



REVIEW

Social Media-Based Surveillance Systems for Health Informatics Using Machine and Deep Learning Techniques: A Comprehensive Review and Open Challenges

Samina Amin¹, Muhammad Ali Zeb¹, Hani Alshahrani^{2,*}, Mohammed Hamdi²,
Mohammad Alsulami² and Asadullah Shaikh³

¹Institute of Computing, Kohat University of Science and Technology, Kohat, 26000, Pakistan

²Department of Computer Science, College of Computer Science and Information Systems, Najran University, Najran, 61441, Saudi Arabia

³Department of Information System, College of Computer Science and Information Systems, Najran University, Najran, 61441, Saudi Arabia

*Corresponding Author: Hani Alshahrani. Email: hmalshahrani@nu.edu.sa

Received: 16 July 2023 Accepted: 14 November 2023 Published: 29 January 2024

ABSTRACT

Social media (SM) based surveillance systems, combined with machine learning (ML) and deep learning (DL) techniques, have shown potential for early detection of epidemic outbreaks. This review discusses the current state of SM-based surveillance methods for early epidemic outbreaks and the role of ML and DL in enhancing their performance. Since, every year, a large amount of data related to epidemic outbreaks, particularly Twitter data is generated by SM. This paper outlines the theme of SM analysis for tracking health-related issues and detecting epidemic outbreaks in SM, along with the ML and DL techniques that have been configured for the detection of epidemic outbreaks. DL has emerged as a promising ML technique that adapts multiple layers of representations or features of the data and yields state-of-the-art extrapolation results. In recent years, along with the success of ML and DL in many other application domains, both ML and DL are also popularly used in SM analysis. This paper aims to provide an overview of epidemic outbreaks in SM and then outlines a comprehensive analysis of ML and DL approaches and their existing applications in SM analysis. Finally, this review serves the purpose of offering suggestions, ideas, and proposals, along with highlighting the ongoing challenges in the field of early outbreak detection that still need to be addressed.

KEYWORDS

Social media; epidemic; machine learning; deep learning; health informatics; pandemic

Nomenclature

<i>SM</i>	Social media
<i>CDC</i>	Centers for disease control
<i>DL</i>	Deep learning
<i>SVM</i>	Support vector machine



<i>GRU</i>	Gated recurrent unit
<i>LR</i>	Logistic regression
<i>RF</i>	Random forest
<i>SARS</i>	Severe acute respiratory syndrome
<i>ILI</i>	Influenza-like-illness
<i>NLP</i>	Natural language processing
<i>MAE</i>	Mean absolute error
<i>LiR</i>	Linear regression
<i>CNN</i>	Convolutional neural networks
<i>CBOW</i>	Continuous bag of word
<i>LSTM</i>	Long-short term memory
<i>MSE</i>	Mean square error
<i>TF-IDF</i>	Term frequency inverse document frequency
<i>ANN</i>	Artificial neural networks
<i>NB</i>	Naïve bayes
<i>MLP</i>	Multi-layer perceptron
<i>DT</i>	Decision tree
<i>WHO</i>	World health organization
<i>ML</i>	Machine learning
<i>GFT</i>	Google flu trends
<i>RMSE</i>	Root mean square error
<i>KNN</i>	K-nearest neighbors
<i>OCR</i>	Optical character recognition
<i>NER</i>	Named entity recognition
<i>LLM</i>	Large language models

1 Introduction

In recent years, Web 2.0, SM, and the news media have been extensively utilized to clarify trends in epidemic outbreak initiation and prevalence. Fig. 1 presents widely used SM platforms that have attracted growing user engagement over the past few decades, drawing inspiration from a similar figure designed in [1]. The immense popularity and proliferation of SM have gained social interaction among the public, thus generating a huge volume of information regarding any topic, such as political campaigns, sports, education, and products, etc. Similarly, SM also delivers a massive amount of information regarding epidemic outbreaks, if the outbreaks speedily rise in a region [2–4]. An SM provides a unique way to explore and understand social collaboration and interaction between the public and healthcare responders, now more than ever before. In recent years, SM has attracted extensive interest as a conceivable mechanism to detect and monitor epidemic outbreaks in a region because it can deliver real-time monitoring systems at a lower cost as compared to traditional monitoring systems [5–7].

An epidemic occurs when a viral disease spreads widely and emerges from natural reservoirs to infect people [8,9]. Moreover, it is also used to define active crises that are out of control; for instance, the severe acute respiratory syndrome (SARS) that took the lives of about 9000 people around the world in 2003. Similarly, a flu epidemic occurred in 2011, and a COVID-19 (Corona Virus Disease–19) epidemic outbreak is currently underway. An outbreak is a sudden spike in infectious disease that happens in a society or geographic region, or it may affect many countries and thus can last for a couple of months or even for several years [8]. Each year, some outbreaks are expected, like dengue or

influenza/flu. An outbreak can often be considered a single incidence of a viral or infectious disease. It may be valid if the infectious agent is uncommon (such as COVID-19) or if it has significant public health implications (e.g., bioterrorism viruses such as smallpox or Ebola virus).



Figure 1: Common SM applications

Now, in the era of data, the utilization of SM is common for sharing news, events, daily life activities, or even emotions to express. SM has also played a significant role in real-time analysis and more rapid forecasting has been considered in many areas. This includes disaster prediction [10–13], fake information detection [14], sports activities [15], political campaigns [16,17], sentiment analysis [18], communication [19,20], sarcasm detection [21,22], stock market fluctuations [23], and health surveillance systems [24–26].

In health surveillance systems, SM offers effective resources for epidemic outbreak detection and an active way of coping against the outbreaks [3,27,28]. Numerous studies [29–31] have configured search engines or search queries to develop a method for tracking an epidemic outbreak in a region. Bhattacharya et al. [32] proposed a model for disease surveillance by utilizing SM and developed a belief surveillance mechanism for health promotion. In health care, this type of monitoring is deployed to evaluate a user's level of confidence in the dissemination of health-related information in SM.

1.1 Social Media-Based Surveillance Systems in Health Informatics during Epidemic Outbreaks

SM-based surveillance methods rely on the analysis of publicly available SM data to identify signals of disease outbreaks. These methods can provide real-time information on the spread of diseases and help public health officials to respond quickly and effectively to outbreaks [33–35]. Early detection of seasonal epidemic outbreaks can decrease their influence on daily lives. Therefore, early detection and surveillance systems are important for tracking and attempting to reduce the impact of epidemic outbreaks that become uncontrollable by a prompt reaction. The epidemic outbreak has been the 21st century's most deadly infectious disease. Epidemic outbreaks are infectious diseases that can circulate across the nation as well as across the world if the pandemic assessment hits the extent of

an epidemic and tends to wipe out the whole nation [36,37]. The most searched/explored SM sources for data gathering by [35] are depicted in Fig. 2. Following Fig. 2, it can be observed that Twitter is the most popular SM source for health-related data gathering, configuring Twitter (64%), internet search queries/Google trends/Wikidepedia (15%), Crowdsourcing (4%), Instagram (3%), YouTube (2%), News articles (1%), SM search (1%), and other microblogs (6%) [35].

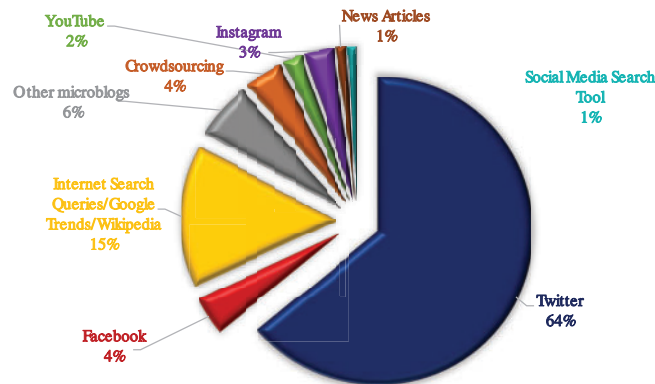


Figure 2: Types of SM platforms explored for health-related data collection

The sudden uptick in world travel and the integrated existence of contemporary civilization has contributed to growing attention to both existing and newly emerging outbreaks of threats. Public health officially requires timely and reliable reports about epidemic outbreaks, intending to take action and detect early warnings [38]. Traditional systems of epidemic outbreak surveillance are primarily designed based on manually compiled virology and medical studies. The traditional outbreak tracking methods of notification from physicians may take days or weeks to compile and deliver, so identifying more rapidly available sources of information is an actual priority. Some prominent outbreaks have happened in the world, such as dengue, influenza/flu, yellow fever, cholera, COVID-19, and several others [39–41].

Several review studies have explored the applications of ML and DL methods in the area of SM-based surveillance systems from different perspectives. For instance, Lamba et al. [42] conducted a review using ML for medical informatic. Al-Garadi et al. [43] explored the prediction of cyberbullying on SM using ML approaches. Riswantini et al. [44] and Gupta et al. [35] conducted a comprehensive review of handling disease outbreaks on SM using ML techniques. In addition, some other reviews were conducted using DL approaches [45–47].

SM-based surveillance techniques have shown great potential for the health sector, especially when combined with the ML and DL methods. According to our knowledge, the existing approaches for detecting or predicting epidemic outbreaks in SM were designed to detect influenza and dengue outbreaks, including seasonal dengue fever, chikungunya, Ebola virus, and influenza or swine flu. This article focuses on reviewing the current methods, strategies, architecture, and framework for the prediction, detection, or classification trends of epidemic outbreaks in SM information. The investigated approaches analyze SM and most of them use Twitter data that has keywords related to a specific epidemic outbreak for quicker identification in an initiative aimed at attaining and promoting public health. In this review, we will also discuss the current state of SM-based surveillance techniques, their applications, and future research directions.

To the best of our knowledge, the novelty of this review lies in its comprehensive exploration of the utilization of both ML and DL techniques in SM-based surveillance systems for health informatics. While there are existing reviews that focus on either ML or DL separately [42,43,45,46]. This review [48] provides a comprehensive analysis of the combined use of these techniques in the context of health-related surveillance on SM platforms.

In addition, this review addresses the specific challenges and open issues that are unique to SM-based health surveillance systems and still need to be addressed for health informatics. It highlights the ethical considerations related to data privacy and the difficulty of distinguishing between reliable health information and misinformation on SM. These discussions provide valuable insights for researchers and practitioners seeking to implement such surveillance systems effectively.

1.2 Research Motivation

The prime motivation behind this analysis is based on the following perspectives:

1. The use of SM has enabled faster monitoring of outbreak patterns compared to traditional data collection methods, facilitating real-time analysis. SM data is a valuable source of information that can aid in tracking epidemic outbreaks.
2. Detection of epidemic outbreaks from SM is a challenging problem, and it is still in its initial phase of enhancement, which needs further exploration. However, it is required to investigate research methods to improve current ML and DL techniques for the detection of epidemic outbreaks in SM.
3. A recent innovation in research conducted on the detection of epidemic outbreaks has encouraged us to conduct a systematic review to investigate, outline, summarize, and assess appropriate research studies.
4. Contribute to helping health organizations in detecting the spread of epidemic outbreaks by extracting information from SM in real-time.

1.3 Research Contribution

The primary contributions of this study are as follows:

1. Classifying epidemic outbreaks into dengue, flu/influenza, Ebola virus, Zika, and COVID-19 for SM-based surveillance system.
2. Analyzing the position of common types of epidemics outbreaks in SM: dengue, flu/influenza, Ebola virus, Zika, and COVID-19 in decision making.
3. Providing an overview of current ML techniques developed for the detection of epidemic outbreaks in SM surveillance systems.
4. Presenting a summary of various DL techniques that can be considered for epidemic outbreak detection in SM surveillance systems.
5. Providing a summary of various feature extraction techniques for better disease classification and detection used for SM text data.
6. Discussing the various techniques used for SM-based surveillance, their applications, and the open challenges that still need to be addressed.
7. This review makes a valuable contribution by exploring new learning models for SM analysis and identifying potential applications of NLP and DL.

The paper is structured into different sections as shown in Fig. 3. Section 2 demonstrates a background study of ML and DL. Section 3 provides a discussion on review methodology consisting of the most relevant works related to epidemic outbreaks and surveillance systems for health informatics in SM. Section 4 provides an overview of the assessment and discussion of research questions that utilize ML and DL approaches using SM platforms and concludes with a summary of the research gaps identified in the literature review and a comparison of the existing proposed solutions. Section 5 presents research implications and Section 6 illustrates open research challenges and future research directions. Finally, Section 7 concludes the review.



Figure 3: Structure of the paper

2 Background

2.1 Machine Learning

ML is a broader domain that involves the study of algorithms and statistical models that enable computer systems to learn from and make predictions or decisions based on data. It is a software technique that utilizes knowledge (experience) to identify or predict patterns within a given dataset [42]. Moreover, ML models discover patterns in the data and gain experience, which helps them perform better over time. In supervised ML, the primary field of study is classification, which aims to determine the appropriate category or class for a particular entity based on its features or parameters. Supervised classifiers utilize methods that capture the interactions between data and present several ML challenges. It entails using massive datasets to train algorithms, which then use these relationships and patterns to predict or decide on fresh data. Some common types of ML include random forest (RF), decision tree (DT), support vector machine (SVM), logistic regression (LR), etc. [42].

Fig. 4 shows the structure of ML-based models, drawing inspiration from a similar figure published in [49]. There are numerous domains and fields where ML can be applied such as speech recognition [50–52], NLP [53,54], sentiment analysis [55,56], and health informatics [57–60], etc.



Figure 4: Structure of ML-based techniques

2.2 Deep Learning

DL is a subset of ML that employs multi-layer neural networks to enable more complex computations [45]. DL models possess a greater capacity to identify relevant information, but their effectiveness also depends on the quality of the data. If the data is well-structured, the DL model will have an

easier time analyzing it. Some common types of neural networks in DL include convolutional neural networks (CNNs), recurrent neural networks (RNNs), long short-term memory networks (LSTMs), gated recurrent units (GRUs), Transformers, etc. [61].

It is motivated by the structure and operation of the human brain and is effective at many different tasks, including speech recognition, computer vision [62–65] health informatics [66–69], speech recognition [70–72], and NLP [73–75]. Fig. 5 shows a structure of DL-based techniques for NLP, drawing inspiration from a similar figure developed in [49].

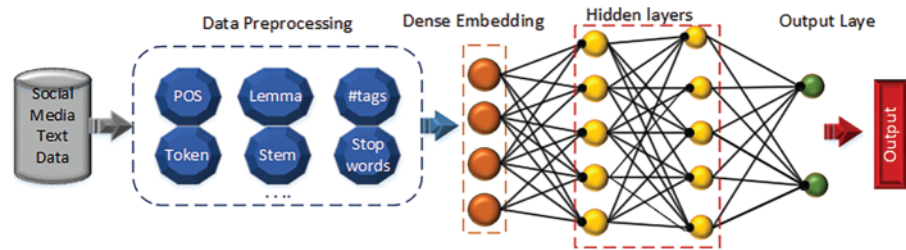


Figure 5: Structure of DL-based techniques

3 Review Methodology

The methodology adopted in this review is presented as follows.

3.1 Review Protocol

In this review, various electronic repositories are explored to search for relevant articles. A proper selection and exclusion strategy is applied to filter the number of retrieved articles. The final selection is based on specific research questions designed for this study, and after a comprehensive analysis, research gaps are reported.

3.2 Research Questions

The following are the primary research questions that will be addressed in this review:

RQ1: What are the common types of epidemic outbreaks reported in SM and their role in information gathering as recognized in the published articles?

RQ2: What are the various ML and DL techniques utilized for the detection and classification of epidemic outbreaks in SM acknowledged in the literature review?

RQ3: Why is DL essential for epidemic outbreak detection and classification in SM, and what are the existing approaches for epidemic outbreak detection and classification with a DL perspective?

RQ4: What are the different feature extraction techniques used to keep the syntactic and semantic relationships among words in SM texts for better detection, which will help to discover the research gaps?

RQ5: Are the SM platforms efficient in the perspective of raising awareness about outbreaks and promoting public health by providing early warnings?

3.3 Scientific Data Sources

This study aims to explore the above-mentioned published works in the past recent years that configure SM data to detect and classify epidemic outbreaks. To identify relevant articles; the scientific libraries explored are Springer Link (<https://link.springer.com/>), ACM digital library (<http://www.acm>.

org/dl), MDPI (<https://www.mdpi.com/>), Google Scholar (<https://scholar.google.com.pk/>), PubMed (<https://pubmed.ncbi.nlm.nih.gov/>), IEEE Xplore (<https://ieeexplore.ieee.org/>), ScienceDirect (<http://www.sciencedirect.com/>), Elsevier (<https://www.elsevier.com/>), PLOS ONE (<https://journals.plos.org/plosone/>) and other. Fig. 6 reflects the distributions of the scientific data sources to conference proceedings, journal articles, and arXiv preprints.

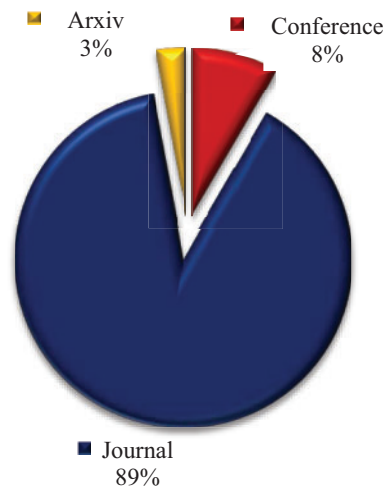


Figure 6: The percentage of publications from various categories (journal articles, conference proceedings, and arXiv)

In the following step, a selection and exclusion search strategy is applied for selection and exclusion to select the most relevant articles.

3.4 Screening Procedure for Article Selection and Exclusion

This section presents the procedure for article selection and exclusion. A systematic search approach based on targeted keywords has been used to collect the most relevant articles. This involves formulating specific search queries, as outlined in Table 1, in order to identify and retrieve relevant literature. The purpose of this review includes research works conducted on SM-based epidemic outbreak detection and classification in real-time using ML and DL techniques. The selection and exclusion strategy, as shown in Fig. 7 and inspired by a similar flowchart published in [76], is used to determine whether an article should be included or excluded. Based on the search strategy, Fig. 8 depicts the number of promising potential articles.

The following key factors form the basis of the selection strategy:

- The respective search queries are related to the work
- Articles relevant to epidemic outbreak detection and classification in SM
- Analyzing SM data for the detection and classification of epidemic outbreaks
- Articles are written in the English language
- Exploring an association between the title of the research article, the keywords designed for this study

Table 1: List of keywords adopted for querying relevant articles

ML+ Flu/influenza, ebola	ML+ Dengue, ML Zika virus	ML + COVID-19	DL Ebola+ Flu/influenza	DL+ Dengue, Zika virus	DL + COVID-19
Flu/influenza detection in SM	Dengue detection in SM	COVID-19 identification in social networks	Epidemic outbreak detection in SM	Detection and classification of dengue in tweets	Early detection of epidemic outbreaks in SM

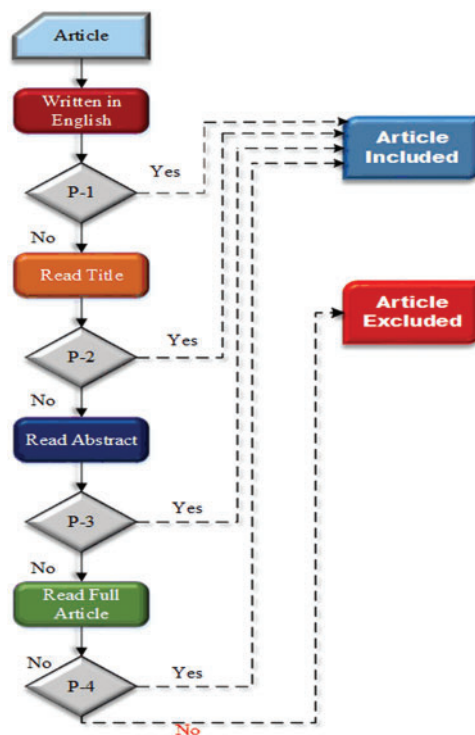


Figure 7: Flowchart: Search strategy for article selection and exclusion

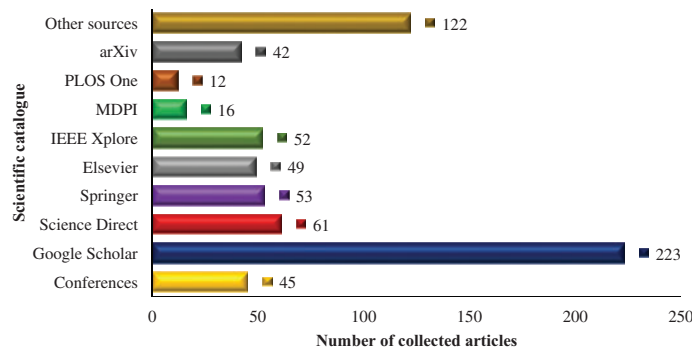


Figure 8: Scientific data sources with numbers of potential articles

A total of 712 articles are found in scientific database searches. Out of 712 articles, 140 are chosen for inclusion in the analysis, 213 duplicates are excluded, 218 articles are removed based on the abstract, and after reading the complete article, 141 are removed by using the specifications depicted in Fig. 9. The selection procedure is focused on various phases: the pre-phase is based on searching for relevant articles from the relevant data sources. Phase-1 is focused on a title-based search to pick appropriate articles. Phase-2 is focused on an abstract-based search to exclude the primary article. Phase-3 relies on a keyword-based search, while Phase-4 employs a comprehensive text-based search strategy, as depicted in Fig. 9, inspired by the PRISMA flow diagram. In addition, Fig. 10 depicts the distribution of publications related to ML and DL over the years, spanning from 2009 to May 2023, within the context of SM-based surveillance systems for health informatics. This visual representation showcases the trends in research output and the evolution of ML and DL applications in the field of healthcare-related surveillance.

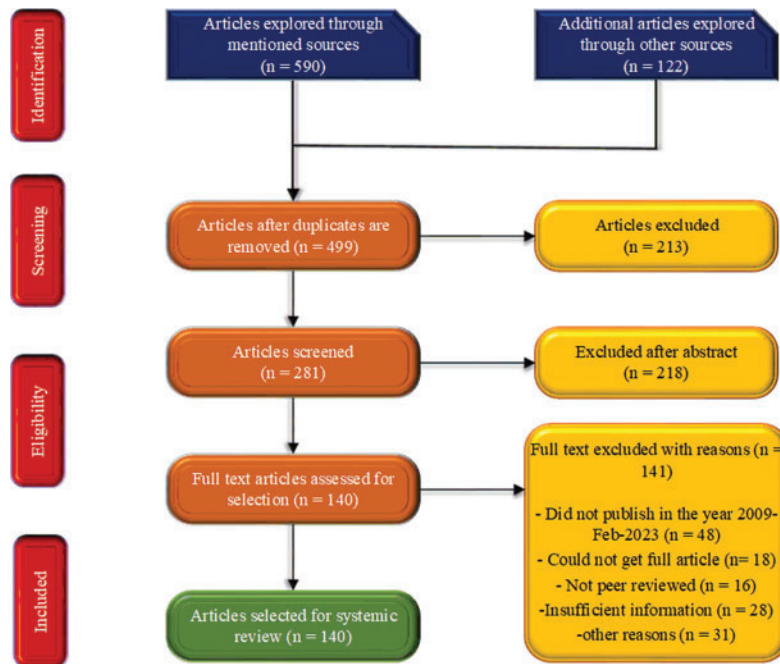


Figure 9: Methodological flow diagram-number of articles reviewed at each phase from primary search to the final number of the selected articles

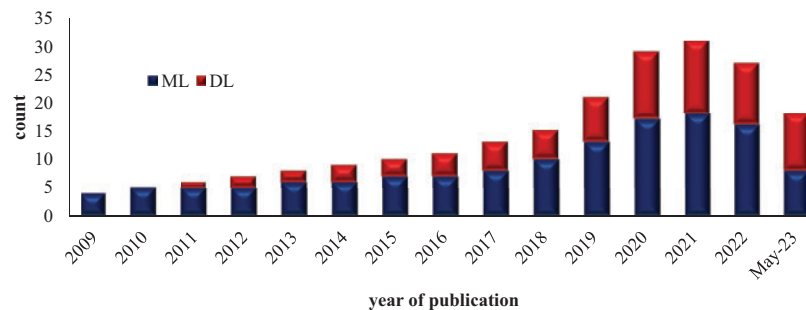


Figure 10: Distribution of ML and DL-based publications for surveillance systems in SM per year (until May 2023)

4 Assessment and Discussion on Research Questions

The research questions posed in this study are thoroughly examined within the context of analyzing ML and DL approaches for the detection of epidemic outbreaks and the provision of early warnings. The answers, to these research questions, are as follows:

RQ1: What are the common types of epidemic outbreaks reported in SM and their role in information gathering as recognized in the published articles?

4.1 Surveillance Systems for Epidemic Outbreaks in Social Media

SM-based surveillance methods, combined with ML and DL techniques, have shown potential for early detection of epidemic outbreaks. This review will discuss the current state of SM-based surveillance methods for early epidemic outbreaks and the role of ML and DL in enhancing their performance. ML and DL techniques can enhance the performance of SM-based surveillance methods by enabling the automated detection of patterns and anomalies in large volumes of data. ML and DL algorithms can be trained on historical data to detect early signals of epidemic outbreaks and to predict the spread of diseases in real-time.

During spontaneous epidemic outbreaks, the public requires access to reliable and timely information on the incidence of the epidemic and its prevention [40]. Many studies have been conducted in the field of SM analysis for detecting public sentiment about epidemic outbreaks [77]. The nature of epidemic outbreaks is characterized by their dynamic and constantly changing temporal and spatial aspects, which can effectively be identified using SM data. This approach can be used to detect diseases such as flu, dengue, and COVID-19, and can aid in promoting public health by detecting patterns and providing early warnings about disease outbreaks [4,78]. This method is much faster than traditional reporting methods where physicians and healthcare professionals report cases of disease to local health centers, which can take days or even weeks before related healthcare professionals/organizations can react and provide resources to control the epidemic [79]. Unfortunately, this delay can result in precious lives being lost before the necessary action can be taken. By utilizing SM analytics, this research aims to decrease the time between the onset and detection of disease outbreaks, enabling a faster response to control the epidemic.

The main concern is public health, and healthcare professionals must stay informed about epidemic outbreaks and diseases that affect their communities in order to make timely and appropriate decisions [80]. In recent years, SM particularly Tweets data have been utilized to have a positive impact on: disease identification such as to predict the present condition of a disease the traditional medically

reported data, and SM data recently reported by the public [81,82], epidemic outbreaks detection [83,25] and estimating the probability of people falling sick [38]. In addition, news media and web blogs have also been utilized to deliver early warnings of augmented disease progression before officially reported data [84,85], as well as for assuring the major transitions in the generative and fertility rate of the epidemic outbreak [85,86].

4.2 Common Types of Epidemic Outbreaks in SM and Their Role in Information Gathering

This section presents the common types of epidemic outbreaks in SM. The review encompasses the following disease outbreaks, as presented in Fig. 11: Influenza/Flu, Dengue fever, COVID-19, and Ebola virus. Each of these diseases is explored to provide a comprehensive understanding of how SM-based approaches with ML and DL techniques have been developed to monitor and detect outbreaks associated with these specific diseases.

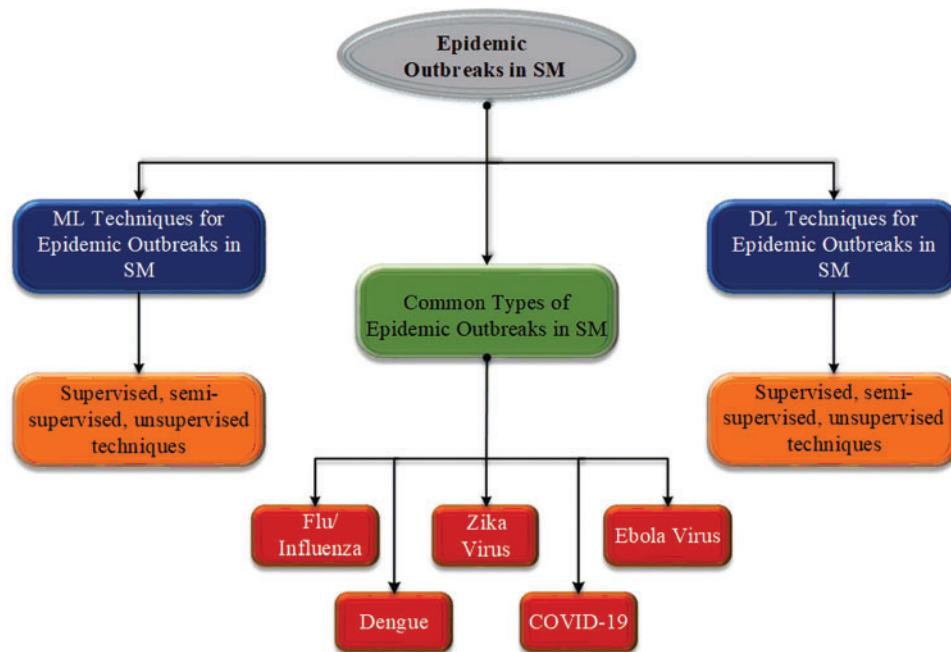


Figure 11: Overview of ML and DL classification for epidemic outbreaks in SM

4.2.1 Flu/Influenza Virus

Flu is a viral infectious outbreak that affects the lungs, respiratory system, and nose of a human being. Flu is also known as influenza; however, it is not similar to the flu viruses of the stomach can influence nausea and vomiting while influenza recovers by itself for most citizens. Yet influenza and its symptoms can also be devastating. Influenza, a respiratory disease, causes a significant number of deaths worldwide each year. The symptoms of flu are usually mild, such as headache, sneezing, fever, sore throat, and coughing [87]. Influenza shots are typically administered during the winter season, and infected individuals are advised to seek medical attention from specialists rather than general practitioners. Failure to treat the flu can lead to serious complications and worsen the patient's condition [88].

It is worth noting that both Wang et al. [89] and Chen et al. [90] utilized DL approaches in their proposed models for epidemic prediction using SM data. Wang et al. employed a partial differential equation (PDE)-based model for influenza prediction, while Chen et al. used two temporal topic models (supervised and unsupervised) to estimate trends in flu outbreaks in South American countries. These studies highlight the potential of DL approaches in accurately predicting and tracking epidemic outbreaks using SM data and suggest that these approaches could be extended to other types of epidemics in the future.

Ginsberg et al. [29] developed an automated approach for exploring search queries associated with influenza. The proposed approach produces more robust approaches for influenza-like-illness surveillance (ILIS), with regional and state-level measures of ILI occurrence in the US, by analyzing the search queries of the public from five years of Google web search log. The growing need global for online search can potentially allow for the model's development in international contexts. They compared the prediction of their proposed approach with the weekly ILI level delivered by Utah State to assess the performance of the state-level prediction produced by the proposed approach. The model was trained in traditional techniques, there is a need to explore the outbreak in depth by deploying advanced approaches.

The increased usage of online social network sites like Twitter's diverse demographic population [91]. For real-time analysis, the SM data generates an effective resource. The main contribution of the study was to propose a framework that utilizes SM data to track an epidemic outbreak in a region and deliver early warnings, even for new outbreaks reliably. The model was evaluated on regression approaches using tweet data. The linear regression (LiR) performed better compared with the Pearson correlation. The limitation and for future work, it was mentioned, there is a need for manual annotation to train the model for the entire SM data.

Carneiro et al. [2] suggested that Google flu trends (GFT) can be utilized for influenza outbreak detection before traditional monitoring mechanisms such as the Centers for disease control and prevention (CDC). In their work, the Google Trends tools, and data processing are described as well as provided a detailed example of how to detect influenza outbreaks in SM. Google search queries-based dataset was utilized for influenza detection on GFT. They suggested that GFT can be used as a blueprint for infectious epidemic outbreaks to provide an early monitoring system. For future direction, however, there is still room to enhance the proposed system for infectious disease outbreaks by utilizing DL approaches.

In [92], ML techniques were used to detect disease outbreaks based on the frequency of mentions of two diseases, flu, and cancer, on Twitter. The study aimed to monitor the spread of epidemics in different US regions by analyzing the number of mentions in each state. The location information was extracted from user timelines to perform the geographic analysis. The novelty of the study was to provide real-time surveillance of disease outbreaks in the US states, but it did not consider important features such as infected people, sentiment about the disease, and alarming situations.

4.2.2 *Ebola Virus*

Several studies have demonstrated the effectiveness of SM-based surveillance methods for early epidemic outbreak detection. For example, a study conducted during the 2014 Ebola outbreak in West Africa showed that Twitter data could be used to detect and track the spread of the disease in real-time [93–95]. For instance, Odlum et al. [96] identified public health information needs that can be accurately through content analysis of SM data, such as tweets, and then met with the right health information. By longitudinal tracking, the present study aimed to evaluate the demand for Ebola

health information at various pandemic durations. Guidry et al. [97] investigated the information and context of posts made on Twitter and Instagram about the Ebola epidemic by the CDC, the WHO, and Médecins Sans Frontières, with a concentrate on the communication methods that were employed during the epidemic. Even though all three health organizations used both forums, the findings point to Instagram as being especially useful for developing relevant, interactions with the publics during times of global health emergencies, as shown by the substantially greater levels of participation on the part of health agencies and the citizens.

Furthermore, using Twitter and news data, Kim et al. [98] proposed sentiment analysis and topic-based content Ebola pandemic with the help of the n-gram Latent Dirichlet Allocation (LDA) topic modeling techniques. Lazard et al. [99] developed a model for detecting the narrative of public concern regarding the Ebola virus from Twitter posts, and Van et al. [100] utilized tweets content to identify public concern regarding the Ebola virus for early warning during the pandemic outbreaks.

4.2.3 *Zika Virus*

Other studies have shown that ML/DL algorithms can improve the accuracy of SM-based surveillance methods for epidemic outbreak detection. For example, a study conducted during the 2016 Zika virus outbreak in Brazil showed that ML algorithms could accurately identify tweets related to the disease, even when they used non-specific terms [101,102].

In order to comprehend how a public health emergency of global importance manifests itself in SM, particularly Twitter, Pruss et al. [103] proposed the relevance of three different sorts of events: those that are location-related, actor-related, and concept-related. The work thereby adds to the body of knowledge about the processes that underlie participation, contributions, and engagement on this SM platform during a disease outbreak. They collected over 6 million tweets referring to the Zika pandemic from December 2015 to March 2016 during the pandemic. Using ML techniques, Daughton et al. [104] looked for concerns and self-disclosures of a particular behavior modification linked to the transmission of disease travel cancellation using tweets about the 2015–2016 Zika virus infection. If Twitter can identify this kind of activity, this method might offer a new source of data for illness modelling. To test the viability of using Twitter data for monitoring the ZIKV pandemic on a national and state (Florida) level, Masri et al. [105] used a recently created method called Cloudberry to filter a random sample of the data. To predict weekly ZIKV infections one week in advance, two auto-regressive models were calibrated using weekly ZIKV case numbers and Zika tweets.

From October 1st, 2015, to February 25th, 2016, Abouzahra et al. [106] gathered 67,000 tweets in both English and Spanish with the hashtags #Zikavirus and #Zika. We used text analytics methods to analyze the tweets and identify the key ideas. We examined the variations in how these ideas were used from one month to the next. Wood [107] developed an ML-based method for debunking and propagating conspiracy theories on Twitter during the 2015–2016 Zika pandemic. They collected around 25,162 tweets on the Zika virus from Twitter to disprove statements and hoaxes that circulate through SM.

4.2.4 *Dengue Outbreak*

Dengue is a contiguous viral epidemic outbreak that is transmitted by mosquitoes. Dengue fever is a viral infection that is spreading globally [108]. Accurate and timely data monitoring is essential to effectively detect dengue outbreaks and assess the impact of preventive interventions [109]. The characteristics of a dengue outbreak will vary from immunosuppressed fever to feared consequences such as viral infection and trauma. In tropical and subtropical regions, the dengue outbreak has

emerged as a challenge for public health promotion, while the dengue outbreak is typically self-limiting [110].

Amin et al. [111] designed an automated approach for the detection and classification of dengue outbreaks in tweet data. In their work, they deployed RNN-based method LSTM to efficiently process the flow of sequence data to classify the tweet messages into dengue positive and negative classes to detect the dengue-infected people. In their proposed work, a comparison was also made among ML including SVM, NB, LR, and DL such as ANN, DNN, and LSMT techniques to find the best approach for outbreak detection from SM data, and the performance of each model is evaluated on test data. They found that the LSTM is the best approach to detecting disease-infected people and analyzing the epidemic outbreak in SM compared to other state-of-art-techniques. For feature extraction, the term frequency-inverse documents frequency (TF-IDF) embedding technique was utilized. In this work, a novel benchmark dataset was designed for outbreak detection collected from 2017–2019. The future research direction mentioned in this work will be to configure word embedding techniques for better disease classification and detection.

Jain et al. [112] introduced an SM-based dengue surveillance and epidemic outbreak detection mechanism incorporating temporal and spatial patterns that help to recognize, classify, and design consumer behavioral trends on SM. Their proposed approach was based on geo-tagging predictive modeling has a major role in the deterrence and monitoring of mosquito-borne disease within limited-resource in a specific region. Tracking public opinions in real-time offers intuitive interfaces or early warnings related to outcomes. In this, LDA-based topic modeling approaches were developed to filter out a similar topic about deterrence, symptoms, and panic. For this purpose, ML classification techniques SVM and Naïve Bayes (NB) were developed. Future research is required to utilize other resources for data like news articles, web blogs, etc. Data that contains emoticons and text in an image may be considered in the future as well.

The increasing number of dengue cases in China has raised significant public health concerns, with the disease spreading to larger regions [113]. The main contribution of their work was to design a timely and accurate approach for dengue prediction in China using state-of-the-art ML techniques by using Baidu search queries and environmental conditions (relative humidity, mean temperature, and precipitation) data collected in the year 2011 to 2014 in Guangdong. To implement the model, they compared and evaluated support vector regression (SVR), least absolute shrinkage, generalized additive, and regression models including gradient boosted regression tree (GBRT), step down-linear regression, selection operator linear regression, root mean square error (RMSE), and negative binomial regression to predict dengue cases. In this work, the proposed SVR approach achieves better performance results to forecast and track dengue outbreaks in comparison with other baseline methods. The features of this work will be helpful for healthcare organizations to identify the initiative needed to improve dengue surveillance. For future direction, the proposed model can be improved using DL methods as well.

4.2.5 COVID-19 Outbreak

COVID-19 is a viral infectious outbreak transmitted by a recently found coronavirus [114–116]. It is extremely contagious, with many patients being able to move into hospitals for testing at the same time, which has significantly affected public healthcare systems. The priority of treatment is also dictated by the severity of the symptoms based on a diagnosis. Clinical experiments indicate that suspected people (people with mild symptoms) can deteriorate rapidly [115,117]. Hence, it is essential to detect early patients' deterioration in order to improve the treatment plan.

Information and news headlines on COVID-19 were quickly posted and shared on SM during the starting months of 2020. The information pattern has been analyzed in SM, and on the web with around 18 years in the infodemiology field, the COVID-19 outbreak has been referred to as the first SM infodemic. However, there is insufficient confirmation about whether and how the SM infodemic has triggered information on COVID-19 and provided early warnings. In several regions of the world, the explosive growth of COVID-19 has been reported [86,118].

Detecting informative tweets related to the COVID-19 outbreak is an important task for providing real-time updates to the public, identifying misinformation, and tracking the spread of the disease [119,120]. For instance, Samuel et al. [121] identified the sentiment of the public related to the COVID-19 outbreak, including its sentiment analysis by utilizing the tweets data on COVID-19. In their work, they used ML techniques NB and LR to identify public sentiment on COVID-19. After that, the effectiveness of the analysis is compared in classifying COVID-19 tweets in the United States. For limited tweet data, the NB approach achieved 91% accuracy while the LR approach achieved 74% accuracy for the shorter tweet. However, both approaches showed comparatively poor performance for longer tweets. There is a need to improve the performance of the ML techniques on long tweets. The performance can be improved with the help of DL techniques. Kabir et al. [122] presented a method that utilizes ML and topic modelling techniques to analyze user sentiment and public posts related to COVID-19 on SM.

Ardabili et al. [123] conducted an interesting study on predicting the COVID-19 outbreak by comparing the performance of ML and soft computing techniques. The authors used MLP and ANFIS models to predict the pandemic outbreak in five countries (Italy, Germany, Iran, USA, and China) by training the models on a dataset obtained from <https://www.worldometers.info/coronavirus/#countries>. The study found that ML techniques, particularly MLP, were effective in modeling the pandemic outbreak. However, future research should also focus on modeling the fatality rate to aid in planning new facilities in affected countries. Additionally, the use of DL techniques can aid in the detection of infected individuals.

Khanday et al. [124] built an effective approach for textual clinical data classification by utilizing ML approaches. The clinical textual data are classified into four classes with the help of classical and ensemble ML techniques. Furthermore, for feature extraction, they used TF-IDF, and Bag of Words (BOW), and for the classification, they used LR, Multinomial Naïve Bayes (MNB), SVM, and Decision Tree (DT). The data were classified into, i.e., COVID, SARS, COVID, and ARDS (Acute Respiratory Distress Syndrome). In the end, the comparative analysis among ML methods showed that the MNB technique outperformed with 96% accuracy the other techniques. For future work, there is a need to increase the accuracy using RNN as the traditional ML techniques are not able to efficiently work on the flow of sequence data. Prabhakar et al. [125] proposed topic modelling techniques for the detection of COVID-19 information from Twitter content.

Nowadays with the help of SM content, a lot of analysis and statistics can be done in case of epidemic outbreaks. Nemes et al. [126] presented an automated approach for predicting and manifesting public sentiment to check the correlation between labels and words of positive and negative sentiment in tweet messages. The analysis was performed with the help of NLP techniques and sentiment classification using the RNN approach. For further processing, they also analyze, visualize, and compile their exploration. The approach that was developed in this work performs accurately on small data even with vague tweet messages in assessing the sentiment polarity of the public regarding COVID-19.

4.3 Summary of Key Performance from Recent Literature

Reviewing the literature, it is possible to evaluate the performance of ML and DL approaches and sort of public tweet messages regarding various epidemic outbreaks on SM. A summary and detailed information on the key performance indicators for qualifying the selected articles using ML techniques are presented in Table 2. Furthermore, Table 3 presents the key performance indicators for qualifying the selected articles using DL techniques.

After reviewing the literature outlined in Tables 2 and 3, it can be concluded that the existing studies did not successfully identify disease-infected individuals from SM texts. There is currently no established methodology or procedure for identifying individuals with a disease from SM information. The literature mainly focuses on detecting the frequency of SM posts regarding a specific disease, rather than distinguishing between disease-infected people and non-disease-infected people in SM posts. To bridge the research gaps, a new approach utilizing DL approaches with word embedding techniques needs to be proposed.

Fig. 12 reveals the most discussed epidemic outbreaks in SM along with the research articles. It shows that the most common outbreak regarding research-related analysis is the flu/influenza, dengue and now it would be COVID-19.

RQ2: What are the various ML and DL techniques utilized for the detection and classification of epidemic outbreaks in SM acknowledged in the literature review?

This section distinguishes the research articles from ML and DL perspectives that are utilized for the detection and classification of epidemic outbreaks in SM acknowledged in the literature review. Fig. 13 reveals that the primary research studies have been conducted in ML configuring SVM (10%), MNB (4%), NB (9%), LR (14%), Linear Regression (10%), DT (10%), LDA (9%), RF (7%), and SVR (8%). Similarly, Fig. 14 shows DL techniques deploying ANNs (29%), DNN (21%), RNN (21%), LSTM (20%), GRU (12%), CNN (3%), and LSTM + CNN (2%<).

Table 2: Summary of the related work using ML techniques

Study	Year	Approach	Accuracy	Outbreak type	Data sources	Infected region	Major insights	Limitations/Future directions
Alessa et al. [91]	2019	LiR	F-measure 89.9%, strong correlation of 96.29%	Flu, ILI	Twitter, CDC	Not specified	<ul style="list-style-type: none"> To create a highly effective and precise method that could efficiently use SM data to monitor flu outbreaks, including new and novel epidemics. The results of this method were compared to the existing data from the Centers for CDC as the baseline, and they showed a strong correlation. 	To train the model for all of the SM data, manual annotation is required.
Wakamiya et al. [127]	2019	SVM, KNN, LiR, LR	Chinese subtask: Flu, Influenza Accuracy-81.0%, F-measure-88.0% English subtask: Accuracy-77.0%, F-measure-85.0% Japanese subtask: Accuracy-72.0%, F-measure-0.820	Flu, Influenza	Twitter	Not specified	<ul style="list-style-type: none"> To classify tweets into disease and non-disease symptoms tweets to track flu outbreak The goal of this study was to analyze the results of groups that participated in a Japanese subtask, an English subtask, and a Chinese subtask, as well as conversations, to understand the challenges that need to be addressed in the field of medical NLP. 	The future direction will be to apply DL approaches.
Wakamiya et al. [128]	2018	Regression, Statistical methods	Accuracy 70%	Swine Flu, H1N1	Twitter	USA	<ul style="list-style-type: none"> To analyze the public sentiment in the US regarding swine flu and H1N1 using regression and statistical approaches 	In the future, ANN will be configured for prediction to analyze tweet data.
Collier et al. [129]	2011	SVM, NB	Accuracy 0.86	Flu, ILI	Twitter	Not specified	<ul style="list-style-type: none"> Analyzing the use opinion in Flu and ILI tweets by using SVM and NB. 	The proposed method will be evaluated on DL techniques.

(Continued)

Table 2 (continued)

Study	Year	Approach	Accuracy	Outbreak type	Data sources	Infected region	Major insights	Limitations/Future directions
Albinati et al. [130]	2017	Regression model, SVM, NB, unigrams, bigrams	Accuracy 0.86	Dengue	Twitter	Brazil	<ul style="list-style-type: none"> The purpose of this study was to monitor the outbreak of flu and ILI in tweets. To achieve this goal, the researchers used supervised learning techniques to classify tweets into four groups based on self-reported protective behaviors and one group based on a self-reported diagnosis. 	Frequency might not be a good indicator to detect infected people in tweets.
Aramaki et al. [83]	2011	SVM	The proposed method shows a high correlation of 0.97	Flu, ILI	Twitter	Not specified	<ul style="list-style-type: none"> This study proposed a model for flu detection in tweets deploying SVM for classification and showed a strong correlation of 0.89 to the gold standard. 	The proposed method will be evaluated on publicly accessible data.
Espina et al. [131]	2017	SVM, Regression	Accuracy of 90.09% and a positive recall value of 77.42%	Dengue, Typhoid Fever	Twitter	Philippine	<ul style="list-style-type: none"> This study aims to improve the existing methods of monitoring epidemic outbreaks. To classify health-related tweets related to dengue cases and typhoid fever in the Philippines 	Future work can be evaluated on real-time data.
Amin et al. [132]		SVM, KNN, DT, RF	Accuracy: SVM 85.0%, KNN 85.0%, DT 84.0%, RF 87.0%	Dengue, Flu	Twitter	Worldwide	<ul style="list-style-type: none"> To design a model for the detection of early warnings of infectious diseases using ML techniques 	For future direction, the proposed model can be improved by DL methods.
Ginsberg et al. [29]	2009	LinR with 4-cross validation	correlation of 0.90	Influenza, ILI	Google search query, CDC	USA	<ul style="list-style-type: none"> To propose an automated method for discovering influenza-related search query need to explore the outbreak in depth. To estimate ILI at regional and state-level actively in the US 	The model was trained on traditional technique, there is a need to explore the outbreak in depth.

(Continued)

Table 2 (continued)

Study	Year	Approach	Accuracy	Outbreak type	Data sources	Infected region	Major insights	Limitations/Future directions
Jain et al. [112]	2018	LDA, SVM, NB	NB 65.6%, SVM 68.3%	Dengue, Chikungunya, Malaria	Twitter	India	<ul style="list-style-type: none"> To provide early outbreak management and surveillance for mosquito-borne viral disease To utilize spatial and temporal information that helps in the identification of user behavior in SM To visualize the users' opinions of SM content To track an infectious disease outbreak in Google Trends To provide real-time outbreak surveillance To design an accurate approach for dengue prediction in China using state-of-the-art ML 	<p>Future research is required to utilize other resources for data like news articles, web blogs, etc. Data that contains emoticons and text in an image may be considered in the future as well.</p>
Carneiro et al. [2]	2012	Supervised and unsupervised techniques	Accuracy is not measured quantitatively	Respiratory Syncytial, West Nile, Avian Influenza, Rash, Bird Flu	Google trends (web search queries)	USA	<ul style="list-style-type: none"> To track an infectious disease outbreak in Google Trends To provide real-time outbreak surveillance To design an accurate approach for dengue prediction in China using state-of-the-art ML 	There is still room to enhance the proposed system for infectious disease outbreaks.
Guo et al. [113]	2017	SVR, NB, RMSE	95% prediction intervals	Dengue	Baidu search queries	China	<ul style="list-style-type: none"> To design an accurate approach for dengue prediction in China using state-of-the-art ML 	For future direction, the proposed model will be improved using DL methods.
Hong et al. [133]	2018	SVM, NB and LR	Not specified	Flu, chickenpox, and measles	Twitter	Australia	<ul style="list-style-type: none"> Used a scalable, automated Cloud-based system for gathering, monitoring, and analyzing SM data to identify disease outbreaks. 	In future, the can be improved using DL methods.
Ahmad et al. [134]	2020	Content analysis method, SPSS software	Content analysis (R = 0.8701)	COVID-19	SM data	Kurdistan Region of Iraq	<ul style="list-style-type: none"> To determine the impact of SM on self-reported mental health To determine the extent of panic about COVID-19 They found a substantial positive statistical correlation between self-reported SM use and the spread of panic related to COVID-19 To discover the user sentiment and the posts shared by the public on COVID-19 in SM and to model public opinion using ML techniques. 	The future guideline is to design seminars for students and staff to determine how to browse and identify health information in the context of epidemic outbreaks.
Kabir et al. [122]	2020	ML, topic modelling	Accuracy is not measured quantitatively	COVID-19	Twitter	USA	<ul style="list-style-type: none"> To discover the user sentiment and the posts shared by the public on COVID-19 in SM and to model public opinion using ML techniques. 	Further experiments need to be done on COVID-19 using SM data.

(Continued)

Table 2 (continued)

Study	Year	Approach	Accuracy	Outbreak type	Data sources	Infected region	Major insights	Limitations/Future directions
Khanday et al. [124]	2020	LR, MNB, SVM, DT	Accuracy-96.2%	COVID-19	Clinical text data	Not specified	<ul style="list-style-type: none"> To classify textual clinical data into four classes with the help of ML techniques i.e., COVID, SARS, and COVID and ARDS 	For future work, there is a need to increase the accuracy using RNN.
Samuel et al. [121]	2020	NB, LR	Accuracy of NB 90%, and regression 74%	COVID-19	Twitter	USA	<ul style="list-style-type: none"> To find the public sentiment in tweets associated with the COVID-19 outbreak 	The performance can be improved with the help of DL techniques.
Punn et al. [135]	2020	Polynomial Regression, RMSE	Accuracy is not measured quantitatively	COVID-19	https://coronavirus.jhu.edu/	Not specified	<ul style="list-style-type: none"> To contribute to the welfare of the living society To develop ML and DL techniques to predict the future accessibility of COVID-19 around the world using real-time information 	In the future, the work can further be enhanced to utilize other ML and DL.
Sivanantham [136]	2021	ML techniques	Accuracy 92.80%, sensitivity 93.20%, specificity 90.42%, precision 95.38%, recall 90.24%, and F-measure 89.78%	COVID-19	Web	Worldwide	<ul style="list-style-type: none"> NLP gathers opinions, which are then categorized as favorable, neutral, and negative. By utilizing several ML and swarm intelligence optimization algorithms. 	In the future, the research work can further be boosted to leverage DL.

Note: All the abbreviations/acronyms are defined in within the paper.

Table 3: Summary of the related work using DL techniques

Study	Year	Approach	Accuracy	Outbreak type	Data sources	Infected region	Major insights	Limitations/Future directions
Du et al. [82]	2018	SVM, RF, KNN, NB, RNN, CNN	CNN-F1-score of 0.8642).	Measles	Twitter	Worldwide	<ul style="list-style-type: none"> Measles outbreak detection in tweets Timely understanding of public perception during the epidemic outbreak 	In the future, the model will be evaluated for influenza and Ebola outbreaks.
Amin et al. [41]	2020	SVM, NB, LR, ANN, DNN, LSTM	Accuracy 94%	Dengue	Twitter	Worldwide	<ul style="list-style-type: none"> The main novelty of this study was to develop an RNN/LSTM to efficiently classify dengue data to detect dengue-infected people in tweets. To design benchmark data for the dengue dataset collected from 2017–19 To propose a method using ANN and optimization techniques for the detection of dengue using tweets 	Future research direction will be to configure word embedding techniques for better classification.
Amin et al. [137]	2021	ANN		Dengue	Twitter	Worldwide	<ul style="list-style-type: none"> To propose a method using ANN and optimization techniques for the detection of dengue using tweets 	DL methods can be used to enhance the work further.
Chae et al. [36]	2018	DNN, LSTM	F-measure-97.0659, 88.7462	Chicknipox, Malaria	Internet search query	South Korea	<ul style="list-style-type: none"> To predict infectious disease outbreaks in SM data To minimize the gaps in reporting on current surveillance mechanisms To provide early monitoring mechanisms 	The data acquisition period was relatively short and had a limited set of parameters for DL. Model performance can be enhanced if the data size is extended.
Ardabili et al. [123]	2020	ML, MLP, ANFIS	Accuracy is not measured quantitatively	COVID-19	https://www.worldometers.info/coronavirus/country	Italy, Germany, Iran, the USA, and China	<ul style="list-style-type: none"> To predict COVID-19 by presenting a comparative analysis of ML and Soft Computing techniques 	For future studies, modeling the fatality rate is needed in preparing for new facilities for nations. Using DL techniques, the infected people can be detected.
Mailla et al. [120]	2021	Majority voting-ensemble DL	91.75% accuracy, 91.14% F1-score	COVID-19	Twitter	Not specified	<ul style="list-style-type: none"> To present an ensemble DL method for classifying Informative tweets To provide COVID-19 sentiment analysis 	The model performance can be enhanced with more datasets in the future.

(Continued)

Table 3 (continued)

Study	Year	Approach	Accuracy	Outbreak type	Data sources	Infected region	Major insights	Limitations/Future directions
Chew et al. [138]	2021	CBOW, TFIDE, DNN	0.00282 RMSE, 0.0021 MAE	COVID-19	Twitter	Not specified	<ul style="list-style-type: none"> To process COVID-19 time-series records and a sizable amount of COVID-19-related Twitter data at the same time to model the growth rate in the number of confirmed COVID-19 cases globally. 	In the future, the work can further be enhanced to utilize other ML and DL.
Alorini et al. [139]	2021	86% accuracy	LSTM-RNN, NLP	COVID-19	Twitter	Worldwide	<ul style="list-style-type: none"> Use NLP methods for opinion mining to glean both critical and supportive tweets about COVID-19. 	In the future, the accuracy can further be improved by applying other DL models.

Note: All the abbreviations/acronyms are defined within the paper.

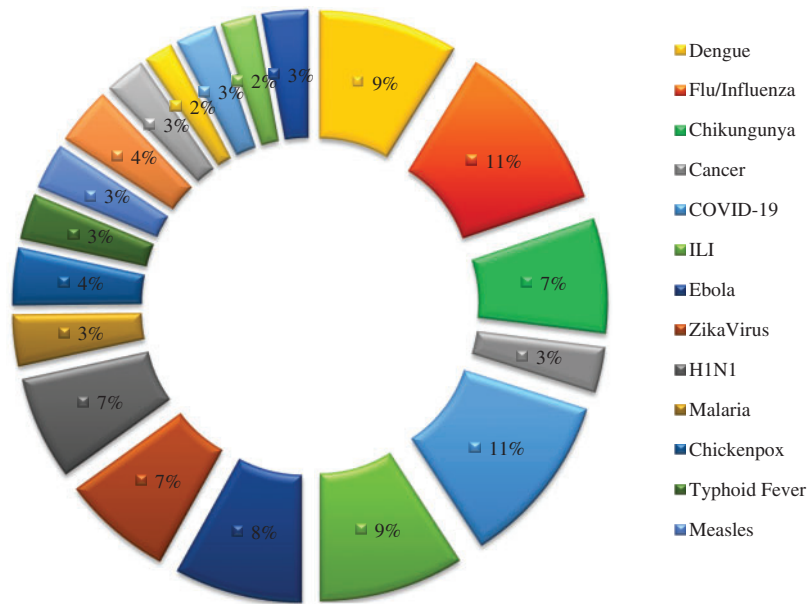


Figure 12: Breakdown of epidemic outbreaks reported in SM

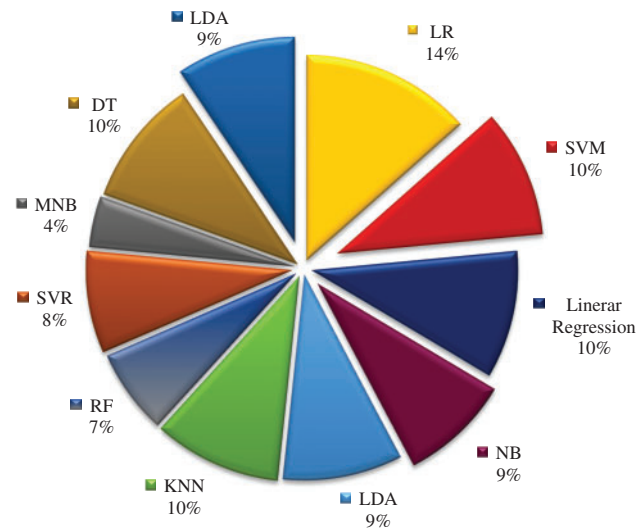


Figure 13: Breakdown of the articles using ML techniques

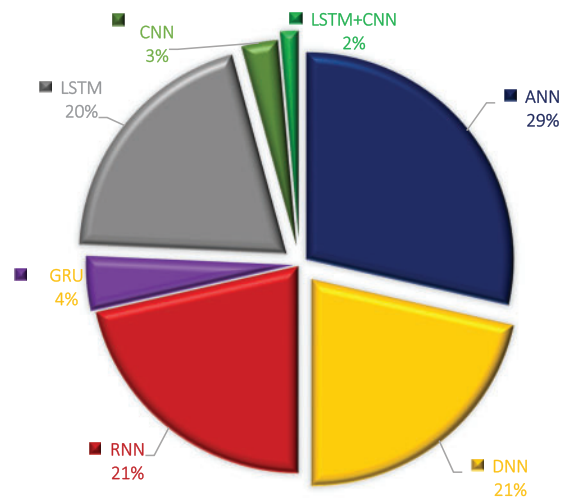


Figure 14: Breakdown of the articles using DL techniques

RQ3: *Why is DL essential for epidemic outbreak detection and classification in SM, and what are the existing approaches for epidemic outbreak detection and classification with a DL perspective?*

Answer 3: Based on our analysis, it is evident that there are significant research gaps in the area of SM analytics for detecting epidemic outbreaks, particularly in terms of real-time disease surveillance for early warning purposes. DL approaches have shown promising results in addressing these challenges. To bridge these gaps, this review proposes a new approach that leverages DL techniques such as RNN/LSTM, CNN, and CNN+LSTM with word embedding techniques. These approaches can be explored further in future research to address the identified research gaps.

RQ4: *What are the different feature extraction techniques used to keep the synthetic and semantic relationships among words in SM texts for better detection, which will help to discover the research gaps?*

Answer 4: After examining the early disease detection approach that utilizes SM information, it can be concluded that existing research models primarily focus on using SM platforms to detect various epidemics, such as seasonal dengue virus, depression, cancer, and flu outbreaks. While the literature uses the term “prediction and detection,” it primarily refers to identifying instances of influenza or swine flu that have already been observed. However, due to the limitations of the current SM monitoring system, new approaches are necessary to effectively detect and monitor epidemic outbreaks in SM.

Based on the existing literature on DL, it is evident that using word embedding techniques to analyze Twitter texts can help capture the semantic and synthetic meaning between words, thus improving classification accuracy. This thesis aims to address the shortcomings of previous studies rather than replace existing disease detection and monitoring systems. However, challenges still exist, such as the limited character count of tweet messages, the prevalence of abbreviations and informal words, grammatical and spelling errors, as well as instances of mixed language and inappropriate sentence structure.

RQ5: *Are the SM platforms efficient in the perspective of raising awareness about outbreaks and promoting public health by providing early warnings?*

Answer 5: Using SM platforms to spread awareness of outbreaks and provide early warnings is a successful strategy. SM can link people to resources and health specialists in real-time, as well as

provide updates and information regarding outbreaks. SM platforms can offer early warnings and outbreak awareness, making them an invaluable source of information during epidemic outbreaks. SM platforms can collect data from human sensors in the form of SM data, which can be analyzed to track and monitor an outbreak. This real-time analysis is faster than traditional methods of data collection and can be deployed to monitor various disease patterns. Additionally, the cost of collecting data through SM is lower than traditional methods [140]. The collaboration of NLP, ML, DL, healthcare analysts, and SM text analysis is proposed as an effective approach for detecting epidemic outbreaks in a region. The review highlights the importance of SM analysis as a valuable surveillance tool during epidemic outbreaks, including flu, dengue, zika, Ebola, and COVID-19. The study emphasizes that real-time information from tweets can alert healthcare professionals and emergency responders to take necessary actions in order to control or monitor an epidemic outbreak.

According to the reviewed studies, social media platforms can be very important for spreading knowledge of outbreaks and promoting public health. Many studies have shown that SM can be an effective and timely approach for identifying and tracking disease outbreaks. Likewise, SM platforms can help disseminate health information, encourage healthy habits, and offer a venue for community involvement and public health campaign participation. The overall results indicate that SM can be a successful tool for spreading health awareness and offering early warnings of outbreaks, despite certain restrictions and difficulties related to its use in public health.

Overall, online SM platforms have the potential to effectively disseminate early alerts and raise public awareness of health issues during epidemics. However, it is important to note that SM can also spread misinformation and rumors, which can be detrimental to public health efforts. As a result, during outbreak scenarios, it is critical to closely monitor and verify information published on SM.

5 Implications

The research holds significant scientific impact and is of great interest to researchers in the fields of health informatics, data science, and public health as follows:

1. Contributing to the advancement of health surveillance methodologies by exploring the potential of SM-based systems.
2. It bridges the gap between ML and DL methodologies, presenting how they can complement each other in the context of health informatics.
3. Addressing the ethical challenges surrounding data privacy and misinformation on SM.
4. Highlighting the open challenges and limitations, the research offers valuable direction for future research developments. Researchers can contribute to addressing these issues by pointing out potential research gaps.

In summary, the research's interdisciplinary approach, relevance to public health issues, and contribution to the field of SM-based health surveillance all contribute to its scientific impact. It makes a substantial and interesting addition to the scientific community and scholars by providing researchers with insightful information, ethical issues, and a roadmap for future study.

6 Open Research Challenges and Future Research Directions

SM-based surveillance techniques combined with ML/DL methods have the potential to revolutionize the health sector by providing real-time information on disease outbreaks and enabling more

effective public health responses. However, there are still several challenges that need to be addressed to fully grasp the potential of these techniques.

1. *Noisy data*: SM data is often noisy, incomplete, and unstructured, which makes it difficult to extract meaningful information. To improve the quality of SM data by reducing noise, filling data gaps, and standardizing data formats. DL models, especially those that include RNNs, LSTM, or transformers, can recognize rapid spikes or hidden patterns in SM data that are related to epidemic outbreaks and crisis occurrences.
2. *Privacy issue*: there are concerns about the use of personal information on SM, and the potential misuse of this information for surveillance purposes. Developing privacy-preserving methods regarding the use of personal information on SM should be addressed for data collection and analysis. Incorporating knowledge from similar tasks or disease transfer learning and multi-task learning in DL and NLP can aid methods in becoming broader.
3. *Data validation*: the data acquired from SM-based surveillance methods require effective validation and verification. To reduce bias in SM data by incorporating data from multiple sources and integrating demographic and geographic data into analyses. The semantic understanding of SM posts can be improved by DL techniques, particularly transformer-based models like BERT, which can comprehend and interpret the context of non-standard language.
4. *Data biasness*: the accuracy of surveillance methods may be impacted by SM data that is biased toward certain demographics or geographical regions. SM-based surveillance systems should be integrated with other data sources, including electronic health records and traditional surveillance methods to provide a more comprehensive model of disease outbreaks.
5. *Misinformation and rumors detection*: SM can also spread misinformation and rumors, which can be detrimental to public health efforts. Therefore, it is important to carefully monitor and verify information shared on SM during outbreak situations. The early detection of misinformation or rumors and concern about a disease, as the sharing of fake news and rumors, has increased with the widespread use of SM. By using advanced DL and NLP techniques, the issue of misinformation or rumors can potentially be addressed.
6. *Mental health detection from SM*: Using the potential of ML and DL techniques, depression, and anxiety can be detected from SM, which can be valuable for psychiatrists and mental health professionals.
7. *Pretrained large language models (LLM)* with contextualized information can be used to improve the performance of the traditional ML and DL models for disease surveillance.
8. Additionally, *optical character recognition (OCR)* can be applied to extract textual data from screenshots of social media posts shared across different platforms.
9. Furthermore, *named entity recognition (NER)* can be used to automatically extract disease-related information from SM texts. These methods can be used to extract names of diseases, medicines, vaccinations, and other related information to develop contextually aware models for disease surveillance through SM.

In conclusion, future research developments should focus on improving the quality of SM data, addressing privacy concerns, reducing bias, developing explainable AI, NLP, pre-trained large language model, ML, and DL methods, and integrating SM-based surveillance with other data sources. The above research directions show that the challenges can be effectively tackled, and innovative techniques can be proposed and integrated into the health informatic systems.

7 Conclusions

The use of online SM platforms is effective in providing early warnings and raising awareness of outbreaks. SM can provide real-time updates and information on outbreaks, as well as connect individuals to health professionals and resources. SM-based surveillance systems, combined with ML and DL approaches, have demonstrated great potential for health organizations. In this paper, the discussions regarding epidemic outbreaks on SM have been highlighted. Applying DL to SM analysis for epidemic outbreaks has become a popular research topic recently. In this study, various ML and DL approaches and their applications in SM analysis have been outlined. For various SM analysis tasks, many of these ML and DL techniques have revealed state-of-the-art results. Shortly, with the advances in SM analysis and DL applications, it has been observed that there will be more exciting research in DL for epidemic outbreaks in SM.

Despite the potential benefits of SM-based surveillance methods and ML/DL techniques for epidemic outbreak detection, there are also some limitations and challenges that need to be addressed (discussed above). These include issues related to data quality, privacy concerns, and the need for effective validation and verification of results.

In conclusion, SM-based surveillance methods combined with ML/DL techniques have shown promise for the early detection of epidemic outbreaks. While there are still some challenges to overcome, these methods have the potential to improve public health responses to disease outbreaks and save lives.

Acknowledgement: The authors are thankful to the Deanship of Scientific Research at Najran University for funding this work, under the Research Groups Funding Program.

Funding Statement: The authors are thankful to the Deanship of Scientific Research at Najran University for funding this work, under the Research Groups Funding Program Grant Code (NU/RG/SERC/12/27).

Author Contributions: The authors confirm their contribution to the paper as follows: study conception and design: Samina Amin, Muhammad Ali Zeb, Hani Alshahrani; data collection: Mohammed Hamdi, Mohammad Alsulami, Asadullah Shaikh; analysis and interpretation of results: Samina Amin, Muhammad Ali Zeb, Hani Alshahrani; draft manuscript preparation: Mohammed Hamdi, Mohammad Alsulami, Asadullah Shaikh. All authors reviewed the results and approved the final version of the manuscript.

Availability of Data and Materials: The datasets used to support the analysis of this study are available within the paper.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

1. Balaji, T. K., Annavarapu, C. S. R., Bablani, A. (2021). Machine learning algorithms for social media analysis: A survey. *Computer Science Review*, 40, 100395.
2. Carneiro, H. A., Mylonakis, E. (2009). Google trends: A web-based tool for real-time surveillance of disease outbreaks. *Clinical Infectious Diseases*, 49(10), 1557–1564. <https://doi.org/10.1086/630200>

3. Matsumoto, R., Yoshida, M., Matsumoto, K., Matsuda, H., Kita, K. (2018). Visualization of the occurrence trend of infectious diseases using Twitter. *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, pp. 511–514. Miyazaki, Japan.
4. Paul, M. J., Sarker, A., Brownstein, J. S., Nikfarjam, A., Scotch, M. et al. (2016). Social media mining for public health monitoring and surveillance. *Pacific Symposium on Biocomputing 2016*, pp. 468–479. Big Island, USA.
5. Jordan, S. E., Hovet, S. E., Fung, I. C. H., Liang, H., Fu, K. W. et al. (2019). Using twitter for public health surveillance from monitoring and prediction to public response. *Data*, 4(1), 6–18. <https://doi.org/10.3390/data4010006>
6. Marques-Toledo, C. D. A., Degener, C. M., Vinhal, L., Coelho, G., Meira, W. et al. (2017). Dengue prediction by the web: Tweets are a useful tool for estimating and forecasting dengue at country and city level. *PLoS Neglected Tropical Diseases*, 11(7), e0005729. <https://doi.org/10.1371/journal.pntd.0005729>
7. Coberly, J. S., Fink, C. R., Elbert, Y., Yoon, I. K., Velasco, J. M. et al. (2014). Tweeting fever: Can twitter be used to monitor the incidence of dengue-like illness in the Philippines? *Johns Hopkins APL Technical Digest*, 32(4), 714–725.
8. Brooke, G., Prischi, F. (2020). Structural and functional modelling of SARS-CoV-2 entry in animal models. *Scientific Reports*, 10(1), 15917. <https://doi.org/10.21203/rs.3.rs-29443/v1>
9. Ianevski, A., Yao, R., Fenstad, M. H., Biza, S., Zusinaite, E. et al. (2020). Potential antiviral options against SARS-CoV-2 infection. *Viruses*, 12(6), 642. <https://doi.org/10.3390/v12060642>
10. Terpstra, T., Stronkman, R., de Vries, A., Paradies, G. L. (2012). Towards a realtime twitter analysis during crises for operational crisis management. *Proceedings of the 9th International Conference on Information Systems for Crisis Response and Management ISCRAM*, Vancouver, Canada.
11. Hernandez, S. A., Sanchez, P. G., Toscano, M. K., Perez, H., Portillo, J. et al. (2019). Using twitter data to monitor natural disaster social dynamics: A recurrent neural network approach with word embeddings and kernel density estimation. *Sensors*, 19(7), 1746. <https://doi.org/10.3390/s19071746>
12. Sakaki, T., Okazaki, M., Matsuo, Y. (2010). Earthquake shakes Twitter users: Real-time event detection by social sensors. *Proceedings of the 19th International Conference on World Wide Web*, pp. 851–860. Raleigh, North Carolina, USA.
13. Earle, P. S., Bowden, D. C., Guy, M. (2011). Twitter earthquake detection: Earthquake monitoring in a social world. *Annals of Geophysics*, 54(6), 708–715. <https://doi.org/10.4401/ag-5364>
14. Ajao, O., Bhowmik, D., Zargari, S. (2018). Fake news identification on twitter with hybrid CNN and RNN models. *Proceedings of the International Conference on Social Media & Society, (SMSociety)*, pp. 226–230. Copenhagen, Denmark.
15. Kovar, A. (2009). Effects of media on sports. *International Journal of Applied Research*, 1(4), 320–323.
16. Jaidka, K., Ahmed, S., Skoric, M., Hilbert, M. (2019). Predicting elections from social media: A three-country, three-method comparative study. *Asian Journal of Communication*, 29(3), 252–273. <https://doi.org/10.1080/01292986.2018.1453849>
17. Salunkhe, P., Surnar, A., Sonawane, S. (2017). A review: Prediction of election using twitter sentiment analysis. *International Journal of Advanced Research in Computer Engineering & Technology (IJARCET)*, 6(5), 723–725.
18. Alaoui, I. E., Gahi, Y., Messoussi, R., Chaabi, Y., Todoskoff, A. et al. (2018). A novel adaptable approach for sentiment analysis on big social data. *Journal of Big Data*, 5(1), 1–18. <https://doi.org/10.1186/s40537-018-0120-0>
19. Dowerah, B. T. (2012). Effectiveness of social media as a tool of communication and its potential for technology enabled connections: A micro-level study. *International Journal of Scientific and Research Publications*, 2(5), 1–10.

20. Subramanian, K. R. (2017). Influence of social media in personal communication. *ACADEMICIA: An International Multidisciplinary Research Journal*, 7(9), 114–124. <https://doi.org/10.5958/2249-7137.2017.00093.3>
21. Mehndiratta, P., Sachdeva, S., Soni, D. (2017). Detection of sarcasm in text data using deep convolutional neural networks. *Scalable Computing*, 18(3), 219–228. <https://doi.org/10.12694/scpe.v18i3.1302>
22. Joshi, A., Tripathi, V., Patel, K., Bhattacharyya, P., Carman, M. (2013). Are word embedding-based features useful for sarcasm detection? arXiv preprint arXiv:1610.00883.
23. Bollen, J., Mao, H., Zeng, X. (2011). Twitter mood predicts the stock market. *Journal of Computational Science*, 2(1), 1–8. <https://doi.org/10.1016/j.jocs.2010.12.007>
24. Romano, S., Martino, S. D., Kanhabua, N., Mazzeo, A., Nejdil, W. (2016). Challenges in detecting epidemic outbreaks from social networks. *2016 30th International Conference on Advanced Information Networking and Applications Workshops (WAINA)*, pp. 69–74. Crans-Montana, Switzerland, IEEE.
25. Wilson, K., Brownstein, J. S. (2009). Early detection of disease outbreaks using the Internet. *CMAJ*, 180(8), 829–831. <https://doi.org/10.1503/cmaj.1090215>
26. Jimeno Yepes, A., MacKinlay, A., Han, B. (2015). Investigating public health surveillance using Twitter. *Proceedings of BioNLP 15*, pp. 164–170. Beijing, China, Association for Computational Linguistics.
27. Iso, H., Wakamiya, S., Aramaki, E. (2016). Forecasting word model: Twitter-based influenza surveillance and prediction. *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*, vol. 1, pp. 76–86. Osaka, Japan, The COLING, 2016 Organizing Committee.
28. Bark, O., Grigoriadis, A., Pettersson, J. A. N., Risne, V., Siitova, A. et al. (2017). *A deep learning approach for identifying sarcasm in text (Master's Thesis)*. Chalmers University of Technology, Gothenburg, Sweden.
29. Ginsberg, J., Mohebbi, M. H., Patel, R. S., Brammer, L., Smolinski, M. S. et al. (2009). Detecting influenza epidemics using search engine query data. *Nature*, 457(7232), 1012–1014. <https://doi.org/10.1038/nature07634>
30. Signorini, A., Segre, A. M., Polgreen, P. M. (2011). The use of twitter to track levels of disease activity and public concern in the U.S. during the influenza A H1N1 pandemic. *PLoS One*, 6(5), e19467. <https://doi.org/10.1371/journal.pone.0019467>
31. Hulth, A., Rydevik, G., Linde, A. (2009). Web queries as a source for syndromic surveillance. *PLoS One*, 4(2), e4378. <https://doi.org/10.1371/journal.pone.0004378>
32. Bhattacharya, S., Tran, H., Srinivasan, P., Suls, J. (2012). Belief surveillance with twitter. *Proceedings of the 4th Annual ACM Web Science Conference*, pp. 43–46. New York, NY, USA: Association for Computing Machinery.
33. Jabeen, F., Khan, F. G., Shah, S., Ahmad, B., Jabeen, S. (2023). Exploration of epidemic outbreaks using machine and deep learning techniques. *International Conference on Cybersecurity, Cybercrimes, and Smart Emerging Technologies*, pp. 289–301. Riyadh, Saudi Arabia.
34. Amin, S., Alharbi, A., Uddin, M. I., Alyami, H. (2022). Adapting recurrent neural networks for classifying public discourse on COVID-19 symptoms in Twitter content. *Soft Computing*, 26(20), 11077–11089. <https://doi.org/10.1007/s00500-022-07405-0>
35. Gupta, A., Katarya, R. (2020). Social media based surveillance systems for healthcare using machine learning: A systematic review. *Journal of Biomedical Informatics*, 108, 103500. <https://doi.org/10.1016/j.jbi.2020.103500>
36. Chae, S., Kwon, S., Lee, D. (2018). Predicting infectious disease using deep learning and big data. *International Journal of Environmental Research and Public Health*, 15(8), 1596. <https://doi.org/10.3390/ijerph15081596>

37. Amin, S., Uddin, M. I., Al-Baity, H. H., Zeb, M. A., Khan, M. A. et al. (2021). Machine learning approach for COVID-19 detection on twitter. *Computers, Materials & Continua*, 68(2), 2231–2247. <https://doi.org/10.32604/cmc.2021.016896>
38. Thapen, N., Simmie, D., Hankin, C., Gillard, J. (2016). DEFENDER: Detecting and forecasting epidemics using novel data-analytics for enhanced response. *PLoS One*, 11(5), 1–19. <https://doi.org/10.1371/journal.pone.01554>
39. Ahmad, T., Haroon, H., Dhama, K., Sharun, K., Khan, F. M. et al. (2020). Biosafety and biosecurity approaches to restrain/contain and counter SARS-CoV-2/COVID-19 pandemic: A rapid-review. *Turkish Journal of Biology*, 44(7), 132–145.
40. Sahni, H., Sharma, H. (2020). Role of social media during the COVID-19 pandemic: Beneficial, destructive, or reconstructive? *International Journal of Academic Medicine*, 6(2), 70–75.
41. Amin, S., Uddin, M. I., Zeb, M. A., Alarood, A. A., Mahmoud, M. et al. (2020). Detecting dengue/flu infections based on tweets using lstm and word embedding. *IEEE Access*, 8, 189054–189068. <https://doi.org/10.1109/access.2020.3031174>
42. Lamba, D., Hsu, W. H., Alsadhan, M. (2021). Predictive analytics and machine learning for medical informatics: A survey of tasks and techniques. In: *Machine learning, big data, and IoT for medical informatics*, pp. 1–35. Amsterdam, Netherlands: Elsevier, Academic Press.
43. Al-Garadi, M. A., Hussain, M. R., Khan, N., Murtaza, G., Nweke, H. F. et al. (2019). Predicting cyberbullying on social media in the big data era using machine learning algorithms: Review of literature and open challenges. *IEEE Access*, 7, 70701–70718.
44. Riswantini, D., Nugraheni, E. (2022). Machine learning in handling disease outbreaks: A comprehensive review. *Bulletin of Electrical Engineering and Informatics*, 11(4), 2169–2186.
45. Lavanya, P. M., Sasikala, E. (2021). Deep learning techniques on text classification using natural language processing (NLP) in social healthcare network: A comprehensive survey. *2021 3rd International Conference on Signal Processing and Communication (ICPSC)*, pp. 603–609. Coimbatore, Tamil Nadu, India.
46. Al-Garadi, M. A., Yang, Y. C., Cai, H., Ruan, Y., O'Connor, K. et al. (2021). Text classification models for the automatic detection of nonmedical prescription medication use from social media. *BMC Medical Informatics and Decision Making*, 21(1), 1–13.
47. Abbas, A. M. (2021). Social network analysis using deep learning: Applications and schemes. *Social Network Analysis and Mining*, 11(1), 106. <https://doi.org/10.1007/s13278-021-00799-z>
48. Jelodar, H., Wang, Y., Orji, R., Huang, S. (2020). Deep sentiment classification and topic discovery on novel coronavirus or COVID-19 online discussions: NLP using LSTM recurrent neural network approach. *IEEE Journal of Biomedical and Health Informatics*, 24(10), 2733–2742.
49. Dang, N. C., Moreno-García, M. N., de la Prieta, F. (2020). Sentiment analysis based on deep learning: A comparative study. *Electronics*, 9(3), 483.
50. William, P., Gade, R., Chaudhari, R., Pawar, A. B., Jawale, M. A. et al. (2022). Machine learning based automatic hate speech recognition system. *2022 International Conference on Sustainable Computing and Data Communication Systems (ICSCDS)*, pp. 315–318. Erode, Tamil Nadu, India.
51. Vashisht, V., Pandey, A. K., Yadav, S. P. (2021). Speech recognition using machine learning. *IEIE Transactions on Smart Processing & Computing*, 10(3), 233–239.
52. Deng, L., Li, X. (2013). Machine learning paradigms for speech recognition: An overview. *IEEE Transactions on Audio, Speech, and Language Processing*, 21(5), 1060–1089.
53. Razno, M. (2019). Machine learning text classification model with NLP approach. *Computational Linguistics and Intelligent Systems*, 2, 71–73.
54. Ofer, D., Brandes, N., Linial, M. (2021). The language of proteins: NLP, machine learning & protein sequences. *Computational and Structural Biotechnology Journal*, 19, 1750–1758.

55. Malviya, S., Tiwari, A. K., Srivastava, R., Tiwari, V. (2020). Machine learning techniques for sentiment analysis: A review. *SAMRIDDHI: A Journal of Physical Sciences, Engineering and Technology*, 12(2), 72–78.
56. Jagdale, R. S., Shirsat, V. S., Deshmukh, S. N. (2019). Sentiment analysis on product reviews using machine learning techniques. *Cognitive Informatics and Soft Computing: Proceeding of CISC 2017*, pp. 639–647. Singapore, Springer. https://doi.org/10.1007/978-981-13-0617-4_61
57. Chen, L. Y., Pierson, E., Rose, S., Joshi, S., Ferryman, K. et al. (2021). Ethical machine learning in healthcare. *Annual Review of Biomedical Data Science*, 4, 123–144.
58. Gull, S., Mansour, R. F., Aljehane, N. O., Parah, S. A. (2021). A self-embedding technique for tamper detection and localization of medical images for smart-health. *Multimedia Tools and Applications*, 80(19), 29939–29964. <https://doi.org/10.1007/s11042-021-11170-x>
59. Shen, D., Wu, G., Suk, H. I. (2017). Deep learning in medical image analysis. *Annual Review of Biomedical Engineering*, 19, 221–248.
60. Shorten, C., Khoshgoftaar, T. M. (2019). A survey on image data augmentation for deep learning. *Journal of Big Data*, 6(1), 60. <https://doi.org/10.1186/s40537-019-0197-0>
61. Voulodimos, A., Doulamis, N., Doulamis, A., Protopapadakis, E. (2018). Deep learning for computer vision: A brief review. *Computational Intelligence and Neuroscience*, 2018, 7068349. <https://doi.org/10.1155/2018/7068349>
62. Razzak, M. I., Naz, S., Zaib, A. (2018). Deep Learning for Medical Image Processing: Overview, Challenges and the Future. In: Dey, N., Ashour, A., Borra, S. (Eds.), *Lecture notes in computational vision and biomechanics*, vol. 26. Cham: Springer, https://doi.org/10.1007/978-3-319-65981-7_12
63. Hatt, M., Parmar, C., Qi, J., Naqa, E.I (2019). Machine (deep) learning methods for image processing and radiomics. *IEEE Transactions on Radiation and Plasma Medical Sciences*, 3(2), 104–108.
64. Shakeel, P. M., Burhanuddin, M. A., Desa, M. I. (2019). Lung cancer detection from CT image using improved profuse clustering and deep learning instantaneously trained neural networks. *Measurement*, 145, 702–712.
65. Shi, F., Wang, J., Shi, J., Wu, Z., Wang, Q. et al. (2020). Review of artificial intelligence techniques in imaging data acquisition, segmentation, and diagnosis for COVID-19. *IEEE Reviews in Biomedical Engineering*, 14, 4–15.
66. Kalaivani, N., Manimaran, N., Sophia, S., Devi, D. D. (2020). Deep learning based lung cancer detection and classification. *IOP Conference Series: Materials Science and Engineering*, 994(1), 012026.
67. Chehade, A. H., Abdallah, N., Marion, J. M., Oueidat, M., Chauvet, P. (2022). Lung and colon cancer classification using medical imaging: A feature engineering approach. *Physical and Engineering Sciences in Medicine*, 45(3), 729–746. <https://doi.org/10.1007/s13246-022-01139>
68. Chugh, G., Kumar, S., Singh, N. (2021). Survey on machine learning and deep learning applications in breast cancer diagnosis. *Cognitive Computation*, 13, 1451–1470. <https://doi.org/10.1007/s12559-020-09813-6>
69. Hesamian, M. H., Jia, W., He, X., Kennedy, P. (2019). Deep learning techniques for medical image segmentation: Achievements and challenges. *Journal of Digital Imaging*, 32, 582–596.
70. Mukhamadiyev, A., Khujayarov, I., Djuraev, O., Cho, J. et al. (2022). Automatic speech recognition method based on deep learning approaches for Uzbek language. *Sensors*, 22(10), 3683.
71. Nassif, A. B., Shahin, I., Attili, I., Azzeh, A. M., Shaalan, K. (2022). Speech recognition using deep neural networks: A systematic review. *IEEE Access*, 7, 19143–19165.
72. Zhang, Z., Geiger, J., Pohjalainen, J., Mousa, A. E. D., Jin, W. et al. (2018). Deep learning for environmentally robust speech recognition: An overview of recent developments. *ACM Transactions on Intelligent Systems and Technology*, 9(5), 1–28.

73. Kamath, U., Liu, J., Whitaker, J. (2019). *Deep learning for NLP and speech recognition*. Cham, Switzerland: Springer.
74. Strubell, E., Ganesh, A., McCallum, A. (2019). Energy and policy considerations for deep learning in NLP. arXiv preprint arXiv:1906.
75. Young, T., Hazarika, D., Poria, S., Cambria, E. (2018). Recent trends in deep learning based natural language processing. *IEEE Computational Intelligence Magazine*, 13(3), 55–75. <https://doi.org/10.1109/MCI.2018.2840738>
76. Ahmad, H., Asghar, M. Z., Khan, A. S., Habib, A. (2020). A systematic literature review of personality trait classification from textual content. *Open Computer Science*, 10(1), 175–193.
77. Amin, S., Uddin, M. I., AlSaeed, D. H., Khan, A., Adnan, M. (2021). Early detection of seasonal outbreaks from twitter data using machine learning approaches. *Complexity*, 2021, 5520366.
78. Charalambous, A. (2019). Social media and health policy. *Asia-Pacific Journal of Oncology Nursing*, 6(1), 24–27. <https://doi.org/10.4103/apjon.apjon>
79. Moorhead, S. A., Hazlett, D. E., Harrison, L., Carroll, J. K., Irwin, A. et al. (2013). A new dimension of health care: Systematic review of the uses, benefits, and limitations of social media for health communication. *Journal of Medical Internet Research*, 15(4), 1–16. <https://doi.org/10.2196/jmir.1933>
80. Alessa, A., Faezipour, M. (2018). A review of influenza detection and prediction through social networking sites. *Theoretical Biology and Medical Modelling*, 15(2), 1–27. <https://doi.org/10.1186/s12976-017-0074-5>
81. Zivkovi, M. (2010). Flu detector-tracking epidemics on Twitter. *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pp. 599–602. Barcelona, Spain. <https://doi.org/10.1007/978-3-642-15939-8>
82. Du, J., Tang, L., Xiang, Y., Zhi, D., Xu, J. et al. (2018). Public perception analysis of tweets during the 2015 measles outbreak: Comparative study using convolutional neural network models. *Journal of Medical Internet Research*, 20(7), e236.
83. Aramaki, E., Sachiko, M., Morita, M. (2011). Twitter catches the flu: Detecting influenza epidemics using twitter the university of Tokyo. *Proceedings of the Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics*, pp. 1568–1576. Edinburgh, Scotland, UK.
84. Yousefinaghani, S., Dara, R., Poljak, Z., Bernardo, T. M. (2019). The assessment of Twitter’s potential for outbreak detection: Avian influenza case study. *Scientific Reports*, 9(1), 1–17. <https://doi.org/10.1038/s41598-019-54388-4>
85. Samaras, L., García-Barriocanal, E., Sicilia, M. A. (2020). Comparing Social media and Google to detect and predict severe epidemics. *Scientific Reports*, 10(1), 1–11. <https://doi.org/10.1038/s41598-020-61686-9>
86. Hossain, M. P., Junus, A., Zhu, X., Jia, P., Wen, T. H. et al. (2020). The effects of border control and quarantine measures on the spread of COVID-19. *Epidemics*, 32(5), 100397. <https://doi.org/10.1016/j.epidem.2020.100397>
87. Zaraket, H., Melhem, N., Malik, M., Khan, W. M., Dbaibo, G. et al. (2018). Review of seasonal influenza vaccination in the eastern mediterranean region: Policies, use and barriers. *Journal of Infection and Public Health*, 12(4), 472–478. <https://doi.org/10.1016/j.jiph.2018.10.009>
88. Paul, M. J., Dredze, M. (2017). Social monitoring for public health. In: *Synthesis lectures on information concepts, retrieval, and services*, vol. 9, no. 5, pp. 1–183. <https://doi.org/10.2200/s00791ed1v01y201707icr060>
89. Wang, Y., Xu, K., Kang, Y., Wang, H., Wang, F. et al. (2020). Regional influenza prediction with sampling twitter data and PDE model. *International Journal of Environmental Research and Public Health*, 17(3), 678. <https://doi.org/10.3390/ijerph17030678>
90. Chen, L., Tozammel, K. S. M. H., Butler, P., Ramakrishnan, N., Prakash, B. A. (2016). Syndromic surveillance of Flu on Twitter using weakly supervised temporal topic models. *Data Mining and Knowledge Discovery*, 30(3), 681–710. <https://doi.org/10.1007/s10618-015-0434-x>

91. Alessa, A., Faezipour, M. (2019). Preliminary 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*, 5(2), 1–17. <https://doi.org/10.2196/12383>
92. Lee, K., Agrawal, A., Choudhary, A. (2013). Real-time disease surveillance using twitter data: Demonstration on flu and cancer. *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 1474–1477. Chicago, Illinois, USA. <https://doi.org/10.1145/2487575.2487709>
93. Coltart, C. E. M., Lindsey, B., Ghinai, I., Johnson, A. M., Heymann, D. L. (2017). The Ebola outbreak, 2013–2016: Old lessons for new epidemics. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 372(1721), 20160297.
94. Odlum, M., Yoon, S. (2015). What can we learn about the Ebola outbreak from tweets? *American Journal of Infection Control*, 43(6), 563–571.
95. Crook, B., Glowacki, E. M., Suran, M., Harris, J. K., Bernhardt, J. M. (2016). Content analysis of a live CDC Twitter chat during the 2014 Ebola outbreak. *Communication Research Reports*, 33(4), 349–355.
96. Odlum, M., Yoon, S. (2018). Health information needs and health seeking behavior during the 2014–2016 Ebola outbreak: A Twitter content analysis. *PLoS Currents*, 10. <https://doi.org/10.1371/currents.outbreaks.fa814fb2bec36e29b718ab6af66124fa>
97. Guidry, J. P. D., Jin, Y., Orr, C. A., Messner, M., Meganck, S. (2017). Ebola on Instagram and Twitter: How health organizations address the health crisis in their social media engagement. *Public Relations Review*, 43(3), 477–486. <https://doi.org/10.1016/j.pubrev.2017.04.009>
98. Kim, E. H. J., Jeong, Y. K., Kim, Y., Kang, K. Y., Song, M. (2015). Topic-based content and sentiment analysis of Ebola virus on Twitter and in the news. *Journal of Information Science*, 42(6), 763–781. <https://doi.org/10.1177/0165551515608733>
99. Lazard, A. J., Scheinfeld, E., Bernhardt, J. M., Wilcox, G. B., Suran, M. (2015). Detecting themes of public concern: A text mining analysis of the Centers for Disease Control and Prevention’s Ebola live Twitter chat. *American Journal of Infection Control*, 43(10), 1109–1111.
100. van Lent, L. G. G., Sungur, H., Kunneman, F. A., van de Velde, B., Das, E. (2017). Too far to care? Measuring public attention and fear for Ebola using Twitter. *Journal of Medical Internet Research*, 19(6), e193.
101. Stefanidis, A., Vraga, E., Lamprianidis, G., Radzikowski, J., Delamater, P. L. et al. (2017). Zika in Twitter: Temporal variations of locations, actors, and concepts. *JMIR Public Health and Surveillance*, 3(2), e6925.
102. Vijaykumar, S., Nowak, G., Himelboim, I., Jin, Y. (2018). Virtual Zika transmission after the first US case: Who said what and how it spread on Twitter. *American Journal of Infection Control*, 46(5), 549–557.
103. Pruss, D., Fujinuma, Y., Daughton, A. R., Paul, M. J., Arnot, B. et al. (2019). Zika discourse in the Americas: A multilingual topic analysis of Twitter. *PLoS One*, 14(5), e0216922.
104. Daughton, A. R., Paul, M. J. (2019). Identifying protective health behaviors on Twitter: Observational study of travel advisories and Zika virus. *Journal of Medical Internet Research*, 21(5), e13090. <https://doi.org/10.2196/13090>
105. Masri, S., Jia, J., Li, C., Zhou, G., Lee, M. C. et al. (2019). Use of Twitter data to improve Zika virus surveillance in the United States during the 2016 epidemic. *BMC Public Health*, 19(1), 761. <https://doi.org/10.1186/s12889-019-7103-8>
106. Abouzahra, M., Tan, J. (2021). Twitter vs. Zika—The role of social media in epidemic outbreaks surveillance. *Health Policy and Technology*, 10(1), 174–181. <https://doi.org/10.1016/j.hlpt.2020.10.014>
107. Wood, M. J. (2018). Propagating and debunking conspiracy theories on Twitter during the 2015–2016 Zika virus outbreak. *Cyberpsychology, Behavior, and Social Networking*, 21(8), 485–490.
108. DeLima, T. F. M., Lana, R. M., SennaCarneiro, T. G., Codeço, C. T., Machado, G. S. et al. (2016). Dengueme: A tool for the modeling and simulation of dengue spatiotemporal dynamics. *International Journal of Environmental Research and Public Health*, 13(9), 1–21. <https://doi.org/10.3390/ijerph13090920>

109. Toan, N. T., Rossi, S., Prisco, G., Nante, N., Viviani, S. (2015). Dengue epidemiology in selected endemic countries: Factors influencing expansion factors as estimates of underreporting. *Tropical Medicine and International Health*, 20(7), 840–863. <https://doi.org/10.1111/tmi.12498>
110. Hasan, S., Sami, F. J., Munther, A. (2016). Dengue virus: A global human threat: Review of literature. *Journal of International Society of Preventive & Community Dentistry*, 6(1), 1–6.
111. Amin, S., Alouffi, B., Uddin, M. I., Alosaimi, W. (2022). Optimizing convolutional neural networks with transfer learning for making classification report in COVID-19 chest X-rays scans. *Scientific Program*, 2022, 5145614. <https://doi.org/10.1155/2022/5145614>
112. Jain, V. K., Kumar, S. (2018). Effective surveillance and predictive mapping of mosquito-borne diseases using social media. *Journal of Computational Science*, 25, 406–415. <https://doi.org/10.1016/j.jocs.2017.07.003>
113. Guo, P., Liu, T., Zhang, Q., Wang, L., Xiao, J. (2017). Developing a dengue forecast model using machine learning: A case study in China. *PLoS Neglected Tropical Diseases*, 11(10), 1–22. <https://doi.org/10.1371/journal.pntd.0005973>
114. Ding, X., Yin, K., Li, Z., Lalla, R. V., Ballesteros, E. et al. (2020). Ultrasensitive and visual detection of SARS-CoV-2 using all-in-one dual CRISPR-Cas12a assay. *Nature Communications*, 11(1), 1–10. <https://doi.org/10.1038/s41467-020-18575-6>
115. Huang, A. T., Garcia-Carreras, B., Hitchings, M. D., Yang, B., Katzelnick, L. C. et al. (2020). A systematic review of antibody mediated immunity to coronaviruses: Kinetics, correlates of protection, and association with severity. *Nature Communications*, 11(1), 1–16. <https://doi.org/10.1038/s41467-020-18450-4>
116. Erkkilä, T., Luoma-aho, V. (2023). Alert but somewhat unaligned: Public sector organisations' social media listening strategies during the COVID-19 pandemic. *Journal of Communication Management*, 27(1), 120–135. <https://doi.org/10.1108/JCOM-02-2022-0015>
117. Sarwar, S., Waheed, R., Sarwar, S., Khan, A. (2020). COVID-19 challenges to Pakistan: Is GIS analysis useful to draw solutions? *Science of the Total Environment*, 730, 139089. <https://doi.org/10.1016/j.scitotenv.2020.139089>
118. Anand, P., Yadav, A., Debata, P., Bachani, S., Gupta, N. et al. (2020). Clinical profile, viral load, management and outcome of neonates born to COVID 19 positive mothers: A tertiary care centre experience from India. *European Journal of Pediatrics*, 180, 547–559. <https://doi.org/10.1007/s00431-020-03800-7>
119. Tsao, S. F., Chen, H., Tisseverasinghe, T., Yang, Y., Li, L. et al. (2021). What social media told us in the time of COVID-19: A scoping review. *The Lancet Digital Health*, 3(3), e175–e194. [https://doi.org/10.1016/S2589-7500\(20\)30315-0](https://doi.org/10.1016/S2589-7500(20)30315-0)
120. Malla, S., A., P. J. A. (2021). COVID-19 outbreak: An ensemble pre-trained deep learning model for detecting informative tweets. *Applied Soft Computing*, 107, 107495. <https://doi.org/10.1016/j.asoc.2021.107495>
121. Samuel, J., Ali, G. G. M. N., Rahman, M. M., Esawi, E., Samuel, Y. (2020). COVID-19 public sentiment insights and machine learning for tweets classification. *Information*, 11(6), 1–22. <https://doi.org/10.3390/info11060314>
122. Kabir, M. Y., Madria, S. (2020). CoronaVis: A real-time COVID-19 tweets data analyzer and data repository. arXiv preprint arXiv:2004.13932
123. Ardabili, S. F., Mosavi, A., Ghamisi, P., Ferdinand, F., Varkonyi-Koczy, A. R. et al. (2020). COVID-19 outbreak prediction with machine learning. *Algorithms*, 13(10), 249.
124. Khanday, A. M. U. D., Rabani, S. T., Khan, Q. R., Rouf, N., Mohi, M. U. D. (2020). Machine learning based approaches for detecting COVID-19 using clinical text data. *International Journal of Information Technology*, 12(3), 731–739. <https://doi.org/10.1007/s41870-020-00495-9>
125. Prabhakar, K. D. R., Prasad, D. A. V. (2020). Informational flow on Twitter-Corona virus outbreak-topic modelling approach. *International Journal of Advanced Research in Engineering and Technology (IJARET)*, 11(3), 128–134.

126. Nemes, L., Kiss, A. (2021). Social media sentiment analysis based on COVID-19. *Journal of Information and Telecommunication*, 5(1), 1–15. <https://doi.org/10.1080/24751839.2020.1790793>
127. Wakamiya, S., Morita, M., Kano, Y., Ohkuma, T., Aramaki, E. (2019). Tweet classification toward twitter-based disease surveillance: New data, methods, and evaluations. *Journal of Medical Internet Research*, 21(2). <https://doi.org/10.2196/12783>
128. Wakamiya, S., Kawai, Y., Aramaki, E. (2018). Twitter-based influenza detection after flu peak via tweets with indirect information: Text mining study. *Journal of Medical Internet Research*, 20(9), 1–27. <https://doi.org/10.2196/publichealth.8627>
129. Collier, N., Son, N. T., Nguyen, N. M. (2011). OMG U got flu? Analysis of shared health messages for bio-surveillance analysis of shared health messages for bio-surveillance. *Journal of Biomedical Semantics*, 2(5), 1–10.
130. Albinati, J., Meira, W., Pappa, G. L., Teixeira, M., Marques, C. T. (2017). Enhancement of epidemiological models for dengue fever based on Twitter data. *ACM International Conference Proceeding Series*, pp. 109–118. Tacoma, Washington, USA. <https://doi.org/10.1145/3079452.3079464>
131. Espina, K., Regina, M., Estuar, J. E. (2017). Infodemiology for syndromic surveillance of dengue and typhoid fever in the Philippines. *Procedia Computer Science*, 121(1), 554–561. <https://doi.org/10.1016/j.procs.2017.11.073>
132. Amin, S., Uddin, M. I., Duaa, H., Khan, A., Adnan, M. (2021). Early detection of seasonal outbreaks from twitter data using machine learning approaches. *Complexity*, 2021, 5520366.
133. Hong, Y., Sinnott, R. O. (2018). A social media platform for infectious disease analytics. *Computational Science and Its Applications-ICCSA 2018*, pp. 526–540. Melbourne, VIC, Australia.
134. Ahmad, A. R., Murad, H. R. (2020). The impact of social media on panic during the COVID-19 pandemic in iraqi kurdistan: Online questionnaire study. *Journal of Medical Internet Research*, 22(5), 1–11. <https://doi.org/10.2196/19556>
135. Punn, N. S., Sonbhadra, S. K., Agarwal, S. (2020). COVID-19 epidemic analysis using machine learning and deep learning algorithms. *medRxiv*. <https://doi.org/10.1101/2020.04.08.20057679>
136. Sivanantham, K. (2021). Sentiment analysis on social media for emotional prediction during COVID-19 pandemic using efficient machine learning approach. In: *Computational intelligence and healthcare informatics*, pp. 215–233. Germany: Wiley Weinheim.
137. Amin, S., Uddin, M. I., Zeb, M. A., Alarood, A. A., Mahmoud, M. et al. (2021). Detecting information on the spread of dengue on twitter using artificial neural networks. *Computers, Materials & Continua*, 67(1), 1317–1332. <https://doi.org/10.32604/cmc.2021.01473>
138. Chew, A. W. Z., Pan, Y., Wang, Y., Zhang, L. (2021). Hybrid deep learning of social media big data for predicting the evolution of COVID-19 transmission. *Knowledge-Based Systems*, 233, 107417.
139. Alorini, G., Rawat, D. B., Alorini, D. (2021). LSTM-RNN based sentiment analysis to monitor COVID-19 opinions using social media data. *ICC 2021-IEEE International Conference on Communications*, pp. 1–6. Montreal, Canada.
140. Al-garadi, M. A., Khan, M. S., Varathan, K. D., Mujtaba, G., Al-Kabsi, A. M. (2016). Using online social networks to track a pandemic: A systematic review. *Journal of Biomedical Informatics*, 62, 1–11. <https://doi.org/10.1016/j.jbi.2016.05.005>