Engineering and Technology Journal e-ISSN: 2456-3358

Volume 09 Issue 04 April-2024, Page No.- 3749-3755

DOI: 10.47191/etj/v9i04.12, I.F. – 8.227

A Literature Survey on Demographics and Topic Impacts on the co-spread of Healthcare prediction based on Twitter(X) Data

Ganesh Nedunuri[1], Sumanth Geddam[2], Srujana Inturi[3]

[1,2,3] Chaitanya Bharathi Institute of Technology

ABSTRACT: Big data is being used much more frequently to monitor trends in business, economics, society, and health as a result of more individuals utilizing Twitter and other social media. Businesses utilize sophisticated statistical techniques to forecast and monitor events such as disease outbreaks and trends in mental health by combining a wide range of data sets from social media. This kind of data utilization has benefits, but there are drawbacks as well, such as the rapidity with which false information can proliferate on Twitter. The approach that is being suggested emphasizes the necessity of improving fact-checking procedures, especially in the healthcare industry where false information can have detrimental effects. This underscores the growing importance of reliable data, information, and knowledge flows in shaping economies while acknowledging the imperative to address challenges associated with the information revolution.

KEYWORDS: Natural Language processing,Social Media data,Topic Modeling,Text Visualisation, Text Summarization, Machine Learning.

I. INTRODUCTION

Twitter is a social media network that is actively used by about 530 million people to express their opinions and thoughts about a wide range of issues, including public health events and healthcare. Because it is constantly updated with information from both public and private sources, this platform has shown to be a valuable source of health-related information. Twitter serves as a valuable real-time source of global public health information for researchers, particularly in the field of public health research. Researchers can improve their understanding of user behavior and communication by grouping Twitter data into pertinent categories.

The primary objective of the public health sector is to safeguard and promote community health. Public health experts work to prevent diseases and mitigate their effects. They closely monitor outbreaks of infectious diseases and investigate why certain individuals are more susceptible to illness than others. Epidemiologists are particularly interested in the dynamics of health conditions within populations. Public health agencies engage in public health surveillance, which involves the systematic and continuous collection, management, analysis, and interpretation of data. This information is then distributed to programs that support public health initiatives.

Because to a confluence of socioeconomic, environmental, and ecological forces, disease outbreaks are becoming more widespread and varied globally. Digital data analysis is one of the new methods being developed for outbreak detection. Improving surveillance and responding quickly are crucial in combating recently emerging infectious diseases such as Zika and Ebola. Technological advancements may enable the prediction and forecasting of illness patterns, such as COVID- 19.

The main goals of this research article are to lessen the public's exposure to false information by analyzing and forecasting healthcare-related tweets.

II. RELATED WORK

With a focus on diagnosing mental health difficulties and public health disasters, particularly during the Covid-19 outbreak, the researchers in [1] conducted a study to ascer- tain the potential utility of Twitter analytics as a prediction tool in healthcare. They assessed previous scholarly research and applications pertaining to Twitter's predictive powers in healthcare through the use of social network analysis (SNA) and a systematic literature review (SLR). The aim of the project was to separate valuable insights from the abundance of misinformation that was spreading on Twitter and to highlight the significance of double-checking material published on the platform.

Machine learning, social network analysis techniques, and big data analytics are some of the technologies used in [1].These tools aid in the processing and analysis of Twitter's massive unstructured data volume, enabling academics to draw important conclusions.For tasks like sentiment analysis, identifying false news, and identifying misinformation, machine learning algorithms are used. It highlighted the function

of fact-checking systems in preventing the dissemination of false information. Together, these technologies allow for a comprehensive knowledge and application of Twitter analytics in the healthcare industry.

Aakansha R. K. Gupta [2] investigated the potential for social media data, particularly from Twitter, to be utilized for real-time surveillance in healthcare. The authors con- ducted a comprehensive review of 1240 publications from 2010 to 2018, focusing on 148 articles that utilized machine learning (ML) techniques to analyze health-related text in social media. The primary objective was to identify trends and technologies employed in this field. According to the researchers, Twitter was the most often used data source for health surveillance, appearing in 64% of the studies, while Support Vector Machine (SVM) was the most often used machine learning method, appearing in 33% of the studies. Twenty-four percent of the research were devoted to tracking influenza-like or flu-like illnesses.Even though the prediction of diseases has improved with the integration of social media data; nonetheless, obstacles like noise, demographic bias, and privacy concerns still exist.

This research makes recommendations for future work, such as integrating online data with physical conditions, incorpo- rating additional features for analysis, exploring video content analysis, improving topic modeling [1], and implementing predictive models using multiple social media platforms for more accurate epidemic predictions.

In [3], H. A. Gregoire Burel examined how well fact-checking worked to stop the COVID-19 pandemic's falsehoods from spreading on Twitter. It aims to comprehend the relationship between disinformation and fact-checks by analyzing trends in various subjects and user demographics.The intention is to offer knowledge that will enable fact-checkers to more effectively target false information and increase the overall impact of correction information in public debate.

The researchers used sophisticated analytical techniques to investigate the reciprocal link between disinformation and fact- checks, including spread variance analysis, impulse response modeling, and causal analysis. These methods facilitate the investigation of the thematic and chronological dimensions of disinformation that circulated within the given time frame. The study makes use of a massive dataset—nearly seventeen times larger than earlier research—that includes tweets from December 2019 to January 2021. To offer a thorough grasp of the dissemination patterns, the authors additionally incor- porated demographic data and made use of the COVID-19 subjects determined by the Poynter Institute's Corona Virus Facts Alliance. The amalgamation of various approaches and data sources facilitates a comprehensive investigation of the influence of fact-checking on the propagation of disinforma- tion within the framework of the epidemic.

T. D. Jitendra Vikram Tembhurne and Md. Moin Almin tackled the growing problem of fake news spreading on social media in [4] by putting out the innovative multi-channel deep learning model, Mc-DNN, for effective and precise detection.The primary goal is to enhance the capabilities of existing fake news detection systems, leveraging both news headlines and articles separately as distinct channels. The aim is to contribute a sophisticated solution that surpasses the limitations of previous models, achieving a high level of accuracy in distinguishing between real and fake news.

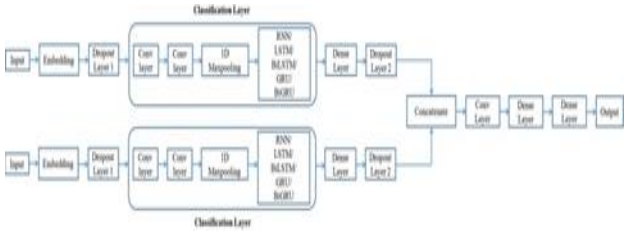


Fig. 1. Model proposed for fake news detection in [4]

Modern advances in machine learning and artificial intel- ligence are used in the Multi channel-Deep Neural Network model that has been suggested.The core of the model is made up of Convolutional Neural Networks, which are effective instruments for automatically extracting features from news articles and headlines. The model's capacity to identify tem- poral dependencies in the news content is further improved with the addition of recurrent neural networks, such as Gated Recurrent Unit and Long Short-Term Memory. The ensemble architecture, inspired by previous literature, maximizes the model's accuracy by combining the strengths of different neu- ral network components. The research strategically leverages the popularity and applicability of CNNs in text classification, providing optimal results for fake news detection. The model's

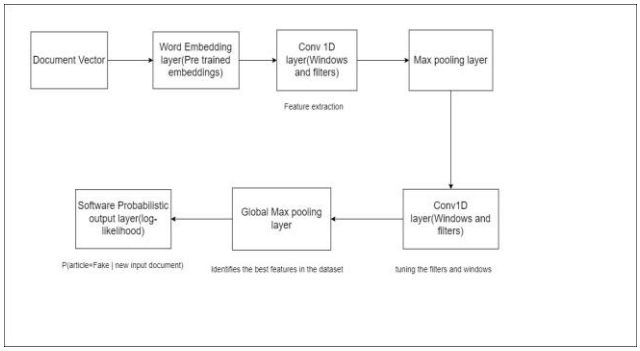


Fig. 2. proposed model for fake news detection

effectiveness is demonstrated through rigorous experimenta- tion and comparison with state-of-the-art machine learning and deep learning techniques, showcasing its superiority in accuracy on significant datasets. In the era of digital informa- tion, the authors' integration of cutting-edge technology has resulted in a reliable and efficient solution for the crucial duty of detecting fake news.

The goal of [5] is to stop misleading information from spread- ing, particularly during the COVID-19 epidemic, with a spe- cific focus on social media in Brazil. The authors introduced a novel method inspired by Markov chains to identify and cluster topics in tweets from fact-checking organizations. Through data analysis, the authors hope to identify patterns, parallels, and divergences in the subjects these organizations covered throughout the pandemic, shedding light on the complex relationship between politics and health crises.

A computational technique influenced by Markov chains was used by M. G. Q. Wilson Ceron a, Mathias-Felipe de-Lima-Santos in [5] to find and group topics in tweets. This technique entails gathering and examining information from two Brazilian fact-checking Twitter accounts. Time- series, subject, and trend analysis are all incorporated into the analysis, offering a multidimensional approach. The authors highlight the potential of their computational model to tackle the challenge of information overload during an "infodemic," providing valuable insights into the topics circulated during a critical period.

I. C.-H. F. H. L. K.-W. F. Sophie E. Jordan, Sierra E. Hovet and Z. T. H in [6] explored how Twitter can be used as a valuable resource for public health research. It focused on classifying Twitter data to understand how people discuss health topics, aiming to utilize this information for global public health surveillance. The goal is to uncover patterns, reactions, and trends related to health issues, such as Zika and Ebola, by analyzing the vast pool of real-time data available on Twitter. Authors in [6] employed various technologies, including data mining and tweet classification, to extract meaningful insights from Twitter. They analyzed short-text updates from users worldwide to categorize information and understand public reactions. It highlights the challenges in using different methods for classifying tweets, with a focus on sentiment anal- ysis and user classification. Despite the potential, the authors emphasized the lack of standardization in these methods and the need to improve accuracy.

M. E. Marouane Birjali, Abderrahim Beni-Hssane in [7] identified suicidal sentiments expressed on Twitter through the application of sentiment analysis. This involves the integration of machine learning algorithms and semantic analysis. The overarching goal is to construct a specialized vocabulary related to suicide and leverage this knowledge to predict and prevent instances of suicidal ideation within the realm of social media.

The auhtors employed several key technologies. Weka serves as a pivotal tool for data mining, facilitating the classification of data and extracting pertinent information from Twitter. The application of machine learning algorithms within Weka enhances the analysis of user-generated content on the platform.

The collection of Twitter data is made possible through the utilization of Twitter4J providing valuable insights into sentiments expressed by users on the social network.

Semantic analysis is conducted using WordNet, a lexical database, to calculate the semantic similarity between tweets. This step is crucial for discerning the affective content and sentiments associated with suicidal thoughts.

The authors acknowledged the importance of social net- works, particularly Twitter, as platforms for the extraction and analysis of sentiments related to depression and suicide. Recognizing the prevalence of risk factors on these platforms, [7] emphasized the need for continuous monitoring.

By examining more than 1.25 million English tweets sent between January 2020 and May 2021, the authors of [8] attempted to comprehend the sentiments and viewpoints of the public during the various phases of the Covid-19 pandemic. The primary goal is to predict tweet popularity, measured by retweets, using advanced machine learning techniques.

The researchers utilized a range of technologies, including the Crystal Feel algorithm and machine learning techniques like LDA [1], to assess the content of tweets. Tweets were used to extract five sets of content features, including topic analysis and TF-IDF vectorizers. After that, supervised ma- chine learning methods, such as an ensemble voting classifier, were given these features in order to create a predictive model for tweet retweetability. The authors emphasized the importance of understanding emotional intensity in tweets and highlights the popularity of emotional content compared to purely informational tweets about the Covid-19 pandemic.

To address the global issue of misinformation during the COVID-19 pandemic, D. Kim in [9] created a dataset called FibVID (Fake news Information-broadcasting dataset of COVID-19). FibVID provides user data, relevant tweets via Twitter, and annotated news pieces from fact-checking websites. The purpose of this dataset is to identify news propagation trends, identify user traits in fake news diffusion, and suggests applications for future research. The exploratory analysis covers topics, claim propagation, and user influence. This paper highlights the significance of FibVID in advanc- ing research on misinformation by offering a comprehensive dataset that spans various dimensions of the information ecosystem. The dataset's multi-modal nature, combining tex- tual and social network information, allows for a holistic understanding of how misinformation spreads. The authors emphasize the potential impact of FibVID on developing more effective and nuanced strategies for combating fake news during health crises and beyond. By shedding light on the intricate interplay between news content, social interactions, and user characteristics, Opportunities for interdisciplinary cooperation and the creation of focused solutions to lessen the negative consequences of disinformation are made possible by FibVID.

F. N. Yasmim Mendes Roch in [10] showed how misinformation spread through social media platforms affects people's mental health and contributes to anxiety, fear, panic, and other psychological disorders.

The influence of fake news on patients and healthcare workers worldwide was examined by the authors in a number of studies. The psychological pain brought on by false information is highlighted, including feelings of anxiety, fear, panic, melancholy, and exhaustion. Additionally, it highlights the significance of comprehending infodemic expertise [5] and the possible harm it may cause to public health.

Authors employed various methodologies, including online surveys, interviews, and queries in various databases. The find- ings suggest that misinformation on social media platforms, especially through popular channels like Facebook, YouTube, WhatsApp, and Twitter, has led to a range of psychological reactions in individuals of different ages and educational backgrounds.

[11], M. Aman presented a comprehensive language model-based system for identifying false content in created images and movies. The algorithm uses the Llama language model for fine-tuning and splits the instruction pool into alignment and specific tasks. The model aligns with human intentions and values and detects fake news through binary classification. The algorithm uses ChatGPT to generate instructions batch and employs parameter-efficient fine-tuning to optimize the algorithm for applications.

The authors of this research used novel approaches and algorithms along with a comprehensive literature analysis to explore the predictive potential of Twitter data. In particular, the study uses topic modeling with the LDA algorithm, enabling the identification of key topics within the Twitter data. Additionally, a text-to-network analysis, involving keyword co-occurrence methodology, is utilized to create a network graph based on the abstracts and titles of the research papers, facilitating the identification of influential keywords and the inference of topics based on keyword co-occurrences. Graph analytics techniques, including betweenness centrality and modularity class calculations, are also employed to identify influential keywords and communities within the network graph. These approaches enable a comprehensive analysis of the Twitter data, extraction of relevant topics, and insights into the predictive potential of Twitter across diverse application domains [12].

[13] A deep learning-based classifier that determines a tweet or post's position on a claim is proposed by the author in this work. Three APIs—Tweepy, Selenium, and Beau- tiful Soup—are used in a three-step process for gathering data as part of the technique. Four stance detectors—BERT- Support Vector Machine, BERT-Random Forest, BERT-Neural Networks, and BERT-Convolution Neural Network—are then trained and compared using the reaction-based labeled dataset that was created using the data that was gathered. The results of these models are contrasted with the top-ranked teams' scores from the Sem-Eval 2016 tournament. According to the findings, the BERT-CNN performed better than the initial systems. This study's approach uses the pre-trained BERT model to apply sentence embedding techniques for text representation. In addition to performing better than current systems, the suggested model can help detect false content on social media.

According to [14] It looked into how social media spreads fake news and how culture and demography affect that. According to the study, age has a bigger impact on whether or not fake news is accepted in a given culture than either gender or level of education. Culture was also found to have the biggest impact on the spread of fake news. The paper underscored the importance of considering sociocultural factors when an- alyzing the impact of fake news. Employing a quantitative approach, the paper conducted an online survey through social media and utilized confirmatory factor analysis via structural equation modeling with AMOS to explore the relationships between different factors and the acceptance of fake news.

A multilayer network-based method was presented by Gian-luca Bonifazi and Bernardo Brev in [15] to examine user conversations and interactions on social media sites, with particular emphasis on the Twitter environment. In this a multilayer network model was incorporated for different types of user relationships and interactions, such as retweets, replies, likes, and mentions. The single levels of the multilayer net- work were examined, and pertinent hashtags pertaining to pro-

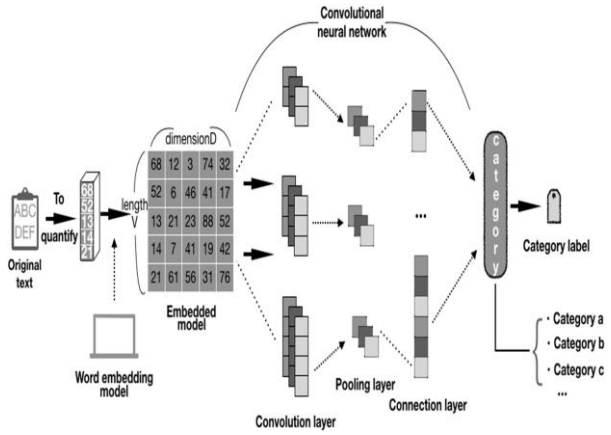
, anti-, and neutral viewpoints were extracted. The author elucidated the distinctions between the single network and multilayer network approaches with regard to information extraction.This approach was applied to identify the most influencing users in the social media.

III. COMPARISON STUDY

Latent thematic structures within a vast corpus of document data can be found using topic modeling, a potent statistical tool in machine learning and natural language processing. Topic modeling's main objective is to locate and extract relevant topics that are indicative of the underlying themes found in the textual data.Latent Dirichlet Allocation is a probabilistic model, one of the most widely used approaches for topic modeling, which assumes that every text is a mixture of a finite number of themes and that every word in the document is assigned to one of the topics. Latent patterns in textual data can be discovered without the requirement for prior subject- matter expertise when combined with other methods like non- negative matrix factorization.

A. Latent Dirichlet Allocation

In natural language processing, latent topics are often identified in large text corpora using a probabilistic generative model known as Latent Dirichlet Allocation. According to the fundamentals of LDA, every document is a collection of several themes, and each topic is a distribution of words. The concept presupposes a two-step generative process for document creation:



1. Topic Assignment:

Topic proportions sum up to one in each document, which is represented as a mixture of themes.

A subject is probabilistically assigned for every word in a document according to the topic proportions in the document.

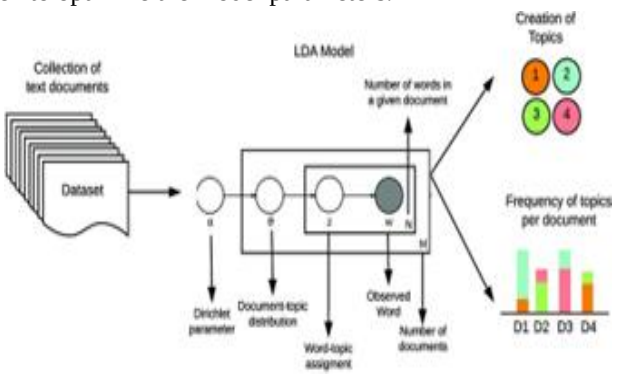
2. Word Assignment:

For each word, a specific term is then chosen from the selected topic's word distribution. The key parameters in LDA include:

Number of Topics (K): This user-specified parameter determines the assumed number of topics present in the corpus.

Dirichlet Priors (α and β): These hyperparameters influence the shape of the topic distribution within documents (α) and the word distribution within topics (β).

Through the use of iterative methods like Gibbs Sampling and Variational Inference, the latent themes and word distributions that correspond with them are estimated during the LDA training process. Through iterative adjustments to the allocations of themes to words and documents, these methods seek to optimize the model parameters.



In practical terms, when applied to a document corpus, LDA can reveal underlying thematic structures, which enables to identify prevalent topics and their associated words. As a result, a distribution over subjects and a set of topics, each represented as a probability distribution over words, are produced for each document.

Application areas for LDA include content recommendation, information retrieval, and document clustering. Understanding the structure and substance of textual material can be facili- tated by using this helpful tool, which can reveal hidden topic patterns in massive datasets.

B. Convolutional Neural Network (ANN)

A Convolutional Neural Network (CNN) adapted for text classification processes sequential data by applying convolutional filters to extract features. Activation functions and pooling layers enhance its ability to discern complex patterns. Flattened features are fed into fully connected layers, and a softmax layer converts outputs into probabilities for classification. During training, the CNN optimizes weights through backpropagation, minimizing a loss function. Trained on text, it excels in recognizing patterns and making accurate predictions. Originally designed for images, CNN's adaptability showcases its effectiveness in diverse domains, emphasizing its significance in modern deep learning research.

C. Support Vector Machine (SVM)

Regression and classification problems are addressed by SVM, a robust and adaptable supervised learning model. Identifying a hyperplane in a high-dimensional space that maximizes the margin—that is, the distance between the nearest data point for each class and the hyperplane—is the primary objective. Since kernel functions transfer the data into higher-dimensional spaces, SVMs are highly efficient for linearly separable data and can handle nonlinear separations. With the use of these kernels, the SVM is able to determine the optimal separating hyperplanes. The support vectors, or subset of training data nearest to the hyperplane, are crucial for determining the hyperplane's location.

D. Two layer ensemble model

A two-layer ensemble model for text classification involves combining predictions from two distinct models to enhance overall performance. The first layer typically consists of di- verse base models. Each base model captures unique features in the input text. The second layer, known as the meta- classifier, aggregates the predictions from the base models to make the final classification decision. This ensemble approach leverages the strengths of different models, promoting more robust and accurate text classification results.

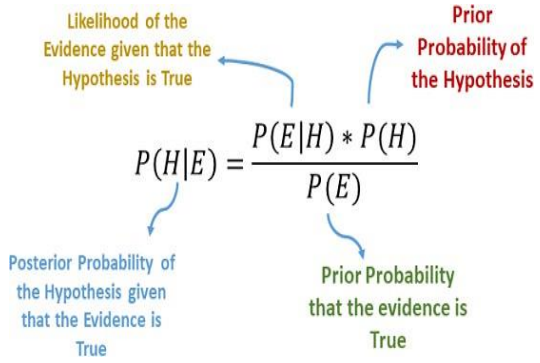
E. Bi-directional LSTM

Long short-term memory networks that are bidirectional handle incoming data both forward and backward. Each LSTM unit captures different aspects of sequential information, al- lowing the model to understand context from

preceding and succeeding words simultaneously. BI-LSTMs are especially useful for text classification tasks because of its bidirectional approach, which improves the network's comprehension of long-range dependencies and contextual complexities in the input text.

F. Naive Bayes

The Naive Bayes machine learning technique is used to categorize news as authentic or fraudulent. It uses the probabilities of the attributes to calculate the likelihood that a given data point belongs to a particular class. It is based on the Bayes theorem. The "naive" part of the technique is explained by the feature independence assumption, which streamlines computations and often produces meaningful results. After determining the probabilities of each class based on the features of the data, Naive Bayes classifies the data and awards a point to the class with the highest probability.



IV. COMPARISON ANALYSIS

Machine Learning (ML) models provide a practical method to the classification of fake news. These models often involve the use of Support Vector Machines and Naive Bayes, two popular ML techniques, together with careful feature engi- neering. These models excel in scenarios where interpretability is crucial, providing insights into the factors driving classifi- cation decisions. Feature engineering may encompass word frequency, TF-IDF, and sentiment analysis, allowing these models to operate effectively even in situations with limited data.

Deep Learning (DL) models, with their capacity to automat- ically learn hierarchical representations, introduce a paradigm shift.Models such as CNN and BiLSTM architectures, offer a different perspective. CNNs are highly proficient in automat- ically identifying patterns and features in text, which makes them ideal for jobs involving the classification of fake news. BILSTM networks improve the model's comprehension of contextual information contained in the language by capturing long-range dependencies in both directions.

LDA is still a widely used topic modeling technique in machine learning. probabilistic generative model Assuming that a collection of documents has a combination of topics, LDA offers interpretable insights into the underlying themes contained in the papers.

On the other hand, BERT, which is commonly used for classification tasks, can also be adapted for topic modeling by framing it as a classification problem. BERT's bidirec- tional training allows it to generate contextual embeddings that capture intricate semantic relationships between words, making it a potent tool for understanding complex topics and relationships within large text corpora.

CONCLUSIONS

Our analysis of Twitter talks about health has shed important light on the variables driving these exchanges. We have improved our understanding of the many viewpoints present in online health debate by looking at factors like age and geog- raphy. This research represents a significant step towards im- proving healthcare decision-making by recognizing the impact of these factors. It is evident that a thorough understanding of these aspects is essential for making informed healthcare choices in the digital realm.

Furthermore, our research has made noteworthy technolog- ical advancements. The development of predictive models and advanced techniques for identifying false health information on Twitter contributes to creating a safer online environment for accessing healthcare information.This protects people from dangerous false information and gives them the confidence to take part in online health debates.

Our study aims to create a digital environment in the future where individuals can trust the health information they get online. By leveraging Twitter data and demographic insights, we seek to promote more informed healthcare decisions. The integration of precise disease tracking, demographic-informed strategies, and a proactive stance against misinformation lays the groundwork for a digital healthcare space that is both informative and secure.

REFERENCES

1. S.S.-A.Enrique Cano-Marin, Marc¸al Mora-Cantallops, "The power of big data analytics over fake news: A scientometric review of twitter as a predictive system in healthcare," Technological Forecasting and Social Change, vol. 190, 2023.
2. R.K.Aakansha Gupta, "Social media based surveillance systems for healthcare using machine learning: A systematic review," in Journal of Biomedical Informatics, 2020.
3. H. A. Gre´goire Burel, Tracie Farrell, "Demographics and topics impact on the co-spread of covid-19 misinformation and fact-checks on twitter," Information Processing Management, vol. 58, 2021.
4. T. D. Jitendra Vikram Tembhurne, Md. Moin Almin, "Mc-dnn: Fake news detection using multi-

channel deep neural networks," in Interna- tional Journal on Semantic Web Information Systems, vol. 18, 2022.

5. M. G. Q. Wilson Ceron a, Mathias-Felipe de-Lima-Santos, "Fake news agenda in the era of covid-19: Identifying trends through fact-checking content," in Online Social Networks and Media, vol. 21, pp. 1080–1083, 2021.
6. I. C.-H. F. H. L. K.-W. F. Sophie E. Jordan, Sierra E. Hovet and Z. T. H. Tse, "Using twitter for public health surveillance from monitoring and prediction to public response," in Big Data and Digital Health, vol. 4, 2019.
7. M. E. Marouane Birjali, Abderrahim Beni-Hssane, "Machine learning and semantic sentiment analysis based algorithms for suicide sentiment prediction in social networks," in Procedia Computer Science, vol. 113,pp. 65–72, 2017.
8. M. E. Marouane Birjali, Abderrahim Beni-Hssane, "Predicting the pop- ularity of tweets by analyzing public opinion and emotions in different stages of covid-19 pandemic," in International Journal of Information Management Data Insights, vol. 2, pp. 65–72, 2022.
9. A. J.-L. S. J. Y.-P. E. C. D. Kim, J., "Fibvid: Comprehensive fake news diffusion dataset during the covid-19 period," in International Journal of Telematics and Informatics, vol. 64, 2021.
10. G. A. D. C. H. d. O. F. D. L. L. D. d. F. N. Yasmim Mendes Rocha1, Gabriel Aca´cio de Moura, "The impact of fake news on social media and its influence on health during the covid-19 pandemic: a systematic review," in Journal of Public Health: From Theory to Practice, vol. 31,pp. 1007–1016, 2023.
11. M. Aman, "Large language model based fake news detection," vol. 231, pp. 740–745, 2024.
12. S.S.-A.Enrique Cano-Marin, Marc¸al Mora-Cantallops, "Twitter as a predictive system: A systematic literature review," in Journal of Business Research, vol. 257, 2023.
13. N. A. Mudasir Ahmad Wani and P. Bours, "Impact of unreliable content on social media users during covid-19 and stance detection system," in Electronic Solutions for Artificial Intelligence Healthcare, 2021.
14. G. Rampersad and T. Althiyabi, "Fake news: Acceptance by demograph- ics and culture on social media," in JOURNAL OF INFORMATION TECHNOLOGY POLITICS, vol. 17, pp. 1–11, 2020.
15. S. C. E. C.-L. V. Gianluca Bonifazi, Bernardo Breve, "Investigating the covid-19 vaccine discussions on twitter through a multilayer network-based approach," in Information Processing Management, vol. 59, 2022.