DOI: 10.14421/ijid.2025.5196 IJID (International Journal on Informatics for Development), *e-ISSN: 2549-7448*
Vol. 14, No. 2, December 2025, Pp. 629-639

Twitter (X) Sentiment Analysis on Monkeypox: A Systematic Literature Review

Hazel Chamboko
Department of Informatics and Analytics
National University of Science and Technology
Bulawayo, Zimbabwe
hazelchamboko@gmail.com

Belinda Ndlovu
Department of Informatics and Analytics
National University of Science and Technology
Bulawayo, Zimbabwe
belinda.ndlovu@nust.ac.zw

Kudakwashe Maguraushe
School of Computing
University of South Africa
Johannesburg, South Africa
magark@unisa.ac.za

Article History
Received May 2nd, 2025
Revised July 14th, 2025
Accepted July 14th, 2025
Published December 2025

***Abstract*— Monkeypox has a risk of growing into a global threat. Understanding public sentiments is crucial for effective emergency responses, as it helps counter misinformation, enhance communication, and improve the retention and application of public health information. This systematic review of literature aims to provide foundations for identifying existing algorithms, commonly used data collection methods, and pre-processing techniques applied to Twitter discussions on Mpox. The review followed the PRISMA guidelines. Relevant literature was retrieved from ScienceDirect, IEEE, PubMed, and Springer databases, resulting in 15 studies that met the inclusion criteria. Most preprocessing methods include stop word removal, lemmatisation, and tokenisation; commonly used data collection methods include Twitter API, Academic API V2, Snscrape, Twint, and Tweepy. Classification of sentiment tended to be hybrid models like CNN-LSTM or transformer-based models such as BERT, which also perform well in dealing with complex linguistic patterns. These recent models, additionally, addressed other very important issues like misinformation detection, irony, and bot-generated content, which earlier models would often fail to tackle. Despite these advancements, much work still needs to be done in improving the accuracy, generalizability, and interpretability of sentiment analysis models in live monitoring of public health.**

***Keywords*—deep learning; machine learning; public sentiment; PRISMA; transformer-based models**

1 INTRODUCTION

Monkeypox (Mpox), a viral zoonotic disease with recent global outbreaks, has become a major public health concern [1]. As of June 2024, 116 countries have reported Mpox cases, with a total of 99,176 cases [2]. Researchers have worked on emotion classification worldwide to assist policymakers and public health in understanding how people view the pandemic using text mining and emotion classification tools [3]. The vast amount of data accessible daily has prompted the use of Natural Language Processing (NLP) tools for text analytics. Evaluating the public's perspective of infectious diseases is essential for the government and policymakers when developing mitigation efforts to restrict the virus [4]. Sentiment analysis, commonly known as opinion mining, is within the topic of NLP [5].

One would wonder why sentiment analysis is important during pandemics, or even in areas that need classification of what people may think about a product. It aims to analyse and understand the feelings, emotions, opinions, and attitudes that people express regarding a specific topic or subject [6]. Today, public opinion and experience are all over the internet, and sentiment analysis allows researchers to understand and use the experiences effectively to enhance healthcare quality [7].

The National Research Council (NRC) method effectively classifies the emotions and sentiments of frequently used words. However, the method still has limitations in analysing whether the user uses fugitive words like satire [8]. In 2024, [9] used four main algorithms: Support Vector Machine, Naïve Bayes, Logistic Regression, and Random Forest, along with Bag-of-words (BOW) and Term Frequency-Inverse Document Frequency (TF-IDF) for feature extraction. However, they were unable to detect nuanced emotions and sarcasm. They proposed the creation of advanced machine learning (ML) models to address hidden perspectives, including sarcasm [9].

Moreover, the Valence Aware Dictionary sEntiment Reasoner (VADER) approach has the limitation of not being able to detect bot-generated tweets [10]. In contrast to how [11] utilized VADER, it works well again with TextBlob and Flair for stigmatisation detection and disinformation; nevertheless, the model requires a large dataset for it to be accurate enough [12]. In 2023, [13] used the same techniques [14] used and could not detect bot-generated tweets. They suggested a comprehensive study across multiple languages and advanced models focusing on automated tweets. In data collection, the Twitter API is the most frequently used data collection method. However, the Twitter API has API restrictions [15].

Systematic survey papers dating from 2020 have in the past discussed in greater detail various algorithms of sentiment analysis, especially in terms of classical machine-learning models such as Naïve Bayes, SVM, and Decision Trees. However, since those reviews were conducted, a major shift took place in NLP from the traditional models to deep learning and then to transformer models like BERT and RoBERTa [16], [17]. Those newer models drastically changed sentiment analysis in the context of short, noisy texts like tweets. Hence, there is a need now for a thorough revisit of the currently existing algorithms in sentiment analysis to address recent innovations and analyse how they perform against each other in public health surveillance on Twitter. Limitations such as API rate limits, keyword bias, and language constraints affect the reliability and representativeness of the collected data [15]. Scholars nowadays have also used third-party tools (Snscrape and Twint, for example) and custom scraping methods to bypass these limitations. However, despite the improvements witnessed, few reviews have systematically analysed or documented this data collection evolution, particularly for cases such as Mpox, where real-time tracking is imperative. Recent advancements in sentiment analysis have transcended the traditional lexicon-building methodology and begun to employ machine learning, deep learning, and hybrid approaches, CNN-LSTM and BERT-type models included [6], [18]. Earlier reviews have described these techniques, but for disease-related conversations on Twitter, none systematically compare their strengths and drawbacks. Preprocessing is also an important issue not adequately addressed, as Twitter provides ample examples of noisy and informal text requiring cleaning steps, including tokenisation [19], [20]. This review, hence, attempts to fill in these gaps by looking at the methodological choices and the preprocessing strategies essential for public health sentiment analysis.

This study seeks to review existing literature to answer the following questions systematically:

1. What existing algorithms are used for sentiment analysis and classification?
2. What are the frequently used methods to collect data for sentiment analysis?
3. What approaches and features have been adopted by other researchers?
4. What techniques are researchers using to preprocess Twitter data?

2 METHOD

Numerous frameworks exist to guide systematic literature reviews, including Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) [21], the Cochrane Handbook [22], the Joanna Briggs Institute (JBI) Reviewer's Manual [23], and the Campbell Collaboration Guidelines [24]. These methods present different complexities and scopes but mainly seek to promote transparency in reporting and replicability and ensure the highest level of methodological precision. The researcher chose the PRISMA methodology as it enhances transparency in systematic reviews [16]. PRISMA serves well because it concentrates on reviews involving multiple databases in which precise documentation of search and screening procedures is of principal concern [25], and this certainly coincides with the objectives of this Mpox-focused Twitter sentiment analysis. Thus, its application implies that the review process is systematic, reproducible, and acceptable internationally, thereby lending weight to the results [26]. A thorough search was done across the following databases to collect relevant information: Science Direct (17 papers), IEEE (52 papers), PubMed (15 papers) and Springer (13 papers). A search term including (“monkeypox” OR “mpox”) and (“sentiment analysis” or “text classification” or “opinion mining”) AND “twitter” was tailored to these libraries' syntax to extract relevant papers. The search was conducted on the 29th of December 2024.

2.1 *Inclusion and Exclusion Criteria*

The studies used this criterion: Papers had to be peer-reviewed journal papers written in English. Papers must have been published between 2019 to 2024, and they mainly focused on sentiment analysis/opinion mining on Twitter data.

2.2 *Screening the Studies*

As illustrated in the PRISMA diagram in Fig.1 1, 97 articles were retrieved from the 4 libraries. The articles were downloaded and stored in a reference manager. Mendeley was used to deduplicate all records, and 30 articles were removed. 33 articles were removed after reading abstracts and concluding that they were irrelevant. After all the screening based on abstracts, 48 papers were left. The researcher removed 14 as they were unrelated to sentiment analysis, 11 were published before 2019, and 8 were not written in English. 15 papers met final inclusion.

3 RESULT AND DISCUSSION

This section presents the results of the 15 articles that were included in the study, using a PRISMA flow diagram. It is shown in Table 1.

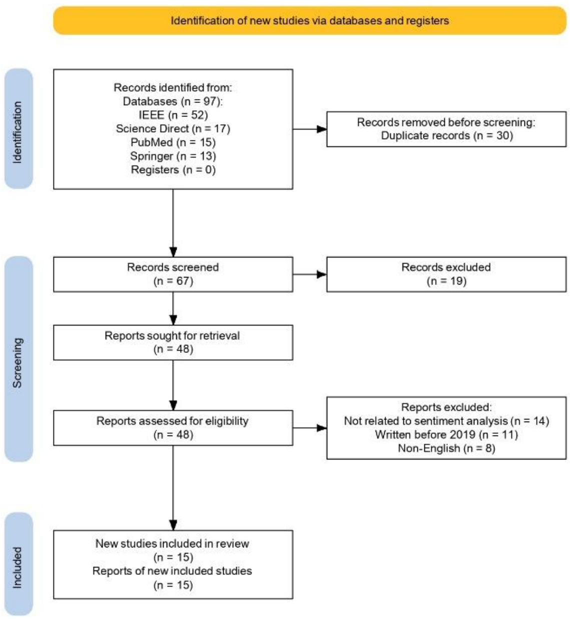

Figure 1. Adapted PRISMA diagram [21]

Table 1. Papers that Met the Inclusion Criteria

Author	Origin	Algorithm	Approach	Adopted Features	Data Collection Method	Preprocessing Techniques.
[8]	Indonesia	• NRC Lexicon	• Lexicon	• Word Frequency • Lexicon Features	• Twitter API	• Stop Word Removal, • Case Normalisation
[13]	USA	• VADER • TextBlob • Flair	• Lexicon	• Lexicon Features	• Twitter’s academic research API v2	• Stop Word Removal
[11]	Netherlands	• VADER	• Lexicon	• Hashtag Frequency • Word Frequency	• Hydrator App	• Tokenization • Stop word removal • Word Frequency analysis
[14]	Netherlands	• VADER • TextBlob • Flair	• Lexicon	• Lexicon Features	• Twitter’s academic research API v2	• Case folding
[4]	USA	• CNN • LSTM • BiLSTM • CNN-LSTM	• Hybrid • Deep Learning	• NRC Lexicon Features	• Twitter API • Tweepy Library •	• Stop Word Removal •
[20]	USA	• VADER • TextBlob • Logistic Regression • SVM • Random Forest • KNN • Multilayer Perception • Naïve Bayes • XGBoost	• Lexicon • Machine Learning	• Word Frequency • TF-IDF • Word Cloud	• Twitter API • Tweepy Library	• Stemming • Lemmatisation, Retweet and user tag removal • Emoji and text conversion
[19]	Peru	• CNN-LSTM	• Hybrid	• Word lemmas • Sentiment Labelling	• Twitter API	• Stemming • Tokenisation

	Spain			• Bag of Words		
[27]	USA	• Multilingual XLM-roBERTa-base • DistilRoBERTA	• Transformers		• Snscrape	• Tokenisation • Stop Word Removal • Lemmatisation • Stemming
[27]	Singapore and Ireland.	• BERT • BERTopic.	• Transformers		•	• Stop word removal • Stemming • Lemmatisation • Tokenisation
[28]	India	• TextBlob • LDA.	• Lexicon	• Syntactic Parsing • Part-of-Speech Tagging • Lexicon Features	• Twint	• Stop word removal • Stemming • Lemmatisation
[29]	Switzerland	• FastText • VADER •	• Lexicon	• N-grams, • Out-of-vocabulary words	• Snscrape • Twitter- API	• Tokenisation • Stemming • Lemmatisation
[30]	Croatia	• AFINN	• Lexicon	• Lexicon Features	• Twitter academic API	• Tokenization
[18]	USA	• VADER • TextBlob • Azure Machine Learning	• Lexicon • Machine Learning	• Word Frequency • Lexicon Features	• Twitter API	• Lemmatisation • Tokenisation • Stop word removal • Word frequency analysis
[31]	Switzerland	• TextBlob • VADER • Random Forest • Logistic Regression • Decision Trees • LSTM-GRNN	• Lexicon • Machine Learning • Hybrid	• Lexicon Features, • TF-IDF • Unigrams • Trigrams • Bigrams	• Tweepy Library	• Noise removal • Case normalisation • Stop word removal • Lemmatization • Tokenization
[32]	Brazil	• Random Forest • Linear Support Vector Machine • Logistic Regression • SBERT • mUSE	• Machine Learning, • Transformers	• Bigrams • Unigrams, • N-grams	• Twitter API	• Stop word removal

3.1 *Number of Studies by Data Collection Method*

Figure 2 shows a bubble chart that illustrates the tools and libraries used to collect Twitter data. The Twitter API has several associated tools, such as Tweepy and Hydrator App, that rely on the Twitter API to function. 7 studies used the Twitter API, and 3 studies employed the Twitter academic API. Larger bubbles (e.g., Twitter API) indicate more frequent usage, while smaller ones, i.e., Twint, Snscrape, show relatively lower adoption as they were each used in 1 or 2 studies, respectively.

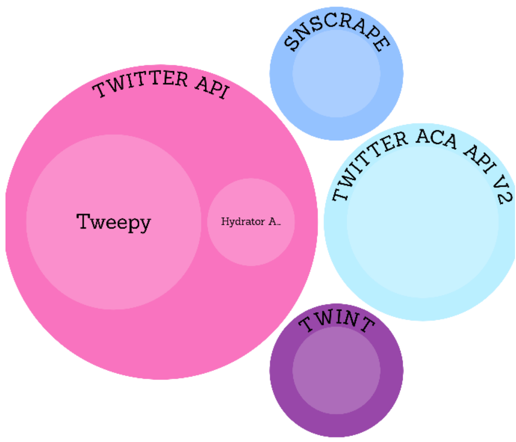

Figure 2. Number of studies per data collection method

3.2 *Number of Studies by Data Processing Technique*

Figure 3 shows a bar chart highlighting the frequency of data preprocessing techniques, namely tokenisation (9 studies), stop word removal (9 studies), stemming (7 studies), case folding (4 studies), word frequency analysis (2 studies) and lemmatisation (8 studies).

From Figure 3, tokenisation and stop word removal show the highest usage, reflecting their essential role in handling short, unstructured tweet data.

3.3 Number of Studies per Approach

Figure 4 is a doughnut chart illustrating the distribution of various sentiment analysis approaches. Figure 4 illustrates that lexicon-based methods prevail because they are often simpler and more interpretable, making them particularly attractive for low-resource languages.

Six lexicon methods were used, including Flair and TextBlob. Transformers are catching up and may soon surpass lexicon-based methods due to advances in NLP and pre-trained models. Five Transformers were identified, including BERT and DistilRoBERTa. Six deep learning models were identified, including Long Short-Term Memory (LSTM) and Convolutional Neural Networks (CNNs). Deep learning approaches were used, but seem to be declining in preference compared to the other methods. Machine learning approaches were the most used in hybrid architectures.

3.4 Features Adopted

Figure 5 illustrates the distribution of adopted features across the included studies, showing the frequency with which, each feature was utilised. Word frequency was the second most frequently used feature, cited in 3 studies, reinforcing its foundational role in NLP.

TF-IDF and N-grams were each adopted in 2 studies, reflecting their usefulness in capturing term importance and local context within text. A variety of other features, including hashtag frequency, bigrams and unigrams, were each used in 1 study. This suggests a wide diversity in feature engineering approaches, with some features being tailored to specific research goals or datasets. The following discusses and answers the research questions.

3.5 Existing Algorithms in Sentiment Analysis

Based on the existing algorithms that were adopted in various studies, as depicted in Table 1, a comprehensive discussion was conducted.

3.5.1 Convolutional Neural Networks (CNN): Convolutional Neural Networks are distinct from fully connected networks in that the neurons in each layer are connected, creating a three-dimensional structure from the input to the convolutional and pooling layers [6]. As noted in [6], the CNN model produced promising outcomes in fitting the training data, as there was a decrease in training loss and a consistent improvement in training accuracy. However, the CNN model struggled to generalise to test data, indicated by fluctuations in test loss and consistently lower test accuracy compared to training accuracy. Findings in [6] suggest that CNNs' application to textual data might overlook crucial features, leading to overfitting. CNN can be useful for analysing small sentences. However, it is advised to use Recurrent Neural Networks (RNN) when the length of the sentences increases [33].

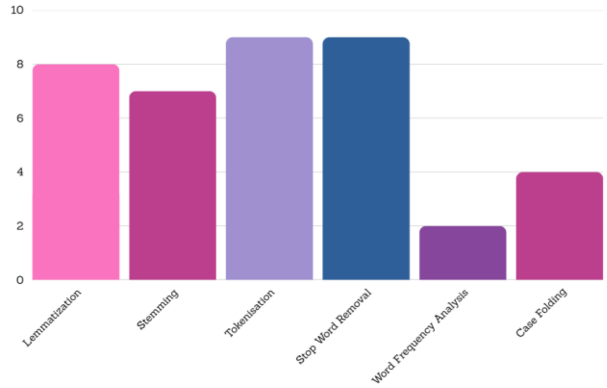

Figure 3. Studies by data processing techniques

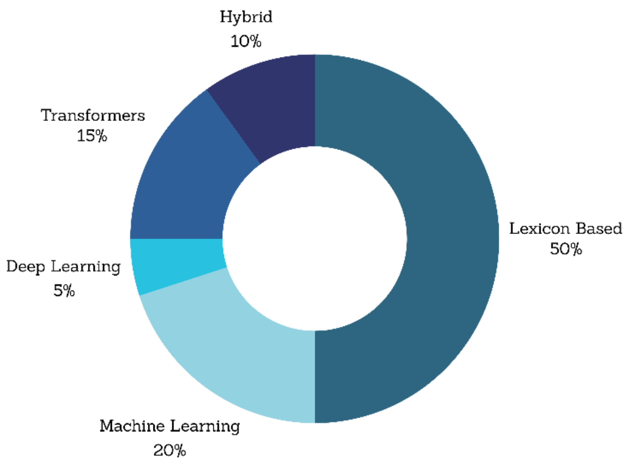

Figure 4. Number of studies per approach

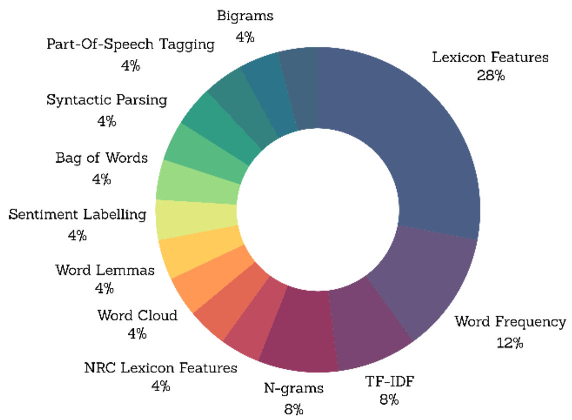

Figure 5. Features adopted

3.5.2 *Hybrid CNN-LSTM:* As explored in [4], CNN-LSTM represents a strong alternative for emotion classification. The hybrid approach surpassed traditional methods in detecting complex emotional states, making it a valuable tool for public health sentiment analysis [19]. The study [3] combined CNN-LSTM design with hyperparameter tunings and obtained a higher accuracy for the datasets related to Monkeypox. Similarly, [19] used a CNN-LSTM model, and it produced good results on accuracy, specificity, recall, and F1 score metrics. However, [19] recommended BERT algorithms in sentiment analysis of pandemic-related information. The study [6] combined CNN and LSTM, and it outperformed both individual models in terms of stability and generalisation. Higher accuracy and a smaller gap between the training and test accuracy demonstrate the improvement. However, all the models used in [6] struggled to interpret ironic expressions, and they suggested a need for more nuanced sentiment categorisation and word embedding techniques due to language complexity.

3.5.3 *Long Short-Term Memory:* Long Short-Term Memory (LSTM) is a subcategory of the recurrent neural network (RNN); its functionality is almost the same as RNN [19]. Its strength lies in its ability to detect contextual information in long texts, although it fails to capture crucial parts of the dataset [34]. The study [6] also used the LSTM model and compared it to the hybrid CNN-LSTM model. He suggested that with LSTM's ability to process sequential data, it is suitable for textual analysis. However, LSTMs focus on long-distance dependencies in text but struggle with detecting local features, which negatively affect the classification of sentiments [35].

3.5.4 *Flair:* Flair is an NLP framework with embeddings that are produced from a character language model trained by predicting the next word based on previous words [36]. Flair demonstrated higher performance than lexicon-based approaches in a study conducted by [14]. However, Flair's classification was constrained to binary polarity, limiting its ability to detect a broader range of nuanced emotions. This method was also used in [13] and the technique was limited to issues like bot detection and multilingual analysis.

3.5.5 *NRC Lexicon Method: The* National Research Council (NRC) is a sentiment lexicon containing over 14,000 English words and their associations with two sentiments (positive, negative) and eight basic emotions [37]. The study [8] found that NRC lexicon methods effectively classified the emotions and sentiments of frequently used words, as they identified these words and classified them according to emotions; fear became their most prevalent emotion. It was also able to discriminate between positive and negative sentiments expressed in the tweets. However, [8] mentioned that this method has difficulty handling figurative language, which significantly limits its ability to identify subtle emotional expressions.

3.5.6 *Valence Aware Dictionary for Sentiment Reasoning (VADER) Approach:* VADER is a popular and quicker lexicon-based model for sentiment analysis of various text forms [38]. VADER creates and then experimentally validates a sentiment lexicon that is particularly sensitive to microblog-like situations by combining qualitative and quantitative methodologies [39]. The study [11] emphasized that VADER is made to function well on social media material, which frequently uses slang, casual language, and acronyms. The study [32] clarified that VADER includes a wide range of lexicons, including slang words and abbreviations, making it well-suited for modern social media content. The study [11] also mentioned that their study did not specifically examine a broader spectrum of emotions like joy, sadness, rage, or fear, which could be required to capture nuanced emotions, and instead only concentrated on positive, negative, and neutral thoughts. Similarly, [13] used VADER, but the study did not account for bot-generated tweets, and it was limited to English-language tweets. The study [14] reported that VADER achieved a lower accuracy. These models could easily classify basic polarities (positive, neutral, or negative), but they had trouble differentiating between complex emotions, especially anxiety and fear. This drawback stems from their dependence on static lexicons, which are unable to adequately represent the dynamic and ever-changing nature of discourse on social media [14]. Similarly, [31] indicated that more tweets with neutral sentiments from TextBlob came out negative when VADER was employed.

3.5.7 *TextBlob:* TextBlob is a Python package that offers a straightforward API for exploring basic tasks in NLP [36]. Since TextBlob can be a faster library, data scientists prefer to use it, and its straightforward API makes it easier to perform many common text processing and NLP tasks, including language translation, Parts of Speech (POS) tagging, tokenisation, phrase extraction, classification, sentiment analysis, and more [39]. The study [13] used TextBlob in their sentiment analysis of monkeypox, but it did not account for bot-generated tweets. Moreover, [31] found that ML models trained on TextBlob-annotated data performed better than those trained on VADER or AFINN-annotated data. However, their study concluded that TextBlob relies on a fixed lexicon and does not incorporate intensifiers or negation handling as effectively as VADER.

3.5.8 *Traditional Machine Learning Models:* Traditional machine learning models were also used by several research studies, and these include:

3.5.8.1 *Random Forest (RF):* Random forest is an ensemble model that combines the output of subtrees to provide forecasts with a high degree of accuracy [31]. This algorithm creates a large number of classification decision trees, suggesting that the majority of the trees choose the class category [20]. The study [20] used RF, and it performed well. Similarly, [40] confirmed that ML methods like RF have good accuracy levels, although they are not popular. Moreover, [9] confirmed that RF also performed well in the classification of monkeypox tweets as compared to other ML models. However, [9] recommended more sophisticated ML models that can account for nuanced sentiments, sarcasm, and context-specific meanings to improve accuracy.

3.5.8.2 *Logistic Regression (LR):* Logistic regression is a technique that estimates the probability of an event happening [20]. Logistic regression links absolute dependent variables with one or more independent variables, using a logistic function to calculate probabilities. The dependent variable is usually called the target class [31]. Logistic regression was used in the study of [3] and it displayed the best results among all the machine learning models they had worked on; however, deep learning and hybrid models performed better than logistic regression. The study [31] confirmed that it is less prone to overfitting, especially when using regularisation techniques. However, their research found that it assumes a linear link between independent variables and the log odds of the dependent variable. This may not always hold and making it less suitable for handling complex data relationships compared to advanced models like deep learning. The study [20] also used logistic regression, and they discovered that it is slightly similar to the support vector machine model. However, they suggested the use of deep learning models as they assume they perform better than logistic regression and other traditional ML models.

3.5.8.3 *Support Vector Machine (SVM):* Support Vector Machine is a strong model that creates limits between classes by categorising data into one of the assumed classes [20]. It is known for its capability to work with high-dimensional data and complex distributions [41]. According to the literature review conducted by [5], a study employed SVM, multinomial Naïve Bayes, and rule-based classification methods to classify attitudes from a dataset of 134,194 Arabic tweets that had been automatically labelled with emojis and concluded that SVM outperformed the other methods. The study [42] mentioned that the accuracy of the SVM technique surpassed the other algorithms, including CNN and Deep Neural Networks (DNN). The study [20] testified that the model, which applied SVM, lemmatisation, CountVectorizer, and TextBlob annotation, emerged as the best model, with an accuracy of about 0.9348. However, [5] suggested that deep learning models such as CNN and LSTM may outperform SVM when handling complex text representations and [44] highlighted that SVMs are less efficient in managing extensive datasets compared to methods such as Random Forest.

3.5.8.4 *Naïve Bayes (NB):* Naïve Bayes is a probabilistic classifier that employs conditional probability to assess the likelihood of its input belonging to each class [20]. Removed text describing more information about how the algorithm works. The study [9] used NB for their sentiment analysis on monkeypox tweets, but it was the least performing model amongst the four they had worked on. Similarly, [20] and [3] realized that it could not extract relevant features from text embeddings correctly, and it could not detect the sequence of tweets to learn how the language used in the tweets changes over time.

3.5.8.5 *XGBoost:* eXtreme Gradient Boosting (XGBoost) is a gradient-boosting decision tree known for delivering good results in many ML tasks [44]. The study [20] used XGBoost, and it was the least-performing model among all the models they had integrated with CountVectorizer. These researchers in [20] discovered that when integrated with TF-IDF, XGBoost performed better than K-Nearest Neighbour but not better than the rest of the models they used. XGBoost was used as the meta-classifier within the ensemble framework by [45]. It had higher predictive power and was faster than traditional gradient tree-boosting algorithms.

3.5.8.6 *K Nearest Neighbours (KNNs):* K Nearest Neighbours (KNN) is a machine learning algorithm used for classification as well as in evidence retrieval, pattern recognition, and regression tasks [20]. The algorithm has numerous advantages, like training speed, ease of employment, and effectiveness on large datasets, making it a good choice for classification [46]. When integrated with CountVectorizer, KNN works better than XGBoost, but performs badly with TF-IDF [20]. KNN was tested in different datasets, and it performed well, especially with SMOTE-based feature selection [47]. The study [9] mentioned that KNN was also used to classify sentiment toward COVID-19 vaccines, and it helped analyse general sentiment. KNN was also evaluated by [45], and it achieved the lowest performance (in terms of accuracy and F1-score) compared to the other individual classifiers, indicating that it was more vulnerable to noise and less robust for the sentiment analysis tasks they had considered.

3.5.8.7 *Decision Trees (DT):* DT is an ML model used in classification and regression problems [31]. The study [18] incorporated DT and integrated it with Doc2Vec and Azure, but out of the 42 models they built, it was the least-performing model. Similarly, [3] used DT in their study, and it was the second-best model amongst the machine learning models they had used. However, deep learning and hybrid models performed better than the decision tree. The study [45] also evaluated Decision Trees, and they achieved a moderate performance.

3.5.8.8 *Multilayer Perception (MLP):* The Multilayer Perceptron (MLP) is a neural network model that uses a mathematical function to learn complex features in data. It follows a feedforward approach, combining inputs and weights in a weighted sum before applying an activation function [44]. The study [20] evaluated MLP and it was among the better-performing algorithms, but it was outperformed by SVM when combined with lemmatisation, CountVectorizer and TextBlob annotations. MLP was used by [48] and it had challenges of overfitting as the dataset was small and the model could not generalise the data.

3.5.8.9 *Bidirectional Encoder Representations from Transformers (BERT):* Bidirectional Encoder Representations from Transformers (BERT) is well-known for its capability to understand contextualised text, making it the best tool for detecting the distinctions of diverse programming languages [49]. A sentiment analysis of Greek clinical conversation was performed by [50] and BERT was the best-performing algorithm. The model attained an accuracy of 0.9548, showing its efficiency in detecting sentiment tones in clinical conversations. The study [51] also used BERT in a sentiment analysis of product designs, and it was the best amongst three other models. Similarly, [18] used BERT as a deep learning classifier for stance detection in tweets related to COVID-19 vaccination, and it outperformed all other classifiers. The study [32] used Sentence BERT (SBERT) for sentiment analysis. A pooling layer was added to BERT's output to create sentence embeddings, and fine-tuning was done using a Siamese network structure. SBERT gave results that were better than using unigrams and bigrams. Sentiment analysis of Portuguese tweets was the only challenge because of the limited availability of annotated datasets.

In summary, it is in the review that the studied articles showed a clear transition of sentiment classification approaches, from traditional ML approaches like Naïve Bayes and SVM towards deep learning (e.g., CNN-LSTM) and transformer-based approaches (e.g., BERT and SBERT) that provide better contextualization of sentiment classes, particularly when dealing with Twitter data for public health.

3.6 Frequently Used Methods.

A discussion of the frequently used methods was also conducted.

3.6.1 *SnScrape:* SnScrape is a Python library that utilises web scraping to collect data from social media platform webpages. [52]. For real-time data, [29] used SnScrape to retrieve tweets. SnScrape collects the URL, publication date, total of favourite tweets, total of retweets, username, the name linked with the username, tweet text, tweet ID, and user description [53]. This library is important for extracting targeted Twitter data for study purposes, as it can be used to search for tweets based on exact search terms and selected time frames [54].

3.6.2 *Twitter API and Hydrator App:* The Twitter API, which was published in 2006, was used to get information about the location and account holder's conversational data, as Snscrape cannot capture geographical information [29]. The study [18] also collected datasets every day from public tweets using the Twitter API. The hydrator app was used by [11] to attain tweets and information matching to each Tweet ID using the Twitter API. Similarly, [8] also extracted tweets from Twitter using Rtweet incorporated with RStudio. One of the challenges [18] was that Twitter's API limits researchers to searching public tweets from the past 7–9 days, preventing access to older live tweets.

3.6.3 *Twitter Academic API v2s:* Twitter's academic research API v2 was also used for data collection [14]. The Twitter academic API was accessed using R (V.4.0.5) programming language [30]. The study [19] and [13] also used the Twitter academic API using monkeypox-related keywords.

3.6.4 *Tweepy:* Tweepy is another Python library that interacts with the Twitter API and performs tasks such as tweet search, user data retrieval, and tweet posting. It controls Twitter's OAuth authentication protocol, which allows developers to confirm and approve their applications to access user data on Twitter [52]. To retrieve the tweets from Twitter, the Tweepy library was used via the Twitter developer account by [32]. The studies [55] and [20] used Tweepy to access the Twitter API and gather tweets and related information.

3.6.5 *Twint:* Twint, a Python Twitter scraping tool used without the need for the Twitter API, was used by [56]. The study [57] also retrieved tweets using the Twint project, as it offers the extraction of tweets with no limit. By using the geographical filter on Twint, [58] retrieved tweets from India.

To summarise, the authors depended on the Twitter API, sometimes through libraries like Tweepy; others had to use tools such as Snscrape and Twint to circumnavigate API restrictions and to capture historical or geotagged tweets. This is a testament to a recent, growing demand for an unfettered and flexible entrance to Twitter data in sentiment research.

3.7 Approaches and Features Adopted

The approaches and features adopted were discussed in detail.

3.7.1 *Lexicon-Based Approaches:* The lexicon-based approach makes use of sentiments to describe the polarity of text as it can be positive, negative or neutral [59]. The study [8] used the NRC lexicon in RStudio in the Syuzhet package to analyse tweet emotions on the dataset they used. Similarly, [13] used VADER and TextBlob as their analysers of unlabeled data. The studies [11], [14], [29], [31] also employed the lexical-based approaches. Lexicon Features, including word frequency, hashtag frequency, Term Frequency–Inverse Document Frequency, Part-of-Speech tagging, N-grams and Out-of-Vocabulary Words were used in Lexicon-based approaches by [11], [14], [29], [30], [31].

3.7.2 *Machine Learning Based Approaches:* This approach is when an algorithm trains a classifier from data that has been manually labelled [59]. The study used [18] Azure Machine Learning to calculate their sentiment scores. Moreover, [32] experimented using machine learning models. According to their studies, DT and LR perform better on TextBlob labelled datasets. Similarly, [20] also used a machine learning based approach and employed a word cloud in his study.

3.7.3 *Hybrid-Based Approaches:* This is usually the integration of both lexicon methods and machine learning or deep learning methods [59]. The study [32] used the Long Short-Term Memory-Gated Recurrent Neural Network, and it performed extremely well compared to the machine learning based approaches. Hybrid-based approach was also implemented by [4]. They used CNN-LSTM, and it performed better than the other models in other studies. Similarly, [19] used the same model, and it performed extremely well.

3.7.4 *Transformer-based Approaches:* Transformer-based approaches were created because it is hard to put the learning process of RNNs and LSTM in parallel form, and these models easily fail to recall after being trained on larger datasets [60]. The study [28] used BERT and BERTopic for topic modelling, and it performed well. Moreover, [33] used mUSE and SBERT for their classification. Transformers like Multilingual XLM-RoBERTA-base and DistilRoBERTA were employed by [27].

In summary, different studies have adopted sentiment analysis approaches such as lexicon-based, machine learning-based, hybrid, and transformer-based models. There is a shift in notice towards hybrid approaches and transformer-based models because of their superior performance in complex, multilingual settings.

3.8 Techniques used to Preprocess Twitter Data

The techniques used to preprocess Twitter Data were discussed.

3.8.1 *Tokenisation:* Tokenisation identifies linguistic units and converts them into numerical IDs for vectorisation and mathematical processing [61]. The studies [30] and [20] performed tokenisation as their first step in their preprocessing of Twitter data. Tokenisation was also performed in the studies of [19] together with [18] and [15]. Similarly, [10] also performed tokenisation using the VADER approach.

3.8.2 *Stemming:* Stemming groups similar words with the same root or base form by cutting off suffixes, keeping the core meaning unchanged [62]. Porter Stemmer was used by [30] to perform stemming in their studies. Port Stemmer was also used by [18] to reduce words into their base, word stem, or root form. Word stemming was also done by [20] and [19] to replace retrieved with the root word through the process.

3.8.3 *Lemmatisation:* Unlike stemming, lemmatisation produces a real word form as it removes the inflectional endings and gives back the base or dictionary form of the word, helping to reduce its variations [62]. The study [20] observed that lemmatisation models showed better results than the stemmed models in all the cases. The UDpipe library was used by [19] for lemmatisation. The studies [18] and [30] used lemmatisation to lessen inflected words correctly, making sure that the root word belongs to the language, using WordNetLemmatizer.

3.8.4 *Stop Word Removal*: Stop words are words that often occur in a manuscript but have no important semantic relation to their context [63]. Stop words were removed by [30] to get a consistent representation of samples. The study [20] removed stop words using a Python library called stopword, to avoid noise in the data. The Flair analyser was used by [12] to remove stop words. The study [3] used stop-word removal to filter the data by removing any unnecessary words, using a list of stop words. A SkLearn package, 'stopwords', to calculate tweet sentiment was used by [18] for stop word removal. The Natural Language Toolkit (NLTK) library functions were used by [32] for stop word removal. Elimination of stop words was performed automatically by the VADER operator.

3.8.5 *Case Folding and Normalisation:* Case folding is converting the letters of a word to lowercase [64]. Customised code was used to convert all strings to lowercase by [30]. After cleaning the data, [15] made sure all words in the tweet data were changed to lowercase. Similarly, [8] also changed tweets to lowercase.

3.8.6 Word Frequency Analysis: Word Frequency Analysis, also known as Term Frequency, is a method that involves counting the existence of every word within a dataset [65]. Exploratory data analysis was done by [20] using word frequency and word cloud. The study [19] performed a primary analysis to comprehend the meaning of words and how frequently they were used using a bag-of-words model on the dataset. Word frequency was analysed to get more information and insight based on the Word Frequency Table from the reference [18]. The study [66] used Natural Language Toolkit (NLTK) for word segmentation to find high-frequency vocabulary and content focus in the text. The study [8] analyzed data by calculating the frequency of each word that appeared and presented the results in a table and a word frequency diagram.

To conclude, tokenisation, stop word removal, lemmatisation, and stemming were the most common preprocessing steps. Many studies also involved case normalisation, emoji cleaning, and word frequency analysis, which are essential components in the preprocessing stage to clean noisy and irregular Twitter data for reliable sentiment classification.

3.9 Limitations of the study

Despite the findings, this study has limitations. The review only considered peer-reviewed articles in English. This may have left out important research published in other languages. Furthermore, some of the included studies used small or geographically limited datasets, which may affect the broader applicability of the results.

4 CONCLUSION

This review was initiated to examine the algorithms largely adopted in Twitter sentiment analysis of monkeypox. Data collection methods, adopted features and approaches, and preprocessing techniques were also areas of interest. The reviews show a tremendous transformation favouring the hybrid CNN-LSTM models and Transformer architectures like BERT and DistilRoBERTa for better contextual understanding and sentiment classification, as opposed to traditional machine learning models such as SVM or Naïve Bayes. Lexicon-based sentiment models such as VADER and TextBlob are still used; however, they lack nuanced emotion interpretation and information discernment.

Regarding data collection, most studies resorted to the Twitter API or Twitter Academic API operations, Twint, and SNScrape to circumvent limits on access and rates. Preprocessing techniques were performed to ensure correct signal extraction; since tweets contain highly noisy and informal syntax, the preprocessors included tokenisation, stop word removal, lemmatisation, stemming, case normalisation, and word frequency analysis. The popular features were n-gram features, TF-IDF, and lexicon scores to enhance the input quality for classifier building.

The most consistent and frequently observed trend across studies was the superior performance of hybrid and Transformer-based models in dealing with sentiment tasks in public health. However, limitations were also detected. Many studies did not conduct any multilingual analysis; few, if any, considered bot or ironic detection; and datasets used were restricted to English. Furthermore, most did not assess how well their models could stand the test of time or particular events.

In conclusion, the highest potential in health crisis-like sentiment analysis seems to be geared toward hybrids and Transformers, such as those for mpox. Future research should aim at handling data in multiple languages, detecting bots, treating sarcasm better, and refining the fine-tuning of the Transformer architecture for a more synchronous way of observing public health.

CREDIT AUTHOR STATEMENT

The review work "Twitter (X) Sentiment Analysis on Monkeypox" was made possible through the combined efforts of three authors: Hazel Chamboko (HC), Belinda Ndlovu (BN), and Kudakwashe Maguraushe (KM): Conceptualization; HC, BN, and KM: Preliminary draft writing; HC: Methodology; HC, BN, and KM: Software; HC: Validation; BN and KM: Writing, Reviewing, and Editing; BN and KM: Supervision.

COMPETING INTERESTS

There are no competing interests.

DECLARATION OF GENERATIVE AI AND AI-ASSISTED TECHNOLOGIES IN THE WRITING PROCESS

During the preparation phase of the manuscript, AI tools, including Grammarly, might have been applied toward language and grammar improvement. The authors confirm that all content, including ideas, interpretations, analyses, and conclusions, is their own and that no AI tools were used to generate any text, manipulate data, or perform any part of the literature review. In the end, the full version of the manuscript was reviewed and accepted entirely by the authors.

ACKNOWLEDGMENT

We would like to acknowledge the valuable insights, feedback and substantial edits of the reviewers, which significantly enhanced the quality of this manuscript.

REFERENCES

[1] B. Ncube, M. Dzikiti, A. Nyoni, M. Ncube, B. Mutunhu Ndlovu, and S. Dube, "Effectiveness of Machine Learning algorithms in predicting Monkey Pox (Mpox): A Systematic Literature Review," in *Proceedings of the International Conference on Industrial Engineering and Operations Management*, Michigan, USA: IEOM Society International, Jul. 2024. doi: 10.46254/EU07.20240072.

[2] WHO, "Multi-country outbreak of mpox," Jun. 2024.

[3] K. K. Mohbey, G. Meena, S. Kumar, and K. Lokesh, "A CNN-LSTM-Based Hybrid Deep Learning Approach for Sentiment Analysis on Monkeypox Tweets," *New Gener Comput*, vol. 42, no. 1, pp. 89–107, Mar. 2024, doi: 10.1007/s00354-023-00227-0.

[4] R. Olusegun, T. Oladunni, H. Audu, Y. A. O. Houkpati, and S. Bengesi, "Text Mining and Emotion Classification on Monkeypox Twitter Dataset: A Deep Learning-Natural Language Processing (NLP) Approach," *IEEE Access*, vol. 11, pp. 49882–49894, 2023, doi: 10.1109/ACCESS.2023.3277868.

[5] T. Alqurashi, "0292 www.etasr.com Alqurashi: Arabic Sentiment Analysis for Twitter Data: A Systematic Literature Review Arabic Sentiment Analysis for Twitter Data: A Systematic Literature Review," 2023. [Online]. Available: www.etasr.com

[6] Y. Wang, "Sentiment Analysis in Analyzing Monkeypox-Related Tweets Based on Deep Learning," in *Proceedings of the 1st International Conference on Engineering Management, Information Technology and Intelligence*, SCITEPRESS - Science and Technology Publications, 2024, pp. 215–220. doi: 10.5220/0012923200004508.

[7] A. Yazdani, M. Shamloo, M. Khaki, and A. Nahvijou, "Use of sentiment analysis for capturing hospitalized cancer patients' experience from free-text comments in the Persian language," *BMC Med Inform Decis Mak*, vol. 23, no. 1, Dec. 2023, doi: 10.1186/s12911-023-02358-2.

[8] J. Mantik *et al.*, "Analysis of twitter user sentiment on the monkeypox virus issue using the nrc lexicon," Online, 2023.

[9] M. Doifode and D. Vidya, "IJCRT24A3076," no. 3 March 2024, pp. 1–5, Mar. 2024.

[10] N. Thakur, "Sentiment Analysis and Text Analysis of the Public Discourse on Twitter about COVID-19 and MPox," *Big Data and Cognitive Computing*, vol. 7, no. 2, Jun. 2023, doi: 10.3390/bdcc7020116.

[11] N. Thakur, "Sentiment Analysis and Text Analysis of the Public Discourse on Twitter about COVID-19 and MPox," *Big Data and Cognitive Computing*, vol. 7, no. 2, Jun. 2023, doi: 10.3390/bdcc7020116.

[12] V. S. Dsouza *et al.*, "A sentiment and content analysis of tweets on monkeypox stigma among the LGBTQ+ community: A cue to risk communication plan," *Dialogues in Health*, vol. 2, Dec. 2023, doi: 10.1016/j.dialog.2022.100095.

[13] P. Rajkhowa *et al.*, "Factors Influencing Monkeypox Vaccination: A Cue to Policy Implementation," *J Epidemiol Glob Health*, vol. 13, no. 2, 2023, doi: 10.1007/s44197-023-00100-9.

[14] V. S. Dsouza *et al.*, "A sentiment and content analysis of tweets on monkeypox stigma among the LGBTQ+ community: A cue to risk communication plan," *Dialogues in Health*, vol. 2, Dec. 2023, doi: 10.1016/j.dialog.2022.100095.

[15] K. K. Agustiningsih, E. Utami, and H. Al Fatta, "Sentiment Analysis of COVID-19 Vaccine on Twitter Social Media: Systematic Literature Review," in *Proceedings - 2021 IEEE 5th International Conference on Information Technology, Information Systems and Electrical Engineering: Applying Data Science and Artificial Intelligence Technologies for Global Challenges During Pandemic Era, ICITISEE 2021*, Institute of Electrical and Electronics Engineers Inc., 2021, pp. 121–126. doi: 10.1109/ICITISEE53823.2021.9655960.

[16] A. Albladi, M. Islam, and C. Seals, "Sentiment Analysis of Twitter data using NLP Models: A Comprehensive Review," *IEEE Access*, 2025, doi: 10.1109/ACCESS.2025.3541494.

[17] A. Ligthart, C. Catal, and B. Tekinerdogan, "Systematic reviews in sentiment analysis: a tertiary study," *Artif Intell Rev*, vol. 54, no. 7, pp. 4997–5053, Oct. 2021, doi: 10.1007/s10462-021-09973-3.

[18] M. Qorib, T. Oladunni, M. Denis, E. Ososanya, and P. Cotae, "Covid-19 vaccine hesitancy: Text mining, sentiment analysis and machine learning on COVID-19 vaccination Twitter dataset," *Expert Syst Appl*, vol. 212, Feb. 2023, doi: 10.1016/j.eswa.2022.118715.

[19] O. Iparraguirre-Villanueva *et al.*, "The Public Health Contribution of Sentiment Analysis of Monkeypox Tweets to Detect Polarities Using the CNN-LSTM Model," *Vaccines (Basel)*, vol. 11, no. 2, Feb. 2023, doi: 10.3390/vaccines11020312.

[20] S. Bengesi, T. Oladunni, R. Olusegun, and H. Audu, "A Machine Learning-Sentiment Analysis on Monkeypox Outbreak: An Extensive Dataset to Show the Polarity of Public Opinion From Twitter Tweets," *IEEE Access*, vol. 11, pp. 11811–11826, 2023, doi: 10.1109/ACCESS.2023.3242290.

[21] D. Moher, A. Liberati, J. Tetzlaff, and D. G. Altman, "Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement," *International Journal of Surgery*, vol. 8, no. 5, pp. 336–341, 2010, doi: 10.1016/j.ijsu.2010.02.007.

[22] L. Xiu-Xia, Z. Ya, C. Yao-Long, Y. Ke-Hu, and Z. Zong-Jiu, "The reporting characteristics and methodological quality of Cochrane reviews about health policy research," Apr. 01, 2015, *Elsevier Ireland Ltd*. doi: 10.1016/j.healthpol.2014.09.002.

[23] T. Grammatopoulos, J. W. S. Hunter, Z. Munn, J. C. Stone, and T. H. Barker, "Reporting quality and risk of bias in JBI systematic reviews evaluating the effectiveness of interventions: A methodological review protocol," Jan. 13, 2023, *Lippincott Williams and Wilkins*. doi: 10.11124/JBIES-22-00317.

[24] J. M. Campbell, S. Kavanagh, R. Kurmis, and Z. Munn, "Systematic Reviews in Burns Care: Poor Quality and Getting Worse," 2017, *Lippincott Williams and Wilkins*. doi: 10.1097/BCR.0000000000000409.

[25] N. R. Haddaway, M. J. Page, C. C. Pritchard, and L. A. McGuinness, "PRISMA2020: An R package and Shiny app for producing PRISMA 2020-compliant flow diagrams, with interactivity for optimised digital transparency and Open Synthesis," *Campbell Systematic Reviews*, vol. 18, no. 2, Jun. 2022, doi: 10.1002/cl2.1230.

[26] A. P. Siddaway, A. M. Wood, and L. V Hedges, "How to Do a Systematic Review: A Best Practice Guide for Conducting and Reporting Narrative Reviews, Meta-Analyses, and Meta-Syntheses," *Annu. Rev. Psychol*, vol. 70, pp. 747–770, 2019, doi: 10.1146/annurev-psych-010418.

[27] E. Navori, "Using Natural Language Processing to Understand the Political Perception of Disease: a case study on Ebola, Monkeypox (Mpox), and Zika," 2023.

[28] Q. X. Ng, C. E. Yau, Y. L. Lim, L. K. T. Wong, and T. M. Liew, "Public sentiment on the global outbreak of monkeypox: an unsupervised machine learning analysis of 352,182 twitter posts," *Public Health*, vol. 213, pp. 1–4, Dec. 2022, doi: 10.1016/j.puhe.2022.09.008.

[29] P. Sv and R. Ittamalla, "What concerns the general public the most about monkeypox virus? – A text analytics study based on Natural Language Processing (NLP)," *Travel Med Infect Dis*, vol. 49, Sep. 2022, doi: 10.1016/j.tmaid.2022.102404.

[30] S. Albahli and M. Nawaz, "TSM-CV: Twitter Sentiment Analysis for COVID-19 Vaccines Using Deep Learning," *Electronics (Switzerland)*, vol. 12, no. 15, Aug. 2023, doi: 10.3390/electronics12153372.

[31] R. Marcec and R. Likic, "Using Twitter for sentiment analysis towards AstraZeneca/Oxford, Pfizer/BioNTech and Moderna COVID-19 vaccines," *Postgrad Med J*, vol. 98, no. 1161, pp. 544–550, Jul. 2022, doi: 10.1136/postgradmedj-2021-140685.

[32] A. A. Reshi *et al.*, "COVID-19 Vaccination-Related Sentiments Analysis: A Case Study Using Worldwide Twitter Dataset," *Healthcare (Switzerland)*, vol. 10, no. 3, Mar. 2022, doi: 10.3390/healthcare10030411.

[33] K. Garcia and L. Berton, "Topic detection and sentiment analysis in Twitter content related to COVID-19 from Brazil and the USA," *Appl Soft Comput*, vol. 101, Mar. 2021, doi: 10.1016/j.asoc.2020.107057.

[34] A. Ameur, S. Hamdi, and S. Ben Yahia, "Sentiment Analysis for Hotel Reviews: A Systematic Literature Review," *ACM Comput Surv*, vol. 56, no. 2, Feb. 2023, doi: 10.1145/3605152.

[35] Syaiful Imron, E. I. Setiawan, Joan Santoso, and Mauridhi Hery Purnomo, "Aspect Based Sentiment Analysis Marketplace Product

Reviews Using BERT, LSTM, and CNN," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 7, no. 3, pp. 586–591, Jun. 2023, doi: 10.29207/resti.v7i3.4751.

[36] M. El Azzouzi *et al.*, "Automatic de-identification of French electronic health records: a cost-effective approach exploiting distant supervision and deep learning models," *BMC Med Inform Decis Mak*, vol. 24, no. 1, Dec. 2024, doi: 10.1186/s12911-024-02422-5.

[37] J. Kunsabo and J. Dobša, "A Systematic Literature Review on Topic Modelling and Sentiment Analysis," Sep. 2022.

[38] K. Barik and S. Misra, "Analysis of customer reviews with an improved VADER lexicon classifier," *J Big Data*, vol. 11, no. 1, Dec. 2024, doi: 10.1186/s40537-023-00861-x.

[39] R. Hernández-Pérez, P. Lara-Martínez, B. Obregón-Quintana, L. S. Liebovitch, and L. Guzmán-Vargas, "Correlations and Fractality in Sentence-Level Sentiment Analysis Based on VADER for Literary Texts," *Information (Switzerland)*, vol. 15, no. 11, Nov. 2024, doi: 10.3390/info15110698.

[40] W. Aljedaani *et al.*, "Sentiment analysis on Twitter data integrating TextBlob and deep learning models: The case of US airline industry," *Knowl Based Syst*, vol. 255, Nov. 2022, doi: 10.1016/j.knosys.2022.109780.

[41] B. Ondara, S. Waithaka, J. Kandiri, and L. Muchemi, "Machine Learning Techniques, Features, Datasets, and Algorithm Performance Parameters for Sentiment Analysis: A Systematic Review," *Open Journal for Information Technology*, vol. 5, no. 1, pp. 1–16, Jul. 2022, doi: 10.32591/coas.ojit.0501.01001o.

[42] P. W. Hardjita, Nurochman, and R. Hidayat, "Sentiment Analysis of Tweets on Prakerja Card using Convolutional Neural Network and Naive Bayes," *IJID (International Journal on Informatics for Development)*, vol. 10, no. 2, pp. 82–91, Jan. 2022, doi: 10.14421/ijid.2021.3007.

[43] A. M. Alayba, V. Palade, M. England, and R. Iqbal, "Arabic Language Sentiment Analysis on Health Services."

[44] N. W. C. Mukura and B. Ndlovu, "Performance Evaluation of Artificial Intelligence in Decision Support System for Heart Disease Risk Prediction," 2023.

[45] E. Mankolli, "Reducing the complexity of candidate selection using Natural Language Processing," in *International Conference on Systems, Signals, and Image Processing*, IEEE Computer Society, 2022. doi: 10.1109/IWSSIP55020.2022.9854488.

[46] R. H. Hama Aziz and N. Dimililer, "SentiXGboost: enhanced sentiment analysis in social media posts with ensemble XGBoost classifier," *Journal of the Chinese Institute of Engineers, Transactions of the Chinese Institute of Engineers,Series A*, vol. 44, no. 6, pp. 562–572, 2021, doi: 10.1080/02533839.2021.1933598.

[47] Y. Handika, I. F. Hanif, and F. N. Hasan, "Analysis of Public Sentiment Towards POLRI's Performance using Naive Bayes and K-Nearest Neighbors," *IJID (International Journal on Informatics for Development)*, vol. 13, no. 1, pp. 386–399, Jun. 2024, doi: 10.14421/ijid.2024.4500.

[48] A. Ghallab, A. Mohsen, and Y. Ali, "Arabic Sentiment Analysis: A Systematic Literature Review," 2020, *Hindawi Limited*. doi: 10.1155/2020/7403128.

[49] M. Chowdhury, A. Laskar, T. Ahmad, and A. T. Wasi, "MysticCIOL@DravidianLangTech 2025: A Hybrid Framework for Sentiment Analysis in Tamil and Tulu Using Fine-Tuned SBERT Embeddings and Custom MLP Architectures," 2025.

[50] "JTPES-2024-4-6_3," *Journal of Theory and Practice of Engineering Science*, no. Volume 4 Issue6, 2024, Jun. 2024.

[51] M. E. Chatzimina, H. A. Papadaki, C. Pontikoglou, and M. Tsiknakis, "A Comparative Sentiment Analysis of Greek Clinical Conversations Using BERT, RoBERTa, GPT-2, and XLNet," *Bioengineering*, vol. 11, no. 6, Jun. 2024, doi: 10.3390/bioengineering11060521.

[52] M. K. Shaik Vadla, M. A. Suresh, and V. K. Viswanathan, "Enhancing Product Design through AI-Driven Sentiment Analysis of Amazon Reviews Using BERT," *Algorithms*, vol. 17, no. 2, Feb. 2024, doi: 10.3390/a17020059.

[53] Eleana Jerez-Villota, Graciela Guerrero, and Omar Quimbita, *Evaluation and Comparison of Tweppy and Snscrape Libraries for Efficient Data Gathering on X*, vol. 1190. in Lecture Notes in Networks and Systems, vol. 1190. Cham: Springer Nature Switzerland, 2024. doi: 10.1007/978-3-031-74825-7.

[54] S. Ong, J. Koo, K. A. Johannson, C. J. Ryerson, and G. C. Goobie, "Content Analysis of Idiopathic Pulmonary Fibrosis-related Information on Twitter," *ATS Sch*, vol. 2, no. 4, pp. 576–587, Dec. 2021, doi: 10.34197/ats-scholar.2022-0054OC.

[55] I. S. M. Fadhil, M. H. M. Yusof, I. A. Khalid, S. H. Teoh, and A. A. Almohammedi, "Sentiment analysis comparisons across selected ml models: application on Malaysia online banking twitter data," in *Procedia Computer Science*, Elsevier B.V., 2024, pp. 979–988. doi: 10.1016/j.procs.2024.10.326.

[56] R. Olusegun, T. Oladunni, H. Audu, Y. A. O. Houkpati, and S. Bengesi, "Text Mining and Emotion Classification on Monkeypox Twitter Dataset: A Deep Learning-Natural Language Processing (NLP) Approach," *IEEE Access*, vol. 11, pp. 49882–49894, 2023, doi: 10.1109/ACCESS.2023.3277868.

[57] P. SV *et al.*, "Twitter-Based Sentiment Analysis and Topic Modeling of Social Media Posts Using Natural Language Processing, to Understand People's Perspectives Regarding COVID-19 Booster Vaccine Shots in India: Crucial to Expanding Vaccination Coverage," *Vaccines (Basel)*, vol. 10, no. 11, Nov. 2022, doi: 10.3390/vaccines10111929.

[58] K. Rahul, B. R. Jindal, K. Singh, and P. Meel, "Analysing Public Sentiments Regarding COVID-19 Vaccine on Twitter," in *2021 7th International Conference on Advanced Computing and Communication Systems, ICACCS 2021*, Institute of Electrical and Electronics Engineers Inc., Mar. 2021, pp. 488–493. doi: 10.1109/ICACCS51430.2021.9441693.

[59] P. Sv, J. Tandon, Vikas, and H. Hinduja, "Indian citizen's perspective about side effects of COVID-19 vaccine – A machine learning study," *Diabetes and Metabolic Syndrome: Clinical Research and Reviews*, vol. 15, no. 4, Jul. 2021, doi: 10.1016/j.dsx.2021.06.009.

[60] Azeema Sadia, Fariha Khan, and Fatima Bashir, "An Overview of Lexicon-Based Approach For Sentiment Analysis," 2018.

[61] S. Tabinda Kokab, S. Asghar, and S. Naz, "Transformer-based deep learning models for the sentiment analysis of social media data," *Array*, vol. 14, Jul. 2022, doi: 10.1016/j.array.2022.100157.

[62] E. Gow-Smith, H. T. Madabushi, C. Scarton, and A. Villavicencio, "Improving Tokenisation by Alternative Treatment of Spaces," Apr. 2022, [Online]. Available: http://arxiv.org/abs/2204.04058

[63] R. Pramana, Debora, J. J. Subroto, A. A. S. Gunawan, and Anderies, "Systematic Literature Review of Stemming and Lemmatization Performance for Sentence Similarity," in *Proceedings of the 2022 IEEE 7th International Conference on Information Technology and Digital Applications, ICITDA 2022*, Institute of Electrical and Electronics Engineers Inc., 2022. doi: 10.1109/ICITDA55840.2022.9971451.

[64] A. Habberrih and M. Ali Abuzaraida, "Sentiment Analysis of Libyan Dialect Using Machine Learning with Stemming and Stop-words Removal," in *5TH INTERNATIONAL CONFERENCE ON COMMUNICATION ENGINEERING AND COMPUTER SCIENCE (CIC-COCOS'24)*, Cihan University-Erbil, 2024, pp. 259–264. doi: 10.24086/cocos2024/paper.1171.

[65] C. A. Nurhaliza Agustina, R. Novita, Mustakim, and N. E. Rozanda, "The Implementation of TF-IDF and Word2Vec on Booster Vaccine Sentiment Analysis Using Support Vector Machine Algorithm," in *Procedia Computer Science*, Elsevier B.V., 2024, pp. 156–163. doi: 10.1016/j.procs.2024.02.162.

[66] M. Kumar, L. Khan, and H.-T. Chang, "Evolving techniques in sentiment analysis: a comprehensive review," *PeerJ Comput Sci*, vol. 11, p. e2592, Jan. 2025, doi: 10.7717/peerj-cs.2592.

[67] H. Xu, R. Liu, Z. Luo, and M. Xu, "COVID-19 vaccine sensing: Sentiment analysis and subject distillation from twitter data," *Telematics and Informatics Reports*, vol. 8, Dec. 2022, doi: 10.1016/j.teler.2022.100016.