantler, Theoretical form factor, attenuation, and scattering tabulation for Z = 1–92 from E = 1–10 eV to E = 0.4–1.0 MeV. J. Phys. Chem. Ref. Data, 24 (1995), pp. 71- 643. https://doi.org/10.1063/1.555974

[38] Shivaramu, V. Ramprasath, Effective atomic numbers for photon energy absorption and energy dependence of some thermoluminescent dosimetric compounds. Nucl. Instrum. Methods Phys. Res., Sect. B, 168 (2000), pp. 294-304. https://doi.org/10.1016/S0168-583X(99)01100-3

[39] S. Gowda, S. Krishnaveni, T. Yashoda, T.K. Umesh, R. Gowda, Photon mass attenuation coefficients, effective atomic numbers and electron densities of some thermoluminescent dosimetric compounds. Pramana, 63 (2004), pp. 529-541. https://doi.org/10.1007/BF02704481

[40] S. Agostinelli, J. Allison, K. Amako, J. Apostolakis, H. Araujo, P. Arce, M. Asai, D. Axen, S. Banerjee, G. Barrand, et al. Geant4—A simulation toolkit. Nucl. Instrum. Methods Phys. Res., Sect. A, 506 (2003), pp. 250–303. https://doi.org/10.1016/S0168-9002(03)01368-8

[41] C.T. Chantler, Detailed tabulation of atomic form factors, photoelectric absorption and scattering cross section, and mass attenuation coefficients in the vicinity of absorption edges in the soft X-ray (Z = 30–36, Z = 60–89, E = 0.1 keV–10 keV), addressing convergence issues of earlier work. J. Phys. Chem. Ref. Data, 29 (2000), pp. 597-801. https://doi.org/10.1063/1.1321055

[42] V.P. Singh, M.E. Medhat, N.M. Badiger, Photon attenuation coefficients of thermoluminescent dosimetric materials by Geant4 toolkit, XCOM program and experimental data: a comparison study. Ann. Nucl. Energy, 68 (2014), PP. 96-100. https://doi.org/10.1016/j.anucene.2014.01.011

[43] Shivaramu, Effective atomic numbers for photon energy absorption and photon attenuation of tissues from human organs. Med. Dosim. 27 (2002), PP.1-9. https://doi.org/10.1016/S0958-3947(01)00078-4

[44] Z-G. Yin, H-Y. Qian, J. Jia, N. Zhou, L-Q. Feng, (E)-1-Benzylidene-2-(4-nitrophenyl)hydrazine, Acta Crystallogr., Sect. E 62 (2006), pp. o4913–o4914. https://doi:10.1107/S1600536806040372

Islamic University Journal of Applied Sciences (IUJAS)
https://journals.iu.edu.sa/jesc

Volume VIII, Issue II, December 2026, Pages 281-325

Recent Advances in Conformational Analysis and Spectral techniques, Compiled with Quantum Mechanical Calculations for Some Compounds

Usama. A. Soliman [1] *

[1] Chemistry Division, Department of Physical Sciences, College of Science, Jazan University, P.O. Box. 114, Jazan, 45142, Kingdom of Saudi Arabia.

*Corresponding author: (Usama A. Soliman) , *Email Address*: usoliman@jazanu.edu.sa

Abstract

Currently, quantum mechanical (QM) calculations are essential to support diverse chemical investigations by predicting electronic structures, reaction energies, and spectroscopic properties. Precise analysis of molecular conformations enables a comprehensive understanding of the structural, physicochemical properties of molecular systems. This decade has witnessed significant advancements in computational chemistry and spectroscopic techniques that have revolutionized conformational analysis, providing more accurate structural elucidation. Combination of QM calculations with experimental spectrum data (IR, NMR, UV-Vis, and Raman) has markedly improved the precision of conformational analysis and interpretation of spectral data. This review highlights significant advancements in conformational analysis, spectroscopic techniques, and their synergistic application with QM calculations, particularly density functional theory (DFT). The review presents case studies of chemically and biologically relevant molecules to illustrate the effectiveness of integrating experimental spectra and computational methods for conformational analysis and structural elucidation. This review also highlights the current limitations and proposes suggestions for future methodological improvements.

Keywords: Conformational analysis; Quantum mechanical calculations; IR/Raman; NMR; UV-Vis spectra.

https://doi.org/10.63070/jesc.2026.044
Received 20 May 2026; Revised 10 October 2026; Accepted 18 October 2026;
Available online 25 October 2026.
Published by Islamic University of Madinah on behalf of *Islamic University Journal of Applied Sciences.*
This is an open access article published under the Creative Commons Attribution 4.0 International license (CC BY 4.0): https://creativecommons.org/licenses/by/4.0/

1. Introduction

Conformational analysis and spectroscopy are essential tools for understanding the structures and properties of organic/inorganic compounds, pharmaceuticals, biomolecules, natural products, and polymers. These approaches enable researchers to ascertain the three-dimensional arrangement of atoms inside a molecule, which is essential for predicting physical, chemical, and biological aspects. Recent advances in computational chemistry, machine learning, and spectroscopic techniques have considerably improved the precision and efficacy of conformational analysis. The determination of molecular structure and conformational dynamics is a cornerstone of modern chemistry, with applications in drug design [1-3], catalysis [4, 5], and materials science [6-8].

X-ray crystallography and spectroscopic techniques (NMR, IR, Raman, UV-Vis, and MS) are the primary experimental methods for investigating molecular structure and conformations [9, 10]; however, they possess limitations. For example, X-ray crystallography requires high-quality single crystals, which may be difficult or impossible to obtain for several substances, including flexible molecules, amorphous materials, or proteins with low solubility [11]. Also, the accurate interpretation of the experimental vibrational and NMR spectra may be challenging, particularly for complex materials such as biomolecules and polymers, owing to band/signal overlap and the experimental conditions affecting the spectra [12, 13]. To overcome these challenges, quantum mechanical (QM) methods, particularly density functional theory (DFT), have become indispensable for predicting molecular conformations and their spectroscopic signatures [14]. NMR spectroscopy can distinguish between conformers, but it requires chemical shifts calculated by quantum mechanics to provide precise structural assignments [15]. Interpretation of complex peak patterns in IR and Raman spectra requires DFT-based vibrational frequency calculations [16]. UV-Vis and circular dichroism (CD) spectra are frequently analyzed by time-dependent (TD) DFT calculations to associate electronic transitions with molecular structure [17]. Computational chemistry involves the use of computer software to predict the electronic structure of a chemical species (atom, molecule, ion, or crystal) and its properties based on the concepts of quantum mechanics, molecular mechanics (MM), and molecular dynamics (MD) simulations [18]. This field of chemistry has diverse applications, including drug development, material science, and environmental chemistry, focusing on the prediction of molecular structures, energetics, and spectroscopic behavior [19]. Recent developments in computational chemistry allow for precise predictions of conformations and molecular properties, such as vibrational frequencies (IR/Raman), electronic transitions (UV-Vis), and NMR chemical shifts, which enable direct comparison with experimental data [17].

QM methods, such as DFT and *ab initio* calculations, can predict geometries without requiring crystallization, thereby allowing for structural investigations of amorphous or non-crystalline compounds. Also, MD simulations effectively model flexible systems in solution [20], providing a connection between solid-state (X-ray) and solution (NMR) structures. DFT frequency calculations and normal coordinate analysis enable accurate assignment of vibrational modes to their respective observed IR/Raman bands, which are significant in distinguishing conformers based on their fingerprint regions [21, 22]. DFT-based NMR chemical shift calculations have clarified ambiguities in the conformational preferences of flexible organic molecules [23]. TD-DFT calculations of electronic circular dichroism (ECD) spectrum aid in the determination of absolute stereochemistry [24].

Despite the advancements in QM calculations, challenges persist, including the accurate modeling of weak interactions (such as dispersion forces and hydrogen bonding) and large, flexible systems. Recent advancements in machine learning-augmented quantum mechanics methods, improved sampling algorithms, and hybrid QM/MM approaches are advancing the field of conformational analysis. The integration of experimental spectral data with QM calculations has demonstrated significant efficacy, enabling researchers to validate theoretical models and refine conformational predictions [25]. This synergy has become indispensable, with particularly transformative impacts in several chemical disciplines, including chemical synthesis, drug discovery, catalysis, and materials science; see Table 1.

This review explores the recent advances in conformational analysis, emphasizing the integration of spectroscopic techniques, including NMR, IR, Raman, and UV-Vis, with QM computations such as DFT, *ab initio*, and coupled-cluster approaches. This work highlights the significant impact of the QM-spectral combination across several fields, including materials research, drug development, and catalysis. A review of recent developments in QM methods for conformational studies will be conducted. The article reviews recent case studies that illustrate how this combined approach has yielded significant insights into conformational analysis. Finally, this review will subsequently address the challenges associated with this topic, offering a framework for future research in this field.

Table 1 Impacts of the integrated QM-experimental approach in different fields.

Field	Combined approach	Examples of impacts	Ref.
Organic/ Inorganic synthesis	QM/NMR/ IR/UV-Vis	• Support the structure investigated of novel compounds. • Regioselectivity and thermodynamically preferred reaction route. • Assigning the preferred coordination sites of ligands in metal complexes and studying structural isomerism.	[26-28]
Drug discovery	QM/MD/X-ray/NMR/ TD-DFT/ ECD	• Identifying bioactive conformations of drug candidates • Refining H-atom positions in protein active sites (e.g., hydrogen-bond networks in enzymes). • Resolving the dynamics of drug molecules in lipid bilayers.	[1-3]
Natural products	QM/NMR/ TD-DFT /ECD	• Structural elucidation of natural products. • Determination of the absolute configuration. • Mechanisms involved in natural product biosynthesis.	[24, 29]
Catalysis	QM/MD/ IR/Raman/ IRC	• Evaluation of multiple reaction pathways and identification of transient intermediates. • Elucidating transition-state geometries using Intrinsic Reaction Coordinate (IRC) calculations.	[4, 5]
Materials Science	DFT/ TD-DFT/IR/ UV-Vis	• Characterization of novel polymers and nanomaterials. • Designing organic semiconductors with tailored optoelectronic properties.	[6-8]

2. Conformational Analysis and Its Significance

Conformational analysis is a central aspect of structural chemistry that explores the spatial arrangements of atoms within a molecule and the differences in their energies. Molecular conformations denote different three-dimensional arrangements of atoms in a molecule that maintain the same covalent connectivity and can be classified into rotational isomers and cyclic conformations. The rotational isomers arise from the rotation around single bonds, exemplified by the conformers analyzed for L-histidine methyl ester [30], as illustrated in Figure 1. Whereas, cyclic conformations are formed due to ring puckering in cyclic molecules such as the three conformers (Figure 2) investigated for 1,3-dimethyl-2-imidazolidinone due to the flexibility of the five-membered ring [31]. Molecular conformations exist in dynamic equilibrium, with populations determined by their relative Gibbs free energies (ΔG) and temperature (T) according to the Boltzmann distribution [32]. The conformational landscape can be represented as a potential energy surface (PES), covering all possible conformations along with their relative stabilities [17].

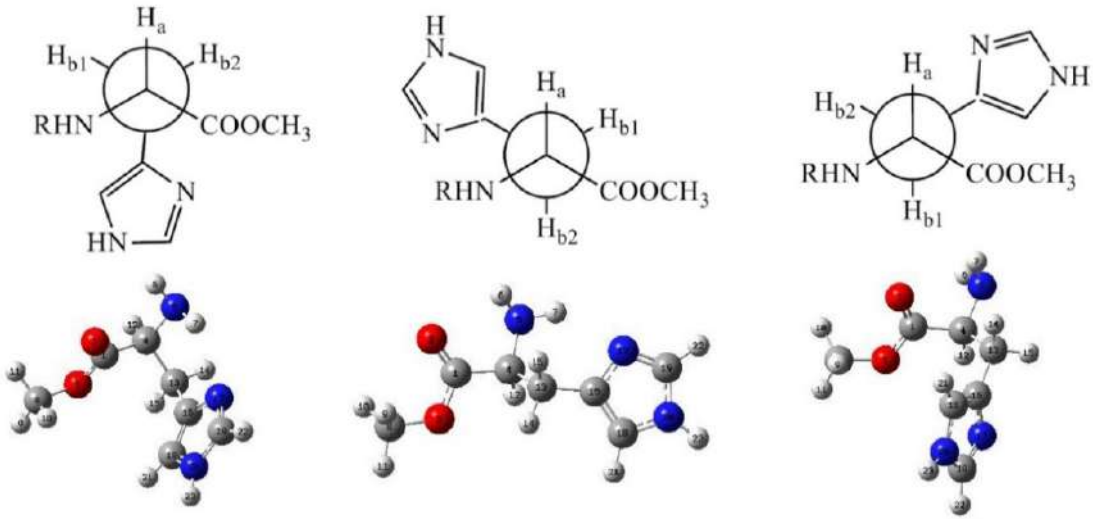


Figure 1 Newman projections representing the three rotational isomers (Top) of L-histidine methyl ester, R=H and their optimized geometries (bottom) at ωB97X-D/aug-cc-pVTZ level reported by Ref. **[30]**.

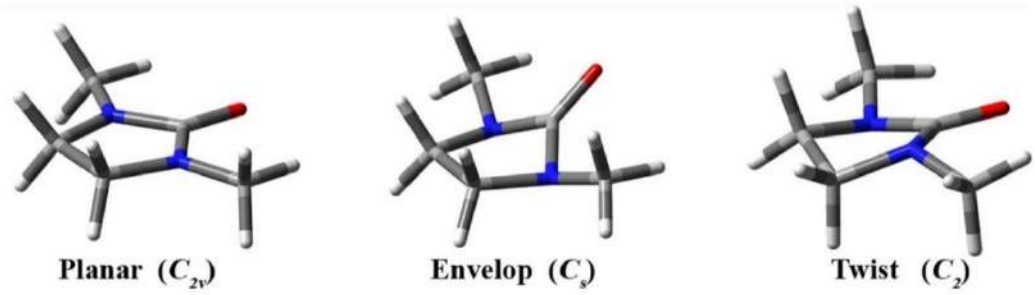


Figure 2 Equilibrium geometries of the proposed cyclic conformers of 1,3-Dimethyl-2-imidazolidinone using the MP2/6-311++G (d,p) calculations [31].

In 1932, Mizushima and Higasi began the historical development of conformational analysis by reporting the existence of *anti* and *gauche* conformers in 1,2-dichloroethane supported by Raman spectroscopy [33]. Barton significantly contributed in the 1950s by applying conformational analysis to cyclohexane derivatives, establishing a foundation for its use in natural product chemistry [34]. During the 1960s and 1970s, conformational analysis expanded into multiple disciplines, notably in the examination of cycloalkanes and alkaloids, emphasizing the distinct characteristics of each compound class [34, 35]. During the 1990s, techniques such as principal component analysis and essential dynamics analysis were developed to examine protein conformational ensembles, associating conformational alterations with functional mechanisms [36]. From the beginning of the year 2000, conformational analysis has experienced notable growth, primarily attributed to advancements in computational methods. The precision and efficiency of conformational studies improved significantly through the integration of experimental approaches and advanced computational methods. The

advancement of computational tools has enhanced researchers' understanding of molecular behavior and interactions, enabling them to analyze complex molecular systems [36, 37].

The stability of molecular conformers is influenced by several factors, including steric effects, intramolecular interactions, and environmental conditions. These factors collectively affect the energy landscape and the favored conformations in different environments. The stability of conformers is notably influenced by intramolecular hydrogen bonds; for example, N–H···O=C bonds increase the stability of *cis*-Amide conformers in proline-rich peptides [38]. Ring strain is a significant factor influencing the conformational preferences of cyclic molecules, such as cyclohexanes, which adopt chair conformations to reduce angle strain [39]. The stability of conformers is also affected by the orientation of substituents, such as axial and equatorial configurations. For example, 5-methyltetrahydro-1,3-oxazine favors an equatorial methyl group [40]. Concerning the environmental effect, the *gauche* conformer of butyronitrile exhibits greater stability under specific conditions of temperature and pressure [41].

Assessing conformational preferences is crucial for understanding molecular behavior, as slight structural alterations can substantially impact function. The study of molecular conformations is significant across various fields, such as drug discovery, molecular biology, materials science, and catalysis. Conformational analysis is essential in the design of drug molecules, as it identifies the optimal conformation for binding to a target protein [3]. The synthesis of natural products such as vitamin B12 highlights the significance of managing conformational preferences to attain specific chemical reactivities and biological activities [42], where conformational design principles are utilized to attain specific molecular structures. For enzymatic catalysis, studying conformational changes is vital for understanding enzyme mechanisms and developing inhibitors. These changes could overcome trade-offs between substrate specificity and product release, thereby enhancing catalytic efficiency [43]. In peptides, exploring the conformational preferences can inform the development of drugs that exhibit improved potency and selectivity. The incorporation of 2,3-methanopipecolic acids in peptides significantly increases the prevalence of the *cis* isomer, which can mimic proline's role in biological processes and assist in the design of novel bioactive compounds [44].

Conformational analysis has significance in material science, shown by its applications in many fields, including organic semiconductors, liquid-crystal, polymers, and metal-organic frameworks (MOFs). Using the 3D electron diffraction (3D ED), Ph-anti-benzothieno[5,6-b]benzothieno[3,2-b]thiophene, a novel organic semiconductor, exhibited complex conformational polymorphisms that influence its crystallization behavior and potential applications [45]. The exploration of liquid crystalline materials is facilitated by advanced computational techniques for rapid conformational analysis, enabling the development of new materials with specific properties [46]. Conformational

analysis elucidates the relationship between polymer structure and properties, facilitating the design of polymers with tailored characteristics for specific applications [47]. The porosity and surface area of MOFs are significantly influenced by the conformation of organic linkers, which are essential for applications such as catalysis and adsorption [48].

Conformational analysis provides valuable insights into molecular structures and dynamics; however, several critical unresolved issues remain. These include the accurate consideration of solvent effects [49], the handling of flexible structures with over ten rotatable bonds [50], and the prediction of temperature-dependent populations [51]. Recent advancements in machine learning force fields [52] and improved sampling techniques [53] are addressing these challenges, facilitating a more comprehensive exploration of conformational spaces.

3. Quantum Mechanical (QM) Methods

QM calculations are the cornerstone of computational chemistry, serving as a crucial tool for predicting and understanding the electronic structure and properties of molecules and materials [54], therefore aligning with both experimental and theoretical methodologies. Recent developments in computer hardware and software have greatly improved the potential of computational chemistry, allowing for more intricate and precise simulations [55]. Commonly used software packages for QM calculations include Gaussian [56], QChem [57], GAMESS [58], Psi4 [59], and NWChem [60]. All these software offer similar fundamental functionalities, with notable distinctions related to their user-friendliness, efficiency, and accessibility (free or commercial). The primary QM methods involve *ab initio* approaches (such as Hartree-Fock and post-Hartree-Fock), semiempirical methods, and DFT, each possessing its own characteristics and applications. The integration of QM with MM in QM/MM methods has broadened the applicability of QM calculations to larger and more complex systems.

Computational chemistry has progressed significantly since the first accurate simulations of gas-phase small molecules [61], shown by the receipt of two Nobel Prizes. In 1998, Walter Kohn and John Pople were awarded the first prize for their pioneering contributions to the development of DFT [62] and computational techniques in quantum chemistry [63]. The second prize was awarded in 2013 for the advancements of computational QM/MM models that simulate complex biological systems [64-66]. Currently, the extensive application of computational chemistry in medical and materials research is feasible [67, 68] due to advancements in methodologies and computational capabilities.

3.1 *Ab initio* Calculations

Ab initio calculations, derived from the Latin term meaning "from first principles" are QM methods that solve the Schrödinger equation without the use of empirical parameters, depending exclusively on fundamental physical constants. *Ab initio* Hartree-Fock (HF) calculations involve the approximation of the total molecular wavefunction Ψ as a Slater determinant of occupied spin orbitals.

The spatial orbitals are expressed as a linear combination of basis functions. HF calculations recover the bulk (≈99%) of the total electronic energy but neglects electron correlation[69], which is typically a few tenths of a hartree for small molecules (≈ 0.37 E_h for water). Because this missing energy is comparable to or larger than the energy differences of chemical interest, HF errors in reaction energies, barrier heights and non-covalent interactions can reach tens of kcal/mol, and harmonic frequencies are systematically overestimated by ~10% [70].

Post-HF approaches, including Møller-Plesset perturbation theory (MP), configuration interaction (CI), and coupled cluster (CC) methods, provide better treatment of electron correlation. These approaches methodically calculate electron correlation energy, necessary for precise electronic structure computations [71]. *Ab initio* calculations are employed to calculate molecular geometries, energies, vibrational frequencies, and spectra. These properties are crucial for understanding chemical reactions and interactions, including enzyme-substrate binding and reaction equilibrium [69].

3.2 Density Functional Theory (DFT)

DFT is currently one of the most widely used and well-established QM techniques for calculating electronic structures because it achieves an excellent compromise between computational cost and accuracy [72]. This theory is based on the Hohenberg-Kohn theorems, which state that all ground state properties of a system are determined by its electron density [62]. The practical applications of DFT involve approximations such as LDA and meta-GGA, especially concerning the exchange-correlation energy, a fundamental part of the theory. The precision of DFT calculations is significantly dependent on the selection of the exchange-correlation functional, and research continues to advance the development of more accurate functionals [73]. Recent advancements include hybrid and double-hybrid functionals, such as B3LYP [74, 75] and ωB97XD [76, 77], which enhance the accuracy of electronic structure predictions. Additionally, dispersion corrections, including DFT-D3 [76] and DFT-D4 [78] are applied for weak interactions in flexible molecules.

The HF theory has several deficiencies and inadequately describes most chemical reactions [70]. Both MP and CI methods provide accurate electronic energies for small atoms and molecules using large basis sets; however, they become too computationally expensive for larger systems. Modern DFT approaches offer greater precision compared to HF calculations with lower computational costs (Figure 3). Thus, DFT is preferred in handling large systems effectively, which makes it appropriate for investigating complex materials and biological systems. Owing to this effectiveness, DFT calculations have been extensively applied in cooperative experimental and computational studies [19].

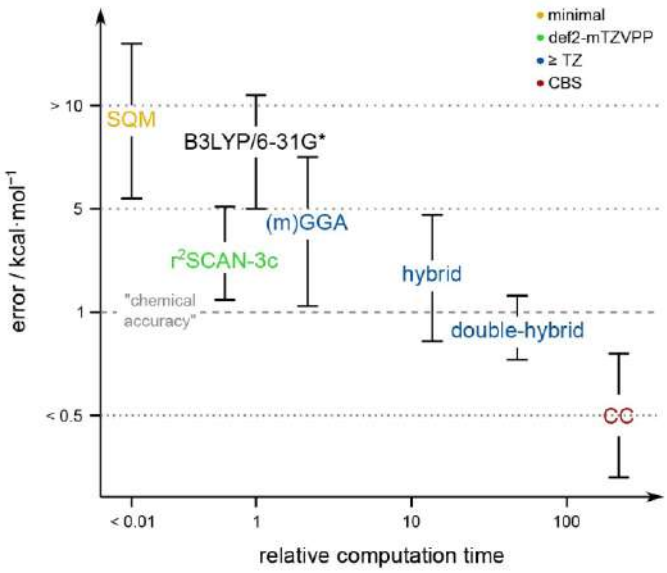


Figure 3. Accuracy in typical thermochemical applications vs. computation time (logarithmic scale) for common quantum chemistry methods [79].

3.3 General considerations in QM calculations

In computational chemistry, it is essential to deal with computational requirements with methodological accuracy and reliability. To achieve this objective, it is crucial to know the advantages/limitations of various approaches, consider the system size, and recognize the targeted properties, including structures, reaction energies, and conformational energies. Identifying suitable theoretical frameworks is essential for precisely characterizing a chemical system. Bursch and coworkers [79] have delineated important aspects to consider when conducting QM computations, the most notable of which are as follows:

1. A key issue in DFT is the selection of the exchange-correlation energy functional; hence, it is advisable to choose a functional carefully, considering the specific chemical system and the job at hand, rather than relying on popularity.
2. The application of the atomic-orbital basis set is essential. The key characteristic of basis sets is their completeness (size), which refers to the number of functions used to represent an electron. An appropriate basis set size (≥ triple-zeta; TZ) is crucial for energy-related features. It is essential to consider polarization functions for flexibility beside diffuse functions for anions, dipole moments, and polarizabilities. Finally, implement efficient core potentials for heavy elements (Z>36).
3. An additional significant aspect is the system's structural flexibility. Highly flexible structures may not be adequately represented by a single molecular property, such as energy or NMR spectra. It is recommended to evaluate the accessibility of relevant conformers for a specific system via conformational analysis.

4. Structure optimizations represent the beginning step of most QM studies, typically starting from either experimental data or an initial estimate using semiempirical or force-field methodologies. The most common user errors involve initial structures and input mistakes that lead to incorrect charge, spin, or protonation states, significantly altering the results. Consequently, the input must be carefully verified prior to initiating the computational procedure.
5. The calculation of vibrational frequencies is essential for determining zero-point vibration energies and for making thermostatistical corrections to enthalpy and free energy. It is essential for predicting IR/Raman spectra. For structure optimization and vibrational frequency calculations, it is necessary to employ the same theory level, or alternatively, a second set of optimized structures derived from this lower level must be utilized.
6. The calculated electronic energies serve as the principal indicator for the examination of conformers, reaction energies, barriers, and noncovalent interactions. These computations need larger basis sets for convergent results compared to structures or frequencies.

The expected performance of the most commonly used methods for conformational energies, NMR shifts, vibrational spectra and electronic spectra is compared in **Table 2**, together with their relative computational cost and recommended applications.

Table 2. Comparison of the main QM methods for conformational energies and spectroscopic properties

Method	Conformer energies	NMR shifts	Vibrational spectra	Electronic spectra	Cost	Recommended use
B3LYP-D3(BJ)	Good with dispersion correction; poor without it for folded/π-stacked structures	Reliable with GIAO; widely used	Very good after scaling	Valence states good; charge-transfer (CT) and Rydberg states underestimated	Low	Routine geometries, frequencies, NMR
M06-2X	Good for non-covalent interactions; grid-sensitive	Good	Good (own scale factors)	Better than B3LYP for CT owing to 54% HF exchange	Low–moderate	Energetics of flexible/dispersion-bound systems
ωB97X-D	Good; built-in dispersion	Good	Good after scaling	Good for CT/Rydberg (range separation)	Low–moderate	General-purpose, ECD/UV-Vis
MP2 (DLPNO-MP2)	Good but overbinds π-stacking/dispersion	Improves difficult cases; costly	Good for small systems	Not applicable	Moderate–high	Check on DFT energies, small/medium systems
CCSD(T) / DLPNO-CCSD(T)	Reference quality (“chemical accuracy”)	Benchmark only	Limited by cost	EOM-CCSD as benchmark	Very high	Single-point energies on DFT geometries
TD-DFT	–	–	–	Typically 0.2–0.3 eV for valence states; larger errors for CT, Rydberg, double excitations	Moderate	UV-Vis/ECD of medium-large molecules

3.4 Time-Dependent Density Functional Theory (TD-DFT)

TD-DFT computations expand traditional DFT to include time-dependent systems, facilitating the calculation of electronic excited states using linear response theory [80]. The approach provides excitation energies (often within 0.2-0.3 eV of experimental values), oscillator strengths, and charge-transfer characteristics [81]. TD-DFT is widely employed to predict the electronic circular dichroism (ECD) and UV-Vis spectra (electronic transitions) of organic and biological molecular systems [17]. Additionally, it is utilized to calculate the optical spectra of semiconductors and insulators, considering the impact of excitons, which are bound electron–hole pairs [82]. In the field of photonics, TD-DFT is helpful in examining the interaction between light and electrons, a crucial aspect for the development of devices that merge photonics with electronics [83].

The accuracy of TD-DFT depends strongly on the functional and on the nature of the excited state. The 0.2–0.3 eV error typically quoted applies to low-lying valence excitations of organic molecules. Global hybrids such as B3LYP underestimate charge-transfer and Rydberg excitation energies, an issue partly cured by range-separated hybrids (CAM-B3LYP, ωB97X-D). In the adiabatic linear-response approximation, states with significant double-excitation character are missing, and near-degenerate states may be mixed or wrongly ordered, which directly affects the sign pattern of ECD spectra. Solvent effects are described by linear-response PCM only approximately and neglect specific interactions, vibronic structure and band broadening are usually replaced by empirical line shapes, and ECD spectra are highly sensitive to conformer populations. Benchmarking against wavefunction methods (EOM-CCSD, DLPNO-STEOM-CCSD), testing several functionals, and Boltzmann-averaging over conformers are therefore recommended.

3.5 Potential Energy Surface (PES) Scans and Conformational Analysis

The combination of potential energy surface (PES) scan and conformational analysis with QM calculations provides significant understanding into molecular behavior. PES scans assess the variations in the energy of a molecule as a function of its geometric parameters, such as bond lengths, angles, or dihedral angles, to identify stable conformations and transition states [17]. A relaxed PES scan involves incrementally altering a particular structural parameter of the molecule followed by geometry optimization at each scanning point. The scan of PES can be carried out in either one or multiple dimensions, based on the complexity of the molecule. PES scans are essential for analyzing molecular flexibility, reaction pathways, and torsional barriers [56]. Developments in algorithms (e.g., CREST, meta-dynamics) and hybrid QM/MM methodologies persistently enhance efficiency and precision [84].

The conformational study of *meso*-2,3-dimercaptosuccinic acid was documented by Shaaban *et al.* [22], indicating that the *trans* conformer was preferred based on the PES scan during internal rotation around the central C–C bond, using B3LYP/6-31++G(d,p) calculations; refer to **Figure 4A**. The authors performed three PES scans for the *trans* conformer by rotating the *carboxyl*, *thiol*, and *hydroxyl* groups to explore additional conformers. Accordingly, the global minimum (**Figure 4B**) has the carbonyl *staggered* relative to the CCS angle but *trans* to C–H, whereas the O–H is positioned *cis* to the carbonyl, and the S–H is *trans* to C–H and *staggered* to the CCC angle.

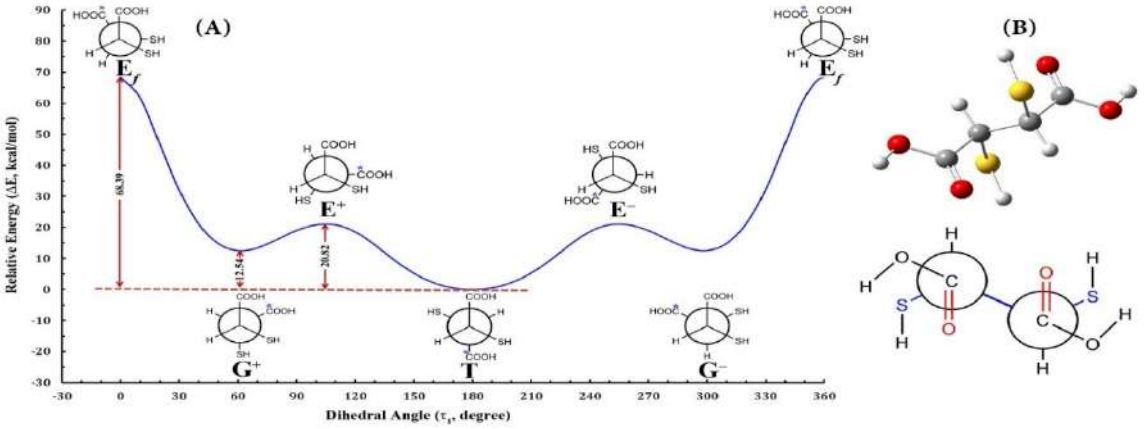


Figure 4. (A) PES scan during internal rotation around the central C–C bond for *meso*-DMSA obtained from B3LYP/6-311++G**[22]; **(B)** The global minimum for the *trans* conformer [22].

Employing B3LYP/6-311++G (d,p) calculations, Reference [31] documented a two-dimensional potential energy surface scan for 1,3-dimethyl-2-imidazolidinone by varying two dihedral angles (**Figure 5**) to determine the *planar*-to-*twist* barrier (42.6 kcal/mol).

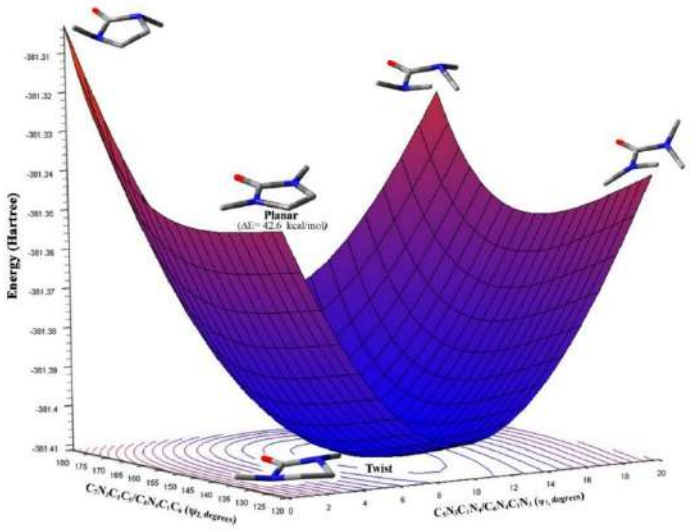


Figure 5. 2D-PES scan for ring puckering of 1,3-Dimethyl-2-imidazolidinone predicted from the B3LYP/6-311++G(d,p) calculations [31].

3.6 Machine Learning-Assisted Conformational Analysis

Machine learning (ML) is a crucial subset of artificial intelligence (AI) that focuses on the development of algorithms that allow computers to learn from data. This entails the application of statistical models for data analysis, pattern identification, and making predictions or decisions based on these patterns [85]. The use of ML in quantum mechanics calculations has revolutionized computational chemistry by accelerating the speed of accurate simulations and reducing computational costs. ML approaches, including neural networks, kernel methods, and Gaussian processes, are progressively used to predict molecular characteristics, optimize geometries, and substitute costly quantum mechanical calculations [86]. Grimme *et al.* developed a novel conformational searching approach, CREST, which offers a rapid alternative that rectifies inaccuracies seen in conventional conformational searching techniques [87]. CREST integrates the semiempirical GFN-xTB methods with metadynamics-based sampling to produce extensive conformational ensembles at a much-reduced cost compared to conventional *ab initio* techniques [88]. CREST is therefore not itself a machine-learning method, but an automated, semiempirical sampling tool that is often combined with ML-based refinement. The ML approaches relevant to conformational analysis are diverse and are best distinguished according to what the model predicts and how it is used. They can be classified into four categories: ML interatomic potentials/force fields, ML property prediction, ML–QM hybrids (Δ-learning), and ML-assisted conformational sampling, as summarised in **Table 3**. The main limitations of ML approaches should also be recognised: limited transferability beyond the training set, the scarcity of data for flexible and solvated systems, and the need for validation against QM calculations and experiment.

Table 3 Classification of machine-learning approaches used in conformational analysis

Category	Description	Examples
ML interatomic potentials / force fields	Neural-network or kernel potentials trained on DFT data to provide near-DFT energies and forces at force-field cost	ANI-2x [89], AIMNet2 [90], MACE-OFF [91]
ML property prediction	Models predicting spectra or shifts directly from structure or from low-level QM descriptors	iShiftML [110], CASCADE [92], DP4-AI [93]
ML–QM hybrids (Δ-learning)	ML corrects a cheap QM method towards a higher level (e.g., DFT → CCSD(T))	Δ-ML approaches reviewed in [86]
ML-assisted conformational sampling	Generative or active-learning models proposing conformers; ML potentials driving MD/metadynamics. Non-ML reference: CREST (GFN-xTB) [87, 88]	Torsional diffusion [94], GeoMol [95]

4. Integration of Spectral Techniques with QM Calculations

Spectral techniques, such as NMR, IR, Raman, UV-Vis, and CD spectroscopy, are indispensable for structural elucidation. Recent advancements in these approaches have markedly improved our ability to investigate molecular structure and conformations with extraordinary accuracy in exploring complex structures. **Table 4** highlights advanced spectroscopic methods and their applications in conformational analysis, enabling future advancements in structural biology and materials science.

The combined use of spectroscopic techniques and computational chemistry has become a cornerstone of conformational analysis. This synergy allows researchers to investigate the molecular structures and interpret spectral data with high accuracy. Barone *et al.* identified many aspects that make QM calculations an indispensable tool in the field of molecular spectroscopy [16]. The precise interpretation of spectral data requires comprehensive computational analysis. For instance, analyzing an experimental vibrational spectrum of a flexible molecule involves QM calculations, followed by spectral predictions and simulations to ascertain the relevant conformers. Experimental spectroscopy aims to elucidate the structure and bonding of molecules, with measurements focused on the frequencies of light absorbed by the molecular system. Computational spectroscopy serves as a connection between experimental data and fundamental physical properties by offering theoretical frameworks that relate observable measurements to molecular characteristics. QM calculations and experimental spectroscopy can serve as benchmarks for one another. Experimental spectroscopy exhibits high sensitivity to the electronic structure of a system, making it an effective method for verifying the reliability and accuracy of theoretical predictions and validating quantum chemistry calculations [16].

Table 4 Advanced Spectroscopic Techniques in Conformational Analysis

Spectral Technique	Applications
Raman optical activity (ROA) and VCD	Determination of absolute configurations and predominant conformations of several chiral compounds in solution [96].
2D-IR Spectroscopy	Monitoring ultrafast conformational dynamics, including protein folding and ligand binding [97].
Tip-enhanced Raman Spectroscopy (TERS)	Investigation of the conformational heterogeneity present in single molecules [98].
Ultrafast NMR spectroscopy	Monitoring conformational changes in real-time [99].

Multidimensional NMR	The analysis of the conformations of complex biological molecules, including proteins and RNA [100]
In-Cell NMR Spectroscopy	Conformational investigation of protein *in situ* within living cells [101]
Single-Molecule Fluorescence Spectroscopy	Investigation of protein folding and dynamics, yielding insights into conformational states and their dynamics [102].
Synchrotron Radiation CD (SRCD)	Investigation the secondary structure of proteins to understand its conformational behavior and stability at various conditions [103].

4.1 Vibrational Spectroscopy (IR/Raman)

Vibrational spectroscopy, involving infrared (IR) and Raman spectroscopy, is employed to analyze the vibrational modes of molecules, providing valuable insights into their structural and chemical properties. QM approaches are crucial for the accurate interpretation of the vibrational spectrum, which aids in conformational study. DFT approaches, particularly those using hybrid functionals like B3LYP [74, 75] and dispersion-corrected variants such as ωB97XD [76, 77], effectively predict IR and Raman spectra, while the use of anharmonic corrections (VPT2) improves accuracy [104].

The objective of the frequency calculations is to assign vibrational motions to the observed IR/Raman band. Normal coordinate analysis (NCA) is a method that effectively aids in the interpretation of the vibrational spectrum through mathematical representations of molecular vibrations using internal coordinates, which are well-known to spectroscopists. The process involves constructing a set of symmetry coordinates derived from combinations of internal coordinates, followed by the application of methods such as Wilson's FG matrix method for NCA [105]. The predicted potential energy distribution (PED) in the NCA indicates the percentage contribution of each symmetry coordinate to the calculated wavenumbers. In 2013, Jamróz [106] developed the computer code VEDA , vibrational energy distribution analysis, which performs NCA directly from the output file of the widely utilized Gaussian package in quantum chemistry. According to findings by Mohamed *et al.* [21], this algorithm cannot provide adequate accuracy to deliver comprehensive vibrational assignments for all fundamentals due to neglect of symmetry elements while mixing molecular vibrations. A comparison between the experimental vibrational spectra and theoretical predictions was essential for accurate assignment of the observed Raman and IR bands, as well as for a thorough understanding of the molecular geometrical structure. Consequently, the theoretical IR/Raman spectra can be simulated utilizing vibrational frequencies, Raman scattering activities, and infrared intensities derived from QM calculations. The theoretical framework for simulating IR and Raman spectra was

partially outlined by Mohamed *et al.* [107]. Shaaban and coworkers simulated the IR and Raman spectra for the most stable conformer of the *keto* 4-Methyl-2-hydroxyquinoline using DFT (B3LYP, ωB97XD, and mPW1PW91) calculations with a 6–311+G(2d,p) basis set [108]. Agreement in band positions after scaling and agreement in band intensities are different criteria. Harmonic frequencies depend mainly on the potential-energy curvature and are corrected by scale factors, whereas IR intensities depend on dipole-moment derivatives and Raman activities on polarizability derivatives, which are more sensitive to the functional and basis set (diffuse functions are particularly important for Raman). The double-harmonic approximation ignores anharmonic effects such as Fermi resonances, which redistribute intensity. Experimental conditions also matter: Raman intensities depend on the excitation wavelength and temperature (calculated activities must be converted to scattering intensities), and IR spectra recorded as solids or by ATR are affected by intermolecular interactions, packing and band-shape distortions, while matrix isolation shifts bands. The choice of line shape and width in simulated spectra further changes apparent intensities. Thus, the weaker simulated intensities below 1600 cm^{-1} in **Figure 6**, caused by the very high computed ν(C=O) intensity (860 km/mol), reflect both calculation limits and solid-state effects.

4.2 Nuclear Magnetic Resonance (NMR) Spectroscopy

Nuclear magnetic resonance (NMR) is a principal method used to elucidate the molecular structure of organic molecules. QM computations are essential for the assignment of NMR spectra and the investigation of molecular conformations, especially for flexible and complex molecules [23]. DFT using the Gauge-Including Atomic Orbital (GIAO) methodology is a widespread QM technique applied to calculate NMR parameters, such as chemical shifts and spin-spin coupling constants [104]. The DP4+ probability method is a significant technique that combines QM calculations with NMR data to improve the accuracy of conformational predictions [109]. In the context of structural elucidation, the implementation of ML models, such as the iShiftML, in computational NMR has resulted in exceptional accuracy in predicting chemical shifts at significantly reduced computational costs, thereby enhancing the correlation between experimental and calculated data to ascertain the most probable structures [110].

Comparative analyses of experimental and computationally predicted values for NMR chemical shifts and coupling constants have been utilized in various structural elucidation challenges. These include the determination of relative stereochemistry in flexible organic compounds [111], elucidation and reassignment of natural product structures [112], identification of byproducts in synthetic reactions [113], inference of macromolecular conformations in cyclic peptides [114], and correction of misassignments in the literature [115]. Shen *et al.* conducted a recent study analyzing

literature from the past decade, identifying over 200 cases of misassigned natural compounds [116]. NMR misassignment errors constituted the most common type of mistake, representing 23% of the total errors [117].

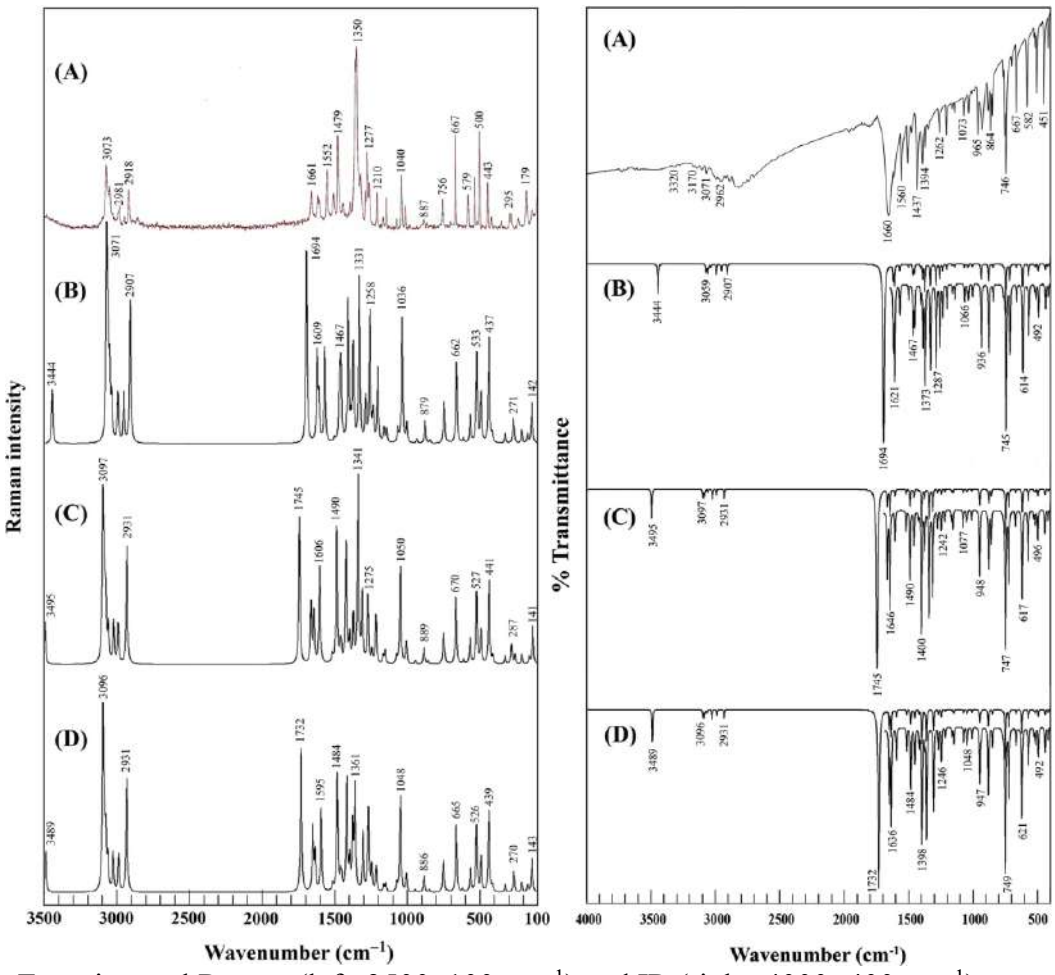


Figure 6. Experimental Raman (left, 3500–100 cm^{-1}) and IR (right, 4000−400 cm^{-1}) spectrum of 4-Methyl-2-hydroxyquinoline (A) compared to the simulated spectrum for *keto*- MQO using scaled frequencies from B3LYP (B), ωB97XD (C) and mPW1PW91 (D) methods at 6–311+*G* (2d,p) basis set [108].

Hexacyclinol is a natural product recognized as a significant example of utilizing NMR calculations for structural reassignment and elucidation. After its isolation and initial investigation of the molecular structure (**Figure 7A**), La Clair *et al.* in 2006 presented a total synthesis of hexacyclinol [118]. However, the measured NMR spectra of the synthetic product were inconsistent with the initially proposed structure, resulting in a controversy within the literature. Resolution was achieved when Rychnovsky proposed a revised structure (**Figure 7B**) according to calculations of ^{13}C NMR

chemical shifts [119]. That year, this structure was validated using total synthesis and X-ray crystallography [120]. This case illustrates that the calculation of NMR parameters can effectively prevent the misassignment of natural product structures, thereby conserving valuable time and resources that would then be spent on total synthesis for structural elucidation [112, 121].

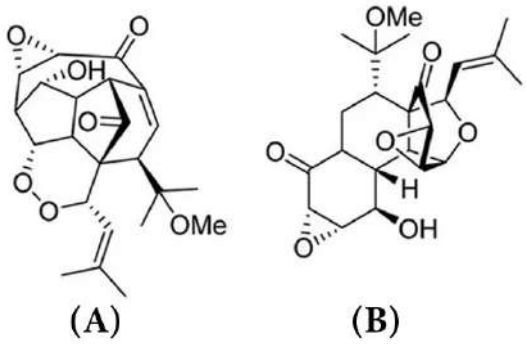


Figure 7. Originally proposed (A) and revised (B) structure of hexacyclinol [112].

4.3 UV-Vis and CD Spectroscopy

Ultraviolet-visible (UV-Vis) spectroscopy is a key analytical method employed to investigate electronic transitions in molecules, offering information about their electronic structure and conjugation. TD-DFT calculations with functionals such as B3LYP and CAM-B3LYP lead to accurate predictions of UV-visible spectra by calculating vertical electronic transitions and oscillator strengths, resulting in peak locations (λ_{max}) and intensities (ε), respectively [104]. Wavefunction-based approaches, like EOM-CCSD, provide enhanced precision for complex systems, but at a higher computational cost [122]. Electronic circular dichroism (ECD) spectroscopy is an effective chiroptical method for determining the absolute configuration of chiral molecules through the measurement of differential absorption of left- and right-circularly polarized light [123]. ECD spectra exhibit significant sensitivity to molecular conformation, thereby requiring TD-DFT calculations for different conformers to aid in precise assignments [124].

The UV-Vis spectral study of tetrahydroquinolines (THQs) revealed that TD-DFT effectively assigned long-wavelength absorption bands to $\pi \rightarrow \pi^*$ transitions, showing strong correlation with experimental findings [125]. The investigation of cyclohexane derivatives indicated that TD-DFT effectively simulates far-ultraviolet (FUV) spectra, elucidating the influence of *axial* and *equatorial* substitutions on electronic transitions, essential for comprehending conformational preferences [126]. The absolute configuration of racemic chlorpheniramine was found via chiral HPLC analysis and TD-DFT [127]. TD-DFT computations were used to examine the UV and CD spectra of selenium/tellurium chelates, yielding insights into their chiroptical characteristics [128].

5. Case Studies

This section presents four carefully selected cases in which spectral analysis, along with QM computations, yields precise conformational analysis.

The four case studies were selected because they (i) pair a quantitative experimental observable with calculations at an explicitly stated level of theory, (ii) together cover the main spectroscopic probes discussed in Section 4 (vibrational, including matrix-isolation, NMR chemical shifts and *J*-couplings, and chiroptical ECD/OR), (iii) span different molecular classes (organosilicon ring, small conjugated nitrile, a flexible drug, and a natural product) and (iv) illustrate different outcomes, namely agreement with the lowest-energy conformer (Sections 5.1–5.2) and cases in which the global minimum was not the experimentally observed species (Sections 5.3–5.4). Section 5.2 is based on the authors' earlier work [21]; it was included because it provides a complete vibrational (NCA/PED) assignment.

5.1 1–chloro-1-chloromethylsilacyclohexane

Recent research by Stocka *et al.* (2022) on 1-chloro-1-chloromethylsilacyclohexane (1-Cl-1-ClMSiCH) illustrates the integration of vibrational spectroscopy (Raman, ATR-FTIR, and matrix isolation FTIR) and QM calculations for conformational analysis [129]. This study theoretically proposed 38 canonical conformers based on ring conformation (chair, boat, envelope, or skew-boat), the *axial*/*equatorial* positioning of the Si–Cl bond, and the *cis*/*trans*/*gauche* orientations of Cl and CH_2Cl. Using DFT-B3LYP and *ab initio*-MP2 calculations utilizing the aug-cc-pVTZ basis set, the authors determined that the chair axial *trans* conformation (**Figure 8a**) represents the global energy minimum structure. With an energy increase of 1.04 kcal/mol, the chair equatorial *trans* conformer was the second most stable, followed by axial *gauche* (1.39 kcal/mol) and equatorial *gauche* (1.65 kcal/mol). Owing to the small energy differences, the authors stated that all four lowest energy chair conformers (**Figure 8**) coexist at room temperature, confirmed by the experimentally observed vibrational spectra. Through the comparison of experimental spectra with the calculated spectra of various conformers, the researchers successfully assigned specific bands to the four lowest energy chair conformers [129].

Figure 9 presents the ATR-FTIR and Raman spectra, along with those calculated for the most stable conformers. From ATR FTIR spectrum, the existence of the equatorial *trans* and axial *gauche* conformers is evidenced by weak shoulder observed at 791 cm^{-1} and the shape of the right slope of the 777 cm^{-1} band. The observed Raman band at 753 cm^{-1} supports the presence of low energy conformers, as no calculated axial *trans* conformer bands are found at this wavenumber. The authors performed matrix isolation experiments to clarify the overlapping bands in the ATR-FTIR spectrum. The matrix isolation spectra are shown in **Figure 10**, particularly in the 900-500 cm^{-1} range, where significant discrepancies among the computed spectra of the four conformers are evident.

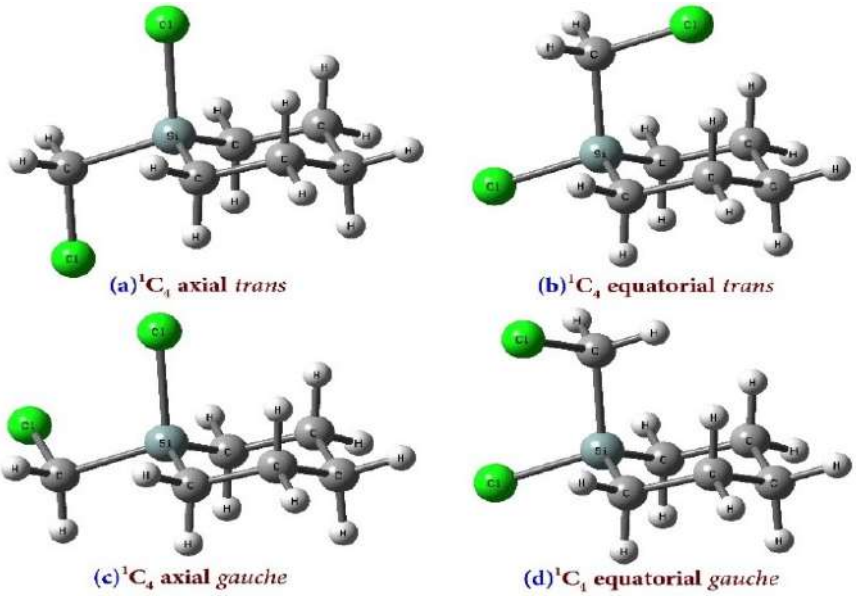


Figure 8. Chair conformers of the 1-Cl-1-ClMSiCH in axial/equatorial, *trans/gauche* conformation [129].

Three distinct bands were identified in the IR spectra of Ne/N_2 matrices at 804 cm^{-1}, 786 cm^{-1}, and 772 cm^{-1}, which are identified to the axial *trans* conformer and assigned to ring deformation modes and Si-C stretching. The very strong IR band detected at 778 cm^{-1} in nitrogen and 780 cm^{-1} in neon matches the calculated wavenumber of 786 cm^{-1} for the equatorial *gauche* conformer, attributed to the stretching vibrations of Si−C and C−C bonds.

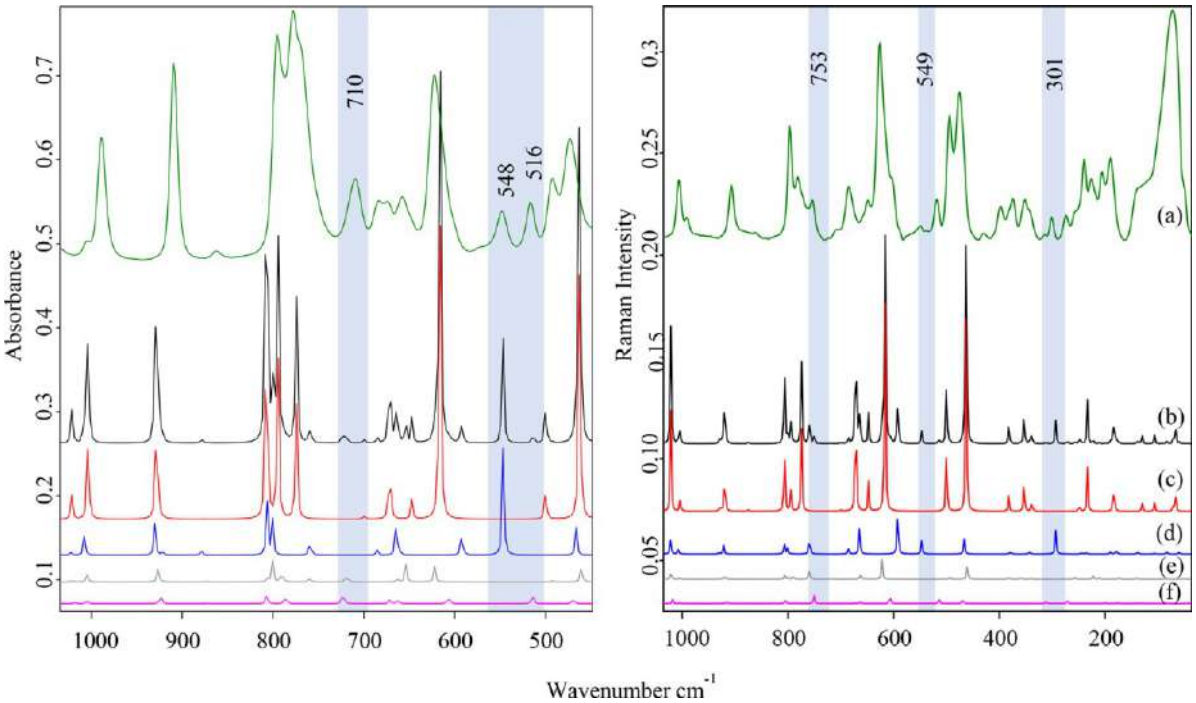


Figure 9. a) ATR (left) and Raman (right) experimental spectra of 1-Cl-1-ClMSiCH, calculated spectra of b) all conformers, c) axial *trans*, d) equatorial *trans*, e) axial *gauche*, f) equatorial *gauche*. This figure taken from Ref. [129].

The authors predicted the abundance of conformers by analyzing the integral intensity ratios of the spectral bands associated with different conformers. The nitrogen matrix concentrations are as follows: 81.3% for axial *trans*, 12.2% for equatorial *trans*, 2.6% for axial *gauche*, and 3.9% for equatorial *gauche*. The values are in qualitative agreement with those theoretically computed using the Boltzmann distribution at room temperature: 73%, 12%, 9%, and 6%, respectively. However, the theoretical populations (73/12/9/6%) differ appreciably from the matrix-derived ones, since matrix deposition may freeze non-equilibrium populations.

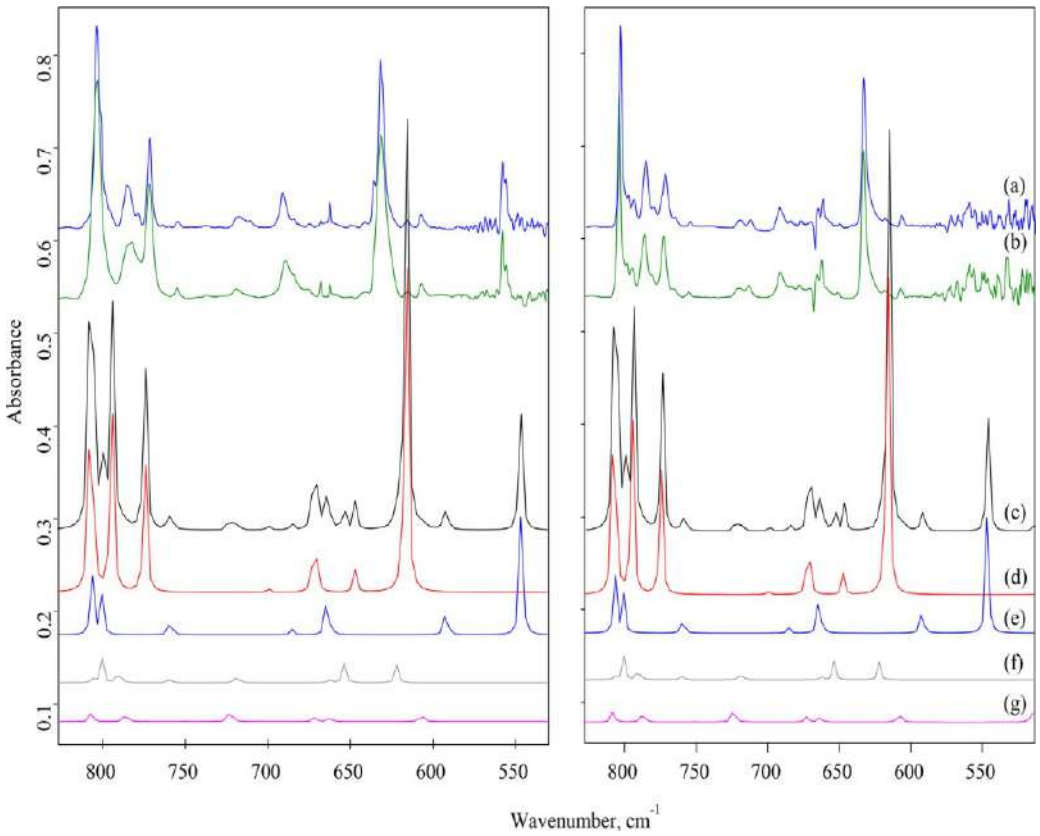


Figure 10 Infrared absorption spectra of 1-Cl-1-ClMSiCH isolated in neon at 3 K (left panel) and nitrogen (right panel) at 17 K, (b) after annealing at 9 K for neon and 35 K for nitrogen; and calculated spectra of (c) all conformers, (d) axial *trans*, (e) equatorial *trans*, (f) axial *gauche*, (g) equatorial *gauche*. This figure taken from Ref. [129]**.**

5.2 Centrosymmetric (2Z,4Z)-Hexa-2,4-dienedinitrile

The conformational analysis and detailed vibrational (IR/Raman) assignment for centrosymmetric (2Z,4Z)-Hexa-2,4-dienedinitrile (HDDN) was reported by Mohamed *et al.* [21], supported by DFT and *ab initio* calculations. As shown in **Figure 11**, the authors proposed three conformers for HDDN that correspond to the C_{2h} (*trans*), C_{2v}, and C_2 (*gauche*) point groups, attributed to the internal rotations around the central C−C single bond. DFT (B3LYP, M05-2X, and M06-2X)

and *ab initio* (MP2 and MP4) calculations utilizing 6-31G(d) and 6-311+G(d,p) basis sets indicated that the *trans* HDDN (C_{2h}) is the most stable conformer. The C_{2v} and *gauche* (C_2) conformers were determined to be energetically less favorable than the most stable C_{2h} by 12.18–13.70 and 6.51–8.18 kcal/mol, respectively. The C_{2v} conformer was identified as a transition state characterized by an imaginary wavenumber. These findings are confirmed by PES scans for internal rotation around the central C–C bond. The preference for the C_{2h} conformer coincides well with Raman and IR spectral analysis.

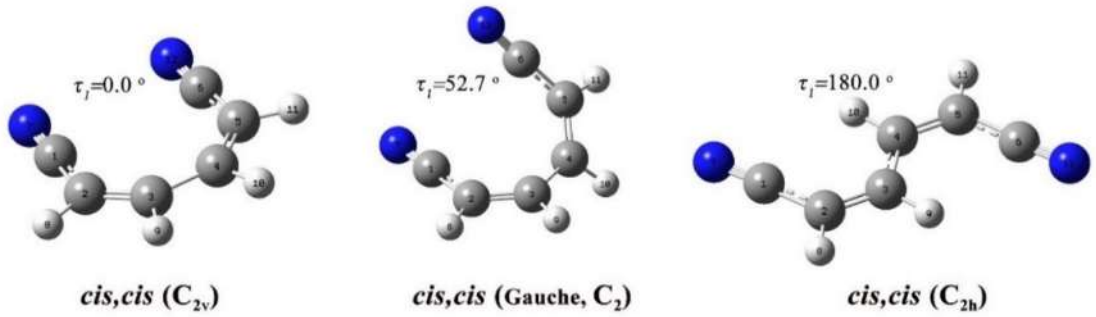


Figure 11. Equilibrium geometries of the proposed conformers of HDDN **[21]**.

This study confirmed the preference for the *trans* HDDN conformer in solution by comparing experimental and estimated spin-spin coupling constants [21]. The estimated/observed J value (H_9-H_{10}) of 10.19/10.8 Hz clearly supports the C_{2h} conformer in deuterated chloroform over the *gauche*-C_2 conformer (J = 4.81 Hz) after the exclusion of the transition state (C_{2v}).

The identification of the C_{2h} conformer was achieved through the analysis of observed Raman and IR spectra, in conjunction with the calculated vibrational frequencies. The single wavenumber is estimated at 999 cm^{-1} for the *trans* isomer, while three distinct estimated bands are observed at 998, 983, and 980 cm^{-1} for the *gauche* form. The strong Raman band observed at 962 cm^{-1} is neither split nor composite, indicating that it corresponds to the δ_{oop} C−H (ν_{12}) mode of the *trans* conformer. The authors presented comprehensive vibrational assignments of HDDN, utilizing NCA and PEDs in conjunction with calculated vibrational frequencies and force constants for the *trans* conformer exhibiting C_{2h} symmetry [21]. Consequently, 48 internal coordinates (**Figure 12**) were utilized to derive 42 symmetry coordinates (**Table 5**) employing the conventional Wilson GF method. Due to the mutual exclusion principle, we have categorized the thirty fundamental vibrations into 15 g modes (11 A_g + 4 B_g), which are only Raman active, and alone IR active 15 u modes (5 A_u + 10 B_u). **Tables 6/7** provide the observed Raman/IR bands along with their precise assignments to the normal modes of *trans* HDDN.

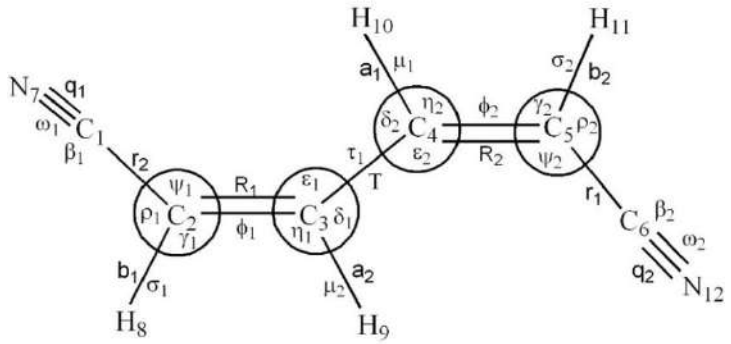


Figure 12. Atom numbering and internal coordinates definition of *trans* HDDN [21].

Table 5. Symmetry coordinates* for HDDN (*trans* conformer, C_{2h}) [21].

Species	ν_i	Definition	Symmetry coordinate*
A_g	ν_1	C–H symmetric stretch	$S_1 = a_1 + a_2 + b_1 + b_2$
	ν_2	C–H symmetric stretch	$S_2 = a_1 + a_2 - b_1 - b_2$
	ν_3	C≡N symmetric stretch	$S_3 = q_1 + q_2$
	ν_4	C=C symmetric stretch	$S_4 = R_1 + R_2$
	ν_5	C–$H_{terminal}$ *in-plane* bend	$S_5 = \rho_1 - \gamma_1 + \rho_2 - \gamma_2$
	ν_6	C–$H_{central}$ *in-plane* bend	$S_6 = \eta_1 - \delta_1 + \eta_2 - \delta_2$
	ν_7	C–$C_{central}$ stretch	$S_7 = T$
	ν_8	C–$C_{terminal}$ symmetric stretch	$S_8 = r_1 + r_2$
	ν_9	C=C–C *in-plane* bend	$S_9 = 2\varepsilon_1 - \eta_1 - \delta_1 + 2\varepsilon_2 - \eta_2 - \delta_2$
	ν_{10}	C–C≡N *in-plane* bend	$S_{10} = \beta_1 + \beta_2$
	ν_{11}	C=C–C *in-plane* bend	$S_{11} = 2\psi_1 - \rho_1 - \gamma_1 + 2\psi_2 - \rho_2 - \gamma_2$
B_g	ν_{12}	C–H *out-of-plane* bend	$S_{12} = \sigma_1 + \sigma_2 - \mu_1 - \mu_2$
	ν_{13}	C–H *out-of-plane* bend	$S_{13} = \sigma_1 + \sigma_2 + \mu_1 + \mu_2$
	ν_{14}	CCCC twist (C=C wag)	$S_{14} = \phi_1 - \phi_2$
	ν_{15}	C-C≡N *out-of-plane* bend (C–$C_{terminal}$ wag)	$S_{15} = \omega_1 - \omega_2$
A_u	ν_{16}	C–H *out-of-plane* bend	$S_{16} = \sigma_1 - \sigma_2 + \mu_1 - \mu_2$
	ν_{17}	C–H *out-of-plane* bend	$S_{17} = \sigma_1 - \sigma_2 - \mu_1 + \mu_2$
	ν_{18}	C-C≡N *out-of-plane* bend (C–$C_{terminal}$ wag)	$S_{18} = \omega_1 - \omega_2$
	ν_{19}	CCCC wag (C=C wag)	$S_{19} = \phi_1 + \phi_2$
	ν_{20}	C-$C_{central}$ torsion	$S_{20} = \tau_1$
B_u	ν_{21}	CH antisymmetric stretch	$S_{21} = a_1 - a_2 - b_1 + b_2$
	ν_{22}	CH antisymmetric stretch	$S_{22} = a_1 - a_2 + b_1 - b_2$
	ν_{23}	C≡N antisymmetric stretch	$S_{23} = q_1 - q_2$

Species	ν_i	Definition	Symmetry coordinate*
	ν_{24}	C=C antisymmetric stretch	$S_{24} = R_1 - R_2$
	ν_{25}	C–$H_{central}$ *in-plane* bend	$S_{25} = \eta_1 - \delta_1 - \eta_2 + \delta_2$
	ν_{26}	C–$H_{terminal}$ *in-plane* bend	$S_{26} = \rho_1 - \gamma_1 - \rho_2 + \gamma_2$
	ν_{27}	C-$C_{terminal}$ antisymmetric stretch	$S_{27} = r_1 - r_2$
	ν_{28}	C=C–$C_{terminal}$ antisymmetric bend	$S_{28} = 2\psi_1 - \rho_1 - \gamma_1 - 2\psi_2 + \rho_2 + \gamma_2$
	ν_{29}	C-C≡N antisymmetric bend	$S_{29} = \beta_1 - \beta_2$
	ν_{30}	C=C–$C_{central}$ antisymmetric bend	$S_{30} = 2\varepsilon_1 - \eta_1 - \delta_1 - 2\varepsilon_2 + \eta_2 + \delta_2$

*Symmetry coordinates are not normalized, for internal coordinates definitions, see **Figure 12**.

Table 6. Observed Raman bands and calculated symmetric wavenumbers (A_g and B_g) for *trans* HDDN [21].

Sp.	ν_i	Approximate description [a]	B3LYP/6-311+G(d,p)					Observed Raman
			Unscaled	Scaled [b]	R act. [c]	Rel. Int. [c]	PED	
A_g	ν_1	ν_s C–H	3187	3041	252	0.65	$87S_1$ $12S_2$	3050 sh, w
	ν_2	ν_s C–H	3166	3021	39.7	0.1	$88S_2$ 12 S_1	3045 w
	ν_3	ν_s C≡N	2321	2226	1171	5.14	$89S_3$ $11S_8$	2216 m
	ν_4	ν_s C=C	1665	1647	1405	10	$60S_4$ $20S_6$ $11S_7$	1613 vs
	ν_5	δ_{ip} C–H_{term}	1452	1395	21.8	0.19	$45S_5$ $32S_6$	-
	ν_6	δ_{ip} C–H_{cent}	1277	1230	336.6	3.41	$45S_6$ $23S_5$ $17S_7$ $28S_7$ $27S_5$ $11S_8$	1236 s
	ν_7	ν_s C–C_{cent}	1117	1093	8.5	0.1	$18S_9$	-
	ν_8	ν_s C–C_{term}	894	888	18.2	0.29	$60S_8$ $12S_4$ $12S_7$ $26S_9$ $22S_7$ $13S_{10}$	880 m
	ν_9	δ_{ip} C=C–C_{cent}	762	754	0.5	0.01	$28S_{11}$	-
	ν_{10}	δ_{ip} C-C≡N	407	404	6.7	0.3	$52S_{10}$ $32S_9$ $53S_{11}$ $13S_9$	403 m
	ν_{11}	δ_{ip} C=C–C_{term}	136	135	20.6	5.16	$32S_{10}$	156 m
B_g	ν_{12}	δ_{oop} C–H	999	999	13.2	0.18	$100S_{12}$	962 m
	ν_{13}	δ_{oop} C–H	819	819	2.8	0.05	$88S_{13}$	801 w
	ν_{14}	δ_{twist} CCCC	616	616	13.2	0.34	$84S_{14}$ $15S_{15}$	592 s
	ν_{15}	γ C–C_{term}	321	320	0.6	0.04	$63S_{15}$ $35S_{14}$	310 m

[a] Notations for fundamentals: s, symmetric; as, antisymmetric; ip, in-plane; ν, stretch; δ, bending and γ, wagging.
[b] Scaling factors of 0.91 and 0.90 for CH/C≡N stretches and δ_{ip} CH, respectively; 1.0 for heavy atom stretches and bends; 1.0 for γ CH.
[d] Calculated Raman activities in $Å^4$/mol. Relative intensities were scaled out of 10 from NCA.

Table 7. Observed infrared bands and calculated anti-symmetric wavenumbers (A_u and B_u) for *trans* HDDN [21].

Sp.	ν_i	Approximate description [a]	B3LYP					Observed IR
			Unscaled	Scaled [b]	IR int. [c]	Rel. Int. [c]	PED	
A_u	ν_{16}	δ_{oop} C–H	1025	1025	1	0.14	$100S_{16}$	982 m
	ν_{17}	δ_{oop} C–H	784	784	68.3	10	$71S_{17}$ $10S_{18}$ $60S_{18}13S_{17}$	750 vs
	ν_{18}	□ C–C_{term}	478	478	3.8	0.56	$32S_{19}$ $53S_{19}$ $25S_{18}$	468 s
	ν_{19}	τ CCCC	248	248	9.7	1.42	$18S_{20}$ $66S_{20}$	245 s
	ν_{20}	τ $CCCC_{cent}$	57	57	11.9	1.75	$10S_{17}23S_{19}$	-
B_u	ν_{21}	ν_{as} C–H	3189	3043	1.7	0.25	$97S_{21}$	3072 m
	ν_{22}	ν_{as} C–H	3174	3028	7.7	1.13	$98S_{22}$	3057 m
	ν_{23}	ν_{as} C≡N	2322	2227	39	5.71	$89S_{23}$ $11S_{27}$	2225/2213 s
	ν_{24}	ν_{as} C=C	1615	1606	4.7	0.69	$82S_{24}$	1573/1555 m
	ν_{25}	δ_{ip} C–H_{cent}	1381	1316	3.9	0.57	$54S_{25}$ $37S_{26}$	1346 m
	ν_{26}	δ_{ip} C–H_{term}	1233	1176	2	0.29	$53S_{26}$ $36S_{25}$	1200 s
	ν_{27}	ν_{as} C–C_{term}	965	955	11.8	1.73	$65S_{27}$	942 s
	ν_{28}	δ_{ip} C=C–C_{term}	644	638	1.1	0.15	$49S_{28}$ $26S_{29}$	632 s
	ν_{29}	δ_{ip} C-C≡N	328	325	0.3	0.04	$51S_{29}$ $42S_{30}$	323 s
	ν_{30}	δ_{ip} C=C–C_{cent}	100	99	16.4	2.4	$46S_{30}$ 35_{28} $20S_{29}$	-

[a] Notations for fundamentals: s, symmetric; as, antisymmetric; ip, in-plane; ν, stretch; δ, bending and γ, wagging.
[b] Scaling factors of 0.91 and 0.90 for CH/C≡N stretches and δ_{ip} CH, respectively; 1.0 for heavy atom stretches and bends; 1.0 for γ CH.
[c] Calculated infrared intensities in km/mol. Relative intensities were scaled out of 10 from NCA.

5.3 Chloroquine (CQ) drug

Hernandes and coworkers reported the conformational study of Chloroquine (CQ, **Figure 13**) in solution by DFT calculations and NMR spectral analysis in three solvents: water, chloroform, and DMSO [130]. CQ has served as a standard antimalarial drug since the early 1950s and has received approval from health authorities in certain countries for its experimental use in treating COVID-19. The authors investigated 30 trial conformers for CQ through a comprehensive conformational search, varying eight dihedral angles (as shown in **Figure 13**) to identify all plausible conformations that may exist in solution. A total of 30 proposed molecular conformations were fully optimized using the ωB97x-D/6-31G(d,p)-PCM calculations.

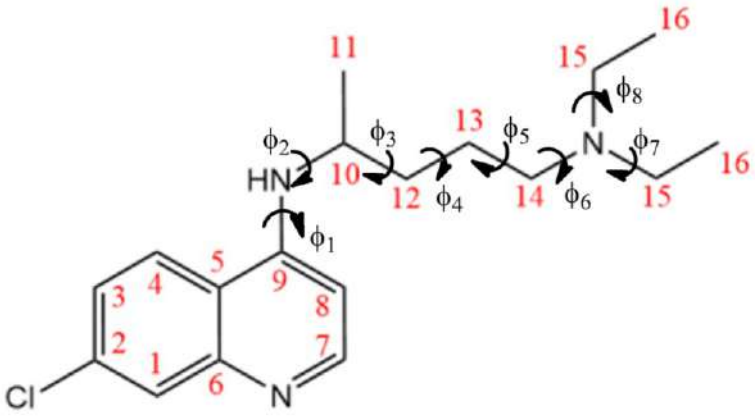


Figure 13. Structural formulas of chloroquine and the numbering adopted. Torsion angles (ϕ_1, ϕ_2, ϕ_3, ϕ_4, ϕ_5, ϕ_6, ϕ_7 and ϕ_8) are relevant for conformational analysis [130].

The optimized geometries for all suggested conformations (**M1-M30**) are shown in **Figure 14**. Using DFT calculations, ^{1}H NMR chemical shifts for **M1-M30** conformers were computed to identify the best match with those experimentally observed in solutions. The computational results indicate that conformer **M24** represents the global minimum-energy structure (lowest relative ΔG value) across all three solvents. It is stabilized by intramolecular hydrogen bonding between the hydrogen of the N−H group and a nitrogen atom attached to two alkyl groups.

The experimental ^{1}H NMR data were acquired in three deuterated solvents, D_2O, $CDCl_3$, and DMSO-d_6, which served as references for comparison with computed chemical shifts. This approach aims to identify the best correspondence with experimental results through the analysis of RMSD values and the experimental ^{1}H NMR profile. **Figure 15** illustrates the experimental ^{1}H NMR spectra in comparison with those computed for selecting optimized conformers. Accordingly, the global minimum-energy structure, **M24**, was not validated in all solvents owing to its inconsistent NMR profile, notably its undetected large chemical shift for N–H (~9 ppm). As shown in **Figure 15**, the optimized conformation **M12** demonstrates the most accurate alignment with the observed ^{1}H NMR profile in D_2O. However, structure **M8** exhibits a nearly perfect match with the experimental ^{1}H NMR spectrum in chloroform and DMSO (**Figure 16**), demonstrating the solvent's influence on conformational preferences. These findings indicate that, for chloroquine, spectroscopic analysis (NMR matching) was a more discriminating criterion than the calculated relative free energies for ascertaining solution-phase conformations, underscoring **M12** and **M8** as significant structures for forthcoming drug interaction investigations.

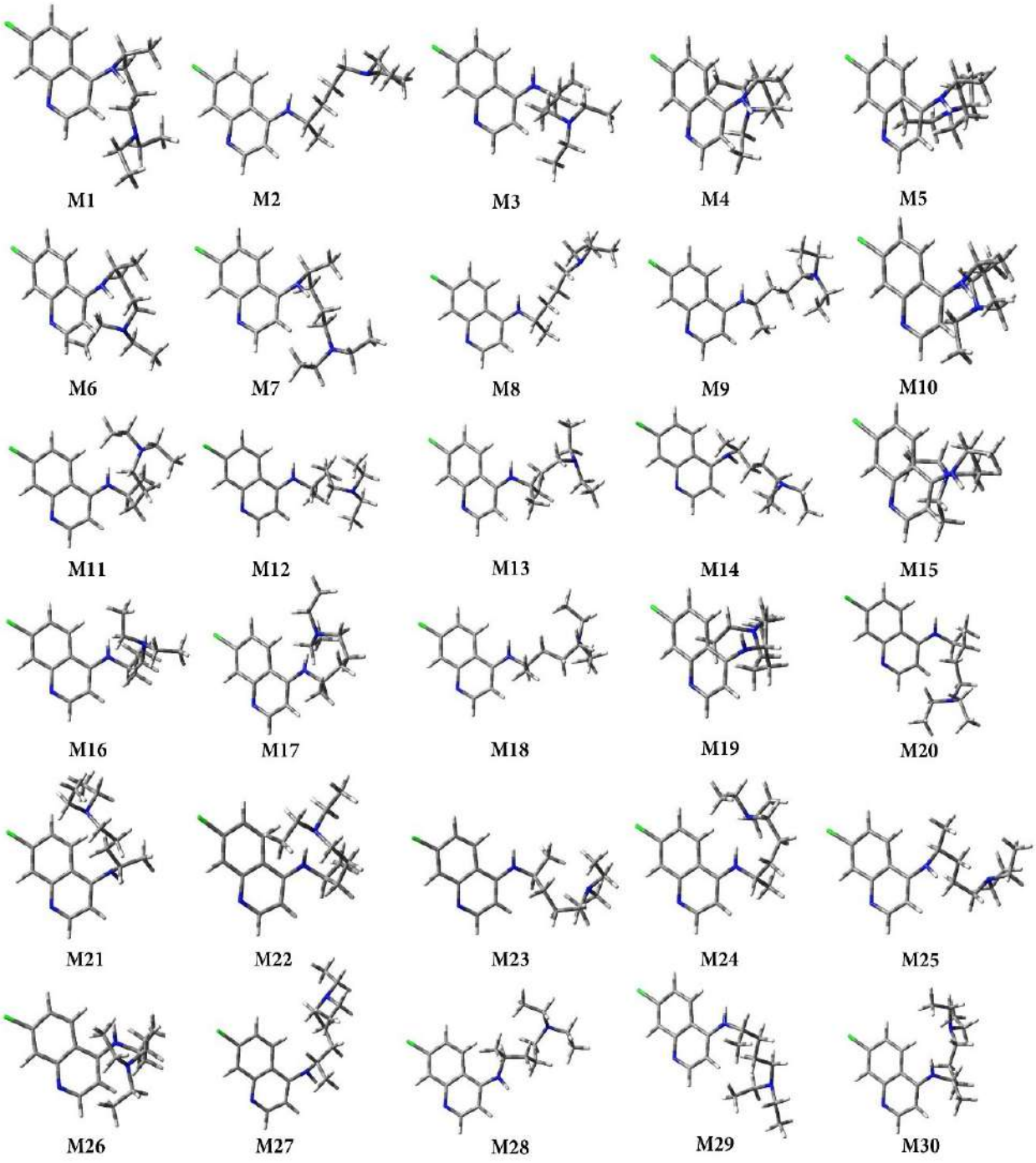


Figure 14. ωB97x-D/6-31G(d,p)-PCM-Water fully optimized structure of chloroquine (conformations **M1**–**M30).** This figure sketched using structure form supplement materials of Ref. [130]

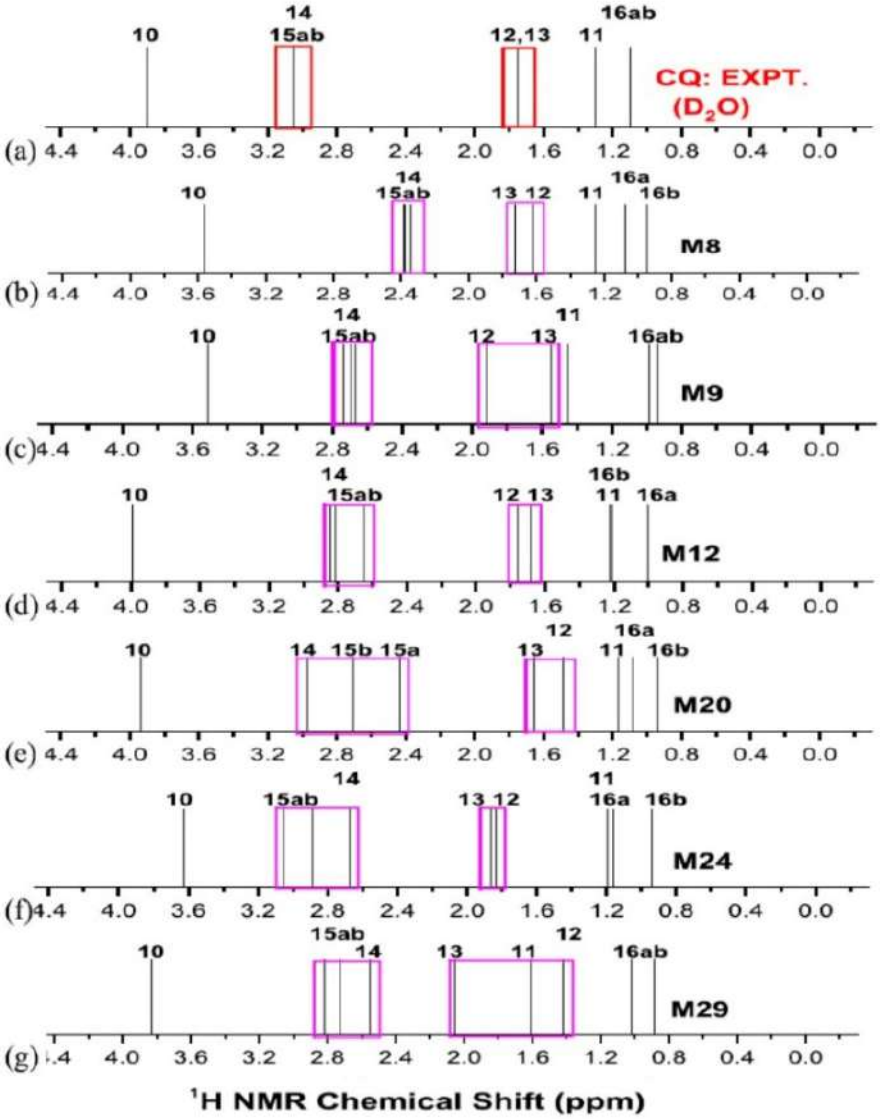


Figure 15. Experimental in D_2O (a) and B3LYP/6-31G(d,p)−PCM−water ^{1}H NMR spectra (b−g) for selected optimized structures of CQ. This figure taken from Ref. [130].

5.4 Absolute Configuration of Fusariumin A

The absolute configuration of Fusariumin A has been confirmed to be (2'R,3'R) through a combination of NMR techniques utilizing residual dipolar coupling (RDC) and chiroptical ECD spectroscopy, supported by TD-DFT calculations [131, 132]. Fusariumin A (**Figure 17A**) is an alkylpyrrole derivative extracted from the fungus Fusarium sp.

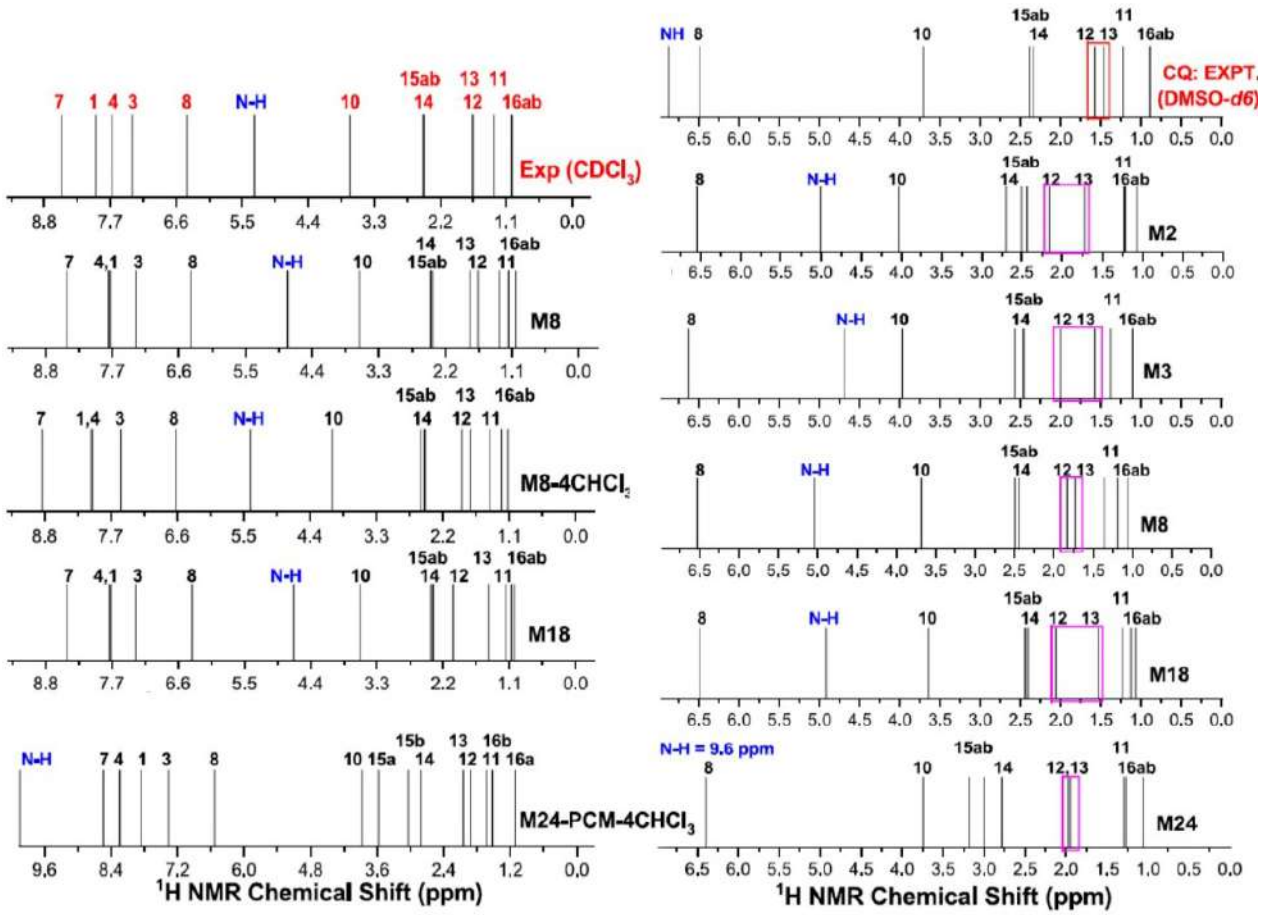


Figure 16. Experimental (left in $CDCl_3$; right in DMSO-d_6) and calculated ^{1}H NMR spectra for various optimized structures of CQ. This Figure taken from Ref. [130]

This compound includes two adjacent stereocenters within a flexible alkyl chain, and their relative configuration could not be established through conventional NMR methods. A systematic conformational search discovered the possible conformers of the 2'R,3'R diastereomer of Fusariumin A. Considering the high flexibility of the alkyl chains in Fusariumin A, this search focused solely on the rotations around the three bonds nearest to the stereocenters, while maintaining the two alkyl chains in the most stable all-*trans* conformation. The rotatable bonds included in the conformational search are indicated in red in Figure 17A. This technique yielded 12 conformers for the 2'R,3'R diastereomer. Of these, only conformer 11 (Figure 17C) demonstrated a satisfactory fit with the experimental RDCs and was consistent with all measured NOEs and J-couplings [131]. The conformer with the lowest energy, as determined by DFT calculations, was conformer **2**; nevertheless, it was incorrect since it did not correspond with the observed RDCs and NOEs (Figure 17B). Conformer **11** was the second-lowest energy structure, with an energy increase of 1.04 kcal/mol (ΔG = 1.1 kcal/mol). TD-DFT computations (B3LYP/6-31G*) were used to predict the ECD spectra of conformations **2** and **11** in DMSO employing the IEFPCM solvation model, in order to ascertain the exact configuration. The

predicted ECD spectrum for conformer **11** exhibited a negative ECD band that coincided closely with the experimental spectrum (**Figure 17D**), thus confirming the configuration of conformer **11** for natural Fusariumin A. The predicted ECD spectrum of conformer **2**, identified as the lowest energy conformer through DFT calculations, exhibits an opposite sign in comparison to that of conformer 11 (**Figure 17E**).

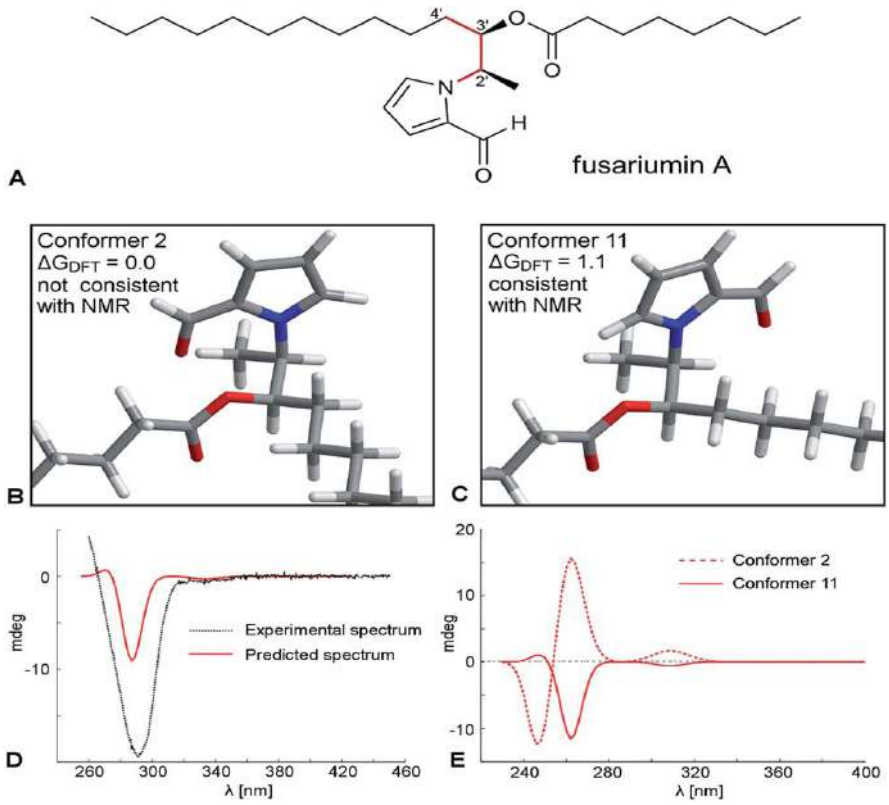


Figure 17. (A) Structure of Fusariumin A. (B) conformer **2** (C) conformer **11** (D) Experimental ECD spectrum of Fusariumin A as compared to the predicted ECD spectrum of the (2'R,3'R) enantiomer. (E) Predicted ECD spectra for Conformers **2** and **11**. This figure adapted from Ref. [132].

These findings have been verified by the optical rotation (OR) predictions at five distinct wavelengths. The anticipated OR values for conformer **11** closely aligned with the observed results, however those for conformer **2** had an opposite sign.

Populations follow from the Boltzmann distribution of relative free energies, so the energy type matters: electronic energies neglect vibrational and entropic contributions; Δ(E+ZPE) corrects for zero-point effects; and ΔG adds thermal enthalpy and entropy, normally from harmonic frequencies within the rigid-rotor/harmonic-oscillator approximation at the stated temperature. Low-frequency modes (<100 cm^{-1}) can introduce errors of several kcal/mol in entropy and are often treated with a quasi-harmonic correction. Because a 1 kcal/mol error changes a population ratio by a factor of ~5 at 298 K, for example, that the theoretical populations in Section 5.1 (73/12/9/6%) differ appreciably

from the matrix-derived ones (81.3/12.2/2.6/3.9%), since matrix deposition may freeze non-equilibrium populations.

These case studies, in particular chloroquine [130] and Fusariumin A [131, 132], show that agreement between Boltzmann-weighted calculated and experimental spectra is a necessary complement to, and in some cases a more discriminating criterion than, calculated relative energies. Calculated stabilities remain essential because they depend on the level of theory, the solvent model and the thermal corrections, whereas spectra may also mismatch for these same reasons. The most reliable assignments (e.g., DP4/DP4+ [32, 133], hexacyclinol [119, 120]) combine both criteria, and detailed protocols for the calculation of NMR chemical shifts are available [134].

6. Practical Guidelines for Method Selection

This section summarises typical starting points for selecting QM methods, basis sets and spectroscopic techniques for different classes of compounds **(Table 8).** These are typical starting points that should be validated against a higher-level reference for the system at hand.

Table 8. Practical guidelines for the selection of QM methods and basis sets

Task / property	Suggested level of theory	Suggested basis set	Key practical points
Conformer search	CREST (GFN2-xTB + metadynamics) or force-field/MD; then DFT re-optimisation	GFN2-xTB; then def2-SVP / 6-31G(d,p)	Use an energy window of ~5–6 kcal/mol; remove duplicates; check convergence for >10 rotatable bonds
Geometry + frequencies	Hybrid or range-separated hybrid DFT with dispersion (B3LYP-D3(BJ), ωB97X-D, M06-2X) or r^{2}SCAN-3c	Double-zeta with polarisation to def2-TZVP	Same level for optimisation and Hessian; confirm no imaginary frequencies; apply scale factors
Relative conformer energies	DLPNO-CCSD(T) or double-hybrid DFT single points on DFT geometries; dispersion-corrected hybrid DFT if cost-limited	Triple-/quadruple-zeta (def2-TZVPP, aug-cc-pVTZ) or CBS extrapolation	Include ZPE and thermal corrections to obtain ΔG; address BSSE (gCP/counterpoise)
NMR shifts and *J* couplings	GIAO with e.g. mPW1PW91, ωB97X-D or WP04 (^{1}H); DP4+ analysis	6-311+G(2d,p), def2-TZVP or pcS-*n*	Boltzmann-average all relevant conformers; implicit solvent matching the experiment; linear scaling
IR / Raman	Hybrid DFT (B3LYP-D3, ωB97X-D); VPT2 where anharmonicity matters	6-311+G(d,p), def2-TZVP (diffuse functions for Raman activities)	Compare frequencies and intensities separately; PED/NCA for assignment; matrix or solid-state effects
UV-Vis / ECD	TD-DFT with range-separated hybrids (CAM-B3LYP, ωB97X-D) or PBE0; EOM-CCSD / DLPNO-STEOM-CCSD for benchmarking	def2-TZVPD, aug-cc-pVDZ, 6-311+G(d,p)	Include enough states; Boltzmann-weight conformers; check charge-transfer character; implicit/explicit solvent

In terms of compound classes, small rigid organics are adequately treated by a single DFT workflow; flexible molecules and drug-like compounds require a conformer search plus Boltzmann averaging; chiral natural products need ECD/OR/NMR (DP4+) combined analysis; and metal complexes need functionals and effective core potentials benchmarked for the metal in question.

7. Challenges, Future Perspectives and Conclusions

Despite considerable advancements in the integration of conformational analysis, spectroscopic methods, and quantum mechanical (QM) computations, numerous challenges remain. Addressing these limitations is essential for improving accuracy, efficiency, and applicability in various chemical systems. A key challenge is the balance between accuracy and computational cost. Advanced QM methods involve CCSD and CASSCF, providing highly precise results; nonetheless, they are computationally expensive for large molecules with more than 50 atoms. DFT functionals, such as B3LYP, M06-2X, and ωB97XD, enable a reasonable balance; however, they may be inadequate for systems characterized by strong electron correlation or dispersion-dominated interactions [84]. Challenges continue also in the precise modeling of solvent effects and temperature-dependent behavior. Solvation models, such as PCM and SMD, provide approximations of bulk solvent effects while ignoring specific interactions between solute and solvent. Explicit solvent modeling, whether through molecular dynamics or QM/MM, enhances accuracy while also elevating computational requirements [135]. Most QM calculations assume static structures at absolute zero; however, spectroscopic properties are sensitive to temperature variations. Thermal fluctuations may be addressed by path-integral methods or MD [136]. QM approaches struggle with large-sized, flexible, or multi-component systems (e.g. biomolecules and supramolecular assemblies). Employing hybrid QM/MM and coarse-grained models helps the resolution of these challenges [137].

Implicit models (PCM, SMD, COSMO) are usually sufficient when the solvent is aprotic and weakly interacting (e.g., $CDCl_3$, CCl_4, cyclohexane), when the solute lacks strong hydrogen-bond donors/acceptors, and when relative energies of neutral conformers or non-exchangeable NMR shifts are the target. They become inadequate for (i) protic or strongly coordinating solvents (water, DMSO, alcohols) where specific hydrogen bonds or solvent bridges stabilise particular conformers, (ii) charged or zwitterionic species, (iii) exchangeable protons (N–H, O–H) whose chemical shifts depend strongly on hydrogen bonding, and (iv) chiroptical properties of chromophores with direct solute–solvent contacts. In these situations the cluster–continuum approach (a few explicit solvent molecules plus a continuum) is the minimum, with *ab initio*/classical MD sampling and averaging of QM properties over snapshots, or QM/MM, required when the solvation shell is dynamic. The chloroquine case (Section 5.3), where the N–H shift of the lowest-energy conformer **M24** was not observed experimentally, is a typical example in which specific solvent interactions may contribute to the mismatch.

Future research should integrate computational methods with experimental validation to improve conformational analysis and spectral predictions. QM methods and ML-accelerated calculations will enhance accuracy while minimizing computational costs. Hybrid QM/MM and

fragment-based methods will enable the study of large biomolecules and materials, while explicit solvation models and molecular dynamics will clarify the effects of solvent and temperature. Automated workflows that integrate machine learning, quantitative methods, and experimental data will enhance spectral interpretation, minimizing human bias and improving reproducibility. Innovations such as quantum computing and explainable AI present potential solutions for addressing complex chemical systems and dynamic processes. Collaborative efforts among computational chemists, spectroscopists, and data scientists are crucial for enhancing these methodologies, thereby maintaining their significance in drug discovery, materials science, and catalysis. Utilizing these advancements will allow researchers to achieve greater precision, faster predictions, and deeper understanding of molecular behavior.

Despite the progress summarised here, four issues limit the reliability of QM-assisted conformational analysis today. (i) Free-energy errors: conformer populations depend exponentially on ΔG (1 kcal/mol ≈ a factor of 5 at 298 K), yet errors of 1–2 kcal/mol remain common for flexible, dispersion-dominated systems and are amplified by harmonic, rigid-rotor thermal corrections. (ii) Solvent: implicit models cannot reproduce specific solute–solvent hydrogen bonding, as the chloroquine case shows; routine, affordable explicit-solvent or QM/MM-MD protocols are still lacking. (iii) Sampling: conformer searches for molecules with more than ~10 rotatable bonds are rarely converged, and results are seldom reported with convergence tests. (iv) Reproducibility: studies differ in functional, basis set, solvent model and thermal treatment, hampering comparison. We therefore recommend (a) benchmarking against a higher-level reference on a representative subset of conformers, (b) reporting Boltzmann-averaged rather than single-conformer spectra, (c) reporting the type of energy used for populations, and (d) depositing geometries and input files. Machine-learning potentials and automated workflows are the most promising route to meet these requirements, but they need rigorous, system-class-specific validation before they replace DFT in routine use.

The integration of spectroscopic techniques with QM computations has significantly revolutionized the field of conformational analysis, providing structural elucidation with an unprecedented degree of precision. Methods based on DFT calculations alongside NMR, IR, and UV-Vis spectroscopy offer reliable conformational insights for various compounds. Recent DFT functionals now provide precise predictions of NMR chemical shifts (within 0.1–0.2 ppm for ^{1}H NMR) and vibrational frequencies (within 5 cm^{-1} when anharmonic corrections are included). This review's case studies indicate that agreement between Boltzmann-weighted calculated and experimental spectra is a necessary complement to, and in some cases a more discriminating criterion than, calculated relative energies for ascertaining conformational preference among many feasible conformations; the

most reliable assignments combine both criteria. Prospective advancements in computational methodologies and machine learning are anticipated to further enhance these investigations, thereby increasing their applications in drug discovery, biochemistry, and materials science.

Funding

The authors declare that this research received no external funding.

Conflicts of Interest

The authors declare no conflict of interest.

References

[1] B. Dittrich, J. Lübben, S. Mebs, A. Wagner, P. Luger, and R. Flaig, "Accurate Bond Lengths to Hydrogen Atoms from Single-Crystal X-ray Diffraction by Including Estimated Hydrogen ADPs and Comparison to Neutron and QM/MM Benchmarks," Chemistry – A European Journal, vol. 23, no. 19, pp. 4605–4614, Mar. 2017, doi: 10.1002/chem.201604705.

[2] M. Monti, L. Biancorosso, and E. Coccia, "Time-Resolved Circular Dichroism in Molecules: Experimental and Theoretical Advances," Molecules, vol. 29, no. 17, p. 4049, Aug. 2024, doi: 10.3390/molecules29174049.

[3] G. J. Sharman, "Conformation and Stereochemical Analysis of Drug Molecules," eMagRes, pp. 105–116, Mar. 2015, doi: 10.1002/9780470034590.emrstm1392.

[4] S. G. Podkolzin, G. B. Fitzgerald, and B. E. Koel, "Combining Vibrational Spectroscopies with Quantum Chemical Calculations for Molecular-Level Understanding of Reaction Mechanisms on Catalytic Surfaces," Applications of Molecular Modeling to Challenges in Clean Energy, pp. 153–176, Jan. 2013, doi: 10.1021/bk-2013-1133.ch009.

[5] S. Maeda, Y. Harabuchi, Y. Ono, T. Taketsugu, and K. Morokuma, "Intrinsic reaction coordinate: Calculation, bifurcation, and automated search," International Journal of Quantum Chemistry, vol. 115, no. 5, pp. 258–269, Sep. 2014, doi: 10.1002/qua.24757.

[6] J. Anan, E. Adu Fosu, C. Obuah, M. Kojo Ainooson, A. Aniagyei, L. Hamenu, A. Oppong, A. Muller, "A DFT and TD-DFT studies of the photosensitizing capabilities of thiophene-based dyes," Computational and Theoretical Chemistry, vol. 1237, p. 114633, Jul. 2024, doi: 10.1016/j.comptc.2024.114633.

[7] H. Etabti, M. Er-Rajy, A. Fitri, A. Ammasi, M. Shkir, A.T. Benjelloun, M. Benzakour, M. McHarfi, "Computational Analysis of Triazolone-Based Dyes in DSSCs: Exploring Acceptor Terminal for Enhanced Photovoltaic Performance," Journal of Fluorescence, vol. 35, no. 10, pp. 9333–9357, Mar. 2025, doi: 10.1007/s10895-025-04231-z.

[8] R. Diana, U. Caruso, F. S. Gentile, L. Di Costanzo, P. Musto, and B. Panunzi, "Structural feature of thermo-induced fluorochromism in a 1D zinc coordination polymer. A cross-analysis by PL

and FTIR spectroscopy, and DFT formalism," Dyes and Pigments, vol. 202, p. 110247, Jun. 2022, doi: 10.1016/j.dyepig.2022.110247.

[9] F. H. Allen, "The Cambridge Structural Database: a quarter of a million crystal structures and rising," Acta Crystallographica Section B Structural Science, vol. 58, no. 3, pp. 380–388, May 2002, doi: 10.1107/s0108768102003890.

[10] James W. Robinson, Eileen Skelly Frame, George M. Frame II, "Introduction to Spectroscopy," Undergraduate Instrumental Analysis, pp. 92–143, Jul. 2014, doi: 10.1201/b15921-6.

[11] J. Drenth, Principles of Protein X-Ray Crystallography. Springer New York, 2007. doi: 10.1007/0-387-33746-6.

[12] J. CAVANAGH, W. J. FAIRBROTHER, A. G. PALMER, M. RANCE, and N. J. SKELTON, "SEQUENTIAL ASSIGNMENT, STRUCTURE DETERMINATION, AND OTHER APPLICATIONS," Protein NMR Spectroscopy, pp. 781–817, 2007, doi: 10.1016/b978-012164491-8/50012-9.

[13] B. Smith, "Spectral Interpretation Aids," Infrared Spectral Interpretation, pp. 195–205, Feb. 2018, doi: 10.1201/9780203750841-9.

[14] A. D. Becke, "Density-functional thermochemistry. III. The role of exact exchange," The Journal of Chemical Physics, vol. 98, no. 7, pp. 5648–5652, Apr. 1993, doi: 10.1063/1.464913.

[15] T. Bally and P. R. Rablen, "Quantum-Chemical Simulation of ¹H NMR Spectra. 2. Comparison of DFT-Based Procedures for Computing Proton–Proton Coupling Constants in Organic Molecules," The Journal of Organic Chemistry, vol. 76, no. 12, pp. 4818–4830, May 2011, doi: 10.1021/jo200513q.

[16] V. Barone, S. Alessandrini, M. Biczysko, J.R. Cheeseman, D.C. Clary, A.B. McCoy, R.J. Dirisio, F. Neese, M. Melosso, C. Puzzarini, "Computational molecular spectroscopy," Nature Reviews Methods Primers, vol. 1, no. 1, May 2021, doi: 10.1038/s43586-021-00034-1.

[17] J. H. Jensen, "Computational Chemistry," Jan. 2017, doi: 10.14293/s2199-1006.1.sor-chem.clywqk5.v1.

[18] P. Atkins, J. de Paula, and J. Keeler, "Computational chemistry," Atkins' Physical Chemistry, Dec. 2022, doi: 10.1093/hesc/9780198847816.003.0049.

[19] G. Singh and S. Kumar, "Computational Chemistry: Exploring Molecular Properties and Reactivity," Innovations in Science and Technology: Shaping a Sustainable Future (Vol. 1), pp. 87–97, May 2024, doi: 10.9734/bpi/mono/978-81-973809-6-9/ch8.

[20] J. Hénin, T. Lelièvre, M. R. Shirts, O. Valsson, and L. Delemotte, "Enhanced Sampling Methods for Molecular Dynamics Simulations [Article v1.0]," Living Journal of Computational Molecular Science, vol. 4, no. 1, p. 1583, Dec. 2022, doi: 10.33011/livecoms.4.1.1583.

[21] T. A. Mohamed, I. A. Shaaban, U. A. Soliman, and W. M. Zoghaib, "Computational studies, NMR, Raman and infrared spectral analysis of centrosymmetric (2Z,4Z)-Hexa-2,4-dienedinitrile," Journal of Theoretical and Computational Chemistry, vol. 18, no. 02, p. 1950011, Mar. 2019, doi: 10.1142/s0219633619500111.

[22] I. A. Shaaban, A. Karoyo, L. D. Wilson, and T. A. Mohamed, "Raman and DRIFT spectra, vibrational assignments and quantum mechanical calculations of centrosymmetric meso -2,3-Dimercaptosuccinic acid," Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, vol. 183, pp. 275–283, Aug. 2017, doi: 10.1016/j.saa.2017.04.017.

[23] G. Lauro, P. Das, R. Riccio, D. S. Reddy, and G. Bifulco, "DFT/NMR Approach for the Configuration Assignment of Groups of Stereoisomers by the Combination and Comparison of

Experimental and Predicted Sets of Data," The Journal of Organic Chemistry, vol. 85, no. 5, pp. 3297–3306, Jan. 2020, doi: 10.1021/acs.joc.9b03129.

[24] A. Mándi, J. Wu, and T. Kurtán, "TDDFT-ECD and DFT-NMR studies of thaigranatins A–E and granatumin L isolated from *Xylocarpus granatum*" RSC Advances, vol. 10, no. 53, pp. 32216–32224, 2020, doi: 10.1039/d0ra03725g.

[25] M. Habgood, T. James, and A. Heifetz, "Conformational Searching with Quantum Mechanics," Quantum Mechanics in Drug Discovery, pp. 207–229, 2020, doi: 10.1007/978-1-0716-0282-9_14.

[26] R. Fouad, I. A. Shaaban, T. E. Ali, M. A. Assiri, and S. S. Shenouda, "Co(ii), Ni(ii), Cu(ii) and Cd(ii)-thiocarbonohydrazone complexes: spectroscopic, DFT, thermal, and electrical conductivity studies" RSC Advances, vol. 11, no. 60, pp. 37726–37743, 2021, doi: 10.1039/d1ra06902k.

[27] I. A. Shaaban, T. E. Ali, M. A. Assiri, A. M. Fouda, S. M. Eledfawy, and N. M. Hassanin, "Regioselective cyclization reaction of 2-imino-2H-chromene-3-carboxamide with triethyl phosphonoacetate; a combined spectral and computational studies," Journal of Molecular Structure, vol. 1199, p. 126935, Jan. 2020, doi: 10.1016/j.molstruc.2019.126935.

[28] S. Bondock, T. Albarqi, I.A. Shaaban, M.M. Abdou, "Novel asymmetrical azines appending 1,3,4-thiadiazole sulfonamide: synthesis, molecular structure analyses, in silico ADME, and cytotoxic effect," RSC Advances, vol. 13, no. 15, pp. 10353–10366, 2023, doi: 10.1039/d3ra00123g.

[29] F.L.P. Costa, A.C.F. De Albuquerque, R.G. Fiorot, L.M. Lião, L.H. Martorano, G.V.S. Mota, A.L. Valverde, J.W.M. Carneiro, F.M. Dos Santos Junior, "Structural characterisation of natural products by means of quantum chemical calculations of NMR parameters: new insights," Organic Chemistry Frontiers, vol. 8, no. 9, pp. 2019–2058, 2021, doi: 10.1039/d1qo00034a.

[30] C. B. Braga and R. Rittner, "Combined Utilization of ^{1}H NMR, IR, and Theoretical Calculations To Elucidate the Conformational Preferences of Some l-Histidine Derivatives," The Journal of Physical Chemistry A, vol. 121, no. 3, pp. 729–740, Jan. 2017, doi: 10.1021/acs.jpca.6b12515.

[31] I. A. Shaaban, "Conformational analysis, infrared/Raman spectral assignment, and electronic structural studies of 1,3-dimethyl-2-imidazolidinone using quantum chemical calculations," Journal of Molecular Structure, vol. 1175, pp. 708–720, Jan. 2019, doi: 10.1016/j.molstruc.2018.08.015.

[32] S. G. Smith and J. M. Goodman, "Assigning Stereochemistry to Single Diastereoisomers by GIAO NMR Calculation: The DP4 Probability," Journal of the American Chemical Society, vol. 132, no. 37, pp. 12946–12959, Aug. 2010, doi: 10.1021/ja105035r.

[33] S. TOYOTA, "Discovery of internal rotation and conformers of 1,2-dichloroethane: the dawn of the concept of conformation," Proceedings of the Japan Academy, Series B, vol. 100, no. 2, pp. 101–113, Feb. 2024, doi: 10.2183/pjab.100.003.

[34] G. A. Morrison, "Conformational Analysis of Some Alkaloids," Fortschritte der Chemie Organischer Naturstoffe / Progress in the Chemistry of Organic Natural Products / Progrès dans la Chimie des Substances Organiques Naturelles, pp. 269–317, 1967, doi: 10.1007/978-3-7091-8164-5_9.

[35] V. Dragojlovic, "Conformational analysis of cycloalkanes," ChemTexts, vol. 1, no. 3, Aug. 2015, doi: 10.1007/s40828-015-0014-0.

[36] S. Hayward, “A Retrospective on the Development of Methods for the Analysis of Protein Conformational Ensembles,” The Protein Journal, vol. 42, no. 3, pp. 181–191, Apr. 2023, doi: 10.1007/s10930-023-10113-9.

[37] Y. Zhou, I. Limbu, M. J. Garson, and E. H. Krenske, “Conformational Sampling in Computational Studies of Natural Products: Why Is It Important?,” Journal of Natural Products, vol. 87, no. 10, pp. 2543–2549, Sep. 2024, doi: 10.1021/acs.jnatprod.4c00852.

[38] J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Žídek, A. Potapenko, A. Bridgland, C. Meyer, S.A.A. Kohl, A.J. Ballard, A. Cowie, B. Romera-Paredes, S. Nikolov, R. Jain, J. Adler, T. Back, S. Petersen, D. Reiman, E. Clancy, M. Zielinski, M. Steinegger, M. Pacholska, T. Berghammer, S. Bodenstein, D. Silver, O. Vinyals, A.W. Senior, K. Kavukcuoglu, P. Kohli, D. Hassabis, “Highly accurate protein structure prediction with AlphaFold,” Nature, vol. 596, no. 7873, pp. 583–589, Jul. 2021, doi: 10.1038/s41586-021-03819-2.

[39] L. C. Allen, “Electronegativity is the average one-electron energy of the valence-shell electrons in ground-state free atoms,” Journal of the American Chemical Society, vol. 111, no. 25, pp. 9003–9014, Dec. 1989, doi: 10.1021/ja00207a003.

[40] V. V. Kuznetsov, “Relative stability of 5-methyl- and 3,5-dimethyltetrahydro-1,3-oxazine conformers,” Russian Journal of Organic Chemistry, vol. 52, no. 3, pp. 409–413, Mar. 2016, doi: 10.1134/s1070428016030192.

[41] M. I. Alomari, “Computational study of n-butyronitrile in gas and condensed phases: conformational relative stability and thermal properties,” Journal of Chemical Sciences, vol. 130, no. 12, Nov. 2018, doi: 10.1007/s12039-018-1573-6.

[42] R. Chen, Y. Shen, S. Yang, and Y. Zhang, “Conformational Design Principles in Total Synthesis,” Angewandte Chemie International Edition, vol. 59, no. 34, pp. 14198–14210, Jun. 2020, doi: 10.1002/anie.202003735.

[43] O. Rivoire, “A role for conformational changes in enzyme catalysis,” Biophysical Journal, vol. 123, no. 12, pp. 1563–1578, Jun. 2024, doi: 10.1016/j.bpj.2024.04.030.

[44] L. Ricci, L. Sernissi, D. Scarpi, F. Bianchini, A. Contini, and E. G. Occhiato, “Synthesis and conformational analysis of peptides embodying 2,3-methanopipecolic acids,” Organic & Biomolecular Chemistry, vol. 15, no. 32, pp. 6826–6836, 2017, doi: 10.1039/c7ob01617d.

[45] H. Kurokawa, S. Maki-Yonekura, K. Takaba, T. Higashino, S. Inoue, T. Hasegawa, K. Yonekura, “3D Electron Diffraction Structure of an Organic Semiconductor Reveals Conformational Polymorphism,” Journal of the American Chemical Society, vol. 147, no. 7, pp. 5669–5678, Feb. 2025, doi: 10.1021/jacs.4c12734.

[46] J. Hobbs, C. J. Gibb, E. Cruickshank, R. Walker, and R. J. Mandle, “Rapid conformational analysis of semi-flexible liquid crystals,” Liquid Crystals, vol. 51, no. 6, pp. 1022–1034, Apr. 2024, doi: 10.1080/02678292.2024.2333323.

[47] Y. Sasanuma, “Applications: Conformational Analysis and Elucidation of Structure–Property Relationships of Polymers,” Conformational Analysis of Polymers, pp. 207–213, Mar. 2023, doi: 10.1002/9781119716655.part5.

[48] E. Sharmin and F. Zafar, “Introductory Chapter: Metal Organic Frameworks (MOFs),” Metal-Organic Frameworks, Oct. 2016, doi: 10.5772/64797.

[49] J. Tomasi, B. Mennucci, and R. Cammi, “Quantum Mechanical Continuum Solvation Models,” Chemical Reviews, vol. 105, no. 8, pp. 2999–3094, Jul. 2005, doi: 10.1021/cr9904009.

[50] K. S. Watts, P. Dalal, R. B. Murphy, W. Sherman, R. A. Friesner, and J. C. Shelley, "ConfGen: A Conformational Search Method for Efficient Generation of Bioactive Conformers," Journal of Chemical Information and Modeling, vol. 50, no. 4, pp. 534–546, Apr. 2010, doi: 10.1021/ci100015j.

[51] A. Fernandez-Ramos, B. A. Ellingson, B. C. Garrett, and D. G. Truhlar, "Variational Transition State Theory with Multidimensional Tunneling," Reviews in Computational Chemistry, pp. 125–232, Jan. 2007, doi: 10.1002/9780470116449.ch3.

[52] W. Plazinski, A. Plazinska, and A. Brzyska, "Efficient sampling of high-energy states by machine learning force fields," Physical Chemistry Chemical Physics, vol. 22, no. 25, pp. 14364–14374, 2020, doi: 10.1039/d0cp01399d.

[53] G. A. Tribello, M. Bonomi, D. Branduardi, C. Camilloni, and G. Bussi, "PLUMED 2: New feathers for an old bird," Computer Physics Communications, vol. 185, no. 2, pp. 604–613, Feb. 2014, doi: 10.1016/j.cpc.2013.09.018.

[54] I.N. Levine, "Quantum Chemistry. 7th Edition", Pearson Education, New York. 2014.

[55] D. G. Truhlar and V. Mckoy, "Computational chemistry," Computing in Science & Engineering, vol. 2, no. 6, pp. 19–21, Nov. 2000, doi: 10.1109/mcise.2000.881703.

[56] M.J. Frisch, G.W. Trucks, H.B. Schlegel, G.E. Scuseria, M.A. Robb, J.R. Cheeseman, G. Scalmani, V. Barone, G.A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A.V. Marenich, J. Bloino, B.G. Janesko, R. Gomperts, B. Mennucci, H.P. Hratchian, J.V. Ortiz, A.F. Izmaylov, J.L. Sonnenberg, Williams, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V.G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J.A. Montgomery Jr., J.E. Peralta, F. Ogliaro, M.J. Bearpark, J.J. Heyd, E.N. Brothers, K.N. Kudin, V.N. Staroverov, T.A. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A.P. Rendell, J.C. Burant, S.S. Iyengar, J. Tomasi, M. Cossi, J.M. Millam, M. Klene, C. Adamo, R. Cammi, J.W. Ochterski, R.L. Martin, K. Morokuma, O. Farkas, J.B. Foresman, D.J. Fox, in, Wallingford, CT, 2016.

[57] Y. Shao, Z. Gan, E. Epifanovsky, A.T.B. Gilbert, M. Wormit, J. Kussmann, A.W. Lange, A. Behn, J. Deng, X. Feng, D. Ghosh, M. Goldey, P.R. Horn, L.D. Jacobson, I. Kaliman, R.Z. Khaliullin, T. Kuś, A. Landau, J. Liu, E.I. Proynov, Y.M. Rhee, R.M. Richard, M.A. Rohrdanz, R.P. Steele, E.J. Sundstrom, H.L. Woodcock, P.M. Zimmerman, D. Zuev, B. Albrecht, E. Alguire, B. Austin, G.J.O. Beran, Y.A. Bernard, E. Berquist, K. Brandhorst, K.B. Bravaya, S.T. Brown, D. Casanova, C.-M. Chang, Y. Chen, S.H. Chien, K.D. Closser, D.L. Crittenden, M. Diedenhofen, R.A. Distasio, H. Do, A.D. Dutoi, R.G. Edgar, S. Fatehi, L. Fusti-Molnar, A. Ghysels, A. Golubeva-Zadorozhnaya, J. Gomes, M.W.D. Hanson-Heine, P.H.P. Harbach, A.W. Hauser, E.G. Hohenstein, Z.C. Holden, T.-C. Jagau, H. Ji, B. Kaduk, K. Khistyaev, J. Kim, J. Kim, R.A. King, P. Klunzinger, D. Kosenkov, T. Kowalczyk, C.M. Krauter, K.U. Lao, A.D. Laurent, K.V. Lawler, S.V. Levchenko, C.Y. Lin, F. Liu, E. Livshits, R.C. Lochan, A. Luenser, P. Manohar, S.F. Manzer, S.-P. Mao, N. Mardirossian, A.V. Marenich, S.A. Maurer, N.J. Mayhall, E. Neuscamman, C.M. Oana, R. Olivares-Amaya, D.P. O'Neill, J.A. Parkhill, T.M. Perrine, R. Peverati, A. Prociuk, D.R. Rehn, E. Rosta, N.J. Russ, S.M. Sharada, S. Sharma, D.W. Small, A. Sodt, T. Stein, D. Stück, Y.-C. Su, A.J.W. Thom, T. Tsuchimochi, V. Vanovschi, L. Vogt, O. Vydrov, T. Wang, M.A. Watson, J. Wenzel, A. White, C.F. Williams, J. Yang, S. Yeganeh, S.R. Yost, Z.-Q. You, I.Y. Zhang, X. Zhang, Y. Zhao, B.R. Brooks, G.K.L. Chan,

D.M. Chipman, C.J. Cramer, W.A. Goddard, M.S. Gordon, W.J. Hehre, A. Klamt, H.F. Schaefer, M.W. Schmidt, C.D. Sherrill, D.G. Truhlar, A. Warshel, X. Xu, A. Aspuru-Guzik, R. Baer, A.T. Bell, N.A. Besley, J.-D. Chai, A. Dreuw, B.D. Dunietz, T.R. Furlani, S.R. Gwaltney, C.-P. Hsu, Y. Jung, J. Kong, D.S. Lambrecht, W. Liang, C. Ochsenfeld, V.A. Rassolov, L.V. Slipchenko, J.E. Subotnik, T. Van Voorhis, J.M. Herbert, A.I. Krylov, P.M.W. Gill, M. Head-Gordon, "Advances in molecular quantum chemistry contained in the Q-Chem 4 program package," Molecular Physics, vol. 113, no. 2, pp. 184–215, Sep. 2014, doi: 10.1080/00268976.2014.952696.

[58] G.M.J. Barca, C. Bertoni, L. Carrington, D. Datta, N. De Silva, J.E. Deustua, D.G. Fedorov, J.R. Gour, A.O. Gunina, E. Guidez, T. Harville, S. Irle, J. Ivanic, K. Kowalski, S.S. Leang, H. Li, W. Li, J.J. Lutz, I. Magoulas, J. Mato, V. Mironov, H. Nakata, B.Q. Pham, P. Piecuch, D. Poole, S.R. Pruitt, A.P. Rendell, L.B. Roskop, K. Ruedenberg, T. Sattasathuchana, M.W. Schmidt, J. Shen, L. Slipchenko, M. Sosonkina, V. Sundriyal, A. Tiwari, J.L. Galvez Vallejo, B. Westheimer, M. Włoch, P. Xu, F. Zahariev, M.S. Gordon, "Recent developments in the general atomic and molecular electronic structure system," The Journal of Chemical Physics, vol. 152, no. 15, Apr. 2020, doi: 10.1063/5.0005188.

[59] R.M. Parrish, L.A. Burns, D.G.A. Smith, A.C. Simmonett, A.E. Deprince, E.G. Hohenstein, U. Bozkaya, A.Y. Sokolov, R. Di Remigio, R.M. Richard, J.F. Gonthier, A.M. James, H.R. McAlexander, A. Kumar, M. Saitow, X. Wang, B.P. Pritchard, P. Verma, H.F. Schaefer, K. Patkowski, R.A. King, E.F. Valeev, F.A. Evangelista, J.M. Turney, T.D. Crawford, C.D. Sherrill, "Psi4 1.1: An Open-Source Electronic Structure Program Emphasizing Automation, Advanced Libraries, and Interoperability," Journal of Chemical Theory and Computation, vol. 13, no. 7, pp. 3185–3197, Jun. 2017, doi: 10.1021/acs.jctc.7b00174.

[60] M. Valiev, E.J. Bylaska, N. Govind, K. Kowalski, T.P. Straatsma, H.J.J. Van Dam, D. Wang, J. Nieplocha, E. Apra, T.L. Windus, W.A. De Jong, "NWChem: A comprehensive and scalable open-source solution for large scale molecular simulations," Computer Physics Communications, vol. 181, no. 9, pp. 1477–1489, Sep. 2010, doi: 10.1016/j.cpc.2010.04.018.

[61] H. F. Schaefer, "Methylene: A Paradigm for Computational Quantum Chemistry," Science, vol. 231, no. 4742, pp. 1100–1107, Mar. 1986, doi: 10.1126/science.231.4742.1100.

[62] W. Kohn, "Nobel Lecture: Electronic structure of matter—wave functions and density functionals," Reviews of Modern Physics, vol. 71, no. 5, pp. 1253–1266, Oct. 1999, doi: 10.1103/revmodphys.71.1253.

[63] J. A. Pople, "Nobel Lecture: Quantum chemical models," Reviews of Modern Physics, vol. 71, no. 5, pp. 1267–1274, Oct. 1999, doi: 10.1103/revmodphys.71.1267.

[64] M. Karplus, "Citing 'Development of Multiscale Models for Complex Chemical Systems: From H+H2 to Biomolecules (Nobel Lecture)'," Angewandte Chemie International Edition, vol. 53, no. 38, pp. 9992–10005, Jul. 2014, doi: 10.1002/anie.201403924.

[65] M. Levitt, "Birth and Future of Multiscale Modeling for Macromolecular Systems (Nobel Lecture)," Angewandte Chemie International Edition, vol. 53, no. 38, pp. 10006–10018, Aug. 2014, doi: 10.1002/anie.201403691.

[66] A. Warshel, "Multiscale Modeling of Biological Functions: From Enzymes to Molecular Machines (Nobel Lecture)," Angewandte Chemie International Edition, vol. 53, no. 38, pp. 10020–10031, Jul. 2014, doi: 10.1002/anie.201403689..

[67] X. Dai and Y. Chen, "Computational Biomaterials: Computational Simulations for Biomedicine," Advanced Materials, vol. 35, no. 7, Dec. 2022, doi: 10.1002/adma.202204798.

[68] S. G. Louie, Y.-H. Chan, F. H. da Jornada, Z. Li, and D. Y. Qiu, "Discovering and understanding materials through computation," Nature Materials, vol. 20, no. 6, pp. 728–735, May 2021, doi: 10.1038/s41563-021-01015-1.

[69] E. G. Lewars, "Ab initio Calculations," Computational Chemistry, pp. 175–390, Sep. 2010, doi: 10.1007/978-90-481-3862-3_5.

[70] C.J. Cramer, "Essentials of Computational Chemistry: Theories and Models," Wiley, 2005.

[71] J. Townsend, J. K. Kirkland, and K. D. Vogiatzis, "Post-Hartree-Fock methods: configuration interaction, many-body perturbation theory, coupled-cluster theory," Mathematical Physics in Theoretical Chemistry, pp. 63–117, 2019, doi: 10.1016/b978-0-12-813651-5.00003-6.

[72] E. F. Silva-Júnior, T. M. Aquino, and J. X. Araújo-Júnior, "Quantum Mechanical (QM) Calculations Applied to ADMET Drug Prediction: A Review," Current Drug Metabolism, vol. 18, no. 6, Jul. 2017, doi: 10.2174/1389200218666170316094514.

[73] M. Aamir Iqbal, N. Ashraf, W. Shahid, D. Afzal, F. Idrees, and R. Ahmad, "Fundamentals of Density Functional Theory: Recent Developments, Challenges and Future Horizons," Density Functional Theory - Recent Advances, New Perspectives and Applications, May 2022, doi: 10.5772/intechopen.99019.

[74] A. D. Becke, "Density-functional exchange-energy approximation with correct asymptotic behavior," Physical Review A, vol. 38, no. 6, pp. 3098–3100, Sep. 1988, doi: 10.1103/physreva.38.3098.

[75] C. Lee, W. Yang, and R. G. Parr, "Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density," Physical Review B, vol. 37, no. 2, pp. 785–789, Jan. 1988, doi: 10.1103/physrevb.37.785.

[76] J.-D. Chai and M. Head-Gordon, "Systematic optimization of long-range corrected hybrid density functionals," The Journal of Chemical Physics, vol. 128, no. 8, Feb. 2008, doi: 10.1063/1.2834918.

[77] J.-D. Chai and M. Head-Gordon, "Long-range corrected hybrid density functionals with damped atom–atom dispersion corrections," Physical Chemistry Chemical Physics, vol. 10, no. 44, p. 6615, 2008, doi: 10.1039/b810189b.

[78] E. Caldeweyher, J.-M. Mewes, S. Ehlert, and S. Grimme, "Extension and evaluation of the D4 London-dispersion model for periodic systems," Physical Chemistry Chemical Physics, vol. 22, no. 16, pp. 8499–8512, 2020, doi: 10.1039/d0cp00502a.

[79] M. Bursch, J.-M. Mewes, A. Hansen, and S. Grimme, "Best Practice DFT Protocols for Basic Molecular Computational Chemistry," Apr. 2022, doi: 10.26434/chemrxiv-2022-n304h.

[80] M. E. CASIDA, "Time-Dependent Density Functional Response Theory for Molecules," Recent Advances in Density Functional Methods, pp. 155–192, Nov. 1995, doi: 10.1142/9789812830586_0005.

[81] A. Dreuw and M. Head-Gordon, "Single-Reference ab Initio Methods for the Calculation of Excited States of Large Molecules," Chemical Reviews, vol. 105, no. 11, pp. 4009–4037, Oct. 2005, doi: 10.1021/cr0505627.

[82] J. R. Williams, N. Tancogne-Dejean, and C. A. Ullrich, "Time-Resolved Exciton Wave Functions from Time-Dependent Density-Functional Theory," Journal of Chemical Theory and Computation, vol. 17, no. 3, pp. 1795–1805, Feb. 2021, doi: 10.1021/acs.jctc.0c01334.

[83] B. Ait Hammou, A. El Kaaouachi, E. Hassan Mounir, H. Mabchour, A. El Oujdi, A. Echchelh, S. Dlimi, D. Ennajih, "DFT and TDDFT Calculations of Ground and Excited States of Photoelectron Emission," Density Functional Theory - New Perspectives and Applications, Jan. 2024, doi: 10.5772/intechopen.111611.

[84] S. Grimme, "Exploration of Chemical Compound, Conformer, and Reaction Space with Meta-Dynamics Simulations Based on Tight-Binding Quantum Chemical Calculations," Journal of Chemical Theory and Computation, vol. 15, no. 5, pp. 2847–2862, Apr. 2019, doi: 10.1021 /acs.jctc.9b00143.

[85] M. Paluszek and S. Thomas, "An Overview of Machine Learning," MATLAB Machine Learning Recipes, pp. 1–19, 2024, doi: 10.1007/978-1-4842-9846-6_1.

[86] O. A. von Lilienfeld, K.-R. Müller, and A. Tkatchenko, "Exploring chemical compound space with quantum-based machine learning," Nature Reviews Chemistry, vol. 4, no. 7, pp. 347–358, Jun. 2020, doi: 10.1038/s41570-020-0189-9.

[87] S. Grimme, C. Bannwarth, S. Dohm, A. Hansen, J. Pisarek, P. Pracht, J. Seibert, F. Neese, "Fully Automated Quantum-Chemistry-Based Computation of Spin–Spin-Coupled Nuclear Magnetic Resonance Spectra," Angewandte Chemie International Edition, vol. 56, no. 46, pp. 14763–14769, Oct. 2017, doi: 10.1002/anie.201708266.

[88] P. Pracht, S. Grimme, C. Bannwarth, F. Bohle, S. Ehlert, G. Feldmann, J. Gorges, M. Müller, T. Neudecker, C. Plett, S. Spicher, P. Steinbach, P.A. Wesołowski, F. Zeller, "CREST—A program for the exploration of low-energy molecular chemical space," The Journal of Chemical Physics, vol. 160, no. 11, Mar. 2024, doi: 10.1063/5.0197592.

[89] C. Devereux, J. S. Smith, K. K. Huddleston, K. Barros, R. Zubatyuk, O. Isayev, and A. E. Roitberg, "Extending the Applicability of the ANI Deep Learning Molecular Potential to Sulfur and Halogens," Journal of Chemical Theory and Computation, vol. 16, pp. 4192–4202, Jun. 2020, doi: 10.1021/acs.jctc.0c00121.

[90] D. M. Anstine, R. Zubatyuk, and O. Isayev, "AIMNet2: a neural network potential to meet your neutral, charged, organic, and elemental-organic needs," Chemical Science, vol. 16, pp. 10228–10244, 2025, doi: 10.1039/D4SC08572H.

[91] D. P. Kovács, J. H. Moore, N. J. Browning, I. Batatia, J. T. Horton, Y. Pu, V. Kapil, W. C. Witt, I.-B. Magdău, D. J. Cole, and G. Csányi, "MACE-OFF: Short-Range Transferable Machine Learning Force Fields for Organic Molecules," Journal of the American Chemical Society, vol. 147, pp. 17598–17611, May 2025, doi: 10.1021/jacs.4c07099.

[92] Y. Guan, S. V. S. Sowndarya, L. C. Gallegos, P. C. St. John, and R. S. Paton, "Real-time prediction of 1H and 13C chemical shifts with DFT accuracy using a 3D graph neural network," Chemical Science, vol. 12, pp. 12012–12026, 2021, doi: 10.1039/D1SC03343C.

[93] A. Howarth, K. Ermanis, and J. M. Goodman, "DP4-AI automated NMR data analysis: straight from spectrometer to structure," Chemical Science, vol. 11, pp. 4351–4359, 2020, doi: 10.1039/D0SC00442A.

[94] B. Jing, G. Corso, J. Chang, R. Barzilay, and T. Jaakkola, "Torsional Diffusion for Molecular Conformer Generation," Advances in Neural Information Processing Systems 35 (NeurIPS 2022), 2022, arXiv:2206.01729.

[95] O.-E. Ganea, L. Pattanaik, C. W. Coley, R. Barzilay, K. F. Jensen, W. H. Green, and T. S. Jaakkola, "GeoMol: Torsional Geometric Generation of Molecular 3D Conformer Ensembles," Advances in Neural Information Processing Systems 34 (NeurIPS 2021), 2021, arXiv:2106.07802.

[96] A.S. Perera, C.D. Carlson, J. Cheramy, Y. Xu, Citing, "Infrared and vibrational circular dichroism spectra of methyl β-D-glucopyranose in water: The application of the quantum cluster growth and clusters-in-a-liquid solvation models," Chirality, vol. 35, no. 10, pp. 718–731, May 2023, doi: 10.1002/chir.235

[97] P. Hamm, M. Zanni,"Concepts and Methods of 2D Infrared Spectroscopy," Vol. 1. Cambridge: Cambridge University Press, 2011.

[98] W. Wang and V. Deckert, "Tip-Enhanced Raman Spectroscopy," Raman Spectroscopy, pp. 233–289, 2024, doi: 10.1007/978-981-97-1703-3_11.

[99] R. Franco, S. Gil-Caballero, I. Ayala, A. Favier, and B. Brutscher, "Probing Conformational Exchange Dynamics in a Short-Lived Protein Folding Intermediate by Real-Time Relaxation–Dispersion NMR," Journal of the American Chemical Society, vol. 139, no. 3, pp. 1065–1068, Jan. 2017, doi: 10.1021/jacs.6b12089.

[100] M.-K. Lee, M. Gal, L. Frydman, and G. Varani, "Real-time multidimensional NMR follows RNA folding with second resolution," Proceedings of the National Academy of Sciences, vol. 107, no. 20, pp. 9192–9197, May 2010, doi: 10.1073/pnas.1001195107.

[101] V. Kumar, "In-cell NMR spectroscopy: A tool to study cellular structure biology," Advances in Protein Molecular and Structural Biology Methods, pp. 267–276, 2022, doi: 10.1016/b978-0-323-90264-9.00017-9.

[102] S. Basak and K. Chattopadhyay, "Studies of protein folding and dynamics using single molecule fluorescence spectroscopy," Physical Chemistry Chemical Physics, vol. 16, no. 23, p. 11139, 2014, doi: 10.1039/c3cp55219e.

[103] P. S. Kumagai, A. P. U. Araujo, and J. L. S. Lopes, "Going deep into protein secondary structure with synchrotron radiation circular dichroism spectroscopy," Biophysical Reviews, vol. 9, no. 5, pp. 517–527, Aug. 2017, doi: 10.1007/s12551-017-0314-2.

[104] A. A. Altaf, S. Kausar, and A. Badshah, "Spectral Calculations with DFT," Density Functional Calculations - Recent Progresses of Theory and Application, May 2018, doi: 10.5772/intechopen.71080.

[105] E.B. Wilson, J.C. Decius, P.C. Cross, "Molecular Vibrations: The Theory of Infrared and Raman Vibrational Spectra," Vol. 102. New York: McGraw-Hill, 1955.

[106] M. H. Jamróz, "Vibrational Energy Distribution Analysis (VEDA): Scopes and limitations," Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, vol. 114, pp. 220–230, Oct. 2013, doi: 10.1016/j.saa.2013.05.096.

[107] T. A. Mohamed, G. A. Guirgis, Y. E. Nashed, and J. R. Durig, "Spectra and structure of silicon containing compounds," Vibrational Spectroscopy, vol. 30, no. 2, pp. 111–120, Nov. 2002, doi: 10.1016/s0924-2031(02)00003-6.

[108] I. A. Shaaban, M. A. Assiri, T. E. Ali, and T. A. Mohamed, "Keto-enol tautomerism, spectral (infrared, Raman and NMR) studies and Normal coordinate analysis of 4-Methyl-2-hydroxyquinoline using quantum mechanical calculations," Journal of Molecular Structure, vol. 1252, p. 132137, Mar. 2022, doi: 10.1016/j.molstruc.2021.132137.

[109] C. Cuadrado, F. Cen-Pacheco, and A. H. Daranas, "Computationally Assisted Analysis of NMR Chemical Shifts as a Tool in Conformational Analysis," Organic Letters, vol. 26, no. 31, pp. 6529–6534, Jun. 2024, doi: 10.1021/acs.orglett.4c01642.

[110] J. Li, J. Liang, Z. Wang, A.L. Ptaszek, X. Liu, B. Ganoe, M. Head-Gordon, T. Head-Gordon, "Highly Accurate Prediction of NMR Chemical Shifts from Low-Level Quantum Mechanics Calculations Using Machine Learning," Journal of Chemical Theory and Computation, vol. 20, no. 5, pp. 2152–2166, Feb. 2024, doi: 10.1021/acs.jctc.3c01256.

[111] G. Saielli, K. C. Nicolaou, A. Ortiz, H. Zhang, and A. Bagno, "Addressing the Stereochemistry of Complex Organic Molecules by Density Functional Theory-NMR: Vannusal B in Retrospective," Journal of the American Chemical Society, vol. 133, no. 15, pp. 6072–6077, Mar. 2011, doi: 10.1021/ja201108a.

[112] A. C. F. de Albuquerque, L. H. Martorano, and F. M. dos Santos, "Are we still chasing molecules that were never there, The role of quantum chemical simulations of NMR parameters in structural reassignment of natural products," Frontiers in Natural Products, vol. 2, Jan. 2024, doi: 10.3389/fntpr.2023.1321043.

[113] H. Shi, I.N. Michaelides, B. Darses, P. Jakubec, Q.N.N. Nguyen, R.S. Paton, D.J. Dixon, "Total Synthesis of (−)-Himalensine A," Journal of the American Chemical Society, vol. 139, no. 49, pp. 17755–17758, Nov. 2017, doi: 10.1021/jacs.7b10956.

[114] Q.N.N. Nguyen, J. Schwochert, D.J. Tantillo, R.S. Lokey ,"Using ^{1}H and ^{13}C NMR chemical shifts to determine cyclic peptide conformations: a combined molecular dynamics and quantum mechanics approach," Physical Chemistry Chemical Physics, vol. 20, no. 20, pp. 14003–14012, 2018, doi: 10.1039/c8cp01616j.

[115] J. K. Sader and J. E. Wulff, "Reinvestigation of a robotically revealed reaction," Nature, vol. 570, no. 7762, pp. E54–E59, Jun. 2019, doi: 10.1038/s41586-019-1176-5.

[116] S.-M. Shen, G. Appendino, and Y.-W. Guo, "Pitfalls in the structural elucidation of small molecules. A critical analysis of a decade of structural misassignments of marine natural products," Natural Product Reports, vol. 39, no. 9, pp. 1803–1832, 2022, doi: 10.1039/d2np00023g.

[117] F. Wang, "Future of computational molecular spectroscopy—from supporting interpretation to leading the innovation," Physical Chemistry Chemical Physics, vol. 25, no. 10, pp. 7090–7105, 2023, doi: 10.1039/d3cp00192j.

[118] J.J. La Clair,"Total Syntheses of Hexacyclinol, 5-epi-Hexacyclinol, and Desoxohexacyclinol Unveil an Antimalarial Prodrug Motif," Angewandte Chemie, vol. 118, no. 17, pp. 2835–2839, Apr. 2006, doi: 10.1002/ange.200504033.

[119] S. D. Rychnovsky, "Predicting NMR Spectra by Computational Methods: Structure Revision of Hexacyclinol," Organic Letters, vol. 8, no. 13, pp. 2895–2898, Jun. 2006, doi: 10.1021/ol0611346.

[120] J. A. Porco, S. Su, X. Lei, S. Bardhan, and S. D. Rychnovsky, "Total Synthesis and Structure Assignment of (+)-Hexacyclinol," Angewandte Chemie International Edition, vol. 45, no. 35, pp. 5790–5792, Aug. 2006, doi: 10.1002/anie.200602854.

[121] I. Cortés, C. Cuadrado, A. Hernández Daranas, and A. M. Sarotti, "Machine learning in computational NMR-aided structural elucidation," Frontiers in Natural Products, vol. 2, Jan. 2023, doi: 10.3389/fntpr.2023.1122426.

[122] J. F. Stanton and R. J. Bartlett, "The equation of motion coupled-cluster method. A systematic biorthogonal approach to molecular excitation energies, transition probabilities, and excited state properties," The Journal of Chemical Physics, vol. 98, no. 9, pp. 7029–7039, May 1993, doi: 10.1063/1.464746.

[123] A. G. Petrovic, N. Berova, and J. L. Alonso-Gómez, "Absolute Configuration and Conformational Analysis of Chiral Compounds via Experimental and Theoretical Prediction of Chiroptical Properties: ORD, ECD, and VCD," Structure Elucidation in Organic Chemistry, pp. 65–104, Dec. 2014, doi: 10.1002/9783527664610.ch3.

[124] S. Superchi, P. Scafato, M. Gorecki, and G. Pescitelli, "Absolute Configuration Determination by Quantum Mechanical Calculation of Chiroptical Spectra: Basics and Applications to Fungal Metabolites," Current Medicinal Chemistry, vol. 25, no. 2, pp. 287–320, Jan. 2018, doi: 10.2174/0929867324666170310112009.

[125] M. V. Cooke, I. Malvacio, W. J. Peláez, A. J. Pepino, M. R. Mazzieri, and G. A. Argüello, "TD-DFT calculations of UV absorption bands and their intensities in the spectra of some tetrahydroquinolines," RSC Advances, vol. 5, no. 33, pp. 26255–26262, 2015, doi: 10.1039/c5ra00249d.

[126] Y. Morisawa, Y. Higaki, and Y. Ozaki, "Far-Ultraviolet Spectroscopy and Quantum Chemical Calculation Studies of the Conformational Dependence on the Electronic Structure and Transitions of Cyclohexane, Methyl and Dimethyl Cyclohexane, and Decalin; Effects of Axial Substitutions on the Electronic Transitions," The Journal of Physical Chemistry A, vol. 125, no. 37, pp. 8205–8214, Sep. 2021, doi: 10.1021/acs.jpca.1c05688.

[127] M. A. Muñoz, "Chiral Chromatographic Analysis and Absolute Stereochemistry of Racemic Chlorpheneramine Using Chiral HPLC-Cd-or and TD-DFT Calculations," Journal of the Chilean Chemical Society, vol. 63, no. 3, pp. 4155–4158, 2018, doi: 10.4067/s0717-9707201800030 4155.

[128] Y. Wang, J. Fleischhauer, S. Bausch, M. Sebestian, and P. H. Laur, "Conformational Analysis and TDDFT Calculations of the Chiroptical Properties of Tris[1,2-propanediolato(2-)-κ O ,κ O ′]-selenium/tellurium and Related Compounds," Enantiomer, vol. 7, no. 6, pp. 343–374, Nov. 2002, doi: 10.1080/10242430215700.

[129] J. Stocka, R. Platakyte, T.M.C. McFadden, J. Ceponkus, V. Aleksa, A.G. Hanna, V. Sablinskas, P. Rodziewicz, G.A. Guirgis , "Conformational diversity of 1-chloro-1-chloromethylsila cyclohexane with experimental (Raman and IR) and computational (DFT, MP2) methods," Journal of Molecular Structure, vol. 1249, p. 131644, Feb. 2022, doi: 10.1016/j.molstruc. 2021.131644.

[130] I. S. Hernandes, H. C. Da Silva, H. F. Dos Santos, and W. B. De Almeida, "Unveiling the Molecular Structure of Antimalarial Drugs Chloroquine and Hydroxychloroquine in Solution through Analysis of ^{1}H NMR Chemical Shifts," The Journal of Physical Chemistry B, vol. 125, no. 13, pp. 3321–3342, Mar. 2021, doi: 10.1021/acs.jpcb.1c00609.

[131] L.-Y. Liu, H. Sun, C. Griesinger, and J.-K. Liu, "The Use of a Combination of RDC and Chiroptical Spectroscopy for Determination of the Absolute Configuration of Fusariumin A from the Fungus Fusarium sp.," Natural Products and Bioprospecting, vol. 6, no. 1, pp. 41–48, Jan. 2016, doi: 10.1007/s13659-015-0084-0.

[132] L. Grauso, R. Teta, G. Esposito, M. Menna, and A. Mangoni, "Computational prediction of chiroptical properties in structure elucidation of natural products," Natural Product Reports, vol. 36, no. 7, pp. 1005–1030, 2019, doi: 10.1039/c9np00018f.

[133] N. Grimblat, M. M. Zanardi, and A. M. Sarotti, "Beyond DP4: an Improved Probability for the Stereochemical Assignment of Isomeric Compounds using Quantum Chemical Calculations of NMR Shifts," The Journal of Organic Chemistry, vol. 80, no. 24, pp. 12526–12534, Dec. 2015, doi: 10.1021/acs.joc.5b02396.

[134] P. H. Willoughby, M. J. Jansma, and T. R. Hoye, "A guide to small-molecule structure assignment through computation of (1H and 13C) NMR chemical shifts," Nature Protocols, vol. 9, no. 3, pp. 643–660, Mar. 2014, doi: 10.1038/nprot.2014.042.

[135] A. V. Marenich, S. V. Jerome, C. J. Cramer, and D. G. Truhlar, "Charge Model 5: An Extension of Hirshfeld Population Analysis for the Accurate Description of Molecular Interactions in Gaseous and Condensed Phases," Journal of Chemical Theory and Computation , vol. 8, no. 2, pp. 527–541, Jan. 2012, doi: 10.1021/ct200866d.

[136] Y. Hu, P. Ounkham, O. Marsalek, T. E. Markland, B. Krishmoorthy, and A. E. Clark, "Persistent Homology Metrics Reveal Quantum Fluctuations and Reactive Atoms in Path Integral Dynamics," Frontiers in Chemistry, vol. 9, Mar. 2021, doi: 10.3389/fchem.2021.624937.

[137] J. Liu, L.-W. Qi, J. Z. H. Zhang, and X. He, "Fragment Quantum Mechanical Method for Large-Sized Ion–Water Clusters," Journal of Chemical Theory and Computation, vol. 13, no. 5, pp. 2021–2034, Apr. 2017, doi: 10.1021/acs.jctc.7b00149.

Islamic University Journal of Applied Sciences (IUJAS)

https://journals.iu.edu.sa/jesc

Volume VIII, Issue II, December 2026, Pages 326- 345

Variable Thermal Conductivity of Sutterby Nanofluid flow: a Numerical Study

Saima Riasat*, Saira Iqbal and Sadaf Riasat

Department of Mathematical Sciences Fatima Jinnah Women University, Rawalpindi, Pakistan

*Corresponding author: (Saima Riasat), Email Address: **saimaqau@live.com**

Abstract

The current research focuses on the impact of variable thermal conductivity on electro-magnetohydrodynamic (EMHD) flow of a Sutterby nanofluid in a Darcy-Forchheimer porous media with a stretching sheet. The impacts of thermal radiation and Christov-Cattaneo heat flux model are also considered in the study. The similarity approach is used to transform the governing mathematical model into a set of nonlinear ODEs which are then numerically solved by applying the BVP4C method. The detailed parametric studies are performed to provide more insights into the fluid flow and heat transfer behavior. It is found that the increase in magnetic field and velocity slip parameters decreases the fluid velocity. Also, the rise in the variable thermal conductivity parameter increases the temperature profile. The impacts of several controlling parameters on the velocity, temperature, concentration, skin friction coefficient, heat transfer rate, and Nusselt number are discussed and illustrated graphically and in tabular form.

Keywords: Electro-magnetohydrodynamic; Thermal radiation; Sutterby nanofluid; Darcy-Forchheimer porous media.

https://doi.org/10.63070/jesc.2026.045

Received 09 August 2026; Revised 02 October 2026; Accepted 10 October 2026. Available 16 October 2026.
Published by Islamic University of Madinah on behalf of *Islamic University Journal of Applied Sciences.*
This is an open access article published under the Creative Commons Attribution 4.0 International license (CC BY 4.0): https://creativecommons.org/licenses/by/4.0/

Nomenclature

c Positive constant $\left(s^{-1}\right)$

Q Heat source

B_0 Uniform magnetic field

Ha Magnetic parameter

Ec Eckert number

De Deborah number

Le Lewis number

F^* Inertia parameter

E Electric parameter

k_p Porosity parameter

Re Reynolds number

C^* Concentration

C_∞ Ambient concentration

C_W Concentration of the wall

Nt Thermophoretic parameter

Pr Prandtl number

Nb Brownian ratio parameter

R Radiation parameter

T_1 fluid temperature

T_∞ Ambient temperature

T_W wall temperature

k^* Mean absorption coefficient

D_T Thermophoretic diffusion co-efficient.

D_B Brownian diffusion co-efficient.

N_0 Slip length

K_0 Thermal jump length

C_b Drag coefficient

b^2 Consistency index

x, y Cartesian coordinates

U_W Stretching velocity

Greek letters

ε Thermal conductivity parameter

η Dimensionless similitude parameter

γ Thermal relaxation parameter

α Thermal diffusivity

λ^* Thermal relaxation time

σ Stefan Boltzmann constant

γ_1 Chemical reaction parameter

υ Fluid kinematic viscosity

ϕ Dimensionless concentration

ρ Fluid density

β Thermal slip parameter

σ^* Magnetic permeability

θ Dimensionless temperature

τ Nanoparticles to base fluid heat capacity ratio

δ Velocity slip parameter

Subscripts

f Fluid

∞ Ambient state

W On the sheet surface

1. Introduction

Nanofluids are the colloidal suspensions of base fluid with nanoparticles. Nanoparticles are the particles of dimensions approximately 1-100nm in size [1]. Nanofluids are fluids that are considered to enhance thermal conduction and heat convective transmission. Due to their significant applications in bubble absorption technologies [2], reactors [3], coolants [4], pharmaceutical procedure [5], heat exchangers [6], and solar collectors [7], nanofluids have attracted huge interest in recent years. Nanofluids are the term used by Choi et al. [8] to describe the dispersion of tiny nanoparticles within the base fluid. Bachok et al. [9] analyzed two-dimensional nanofluid flow on top of an expanding sheet. The water-based nanofluids flow along a stretched surface signified by a magnetic field was analyzed by Ali et al. [10] and Hamad [11]. Ibrahim et al. [12] analyzed MHD stagnation point nanofluid flow across a stretching surface.

Non-Newtonian fluids have recently piqued the interest of many analysts in the logical and design fields, as they have significant role in industrial and engineering mechanism. The Sutterby fluid [13] is a non-Newtonian fluid, and Sutterby published the basic equations for it in 1966. Due to its dual nature, that describes the behavior of both pseudo-plastic and dilatant fluid, the Sutterby fluid model has attracted the attention of researchers. These fluids include methylcellulose, hydroxyethyl cellulose, and carboxy-methylcellulose. A numeric investigation using Sutterby fluid theory with helical flow was presented by Eissa and Batra [14]. The mixed radiative and convective flow for squeezed Sutterby fluid through a squeezed channel was analyzed by Ahmad et al. [15]. The magnetic field influence of the Sutterby fluid combined with a mass and heat transition was described by Bilal et al. [16]. The Sutterby fluid was discussed in terms of entropy formation by Hayat et al. [17]. The impact of a linearly stretchy surface on the Sutterby nano-liquids flow is examined by Mir et al. in their study from [18]. On a stretching sheet, Bilal et al. [19] investigate a radiative Sutterby fluid with slip boundary conditions.

Because of the numerous uses in industry and engineering, many experts have long been immersed in evaluating fluid flow across porous space. Darcy's theory is commonly used to simulate and evaluate any flow across a permeable medium. The fluid flow in a porous material turns non-linear at higher velocities, and so the inertia impacts may not be neglected. Whereas conventional Darcy's theory is adequate for lower porosity and velocity, it is inadequate for inertial and boundary impacts at higher flow rates. Hence, Forchheimer [20] incorporated these concerns into the Darcian velocity equation by inserting an extra term using square velocity. This important development within law created a brand-new field for study with porous medium. Muskat [21] named the additive term within the Darcy

law the Forchheimer term. Khan et al. [22] analyzed the consequence of viscous dissipation and ohmic heating on ferrofluid Darcy-Forchheimer flow across a stretchable surface. Eid and Mabood [23] illustrated Darcy-Forchheimer motion on a dusty nanofluid over a expanding surface by using porosity media and magnetic field effect. Ramzan et al. [24] investigated the MHD Darcy Forchheimer flow of Williamson nanofluid along a stretching sheet.

Heat transfer is a natural phenomenon that occurs as a consequence of a temperature difference in the same bodies and among the different bodies. In the past centuries, Fourier's principle of heat conduction was used to examine the characteristics in heat transport. One of the model's major flaws are that it creates a parabolic heat equation, which causes a minor disruption to be sensed quickly across the whole substance. To overcome such barrier, Cattaneo [25] modified Fourier law by incorporating thermal relaxation time that permits heat to get transmitted by the slow propagation of thermal waves. The model was recognized as CC heat flux model. Straughan [26] investigate the Christov-Cattaneo model using thermal convection. Zampoli and Tibulle[27] investigated the incompressible fluid flow with Christov-Cattaneo heat flux model. Han et al. [28] used the Cattaneo model to examine the Maxwell boundary layer fluid flow. Hayat et al. [29] studied the effectiveness of Cattaneo model in flow across a expanding surface using varying thickness.

The amount of heat transfer from any material is referred as thermal conductivity. Another important attribute of thermal conductivity is that it varies with temperature. It is utilized in a variety of applications, including steam generators, laminating, concrete heating, electrolytes, moulding blow, and catalysis. Chiam [30-31] analyze the variable thermal conductivity effect on energy transmission. Gbadeyan et al. [32] investigated the 2D Casson nanofluid variable thermal conductivity and diffusivity moving across a stretchable sheet. Hamid et al. [33] employed the Galerkin technique to acquire the mathematical approximation for a nanofluid using variable thermal conductivity passing along a expanding sheet. Salawu and Dada [34] examine the impact of a magnetic field using varying thermal conductivity on an incompressible fluid passing through an expanding sheet. Abel et al. [35] analyze the thermal radiation effect on fluctuating thermal conductivity and temperature dependent thermal source.

Electro-magnetohydrodynamics is the dynamic description of the interaction of electrically conducting fluids with the influence of magnetic and electric fields. Having its various uses in magnetic drug targeting, biomedical engineering, engineering, and geophysics and in many other areas, EMHD has gained considerable attention over the years. The idea of flow control with EMHD was firstly used by Resler and Sears [36]. In micro-channel flow regulation Chakraborty and Paul [37] established a

mathematical model to investigate the EMHD effect. Buren et al. [38] investigate the flow of EMHD Newtonian fluid with corrugated walls in a tiny parallel channel. The Shah et al. [39] explore the CNTs rotating EMHD flow using thermal radiation effect passing through a stretching surface.

The goal of the present study is to examine the effect of variable thermal conductivity on the flow of EMHD Sutterby nanofluid passing through a stretching sheet. The numerical solutions is obtained using bvp4c approach built in MATLAB. The bvp4c method simply imposed the 3-stage Lobatto IIIa formula using a system of finite differences. This formulation used a collocation method having fourth-order accuracy and was widely accepted due to its strong stability properties [40-41]. The effect of various variables on velocity, temperature, and concentration profiles is depicted by employing graphs and tables. At last current findings are then compared to earlier research.

2. Problem Formulation

With the presence of Lorentz force the current study examines the 2-Dimensional EMHD (Electromagnetohydrodynamic) Sutterby nanofluid flow of Darcy-Forchheimer along a stretching sheet and the force is given as:

$$\vec{F} = \sigma^{*}\left(\vec{E} + \left(\vec{J} \times \vec{B}\right)\right). \tag{1}$$

The flow is on a sheet with stretching velocity $U_W = cx$. T_W is temperature constant that is expected on sheet surface. In the vertical axis, a constant magnetic field is considered (see Fig. 1).

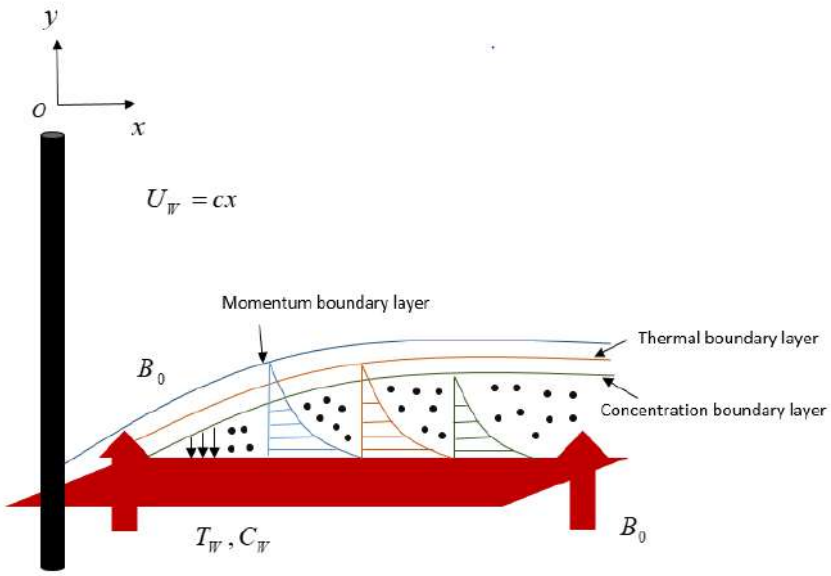


Fig. 1. Flow chart

By using above assumptions, the equations are:

$$\frac{\partial u_1}{\partial x}+\frac{\partial v_1}{\partial y}=0. \tag{2}$$

$$u_1\frac{\partial u_1}{\partial x}+v_1\frac{\partial u_1}{\partial y}=\frac{B_0\sigma^*}{\rho}\left(E_0-B_0u_1\right)+\frac{\upsilon}{2}\frac{\partial^2 u_1}{\partial y^2}\left(1-\frac{M_1b^2}{2}\left(\frac{\partial u_1}{\partial y}\right)^2\right)-Fu_1{}^2-\frac{\upsilon}{\kappa^*}u_1. \tag{3}$$

$$u_1\frac{\partial T_1}{\partial x}+v_1\frac{\partial T_1}{\partial y}=\frac{\mu}{\rho c_p}\left(\frac{\partial u_1}{\partial y}\right)^2+\frac{\partial}{\partial y}\left(\alpha\frac{\partial T_1}{\partial y}\right)+\frac{16\sigma T_\infty{}^3}{3k^*(\rho c)_f}\frac{\partial^2 T_1}{\partial y^2}+\frac{Q_O}{(\rho c_p)}\left(T_1-T_\infty\right) \tag{4}$$

$$+\tau\left[\frac{D_T}{T_\infty}\left(\frac{\partial T_1}{\partial y}\right)^2+D_B\frac{\partial T_1}{\partial y}\frac{\partial C}{\partial y}\right]+\frac{\sigma^*}{\rho c_p}\left(u_1B_0-E_0\right)^2-\lambda^*A*,$$

$$A*=u_1{}^2\frac{\partial^2 T_1}{\partial x^2}+\frac{\partial T_1}{\partial x}\left(u_1\frac{\partial u_1}{\partial x}+v_1\frac{\partial u_1}{\partial y}\right)+\frac{\partial T_1}{\partial y}\left(u_1\frac{\partial v_1}{\partial x}+v_1\frac{\partial v_1}{\partial y}\right)+v_1{}^2\frac{\partial^2 T_1}{\partial y^2}+2u_1v_1\frac{\partial^2 T_1}{\partial x\partial y}. \tag{5}$$

$$u_1\frac{\partial C^*}{\partial x}+v_1\frac{\partial C^*}{\partial y}+k_r\left(C^*-C_\infty\right)=D_B\frac{\partial^2 C^*}{\partial y^2}+\left(\frac{D_T}{T_\infty}\right)\frac{\partial^2 T_1}{\partial y^2}. \tag{6}$$

Here the components of velocity are u_1 and v_1. M_1 represents the deportment index of flow, consistency index is b^2, $F=C_b/x\sqrt{k^*}$ denotes inertia parameter, k^* represents mean absorption coefficient, k_r denotes chemical reaction parameter, E_0 represents the electric field strength, κ^* is the permeability of porosity medium, D_T and D_B are the co-efficient of thermophoretic and Brownian diffusion, C_b denotes the drag parameter, $\alpha=\alpha_\infty\left[1+\varepsilon\theta(\eta)\right]$ represents the variable thermal conductivity, ρ denotes the fluid density, Q_0 denotes heat source, λ^* is the thermal relaxation time and T_1 represents the fluid temperature.

The boundary constraints are:

$$U_W+N_0\frac{\partial u_1}{\partial y}=u_1,\ T_1=T_W+K_0\frac{\partial T_1}{\partial y},\ v_1=0,\ C^*=C_W \text{ at } y=0, \tag{7}$$

$$u_1\to 0,\ T_1\to T_\infty, C^*\to C_\infty \text{ at } y\to\infty.$$

Where the velocity of stretching sheet is $U_W=cx$, T_W denotes the wall temperature, N_0 is the slip length and $N_0=K_0=0$ is observed for no-slip flow.

Using Similarity transformation:

$$u_1 = cxf'(\eta),\ v_1 = -\sqrt{c\upsilon}f(\eta),\ \theta(\eta) = \frac{T_1 - T_\infty}{T_W - T_\infty},\ \phi(\eta) = \frac{C^* - C_\infty}{C_W - C_\infty},\ \eta = \sqrt{\frac{c}{\upsilon}}y. \tag{8}$$

The transformed form of equations are:

$$f''' + 2Ha^2(E - f') - \frac{M_1}{2}\mathrm{Re}\,Def'''(f'')^2 - 2k_p f' - 2(f')^2 - 2F^*(f')^2 + 2ff'' = 0, \tag{9}$$

$$\frac{1}{\mathrm{Pr}}\left(1 + \frac{4}{3}R + \varepsilon\theta\right)\theta'' + EcHa^2(f' - E)^2 + \frac{\varepsilon\theta'^2}{\mathrm{Pr}} + Nt\theta'^2 + Q\theta + Ec(f'')^2 \tag{10}$$
$$+ Nb\theta'\phi' + f\theta' - \gamma\left(ff'\theta' + f^2\theta''\right) + EcHa^2(f' - E)^2 = 0,$$

$$\phi'' + \frac{Nt}{Nb}\theta'' + \mathrm{Pr}\,Le\left[f\phi' - \gamma_1\phi\right] = 0. \tag{11}$$

The transformed boundary constraints are given as::

$$f(\eta) = 0,\ \theta(\eta) = 1 + \beta\theta',\ 1 + \delta f'' = f'(\eta),\ \phi(\eta) = 1 \text{ at } \eta = 0, \tag{12}$$
$$f' \to 0,\ \phi \to 0,\ \theta \to 0 \text{ at } \eta \to \infty.$$

The dimensionless variables occurring in equations

$$\mathrm{Pr} = \frac{\mu c_f}{k},\ Q = \frac{Q_0}{c\left(\rho c_p\right)},\ R = \frac{4\sigma T_\infty^3}{k^* k},\ Nt = \frac{\tau D_T\left(T_W - T_\infty\right)}{\upsilon T_\infty},\ Nb = \frac{\tau D_B\left(C_W - C_\infty\right)}{\upsilon}, \tag{13}$$
$$Nt = \frac{\tau D_T\left(T_W - T_\infty\right)}{\upsilon T_\infty},\ \gamma = \lambda^* c,\ De = b^2c^2,\ Le = \frac{\alpha}{D_B},\ Ha = \sqrt{\frac{\sigma^*}{\rho c}}B_0,\ k_p = \frac{\upsilon}{\kappa^* c},\ \gamma_1 = \frac{k_r}{c},$$
$$F^* = \frac{C_b}{\sqrt{k^*}},\ E = \frac{E_0}{xB_0c},\ \mathrm{Re} = \frac{cx^2}{\upsilon},\ Ec = \frac{c^2x^2}{c_p\left(T_W - T_\infty\right)}.$$

Skin shear stress, mass and heat transfer rate having mass flux j_m, shear stress τ_w and heat flux q_w are:

$$C_f = \frac{\tau_w}{\rho U_w^2},\ Sh_x = \frac{xj_m}{D_B\left(C_w - C_\infty\right)},\ Nu = \frac{xq_w}{k\left(T_w - T_\infty\right)}, \tag{14}$$
$$q_w = -\left[\left(k + \frac{16\sigma T_\infty^3}{3k^*}\right)\frac{\partial T_1}{\partial y}\right],\ \tau_w = -\mu\left[\frac{\partial u_1}{\partial y} + \frac{M_1 b^2}{3}\left(\frac{\partial u_1}{\partial y}\right)^3\right],\ j_m = -D_B\left(\frac{\partial C^*}{\partial y}\right).$$

Equation (8) and (14), gives

$$\begin{aligned} \mathrm{Re}_x^{1/2}\, C_{f_x} &= -\left(f''(0) + (M_1/3)\,\mathrm{Re}\, De \left(f''(0) \right)^3 \right), \\ Nu_x\, \mathrm{Re}_x^{-1/2} &= -\left(1 + \varepsilon\theta(0) + (4/3)R\right)\theta'(0), \\ Sh_x\, \mathrm{Re}_x^{-1/2} &= -\phi'(0). \end{aligned} \tag{15}$$

3. Numerical method of solution

The computational findings of the problem are obtained by employing bvp4c technique in MATLAB coding and problem is converted into first order by defining the new variables.

$$f = y_1, f' = y_2, f'' = y_3, f''' = yy_1, \tag{16}$$

$$\theta = y_4, \theta' = y_5, \theta'' = yy_2, \tag{17}$$

$$\phi = y_6, \phi' = y_7, \phi'' = yy_3. \tag{18}$$

The transformed form of ordinary differential equation is obtained by using equations (16-18) into equations (9-11)

$$yy_1 = \frac{1}{1 - \frac{M_1}{2}\mathrm{Re}\, Dey_3^{\,2}} \left(2y_2^{\,2} - 2Ha^2 \left(E - y_2 \right) - 2y_1 y_3 + 2F^* y_2^{\,2} + 2k_p y_2 \right), \tag{19}$$

$$yy_2 = \frac{1}{1 + \varepsilon y_4 + \frac{4}{3}R} \left(-\varepsilon y_5^{\,2} - EcHa^2 \left(y_2 - E_1 \right)^2 + \mathrm{Pr}\left(-y_1 y_5 + \gamma \left(y_1 y_2 y_5 \right) - Nty_5^{\,2} - Ecy_3^{\,2} - Nby_5 y_7 - Qy_4 \right) \right), \tag{20}$$

$$yy_3 = -\frac{Nt}{Nb} yy_2 - \mathrm{Pr}.Le\left(y_1 y_7 - \gamma_1 y_6 \right). \tag{21}$$

Transformed boundary constraints are:

$$\begin{aligned} &y_1(\eta) = 0,\ y_4(\eta) = 1 + \beta y_5(\eta),\ y_2(\eta) = 1 + \delta y_3(\eta),\ y_6(\eta) = 1 \text{ at } \eta \to 0, \\ &y_2(\eta) = 0,\ y_5(\eta) = 0,\ y_7(\eta) = 0 \text{ at } \eta \to \infty. \end{aligned} \tag{22}$$

4. Results and discussion

This segment depicts the various physical variables impact on various profiles.

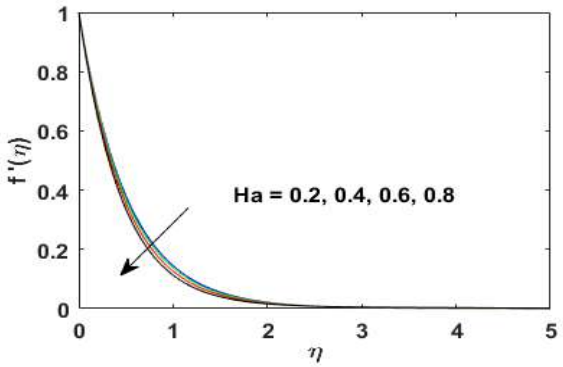


Fig. 2a: Velocity profile $f'(\eta)$ for varying Ha

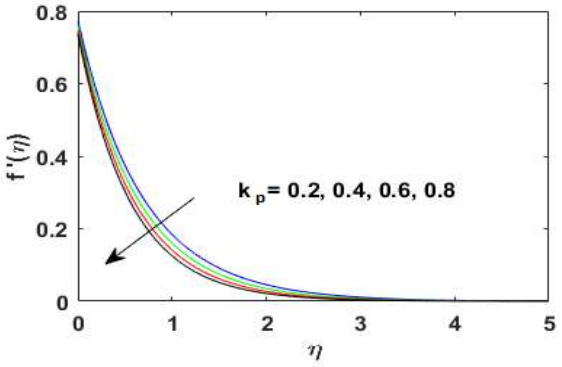


Fig. 2b: Velocity profile $f'(\eta)$ for varying k_p

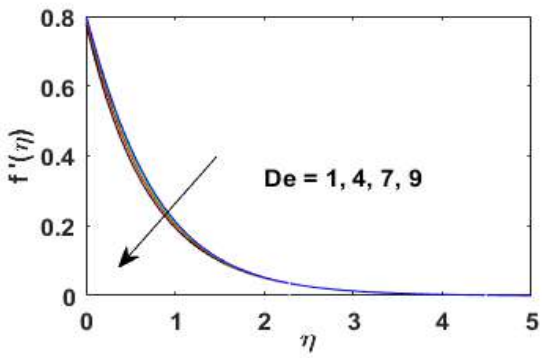


Fig. 3a: Velocity profile $f'(\eta)$ for varying De

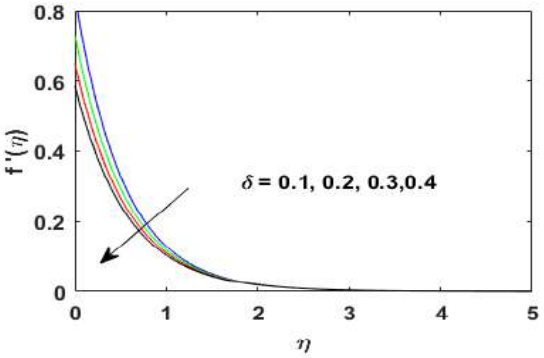


Fig. 3b: Velocity profile $f'(\eta)$ for varying δ

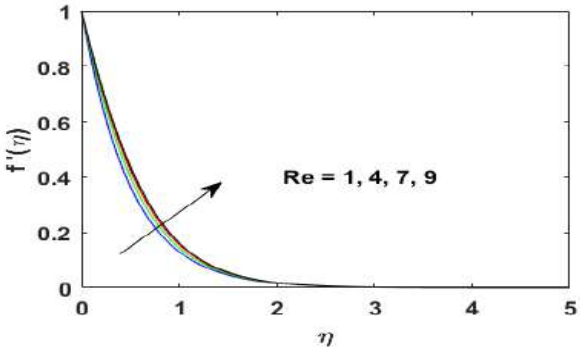


Fig. 4: Velocity profile $f'(\eta)$ for varying Re

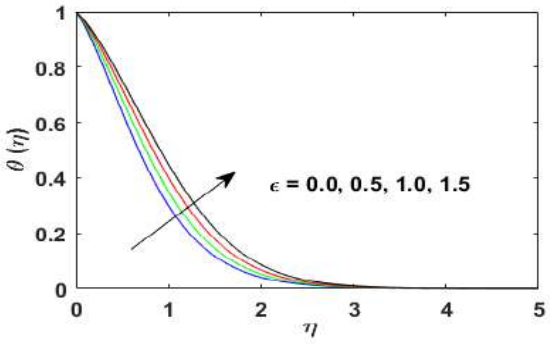


Fig. 5a: Velocity profile $f'(\eta)$ for varying ε

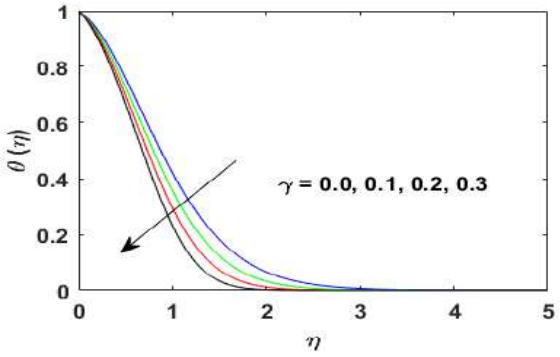


Fig. 5b: Velocity profile $f'(\eta)$ for varying γ

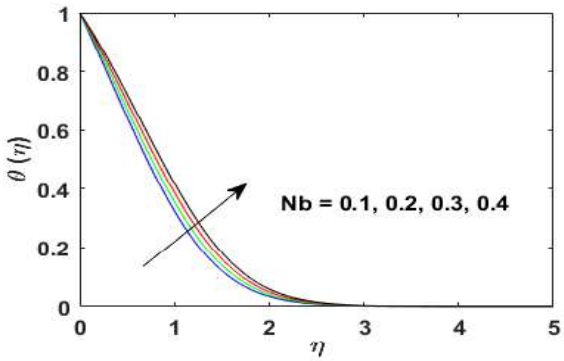


Fig. 6a: Temperature profile $\theta(\eta)$ for varying *Nb*

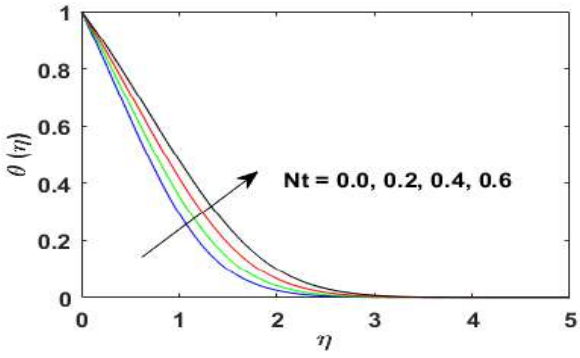


Fig. 6b: Temperature profile $\theta(\eta)$ for varying *Nt*

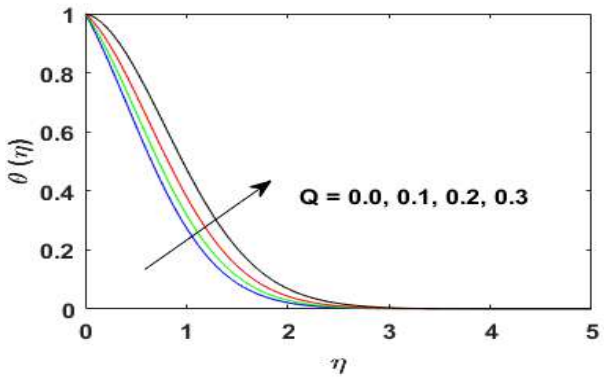


Fig. 7: Temperature profile $\theta(\eta)$ for varying Q

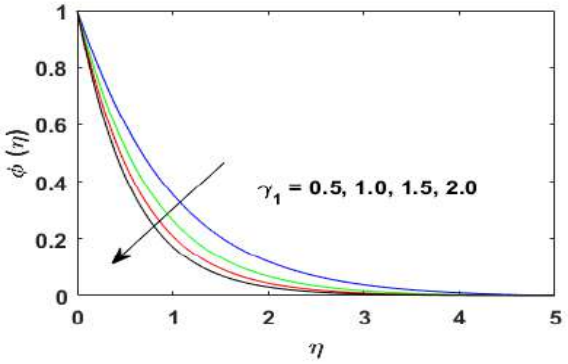


Fig. 8a: Concentration profile $\phi(\eta)$ for varying γ_1

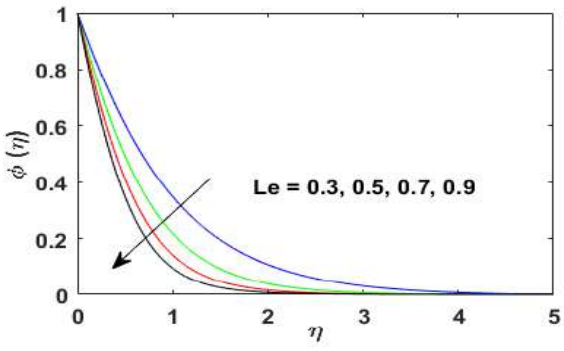


Fig. 8b: Concentration profile $\phi(\eta)$ for varying for Le

Fig 2a. and Fig 2b depict the effect of both the magnetic field parameter Ha and porosity parameter k_p on $f'(\eta)$. It is observed that as the magnetic field variable increases, velocity profile shows decreasing trend (see Fig. 2a). The reason is that the Lorentz force rises physically when Ha increases. Hence increased in resistance to fluid motion that decrease the velocity of fluid. Furthermore, $f'(\eta)$ decelerates as value of k_p is increased (see Fig. 2b). The reason is that porosity parameter is inversely proportion to porous medium permeability. Thus, the lager value of k_p gives greater resistance. The influence of De and velocity slip parameter δ on the $f'(\eta)$ is depicted in Fig.3a and Fig. 3b. The Deborah number measures the material's viscoelastic properties. The larger value of De, means solid substance, the lower value of De means fluid substances. In the figure it is seen that by accelerating the values of De velocity profile decreases (see Fig. 3a). Similarly, velocity profile decreases when the slip factor increases (see Fig. 3b). The reason is that the velocity of fluid along the stretching sheet

are not same with the stretching velocity of sheet due to the slip phenomenon. The impact of Reynolds number Re on $f'(\eta)$ is depicted in Fig. 4. The velocity profile rises as the Reynolds number accelerates. This is due to fact that the Reynold number describes the relationship of inertial and viscous forces, hence the larger value of Reynold number accelerates the velocity profile because of dominance of inertial forces.

In Fig. 5a and Fig. 5b the coefficient of variable thermal conductivity ε and the thermal relaxation γ impact on $\theta(\eta)$ is depicted. In Fig. 5a it is observed that $\theta(\eta)$ rises by increasing the value of the variable thermal conductivity parameter. Increased thermal conductivity implies greater kinetic energy among molecules, leading to a higher number of molecular interactions and, as a result, a direct link among temperature and kinetic energy. Conversely, as the inputs of γ accelerates $\theta(\eta)$ decelerates. This is due to raising thermal relaxation parameter allows material particles to take larger time to exchange heat with nearby particles (see Fig. 5b). In Fig. 6a and Fig. 6b it is noticed that the fluid temperature accelerates as inputs of the Nb and Nt enhances. The reason is that the by enhancing the Brownian motion parameter intensifies the nanofluid thermal conductivity hence temperature profiles increase (see Fig. 6a). In the same way as the value of Nt increases, the thermophoresis force accelerates, leading the nanoparticles to migrate from hot to cold region hence the temperature profile to accelerate (see Fig. 6b). In Fig. 7 the temperature profile fluctuates as the value of heat source parameter Q accelerates. The reason is that when heat generation increases the fluids temperature rises.

In Fig. 8a and Fig. 8b. the impact of scaled chemical reaction parameter γ_1 and the Lewis number Le on the concentration profile is depicted. By changing the values of γ_1, decelerates the $\phi(\eta)$. This is due to fact that rising the impact of γ_1 accelerates the intermolecular mass transfer rate which causes the volumetric reduction in nanoparticles (see Fig. 8a). Moreover, as the Lewis number rises then concentration profile diminishes (see Fig. 8b). The fact is that the rate of mass transfer rises as the Lewis number accelerates.

Tables 1, 2 and 3 demonstrate the impact of different physical variables on Nusselt number, skin friction and mass transfer rate. In Table 1 it can be noticed that skin friction coefficient accelerates for k_p, Re and De, but reduces for M. In Table 2 it is noticed that heat transfer rate accelerates by accelerating the values of R but decelerate with Nb and Nt. In Table 3 it is discovered that $-\phi'(0)$ increases for Nt, Le but decreases for Nb.

Table 1: Impact of k_p, De, Re, M on $\mathrm{Re}_x^{\frac{1}{2}} C_f$.

De	Re	M	k_p	$\mathrm{Re}_x^{\frac{1}{2}} C_f$
-	-	-	0.1	0.7280
0.1	0.1	0.1	0.2	0.7401
-	-	-	0.3	0.7575
-	-	0.1	-	0.7382
0.1	0.1	0.2	0.1	0.7348
-	-	0.3	-	0.7296
-	2	-	-	0.7310
0.2	3	0.1	0.1	0.7490
-	4	-	-	0.8140
0.2	-	-	-	0.7480
0.4	3	0.1	0.1	0.7721
0.6	-	-	-	0.7939

Table 2: Impact of Nb, Nt and R on $Nu_x \mathrm{Re}_x^{-1/2}$

Nt	Nb	R	$Nu_x \mathrm{Re}_x^{-1/2}$
-	0.1	-	0.164715
0.2	0.2	0.2	0.163103
-	0.3	-	0.161498
-	-	0.1	0.170202
0.1	0.1	0.2	0.172369
-	-	0.4	0.175646
0.1	-	-	0.168343
0.2	0.2	0.1	0.166608
0.3	-	-	0.164880

Table 3: Impact of Nt, Le and Nb on $-\phi'(0)$

Nt	Le	Nb	$-\phi'(0)$
-	0.1	-	0.23388
0.1	0.2	0.1	0.244731
-	0.3	-	0.255486
-	-	0.1	0.260419
0.1	0.1	0.2	0.237574
-	-	0.3	0.229952
0.1	-	-	0.234226
0.2	0.2	0.2	0.248435
0.3	-	-	0.264422

5. Conclusions

In this study, the impact of variable thermal conductivity on two-dimensional EMHD flow of Sutterby nanofluid by considering Cattaneo Christov heat flux model in the presence of porous media over a stretching sheet was analyzed. The graphical outcomes are examined by varying the input variables. The findings are:

- The effect of Hartmann number Ha , velocity slip parameters δ, Deborah number De and Porosity parameter k_p decreases the velocity profile.
- The velocity distribution accelerates by increasing the Reynold number Re.
- The accelerating values of γ declines the temperature of fluid but the larger input of Q, Nt, Nb and ε increases the temperature.
- The concentration profile decelerates by using the larger inputs of Le and γ_1 .
- The skin shear stress accelerates by enhancing the values of k_p, Re and De but decelerates with accelerating values of M .
- The Nusselt number diminishes by rising the inputs of Nb and Nt but rises as radiation parameter increases.
- Sherwood number increases for Nt and Le, but decrease with increasing value of Nb .

6. Limitation and Future work

Present study has the following limitations

- ❖ Present study is two dimensional, with no slip boundary conditions.
- ❖ Other thermophysical quantities are not taken variable.
- ❖ Grid independent study is not carried out in the presence analysis.

The present study can be extended to:

- ❖ Carry out the sensitivity analysis.
- ❖ Convergence analysis to verify the authenticity of the numerical solution obtained.
- ❖ The electroosmotic flow can be considered as the electric and magnetic field are both applied.

Funding

The authors declare that this research received no external funding

Data Availability Statement

Data is given withing the manuscript.

Conflicts of Interest

The authors declare no conflict of interest.

References

[1] Sharma, G., Kumar, A., Sharma, S., Naushad, M., Dwivedi, R. P., ALOthman, Z. A., & Mola, G. T. (2019). Novel development of nanoparticles to bimetallic nanoparticles and their composites: a review. *Journal of King Saud University-Science*, *31*(2), 257-269.

[2] Su, F., Deng, Y., & Ma, H. (2015). Numerical analysis of ammonia bubble absorption in a binary nanofluid. *Chemical Engineering Communications*, *202*(4), 500-507.

[3] Zarifi, E., Jahanfarnia, G., & Veysi, F. (2013). Subchannel analysis of nanofluids application to VVER-1000 reactor. *Chemical Engineering Research and Design*, *91*(4), 625-632.

[4] Sreeremya, T. S., Krishnan, A., Mohamed, A. P., Hareesh, U. S., & Ghosh, S. (2014). Synthesis and characterization of cerium oxide based nanofluids: An efficient coolant in heat transport applications. *Chemical Engineering Journal*, *255*, 282-289.

[5] Tripathi, D., & Bég, O. A. (2014). Mathematical modelling of peristaltic pumping of nano-fluids. In *Modelling and Simulation of Diffusive Processes* (pp. 69-95). Springer, Cham.

[6] Ali, B., Hussain, S., Naqvi, S. I. R., Habib, D., & Abdal, S. (2021). Aligned magnetic and bioconvection effects on tangent hyperbolic nanofluid flow across faster/slower stretching wedge with activation energy: Finite element simulation. *International Journal of Applied and Computational Mathematics*, *7*(4), 1-20.

[7] Ali, B., Rasool, G., Hussain, S., Baleanu, D., & Bano, S. (2020). Finite element study of magnetohydrodynamics (MHD) and activation energy in Darcy–Forchheimer rotating flow of Casson Carreau nanofluid. *Processes*, *8*(9), 1185.

[8] Choi, S. U., & Eastman, J. A. (1995). *Enhancing thermal conductivity of fluids with nanoparticles* (No. ANL/MSD/CP-84938; CONF-951135-29). Argonne National Lab., IL (United States).

[9] Bachok, N., Ishak, A., & Pop, I. (2011). Stagnation-point flow over a stretching/shrinking sheet in a nanofluid. *Nanoscale research letters*, *6*(1), 1-10.

[10] Ali, B., Thumma, T., Habib, D., & Riaz, S. (2022). Finite element analysis on transient MHD 3D rotating flow of Maxwell and tangent hyperbolic nanofluid past a bidirectional stretching sheet with Cattaneo Christov heat flux model. *Thermal Science and Engineering Progress*, *28*, 101089.

[11] Hamad, M. A. A. (2011). Analytical solution of natural convection flow of a nanofluid over a linearly stretching sheet in the presence of magnetic field. *International communications in heat and mass transfer*, *38*(4), 487-492.

[12] Ibrahim, W., Shankar, B., & Nandeppanavar, M. M. (2013). MHD stagnation point flow and heat transfer due to nanofluid towards a stretching sheet. *International journal of heat and mass transfer*, *56*(1-2), 1-9.

[13] Sutterby, J. L. (1966). Laminar converging flow of dilute polymer solutions in conical sections: Part I. Viscosity data, new viscosity model, tube flow solution. *AIChE Journal*, *12*(1), 63-68.

[14] Batra, R. L., & Eissa, M. (1994). Helical flow of a Sutterby model fluid. *Polymer-Plastics Technology and Engineering*, *33*(4), 489-501.

[15] Ahmad, S., Farooq, M., Javed, M., & Anjum, A. (2018). Double stratification effects in chemically reactive squeezed Sutterby fluid flow with thermal radiation and mixed convection. *Results in physics*, *8*, 1250-1259.

[16] Bilal, S., Sohail, M., Naz, R., & Malik, M. Y. (2020). Dynamical and optimal procedure to analyze the exhibition of physical attributes imparted by Sutterby magneto-nanofluid in Darcy medium yielded by axially stretched cylinder. *Canadian Journal of Physics*, *98*(1), 1-10.

[17] Hayat, T., Bibi, F., Khan, A. A., & Alsaedi, A. (2021). Entropy production minimization and non-Darcy resistance within wavy motion of Sutterby liquid subject to variable physical characteristics. *Journal of Thermal Analysis and Calorimetry*, *143*(3), 2215-2225.

[18] Mir, N. A., Farooq, M., Rizwan, M., Ahmad, F., Ahmad, S., & Ahmad, B. (2020). Analysis of thermally stratified flow of Sutterby nanofluid with zero mass flux condition. *Journal of Materials Research and Technology*, *9*(2), 1631-1639.

[19] Bilal, S., Sohail, M., Naz, R., Malik, M. Y., & Alghamdi, M. (2019). Upshot of ohmically dissipated Darcy-Forchheimer slip flow of magnetohydrodynamic Sutterby fluid over radiating linearly stretched surface in view of Cash and Carp method. *Applied Mathematics and Mechanics*, *40*(6), 861-876

[20] Forchheimer, P. (1901). Wasserbewegung durch boden. *Z. Ver. Deutsch, Ing.*, *45*, 1782-1788.

[21] Muskat, M. (1938). The flow of homogeneous fluids through porous media. *Soil Science*, *46*(2), 169.

[22] Khan, M. I., Alzahrani, F., & Hobiny, A. (2020). Simulation and modeling of second order velocity slip flow of micropolar ferrofluid with Darcy–Forchheimer porous medium. *Journal of Materials Research and Technology*, *9*(4), 7335-7340.

[23] Eid, M. R., & Mabood, F. (2021). Entropy analysis of a hydromagnetic micropolar dusty carbon NTs-kerosene nanofluid with heat generation: Darcy–Forchheimer scheme. *Journal of Thermal Analysis and Calorimetry*, *143*(3), 2419-2436.

[24] Ramzan, M., Gul, H., & Zahri, M. (2020). Darcy-Forchheimer 3D Williamson nanofluid flow with generalized Fourier and Fick's laws in a stratified medium. *Bulletin of the Polish Academy of Sciences. Technical Sciences*, *68*(2).

[25] Cattaneo, C. (1948). Sulla conduzione del calore. *Atti Sem. Mat. Fis. Univ. Modena*, *3*, 83-101.

[26] Straughan, B. (2010). Thermal convection with the Cattaneo–Christov model. *International Journal of Heat and Mass Transfer*, *53*(1-3), 95-98.

[27] Tibullo, V., & Zampoli, V. (2011). A uniqueness result for the Cattaneo–Christov heat conduction model applied to incompressible fluids. *Mechanics Research Communications*, *38*(1), 77-79.

[28] Han, S., Zheng, L., Li, C., & Zhang, X. (2014). Coupled flow and heat transfer in viscoelastic fluid with Cattaneo–Christov heat flux model. *Applied Mathematics Letters*, *38*, 87-93.

[29] Hayat, T., Farooq, M., Alsaedi, A., & Al-Solamy, F. (2015). Impact of Cattaneo-Christov heat flux in the flow over a stretching sheet with variable thickness. *Aip Advances*, *5*(8), 087159.

[30] Chiam, T. C. (1996). Heat transfer with variable conductivity in a stagnation-point flow towards a stretching sheet. *International Communications in Heat and Mass Transfer*, *23*(2), 239-248.

[31] Chiam, T. (1998). Heat transfer in a fluid with variable thermal conductivity over a linearly stretching sheet. *Acta mechanica*, *129*(1), 63-72.

[32] Gbadeyan, J. A., Titiloye, E. O., & Adeosun, A. T. (2020). Effect of variable thermal conductivity and viscosity on Casson nanofluid flow with convective heating and velocity slip. *Heliyon*, *6*(1), e03076.

[33] Hamid, M., Usman, M., Zubair, T., Haq, R. U., & Wang, W. (2018). Shape effects of MoS2 nanoparticles on rotating flow of nanofluid along a stretching surface with variable thermal conductivity: A Galerkin approach. *International Journal of Heat and Mass Transfer*, *124*, 706-714.

[34] Salawu, S. O., & Dada, M. S. (2016). Radiative heat transfer of variable viscosity and thermal conductivity effects on inclined magnetic field with dissipation in a non-Darcy medium. *Journal of the Nigerian Mathematical Society*, *35*(1), 93-106.

[35] Subhas, A., Prasad, K. V., & Mahaboob, A. (2005). Buoyancy force and thermal radiation effects in MHD boundary layer viscoelastic flow over continuously moving stretching surface. *Int. J. Therm. Sci*, *44*, 465-476.

[36] Resler Jr, E. L., & Sears, W. R. (1958). The prospects for magneto-aerodynamics. *Journal of the Aerospace Sciences*, *25*(4), 235-245.

[37] Chakraborty, S., & Paul, D. (2006). Microchannel flow control through a combined electromagnetohydrodynamic transport. *Journal of Physics D: Applied Physics*, *39*(24), 5364.

[38] Buren, M., Jian, Y., & Chang, L. (2014). Electromagnetohydrodynamic flow through a microparallel channel with corrugated walls. *Journal of Physics D: Applied Physics*, *47*(42), 425501.

[39] Shah, Z., Bonyah, E., Islam, S., & Gul, T. (2019). Impact of thermal radiation on electrical MHD rotating flow of Carbon nanotubes over a stretching sheet. *AIP Advances*, *9*(1), 015115.

[40] Laurent, O. J. (1996). Lobatto methods. *URL: https://homepage. math. uiowa. edu/~ ljay/publications. dir/Lobatto.*

[41] Ascher, U. M., Mattheij, R. M., & Russell, R. D. (1995). *Numerical solution of boundary value problems for ordinary differential equations*. Society for Industrial and Applied Mathematics.

Islamic University Journal of Applied Sciences (IUJAS)

https://journals.iu.edu.sa/jesc

Volume VIII, Issue II, December 2026, Pages 346-366

Form over Substance: A Metric-Decomposition Diagnosis of Group Relative Policy Optimization for Arabic Legal Reasoning

Nawaf Al Enezi [1, *], Abdullah Alshanqiti [1], Ahmad B. Alkhodre [1]

[1] Faculty of Computer and Information Systems, Islamic University of Madinah, Madinah, KSA

451027326@stu.iu.edu.sa ; amma@iu.edu.sa ; aalkhodre@iu.edu.sa

*Corresponding author: Nawaf Al Enezi, *Email Address*: **451027326@stu.iu.edu.sa**

Abstract

Reinforcement learning with verifiable rewards (RLVR) drives much post-training of reasoning models, yet it is judged by a rising reward curve that cannot indicate which of several rule-based verifiers earned the gain. This study makes that observable during training through the weight-zero monitor: the substantive verifier is registered a second time at weight 0.0 and logged alongside the within-group standard deviation of each component, so that it cannot be optimized against. It was applied within an internally registered falsification program of Group Relative Policy Optimization (GRPO) on Qwen3-8B over Saudi commercial rulings. The registered attempt moved the policy without improving it: every greedy completion changed, 54% of held-out dispositions changed, macro-F1 fell by 0.067 (95% CI [−0.136, +0.011]), and repetition halved while the monitor remained flat. Acting on this reading, eight-sample majority voting, which trains nothing, achieved the only significant gain, +0.088 macro-F1 on a frozen 300-case instrument (95% CI [0.043, 0.140]). We derive when a gated composite reward rewards form rather than substance, and we recommend measuring the extractable headroom before allocating compute to policy-gradient training.

Keywords: Arabic legal natural language processing; reinforcement learning with verifiable rewards; group relative policy optimization; reward decomposition; legal judgment prediction; self-consistency.

https://doi.org/10.63070/jesc.2026.046

Received 20 August 2026; Revised 2 October 2026; Accepted 10 October 2026.Available online 16 October 2026.

Published by Islamic University of Madinah on behalf of *Islamic University Journal of Applied Sciences.*

This is an open access article published under the Creative Commons Attribution 4.0 International license (CC BY 4.0): https://creativecommons.org/licenses/by/4.0/

1. Introduction

1.1 Problem statement

Group Relative Policy Optimization (GRPO) replaces the learned value function of proximal policy optimization with a group-relative baseline, and consequently it made reinforcement learning with verifiable rewards (RLVR), i.e., policy-gradient training against deterministic rule-based verifiers, affordable at post-training scale [1], [2]. The standard evidence that such a run has worked is a rising aggregate reward curve. When the reward is a composite (i.e., a weighted sum of several verifiers), this evidence is unsound, since the sum allows easily saturated components (e.g., a format check or a repetition check) to accumulate reward that the substantive component, namely the correct answer, never earns. Throughout this paper, we call form what a verifier can check without reading the case (the well-formedness of the output, its length, its repetition rate) and substance the operative disposition itself, i.e., the quantity the reward exists to improve. This is Goodhart's law [3] in an unusually clean form, as the proxy is an arithmetic sum of verifiers that are each correct on their own, and hence there is no learned judge whose errors could be audited. The parties affected are practitioners who post-train with composite verifiable rewards, most acutely in low-resource domains such as Arabic legal reasoning, where compute budgets are small and a misinterpreted run is costly. Such a practitioner needs a readout during the run, and none of the usual sources provides one: held-out evaluation is expensive and infrequent, the per-step reward is an aggregate and thus ambiguous, and the literature that motivates per-component logging [4] does not indicate whether a particular run is acquiring substance or form.

In this paper, we consider Arabic legal judgment prediction, i.e., predicting the operative disposition of a ruling from the case record, over Saudi first-instance commercial rulings. The published record on this task documents class imbalance and weak minority-class performance [5]–[7], reports that the customary automatic metrics are unreliable in this jurisdiction [8]–[11], and includes field-level critiques finding that much of the field does not measure what it claims to measure [12], [13]. Two properties of that record bear on our instrument. First, the disposition space is dominated by a single class, so a reward term that rewards surface form is easily satisfied while a class-balanced score such as macro-F1 remains unchanged. Second, the substantive endpoint (i.e., which disposition the court reached) already has an exact verifier, so reading it a second time costs nothing. To our knowledge, no existing instrument reports a property of one run at one step, namely which component is being learned now; this paper addresses that gap.

1.2 Contributions

The program reported here was designed to raise the held-out disposition accuracy of a distilled Qwen3-8B student with GRPO, beyond the level that supervised distillation had reached. It did not: the policy changed, but the disposition did not improve. This paper reports the instrument that made that failure observable while it was occurring, the diagnosis it supports, and the decision rule that recovered a gain without training. The originality and main contributions of this paper are outlined as follows:

- We introduce the weight-zero monitor, which duplicates the verifier of the substantive endpoint, computes it on every rollout, and registers it in the objective at weight 0.0, so that it

cannot be optimized against while measuring exactly the quantity the objective is meant to improve. It is logged alongside the within-group standard deviation of every reward component, at the cost of one deterministic function call per rollout.

- We report a diagnosis of the registered run, in which the objective was the disposition verifier itself with three monitors logged at weight zero. The logged aggregate did not rise, the substance signal stayed flat, and the held-out disposition score fell, while repetition halved and completions shortened. The movement, in other words, went into the quantities rewarded by the degenerate-group tie-break. That reading is consistent with the traces rather than established by an ablation.
- We propose a decision rule. Rather than iterating on the reward design, the program measured whether the capability being optimized for was already present in the sampling distribution (a quantity we call extractable headroom) and selected the next intervention accordingly; this selection produced the only significant improvement of the program, from an intervention that updates no parameters. The move has precedents [14]–[16], whose predictors are pre-hoc and aggregate; our monitor is in-training and per-component, and it answers a question they cannot, namely which component of a composite is being rewarded while the run is in progress.

1.3 Scope and paper organization

Regarding scope, our claims apply to the monitoring protocol and the decision rule, not to the accuracy levels observed (one court, one jurisdiction, one base model, and one reward family). The null reported here is therefore a null about this program rather than about verifiable-reward training in general.

2. Related Work

This section states, for each body of work our design rests on, what it settles and what it leaves to us: how composite rewards fail under group-relative optimization (Section 2.1), how such failures are detected while a run is in progress (Section 2.2), and what the Arabic and Saudi legal judgment prediction literature establishes about our testbed (Section 2.3). Our reading is that the first explains failures after the fact, the second is built for learned reward models rather than deterministic verifiers, and the third has met neither.

2.1 Composite rewards under group-relative optimization

GRPO samples a group of completions per prompt, scores each with the reward, and uses the within-group standardized reward as the advantage, dispensing with the learned value function of proximal policy optimization [1], [2]. Reward variance within a group is therefore a precondition for learning rather than a side effect: the expected policy gradient vanishes when the reward standard deviation is small even when the expected reward is far from optimal [17], and under binary rewards the fraction of degenerate groups exceeds the independent-Bernoulli prediction [18]. Entropy collapse and degenerate groups are documented at scale [19].

The line closest to our own concerns what happens when the reward is a sum of several verifiers. GDPO shows that normalizing after aggregation destroys component identity, collapsing distinct reward combinations into identical advantages, and that normalizing each component by its own

standard deviation restores each to exactly its weight [4]; MO-GRPO reaches the same repair for multi-objective settings [20]. The diagnosis has since been sharpened: MDP-GRPO names three separate pathologies of z-score normalization under discrete low-dispersion rewards, including zero-variance collapse [21], and a local covariance law states the condition under which an objective improves at all [22]. Against that, format and length signals alone have been shown to dominate early training and to account for most of the apparent gain [23], which is the specific risk our instrument is built to read. We do not read this literature as unanimous. Against the collapse narrative, standard-deviation normalization has been shown to implement an adaptive gradient with a provably improved convergence rate over unnormalized REINFORCE, governed by the average within-prompt reward standard deviation [24]: the mechanism that destroys resolution in a composite is what makes GRPO converge quickly when variance is healthy, which is why Section 5.3 treats the gate as a variance source rather than a defect.

Two further results bear on the outcome we report. First, the coupling between supervised fine-tuning and subsequent reinforcement learning: the two are proved not to decouple in either order, with the predicted degradation confirmed on a Qwen3 model [25]; checkpoints that have received excessive supervised fine-tuning lose the plasticity that policy-gradient training needs [26], while on-policy reinforcement learning prefers the Kullback–Leibler (KL)-minimal solution among those that solve the task [27]; and GRPO has been characterized as amplifying capabilities a model already has where supervised fine-tuning replaces them [28]. A distilled parent is the configuration these results warn about. Second, a deflationary reading that a reviewer may fairly raise against any GRPO-specific claim: under the standard structural assumptions, GRPO reduces to filtered iterative supervised fine-tuning, which matches it empirically across model families [29]. We do not adjudicate it here; it would not change our null, only its name.

Running against all of the above is a positive result in the setting that ought to have favored ours: deterministic rule-based verifiers applied to intermediate reasoning steps in a guideline-defined domain report up to 20% higher F1 than the prior state of the art and 6.5% over verifiable outcome rewards [30]; Section 5.4 takes that counter-case seriously.

2.2 Detecting reward hacking while a run is in progress

That a detection signal must be held outside the objective is established rather than assumed. The proxy score of a learned reward model keeps rising after the gold score it approximates has peaked [31]; folding a chain-of-thought monitor into the training reward yields obfuscated reward hacking, so a monitorability tax must be paid [32]. A recent survey organizes detection and mitigation by what they intervene on [33], and empirical work supplies invariance-based stress tests with early-warning signals [34].

What this literature does not supply, to our knowledge, is a per-component readout of a live run whose verifier is exact. Its instruments are built for learned reward models, where the monitor is itself a model. The nearest relative is a diagnostic rather than a monitor: the Advantage Collapse Rate quantifies the proportion of batches with ineffective gradients and predicts training stagnation [35]. We regard our contribution as extending that idea from the aggregate to the component, in the deterministic case where the endpoint already has an exact verifier and reading it again costs one function call per rollout.

We note one disagreement our result speaks to. One line argues that adding format or process components to the reward suppresses hacking, in medical question answering [36] and through process-outcome consistency filtering [37]; another finds that such components capture the gradient and bias the policy toward surface properties, with static-analysis rewards shortening completions without improving functional correctness [38]. When a format component is a safeguard and when it is the hack is, in our reading, unresolved; Section 5.3 offers a partial criterion.

2.3 Arabic and Saudi legal judgment prediction

Legal judgment prediction is surveyed in [5] and, for the legal domain more broadly, in [39]; LEXTREME and LexGLUE frame the task and its metrics [6], [7]. For Arabic the record is recent and concentrated. Ammar et al. benchmark fourteen model variants on 10,813 Saudi commercial cases and report that every automatic metric except human evaluation was inconsistent [8]; Kmainasi et al. build an Arabic dataset from the same court and show that low-rank adaptation (LoRA) fine-tuning of small models matches larger ones [9]; Almalki et al. reach 85.62% on a binary accept-or-reject formulation over 19,822 cases with sentence-level AraBERT and aggregation [10]; and ALARB adds 13K cases with court reasoning and cited clauses, demonstrating instruction tuning to near-frontier verdict prediction [11]. Alongside these sit an Arabic legal knowledge benchmark [40] and further Saudi commercial studies [41], [42].

Two findings from this literature shape our evaluation rather than merely preceding it. The first is critical: a substantial part of the field has been argued not to measure what it claims to measure [12], [13], and this has been quantified elsewhere in a low-resource language, where the model with the highest accuracy on an imbalanced legal task was a majority-class predictor scoring 23% macro-F1 while the genuinely best model scored 44% [43]. Our disposition space is dominated by one class, which is why macro-F1 and not accuracy is the criterion throughout this paper. The second concerns reproducibility: standardized evaluation of reasoning claims shows most reported reinforcement-learning gains to be far smaller than claimed, with seeds, decoding parameters, and prompt formatting accounting for much of the difference [44]. We report a null, so it does not threaten our result; it sets the standard our reporting tries to meet.

The literature therefore establishes that composite rewards fail through normalization and gate effects, that the repairs are per-component, that a detection signal must stay outside the objective, and that Arabic legal judgment prediction is dominated by a single class and poorly served by its customary metrics. What it does not provide is a per-step, per-component readout of a live run with an exact verifier, nor a registered program that adjudicates GRPO on this task with class-balanced scores and rules fixed in advance; those two gaps determine Sections 3.2 and 3.3.

3. Methods

3.1 Experimental program and evaluation instruments

The program was specified before execution as a set of single-variable interventions, or arms, branched from one parent checkpoint. Each arm varies exactly one factor, and each is scored on the primary evaluation set by the same rule-based disposition classifier that produced the reference labels. Nine paired evaluations across five intervention families were adjudicated: the registered final attempt, an

independently authored conservative recipe, six further optimizer families, and one arm that changed the learning rate alone; a first probe under corrected numerical settings preceded them as an instrument check and is reported in Section 4.3 but not counted. Every arm carried its decision rule in advance.

The design separates two questions that a single run confounds. The first is mechanical: did the optimizer update the model at all? The second is substantive: did the update change what the model decides on cases it has not seen? Table 2 reports both for every arm, and that separation is what makes a null readable. The student is Qwen3-8B [45], adapted with LoRA [46], in which small trainable matrices are added to the frozen base weights, trained with TRL's GRPO trainer (TRL 0.29.1, PEFT 0.19.1 (parameter-efficient fine-tuning), Transformers 5.12.1, PyTorch 2.11.0) and evaluated through vLLM 0.25.1 [47]; Table 3 lists the configuration. The adapter is a rank-64 LoRA on every linear projection, the condition under which the learning-rate analysis of Section 5.2 can rely on [48]. The corpus is split at the level of the case, so no ruling contributes text to more than one split, and the held-out split is never trained on and never used for checkpoint selection. Prompts longer than 3,072 tokens are dropped rather than truncated, so that no training prompt carries a partial case record; approximately 20,000 unique prompts remain for the optimization stage. A structural audit over the line found fully well-formed output, with mean coverage of the reasons and ruling sections of 0.966 and 0.946 against the source documents. Table 1 collects the data quantities at each stage.

Table 1. Data quantities at each stage. Splitting is at the level of the case, so no ruling contributes text to more than one split.

Stage	n	Role
Distillation traces	1,626	Verified chain-of-thought supervision for the parent checkpoint
Optimization prompts	≈20,000	Policy-gradient stage, after the 3,072-token length filter
Adjudication slice	200	Per-arm paired evaluation (training-lever family); drawn from the held-out split
Frozen instrument	300	Inference-composition family; held out from all training
External comparison set	200	Frontier contrasts; leakage-audited

3.2 The weight-zero monitor

A weight-zero monitor is an ordinary reward function registered with the trainer at weight 0.0: it is computed on every rollout and logged at every step, and it contributes nothing to the advantage, so no optimization pressure can corrupt it and it measures without bias the quantity the composite is intended to improve. The per-component within-group standard deviations sharpen the reading at the same cost: for a binary verifier with success probability p, the within-group standard deviation is $\sqrt{(p(1-p))}$, so a component that is being learned shows up as a collapse of its variance, and one that is being ignored as a flat band. Three properties matter: the monitor must be the endpoint's own verifier rather than a proxy; it must be computed on every rollout from the first step rather than at checkpoints; and it must be logged beside the per-component standard deviation, or a flat monitor is consistent with several mechanisms and identifies none. TRL already logs the mean and standard deviation of every registered reward function and accepts a weight of 0.0 [49]; the contribution is the prescription for using that telemetry. Throughout, we use exactly two terms for this instrument and its output: the weight-zero monitor is the registered verifier itself, and the substance signal is the reading it logs.

Before trusting the monitor, we characterized it off-line on stored rollouts (see Fig. 1). *Calibration.* With the fraction of correct completions in a group held at known values, the monitor's mean reproduces that fraction exactly; the within-group standard deviation is not monotone, peaking at 0.500

when half the group is correct and returning to zero at both extremes, so a reading of 0.433 is consistent with a success probability of 0.25 or 0.75, and direction is read from the mean. *Resolution.* At the 16 completions per update this program used, a single reading near a success probability of one half carries a standard error of 0.124; detecting a true shift of 0.10 requires averaging over roughly 25 steps, and a shift of 0.05 roughly 99 steps, so the 100-step arms can see a large movement in substance and not a small one, and the null of Section 4.2 is a null at that resolution. *Controls.* Permuting the reference labels 4,000 times drops the monitor from 0.450 to 0.192 (95% interval [0.157, 0.230]), so it reads the verifier rather than its own sampling variation; applied to two checkpoints whose capability differs by construction, the distilled parent and the same student after targeted verified injection, it moves in the correct direction, from 0.4500 to 0.5067, a paired difference of 0.0567 with an interval of [−0.0033, 0.1167]: it resolves checkpoints separated by about a tenth and not by a twentieth. Read on the training batch, the step-to-step series is over-dispersed relative to a constant-ability model (permutation p = 0.001) with no trend (p = 0.93): with two cases per update, the reading reflects which cases were drawn, so the monitor should be computed on a probe set held fixed across steps, as Section 5.5 states.

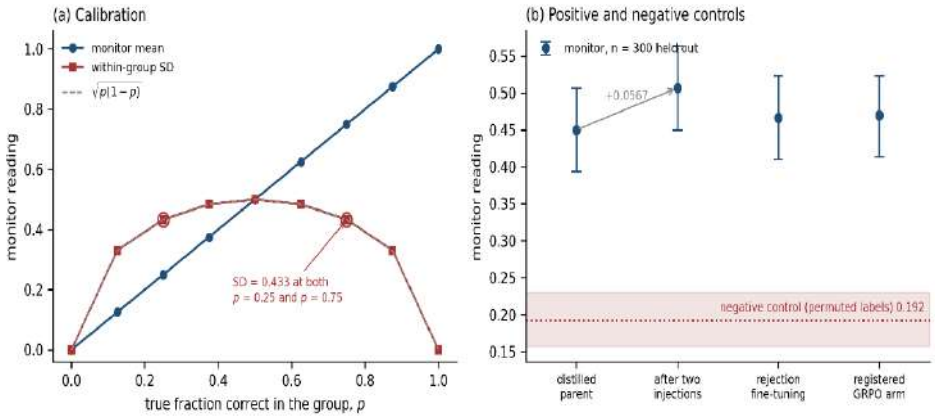


Fig. 1. Validating the weight-zero monitor off-line. (a) Calibration against a known fraction correct: the mean is exact; the within-group standard deviation is non-monotone. (b) Controls on the frozen 300-case instrument: permuting the reference labels collapses the reading to the shaded band, and the monitor separates the distilled parent from the injection-augmented student in the correct direction with an interval that touches zero.

3.3 Data, reward design, and falsification protocol

Our parent checkpoint is a Qwen3-8B student distilled on 1,626 verified chain-of-thought traces over published Saudi first-instance commercial rulings, in the reasoning-distillation line that transfers reasoning from a larger teacher [50]. Each trace was admitted only after clearing deterministic acceptance checks against the source document. Rewards are pure functions of the completion and of the gold verifiable fields: deterministic, GPU-free, and unit-tested. The objective actually optimized in the registered attempt is a single binary driver with three weight-zero monitors: $R = 1.0 \cdot D + 0.0 \cdot M_{\text{dec}} + 0.0 \cdot M_{\text{amt}} + 0.0 \cdot M_{\text{deg}}$. The driver D reads the disposition alone: a completion that fails the format gate (one reasoning block, the required section tags in order, a non-empty operative section) scores −0.0725; one that passes scores $(r - \rho_0)/(1 - \rho_0 - \rho_1)$, where $r \in \{0, 1\}$ records whether the extracted disposition matches the reference and $\rho_0 = 0.02$, $\rho_1 = 0.09$ are the audited reference-label noise rates, so a correct disposition scores 1.101, an incorrect one −0.022, and the gate score lies 0.05 below that of an incorrect answer. When all eight completions of a group score alike, a tie-break of at most 0.1 is

added, 0.06 for the fraction of reference amounts recovered and 0.04 for low repetition; the recorded degenerate-group rate was 0.19. The amount verifier enters the objective only through that tie-break, and the partial-credit driver analyzed in Section 5.3 (−0.5 behind a closed gate, $0.3 + 0.7 \cdot c$ behind an open one) is a design the program specified and no recorded arm ran; the structural partial-credit arm of Table 2 changed only the degenerate-group tie-break, not the driver. The monitors are the disposition match (M_{dec}), the fraction of reference amounts recovered (M_{amt}), and a binary degeneration check (M_{deg}: at least 20 tokens, no token above 30% of the text, Latin at most 5% of letters, CJK at most 0.5% of characters). Table 3 gives the trainer configuration.

Table 3. Configuration of the registered attempt, as recorded in its trainer arguments and adapter configuration.

Setting	Value
Optimizer steps	100
Learning rate	1×10^{-4} peak, linear decay to zero, no warm-up (logged 9.6×10^{-5} at step 5 to 1.0×10^{-6} at step 100)
Divergence penalty	none ($\beta = 0$)
Optimizer	AdamW, no weight decay, gradient clipping at 1.0
Rollouts	8 completions per prompt, 2 prompts per generation batch, 2 optimization passes per batch
Loss	DAPO (decoupled clip and dynamic sampling policy optimization) token-level; clipping 0.2 below, 0.4 above; group-mean advantage without standard-deviation scaling
Sampling	temperature 1.0, top-p 0.95, at most 2,048 tokens, truncated completions masked
Reward weights	[1.0, 0.0, 0.0, 0.0] over (D, M_{dec}, M_{amt}, M_{deg})
Adapter	LoRA, rank 64, $\alpha = 128$, no dropout, on q, k, v, o, gate, up, and down projections
Precision and seed	bfloat16; seed 42

Prompts were drawn from a success band of 0.25 to 0.75 under the parent policy, so that within-group reward variance would be non-zero by construction, which is the precondition for a non-vanishing policy gradient [17]. The run used a group size of 8, with two optimization iterations per generation batch, at sampling temperature 1.0, in bf16 precision, on a single A100 GPU. Its budget was fixed in advance at 100 steps: with two prompts per step and eight completions each, the registered attempt consumed 200 prompts and 1,600 rollouts, approximately 11% of the difficulty band (the trainer recorded 0.107 of an epoch at step 100) and about 1% of the length-filtered pool.

3.4 Intervention arms and external comparison

Five factors were varied one at a time: the numerical settings, the divergence anchor (the KL-penalty coefficient that ties the policy to the reference), the optimizer family, the learning rate, and the step count. Each arm branched from the same parent, was scored on the same evaluation instrument by the same frozen classifier and carried its adjudication rule before it ran. In addition to these training-lever arms, two further families change no parameters and are scored on the frozen 300-case instrument. The first quantifies extractable headroom by comparing eight-sample majority voting, in the self-consistency tradition [51], with greedy decoding, recording mean modal agreement alongside. The second composes voting with retrieval: three precedents per case retrieved by BM25, the lexical baseline of [52], over an index built from the training split only, in the retrieval-augmented tradition [53] and following the precedent-enhanced prompt template [54]; the same system is also decoded greedily so that the increment of the composition over voting alone can be tested. The delivered system is that composition, with no policy-gradient component, run in two independently seeded serving sessions so that sampled-serving variation could be reported as a range. The external comparison reads

it against Gemini 3.1 Pro (google/gemini-3.1-pro-preview) and GPT-5.5 (openai/gpt-5.5) on a 200-case set drawn at a fixed seed from a leakage-audited population disjoint from all training data, audited in the manner urged by [55], with identical prompts, candidate pools, and metrics at temperature 0 and pinned model identifiers. Because the frontier systems are not trained on our tag schema, a fallback that classifies the tail of the free text was added for the comparators and not for our own arms, an asymmetry that runs in the comparators' favor.

3.5 Statistical protocol

Three evaluation sets serve the paper, and each analysis names its set: the fixed 200-case held-out slice carries the nine paired arm evaluations of Section 4.1; the frozen 300-case instrument carries the parameter-free inference-composition results of Section 4.4 (majority voting, the precedent-augmented composition, and their increment); and a separate 200-case leakage-audited set carries the external comparison of Section 4.5.

All headline comparisons in this paper are paired. Metrics receive 95% intervals from 10,000 seeded bootstrap resamples [56]; McNemar's test [57] adjudicates paired per-case correctness with exact mid-p values; any interval containing zero is inconclusive whatever the point estimate. The Holm-Bonferroni correction [57] is applied within three declared contrast families and never across them: the training-lever family (the nine paired evaluations of Section 4.1 on the 200-case slice), the inference-composition family (voting against greedy decoding, voting with three precedents against greedy decoding, and the increment of the composition over voting, on the frozen 300-case instrument), and the external-comparison family (three pairwise system contrasts at n = 200). The resolving limit was stated before any result: at n = 300 and a discordance of 10 to 20%, the paired accuracy half-width is approximately ±0.036 to ±0.051. That proved too narrow for independently trained systems, where realized discordance was 37 to 38%, and on the 200-case external comparison the limit expands to approximately ±0.084. Decision rules were entered before the run each governed, in a dated change-control record kept with the project, following the practice described in [58]. Intervals are reported in one format throughout: the difference first, then the bracketed 95% interval, then the sample size and test, as in Δ macro-F1 = −0.067 (95% CI [−0.136, +0.011], n = 200, paired bootstrap); McNemar outcomes are reported as discordant counts with their mid-p.

4. Results

4.1 Overview of the falsification program

Table 2 shows, for every arm, the evidence that the weights moved, the share of held-out decisions that changed, and the adjudicated change in macro-F1, and Fig. 2 plots the four single-run arms against the parent checkpoint. As can be observed, no interval excludes zero.

Table 2. The falsification program: each arm, the evidence that its weights moved (ΔLoRA, the maximum absolute change in any adapter parameter), its behavioral change on held-out cases, and its adjudication; steps are completed/configured.

Arm	Steps	Evidence that weights moved	Held-out decisions changed	Δ macro-F1 (95% CI)	Adjudication
Corrected probe	85/100	max |ΔLoRA| 3.546×10^{-3}	0%	0.000 [0, 0]	Weights moved; greedy policy frozen

Arm	Steps	Evidence that weights moved	Held-out decisions changed	Δ macro-F1 (95% CI)	Adjudication
Registered final	100/100	max |ΔLoRA| 3.78×10^{-3}	54% of 200	−0.067 [−0.136, +0.011]	Movement without gain
Conservative standalone	100	max |ΔLoRA| 4.35×10^{-4}	Changed	+0.026 [−0.029, +0.081] on slice; frozen-instrument macro-F1 −0.015	Non-significant; frozen-instrument criterion fell
Six optimizer families	7 runs	Valid meters	Varied	No accepted gain in any	Locked null
Reduced learning rate 3×10^{-5}	100	|Δ| tracks the rate	88 of 200	−0.002 [−0.071, +0.081]	Damage removed; no gain beneath it

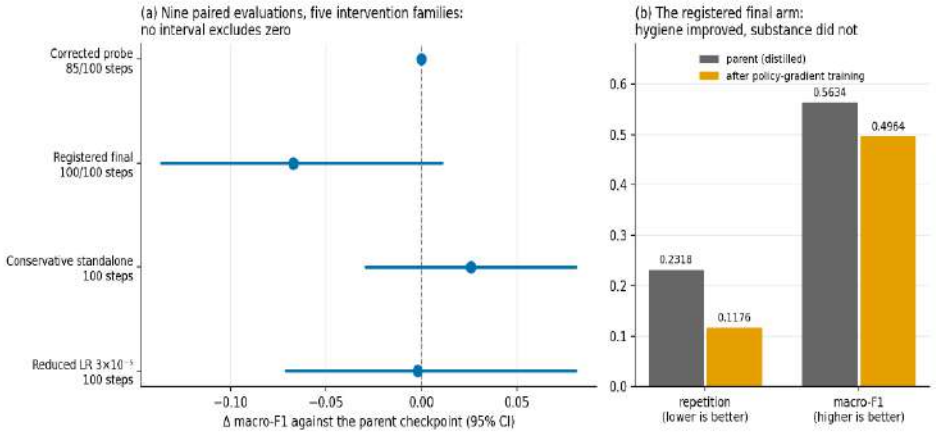


Fig. 2. The falsification program against the parent checkpoint. (a) Change in macro-F1 with 95% bootstrap intervals for the four single-run arms of Table 2, each scored on the fixed 200-case slice; no interval excludes zero. (b) The registered final arm: the repetition rate falls while macro-F1 also falls.

4.2 Registered attempt: policy change without improvement

The registered final attempt ran on rollouts certified clean by the instrument-repair campaign of Section 4.3 and completed its hundred steps at full update scale. Specifically, it moved: every one of the 200 greedy completions differed from the parent's, and 108 of them (54%) carried a different operative ruling. It did not improve: accuracy fell from 0.505 to 0.455 and macro-F1 from 0.5634 to 0.4964, a difference of −0.067 whose 95% bootstrap interval is [−0.136, +0.011] and spans zero; McNemar's test returned 36 against 26 discordant cases at mid-p = 0.21, and the gap was unchanged under alias-tolerant extraction (0.465 against 0.520). One thing improved, and it is form: the repetition rate halved, from 0.2318 to 0.1176, although repetition entered the objective only through the degenerate-group tie-break, which supplied the whole advantage in the 19% of groups whose dispositions scored alike; the substantive term, which the other groups optimized directly, deteriorated in the same run.

A second recipe, authored independently and branched from the same parent, moved the same slice the other way: accuracy rose by 0.035 and macro-F1 by 0.026, with an interval of [−0.029, +0.081] and McNemar p = 0.28; on the frozen 300-case instrument its raw accuracy rose while its macro-F1 fell by 0.015, the pattern expected when predictions shift toward the largest class. Six further optimizer families, run as seven single-variable arms (one family contributed two runs differing only in seed), produced no accepted gain; ranked by point estimate, the arms that concentrated or reshaped the gradient most aggressively sit at the worst end of the range, an ordering that is descriptive only and not evidence of a relationship, since every arm is single-seed and the spread across arms confounds

the intervention with training variance. The last arm changed one number, the learning rate, reduced to 3×10^{-5}, and the loss shrank with it to -0.002 with an interval spanning zero.

4.3 Alternative explanations excluded by the instrument

A null is admissible only under a design that excludes the alternatives, and here each exclusion is a measurement rather than an argument. *Defective machinery.* The earlier probe moved weights by 3.546×10^{-3} and changed 0% of decisions, the signature of a live optimizer behind a frozen generation path; it stopped at 85 of 100 steps because of an environment fault. Seven instrument defects were closed against recorded diagnostic output before the registered attempt ran, and one earlier verdict, which had read positive on a movement of +0.180 in the weight-zero monitor, was retracted when the corrected re-run reduced it to 0% of decisions changed and a macro-F1 difference of 0.000 [0, 0]. *Advantage collapse.* The zero-variance group fraction read 0.0 at all 20 logged readings, against a run-level collapse rate of 0.07, whereas the documented failure regime runs from 0.28 to 0.45. *Ineffective updates.* Full-precision snapshots put the maximum absolute adapter delta at 3.78×10^{-3}, and the cryptographic hash of the step-100 adapter was verified against the canonical adapter. *The weight-zero monitor.* The substance signal stayed flat in every window: 0.431 over the first half of the run and 0.332 over the second, 0.382 over all twenty readings, on the training-batch series that Section 3.2 characterizes and Fig. 4(b) draws; the driver itself, at 0.459 and 0.350 over the same halves, moved with it. A substance signal that does not move while the form terms improve is the form-over-substance signature stated as a measurement, available at every step for one deterministic function call per rollout. *Configuration.* Nine factors were varied across the arms (base model, difficulty band, reward gate, precision, sampling regime, monitor calibration, learning rate, divergence anchor, and step count), and the held-out measurement did not change sign in any consistent direction.

4.4 Extractable headroom and majority voting

Priced as described in Section 3.4, the sampling distribution already contained correct dispositions that greedy decoding was discarding (i.e., the extractable headroom was positive). Eight-sample majority voting, which updates no parameters, produced the program's only significant gain: on the frozen 300-case instrument, accuracy rose from 0.450 to 0.550 and macro-F1 by 0.088, with an uncorrected 95% paired-bootstrap interval of [0.043, 0.140]. Within the inference-composition family of three, the most stringent Holm step tests at $\alpha = 0.0167$; a percentile re-run of the paired bootstrap at that level (10,000 resamples, seed 3407) gives [0.032, 0.162], which excludes zero and reproduces the registered point estimate and lower bound exactly. Mean modal agreement was 0.719 to 0.730, so identical prompts yielded disagreement among samples in roughly 27 to 28 percent of cases. Composing voting with three retrieved precedents reached 0.5667 accuracy and 0.5558 macro-F1, a macro-F1 gain of 0.1728 over the same system decoded greedily (0.3830); self-consistency was unchanged at 0.717 to 0.734, so the retrieval channel added information rather than stability. Across two independently seeded serving sessions the delivered system reaches 0.600 accuracy and a macro-F1 of 0.563 to 0.573, the range recording sampled-serving variation. Extractable headroom was negative before targeted verified injection, positive after the first injection, and saturated within session noise after the second.

4.5 External comparison

Under the protocol of Section 3.4, the composed system reached 0.630 accuracy and 0.6389 macro-F1, against 0.660 accuracy and 0.5712 macro-F1 for Gemini 3.1 Pro and 0.640 accuracy and 0.5664 macro-F1 for GPT-5.5. No pairwise difference in accuracy was rejectable at this sample size after Holm correction. We state this as far as the test licenses it and no further: non-rejection at a resolving limit of approximately ±0.084 is not measured equivalence, and we do not upgrade it to parity anywhere. The macro-F1 differences, which favor the composed system, were not tested: the declared external family of Section 3.5 holds the three per-case accuracy contrasts, and no post-hoc macro-F1 contrast was constructed. After the free-text fallback described in Section 3.4, the residual unscoreable rates were 0.5% for the composed system against 8.0% and 1.0% for the comparators; this is reported as an operational contrast rather than as a quality claim.

4.6 Training traces of the registered attempt

Fig. 4 presents the traces of the registered attempt, drawn from the trainer log as recorded. The gradient norm had a median of 0.485 and rose from 0.28 to 1.00 as the learning rate decayed; no divergence term was logged, because none was configured. The logged aggregate, the unweighted sum of the four functions, moved from 2.49 to a maximum of 2.84 and ended at 1.96, and the driver's batch mean drifted from 0.59 to 0.33. The degeneration term read 1.0 with zero spread throughout, the within-batch standard deviations of the driver and of the substance signal stayed between 0.22 and 0.57, and the mean completion length fell from 982 to 841 tokens. Under the constants of Section 3.3 the gate could not have carried the advantage in this attempt; where the disposition term had no variance, the tie-break did.

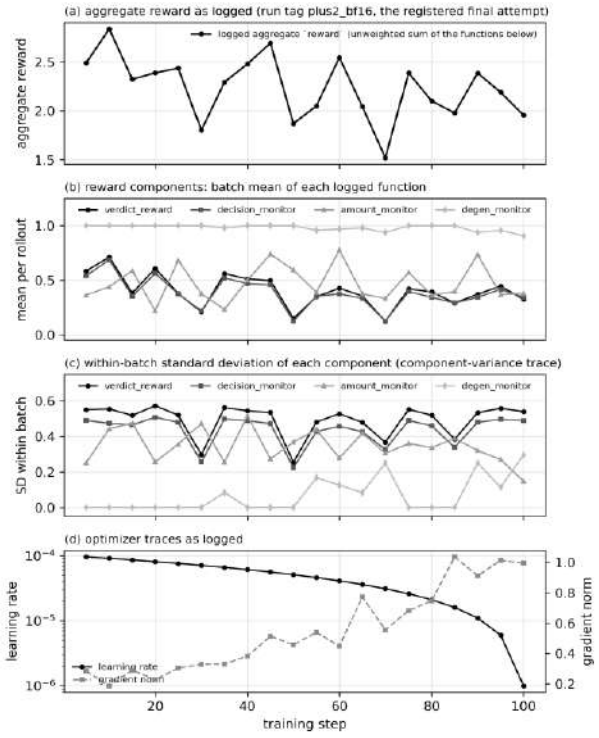


Fig. 4. Training traces of the registered final GRPO attempt on the Qwen3-8B student (100 steps, logged every five steps), drawn from the trainer log as recorded. (a) The aggregate reward as logged, the unweighted sum of the four reward functions. (b) Batch means of the four logged reward functions of Section 3.3: the degeneration term is saturated at 1.0 from the first reading, and the verdict reward and the substance signal fluctuate without trend. (c) Within-batch standard deviation of each function, the component-variance trace of Section 3.2. (d) Learning rate (left, logarithmic) and gradient norm (right).

5. Discussion

5.1 Principal findings and candidate causes

This paper reports three findings. The registered policy-gradient arm explored, and the objective could not rank that exploration into a gain on the substantive endpoint: the policy changed most of its held-out decisions and improved only on form, a repetition rate that halved and completions that shortened, the quantities rewarded by the degenerate-group tie-break. This is the form-over-substance pattern; an aggregate reward curve, had it risen, would have reported it as progress. The weight-zero monitor and its companion traces excluded defective machinery, advantage collapse, ineffective updates, and configuration as explanations while the run was in progress. And quantifying the sampling distribution instead of redesigning the reward showed that correct dispositions were already being sampled and discarded, so that majority voting, which trains nothing, produced the only significant gain, and composing it with retrieved precedents added information rather than stability, consistent with the documented sensitivity of generation to retrieved noise [59].

Several candidate causes are jointly consistent with the traces; the program was built to test two of them (Sections 5.2 and 5.3) and quantified a third, the absence of extractable capability, in advance. A fourth entered the literature after the program closed and has the standing of a proof: supervised fine-tuning and reinforcement learning do not decouple in either order, and reinforcement learning raises the supervised loss, with the predicted degradation confirmed on a Qwen3 model [25]. Our arms are exactly that configuration, and our held-out macro-F1 falls; three cautions attach, since the confirmation is at 0.6B parameters while this program runs at 8B, the theory bounds a direction rather than a magnitude, and the quantity it bounds is the supervised loss on the distillation traces, which this program did not measure. A fifth is related and equally untested here: a checkpoint that has received excessive supervised fine-tuning can lose the plasticity that policy-gradient training needs [26], and on-policy reinforcement learning prefers the KL-minimal solution among those that solve the task [27], which is consistent with a run that moved little from a rigid reference.

5.2 Effect of update scale and the divergence anchor

Two of the program's arms sit at opposite ends of the update-scale axis, and the contrast separates two nulls that a single run would confound. The conservative standalone recipe of Table 2 was configured by its independent author at a peak learning rate of 1×10^{-5} decayed linearly over its hundred steps and with a divergence penalty to the reference. Its log records a divergence with a median of 0.010, a gradient norm with a median of 0.25 that never trended, and an adapter whose largest entry changed by 4.4×10^{-4}, an order of magnitude less than the registered attempt's. The empirical analysis we rely on reports that the optimal learning rate for low-rank adaptation is about ten times the full-fine-tuning optimum, approximately invariant to the adapter rank [48]. It gives no absolute optimum for a configuration like ours, so the implied optimum of approximately 5×10^{-5} is our own inference from it; the conservative recipe's peak is five times lower than that value. Its anchor, moreover, ties the policy to a supervised checkpoint whose disposition accuracy is the 0.46 that the program never moved beyond: anchoring to a reference that lacks the target capability turns the regularizer into a constraint that holds the policy near the failure. The registered attempt is the other case: it ran at a peak of 1×10^{-4}, roughly twice the inferred low-rank optimum, with no divergence penalty, and it moved, the adapter's

largest entry changing by 3.78×10^{-3}, every greedy completion changing, and 54% of held-out dispositions flipping, while macro-F1 fell. Update scale therefore does not explain the registered null; it explains the conservative one. The arm that reduced the registered rate to 3×10^{-5} removed the loss without producing a gain beneath it, which identifies the mechanism of that loss as optimizer drift proportional to update magnitude rather than learning with the wrong sign; an independent controlled null on a supervised web-agent checkpoint reports the same gating by learning rate [16]. Two results bound what a policy-gradient arm can add once the margin is thin: reinforcing only correct responses improves pass@1 while degrading performance at higher k through lost diversity [60], and policy-gradient training with verifiable rewards has been characterized as reweighting existing solution traces rather than discovering new ones [61].

5.3 Advantage capture in gated composite rewards

Group-relative optimization standardizes each completion's summed reward within its group. Writing the standardized advantage as $A = (r - \bar{r})/\sigma_r$, where $r = \Sigma_c\, w_c r_c$ is the weighted sum of the components, each component's share of the direction of A scales with $w_c\sigma_c/\sigma_r$ when the components vary independently, so when one component's weighted standard deviation dominates σ_r the signal points almost entirely along it. This is the mechanism that GDPO [4] and MO-GRPO [20] repair by per-component normalization, that the covariance law of [22] states as the condition for an objective to improve, and that saturation-aware reweighting [62] addresses for components already solved. In the registered attempt the gate penalty was set at 0.05 (Section 3.3) and could not capture the advantage; form entered through the degenerate-group tie-break instead. The closed form that follows concerns the gated partial-credit driver, -0.5 behind a closed gate and $0.3 + 0.7q$ behind an open one, the common shape of composite verifiable rewards, and the one this program specified but never ran. It is a design result, and Fig. 3 reproduces it without any of our stack.

Write p for the probability that a completion is well formed and q for the probability that its disposition is correct. The law of total variance splits the driver's within-group variance into a gate term and a substance term, $\mathrm{Var}(D) = p(1-p)(0.8 + 0.7q)^2 + 0.49\, p\, q(1-q)$, and the two are equal only at $p^*(q) = 1 - 0.49\, q(1-q)/(0.8 + 0.7q)^2$, which is 0.903 at $q = 0.46$ (the disposition accuracy of the supervised reference, an illustrative value) and never falls below 0.898 for any q (Fig. 3b, minimum at $q \approx 0.35$). Below that threshold the gate carries more than half of the driver's variance and the standardized advantage is directed at format: the gap between well-formed and malformed completions is large, between 0.8 and 1.5 points, while substance, reachable only through an open gate, is worth at most 0.7. Fig. 3 reproduces the argument without project data: 400 sampled groups of sixteen Bernoulli draws match the closed form to within sampling error (maximum deviation 0.016 across the sweep), and nothing in the derivation refers to law, to Arabic, or to our corpus.

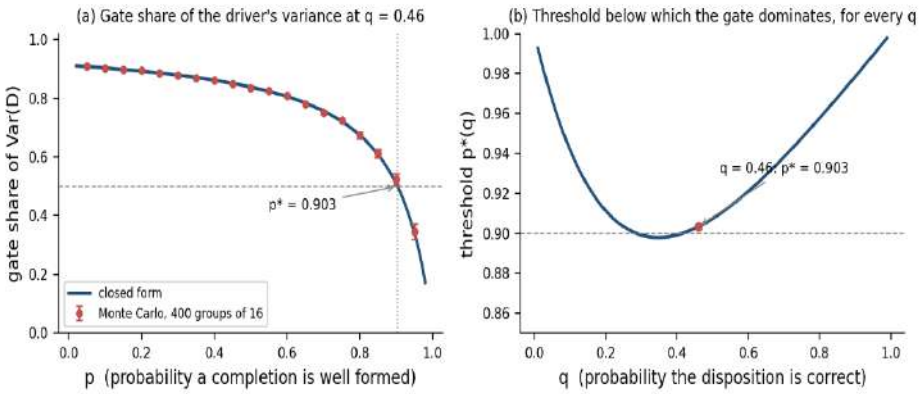


Fig. 3. The gate-variance argument reproduced without project data, using only the constants of the gated partial-credit driver stated in Section 5.3. (a) The gate's share of the driver's within-group variance as the well-formedness probability p is swept at $q = 0.46$, the disposition accuracy of the supervised reference: closed form (line) and Monte Carlo estimate (points, 95% bootstrap intervals) agree, and the share falls below one half only at $p^* = 0.903$, the equal-variance point. (b) The threshold $p^*(q)$: it lies between 0.898 and 1.0 for every q.

The gated design was our own engineering, built to prevent the pathology it invites, and it needs to be accounted for. Multiplying substance by a structure gate denies format reward without substance exposure, and under group normalization the gate then becomes the dominant variance source until nine rollouts in ten are well formed; the registered attempt avoided this by setting its gate penalty at 0.05 and still did not gain. One caveat qualifies the derivation: it treats components as separable, whereas in this design the gated driver is multiplicatively coupled to the structure gate by construction, so the covariance term is not zero and the decomposition is an approximation rather than an identity. The direction of the conclusion is unaffected, because coupling to the dominant-variance component strengthens rather than weakens its capture of the advantage.

5.4 Limitations and threats to validity

Causes that are jointly consistent with the traces constitute a consistency argument, not an ablation. One published result runs directly against the reading this section defends: deterministic rule-based verifiers in a guideline-defined domain are reported to outperform both the prior state of the art and verifiable outcome rewards [30], the favorable case our stack was supposed to be. We take the difference to be structural: that work verifies the process, scoring each intermediate step against the guideline that licenses it, whereas this stack verifies the outcome and gates it behind format, and Section 5.3 derives what gating costs. On that reading the counter-case supports the closed-form argument rather than contradicting the null; we have not run a process-verified arm on this corpus.

Training budget. Each arm saw 200 prompts and 1,600 rollouts, and the objection that the null is an artifact of too little data deserves a direct answer. First, the budget was sufficient to move the policy: every greedy completion changed, 54% of held-out dispositions changed, and the adapter's largest entry changed by 3.78×10^{-3} at a learning rate twice the inferred low-rank optimum, so the null is movement without gain rather than the absence of movement that under-training predicts. Second, Section 3.2 states the resolution at which the null is read: 100 steps resolve a shift of about 0.10 in the substance rate and not a smaller one, and the claim is bounded accordingly. Third, the volume of data is not the binding constraint when headroom exists; policy-gradient training with verifiable rewards has been shown to yield large gains from a single training example when the base policy already samples the correct answer [63], which is exactly the condition that Section 4.4 measured and found

exhausted for the registered parent. Whether a longer run on the full pool would behave differently is a question this program did not test, and Section 6 lists it. *Design and replication.* Every arm is single-seed, so breadth across nine paired evaluations substitutes for repetition within any one of them; small rollout groups leave exploration breadth open as an alternative account; and a full-fine-tuning control would test a different update-capacity hypothesis. A set of intervals that all contain zero remains a null however the variance is apportioned, whereas any ordering of the arms is not.

Model-family specificity. All arms use one base model, from a family on which gains from format-only or random rewards are documented and known not to transfer [64]. Part of that documented family effect has since been attributed to benchmark contamination: leakage-free generators show that only accurate reward signals produce steady gains [65]. A family-specific mechanism therefore cannot be excluded until the arms are repeated on a second base model.

Registration. Decision rules were not registered with a public service; they were entered, before the run each governed, in a dated change-control record in which every intervention carries a numbered change request, the single lever it varies, and the rule that adjudicates it.

Instrument. One rule-based classifier produced the reference labels and scores every arm, so any error in it moves both sides of every comparison. The version used was frozen before the registered attempt ran, and one documented bias, a character-ratio rule in the strict extractor that penalizes shorter outputs, was left in place rather than repaired after the fact.

Scope and reproducibility. Claims of generality apply to the monitoring protocol and the decision rule, not to the accuracy levels observed: one court, one jurisdiction, one base model, one reward family. The corpus cannot presently be redistributed pending a rights and anonymization review, which materially limits independent replication.

5.5 Recommended monitoring protocol and decision rule

Setting the specifics of our run aside, one instrument remains, and we recommend it to researchers training with verifiable rewards in general: weight-zero monitor rewards with per-component group-variance traces, logged from step zero. In our run, three such monitors cost no measurable compute and were the only instruments that distinguished a moving policy from a substance signal that barely moved. Before allocating compute to policy-gradient optimization, quantify the extractable headroom: measure whether the distribution the policy already samples contains correct answers that greedy decoding discards, a quantity that pass@k reads as a diagnostic [66]. Majority voting is itself a statistical certificate of the model's mode, and label-free post-training sharpens the distribution toward that same mode [67]; selection and training therefore compete for one margin rather than two. Selection rules admit sharper forms than plain voting: restricting candidates to high-frequency responses before ranking is minimax-optimal for pass@k inference scaling [68]. Register the verifiers that carry the objective a second time at weight zero, compute them on a probe set held fixed across steps, state the resolution at which their null is a null, and promote checkpoints only on held-out substance metrics. Where that headroom is exhausted, the effective intervention introduces new behavior into the sampling distribution, here targeted verified injection, in the manner of preparing a model for reinforcement learning by injecting behaviors before it [69].

6. Conclusion

This paper addressed the problem of reading a composite verifiable reward while it is being optimized. We have shown that such a reward can move a policy for a hundred steps toward shorter and less repetitive completions without affecting the quantity it was designed to improve, and that no aggregate readout reports this. Towards detecting it during the run, we have presented the least expensive instrument that does so, i.e., the endpoint's own verifier registered a second time at weight zero and logged with the within-group variance of each component. On an internally registered program of nine paired evaluations, this instrument produced a readable null (a registered arm that changed 54% of held-out dispositions while macro-F1 fell by 0.067 with an interval spanning zero) and directed the program to the one intervention that produced a gain, namely eight-sample majority voting, which trains nothing and raised macro-F1 by 0.088 (95% CI [0.032, 0.162] after correction). The claim we defend is narrow: where a substantive endpoint has an exact verifier, we recommend registering it twice, once inside the objective and once outside it at weight zero, so that the run can be interpreted; and the present results suggest quantifying the extractable headroom before policy-gradient optimization is undertaken. The findings are bounded by the limitations of Section 5.4: single-seed arms, one base model and reward family, 200 prompts per arm, and a corpus that cannot yet be redistributed. Given these limitations, future work will proceed in the following directions:

1. Multi-seed replication of the registered and extreme arms, besides larger rollout groups and longer runs over the full prompt pool.
2. Per-component normalization before the weighted sum [4], as a test of the mechanism derived in Section 5.3.
3. A plasticity diagnostic of the distilled parent checkpoint.
4. Repeating the arms on a second base model where the extractable headroom is positive rather than exhausted.

In conclusion, a rising reward curve is an indication rather than a measurement, and on the present results, registering the endpoint's own verifier at weight zero is the minimum instrumentation we would ask a composite-reward run to report.

Ethics Statement

This study involved no human subjects, human data, or personal data; the corpus derives from judicial rulings published as public records. The systems described are research artifacts and are not fit for autonomous adjudication.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Data Availability Statement

Three availability categories apply. (i) Not releasable: the implementation source code, whose export the authors' institution prohibits; as compensating disclosure, Section 3.3 states every reward function

and constant, and Section 5.3 derives the central mechanism in closed form so that it can be checked with algebra. (ii) Temporarily withheld pending review: the corpus, an independent collection assembled from rulings published as public records on the Saudi Ministry of Justice portal, and the trained adapters, both held until a rights and anonymization review concludes. (iii) Released with the article as supplementary material: the script that produces Fig. 3, which uses no project data; the script that produces Fig. 4; and the registered attempt's trainer log as recorded, a 39 kB text file that contains no case text.

Acknowledgement of AI Use

The authors used Claude (Anthropic) as an assistive tool for the following limited purposes: language refinement, structuring of certain sections, and preliminary organization of literature. All intellectual contributions, data interpretation, critical analysis, and final scientific judgments remain solely those of the authors. The AI was not used to generate original research ideas, fabricate data, or replace the authors' scholarly reasoning. The final manuscript was thoroughly reviewed and approved by the authors.

References

[1] Z. Shao et al., "DeepSeekMath: Pushing the Limits of Mathematical Reasoning in Open Language Models," arXiv:2402.03300, 2024.
[2] D. Guo et al., "DeepSeek-R1 Incentivizes Reasoning in LLMs Through Reinforcement Learning," Nature, vol. 645, pp. 633–638, 2025, doi: 10.1038/s41586-025-09422-z.
[3] J. Karwowski, O. Hayman, X. Bai, K. Kiendlhofer, C. Griffin, and J. Skalse, "Goodhart's Law in Reinforcement Learning," in Proc. Int. Conf. Learn. Represent. (ICLR), 2024. arXiv:2310.09144.
[4] S.-Y. Liu et al., "GDPO: Group Reward-Decoupled Normalization Policy Optimization for Multi-Reward RL Optimization," arXiv:2601.05242, 2026.
[5] J. Cui, X. Shen, F. Nie, Z. Wang, J. Wang, and Y. Chen, "A Survey on Legal Judgment Prediction: Datasets, Metrics, Models and Challenges," arXiv:2204.04859, 2022.
[6] J. Niklaus, V. Matoshi, P. Rani, A. Galassi, M. Stürmer, and I. Chalkidis, "LEXTREME: A Multi-Lingual and Multi-Task Benchmark for the Legal Domain," in Findings Assoc. Comput. Linguistics: EMNLP, 2023. arXiv:2301.13126.
[7] I. Chalkidis et al., "LexGLUE: A Benchmark Dataset for Legal Language Understanding in English," in Proc. 60th Annu. Meeting Assoc. Comput. Linguistics (ACL), Vol. 1: Long Papers, Dublin, Ireland, May 2022, pp. 4310–4330, doi: 10.18653/v1/2022.acl-long.297.
[8] A. Ammar, A. Koubaa, B. Benjdira, O. Nacar, and S. Sibaee, "Prediction of Arabic Legal Rulings Using Large Language Models," Electronics, vol. 13, no. 4, art. 764, 2024, doi: 10.3390/electronics13040764.
[9] M. B. Kmainasi, A. E. Shahroor, and A. Al-Ghraibah, "Can Large Language Models Predict the Outcome of Judicial Decisions?" arXiv:2501.09768, 2025.
[10] A. Almalki, S. Alsafari, and N. M. Alotaibi, "Legal Judgment Prediction in the Saudi Arabian Commercial Court," Future Internet, vol. 17, no. 10, art. 439, 2025, doi: 10.3390/fi17100439.
[11] H. Abu Shairah et al., "ALARB: An Arabic Legal Argument Reasoning Benchmark," in Proc. 3rd Arabic Natural Language Processing Conf. (ArabicNLP), Suzhou, China, Nov. 2025, pp. 389–406, doi: 10.18653/v1/2025.arabicnlp-main.32.
[12] M. Medvedeva, M. Wieling, and M. Vols, "Rethinking the field of automatic prediction of court decisions," Artif. Intell. Law, vol. 31, no. 1, pp. 195–212, 2023, doi: 10.1007/s10506-021-09306-3.

[13] M. Medvedeva and P. McBride, "Legal Judgment Prediction: If You Are Going to Do It, Do It Right," in Proc. Natural Legal Language Processing Workshop 2023, Singapore, Dec. 2023, pp. 73–84, doi: 10.18653/v1/2023.nllp-1.9.
[14] F. Kang, M. Kuchnik, K. Padthe, M. Vlastelica, R. Jia, C.-J. Wu, and N. Ardalani, "Quagmires in SFT-RL Post-Training: When High SFT Scores Mislead and What to Use Instead," arXiv:2510.01624, 2025.
[15] L. Zhang and J. Li, "Predicting Inference-Time Scaling Gains from Labeled Validation-Set Output Statistics," arXiv:2606.02981, 2026.
[16] C. Gan, Z. Cai, Y. Liang, H. Wei, S. Ni, and Q. Zhang, "A Learning-Rate-Gated Failure of GRPO in a Small Language and Vision-Language Model Web Agent: A Controlled Null and Its Mechanism," arXiv:2607.12640, 2026.
[17] N. Razin et al., "Vanishing Gradients in Reinforcement Finetuning of Language Models," in Proc. Int. Conf. Learn. Representations (ICLR), 2024. arXiv:2310.20703.
[18] W. Nie et al., "Gradient Starvation in Binary-Reward GRPO: Why Group-Mean Centering Fails and Why the Simplest Fix Works," arXiv:2605.07689, 2026.
[19] Q. Yu et al., "DAPO: An Open-Source LLM Reinforcement Learning System at Scale," in Adv. Neural Inf. Process. Syst. (NeurIPS), 2025. arXiv:2503.14476.
[20] Y. Ichihara et al., "MO-GRPO: Mitigating Reward Hacking of Group Relative Policy Optimization on Multi-Objective Problems," arXiv:2509.22047, 2025.
[21] M. M. Salmani-Zarchi, Z. Rahimi, H. Faili, and M. J. Dousti, "MDP-GRPO: Stabilized Group Relative Policy Optimization for Multi-Constraint Instruction Following," in Proc. 64th Annu. Meeting Assoc. Comput. Linguistics (ACL), Vol. 1: Long Papers, San Diego, CA, USA, 2026, pp. 42784–42797, doi: 10.18653/v1/2026.acl-long.1982. arXiv:2606.06058.
[22] Y. Lu and M. Jiang, "Uncovering Cross-Objective Interference in Multi-Objective Alignment," arXiv:2602.06869, 2026.
[23] R. Xin et al., "Surrogate Signals from Format and Length: Reinforcement Learning for Solving Mathematical Problems without Ground Truth Answers," arXiv:2505.19439, 2025.
[24] C. Ge, C. H. Yin, H. Liang, and J. Zhang, "Why GRPO Needs Normalization: A Local-Curvature Perspective on Adaptive Gradients," in 5th Workshop on Mathematical Reasoning and AI (MATH-AI), Neural Inf. Process. Syst. (NeurIPS), Dec. 2025. arXiv:2601.23135.
[25] X. Niu, B. Bai, W. Han, and W. Zhang, "On the Non-decoupling of Supervised Fine-tuning and Reinforcement Learning in Post-training," arXiv:2601.07389, 2026.
[26] R. Liu, J. Liu, X. Wan, Y. Fu, and L. Pan, "When RL Fails after SFT: Rejuvenating Model Plasticity for Robust SFT-to-RL Handoff," arXiv:2606.09932, 2026.
[27] I. Shenfeld, J. Pari, and P. Agrawal, "RL's Razor: Why Online Reinforcement Learning Forgets Less," arXiv:2509.04259, 2025.
[28] N. Rajani, A. P. Gema, S. Goldfarb-Tarrant, and I. Titov, "Scalpel vs. Hammer: GRPO Amplifies Existing Capabilities, SFT Replaces Them," in Actionable Interpretability Workshop, Int. Conf. Mach. Learn. (ICML), 2025. arXiv:2507.10616.
[29] S. R. Samineni, D. Kalwar, K. Valmeekam, K. Stechly, and S. Kambhampati, "RL in Name Only? Analyzing the Structural Assumptions in RL Post-Training for LLMs," arXiv:2505.13697, 2025.
[30] M. Pronesti, A. Belz, and Y. Hou, "Beyond Outcome Verification: Verifiable Process Reward Models for Structured Reasoning," in Findings of the Association for Computational Linguistics: ACL 2026, 2026. arXiv:2601.17223.
[31] L. Gao, J. Schulman, and J. Hilton, "Scaling Laws for Reward Model Overoptimization," in Proc. Int. Conf. Mach. Learn. (ICML), 2023.
[32] B. Baker et al., "Monitoring Reasoning Models for Misbehavior and the Risks of Promoting Obfuscation," arXiv:2503.11926, 2025.
[33] X. Wang et al., "Reward Hacking in the Era of Large Models: Mechanisms, Emergent Misalignment, Challenges," arXiv:2604.13602, 2026.
[34] I. F. Shihab, S. Akter, and A. Sharma, "Detecting Proxy Gaming in RL and LLM Alignment via Evaluator Stress Tests," arXiv:2507.05619, 2025.

[35] X. He et al., "Advantage Collapse in Group Relative Policy Optimization: Diagnosis and Mitigation," arXiv:2605.21125, 2026.
[36] M. F. Tarek et al., "Reward Hacking Mitigation using Verifiable Composite Rewards," in Proc. 16th ACM Int. Conf. Bioinformatics, Computational Biology, and Health Informatics (BCB), 2025.
[37] C. Ye et al., "Beyond Correctness: Harmonizing Process and Outcome Rewards through RL Training," in MATH-AI Workshop, Neural Inf. Process. Syst. (NeurIPS), 2025. arXiv:2509.03403.
[38] E. Skopin and E. Kotelnikov, "Improving Small Language Models for Code Generation with Reinforcement Learning from Verification Feedback," in Proc. Artificial Intelligence and Natural Language Conf. (AINL), 2026. arXiv:2605.30478.
[39] F. Ariai and G. Demartini, "Natural Language Processing for the Legal Domain: A Survey of Tasks, Datasets, Models, and Challenges," ACM Comput. Surv., 2024.
[40] F. Hijazi et al., "ArabLegalEval: A Multitask Benchmark for Assessing Arabic Legal Knowledge in Large Language Models," in Proc. 2nd Arabic Natural Language Processing Conf. (ArabicNLP), Bangkok, Thailand, Aug. 2024, doi: 10.18653/v1/2024.arabicnlp-1.20.
[41] M. Altherwi, "The Efficacy of Large Language Models (LLM) in Predicting the Outcomes of Commercial Legal Cases based on The Saudi Law: A Preliminary Study," in Proc. 6th Int. Conf. Computing, Networks and Internet of Things (CNIOT), 2025, pp. 1–5, doi: 10.1109/CNIOT65435.2025.11070426.
[42] N. Alangari, N. Alshenaifi, and H. Almuzaini, "Hierarchical Classification of Arabic Legal Cases Using Transformer Architectures and Large Language Models," Electronics, vol. 15, no. 14, art. 2988, 2026, doi: 10.3390/electronics15142988.
[43] V. Ovcharov, "UA-Legal-Bench: A Benchmark for Evaluating Large Language Models on Ukrainian Legal Reasoning," arXiv:2605.29170, 2026.
[44] A. Hochlehnert et al., "A Sober Look at Progress in Language Model Reasoning: Pitfalls and Paths to Reproducibility," arXiv:2504.07086, 2025.
[45] A. Yang et al., "Qwen3 Technical Report," arXiv:2505.09388, 2025.
[46] E. J. Hu, Y. Shen, P. Wallis, Z. Allen-Zhu, Y. Li, S. Wang, L. Wang, and W. Chen, "LoRA: Low-Rank Adaptation of Large Language Models," in Proc. Int. Conf. Learn. Representations (ICLR), 2022. arXiv:2106.09685.
[47] W. Kwon et al., "Efficient Memory Management for Large Language Model Serving with PagedAttention," in Proc. 29th ACM Symp. Operating Systems Principles (SOSP), 2023, pp. 611–626, doi: 10.1145/3600006.3613165.
[48] J. Schulman and Thinking Machines Lab, "LoRA Without Regret," Thinking Machines Lab: Connectionism, Sep. 2025, doi: 10.64434/tml.20250929. [Online]. Available: https://doi.org/10.64434/tml.20250929. Accessed: Aug. 4, 2026.
[49] Hugging Face, "TRL Documentation: Logging (GRPOTrainer logged metrics)," 2026. [Online]. Available: https://huggingface.co/docs/trl/main/logging. Accessed: Sep. 6, 2026.
[50] N. Ho, L. Schmid, and S.-Y. Yun, "Large Language Models Are Reasoning Teachers," in Proc. 61st Annu. Meeting Assoc. Comput. Linguistics (ACL), Vol. 1: Long Papers, Toronto, Canada, 2023, pp. 14852–14882.
[51] X. Wang et al., "Self-Consistency Improves Chain of Thought Reasoning in Language Models," in Proc. 11th Int. Conf. Learn. Representations (ICLR), 2023.
[52] N. Thakur, N. Reimers, A. Rücklé, A. Srivastava, and I. Gurevych, "BEIR: A Heterogeneous Benchmark for Zero-shot Evaluation of Information Retrieval Models," in Proc. Neural Inf. Process. Syst. Datasets and Benchmarks Track, 2021. arXiv:2104.08663.
[53] Y. Gao et al., "Retrieval-Augmented Generation for Large Language Models: A Survey," arXiv:2312.10997, 2023.
[54] Y. Wu et al., "Precedent-Enhanced Legal Judgment Prediction with LLM and Domain-Model Collaboration," in Proc. 2023 Conf. Empirical Methods in Natural Language Processing (EMNLP), Singapore, 2023, pp. 12060–12075.

[55] S. Balloccu, P. Schmidtova, M. Lango, and O. Dusek, "Leak, Cheat, Repeat: Data Contamination and Evaluation Malpractices in Closed-Source LLMs," in Proc. 18th Conf. European Chapter of the Association for Computational Linguistics (EACL), Vol. 1: Long Papers, St. Julian's, Malta, Mar. 2024, pp. 67–93, doi: 10.18653/v1/2024.eacl-long.5.
[56] D. Ulmer, C. Hardmeier, and J. Frellsen, "deep-significance: Easy and Meaningful Statistical Significance Testing in the Age of Neural Networks," in Proc. ML Evaluation Standards Workshop, Int. Conf. Learn. Represent. (ICLR), 2022. arXiv:2204.06815.
[57] S. Riezler and M. Hagmann, Validity, Reliability, and Significance: Empirical Methods for NLP and Data Science, 2nd ed. Cham, Switzerland: Springer, 2024.
[58] E. van Miltenburg, C. van der Lee, and E. Krahmer, "Preregistering NLP Research," in Proc. 2021 Conf. North Amer. Chapter Assoc. Comput. Linguistics: Human Lang. Technol. (NAACL-HLT), 2021, pp. 613–623.
[59] F. Cuconasu et al., "The Power of Noise: Redefining Retrieval for RAG Systems," in Proc. 47th Int. ACM SIGIR Conf. Res. Develop. Inf. Retrieval, Washington, DC, USA, 2024, pp. 719–729.
[60] X. Zhu et al., "The Surprising Effectiveness of Negative Reinforcement in LLM Reasoning," arXiv:2506.01347, 2025.
[61] B. Bai, X. Wang, P. Ye, and T. Chen, "Learning to Explore with Parameter-Space Noise: A Deep Dive into Parameter-Space Noise for Reinforcement Learning with Verifiable Rewards," in ICML Workshop on AI for Math, Jul. 2026. arXiv:2602.02555.
[62] Y. Wang et al., "Learn What's Left, Not What's Mastered: Saturation Aware Advantage Reweighting for Multi-Reward Policy Optimization," arXiv:2608.16072, 2026.
[63] Y. Wang et al., "Reinforcement Learning for Reasoning in Large Language Models with One Training Example," arXiv:2504.20571, 2025.
[64] R. Shao et al., "Spurious Rewards: Rethinking Training Signals in RLVR," arXiv:2506.10947, 2025.
[65] M. Wu et al., "Reasoning or Memorization? Unreliable Results of Reinforcement Learning Due to Data Contamination," in Proc. AAAI Conf. Artificial Intelligence, 2026. arXiv:2507.10532.
[66] Y. Yu, "Pass@k Metric for RLVR: A Diagnostic Tool of Exploration, But Not an Objective," arXiv:2511.16231, 2025.
[67] P. Cordero-Encinar and A. B. Duncan, "Certified Self-Consistency: Statistical Guarantees and Test-Time Training for Reliable Reasoning in LLMs," arXiv:2510.17472, 2025.
[68] Q. Di, K. Ji, X. Li, H. Zhao, and Q. Gu, "Best-of-Majority: Minimax-Optimal Strategy for Pass@k Inference Scaling," in Proc. Int. Conf. Learn. Represent. (ICLR), 2026. arXiv:2510.03199.
[69] Z. Cen, Y.-F. Yao, W. J. Han, Z. Liu, and D. Zhao, "Behavior Injection: Preparing Language Models for Reinforcement Learning," arXiv:2505.18917, 2025.

Islamic University Journal of Applied Sciences (IUJAS)

https://journals.iu.edu.sa/jesc

Volume VIII, Issue II, December 2026, Pages 367-396

AraBit-1.58: Morphology-Aware Embeddings Reduce the Cost of Ternary Quantization in Arabic Text Simplification

Abdullah Alshanqiti [1, *]

[1] Faculty of Computer and Information Systems, Islamic University of Madinah, Madinah, KSA

*Corresponding author: Abdullah Alshanqiti; *Email*: amma@iu.edu.sa ; a.m.alshanqiti@gmail.com

Abstract

Extreme quantization restricts Transformer weights to the ternary set {−1,0,+1}, but whether this transfers to morphologically rich languages is untested. Arabic is demanding, because its clitic-heavy, root-and-pattern morphology yields a vast inventory of surface forms that subword tokenizers segment at linguistically meaningless boundaries. AraBit-1.58 couples a Morphology-Aware Tokenization and Embedding layer (MATE) with a BitNet b1.58 ternary backbone. MATE decomposes each word into proclitics, root, morphosemantic pattern, and enclitics, embeds these separately, and gates to a subword fallback when analysis fails. We evaluate on the SAMER Arabic simplification corpus under a controlled two-by-two design over three seeds that varies only the front end and the weight precision. Ternary quantization compresses weight memory 2.85×, from 104.72 to 36.70 MB, but without a fused ternary kernel it cuts neither latency nor energy. The comparison confirms our pre-specified hypothesis: quantization costs 9.43 SARI with subword input against 7.49 SARI with morphological input, an interaction of +1.93 SARI significant across seeds (paired t(2) =4.55, one-tailed p=0.023). Morphological grounding therefore buys robustness to extreme quantization rather than higher absolute quality, measured on SARI at small scale.

Keywords: Arabic natural language processing; Extreme quantization; 1-bit language models; BitNet; Morphological segmentation; Text simplification; Edge inference.

https://doi.org/10.63070/jesc.2026.047

Received 12 August 2026; Revised 10 September 2026; Accepted 19 September 2026; Available online 10 October 2026.

Published by Islamic University of Madinah on behalf of *Islamic University Journal of Applied Sciences.*

This is an open access article published under the Creative Commons Attribution 4.0 International license (CC BY 4.0): https://creativecommons.org/licenses/by/4.0/

1. Introduction

The computational appetite of modern Transformer architectures [1] has become a practical barrier to their deployment outside well-provisioned data centres. A model that must hold billions of half-precision parameters in memory cannot serve a mobile phone, a classroom tablet, or an offline kiosk. This bottleneck bears particularly heavily on Arabic Natural Language Processing, where the community must contend simultaneously with limited compute budgets and with a language whose morphology is far richer than that of the Indo-European languages around which most efficiency research has been organised.

Arabic is built on a root-and-pattern morphological system. A consonantal root, typically of three radicals, is interleaved with a morphosemantic pattern to produce a stem, and that stem is then wrapped in proclitics and enclitics that encode conjunction, definiteness, preposition, and pronominal reference. A single orthographic word may therefore carry the content of an entire English clause. The consequence for statistical modelling is a surface-form inventory that grows combinatorially, a high type-to-token ratio, and severe data sparsity for any model that treats surface forms as atomic.

Towards making Arabic models viable at the edge, researchers have pursued several efficiency paradigms. Token-mixing attention has been replaced by Fourier transforms [2]. Distilled encoders such as DistilBERT [3] have been used in place of full BERT [4] for downstream Arabic tasks, and sparse Mixture-of-Experts routing [5] has been used to grow capacity without a proportional growth in compute. These lines of work reduce arithmetic or activation cost. None of them reduces the width of the weights themselves. They all continue to store parameters in FP16 or BF16, and the memory footprint therefore remains substantial.

A different frontier was opened by the BitNet line of work. BitNet [6] demonstrated that Transformers can be trained natively with binary weights, and BitNet b1.58 [7] extended the weight alphabet to the ternary set $\{-1,0,+1\}$, an information content of $\log_2 3 \approx 1.58$ bits per weight. Because the alphabet contains only zero and unit magnitudes, the dominant matrix multiplication collapses into signed accumulation. The recent BitNet b1.58 2B4T release [8] showed that this holds at the two-billion-parameter scale, and that such a model can match full-precision peers of comparable size while occupying a fraction of the memory. A recent survey [9] situates these results within the wider binary-network literature.

Extreme quantization is not free. Restricting weights to three values imposes a hard ceiling on how much structure a single linear layer can encode, and the ceiling binds most tightly where the input representation is least informative. This is precisely the situation created by byte-pair encoding (BPE) [10] on Arabic. BPE is trained to minimise corpus description length, not to respect morpheme boundaries, and it routinely splits an Arabic word in ways that separate a root radical from its pattern

or fuse a clitic to a stem. A ternary network fed such a segmentation must first reconstruct morphological identity before it can begin to model meaning, and it must do so with weights that cannot express fine-grained magnitude.

Our position in this paper is that the two problems should not be solved in the same place. Rather than asking the quantized layers to recover morphology implicitly, we inject morphological structure explicitly at the point where the representation is formed. To clarify more, the embedding table in a BitNet architecture is retained at higher precision than the transformer blocks, and it is therefore the natural location for information that ternary weights are poorly suited to encode. Our proposal, AraBit-1.58, exploits this asymmetry deliberately.

In a nutshell, this paper makes the following contributions:

1. We introduce **AraBit-1.58**, the first architecture, to the extent of our knowledge, that combines native ternary quantization with explicit morphological decomposition for a morphologically rich language.
2. We propose **MATE**, a Morphology-Aware Tokenization and Embedding layer that represents each Arabic token by its proclitics, root, pattern, and enclitics rather than by an arbitrary subword split, and that gates gracefully to a subword fallback when morphological analysis fails.
3. We give an explicit parameter and memory accounting that separates ternary from full-precision components, so that the reported compression factor can be audited rather than assumed.
4. We report a two-by-two ablation that holds the backbone, the training data, and the optimisation budget fixed while varying only the input representation and the weight precision, and we pre-specified the interaction it tests before running it.
5. We report latency and energy under two accountings, one for the model in isolation and one end-to-end including morphological preprocessing, and we discuss the gap between them frankly. In our setting both regress under ternary weights, which we report rather than omit.
6. We report a component-level ablation that localises the effect within MATE, showing that clitic separation and the learned gate carry it while the root and pattern embeddings do not, which argues for a leaner front end than the one we built.

What our measurements establish is a precise and, we believe, more useful claim than the one we set out to test. Morphological grounding does not raise absolute generation quality at this scale. It buys something else: robustness. The quality that extreme quantization destroys is measurably smaller when the input carries explicit morphological structure, which is the interaction our hypothesis predicted and which the learned gate’s behaviour independently corroborates. That is a result about where

morphological information earns its place in a low-precision model, and it points toward the regime in which such front ends are worth their cost.

The remainder of the paper is organised as follows. Section 2 reviews efficient Arabic modelling, extreme quantization, and morphology-aware representation, and identifies the gap our work addresses. Section 3 details the proposed architecture. Section 4 describes the experimental setup. Sections 5 to 7 present the results, the ablation study, and an error analysis. Section 8 discusses threats to validity, and Section 9 concludes.

2. Related Work

2.1. Efficiency Paradigms for Arabic Language Models

Work on efficient Arabic modelling has largely followed three routes. The first replaces or simplifies the attention mechanism. FNet [2] showed that an unparameterised Fourier transform can substitute for self-attention with modest quality loss and substantial speedup, and this idea has been carried into Arabic rephrasing and simplification [11]. The second route is knowledge distillation, in which a small student network absorbs the behaviour of a larger teacher [12]. DistilBERT [3] is the canonical instance, and it has been applied successfully to Arabic extractive summarization in a dual-stage arrangement that couples sentence selection with clause-level reduction [13]. The third route is conditional computation, where a Mixture-of-Experts layer activates only a subset of parameters per token [5], an approach adapted to Arabic simplification in [14].

These approaches differ in mechanism but share a limitation. Each reduces the amount of arithmetic performed per token, or the number of parameters engaged per token, while leaving the numerical width of each stored parameter untouched. A distilled or sparsely routed model still stores FP16 or BF16 weights, so its resident memory scales at two bytes per parameter. For a device with a few hundred megabytes of available memory, this is the binding constraint.

2.2. Extreme Quantization and Native 1-bit Models

Quantization research has traditionally been divided between post-training quantization, which compresses an already-trained full-precision model, and quantization-aware training, which anticipates the eventual precision during optimisation. GPTQ [15] and AWQ [16] are strong representatives of the former at four bits, and LLM.int8() [17] handles the eight-bit case with outlier-aware decomposition. Below four bits, post-training methods degrade sharply, because the quantization grid becomes too coarse to preserve the weight distribution of a model that was never trained to tolerate it.

The BitNet line takes the opposite stance and quantizes from the first optimisation step. BitNet [6] introduced the BitLinear layer, which binarises weights and quantizes activations to eight bits while retaining a full-precision master copy for gradient accumulation. BitNet b1.58 [7] generalised the

weight alphabet to $\{-1,0,+1\}$, which adds an explicit zero and therefore permits feature filtering, and reported parity with FP16 LLaMA baselines from roughly the three-billion-parameter scale upward. The 2B4T technical report [8] provided the first open native 1-bit model at scale together with a CPU inference backend. A survey of the broader field is given in [9].

Two properties of this literature bear directly on our design. First, ternary quantization is applied to the linear projections inside the transformer blocks, while the token embedding table, the output head, and the normalisation parameters are retained at higher precision. The compression factor of a BitNet model is therefore never the naive ratio of 16 to 1.58, and any honest efficiency claim must state which components were counted. Second, published parity results are reported at scale. At smaller parameter counts, ternary models are known to trail their full-precision counterparts, which is an important caveat at the scale we work at.

2.3. Morphology-Aware Representation for Arabic

The Arabic NLP community has long held that surface forms are the wrong unit of analysis. MADAMIRA [18] and, more recently, CAMeL Tools [19] provide full morphological analysis and disambiguation, returning for each token a lemma, a root, a pattern, a part-of-speech tag, and a segmentation into clitics and stem. Farasa [20] offers a fast segmenter oriented towards throughput rather than depth, and transformer-based segmentation has also been applied at the clause level to unpunctuated Arabic text [21]. Habash [22] gives the standard treatment of the underlying linguistic system.

Morphological preprocessing has repeatedly been shown to help Arabic models, and AraBERT [23] used Farasa segmentation during pre-training for this reason. What the literature has not examined is whether the benefit of morphological grounding is *constant* across weight precisions. Our hypothesis is that it is not, and that the benefit grows as precision falls, because a low-precision network has less spare capacity with which to reconstruct structure that the tokenizer discarded. This hypothesis is testable, and Section 6 is designed to test it.

2.4. Limitations of the Reviewed Approaches and the Gap Addressed

Glancing over the reviewed literature, three gaps stand out. Efficient Arabic models reduce compute but not weight width. Native 1-bit models reduce weight width but have been developed and validated almost entirely on English, and the released checkpoints acknowledge limited support for non-English languages [8]. Morphology-aware representation has been studied under full precision only. Table 1 positions the reviewed work along these axes and locates our contribution.

Table 1. Positioning of the reviewed approaches. FP denotes full precision, PTQ denotes post-training quantization, and QAT-native denotes quantization applied from the beginning of training.

Approach	Weight precision	Compute reduction	Morphology-aware	Arabic
FNet / FFT-Seq2Seq [2, 11]	FP16	Yes (attention)	No	Yes
DistilBERT summarizer [3, 13]	FP16	Yes (depth)	No	Yes
Switch Transformer [5, 14]	FP16	Yes (sparsity)	No	Yes
AraBERT [23]	FP16	No	Yes (Farasa)	Yes
GPTQ / AWQ [15, 16]	4-bit (PTQ)	Yes (memory)	No	No
BitNet b1.58 [7, 8]	1.58-bit (native)	Yes (arithmetic)	No	No
AraBit-1.58 (this work)	**1.58-bit (native)**	**Yes**	**Yes (CAMeL)**	**Yes**

3. AraBit-1.58: The Proposed Architecture

3.1. Problem Formulation and Notation

Let $X = (x_1, \dots, x_n)$ be an input sequence of n Arabic orthographic tokens, and let $Y = (y_1, \dots, y_m)$ be the target sequence, which in our task is a simplified rendering of X. We seek a parameterised model $p_\theta(Y \mid X)$ in which the overwhelming majority of θ is drawn from the ternary alphabet $\mathcal{T} = \{-1, 0, +1\}$. The constraint is that generation quality remains close to that of a full-precision model of comparable size.

We partition the parameters into two sets. The ternary set θ_t comprises the weights of every linear projection inside the transformer blocks. The full-precision set θ_f comprises the embedding tables, the projection matrix of the MATE layer, the output head, and all normalisation and bias terms. This partition follows the BitNet convention [7, 8], and we make it explicit here because the memory analysis in Section 3.5 depends on it. Figure 1 gives the overall architecture.

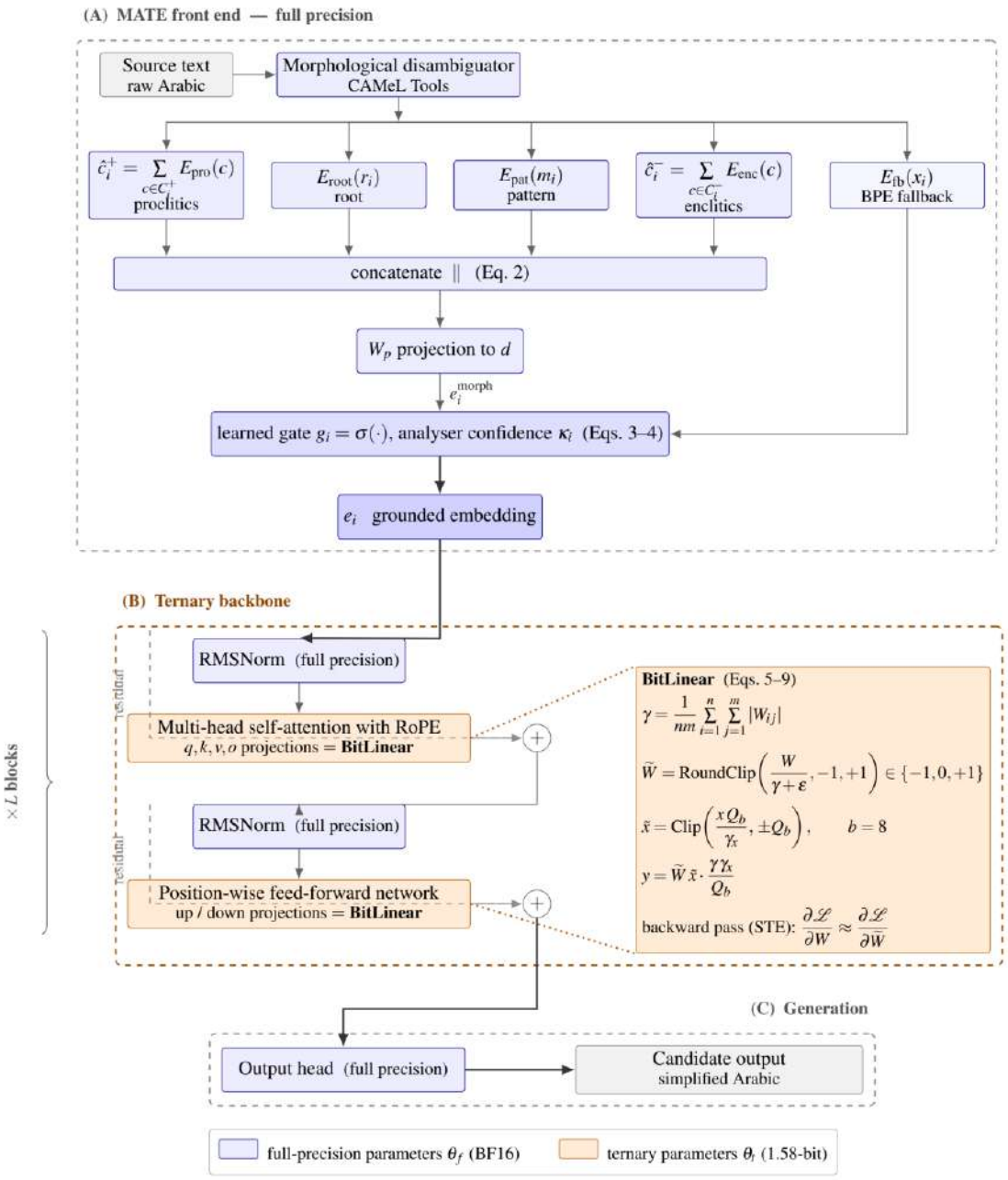


Fig.1. Overall architecture of AraBit-1.58.

Overall architecture of AraBit-1.58. Stage (A) is the MATE front end: a morphological disambiguator returns the proclitic set, root, pattern and enclitic set of each token. The two clitic sets are sum-pooled, the four component vectors are concatenated and projected by W_p, and a learned gate mixes the result with a BPE fallback according to the analyser confidence κ_i. Stage (B) is a stack of L transformer blocks in which every linear projection is a BitLinear layer with ternary weights and eight-bit activations, while normalisation and residual paths are retained at full precision, as is the output head of Stage (C).

3.2 Morphology-Aware Tokenization and Embedding (MATE)

Unlike English, where word formation is largely concatenative, an Arabic word is produced by interleaving a consonantal root with a morphosemantic pattern and then attaching clitics on both sides. Consider the form *wa-sa-yaktubūna-hā*. It decomposes into the proclitics *wa* (and) and *sa* (future marker), the root *k-t-b*, the imperfective pattern that yields *yaktubūna*, and the enclitic *hā* (feminine object pronoun). A BPE tokenizer has no access to any of this and will produce a segmentation determined solely by corpus frequency.

MATE resolves each token by running a morphological disambiguator, for which we use CAMeL Tools [19], over the raw input in context. For token x_i the analyser returns a proclitic sequence C_i^+, a root r_i, a pattern m_i, and an enclitic sequence C_i^-. Because the number of clitics on either side varies, we sum-pool over each clitic set rather than assuming a fixed slot, which gives

$$\hat{c}_i^+ = \sum_{c \in C_i^+} E_{\text{pro}}(c), \qquad \hat{c}_i^- = \sum_{c \in C_i^-} E_{\text{enc}}(c)$$

The four component vectors are concatenated and projected back to the model dimension d by a trainable matrix W_p, giving the morphologically grounded representation

$$e_i^{\text{morph}} = W_p\left[\, \hat{c}_i^+ \;\|\; E_{\text{root}}(r_i) \;\|\; E_{\text{pat}}(m_i) \;\|\; \hat{c}_i^- \,\right] + b_p$$

where $\|$ denotes concatenation. Not every token admits a clean analysis. Foreign words, digits, proper nouns absent from the lexicon, and orthographic noise will either fail to analyse or receive a low-confidence analysis. Discarding these tokens is not acceptable in a generation setting, so MATE retains a compact BPE fallback embedding E_{fb} and mixes the two representations through a learned gate:

$$\begin{aligned} g_i &= \sigma\left(w_g^\top\left[\, e_i^{\text{morph}} \;\|\; E_{\text{fb}}(x_i) \;\|\; \kappa_i \,\right] + b_g\right) \\ e_i &= g_i \odot e_i^{\text{morph}} + (1 - g_i) \odot E_{\text{fb}}(x_i) \end{aligned}$$

where κ_i is the analyser confidence for token x_i and σ is the logistic function. The gate is learned jointly with the rest of the model, so the network can itself decide how far to trust the analyser on a given token. This also yields a diagnostic, since the mean gate value over a corpus reports how much of the representation is in fact being carried by morphology, and we report this quantity in Section 7.

Two consequences of this design deserve emphasis, and the first of them did not hold in the form we originally expected.

The root and clitic inventories are indeed compact. Measured on our pre-training corpus, they contain 3,898 roots, 18 proclitics, and 42 enclitics, so 3,958 rows in total against the 13,699 entries of the BPE control vocabulary. The pattern inventory is not compact. We had anticipated a few hundred abstract templates, on the reasoning that Modern Standard Arabic derives its stems from a closed set of morphosemantic patterns. CAMeL Tools' pattern feature is not that object. It is fully diacritized and

fully inflected for each word form, with the root radicals replaced by numbered slots, so it varies almost as freely as the surface forms themselves. Our corpus yields 55,316 distinct pattern values, four times the size of the BPE vocabulary rather than a fraction of it. We record this because the ambiguity is easy to inherit from a toolkit's feature name, and because any parameter budget built on the abstract-template reading would be wrong by two orders of magnitude.

The consequence for the parameter accounting is that MATE is not more compact than a BPE front end at matched model dimension. The pattern table stays affordable only because its embedding width is narrow, 48 dimensions against the model dimension of 512, giving $55{,}316 \times 48 \approx 2.66\text{M}$ parameters where a full-width table of that height would require 28.3M. Summing the four morphological tables, the BPE fallback table, and the projection W_p gives approximately 7.01M full-precision embedding parameters, which is by construction the figure we matched the control vocabulary against, so the two front ends enter the ablation with equal input-side budgets. Once the learned gate is included, MATE is in fact 0.53M parameters *larger* than the control, as Table 5 records. The honest summary is that MATE redistributes embedding capacity across linguistically meaningful inventories rather than reducing it, and any efficiency argument for the front end has to rest on that redistribution rather than on a smaller parameter count.

Second, morphological decomposition shortens the effective input, because clitics that BPE would render as separate subword tokens are folded into the same position, which reduces sequence length and hence attention cost.

3.3 The Ternary Backbone

Once morphologically grounded embeddings are formed, they pass through a stack of L transformer blocks in which every linear projection is replaced by a BitLinear layer, following [7] and [8]. During the forward pass, a full-precision master weight matrix $W \in \mathbb{R}^{n \times m}$ is quantized to the ternary alphabet by scaling with the mean absolute value and rounding with saturation:

$$\gamma = \frac{1}{nm} \sum_{i=1}^{n} \sum_{j=1}^{m} |W_{ij}|$$

$$\widetilde{W} = \text{RoundClip}\left(\frac{W}{\gamma + \epsilon},\ -1,\ +1\right),$$

$$\text{RoundClip}(x, a, b) = \max\left(a,\ \min\left(b,\ \text{round}(x)\right)\right)$$

where ϵ is a small constant guarding against division by zero. Activations are quantized to $b = 8$ bits per token using absolute-maximum scaling,

$$\gamma_x = \| x \|_\infty, \qquad Q_b = 2^{b-1},$$
$$\tilde{x} = \mathrm{Clip}\left(\frac{x\, Q_b}{\gamma_x},\ -Q_b + \epsilon,\ Q_b - \epsilon\right)$$

and the layer output is rescaled to recover the original dynamic range:

$$y = \widetilde{W}\tilde{x} \cdot \frac{\gamma\, \gamma_x}{Q_b}$$

The rounding operation has zero gradient almost everywhere, so we employ the Straight-Through Estimator [24], which treats the quantizer as an identity map on the backward pass:

$$\frac{\partial \mathcal{L}}{\partial W} \approx \frac{\partial \mathcal{L}}{\partial \widetilde{W}}$$

Gradients are accumulated into the full-precision master copy W, which is retained during training and discarded after it. Only the packed ternary tensor $\widetilde{W}$ is required at inference. Following the 2B4T configuration [8] we use rotary position embeddings [25] and sub-layer normalisation of the RMSNorm family [26], both of which are retained at full precision.

3.4. Training Procedure

Because post-training quantization to 1.58 bits is known to fail [7], the backbone cannot be obtained by compressing an existing Arabic checkpoint. It must be trained natively. Our procedure has two phases.

Phase 1, ternary pre-training. We pre-train the backbone from random initialisation on a general Modern Standard Arabic corpus (Arabic Wikipedia) using a decoder-only architecture with a prefix-LM masking objective: attention is bidirectional over a prefix segment of each sequence and causal over the continuation that must be generated. Every configuration in our comparison, including the BPE ablation, receives an identical token budget and an identical optimiser schedule, so that the ablation varies only the input representation. We use AdamW [27] with a linear warm-up followed by cosine decay, and we note that BitNet training benefits from a larger peak learning rate than full-precision training of the same size, an observation reported in [7] and confirmed in our preliminary runs.

Phase 2, task fine-tuning. The pre-trained backbone is fine-tuned on the parallel complex-simple pairs of the simplification task. The MATE tables and gate are trained jointly throughout both phases rather than frozen, since a gate learned only at fine-tuning time would have no opportunity to calibrate against the analyser errors encountered during pre-training.

3.5. Parameter and Memory Accounting

An efficiency claim is only meaningful if its accounting is stated. Let P_t be the number of ternary parameters and P_f the number of full-precision parameters. With ternary weights packed at $\log_2 3 \approx 1.58$ bits and full-precision parameters stored in BF16 at two bytes, the resident weight memory in bytes is

$$M = P_t \cdot \frac{\log_2 3}{8} + 2P_f$$

The naive compression ratio of 16 to 1.58, that is $10.1\times$, is therefore an upper bound that no real BitNet model attains, because P_f is never zero. The realised ratio depends on the fraction of parameters held in the embedding tables, and that fraction is larger for small models than for large ones. This is a further reason to be cautious about a small scale, and we report the itemised accounting in Table 5 rather than a single headline number.

4. Experimental Setup

4.1. Datasets

This study evaluates a single downstream task, Arabic text simplification. Three corpora are involved, each in a distinct role, and Table 2 lists them together.

The task itself uses the SAMER Corpus [28], the first manually annotated Arabic parallel corpus for text simplification. It comprises 159K words selected from fifteen publicly available Arabic fiction novels, it carries readability annotations at document and word level, and it provides two simplified parallel versions of each text targeting two learner levels. We train and evaluate on the level-5 source to level-3 target direction throughout. SAMER is the appropriate benchmark here because its simplifications are human-authored rather than machine-translated, and because its levelled structure permits reporting quality separately per target level.

The SAMER readability lexicon [29] enters only as an evaluation resource. It supplies the per-lemma readability levels used by the reference-free measure of Section 4.3, and no part of it is used for training.

Pre-training consumes a general Modern Standard Arabic corpus drawn from Arabic Wikipedia, under the procedure of Section 3.4. It carries no simplification supervision, and it is shared identically across all four cells of the ablation.

SAMER is the only evaluation corpus used in this study. Machine-translated Arabic simplification sets such as EW-SEW and WikiLarge [30] exist, but they carry translation artefacts that a natively authored corpus does not, and we prefer a single clean benchmark to a broader but noisier one.

Table 2. Datasets used in this study. Counts are the pairs and sentences actually consumed by our pipeline, with empty-field rows dropped at load time, measured directly rather than read off raw file line counts. Corpora that were planned but not used are omitted from this table and described in the surrounding text instead.

Dataset	Task	Pairs (tr/dev/te)	Source	Ref.
SAMER Corpus	Simplification	14,343 / 2,983 / 3,277	Human-authored fiction	[28]
SAMER readability lexicon	Evaluation	40,686 levelled lemmas	Human-annotated lexicon	[29]
MSA pre-training corpus	Pre-training	15,643,471 BPE tokens, 9,804,440 words, 402,557 sentences, 6,000 articles	Arabic Wikipedia, dump 20231101	—

4.2. Baselines

Our baseline set is designed as a set of controls rather than as a list of competitors. Each entry isolates one factor.

- **B1. AraBART [31], FP16**, and AraT5 [32] where usable (see Section 8). Established full-precision Arabic sequence-to-sequence models. These provide an upper bound on quality, but they are pre-trained on far larger corpora than our proof-of-concept budget permits, so the comparison is informative about deployment rather than about architecture.
- **B2. Same architecture, FP16, with MATE.** An identically sized model with the identical MATE front end and full-precision weights. The gap between B2 and AraBit-1.58 measures the pure cost of ternary quantization.
- **B3. Same architecture, ternary, with BPE.** An identically sized ternary model whose front end is a standard BPE tokenizer of matched vocabulary budget. The gap between B3 and AraBit-1.58 measures the contribution of MATE. This is the central control of the study.
- **B4. Same architecture, FP16, with BPE.** The fourth cell of the two-by-two design. Comparing the B2-versus-B4 gap against the AraBit-versus-B3 gap tests our central hypothesis, namely that morphological grounding matters more at low precision than at full precision.
- **B5. Post-training quantized AraBART.** We planned four-bit GPTQ [15] or AWQ [16], the standard choices, but neither has a working Windows build in our environment and building from source needs a C compiler we do not have – an infrastructure limitation, not a methodological one (Section 8). We instead report four-bit NF4 [33] via bitsandbytes, a genuinely different quantization scheme from GPTQ/AWQ, not a compatible substitute – labeled distinctly throughout (Table 4/5: "AraBART + NF4", never reported as GPTQ). GPTQ

and AWQ remain the realistic efficiency competitors a practitioner would consider, and their absence is a real gap in this study rather than one we can close by substitution.

- **B6. FFT-Seq2Seq [11] and Switch-Arabic [14].** Our own prior models were planned as continuity baselines. They were not re-run under this protocol, and their published numbers were obtained on a different corpus with a different SARI implementation, so we do not tabulate them.

We deliberately exclude commercial large language models from the comparison. They are not reproducible, their parameter counts and training data are undisclosed, and a small ternary model is not in the same deployment class. Where a qualitative reference point is useful, we report it in the discussion rather than in the results tables.

4.3. Evaluation Metrics Quality

We report SARI [34] as the primary simplification metric, together with BERTScore-F1 [35] computed with the Arabic encoder aubmindlab/bert-base-arabertv02 [23] for semantic preservation, and a reference-free readability measure so that simplification is assessed on its own terms rather than only on overlap. For readability we use the levelled lexicon of Al Khalil et al. [29]. Each output word is lemmatised with CAMeL Tools and looked up against the lexicon's per-lemma readability level. We report two quantities for the source, for each system's output, and for the gold reference: the mean level, and the proportion of lemmas at the simplest level, level 1. Both appear in Table 4. Together they let us state directly whether a system's output moved toward or away from the reference's simplicity, rather than only how much it overlaps with it. OSMAN [36] was considered and dropped, because no verified reference implementation was available against which to check a from-scratch one, and the levelled-lexicon measure needs no separate implementation to validate. We do not report BLEU [37], which is unsuitable as a primary simplification metric, having been shown to correlate poorly and sometimes negatively with human simplicity judgements [38]. Every SARI figure in this paper is computed against a single reference per example. Every system in Table 4, own-model and baseline alike, decodes greedily with an identical 64-token generation budget. An earlier internal inconsistency, in which our own cells were capped lower than the baselines, was found and corrected. We verified that the correction changes no reported number, because generation already terminates at an end-of-sequence token well before any of the old, lower caps on nearly every example.

Efficiency. We report resident weight memory in megabytes under the accounting of [eq:mem], peak inference memory in gigabytes, and latency per generated token in milliseconds. Latency is reported twice, once for the model in isolation and once end-to-end including morphological analysis, since the disambiguator is a CPU-bound stage whose cost does not vanish. Energy is reported in joules per

generated sequence, measured through hardware counters, and we state that our figures are measured rather than analytically estimated from arithmetic operation counts.

4.4. Implementation Details

AraBit-1.58 is implemented in PyTorch, with morphological analysis provided by CAMeL Tools [19]. Model size, layer count, hidden dimension, head count, and vocabulary sizes are listed in Table 3, together with the optimiser, the learning-rate schedule, the batch sizes, and the token and epoch budgets of both training phases, so that the runs can be reproduced. Pre-training the twelve backbone runs, four cells at three seeds each, took 166.6 minutes of wall-clock time in total on the hardware named in that table. The full implementation, configuration files, and the scripts that produce every table and figure in this paper are available at https://github.com/AMahfodh/AraBit.

Every result reported in this paper is measured at this scale. Vocabulary sizes are counted directly from the fitted tokenizers and morphological vocabularies over the 6,000-article pre-training corpus of Table *2* rather than estimated. The optimisation block gives the numerical training settings needed to reproduce the runs. Every cell of the ablation receives the same schedule, the same token budget, and the same decoding configuration, so that the comparison varies only the input representation and the weight precision.

Table 3. Model configuration

Component	Setting
Layers L	12
Hidden dimension d	512
Attention heads	8
Root vocabulary	3,898
Pattern vocabulary	55,316, used as-is rather than truncated. This is far larger than our original design estimate, for the reason set out in Section 3.2
Proclitic / enclitic vocabulary	18 / 42
BPE fallback vocabulary (E_fb, MATE's own OOV-piece tokenizer)	8,000
Shared output vocabulary (BPE, identical across all four cells, see Section 8)	13,699
Activation precision	8-bit, absmax per token
Weight precision (transformer blocks)	1.58-bit ternary
Weight precision (embeddings, head, norms)	BF16
Optimisation, identical across all four cells	

Component	Setting
Optimiser	AdamW [27], weight decay 0.1
Learning-rate schedule	Linear warm-up over the first 2% of total steps, then cosine decay to zero
Sequence length / batch size, pre-training	256 tokens / 8 sequences
Pre-training tokens	10M per seed, per cell (3 seeds × 4 cells)
Pre-training wall-clock time	166.6 minutes for all twelve runs
Fine-tuning	3 epochs on the SAMER training split.
Decoding	Greedy, 64-token generation budget, identical for every system in Table 4
Hardware	NVIDIA RTX 3070 (8 GB VRAM), with morphological pre-analysis on a 12-core CPU

5. Results and Discussion

5.1. Generation Quality

Table 4 reports generation quality. Three comparisons matter. The distance between AraBit-1.58 and B2 quantifies what ternary quantization costs when the input representation is held fixed. The distance between AraBit-1.58 and B3 quantifies what morphological grounding buys at ternary precision. The distance between AraBit-1.58 and the FP16 published baselines indicates whether a ternary model at this scale is yet in the useful range for deployment, and we expect this last gap to be the largest, since those baselines carry far greater pre-training budgets.

SARI is computed against a single reference per example. Higher is better for SARI and for BERTScore-F1. The two readability columns are the levelled-lexicon measure of Section *4.3*, reported as the mean SAMER lemma level of the output and as the proportion of its lemmas at level 1. For both of them the target is the gold reference rather than an extremum, and the reference scores 1.71 and 0.583 against the source's 1.95 and 0.541. Readability must be read jointly with SARI and BERTScore rather than on its own, because an output that draws on a narrow, high-frequency vocabulary scores as simple whether or not it simplifies well. The ternary rows illustrate exactly that, and Section *8* sets out the point in full. Own-model rows give mean ± std over 3 seeds on the full 3,277-example test split. AraBART and AraBART+NF4 are single measurements on the same split. AraT5 is excluded because the released checkpoint produced degenerate output that a task prefix did not fix (Section *8*), and AraBART+GPTQ because the toolchain was unavailable on our hardware, both of which we report in text rather than as a broken or empty number. Our earlier Fourier-transform and Switch Transformer models were not re-run under this protocol, so their published numbers, computed on a different corpus with a different SARI implementation, are not comparable and are not tabulated.

Table 4. Generation quality on the SAMER Arabic simplification task.

Model	Precision	SARI	BERTScore-F1	Readability: mean level	Readability: level-1 prop.
Reference points					
Source text	—	—	—	1.95	0.541
Gold reference	—	—	—	1.71	0.583
External baselines					
AraBART [31]	FP16	39.03	0.6412	1.76	0.574
AraBART + NF4 [33]	4-bit	37.65	0.6384	1.76	0.575
Controlled cells (identical backbone, data and budget)					
B4: ours, BPE	FP16	24.01 ± 0.04	0.6220 ± 0.0012	2.24 ± 0.02	0.418 ± 0.011
B2: ours, MATE	FP16	20.93 ± 0.75	0.4989 ± 0.0170	2.08 ± 0.07	0.484 ± 0.019
B3: ours, BPE	1.58-bit	14.58 ± 0.03	0.4541 ± 0.0050	1.98 ± 0.04	0.394 ± 0.010
AraBit-1.58 (MATE)	**1.58-bit**	**13.44 ± 0.06**	**0.3365 ± 0.0052**	**1.82 ± 0.09**	**0.475 ± 0.029**

The AraBit-versus-B3 gap (13.44 vs. 14.58, −1.14 SARI) is small but real, and in the unfavourable direction: adding MATE at ternary precision costs slightly more SARI than it gains here. The B2-

versus-B4 gap is larger and in the same unfavourable direction (20.93 vs. 24.01, -3.08): MATE costs SARI at full precision too, more so than at ternary precision. Neither controlled comparison shows MATE improving absolute simplification quality at this scale – BPE beats MATE at both precisions. What Table 6 and Section 6 show is a narrowing of MATE's shortfall specifically at ternary precision, which is the interaction our hypothesis predicts. That is a different claim from MATE winning outright, and Section 6 reports the test that distinguishes the two. Every own-model cell trails every FP16 external baseline, as expected given the pre-training budget asymmetry (Section 8).

The readability columns require care, and they are the one place in Table 4 where the ranking inverts. Both ternary cells score as simpler than their full-precision counterparts on mean lemma level, 1.82 against 2.08 with morphological input and 1.98 against 2.24 with BPE input, while scoring far worse on SARI and on BERTScore-F1. This is not evidence that ternary quantization improves simplification. The likelier explanation is that short, repetitive ternary output draws on a narrower and more frequent vocabulary, which is precisely what a levelled lexicon rewards. A system can score as simple by degenerating as readily as by simplifying well, and Section 8 develops the point.

The level-1 proportion is informative under the same guard. MATE reaches a higher proportion than BPE at both precisions, 0.484 against 0.418 at full precision and 0.475 against 0.394 at ternary precision, and the full-precision MATE cell comes closest of our four cells to the gold reference value of 0.583. We record the pattern without resting a claim on it, because the degenerate-vocabulary mechanism just described could produce the same ordering. The divergence between SARI and BERTScore-F1 visible across these rows is taken up in Section 6, where the interaction test makes it precise.

5.2. Efficiency

Table 5. Efficiency accounting. Memory follows *[eq:mem]* with ternary and full-precision components itemised. Latency is reported both for the model alone and end-to-end including morphological analysis. The P_f and memory figures for our own cells reflect the corrected output head described in Section *8*. The two MATE cells were re-measured after that correction, and the two BPE cells are unaffected by it. The memory figure for AraBART and AraBART+NF4 is real measured resident weight memory rather than *[eq:mem]*-derived, because NF4 is a different quantization scheme from our ternary BitLinear. The two baselines have no morphological preprocessing stage, so an end-to-end latency distinct from the model-only figure does not apply to them, and we did not instrument them for energy. AraBART+GPTQ was infeasible on our hardware (Section *8*) and is described there rather than tabulated.

Model	P_t (M)	P_f (M)	Mem. (MB)	Lat. model (ms/tok)	Lat. e2e (ms/tok)	Energy (J/seq)
AraBART	—	139.222	556.89	7.5899	n/a	n/a
AraBART + NF4	—	139.222	222.03	12.2630	n/a	n/a
B4: FP16 + BPE	0.000	51.832	103.67	0.8501	0.9049	1.7575
B2: FP16 + MATE	0.000	52.358	104.72	0.8377	1.2143	2.0818
B3: ternary + BPE	37.749	14.084	35.65	1.8374	1.8401	3.0049
AraBit-1.58	**37.749**	**14.609**	**36.70**	**2.0192**	**2.3775**	**3.2503**

The two latency columns are expected to differ substantially. Morphological disambiguation is a CPU-bound stage that runs once per input sentence, and on short inputs its cost can be comparable to the entire ternary forward pass. We report the model-only figure because it is the quantity that isolates the architectural contribution, and we report the end-to-end figure because it is the quantity a deployment engineer must plan around. A paper that reported only the first would be misleading. We also note that the analyser cost amortises over long inputs and can be pipelined, though we did not measure the latency-versus-length curve and therefore make no quantitative claim about the amortisation rate.

The efficiency result is genuinely mixed and we state the unfavourable half plainly. Ternary weights deliver the memory reduction they promise: $103.67 \rightarrow 35.65$ MB with BPE input ($2.91\times$) and $104.72 \rightarrow 36.70$ MB with MATE ($2.85\times$). They deliver neither a latency nor an energy reduction here. End-to-end latency roughly doubles, from 0.9049 to 1.8401 ms/token for the BPE pair and from

1.2143 to 2.3775 ms/token for the MATE pair, and energy rises from 1.7575 to 3.0049 J/sequence and from 2.0818 to 3.2503 J/sequence respectively. The reason is that our BitLinear layers simulate ternary arithmetic in full-precision tensors rather than dispatching to a packed ternary kernel, so the quantization overhead is paid without the corresponding speedup being realised. The memory figures reflect the architecture and would survive a kernel change. The latency and energy figures reflect our implementation and would not. Readers should treat the compression result as the transferable one and treat our timing numbers as an upper bound that a fused kernel would be expected to improve on substantially.

6. Ablation Study

The ablation is the scientific core of this paper, since it is the only part of the design that can distinguish a genuine contribution from a backbone effect. We hold the architecture, the pre-training corpus, the token budget, and the optimiser schedule fixed, and vary only the input representation and the weight precision. This yields the two-by-two design of Table 6.

Mean $\pm$ std over 3 seeds on the full 3,277-example test split. Δ denotes the SARI difference attributable to MATE at each precision, given here as a plain mean-versus-mean difference for readability. Our hypothesis predicts $\Delta_{ternary} > \Delta_{FP16}$, and the paired interaction test reported in the discussion that follows is the one that assesses it.

Table 6. The central two-by-two ablation.

	BPE front end	**MATE front end**	**Δ (MATE gain)**
FP16 weights	24.01 ± 0.04 (B4)	20.93 ± 0.75 (B2)	−3.08
Ternary weights	14.58 ± 0.03 (B3)	13.44 ± 0.06 (AraBit-1.58)	−1.14

Both Δs are negative – MATE costs SARI at both precisions, not a gain. But H1 is not a claim about either Δ's sign. It is a claim about the sign of $\Delta_{\text{ternary}} - \Delta_{\text{FP16}}$, i.e. whether MATE's shortfall narrows at ternary precision. It does: $\Delta_{\text{ternary}} - \Delta_{\text{FP16}} = -1.14 - (-3.08) = +1.94$ by the table's rounded means. Testing this properly requires the paired structure across seeds, because the two Δs are correlated (both share the same BPE-vs-MATE architectural comparison, just at different precisions). Treating each seed's interaction $I(s) = \Delta_{\text{ternary}}(s) - \Delta_{\text{FP16}}(s)$ as one paired observation ($n = 3$) gives mean $I = +1.93$, $t(2) = 4.55$, one-tailed $p = 0.023$ against $H_0: I \leq 0$, with a minimum detectable effect at $\alpha = 0.05$ of 1.24 SARI points – comfortably below the observed effect. A two-level bootstrap (10,000 resamples, over both seeds and test examples) gives a 95% CI of [1.15,2.55], entirely above zero. That bootstrap should be read as characterising example-level variability rather than as

independent confirmation of the seed-level result. All three per-seed interactions ($+1.10$, $+2.20$, $+2.50$) are positive, so any resampling scheme drawing seeds with replacement is arithmetically obliged to return a positive mean, and the interval's lower bound is close to the smallest observed seed. The parametric test and the bootstrap rest on the same three numbers.

H1 is supported on SARI, with two qualifications we state rather than leave to a reader to derive. The sample is small ($n = 3$ seeds, $df = 2$), and our own pre-specified analysis plan set a seed-expansion threshold that the realised minimum detectable effect of 1.83 SARI (two-tailed) crossed. We did not expand, on the grounds that the effect was already significant, and we record that decision here because it was taken after the result was known. The second qualification is that the same interaction computed on BERTScore is $(-0.1176) - (-0.1231) = +0.0055$, roughly half of one standard error and indistinguishable from zero. The effect we report is therefore specific to SARI at this scale, and a reader should weight it accordingly.

Why the two metrics diverge is worth setting out, because the divergence is not arbitrary. Three mechanisms plausibly contribute, and our measurements do not separate them.

The first is what each metric counts. SARI scores edit operations, namely the n-grams added, kept, and deleted relative to the source and the reference. MATE acts directly on which surface forms the model is able to produce, so it operates on the very quantity SARI counts. BERTScore-F1 is a similarity between contextual embeddings of the whole candidate and the whole reference, and it is dominated by global semantic adequacy rather than by local lexical choice.

The second is the range in which our cells sit. At SARI values of 13 to 24, against 39.03 for AraBART, the outputs are far from adequate, and BERTScore is compressed into a narrow band near the low end of its useful range. An interaction worth one to two SARI points maps into a displacement of a few thousandths of cosine similarity inside that band, which the seed-level variation of BERTScore in Table 4, between ± 0.0012 and ± 0.0170, cannot resolve.

The third is the front end of the metric itself. BERTScore-F1 is computed with aubmindlab/bert-base-arabertv02, which consumes raw text and re-tokenizes every candidate and every reference with its own subword vocabulary before embedding them. That encoder is the variant of AraBERT trained without morphological pre-segmentation, so the distinctions MATE introduces at the input are partly re-absorbed into subword pieces by the metric itself. BERTScore is therefore structurally less sensitive than SARI to the manipulation under test.

We do not adjudicate between these explanations, and the defensible reading is the narrow one. The interaction is evidenced on SARI and is not evidenced on BERTScore. Separating the three would require a morphology-aware semantic metric alongside a larger seed budget, which we note as a

direction in Section 9. The finding is an interaction, not an outright win – MATE does not make AraBit-1.58 better than the BPE control in absolute terms (Table 4), it makes ternary quantization cost MATE *less* than it costs plain BPE.

Table 7 then removes MATE components individually, which identifies which part of the morphological signal carries the benefit.

Single seed, ternary precision, and a 200-example evaluation subset throughout, rather than the full split used in Tables *4* and *6*, so these SARI values are not directly comparable with those tables. The full-MATE and plain-BPE rows use the same seed and subset, which keeps the within-table comparison like-for-like. FP params is the full-precision remainder P_f in the sense of Table *5*, read directly from each trained checkpoint.

Table 7. Component-level ablation of the MATE layer.

Configuration	SARI	Δ vs. full	FP params (M)
Full MATE (proclitic + root + pattern + enclitic + gate)	10.81	—	14.609
— pattern embedding	11.08	+0.26	11.929
— root embedding (stem embedding instead)	11.14	+0.33	14.397
— clitic separation (surface form only)	10.56	−0.25	14.583
— learned gate (hard fallback)	10.59	−0.22	14.084
— all (plain BPE)	11.61	+0.79	14.084

The finding is the opposite of the paper's working guess: removing the pattern or root embeddings does not hurt, it slightly *helps* ($+0.26$, $+0.33$ SARI), while removing clitic separation or the learned gate hurts (-0.25, -0.22). At this scale, clitic segmentation and the gate mechanism appear to carry the morphological signal that matters. The root and pattern embeddings add parameters without a corresponding quality benefit, and the pattern table is much the more expensive of the two, accounting for 2.68M full-precision parameters against the root table's 0.21M. A simpler MATE that keeps clitic separation and the gate but drops both root and pattern embeddings would therefore be cheaper and marginally better on this metric. We report this plainly rather than retrofit a narrative: it argues for simplifying the architecture in future work, dropping root/pattern embeddings and keeping clitic separation and the gate. Every MATE configuration, full or ablated, still trails plain BPE in absolute SARI at this single-seed, 200-example scale (11.61), consistent with Table 4's finding that MATE does not win outright – component ablation changes which parts of MATE help each other, not whether MATE beats the BPE control overall.

7. Error Analysis

Quantitative scores alone do not explain where a model fails, and for a morphology-driven architecture the failure modes are informative in their own right. This manuscript reports the analysis that can be

automated, and two further analyses requiring hand annotation are left for a follow-up study (Section 9).

Gate behaviour. The learned gate g_i of [eq:gate] reports how much of each token's representation is carried by morphology rather than by the BPE fallback, which converts a design assumption into a measured quantity. Figure 2 reports it by token category on a 200-example sample. Two of the four categories we had planned to distinguish, digits and Latin-script foreign tokens, do not occur in this corpus of literary Modern Standard Arabic, so we report the two that do rather than present empty cells.

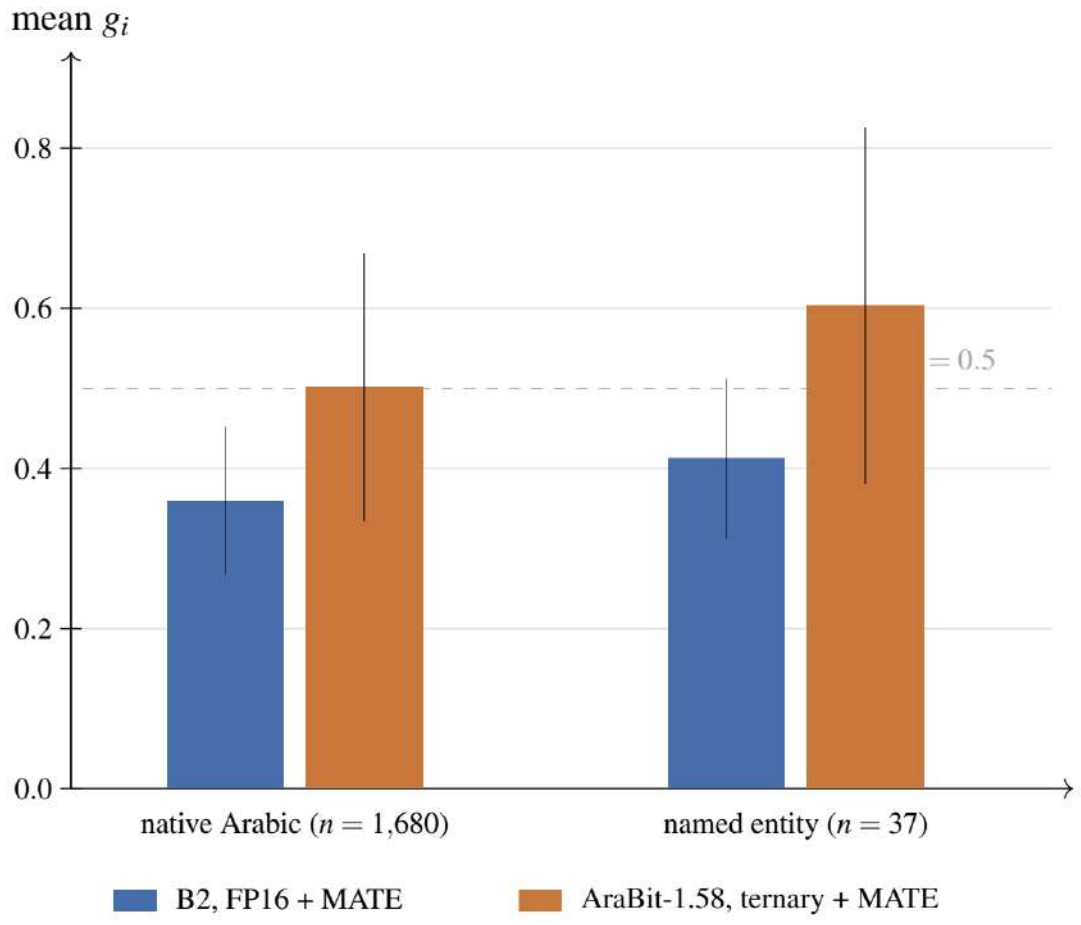


Fig.2. *Token category*

Mean learned gate value by token category, at full precision and at ternary precision. Whiskers show one standard deviation. Higher values indicate greater reliance on the morphological branch relative to the BPE fallback. Digits and Latin-script foreign tokens are absent from this corpus and are therefore not shown.

The measured values are 0.3598 ± 0.0922 (native Arabic) and 0.4120 ± 0.1006 (named entities) at full precision, rising to 0.5014 ± 0.1675 and 0.6030 ± 0.2226 at ternary precision. Two observations follow, one supported and one that we flag as suggestive only.

The supported observation is that the ternary model leans measurably harder on the morphological branch than its full-precision counterpart does. At full precision the gate sits well toward the BPE fallback, and under ternary weights it moves to roughly even. This is a mechanistic reading consistent with the interaction reported in Section 6: when the weights available for reconstructing structure are coarse, the model draws more on structure supplied at the input. We note that the two measurements are consistent, not that either causes the other.

The suggestive observation is that named entities draw slightly higher gate values than native Arabic words at both precisions. With only 37 named entity tokens and standard deviations of 0.10 and 0.22, this difference is not one we would defend, and we report it to document the measurement rather than to interpret it. We also note the variance itself: gate values under ternary weights are far more dispersed than at full precision (0.17–0.22 against 0.09–0.10), so the gate is not converging on a confident policy at this training budget.

8. Threats to Validity and Limitations Scale

This is a proof-of-concept study at a small parameter count. Published parity between ternary and full-precision models has been demonstrated at billion-parameter scale [7, 8], and ternary models are known to trail at smaller sizes. Our results therefore cannot be extrapolated upward, and a favourable result at our scale would be a stronger claim than an unfavourable one. We state this in advance rather than as a retrospective excuse.

Pre-training budget asymmetry. AraBART and AraT5 were pre-trained on corpora orders of magnitude larger than our budget permits. Comparisons against them speak to deployment readiness, not to architectural merit. The architectural claims of this paper rest exclusively on the controlled comparisons B2, B3, and B4, in which every factor other than the one under study is held fixed.

NF4 in place of GPTQ or AWQ. The post-training quantization baseline was planned as GPTQ or AWQ, and neither toolchain was available in our environment. We report four-bit NF4 [33] instead. NF4 is a data-free, distribution-aware weight format, whereas GPTQ and AWQ are calibration-based, so this is not a like-for-like substitute. Every NF4 result is labelled as such, and the calibration-based comparison remains open.

AraT5 excluded from Table 4. The released UBC-NLP/AraT5-base checkpoint produced degenerate output on our data, consisting of repeated emoji and language-name tokens, independently of fine-

tuning, beam search, and decoding hyperparameters. We tested the most plausible benign explanation, that AraT5's multi-task pre-training expects a task-instruction prefix our harness omitted, and found no such convention in its released training script and no change in behaviour across five candidate prefixes. We therefore report this as an observed limitation of that checkpoint on our data rather than tabulate a broken number.

SARI implementation. SARI is implemented directly from Xu et al. [34], because easse has no build for the Python version this pipeline requires. We cross-checked against a working easse installation on 50 examples of our own output, tokenized identically. The two correlate strongly (Pearson $r = 0.74$, Spearman $\rho = 0.77$) without being numerically interchangeable, with a mean absolute per-example difference of 4.3 SARI points concentrated in very short sentences, where the two differ on zero-addition edge cases that easse itself exposes as configurable rather than settled. All SARI values here use one implementation throughout, so the within-paper comparisons on which H1 rests are unaffected. Absolute values should not be compared against easse-based numbers from other papers.

Output vocabulary control. An earlier configuration of this study gave the MATE cells a word-level output vocabulary of 12,000 entries while the BPE cells predicted over 13,699 subword units, which varied output granularity alongside the input representation. Only 56.66% of reference word tokens on the SAMER test split were reachable under that word-level vocabulary, so roughly two in five correct target words were unproducible regardless of representation quality. Every number reported here comes from a corrected design in which all four cells predict over the same 13,699-entry BPE vocabulary, with the MATE cells retrained from scratch under it. We document this because the general point outlasts the paper: an ablation cell that changes the output space is not a controlled comparison of input representations.

Sample size and metric scope. The interaction is estimated from three seeds, giving two degrees of freedom, and is observed on SARI. Our analysis plan set a seed-expansion threshold at a minimum detectable effect of one SARI point, and the realised value was 1.83. We did not expand, because the effect was already significant at the pre-specified level, and we record the decision here because it was taken with the result in view. The interaction does not appear on BERTScore (Section 6). We therefore present the finding as well-evidenced within its stated scope, and a replication at ten or more seeds across both metrics would extend that scope.

Single-reference evaluation. SARI was designed for multiple reference simplifications, and the SAMER split provides one reference per source sentence. Single-reference SARI is noisier and penalises legitimate alternative simplifications, which depresses all systems in Table 4 and compresses the differences between them.

Single-corpus evaluation. All results are obtained on SAMER, a corpus of literary Modern Standard Arabic. Generalisation to other genres, to machine-translated simplification data, and to dialectal Arabic remains to be established.

Pipeline dependence. AraBit-1.58 depends on an external morphological disambiguator, and it therefore inherits that component's errors, its latency, and its domain assumptions. CAMeL Tools is developed and evaluated primarily on Modern Standard Arabic, so our results should not be assumed to transfer to dialectal Arabic, where analyser accuracy degrades and the whole argument of this paper would need to be re-examined.

Metric limitations. SARI, BERTScore, and readability indices each capture a partial view of simplification quality, and none of them is a substitute for human judgement. This manuscript reports automatic metrics only, so no conclusion here has been checked against human judgement. That is a real limitation of the evidence and not a deferred task, and it bears most directly on the readability column discussed next. Concretely, on our full test split the two ternary cells score as *more* readable by the SAMER levelled-lexicon metric than their full-precision counterparts (mean lemma level 1.82 for MATE and 1.98 for BPE, against 2.08 and 2.24 at full precision) despite being the clearly *worse*-quality systems by SARI (13.4–14.6 vs. 20.9–24.0) – almost certainly because low-quality, short, repetitive ternary output draws on a narrower, more frequent vocabulary, not because ternary quantization improves simplification. The Readability column in Table 4 must therefore be read jointly with SARI and BERTScore, never in isolation: a system can score as "simple" by producing degenerate text as easily as by producing a genuinely good simplification.

Reproducibility of energy figures. Energy measurements are sensitive to hardware, driver version, batch size, and ambient conditions. We report the measurement methodology in full and we do not compare our measured joules against analytically estimated figures from other papers, since the two quantities are not commensurable.

9. Conclusion and Future Work

This paper asked whether extreme quantization survives contact with a morphologically rich language, and whether the answer depends on how the input is represented. We proposed AraBit-1.58, which pairs a morphology-aware embedding front end with a natively trained ternary backbone, and we designed a controlled two-by-two study capable of separating the contribution of the front end from that of the backbone.

The answer to the first question is a qualified no at our scale. Ternary quantization costs a great deal of quality here: $24.01 \rightarrow 14.58$ SARI with BPE input and $20.93 \rightarrow 13.44$ with morphological input. It buys a 2.85–$2.91\times$ reduction in resident weight memory, and in our implementation, it buys nothing

in latency or energy, both of which regress for want of a fused ternary kernel. The answer to the second question is a qualified yes. The cost of quantization is smaller with morphological input than with BPE input, by 1.93 SARI, an interaction that is significant across three seeds and that the gate measurements of Section 7 corroborate mechanistically, the ternary model relying more on the morphological branch than its full-precision counterpart does. We state the claim at the width the evidence supports. MATE does not raise absolute quality at either precision, the interaction is observed on SARI and not on BERTScore, and it is estimated from three seeds. Within those bounds the finding is clear: structured input matters more as weight precision falls. Testing it at a scale where the absolute numbers are competitive is the natural next step, and these results indicate where to look.

Two aspects of the design proved consequential. Retaining a learned fallback gate rather than a hard rule allowed the model to tolerate morphological analysis errors without a separate error-correction stage. Separating the memory accounting into ternary and full-precision components revealed that the realised compression factor at small scale is substantially below the naive ratio implied by weight width alone, which is a caution the efficiency literature does not always make explicit.

The approach carries costs that we report rather than minimise. Morphological disambiguation adds approximately 0.36 ms per generated token to the ternary MATE configuration, against 0.003 ms for the BPE control, which is the entire gap between our model-only and end-to-end latency columns. The MATE tables also introduce full-precision parameters that offset part of the ternary backbone’s savings. Contrary to our design expectation, they do not reduce the embedding budget relative to byte-pair encoding. The fully inflected pattern inventory is four times the size of the control vocabulary, and MATE ends up 0.53M parameters larger overall once the learned gate is counted.

Future work proceeds in six directions. First, we intend to replace the external analyser with a jointly trained segmenter, which would remove the pipeline dependency and the latency it imposes. Second, we plan to route sparse Mixture-of-Experts capacity through the ternary backbone, extending our earlier Switch Transformer work [14] into the quantized setting, so that capacity can be increased without a proportional increase in resident memory. Third, we will evaluate on dialectal Arabic, where morphological analysers are markedly less reliable and where the central assumption of this paper is genuinely open to question. Fourth, the component ablation of Section 6 argues for a simpler front end than the one we built, keeping clitic separation and the learned gate while dropping the root and pattern embeddings, and that variant should be trained and evaluated on its own terms rather than only as an ablation cell. Fifth, two of the error-analysis directions of Section 7 require hand annotation and are left for a follow-up study. One correlates per-sentence morphological analyser accuracy, measured against a gold-annotated sample, with per-sentence SARI, which would test whether the learned gate

genuinely absorbs analyser errors or the pipeline remains brittle. The other is a manual failure taxonomy covering hallucination, meaning loss, agreement error, clitic attachment error, and failure to simplify, comparing AraBit-1.58 against the ternary BPE control. Only the gate-behaviour analysis of Section 7, which needs no hand annotation, is reported in this manuscript. Sixth, the divergence between SARI and BERTScore discussed in Section 6 calls for a replication with a larger seed budget and with a morphology-aware semantic metric, which would separate the three explanations we were able to offer but not to distinguish.

Code and Data Availability

The implementation, training and evaluation configurations, and the scripts that regenerate every table and figure reported here are available at https://github.com/AMahfodh/AraBit. The SAMER simplification corpus and the SAMER readability lexicon [28, 29] are distributed by their authors and are not redistributed in that repository. The pre-training corpus is the publicly available Arabic Wikipedia dump identified in Table 2.

Acknowledgements

The author employed Gemini-3.5-Flash for language polishing and readability enhancements, and Claude 3.5 Sonnet for coding and algorithm implementation. The author thoroughly reviewed and verified all AI-assisted outputs and takes full responsibility for the content and integrity of this manuscript.

References

[1] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A.N. Gomez, Ł. Kaiser, and I. Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 5998–6008, 2017.

[2] J. Lee-Thorp, J. Ainslie, I. Eckstein, and S. Ontañón. FNet: Mixing tokens with Fourier transforms. *Proceedings of NAACL-HLT 2022*, 4296–4313, 2022.

[3] V. Sanh, L. Debut, J. Chaumond, and T. Wolf. DistilBERT, a distilled version of BERT: Smaller, faster, cheaper and lighter. *arXiv preprint arXiv:1910.01108*, 2019.

[4] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova. BERT: Pre-training of deep bidirectional transformers for language understanding. *Proceedings of NAACL-HLT 2019*, 4171–4186, 2019.

[5] W. Fedus, B. Zoph, and N. Shazeer. Switch transformers: Scaling to trillion parameter models with simple and efficient sparsity. *Journal of Machine Learning Research*, 23, 1–39, 2022.

[6] H. Wang, S. Ma, L. Dong, S. Huang, H. Wang, L. Ma, F. Yang, R. Wang, Y. Wu, and F. Wei. BitNet: Scaling 1-bit transformers for large language models. *arXiv preprint arXiv:2310.11453*, 2023.

[7] S. Ma, H. Wang, L. Ma, L. Wang, W. Wang, S. Huang, L. Dong, R. Wang, J. Xue, and F. Wei. The era of 1-bit LLMs: All large language models are in 1.58 bits. *arXiv preprint arXiv:2402.17764*, 2024.

[8] S. Ma, H. Wang, S. Huang, X. Zhang, Y. Hu, T. Song, Y. Xia, and F. Wei. BitNet b1.58 2B4T technical report. *arXiv preprint arXiv:2504.12285*, 2025.

[9] L. Liu, Z. Zheng, C. Wang, T. Su, and Z. Yang. Binary neural networks for large language model: A survey. *arXiv preprint arXiv:2502.19008*, 2025.

[10] R. Sennrich, B. Haddow, and A. Birch. Neural machine translation of rare words with subword units. *Proceedings of the 54th annual meeting of the association for computational linguistics*, 1715–1725, 2016.

[11] A. Alshanqiti, A. Alkhodre, A. Namoun, S. Albouq, and E. Nabil. A transformer Seq2Seq model with fast Fourier transform layers for rephrasing and simplifying complex Arabic text. *International Journal of Advanced Computer Science and Applications*, 14, 888–898, 2023.

[12] G. Hinton, O. Vinyals, and J. Dean. Distilling the knowledge in a neural network. *arXiv preprint arXiv:1503.02531*, 2015.

[13] A. Alshanqiti, A. Namoun, A. Alsughayyir, A.M. Mashraqi, A.R. Gilal, and S.S. Albouq. Leveraging DistilBERT for summarizing Arabic text: An extractive dual-stage approach. *IEEE Access*, 9, 135594–135607, 2021.

[14] A. Alsughayyir, and A. Alshanqiti. A light but efficient switch transformer model for Arabic text simplification. *International Journal of Advanced and Applied Sciences*, 12, 71–78, 2025.

[15] E. Frantar, S. Ashkboos, T. Hoefler, and D. Alistarh. GPTQ: Accurate post-training quantization for generative pre-trained transformers. *International conference on learning representations (ICLR)*, 2023.

[16] J. Lin, J. Tang, H. Tang, S. Yang, W.-M. Chen, W.-C. Wang, G. Xiao, X. Dang, C. Gan, and S. Han. AWQ: Activation-aware weight quantization for on-device LLM compression and acceleration. *Proceedings of machine learning and systems (MLSys)*, 2024.

[17] T. Dettmers, M. Lewis, Y. Belkada, and L. Zettlemoyer. LLM.int8(): 8-bit matrix multiplication for transformers at scale. *Advances in neural information processing systems*, 2022.

[18] A. Pasha, M. Al-Badrashiny, M. Diab, A. El Kholy, R. Eskander, N. Habash, M. Pooleery, O. Rambow, and R. Roth. MADAMIRA: A fast, comprehensive tool for morphological analysis and disambiguation of Arabic. *Proceedings of LREC 2014*, 1094–1101, 2014.

[19] O. Obeid, N. Zalmout, S. Khalifa, D. Taji, M. Oudah, B. Alhafni, G. Inoue, F. Eryani, A. Erdmann, and N. Habash. CAMeL tools: An open source python toolkit for Arabic natural language

processing. *Proceedings of the 12th language resources and evaluation conference (LREC)*, 7022–7032, 2020.

[20] A. Abdelali, K. Darwish, N. Durrani, and H. Mubarak. Farasa: A fast and furious segmenter for Arabic. *Proceedings of NAACL-HLT 2016: demonstrations*, 11–16, 2016.

[21] A.M. Alshanqiti, S. Albouq, A.B. Alkhodre, A. Namoun, and E. Nabil. Employing a multilingual transformer model for segmenting unpunctuated Arabic text. *Applied Sciences*, 12, 10559, 2022.

[22] N. Habash. Introduction to Arabic natural language processing. Morgan & Claypool Publishers, 2010.

[23] W. Antoun, F. Baly, and H. Hajj. AraBERT: Transformer-based model for Arabic language understanding. *Proceedings of the 4th workshop on open-source arabic corpora and processing tools (OSACT4), LREC 2020*, 9–15, 2020.

[24] Y. Bengio, N. Léonard, and A. Courville. Estimating or propagating gradients through stochastic neurons for conditional computation. *arXiv preprint arXiv:1308.3432*, 2013.

[25] J. Su, M. Ahmed, Y. Lu, S. Pan, W. Bo, and Y. Liu. RoFormer: Enhanced transformer with rotary position embedding. *Neurocomputing*, 568, 127063, 2024.

[26] B. Zhang, and R. Sennrich. Root mean square layer normalization. *Advances in neural information processing systems*, 2019.

[27] I. Loshchilov, and F. Hutter. Decoupled weight decay regularization. *International conference on learning representations (ICLR)*, 2019.

[28] B. Alhafni, R. Hazim, J.D. Pineros Liberato, M. Al Khalil, and N. Habash. The SAMER Arabic text simplification corpus. *Proceedings of LREC-COLING 2024*, 16079–16093, 2024.

[29] M. Al Khalil, N. Habash, and Z. Jiang. A large-scale leveled readability lexicon for standard Arabic. *Proceedings of LREC 2020*, 3053–3062, 2020.

[30] S.S. Al-Thanyyan, and A.M. Azmi. Simplification of Arabic text: A hybrid approach integrating machine translation and transformer-based lexical model. *Journal of King Saud University – Computer and Information Sciences*, 35, 101662, 2023.

[31] M.K. Eddine, N. Tomeh, N. Habash, J. Le Roux, and M. Vazirgiannis. AraBART: A pretrained Arabic sequence-to-sequence model for abstractive summarization. *Proceedings of the seventh arabic natural language processing workshop (WANLP)*, 31–42, 2022.

[32] E.M.B. Nagoudi, A. Elmadany, and M. Abdul-Mageed. AraT5: Text-to-text transformers for Arabic language generation. *Proceedings of the 60th annual meeting of the association for computational linguistics*, 628–647, 2022.

[33] T. Dettmers, A. Pagnoni, A. Holtzman, and L. Zettlemoyer. QLoRA: Efficient finetuning of quantized LLMs. *Advances in neural information processing systems*, 36, 10088–10115, 2023.

[34] W. Xu, C. Napoles, E. Pavlick, Q. Chen, and C. Callison-Burch. Optimizing statistical machine translation for text simplification. *Transactions of the Association for Computational Linguistics*, 4, 401–415, 2016.

[35] T. Zhang, V. Kishore, F. Wu, K.Q. Weinberger, and Y. Artzi. BERTScore: Evaluating text generation with BERT. *International conference on learning representations (ICLR)*, 2020.

[36] M. El-Haj, and P. Rayson. OSMAN: A novel Arabic readability metric. *Proceedings of LREC 2016*, 250–255, 2016.

[37] K. Papineni, S. Roukos, T. Ward, and W.-J. Zhu. BLEU: A method for automatic evaluation of machine translation. *Proceedings of the 40th annual meeting of the association for computational linguistics*, 311–318, 2002.

[38] E. Sulem, O. Abend, and A. Rappoport. BLEU is not suitable for the evaluation of text simplification. *Proceedings of EMNLP 2018*, 738–744, 2018.