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

Volume VIII, Issue I, July 2026, Pages 46-61

Mapping Research Trends in Generative Artificial Intelligence: A Bibliometric and Network Analysis

Abdulbasid S. Banga [1, *]

[1] College of Computing and Informatics, Saudi Electronic University, Riyadh 11673, Saudi Arabia, a.banga@seu.edu.sa

*Corresponding author: (Abdulbasid S. Banga), *Email Address*: a.banga@seu.edu.sa

Abstract

This study investigates the integration of user-generated social media data into Public Health Surveillance (PHS) through a dynamic, multidimensional analytical framework. Existing approaches largely rely on offline processing of static social network data, leaving a gap in methods capable of handling high-velocity, continuously evolving social streams. We propose a dynamic model that updates multidimensional representations of social data in real time using unsupervised text-mining techniques. By jointly analyzing semantic content and temporal posting patterns, the framework identifies emerging events, topics, and influential or relevant users. To enhance the utility of social data for PHS, we introduce quantitative quality measures that filter low-value or out-of-domain user profiles. The approach is evaluated on a multi-year Twitter data stream, demonstrating its effectiveness in isolating meaningful signals and excluding noisy contributors. We further outline procedures for deriving user profiles from self-descriptions to support targeted filtering. The results show that the proposed model enables robust topic and event detection, audience characterization, and impact assessment. Overall, the dynamic multidimensional framework provides a scalable and adaptable foundation for incorporating social media intelligence into PHS systems.

Keywords: Health surveillance; Social network analysis; Multidimensional analysis; Text mining.

https://doi.org/10.63070/jesc.2026.016
Received 20 November 2025; Revised 23 January 2026; Accepted 21 February 2026;
Available online 04 March 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

PHS is the process of continually gathering, analyzing, and interpreting health-related data, then promptly sharing these findings with public health practitioners, clinicians, and policymakers for use in controlling and preventing disease [1]. In the last ten years, there have been a number of studies showing how social media data can be used along with traditional surveillance methods to provide information on the prevalence of disease, emergencies, indicators of outbreaks, and how people react to official public health messages [2,3].

Historically, public health surveillance has been based primarily on clinic-based data and periodic survey data collected from providers and pharmacies. While social media is not a substitute for official health data, it can add value to the PHS process by providing timely, easily accessible, and inexpensive information that reflects the public's concerns about and behaviors towards health. The benefit of using social media in PHS, however, is not just through collecting such data it also includes the ability to filter, integrate, and analyze a large amount of different and rapidly changing data. Current methods of utilizing these types of data typically use static datasets and ad hoc analyses on each dataset to answer only specific questions. As a result, these types of analyses have limited generalizability and scalability to other public health settings [3,4]. That being the case, their methods and output aren't transferable to other scenarios. Nevertheless, PHS agencies need the creation of versatile technologies equipped with sophisticated techniques to capture, integrate, and analyze social media streams for various objectives [2].

This research is primarily aimed at providing an adaptable analytical tool to assist public health (PH) agencies with analysis of the evolution of social media postings using multiple viewpoints. The method proposed by this program will allow PH agencies to analyze who is posting, what topics are being discussed, and how information flows over time within social media streams. In addition to providing real-time monitoring of emerging patterns, warnings, and events through continuous access to live social media feeds, the system supports both online analysis and historical analysis in order to allow PH practitioners to convert massive amounts of unstructured social media data into actionable public health information.

2. Procedures

Despite only a segment of people who use social media sharing health information explicitly, the total number of social media users can provide helpful data that is in addition to traditional public health data systems, which support PHS efforts [21,54]. Social media (and other digital) posts frequently

reflect public opinion, experience, and reaction towards a given health event and are therefore useful for surveillance purposes. However, while social media data can be valuable to generate predictive and descriptive attributes of health, extracting reliable and valid health data requires systematic methods to filter out and analyze the raw data, which can be heterogeneous, noisy, and ever-evolving over time. Consequently, social media data should be considered as an additional, rather than a replacement, source of information for the established PHS services.

2.1. Receiving parties

Social Media Users There are two different types of social media users, Organizations and Individuals. Organizations use social media to represent themselves including associations, laboratories, government agencies, journals, companies, etc.. Individual Social Media Users include everyone else from health care professionals to the General Public. As a result, lay social media users typically share information related to symptoms, diagnoses, opinions, emotions, and personal experiences. On the other hand, Organizations use social media primarily to make announcements, promote campaigns/activities, disseminate updates regarding studies, perform marketing activities, etc., and provide scientific findings.

Organizational and Individual Social Media Users exhibit distinctly different engagement patterns. Whereas many health organizations are primarily concerned with heightened public awareness of disease prevention practices, research-oriented accounts are focused on encouraging the dissemination of scientific knowledge. In general, Organizational JMRI accounts tend to have much higher levels of engagement and overall credibility than Individual User accounts. From a PHS perspective, the level of health knowledge of all Users could be classified as Expert or Professional and the level of personal engagement could be classified as Patient, Relative, and/or Observer.

2.2. Publishes articles

It is possible to use the words in the postings not just to find and sort them, but also to label them with subjects that provide varying degrees of information about what they include. Medical condition, health campaign, and financial service are the wide categories, skin disease, WHO alert, and AIDS treatment are the narrower ones, but varied topics fulfill one. One can also easily detect annotation elements that refer to real-world named entities, which include: people, places, organizations, diseases, and medications. Such information items as the location of posting users, hashtags (# sign), and screen names (@ sign) should be taken into consideration when annotating postings.

2.3. Polarity of posts

Social media has become very common as people get to express themselves, whether positively or negatively. For social awareness and acceptability analysis, alarm and conflict detection, and processing post contents into polarity values, [19] is a valuable tool.

Post's timeliness. Because users update their accounts as events unfold, PHS can rely on social media for up-to-the-minute news. Health problems that manifest consistently, seasonally, or as a result of an epidemic are the most often mentioned in the postings. Researchers may be able to sort postings by time and filter them based on their temporal dimension (e.g., daily, weekly, monthly, etc.). Tracking the development of facts, identifying responses to announcements, finding fresh information, and even filtering out noisy data would all become much simpler with this. For instance, if the distribution curve of melanoma postings has a peak, it might mean that there are a lot of useless posts in response to a prominent person's diagnosis.

We propose a method wherein various features are represented by semantic annotations automatically generated by intelligent analysis of incoming postings. Additionally, massive volumes of social media data may be filtered and classified by this technique. This allows analysts to categorize the postings and provide various summaries by combining some or all of these elements.

Seasonal ads informing about melanoma prevention strategies, for instance, may pique the attention of PHS officers who want to gauge public response. The resultant collection may be huge and challenging to handle after collecting the relevant articles using keywords. In our method, these postings are tagged by organizations, general lay users, or patients based on the advanced processing. They are also tagged with views that are good or negative in tone, among other things. Next, we may find useful indicators that evaluate social responses from multiple perspectives by imposing conditions on these features and the postings' temporality.

Factors including age, sex, locality, and ethnicity are additional considerations in demographic and epidemiological research pertaining to social media users. There have been several attempts to derive this data from post contents, as lay users do not mention them in their profiles [10,17,55]. At this time, our research tools do not take these factors into consideration.

2.4. A Multi-Factor Analysis of Streams from Twitter

Our method is based on multidimensional data models that are created dynamically over streams of Twitter data [31]. Figure 1 shows the star-based schema that we propose for PHS. To facilitate analysis in accordance with PHS requirements, this schema primarily aims to provide stream-level summaries of themes and events.

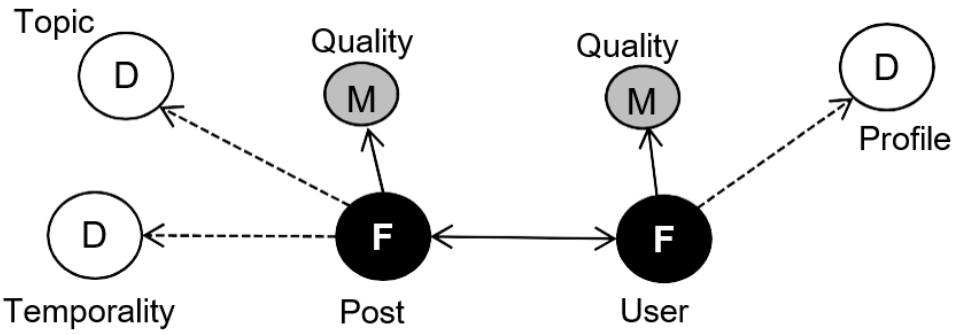


Figure 1. Analyzing Twitter streams using a multi-dimensional model: a brief overview. Dimensions are denoted by (D), facts by (F), and metrics by (M), using the standard nomenclature for multidimensional models. Dimensions are produced dynamically from the streaming data, as shown by the dotted lines.

Two distinct fact types are used to represent the social media data items. The Post Facts include data derived from the post's text, as well as its temporal aspects, subjects, and other quality indicators for tweets, such as the number of responses or retweets. From a different perspective, the User Facts are drawn from the profile data comprising the descriptions along with the other quality factors located at the user tier, e.g., the official verification status of the sending account. Keep in mind that it is possible to get certain user metrics from the sum of a user's post metrics across all of their tweets. The two-headed arrow that goes from "Post" to "User" shows the binary relationship that exists between a user's postings and the tweet's poster. Metrics like a user's total number of responses to relevant posts need this communication.

In the suggested paradigm, the values of the dimensions are determined by the incoming data, making them dynamic. One example is the Topic Dimension, which takes into consideration the primary topics discussed in tweets based on statistical inferences made after processing a certain number of tweets. In particular, in this article, we give a bigram-based temporal analysis study in order to determine these themes within the Twitter stream. Posts may also be characterized by the way the published material changes over time. A tweet may be an opinion piece, a component of a larger news story, or even part of a campaign to raise awareness about health issues. So, to correctly categorize and analyze the stream's data, we need to determine the Temporality Dimension of postings.

The quality of the posts, as well as that of the users, is a significant part of our business. The measurements of the content of the posts (achieved by processing the text part of posts) and the measurements provided by Twitter in the posts metadata (e.g., the number of responses to the tweet or the number of followers of a poster) can be combined to form parameters of the Quality Dimension.

The information's reliability in relation to the current analysis is shown by the quality metrics that are produced. As an example, a large number of people who post often could be spammers whose postings have a strong bias towards certain subjects. Because of complicated elements like metaphors, irony, and sarcasm, or because keywords are ambiguous, some communications might be little more than background noise. This latter point is well recognized in the literature due to the sarcastic usage of some medical terms in other fields, such as politics, such as cancer and metastasis. The purpose of quality metrics is to evaluate the data processing's conformity to the analysis domain.

Finally, there will be a Profile Dimension that the users can use to identify their main objectives upon engaging the social media platform. When doing social network research, user profiling is useful for gaining insight into the demographics of content creators and consumers. A dynamic differentiation between journalists, healthcare providers and services, and concerned citizens is something we'd want to see implemented for PHS. Those who have a genuine interest in the health conditions under observation are considered concerned parties (e.g., cancer survivors, patients' families). Users are profiled based on the little biographical information they supply on their Twitter profiles. In particular, we extract relevant frequent bigrams from user descriptions to identify and manually categorize a collection of individuals. We then use the identified customers to train a neural-network-based classifier to be able to predict the profiles of new users.

We set up a Twitter multi-keyword feed with some of the PHS domain themes (see Table 1) to demonstrate the effectiveness of the proposed streamed multidimensional research methodology. Reproducing a big and diverse stream of tweets for monitoring is the goal of the query.

Table 1. Terms related to the public health surveillance's Twitter feed.

Category	Details
Conditions	cancer, tumor, carcinoma, leukemia, lymphoma, metastasis, sarcoma, skin cancer, melanoma
Treatments	chemotherapy, homeopathy
Observations	UV, Cholesterol, LDL, MRI

Domain coherence and vocabulary variety are two quality measures that we create for user analysis based on the information they post. The word distribution in the tweets serves as the basis for both measures' language models. We find the domain coherence by adding the language model of the user to a distribution of UMLS Meta Thesaurus ® medical words. In our tests, we discovered that UMLS encompassed a lot of medical terms, such as the most commonly used forms in which non-medical

individuals refer to illnesses and ailments. To quantify the range of words used, we propose using the I index by Yule [56] on non-UMLS terms, too.

3. Results

We collected an extensive number of tweets in six months (June to December 2019) with the help of two methods: The Twitter API search method and keywords mentioned in Table 1. In this stream, 395,804 individuals provided a total of 777,778 tweets; 77 percent of them had made only one tweet. In regard to the overall quality of the stream, it is worth mentioning that the obtained data are very redundant; only 281,260 tweets contained original data, whereas the rest were copies of already existing tweets. Because of this, it's easy to see how some streams may favor the most popular tweets. Last but not least, remember that we analyzed user interaction by treating the retweets in the stream independently.

3.1 Users Analysis

While the majority of spam and automated accounts will have an extremely high posting frequency along with low Yule's I's (generally <30), they also tend to post content that does not relate to the health domain. By contrast, individuals who have high Yule's I scores but low domain coherence numbers also tend to be prolific posters; for instance, they will likely have screen names containing streams keyword phrases, yet their posts will have nothing to do with health. Figure 2 illustrates how the filters created using these metrics help to successfully identify and eliminate irrelevant users from consideration. The result of such filtering would eliminate around 15% of total tweets originally collected. Table 2 displays further information on the metrics of the top users.

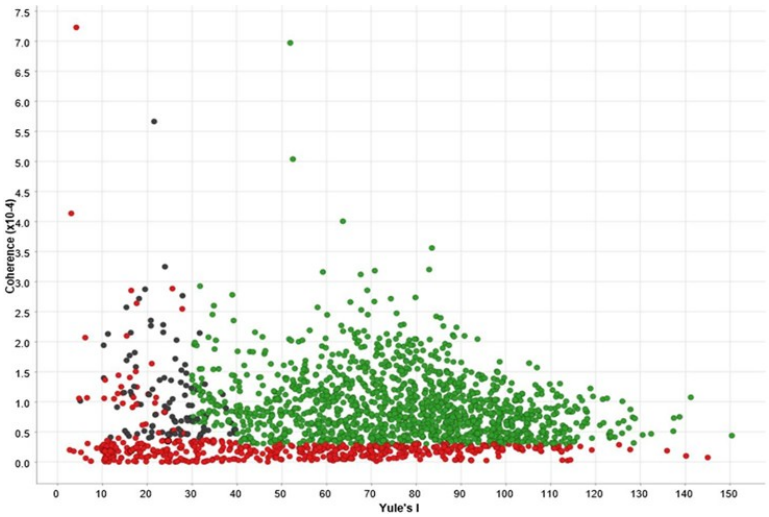


Figure 2. User quality measures according to the number of tweets published in the stream of the two thousand highest users. Positive users are represented by green points, users who need human examination are represented by black points, and users that should be filtered out are represented by red points.

Table 2. Productivity indicators for the top ten users. Stream keywords are bolded in the screen names column. Rows that are shaded represent users who are not in the domain.

Screen Name (User)	#Tweets	Yule's I	Coherence (10^{-4})	Description
Qunol Official	3306	42.3	0.88	Forum
ldl_bot	2556	94.3	0.21	Comics bot
Lymphoma Papers	2110	75.4	1.39	Academic bot
treda10	1740	70.6	1.15	Company CEO
Avos From Mexico	1623	49.9	0.33	Food sales
Eureka Mag	1194	71.2	1.87	Magazine
med vizor	1053	56.6	0.65	Forum
beechnutwx	1010	11.2	0.31	Weather bot
SnoMiddleForkWX	963	29.8	0.32	Weather bot
Sara_sarcoma	947	113.7	0.02	Influencer

3.2 Events and Topics

The identification of health-related events and subjects within the Twitter stream are performed through the application of a temporal frequency analysis on bigrams. In order to create a list of candidates from the data, three conditions must be satisfied for any given bigram:

1. kurtosis ≥ 10 thus, it will exclude all bigrams that do not show sufficient temporal variation

2. peak occurrence ≥ 20 tweets therefore, this allows for inclusion of only those bigrams that were associated with notable spikes of tweet activity

3. temporal window ≤ 10 days as such, this is used to define how long a specific bigram may remain in the Twitter stream.

Applying these thresholds led to the creation of more than 3,000 potential candidates for bigrams. These candidates have been grouped into 260 distinct health-related topics/events based on a graph modularity clustering algorithm [57].

Events account for just around 7% of all tweets, whereas subjects account for 51% of all incoming tweets. Furthermore, there is a correlation between at least one subject and 60% of event-related tweets. Therefore, we need to be mindful of event-related tweets when we analyze certain issues because of the influence they might have.

The most common occurrences identified in English postings are shown in Table 3, while those in Spanish are shown in Table 4. Most of the time, events revolve around news stories involving well-known politicians or athletes. A large number of incidents in Spanish concern youngsters whose illnesses spread to other parts of the world. The quality measures used in these two tables are Y, the average Yule, I, %Ver., the percentage of verified accounts, and OnDom., the average tweets per user

in the stream. In addition, the number of people who tweeted and retweeted about these occurrences is shown in the last column.

Table 3. Top Frequent Events (English)

Event (bigram)	#Tweets	Main Topic	User's Quality (Y, %Ver., On Dom.)	#Users Tweet./Retw.
Alex Trebek	2659	Pancreas cancer	(70, 31%, 6)	2200 / 4535
Carlos Carrasco	659	Leukemia	(72, 28%, 4)	589 / 2473
Tapeworm woman	554	False tumor	(65, 39%, 6)	437 / 332
Ross Perot	357	Leukemia	(85, 23%, 4)	335 / 1093
Bank holidays (chemo canceled)	86	Chemotherapy	(68, 0%, 5)	84 / 27

Table 4. Top Frequent Events (Spanish)

Event (bigram)	#Tweets	Main Topic	User's Quality (Y, %Ver., OnDom.)	#Users Tweet./Retw.
Tabare Vázquez	921	Lung Tumor	(57, 27%, 6)	691 / 1521
Carlos Carrasco	854	Leukemia	(55, 33%, 5)	262 / 622
Francia homeopatía	338	Homeopathy	(59, 11%, 6)	293 / 638
Luis Enrique	282	Cancer	(71, 13%, 3)	102 / 1331
fármaco evita	108	Cancer	(61, 7%, 7)	103 / 84

The number of chosen bigrams for subjects is close to four thousand, which is the same as for events. There are 139 distinct ways to classify these bigrams. Table 5 shows the most common bigrams related to various cancer kinds in English, whereas Table 6 shows the same information for Spanish. The topics containing two words, e.g., "leukemia" and "melanoma were represented by unigrams as opposed to bigrams because their semantics are associated with one word. It should be mentioned that topics involve the incidents relevant to subjects; hence, we need to be cognizant of how these events affect the topics (for instance, the "leukemia" topic). Take note that, owing to awareness initiatives linked with themes, topic-related users tend to be more active than event-related users.

Table 5. Top Frequent Topics (English Subset)

Topic	#Tweets	User's Quality (Y, %Ver., OnDom.)	#Users Tweet./Retw.	%Tweets (Replies, 1st Person)
leukemia	27,000	(76, 6%, 12)	3,200 / 17,300	(16%, 12%)
melanoma	26,791	(78, 8%, 15)	2,700 / 3,364	(14%, 7%)
skin cancer	19,000	(72, 7%, 5)	8,219 / 23,022	(23%, 12%)
brain tumor	18,900	(70, 5%, 5)	6,091 / 11,268	(37%, 28%)
lung cancer	1,500	(75, 4%, 45)	401 / 396	(18%, 8%)

According to Table 5, discussions about leukemia and melanoma made up most of the English-language content considered in this study. There was a considerable amount of public interest in these diseases, as evidenced by the many replies to and sharing of information about the topics. Skin cancer and brain tumors were also widely discussed, and many people responded to posts about those types of cancers by sharing their own experiences. Overall, it appears that users of English-language Twitter were talking about a combination of public awareness efforts and sharing their own experiences with significant cancer types.

Table 6. Top Frequent Topics (Spanish Subset)

Topic	#Tweets	User's Quality (Y, %Ver., OnDom.)	#Users Tweet./Retw.	%Tweets (Replies, 1st Person)
leucemia	12,390	(57, 8%, 4)	1,388 / 11,108	(23%, 7%)
tumor cerebral	4,451	(67, 5%, 3)	1,264 / 9,951	(23%, 16%)
melanoma	3,726	(60, 15%, 9)	208 / 421	(8%, 3%)
cáncer mama	2,800	(58, 4%, 6)	221 / 188	(17%, 12%)
cáncer páncreas	1,100	(57, 3%, 4)	259 / 242	(2%, 3%)

Table 6 shows that the top topic of discussion among Spanish speaking users are conditions related to Leukaemia, while Brain Cancer and Melanoma followed. There are fewer Verified Users discussed in Spanish when compared to English users; however, because there are more retweets in relation to Spanish Tweets, this indicates a more dynamic flow of information in Spanish based on first-hand experiences by those impacted by these conditions and their families and friends.

The first-person perspective and the response percentages for each subject (the final columns of Tables 5 and 6) are two noteworthy aspects that we examine. Politics is a valid subject for around 4% of melanoma-related tweets. By examining the most common subtopics (bigrams) linked to melanoma, this fact has been readily discovered. Compared to the overall melanoma topic's response rate of 14%, we see that 70% of the relevant tweets are responses in this instance. The ratio of persons directly impacted by the illnesses is shown by using the first person; on the other hand.

Lastly, Figure 3 provides an analysis of the subjects and events organized by user profiles, showing the writers (blue) and the audience (red). Section 3 explains that the people who submit relevant tweets are called authors, and that the people who are interested in retweeting such tweets are called the audience. Then, users may be categorized into distinct profiles based on the language models used in their descriptions. The language models in these profiles are presented in Table 7.

Table 7. Language Models for Different Profiles According to the Users' Descriptions

Profile	Definition	Top Words in Users' Descriptions
professional	Specialists	radiosurgery, oncoplastic, oncologist, haemato-oncology, consultants
ex-professional	Retired people	retired, former, senate, viet, colonel
student	Students	student, thesis, undergraduate, engineering, studying
sports	Sportsman	runner, marathon, athlete, skater, rider
religion	Religious terms	amin, allah, savior, hindu, jesus
p. services	Public services	traffic, 24h, breaking, weather, protection
h. services	Health-care services	urgencias, screenings, specialties, uninsured, visitors
concerned	Disease-concerned people	survivor, reminder_ribbon, survivorship, warrior
pseudo	Alternative therapies	reiki, meditation, yoga, ayurveda, remedial
journalist	Press and publications	medline-indexed, issn, journal, indexed, open-access
others	Other people	zombies, yin-yang, weekends, voracious, virgo (all with low probs)

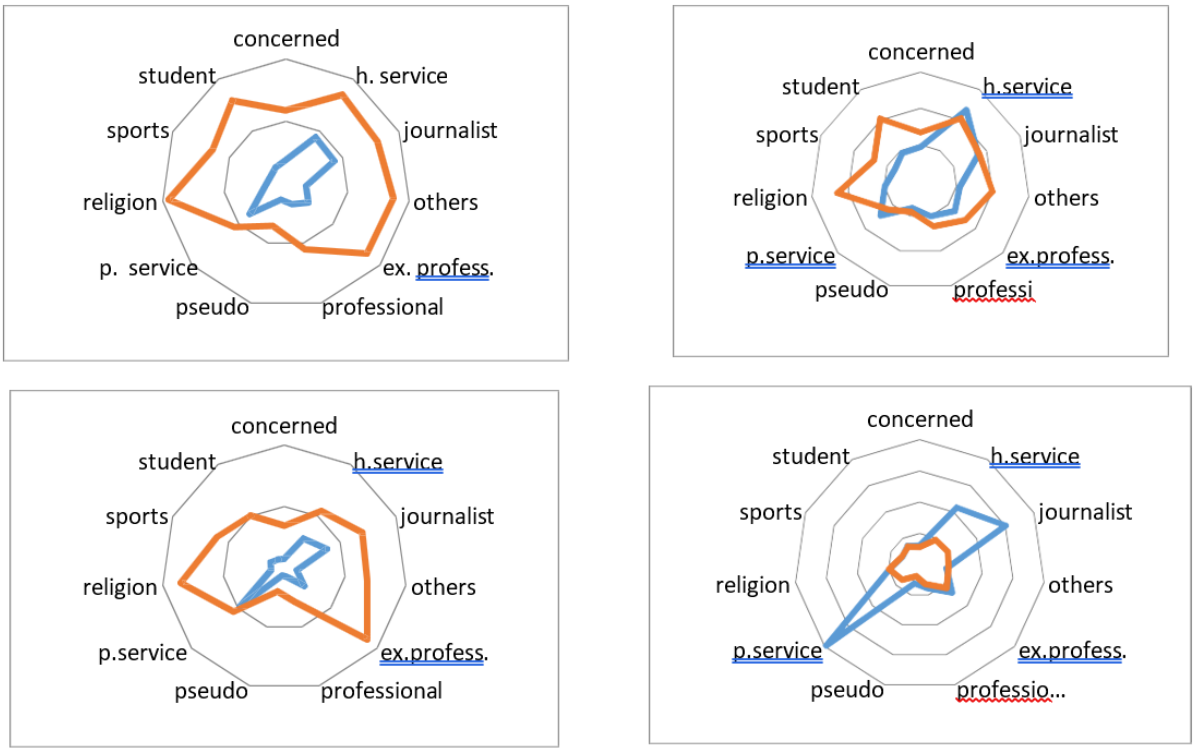


Figure 3. Using blue for writers and red for attendees, we can compare profiles from various events and subjects: (a) First, the subject of leukemia; second, the occurrence involving Carlos Carrasco. part (b) Cancer of the skin (topic) and tapeworm (event) (bottom to top).

Take note of how the profiles linked with various event types and topics change in terms of nature and influence. The postings on the "Carlos Carrasco" case and the "leukemia" issue in Figure 3 (on the left) clearly originate from journalists and the public, or health services. Except for those who employ alternative therapies (pseudo in Figure 3), it similarly depicts a huge audience that is evenly dispersed. On the other hand, the statistic shows that there are more diverse sorts of writers contributing to the "skin cancer" issue. This includes a plethora of tweets generated by public and health services, likely in the context of preventative initiatives. Finally, many writers have reported the unusual incidence of the "tapeworm" (perhaps in an effort to define it or avoid its effects), but only a tiny number of others have retweeted it.

3.3 Contentious and Open to Debate Subjects

Finally, to illustrate the application of the proposed analytical model, we have selected a group of events and subjects the peculiarities of which underlie that they may be observed in the course of PHS works. Table 8 displays a few events and themes that do not have a high number of users or an appropriate audience. You may have noticed that, except for the first subject, none of their audience members have verified accounts. Issues 1 and 2 are campaign-related topics. There are a lot of

messages in Topic 4 and Topic 3; however, the Yule's I scores are poor. For all intents and purposes, they are spam subjects; they do not provide any useful information. Lastly, an example of an unconfirmed occurrence with a miracle treatment is Topic 5.

Table 8. Controversial Events/Topics

Topic/Event	User's Quality (Y, %Ver., OnDom.)	#Users Twitting/Retw.	%Tweets (Replies, 1st Person)
1. black skin	(85, 5%, 2)	180 / 5915	(43%, 22%)
2. force chemo	(97, 0%, 5)	66 / 15	(3%, 0%)
3. *evita metástasis*	(24, 0%, 282)	109 / 185	(0%, 0%)
4. *fármaco frenar*	(25, 0%, 255)	115 / 366	(0%, 0%)
5. *desaparece oración*	(56, 0%, 4)	83 / 3	(0%, 0%)

Regarding the content of the tweets concerning these events/topics, among the hot issues are the color black, that is, the skin, which corresponds to the comment, black skin, as far as skin cancer is concerned, the responses comprise 43%. Only 3% of respondents have voiced a view on the subject of "force chemo," suggesting a lack of strong opposition to a government mandate. Even though numerous people have written on the remaining topics/events in Table 8, none of them have received any replies.

4. Discussion

To enhance the use of social media information in their multidimensional models in the decision support systems, several firms are now using cutting-edge Social Business Intelligence technology to create solutions. In this context, developing systems that generate information in real-time is a significant challenge, as is modifying the analytical models to accommodate dynamic data streams. Using this technology to build health monitoring systems is the primary contribution of this paper's work. Thus, the study presents a multi-dimensional model for analyzing Twitter streams with the purpose of intelligently using public health information and shows how it may be used for PHS tasks.

Posts are characterized by metrics and features that indicate their quality, duration, users, and content in the suggested model. After that, the incoming data stream may be filtered and summarized using these semantic annotations. With the final system in place, we can monitor PHS tasks for mentions of certain themes and events on social media. Incorporating quality measures into the multidimensional model is a strength of the suggested strategy that allows for accurate PHS analysis. Another good feature to be mentioned is the utilization of unguided statistical methods to determine themes, events, and user profiles. This allows for continuous tuning of the multidimensional model to the observed data stream. This research examines application scenarios that show how the suggested tool might be beneficial for PHS-related intelligent analysis of social data and value generation. By using this

conceptual framework, several dimensions may be derived to elucidate occurrences and themes that can be subject to on-demand process monitoring. This is because profiles and user-generated data are summarized and linked to the users' activities. The suggested technique can allow health officers to address the major applications of social media in the PHS: social monitoring, situational awareness, and the surveillance of communications.

Insights into the most noteworthy issues and occurrences connected to the monitored subjects on Twitter, with an emphasis on the kind and caliber of the individuals participating, may be gained from the examined use cases. We can do a better job rating the author’s credibility, for example, is the content from authorities, real users or bots, by looking at the author dimension and quality metrics. A suspected illness epidemic and associated occurrences may be quickly examined by grouping post metrics by often relevant subjects. Our tests, for instance, used data on first-person tweets and responses tagged with "leukemia" to find self-reported treatment experiences. However, event detection provides an opportunity to potentially estimate the present illness incidence rate. During times of crisis, the same approach might be used to identify self-reported need for humanitarian aid. By identifying the user type that generates the data and the people who consume it, as well as by tracking their responses to specific messages, we may learn about their interrelationships and track how people are aware of and perceive PHS events.

5. Conclusions

To aid PHS systems, we have introduced a new approach to multidimensional analysis of Twitter streams in this article. The proposed solution does not need a manual, and it can automatically detect themes and events within the stream according to its semantics and timeline attributes, indicating the capabilities of the solution to provide an actual summary of the contents of these objects and the people who published or interacted with the content. Our experiments illustrate the application of such multidimensional models to PHS tasks based on a real-life long-term cancer stream. Prioritizing data quality and the target audience for observed events and subjects is important to the suggested approach. This study addressed the objectives through a direct examination of the impacts of the multidimensional model. The evidence produced shows how well the model works to develop real-time identification and analysis systems for Health-related Topics, Events and User Response based on Social Media data streams.

In the future, our focus will be on establishing clear key performance indicators (KPIs) for PHS applications. For example: the time from initial detection of an outbreak until public notification is sent , how quickly the topic of public health alert spreads, and measuring the degree of the user's

influence based on their reliability, credibility, and trustworthiness. Additionally, we intend to further develop an automated method for classifying events and topics by linking them to existing medical taxonomies, such as the Unified Medical Language System (UMLS) Meta-thesaurus, as well as utilizing advanced techniques from text mining. Ultimately, the goal of these improvements is to improve the framework's interpretability, robustness, and operational applicability for real-world PHS applications and enhance the overall reliability and value of the proposed system.

Conflicts of Interest: There is no conflict of interest, the writers assert.

Ethical and Privacy Considerations

The present study has been done with only publicly available Twitter data according to the platform's terms of service and data use policies without attempting to identify or deanonymize individual users. Analyses were done using aggregate-level data in order to reduce privacy risks. The present research adheres to the data protection principles that are defined in laws, such as the General Data Protection Regulation (GDPR), which encompass responsible handling, minimizing, and ethical use of user-generated content for public health research.

Funding

The authors declare that this research received no external funding.

Data Availability Statement

Not applicable.

References

[1] B. M. Gomes, C. B. Rebelo, and L. A. De Sousa, “Public health, surveillance systems and preventive medicine in an interconnected world,” in *One Health*, 2022, pp. 33–71. doi: 10.1016/b978-0-12-822794-7.00006-x.

[2] A. Gupta and R. Katarya, “Social media based surveillance systems for healthcare using machine learning: A systematic review,” *Journal of Biomedical Informatics*, vol. 108, p. 103500, Jul. 2020, doi: 10.1016/j.jbi.2020.103500.

[3] O. Edo-Osagie, B. De La Iglesia, I. Lake, and O. Edeghere, “A scoping review of the use of Twitter for public health research,” *Computers in Biology and Medicine*, vol. 122, p. 103770, May 2020, doi: 10.1016/j.compbiomed.2020.103770.

[4] L. Zhou, D. Zhang, C. C. Yang, and Y. Wang, “Harnessing social media for health information management,” *Electronic Commerce Research and Applications*, vol. 27, pp. 139–151, Dec. 2017, doi: 10.1016/j.elerap.2017.12.003.

[5] H. Ögel-Balaban, “The use of online social network sites during the COVID-19 pandemic as a protective or risk factor for well-being of university students,” *Cyberpsychology Journal of Psychosocial Research on Cyberspace*, vol. 16, no. 3, Jun. 2022, doi: 10.5817/cp2022-3-4.

[6] Y. Wang *et al.*, "From virus to viral: content analysis of HIV-related Twitter messages among young men in the U.S.," *BMC Digital Health*, vol. 2, no. 1, Jul. 2024, doi: 10.1186/s44247-024-00102-y.
[7] O. Edo-Osagie, B. De La Iglesia, I. Lake, and O. Edeghere, "A scoping review of the use of Twitter for public health research," *Computers in Biology and Medicine*, vol. 122, p. 103770, May 2020, doi: 10.1016/j.compbiomed.2020.103770.
[8] J. Dietrich *et al.*, "Adverse Events in Twitter-Development of a Benchmark Reference Dataset: Results from IMI WEB-RADR," *Drug Safety*, vol. 43, no. 5, pp. 467–478, Jan. 2020, doi: 10.1007/s40264-020-00912-9.
[9] A. Rawat, M. A. Wani, M. ElAffendi, A. S. Imran, Z. Kastrati, and S. M. Daudpota, "Drug adverse event detection using Text-Based Convolutional Neural Networks (TextCNN) technique," *Electronics*, vol. 11, no. 20, p. 3336, Oct. 2022, doi: 10.3390/electronics11203336.
[10] "Public health surveillance for COVID-19 using Twitter sentiment analysis," *Texila International Journal of Public Health*, vol. 12, no. 3, Sep. 2024, doi: 10.21522/tijph.2013.12.03.art034.
[11] M. Myslín, S.-H. Zhu, W. Chapman, and M. Conway, "Using Twitter to examine smoking behavior and perceptions of emerging tobacco products," *Journal of Medical Internet Research*, vol. 15, no. 8, p. e174, Aug. 2013, doi: 10.2196/jmir.2534.
[12] X. Ji, S. A. Chun, Z. Wei, and J. Geller, "Twitter sentiment classification for measuring public health concerns," *Social Network Analysis and Mining*, vol. 5, no. 1, p. 13, May 2015, doi: 10.1007/s13278-015-0253-5.
[13] A. Alessa and M. Faezipour, "Flu outbreak prediction using Twitter posts Classification and linear regression with Historical Centers for Disease Control and Prevention Reports: Prediction Framework Study," *JMIR Public Health and Surveillance*, vol. 5, no. 2, p. e12383, May 2019, doi: 10.2196/12383.
[14] P. P. Morita, I. Z. Hussain, J. Kaur, M. Lotto, and Z. A. Butt, "Tweeting for health Using Real-time mining and Artificial Intelligence–Based Analytics: Design and development of a big data ecosystem for detecting and analyzing misinformation on Twitter," *Journal of Medical Internet Research*, vol. 25, p. e44356, Mar. 2023, doi: 10.2196/44356.
[15] P. Missier *et al.*, "Tracking dengue epidemics using Twitter content classification and topic modelling," in *Lecture notes in computer science*, 2016, pp. 80–92. doi: 10.1007/978-3-319-46963-8_7.
[16] K. W. Prier, M. S. Smith, C. Giraud-Carrier, and C. L. Hanson, "Identifying Health-Related topics on Twitter," in *Lecture notes in computer science*, 2011, pp. 18–25. doi: 10.1007/978-3-642-19656-0_4.
[17] W. H. Bangyal, M. Iqbal, A. Bashir, and G. Ubakanma, *Polarity Classification of Twitter Data Using Machine Learning Approach*. 2023, pp. 1–6. doi: 10.1109/hccs59561.2023.10452557.
[18] Y. Liscano, L. a. A. Arrieta, J. F. Montenegro, D. Prieto-Alvarado, and J. Ordoñez, "Early warning of infectious disease outbreaks using social media and digital data: A scoping review," *International Journal of Environmental Research and Public Health*, vol. 22, no. 7, p. 1104, Jul. 2025, doi: 10.3390/ijerph22071104.
[19] M. D. Soufi, P. R. Hachesu, and R. Ferdousi, "Oncology Informatics for Lynch Syndrome Research and Care: A Literature review," *JCO Clinical Cancer Informatics*, vol. 6, no. 6, p. e2200087, Nov. 2022, doi: 10.1200/cci.22.00087.
[20] S. Amin, M. I. Uddin, M. A. Zeb, A. A. Alarood, M. Mahmoud, and M. H. Alkinani, "Detecting information on the spread of dengue on Twitter using artificial neural networks," *Computers, Materials & Continua/Computers, Materials & Continua (Print)*, vol. 67, no. 1, pp. 1317–1332, Jan. 2021, doi: 10.32604/cmc.2021.014733.
[21] J. Sleigh, J. Amann, M. Schneider, and E. Vayena, "Qualitative analysis of visual risk communication on twitter during the Covid-19 pandemic," *BMC Public Health*, vol. 21, no. 1, p. 810, Apr. 2021, doi: 10.1186/s12889-021-10851-4.

[22] A. H. Zadeh, H. M. Zolbanin, R. Sharda, and D. Delen, "Social media for nowcasting flu activity: Spatio-Temporal Big data analysis," *Information Systems Frontiers*, vol. 21, no. 4, pp. 743–760, Jan. 2019, doi: 10.1007/s10796-018-9893-0.
[23] A. Z. Klein, A. Magge, K. O'Connor, J. I. F. Amaro, D. Weissenbacher, and G. G. Hernandez, "Toward using Twitter for tracking COVID-19: a natural language processing pipeline and exploratory data set," *Journal of Medical Internet Research*, vol. 23, no. 1, p. e25314, Jan. 2021, doi: 10.2196/25314.
[24] G. Livraga, A. Motta, and M. Viviani, *Assessing User Privacy on Social Media: The Twitter Case Study*. 2022, pp. 1–9. doi: 10.1145/3524010.3539502.
[25] Leon, "Data Warehouse Building: A comprehensive Overview," *BI4Dynamics - Business Intelligence for Microsoft Dynamics*, Jan. 06, 2025. https://www.bi4dynamics.com/data-warehouse-building-a-comprehensive-overview/
[26] E. Blomqvist, "The use of Semantic Web technologies for decision support – a survey," *Semantic Web*, vol. 5, no. 3, pp. 177–201, Jan. 2014, doi: 10.3233/sw-2012-0084.
[27] S. Moon and D. Iacobucci, "Social Media Analytics and its applications in marketing," *Foundations and Trends® in Marketing*, vol. 15, no. 4, pp. 213–292, Mar. 2022, doi: 10.1561/1700000073.
[28] E. French and B. T. McInnes, "An overview of biomedical entity linking throughout the years," *Journal of Biomedical Informatics*, vol. 137, p. 104252, Dec. 2022, doi: 10.1016/j.jbi.2022.104252.
[29] K. Mouyassir, M. Hanine, and H. Ouahmane, "Business Intelligence Model to analyze Social Media through Big Data analytics," *SHS Web of Conferences*, vol. 119, p. 07006, Jan. 2021, doi: 10.1051/shsconf/202111907006.
[30] B. Alabduallah, A. Banga, N. Iqbal, A. Ikram, and H. Diab, "Advancing cryptographic security with a new Delannoy-Derived Chaotic S-Box," *IEEE Access*, vol. 12, pp. 82926–82937, Jan. 2024, doi: 10.1109/access.2024.3410668.
[31] A. Mahajan, B. Kaushik, M. K. I. Rahmani, and A. S. Banga, "A hybrid feature selection and ensemble stacked learning models on Multi-Variant CVD datasets for effective classification," *IEEE Access*, vol. 12, pp. 87023–87038, Jan. 2024, doi: 10.1109/access.2024.3412077.
[32] E. F. Siddiqui, M. Haleem, S. F. Ahmad, A. Salhi, A. T. Zamani, and N. Varish, "A Multi-Layered AI-Driven cybersecurity architecture: integrating entropy analytics, fuzzy reasoning, game theory, and Multi-Agent reinforcement learning for adaptive threat defense," *IEEE Access*, vol. 13, pp. 170235–170257, Jan. 2025, doi: 10.1109/access.2025.3610526.