Research Article BEU Fen Bilimleri Dergisi 15(2): 865-883 (2026) DOI: 10.17798/bitlisfen.1842980

BİTLİS EREN UNIVERSITY
FEN BİLİMLERİ DERGİSİ
ISSN: 2147-3129 / e-ISSN: 2147-3188

SENTIMENT ANALYSIS OF MONKEYPOX TWEETS USING A HYBRID CNN-BiGRU BASED DEEP LEARNING MODEL

Ümit CAN

Munzur University, Computer Engineering Department, Tunceli, Türkiye, ucan@munzur.edu.tr

Article Info

Received: *December 15, 2025*
Revised: *May 2, 2026*
Accepted: *June 2, 2026*

Keywords
Deep Learning,
Machine Learning,
Monkeypox,
Sentiment Analysis,
Social Network.

ABSTRACT

Recently, epidemics have disrupted people's lives. Following the COVID-19 pandemic, cases of monkeypox, a rare disease, began to appear worldwide. These incidents raised societal tension and anxiety. Understanding public attitudes, sentiments, and perceptions regarding the epidemic is essential for developing and implementing effective policies and countermeasures. Online social media platforms are essential venues for expressing feelings and views. In particular, Twitter provides valuable information for measuring societal sentiment. Machine learning and deep learning methods have recently been widely used in social network analysis. Deep learning models, which deliver strong performance, particularly on complex and nonlinear problems, are used in natural language processing and text mining. In this study, a Convolutional Neural Network Bidirectional Gated Recurrent Unit (CNN-BiGRU) hybrid deep learning model was developed by combining a CNN and a BiGRU and applied for the monkeypox sentiment analysis. The dataset was labeled as positive, negative, or neutral with Valence Aware Dictionary and sEntiment Reasoner (VADER). In addition, FastText and Term Frequency-Inverse Document Frequency (TF-IDF) word-embedding methods were used for feature extraction, and the performance of machine-learning and deep-learning methods was evaluated separately. As a result of the experimental studies, the FastText-CNN-BiGRU model obtained the highest results with 0.9382 accuracy, 0.9347 macro precision, 0.9367 macro recall, 0.9355 macro F-score, and 0.9875 macro ROC-AUC. The FastText-CNN model was the second most successful, achieving 0.9303 accuracy, 0.9268 macro precision, 0.9280 macro recall, 0.9271 macro F-score, and 0.9862 macro ROC-AUC. The FastText-BiGRU model was the third most successful, achieving 0.9262 accuracy, 0.9224 macro precision, 0.9255 macro recall, 9236 macro F-score, and 0.9838 macro ROC-AUC.

1. INTRODUCTION

The monkeypox virus is a DNA virus belonging to the orthopoxvirus genus, which includes the variola virus that causes smallpox, the vaccinia virus used in smallpox vaccination, and the bovine pox virus that causes bovine pox [1]. The disease caused by this virus is called monkeypox. Monkeypox virus, a smallpox-like disease, was first identified in 1959 in monkeys housed at a research facility in Copenhagen, Denmark [2]. However, it was in the 1970s that it was identified as a human disease agent. From the 1970s to the present, cases have been reported in Central and West African countries [3]. Initially, very few cases were reported across different regions due to animal exports. However, cases have recently become more widespread, and the disease has been diagnosed in multiple regions and among diverse populations. The disease arrived in the United States in 2003 and was later transmitted to these countries by travelers from Nigeria to Israel, the United Kingdom, and Singapore between 2018 and 2019. Again, in 2021, monkeypox was seen in a person coming from Nigeria to the United States [4]. Due to the devastating effects of the coronavirus pandemic, monkeypox was seen as a new pandemic possibility worldwide and alerted the world. As of February 1, 2023, 85,531 monkeypox cases had been detected worldwide [5].

License: CC BY-NC 4.0

Epidemics have many economic and social impacts on society. Epidemics disrupt economic activities and cause economic costs, especially health costs [6]. Additionally, these diseases generate anxiety and tension within the community and a sense of threat. People in this situation exhibit behaviors that differ from those typical of the general population [7]. Therefore, it is crucial for both the individual and society to manage the emotional and psychological impacts of uncertainty and crises that arise during pandemics. At this point, it is essential to measure and monitor society's emotional state. Sentiment analysis is one method for measuring society's emotions and attitudes [8]. Sentiment analysis is a field of study that tries to identify people's feelings, attitudes, and thoughts toward products, services, events, news, issues, and assets [9]. This field is closely related to computational linguistics, natural language processing, and text mining. During the COVID-19 pandemic, Twitter was among the most important social media platforms on which the public expressed their opinions [10]. Therefore, many studies used this data to better understand the COVID-19 pandemic [11, 12]. During a pandemic, accurately assessing public sentiment or precisely gauging people's views can help organizations battling it develop effective policies [13]. For instance, the success of a widespread vaccination effort may depend on understanding public attitudes toward COVID-19 or monkeypox vaccines.

Various methods were used in sentiment analysis studies. In most studies, machine learning and deep learning methods successfully classified sentiment. In particular, deep learning methods have been more successful at solving various machine learning problems because they can effectively learn from multidimensional, nonlinear data. Therefore, these methods were successful in areas such as image processing, natural language processing, and text mining. Specifically, they construct a document representation for text analysis tasks by learning word embeddings from a corpus of significant texts [10].

Moreover, hybrid techniques have been developed and applied to enhance the effectiveness of deep learning models [14, 15]. This research designed a hybrid deep learning model, a CNN-BiGRU, to identify public sentiment regarding the monkeypox epidemic. Tweets concerning monkeypox were classified as positive, neutral, or negative using the VADER approach. BiGRU, CNN, Random Forest (RF), Gradient Boosting (GB), and Extreme Gradient Boosting (XGBoost) were employed as deep learning and machine learning techniques to assess the model's efficacy. Additionally, two word-embedding methods—FastText and TF-IDF—were employed to extract features from tweets. Using word embedding methods, more compact representations can be obtained from large-scale datasets without data loss, thereby reducing computational cost. The performance of networks with these representations as input is superior to that of networks without them [16]. The models were trained using the feature vectors obtained by these methods and then compared. Based on the classification success metrics, the FastText-CNN-BiGRU model achieved the best performance.

The main contributions of the present study are summarized as follows:

- A CNN-BiGRU hybrid model was developed for the monkeypox sentiment analysis task.
- Sentiment labels were extracted from the Twitter dataset using VADER.
- CNN, BiGRU, RF, XGBoost, and GB methods were compared with the proposed hybrid model.
- Experimental results were compared after the proposed model was trained using word embedding methods like FastText and TF-IDF.
- Experimental results demonstrated that the hybrid model formed by CNN and BiGRU models increased the success rate in the sentiment analysis task with the FastText word embedding technique.

2. BACKGROUND

2.1. Online Social Network

Online social networks enable people to communicate and interact with one another online. These platforms are online environments where users can share information and knowledge and engage in social interactions with others who share similar interests and goals. Social networks were created to cover the internet and mobile devices [17]. Users can share their unique content on social networks and communicate with others. Facebook, Twitter, YouTube, Telegram, TikTok, Instagram, and Pinterest are among the most commonly used social networking platforms. These social networks are how people

communicate today and have changed traditional forms of information sharing [18]. Figure 1 below illustrates the number of users of online social networking platforms, based on the total number of monthly active users worldwide as of January 2025 [19]. Facebook has the highest number of members, with nearly three billion users. YouTube follows with 2.5 billion members. With 556 million members, Twitter is among the top 15 online social platforms by membership.

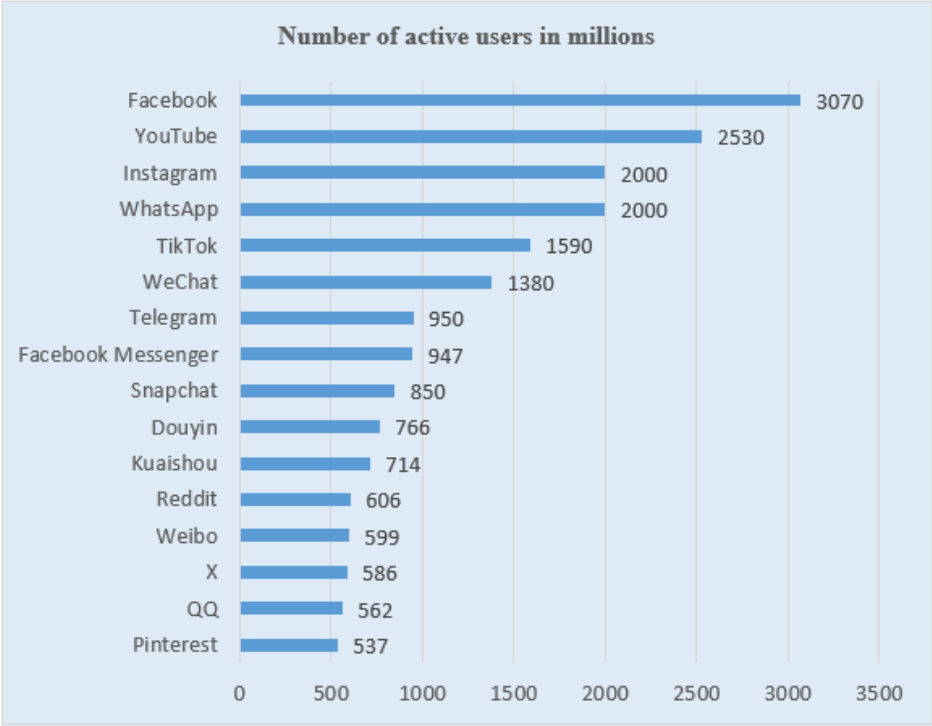


Figure 1. Number of users of online social networks worldwide as of January 2025.

Certain features make social networking sites essential and increase the likelihood of their use. These features include a profile page, a friend network, public commenting, and a private messaging system. Social networking sites are web-based services in which some of a user's information is publicly accessible to other users, enabling them to send friend requests, communicate, and share social media posts [20]. The accessibility and comprehensibility of interfaces and membership on social network platforms make these platforms appealing to many users. Today, many users interact on social networks using their real identities. In this way, social network users can easily share what is happening in their lives, current events, and interests with many of their friends, and express their opinions. They can also share social content, such as videos and images, and be aware of others' sharing. As a result, a large volume of data comprises users' opinions on social, political, and economic issues. Analyzing this data provides vital information for analyzing related issues [21].

2.2. Sentiment Analysis

In general, sentiment analysis—a significant subfield of text mining—attempts to infer from the texts that people write their opinions, attitudes, approaches, or sentiments. Sentiment analysis is a text-mining process that aims to classify the sentiment of a given text [22]. Early work on sentiment analysis, often referred to as emotional polarity, aims to classify text as positive, negative, or neutral. Sentiment analysis is an application of various natural language processing methods. Using text mining and computational linguistics tools helps researchers identify sentiment in text, classify it into sentiment categories, and determine the sentiment polarity of those categories. It analyzes the meaning expressed by a word, sentence, or text [23].

The data needed for sentiment analysis studies can be obtained from various media, such as opinions expressed in microblogs, product, movie, hotel, and restaurant reviews, opinions shared on platforms like Twitter and Facebook, and news articles. Due to sentiment analysis, the massive amounts of text

data collected from social media platforms are transformed into crucial variables that are used in research in a variety of areas, including the general level of happiness in society, customer opinions of goods and companies, the public's opinion of politicians or the government, and public opinions of any epidemic diseases. In sentiment analysis, which aims to extract the emotions users intend to express in text, techniques are broadly classified into two categories: lexicon-based and machine-learning-based [9]. Machine learning-based analysis techniques provide high performance by building supervised learning models. In particular, the success of deep learning models on nonlinear, complex data has brought them to the forefront of sentiment analysis studies [24].

2.3. Related Work

There are various methods in the literature for performing sentiment analysis, including traditional machine learning, deep learning, lexicon-based methods, hybrid methods combining machine learning and lexicon-based approaches, and graph-based methods [10, 25]. On the other hand, deep learning models have recently been applied to a range of problems with success. Because these methods yield better results than deterministic and statistical methods, they have recently been applied across many fields.

There have been numerous sentiment analysis studies on the COVID-19 outbreak; however, since the monkeypox outbreak is relatively recent, the studies in the literature are relatively new. Vijay et al. [26] utilized a dataset of month-oriented and day-oriented aggregated tweets about COVID-19 in India between November 2019 and May 2020. They conducted a sentiment analysis study on this dataset using the TextBlob API. In their sentiment analysis study, Boon and Skunkan [27] used the NRC sentiment lexicon application. With this method, they categorized tweets into various sentiment categories. Moreover, they modeled topics with latent Dirichlet allocation (LDA). Rustam et al. [28] compared the performance of supervised machine learning algorithms, RF, XGBoost classifier, Support Vector Classifier (SVC), Extra Trees Classifier (ETC), and Decision Tree (DT), with the Long Short-Term Memory (LSTM) deep learning method for the Covid-19 sentiment analysis task. They also employed techniques such as term TF-IDF, bag-of-words (BoW), and Global Vectors for Word Representation (GloVe) to evaluate the algorithms' collective performance. Nezhad and Deihimi [29] analyzed 803,278 Farsi tweets posted by users using a CNN-LSTM model to reveal public sentiment about COVID-19 and vaccines from Barekat, Pfizer/BioNTech, AstraZeneca/Oxford, Moderna, and Sinopharm. De Rosis et al. [30] proposed a new model to measure Italians' attitudes during the first four weeks of the COVID-19 pandemic. In this study, tweets were labeled as positive and negative with the Italian Sentimenta tool and manually. Then, predictions were made using the Bidirectional LSTM (Bi-LSTM) model. Basiri et al. [31] proposed a new deep learning-based fusion model for sentiment analysis. This study used a large dataset of tweets from eight countries. They compared their proposed model against Naïve Bayes (NB), Support Vector Machines (SVM), CNN, BiGRU, and FastText. Villavicencio et al. [32] used a dataset of tweets in English and Filipino languages to measure public attitudes toward COVID-19 vaccines in the Philippines. They conducted sentiment analysis using the NB algorithm. As a result of this study, they achieved an accuracy of 81.77%. In another 2021 study [33], a hybrid model, the Heterogeneous SVM (H-SVM), was proposed. This method was compared with the Recurrent Neural Network (RNN) and SVM methods. The hybrid model gave the highest result—Alam et al. [34] conducted sentiment analysis for COVID-19 vaccines with a deep learning-based model. In experimental studies using RNN, LSTM, and Bi-LSTM deep learning models, the Bi-LSTM model achieved the highest accuracy, with 90.83%. Sunitha et al. [35] measured public reactions to COVID-19 in India and Europe using TF-IDF, GloVe, Word2Vec, and FastText embedding feature-extraction methods, and GRU and Capsule Neural Network (CapsNet) deep learning methods. According to the experimental results, they obtained very successful results. In their sentiment analysis study, Chandrasekaran and Hemanth [36] used NB, SVM, Bi-LSTM, and TextBlob methods. According to the experimental results, the Bi-LSTM model achieved the highest accuracy, at 0.87. Singh et al. [37] proposed a new deep learning model for COVID-19 sentiment analysis. They compared their new model against LSTM-RNN, NB, RF, SVM, and Logistic Regression (LR). Their proposed hybrid model achieved the highest accuracy, at 84.56%. Singh et al. [38] employed supervised machine learning methods, including k-Nearest Neighbor (KNN), SVM, RF, and NB, as well as unsupervised methods such as TextBlob, AFINN, and Vader Sentiment, to classify tweets as positive or negative and measure

public sentiment regarding the repeal of Article 370. They achieved highly successful results in experimental studies. AlShammari and AlMansour [39] employed traditional LR, RF, and SVM machine learning models, as well as CNN-based deep learning models, to classify the sentiment of Arabic texts. Here, the emotions and opinions of the company's customers about a telecom company were classified. The CNN algorithm achieved successful results.

Because sentiment analysis on monkeypox is relatively new, only some studies employ deep learning and machine learning methods. Ng et al. [40] conducted sentiment analysis of Twitter posts from May 6, 2022, to July 23, 2022, using the Bidirectional Encoder Representations from Transformers (BERT) method for unsupervised learning. The stigma surrounding mpox among LGBTQ+ Twitter users was mapped and identified by Dsouza et al. [41]. For this task, they used sentiment analysis methods to understand the association of LGBTQ+ individuals with monkeypox on Twitter. Iparraguirre-Villanueva et al. [42] used a dataset of 84,018 tweets from Peru to assess public views on monkeypox. They performed sentiment analysis on this dataset using a CNN-LSTM hybrid model that combines CNN and LSTM architectures. Bengesi et al. [43] conducted a study using Twitter posts to determine public opinion on the monkeypox epidemic. Their study utilized VADER and TextBlob to categorize tweets as positive, negative, or neutral. Then, they used various feature extraction methods and applied RF, LR, kNN, SVM, Multi-layer Perceptron (MLP), NB, and XGBoost to perform sentiment classification. In their study, Mohbey et al. [44] employed a CNN-LSTM hybrid model to determine people's sentiment towards monkeypox. Compared with machine learning and deep learning methods, the hybrid model achieved the highest accuracy of 91%.

3. MATERIAL AND METHOD

The monkeypox virus, which has recently emerged, is among the viral diseases that have raised alarm in numerous regions following the COVID-19 pandemic. The virus, previously known and primarily observed in the Western and Central regions of Africa, has recently begun to affect diverse human populations in different ways. The monkeypox virus, which is found in rodents and the bodies of wild animals, spreads between humans through direct contact with infected animals. Previously, the virus spread from wild animals to humans was slow and limited. Because human-to-human transmission requires prolonged close contact, the disease tended to remain confined to specific areas rather than spread widely. However, the recent increase in its occurrence among people outside Africa has alarmed the world. Numerous scientific studies have been conducted on the virus to deepen understanding of its structure. In this study, a CNN-BiGRU deep learning model was developed for sentiment analysis of the monkeypox tweets. In addition, two different word-embedding models—FastText and TF-IDF were used to enhance the model's performance. The sentiment analysis model created in this study is shown in Figure 2. The steps comprising this model are explained in detail.

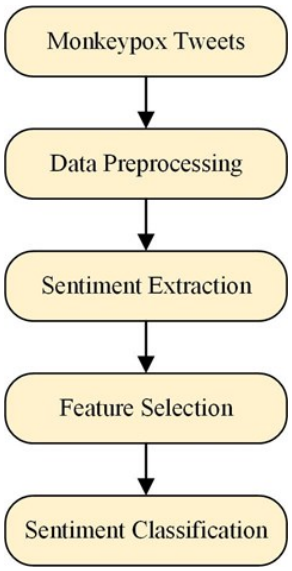


Figure 2. The general model of the monkeypox sentiment analysis.

3.1. Dataset and Data Preprocessing

In this study, a dataset comprising 95,246 tweets about monkeypox was utilized. The dataset comprises eight attributes: date, time, ID, tweet, language, reply count, retweet count, and like count. The dataset used is publicly available on Kaggle [45].

Figure 3 shows the stages of data preprocessing. During data preprocessing, unnecessary symbols, including URLs, punctuation, special characters, stop words, and emojis, were removed from the dataset. Next, tweets were tokenized and stemmed using the NLTK library. Subsequently, lemmatization was applied to the tweets.



Figure 3. Stages of data preprocessing.

After the data preprocessing stages, VADER labeled tweets as positive, negative, or neutral to extract sentiment. The frequency distribution of tweet sentiments is shown in Figure 4. Accordingly, the largest number of samples in the dataset is for the neutral (neu.) class, with 32,336. In addition, there are 25,213 samples in the positive (pos.) class and 23,003 samples in the negative (neg.) class. This distribution indicates that the dataset is imbalanced.

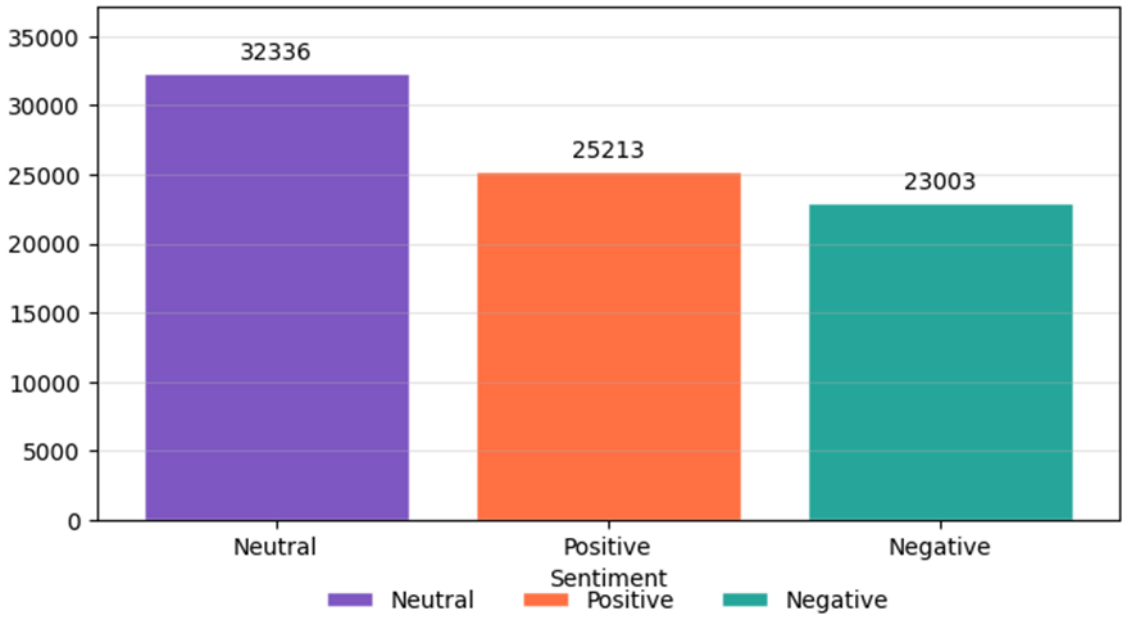


Figure 4. The frequency distribution of tweet sentiments.

Figure 5 shows the distribution of text lengths according to the sentiments of the tweets. The distributions of all three sentiment categories are right-skewed, indicating that most tweets consist of relatively short texts, while only a small number of tweets have longer text lengths. Neutral tweets are more concentrated in shorter text-length ranges, whereas positive and negative tweets display relatively broader distributions.

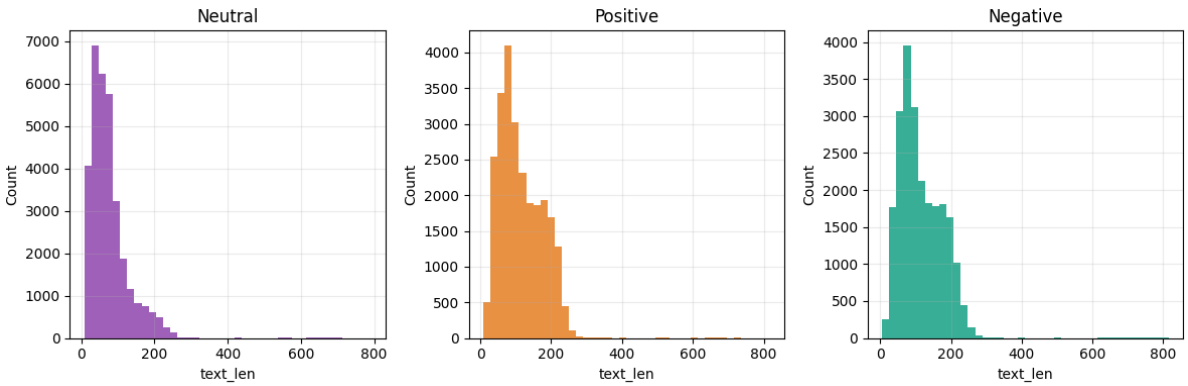


Figure 5. Distribution of text lengths according to the sentiments of the tweets.

Figure 6 shows the distribution of several character lengths according to the sentiments of the tweets. Similar to the text-length distribution, all three sentiment groups exhibit a right-skewed pattern, indicating that most tweets contain a relatively small number of words, while only a limited number of tweets have higher word counts. Neutral tweets are mostly concentrated in shorter word-count ranges. Positive and negative tweets show relatively broader distributions, with more observations extending toward higher word counts.

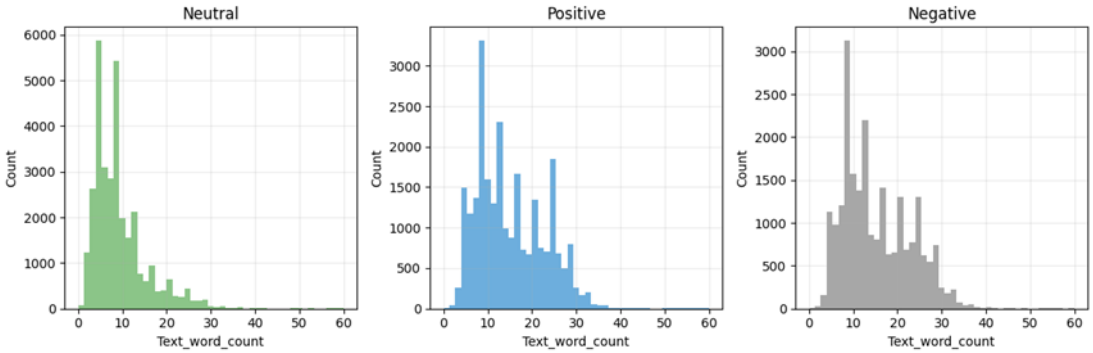


Figure 6. Distribution of several character lengths according to the sentiments of the tweets.

3.2. Word Embedding Methods Used in Sentiment Analysis

Text data cannot be used as raw data in machine learning and deep learning methods. To work with this data, it must first be converted into numerical information that the model can process. Instead of raw text data, word-embedding methods are used to derive representations of the data based on their semantic affinities. In word-embedding methods, a preprocessing step is used to obtain representations of text data provided as input to a model. With these methods, the words or sentences of the language are digitized and converted into vectors. Thus, the text data is represented numerically as a vector in a vector space. A neural network model learns representations from vectors assigned random values during the word-embedding process. It obtains representation vectors for use as input [46].

Word embeddings efficiently represent text using deep learning and machine learning methods. This study trained the designed hybrid model using FastText and TF-IDF, word-embedding representation schemes.

3.2.1. FastText

FastText, an extension of Word2Vec, employs a different approach by breaking words into n-grams and considering multiple letter combinations before feeding them into the neural network. In the N-gram expression, n represents the degree of repetition. In other words, the n expression here provides how many divisions will be made. It allows us to understand how many words or letters there are. After the training, word vectors will be obtained for all n-grams given in the training set. These words can now be represented more accurately, as n-grams of rare-frequency words are less likely to occur [47].

3.2.2. TF-IDF

The TF-IDF method is used to weight terms in text. It is used to express the importance of a term in the text statistically. The ratio of a term's frequency in a document to its total number of unique terms is called TF. The logarithm of the ratio of the total number of sentences in the document to the number of sentences containing the t-term is used to calculate IDF [48].

3.3. Baseline Models

This section provides a concise overview of the machine learning and deep learning models used in this study.

RF: This method, frequently used in classification and regression tasks, can be defined as a collection of tree-type classifiers. Incorporating randomization enhances the Bagging process [49]. This method enables the creation of a prediction set for decision trees by randomly sampling data from the dataset. RF aims to improve accuracy by producing multiple decision trees. In this method, each tree branch is based on the attributes in the dataset. Furthermore, RF can quickly adapt and adjust to the data [50]. It is also a suitable algorithm for classifying high-dimensional data. These properties have made it a popular method for text classification [51].

XGBoost: This model is commonly used with structured datasets and label classification to accelerate computation. XGBoost is a gradient-boosting-based decision tree ensemble designed for scalability and can handle problems at practical application scales while using limited resources. It improves gradient-boosting methods through parallel processing, tree pruning, and hardware optimizations. Furthermore, its parallel processing capacity enables it to run substantially faster, often 10 times faster than the GB method [52].

GB: This algorithm is another commonly used approach to solving machine learning challenges. The GB algorithm's basic concept is to build a stronger learner by iteratively enhancing weak learners. Models are trained sequentially, with each subsequent model addressing the shortcomings of the prior model. Initially, a decision tree is appropriate for the data, and a subsequent decision tree can be trained to forecast new outputs from previous outputs. The GB approach seeks to identify the overall model by minimizing the loss function [53].

CNN: This deep learning model is commonly used in NLP to capture local n-gram patterns and position-invariant features [54]. CNN is a multilayer perceptron architecture first proposed by Fukushima [55]. The traditional CNN architecture consists of five main layers. These are the input, convolution, pooling, fully connected, and output layers. Instead of general matrix multiplication, CNNs use convolution in one or more layers [56]. A CNN model has two essential advantages. The first is to reduce the filter weights, which is vital for learning. The second advantage is weight sharing [57, 58].

BiGRU: The GRU deep learning model is structurally similar to the LSTM model. The sole difference between the two types is the design of the memory cells. Because GRU contains fewer parameters, the memory cell structure is simpler and smaller. Unlike in LSTM, the processes of forgetting and overwriting are merged into a single gate, and the cell's complete state is output. The GRU model's primary advantages are its rapid training and its performance with limited data. GRU is a sort of unidirectional neural network architecture. Unidirectional neural network models are feedforward, with the output always coming from the previous layer. However, in sentiment analysis, the output of the current moment is linked to the preceding and following moments. The BiGRU model combines two GRU architectures [59, 60].

3.4. Developed Hybrid CNN-BiGRU-Based Deep Learning Model

This study aimed to analyze users' sentiments regarding monkeypox by examining tweets about the disease. A hybrid deep learning model was designed to achieve this objective by combining a CNN and a BiGRU model. BiGRU, the bidirectional variant of the GRU model, processes the text sequence in both forward and backward directions, enabling it to capture contextual information in both directions. This helps it to understand how a word at the end of a phrase affects a word at the beginning. Although the CNN architecture is commonly associated with image processing, its capacity to learn local features also makes it useful for text classification. Its capacity to detect distinct patterns in brief texts enables it

to represent semantic clusters accurately. BiGRU, the bidirectional version of the GRU model, processes the text sequence in both forward and backward directions, allowing for bidirectional acquisition of contextual information. This helps the model to learn how a word at the end of a sentence affects the word at the beginning. The proposed hybrid model additionally accepts pre-trained FastText embedding vectors as its first input. The language's semantic structure was represented in the model via the embedding layer, with the goal of achieving more accurate categorization while preserving contextual information. Another input vector was generated using TF-IDF, a statistical word embedding method. The primary objective of the designed hybrid model is to leverage the distinctive characteristics of CNNs and BiGRUs to enhance classification performance. Figure 7 illustrates the architecture of the developed hybrid model.

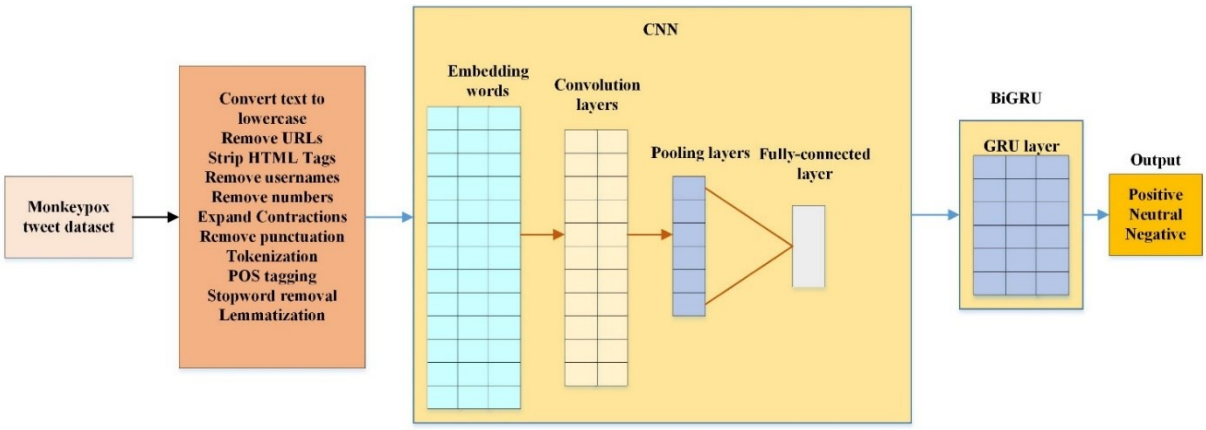


Figure 7. The architecture of the developed hybrid model.

In the proposed model, data preprocessing is performed on the raw Tweet dataset. Subsequently, the FastText and TF-IDF word-embedding approaches extract essential information from tweet texts. The CNN model extracts advanced features from the input data, while the BiGRU model captures contextual information for sentiment analysis.

Following the preprocessing stages, the proposed CNN-BiGRU model receives the twitter dataset as input. The design employs the Keras Tokenizer to tokenize the text, limiting it to the top 50,000 words with an out-of-vocabulary token, then pads it to a specified length of 100 tokens. After that, the sequences are passed on to an embedding layer that uses FastText and is initially loaded with cc.en.300 vectors, which have 300 dimensions. The embedding layer input size is set to the effective tokenizer vocabulary size, and the weights are initialized with the associated embedding matrix. After embedding, SpatialDropout1D (0.2) is used. The CNN block employs a Conv1D layer with 64 filters, kernel size 5, identical padding, and ReLU activation, followed by Batch Normalization. The generated feature maps are fed into a 64-unit BiGRU encoder (return_sequences=True, dropout=0.3, recurrent_dropout=0.3). Next, GlobalMaxPooling1D is used to generate a fixed-length representation, which is then fed through a classification head that includes Dense (64, ReLU) with Batch Normalization and Dropout (0.5), followed by Dense (32, ReLU) with Dropout (0.5). Finally, a Dense (3, softmax) layer generates positive, neutral, and negative classes.

To acquire objective and trustworthy performance predictions, a 10-fold cross-validation process was used. The dataset was divided while maintaining the class distribution in each fold (stratified sampling) for the purposes of cross-validation. In addition, an early stopping mechanism was applied during the training phase to prevent overfitting. Table 1 shows the hyperparameters of the hybrid model. The model was trained using the Adam optimizer (learning rate = 0.001) with a batch size of 256 for up to 10 epochs, and categorical cross-entropy was used as the loss function. Performance metrics were computed separately for each fold, and the reported results correspond to the mean and standard deviation obtained across the 10 folds. This approach provides a more stable and reliable evaluation of the model's performance.

Table 1. Hyperparameter values of the model

Epoch	**10**
Learning_rate	0.001
Optimizer	Adam
Filters in Conv1D layer	64
Kernel_size in Conv1D layer	5
Padding in Conv1D layer	Same
Activation function in Conv1D layer	Relu
Pool_size in MaxPooling1D layer	2
Number of units in the BiGRU layer	64
Dropout (SpatialDropout1D)	0.2
Dropout (GRU dropout)	0.3
Dropout (GRU recurrent_dropout)	0.3
Dropout (Dense layers)	0.5
Number of neurons in the dense layer 1	64
Number of neurons in the dense layer 2	32
Activation function in the Output layer	Softmax
Loss Function	categorical_crossentropy

4. THE EXPERIMENTAL RESULTS

In this study, the proposed CNN-BiGRU hybrid model and the other models used for performance comparison were implemented separately using the FastText and TF-IDF word-embedding methods. In addition, to obtain more reliable and consistent performance predictions, a 10-fold cross-validation strategy was employed. Moreover, a stratified cross-validation scheme was adopted to preserve the class distribution within each fold. With cross-validation, the dataset was divided into ten equal subsets. Then, one subset was reserved for testing. Of the ten subsets, the remaining nine were used for training, with the process repeated. The overall experimental results were obtained through averaging. The word-embedding methods used across all models included FastText and TF-IDF. FastText yielded superior results compared to other approaches.

(a) CNN-BiGRU

3x3		Actual		
		Pos.	Neu.	Neg.
Predicted	Pos.	0.93 ±0.01	0.02 ±0.00	0.05 ±0.02
	Neu.	0.02 ±0.00	0.95 ±0.00	0.02 ±0.00
	Neg.	0.06 ±0.01	0.03 ±0.00	0.93 ±0.02

(b) BiGRU

3x3		Actual		
		Pos.	Neu.	Neg.
Predicted	Pos.	0.93 ±0.01	0.03 ±0.01	0.06 ±0.01
	Neu.	0.01 ±0.00	0.93 ±0.01	0.02 ±0.00
	Neg.	0.06 ±0.01	0.04 ±0.01	0.92 ±0.01

(c) CNN

3x3		Actual		
		Pos.	Neu.	Neg.
Predicted	Pos.	0.91 ±0.01	0.02 ±0.01	0.07 ±0.01
	Neu.	0.02 ±0.00	0.95 ±0.01	0.02 ±0.01
	Neg.	0.07 ±0.01	0.03 ±0.01	0.92 ±0.02

(d) RF

3x3		Actual		
		Pos.	Neu.	Neg.
Predicted	Pos.	0.82 ±0.01	0.02 ±0.00	0.11 ±0.00
	Neu.	0.11 ±0.01	0.95 ±0.01	0.14 ±0.01
	Neg.	0.07 ±0.01	0.03 ±0.00	0.75 ±0.01

(e) XGBoost

3x3		Actual		
		Pos.	Neu.	Neg.
Predicted	Pos.	0.73 ±0.01	0.02 ±0.00	0.09 ±0.01
	Neu.	0.21 ±0.01	0.96 ±0.00	0.28 ±0.01
	Neg.	0.06 ±0.00	0.02 ±0.00	0.63 ±0.01

(f) GB

3x3		Actual		
		Pos.	Neu.	Neg.
Predicted	Pos.	0.62 ±0.01	0.03 ±0.00	0.13 ±0.01
	Neu.	0.32 ±0.01	0.96 ±0.01	0.38 ±0.00
	Neg.	0.06 ±0.00	0.02 ±0.00	0.49 ±0.01

Figure 8. Confusion matrices of the compared models

This study compared the developed hybrid model with well-known machine learning and deep learning models, including XGBoost, GB, RF, CNN, and BiGRU. Experimental results were obtained for each model using the following metrics: Accuracy, Macro-F1, Macro-Precision, Macro-Recall, Macro ROC–AUC, Micro-F1, Weighted-F1, Micro ROC–AUC, and Weighted ROC–AUC. Since FastText performs well, this section presents the FastText confusion matrix and success metrics for the methods.

Figure 8 displays the normalized confusion matrix results for the hybrid model alongside those of the other assessed models. The suggested FastText-CNN-BiGRU model for monkeypox sentiment classification demonstrated the highest overall performance among all approaches. This model accurately categorized 93% of positive tweets, 95% of neutral tweets, and 93% of negative tweets. The results demonstrate that the hybrid model produced equitable predictions across all classes. Following the hybrid model, the BiGRU model achieved the second-best results, accurately classifying 93% of positive tweets, 93% of neutral tweets, and 92% of negative tweets. A marginal decline was observed in the neutral class relative to the hybrid model. Nonetheless, the BiGRU model exhibited a well-rounded overall performance. The CNN model, another deep learning methodology, accurately categorized 91% of positive tweets, 95% of neutral tweets, and 92% of negative tweets. The CNN model achieved high accuracy on neutral samples; however, it showed a minor decrease in performance for the positive class compared to the hybrid design.

Conversely, classical machine learning models yielded inferior and more unbalanced outcomes relative to deep learning models. The RF model accurately categorized 82% of positive tweets, 95% of neutral tweets, and merely 75% of negative tweets. The XGBoost model accurately categorized 73% of positive tweets, 96% of neutral tweets, and 63% of negative tweets. The results demonstrate that the XGBoost model encountered notable difficulty in differentiating negative data. The GB model exhibited the poorest overall performance, accurately predicting 62% of positive tweets, 96% of neutral tweets, and merely 49% of negative tweets.

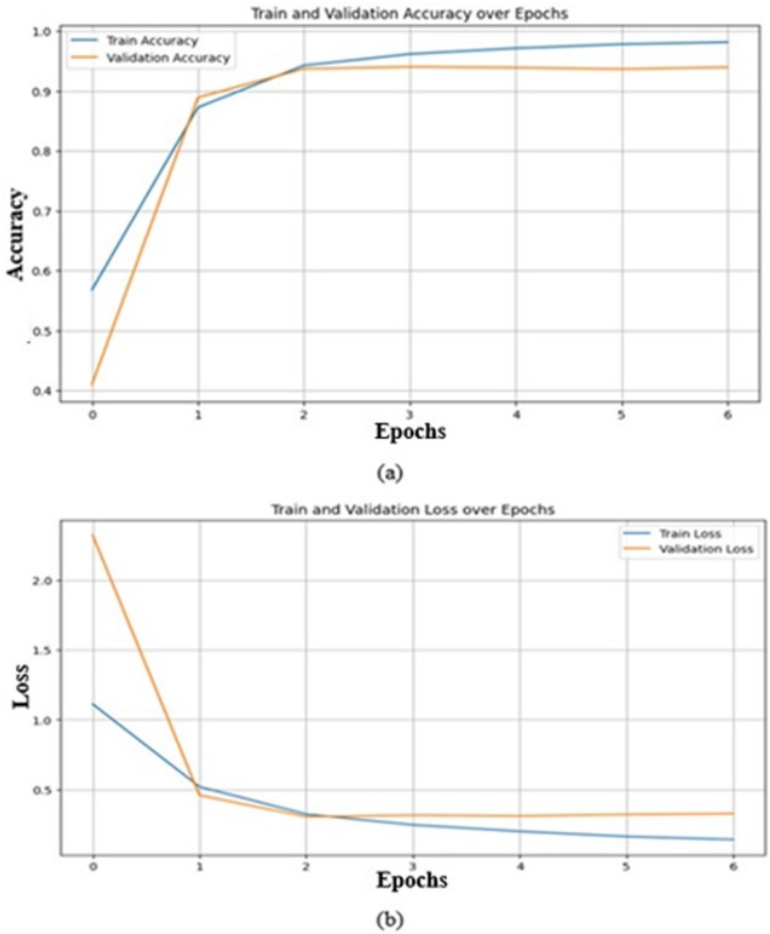


Figure 9. Training and validation accuracy and loss curves of the proposed FastText-CNN-BiGRU model.

The reason BiGRU is more successful than conventional GRU is that it processes input in both directions. BiGRU develops the context that the trained model can use. For example, knowing which words will come right after or just before a word in a sentence allows predictions to be made more successfully. BiGRU is an effective model for capturing bidirectional sequential dependencies between words and sentences.

Figures 9 shows the accuracy and loss curves for the proposed hybrid model trained with FastText word embeddings. The graph shown in Figure 9 (a) indicates that the hybrid model rapidly adapts to the data. In terms of training accuracy, the model exhibits a sharp increase in the first iteration and then continues with an almost stable learning curve after the second iteration. A similar pattern is observed in the validation curve. On the other hand, the loss graph confirms the behavior seen in the accuracy graph. In Figure 9 (b) the loss values decrease sharply during the first two iterations and then follow a plateau trend. Since the early stopping technique was applied during training, it can be observed that the model reached its optimal performance level in a relatively short period of time.

Figure 10 illustrates the ROC curve of the CNN-BiGRU model with FastText word embeddings as input. The figure indicates that the model exhibits exceptional classification performance. The proximity of the macro and micro average values indicates that the hybrid model achieves balanced performance across classes, with no bias towards any specific class. The elevated AUC value for the negative class indicates that the model has a marginally greater efficacy in classifying negative tweets. Likewise, the elevated AUC for the positive sentiment category indicates that the model excels at distinguishing positive tweets. Table 2 presents the models' accuracy, F1 (macro), precision (macro), recall (macro), and ROC-AUC (macro) results along with their standard deviations.

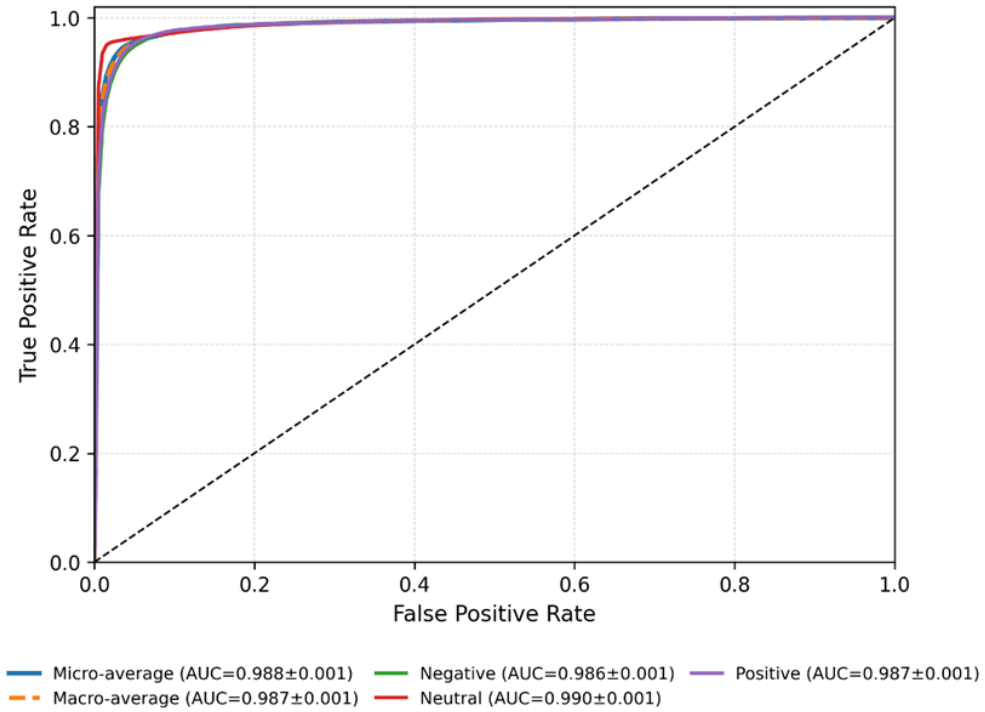


Figure 10. 10-fold CV ROC curve of the proposed hybrid model

Initially, as indicated by the results presented in Table 2, the standard deviation was minimal across all metrics. This outcome indicates that the model yielded stable, consistent results across all folds. The FastText-CNN-BiGRU model attained superior performance across all measures in this table. Following the hybrid model, the FastText-CNN model achieved the optimal outcome. The traditional classification method produced the most favorable outcome with the FastText-RF algorithm and the least favorable outcome with the FastText-GB algorithm.

Table 2. Model performance under 10-fold stratified cross-validation with FastText is reported as Mean ± Std.

Model	Accuracy (Mean ± Std)	F1 Macro (Mean ± Std)	Precision Macro (Mean ± Std)	Recall Macro (Mean ±Std)	ROC–AUC Macro (Mean ± Std)
CNN–BiGRU	0.9382 ± 0.0036	0.9355 ± 0.0037	0.9347 ± 0.0037	0.9367 ± 0.0039	0.9875 ± 0.0011
CNN	0.9303 ± 0.0039	0.9271 ± 0.0039	0.9268 ± 0.0037	0.9280 ± 0.0043	0.9862 ± 0.0009
BiGRU	0.9262± 0.0050	0.9236 ± 0.0051	0.9224 ± 0.0053	0.9255 ± 0.0049	0.9838 ± 0.0010
RF	0.8531 ±-0.0046	0.8462 ± 0.0047	0.8575 ± 0.0045	0.8408 ± 0.0048	0.9579 ± 0.0023
XGBoost	0.7961 ± 0.0058	0.7861 ± 0.0062	0.8249 ± 0.0058	0.7754 ± 0.0061	0.9195 ± 0.0031
GB	0.7218 ± 0.0043	0.7021 ± 0.0047	0.7694 ± 0.0049	0.6931 ± 0.0045	0.8664 ± 0.0034

Upon reviewing Table 3, it is evident that the CNN–BiGRU hybrid model attained superior performance across all specified criteria, as indicated by the results of the 10-fold stratified cross-validation. The model yielded results of around 93.8% for both F1 (Micro) and F1 (Weighted), while the ROC–AUC (Micro) and ROC–AUC (Weighted) values were determined to be 0.9874 and 0.9877, respectively.

The CNN model achieved a second-place ranking, with F1 (Micro) and F1 (Weighted) scores of approximately 93.03 and 0.9306, respectively. The ROC–AUC values, spanning from 0.9863 to 0.9867, demonstrate that the model exhibits robust discriminative capacity. The BiGRU model exhibited performance comparable to CNN; however, a marginal decline was observed in its micro and weighted F1 scores. Meanwhile, machine learning techniques yielded inferior outcomes compared with deep learning models. The RF model performed comparatively well; however, its F1 and ROC–AUC values were still inferior to those of the hybrid and other deep learning models.

Table 3. Other evaluation metric results with 10-fold CV with FastText are reported as Mean ± Std.

Model	F1 Micro (Mean ± Std)	F1 Weighted (Mean ± Std)	ROC–AUC Micro (Mean ± Std)	ROC–AUC Weighted (Mean ± Std)
CNN–BiGRU	0.9382 ± 0.0036	0.9384 ± 0.0031	0.9874 ± 0.0012	0.9877 ± 0.0011
CNN	0.9303 ± 0.0039	0.9306 ± 0.0039	0.9863 ± 0.0012	0.9867 ± 0.0009
BiGRU	0.9262 ± 0.0050	0.9266 ± 0.0049	0.9837 ± 0.0011	0.9842 ± 0.0008
RF	0.8462 ± 0.0047	0.8511 ± 0.0047	0.9563 ± 0.0022	0.9589 ± 0.0023
XGBoost	0.7961 ± 0.0058	0.7914 ± 0.0060	0.9193 ± 0.0028	0.9206 ± 0.0031
GB	0.7218 ± 0.0043	0.7106 ± 0.0046	0.8664 ± 0.0034	0.8679 ± 0.0034

As can be seen in Table 4, the standard deviation was initially low for all parameters. This result shows that the model produced reliable, consistent results across all folds. In this table, the TFIDF-CNN-BiGRU model outperformed all other models across all metrics. The hybrid model produced the best results, followed by the TFIDF-BiGRU and TFIDF-CNN models. With the TF-IDF-RF algorithm, the conventional classification approach yielded the best results, whereas the TF-IDF-GB technique yielded the worst results.

Table 4. The overall model's performance under 10-fold stratified cross-validation using TF-IDF, presented as Mean ± Std.

Model	Accuracy (Mean ± Std)	F1 Macro (Mean ± Std)	Precision Macro (Mean ± Std)	Recall Macro (Mean ± Std)	ROC–AUC Macro (Mean ± Std)
CNN–BiGRU	0.9286 ± 0.0062	0.9247 ± 0.0065	0.9240 ± 0.0062	0.9256± 0.0068	0.9857 ± 0.0019
BiGRU	0.9207 ± 0.0046	0.9170 ± 0.0046	0.9163 ± 0.0047	0.9181± 0.0045	0.9826 ± 0.0008
CNN	0.9037 ± 0.0048	0.8985 ± 0.0053	0.8999 ± 0.0051	0.8975± 0.0054	0.9746 ± 0.0018
RF	0.8538 ± 0.0051	0.8471 ± 0.0055	0.8585 ± 0.0047	0.8416± 0.0057	0.9579 ± 0.0019
XGBoost	0.7963 ± 0.0040	0.7862 ± 0.0044	0.8244 ± 0.0033	0.7757± 0.0045	0.9193 ± 0.0027
GB	0.7217 ± 0.0030	0.7021 ± 0.0036	0.7694 ± 0.0026	0.6930± 0.0035	0.8664 ± 0.0040

Similarly, Table 5 shows that the results for the models' F1 (Micro), F1 (Weighted), ROC-AUC (Micro), and ROC-AUC (Micro) are comparable to those in Table 4.

Table 5. Detailed aggregated evaluation metrics (10-fold CV, mean ± std) with TF-IDF.

Model	F1 (Micro) (Mean ± Std)	F1 (Weighted) (Mean ± Std)	ROC–AUC (Micro) (Mean ± Std)	ROC–AUC (Weighted) (Mean ± Std)
CNN–BiGRU	0.9286 ± 0.0062	0.9288 ± 0.0062	0.9863 ± 0.0019	0.9863 ± 0.0018
BiGRU	0.9207 ± 0.0046	0.9209 ± 0.0046	0.9832 ± 0.0007	0.9832 ± 0.0008
CNN	0.9037 ± 0.0048	0.9034 ± 0.0049	0.9757 ± 0.0017	0.9753 ± 0.0017
RF	0.8538 ± 0.0051	0.8519 ± 0.0053	0.9563 ± 0.0020	0.9588 ± 0.0019
XGBoost	0.7963 ± 0.0040	0.7916 ± 0.0043	0.9193 ± 0.0023	0.9204 ± 0.0027
GB	0.7217 ± 0.0030	0.7105 ± 0.0034	0.8692 ± 0.0033	0.8679 ± 0.0039

Figure 11 presents the SHapley Additive exPlanations (SHAP) results obtained for the FastText-CNN-BiGRU model, which achieved the best performance in the experimental studies, based on the sample data provided in Table 6. An example sentence, such as "great monkey pox chyna virus poor critter," shows that the word "great" contributes to the positive sentiment, whereas the word "poor" introduces a negative sentiment component. In the second example, expressions such as "vaccines" and "lowering immune system" appear to reinforce strong negative sentiment in the model's decision. In the final example, words such as "trust" and "protections" emerge as positive indicators that contribute to the prediction.

Table 6. Three examples were selected to explain the model using the SHAP method

True	Pred	Pred_prob	Text
Positive	Positive	0.9090	great monkey pox chyna virus poor critter
Positive	Positive	0.5764	angierasmussen wondering vaccines may lowering immune systems viruses running wild monkeypox polio etc
Positive	Positive	0.9066	cdcdirector employees cdcgov still working home children told go back school covidmonkeypox protections place resign rochelle trust



Figure 11. Results obtained using the SHAP method

The SHAP method assigns an importance value to each feature in the posts for a specific prediction. According to the SHAP results presented in Figure 11, the local contribution of each word within the

given example is shown. Even though all of the examples are classified as positive, some words may support the prediction of the positive class within the same instance, while others may produce the opposite effect [61]. For example, the fact that the term vaccines in the second example contributes negatively does not indicate that the model encodes the concept of vaccination as negative. Rather, it may reflect contextual patterns of uncertainty regarding vaccine accessibility, availability, and safety. Similarly, studies using mpox-related social network data have shown that discussions about vaccines are shaped by both the preventive public health rationale and the axes of hesitation, risk perception, and concerns about side effects [62]. Based on these findings, it can be argued that the model does not simply label these terms as inherently positive or negative, but instead learns the discourse patterns around these concepts in the training data across the dimensions of risk, protection, vulnerability, and uncertainty [63-64].

5. DISCUSSION

This study developed a CNN-BiGRU hybrid model trained with FastText and TF-IDF embeddings. Furthermore, to evaluate the efficacy of this model, CNN, BiGRU, RF, XGBoost, and GB methods were trained using the same FastText and TF-IDF embeddings. This facilitated an equitable assessment of the models' efficacy. The CNN-BiGRU model using FastText embeddings achieved superior performance across all metrics. This illustrates that integrating CNNs' capacity to extract local features with BiGRUs' bidirectional learning yields effective outcomes. CNN's superior performance over the BiGRU model in extracting features from audio and video data demonstrates its efficacy in this domain. The BiGRU model exhibited superior performance across the training and prediction stages for tasks such as text classification, natural language processing, time-series analysis, and sentiment analysis. The FastText-BiGRU model ranked second most effective. Moreover, the FastText approach demonstrated efficacy in sentiment identification. From the perspective of the CNN model, it is noted that CNN attains superior outcomes using TF-IDF. Upon analyzing machine learning techniques distinct from deep learning models, it has been noted that model efficacy markedly improves with TF-IDF word embeddings. RF is more effective with tabular datasets. RF operates using a combination of categorical and numerical variables, as determined by a decision tree. The RF algorithm has proven to be the most effective machine learning algorithm in conjunction with TF-IDF. This suggests that the TF-IDF framework is highly compatible with RF learning. Although the XGBoost algorithm outperformed RF with FastText, it underperformed when using TF-IDF embedding. The GB algorithm received the lowest ranking across all categories.

Sentiment analysis requires an understanding of the structure and context of language; factors such as word order, word choice, and sentence structure are essential. Deep learning models like BiGRU can better capture more complex and contextual language features. BiGRU can simultaneously examine the preceding and following parts of words and phrases in the text. This is an advantage in sentiment analysis tasks where the meaning of the word or phrase is often closely related to its context. Traditional methods, such as RF, XGBoost, and GB, can have greater difficulty capturing this relational information. Although traditional methods work with large datasets, they are less flexible and less capable of learning than deep learning models.

As a result, this study proposes a hybrid sentiment analysis model that combines a CNN and a BiGRU to leverage their respective strengths. With the CNN model, local information is sufficiently captured and learned. Then, the BiGRU layer, which captures long-range dependencies, helps the model capture long-term dependencies in the input sequence. FastText and TF-IDF word-embedding approaches are also employed in the proposed model to extract essential features for sentiment analysis. In this way, the success of popular word-embedding approaches within the proposed model can be compared.

In Table 7, the results obtained in this study are compared with previous studies on monkeypox analysis in the literature using the accuracy metric. According to this comparison, it can be seen that the hybrid model proposed in this study achieves better performance than many previous studies on monkeypox sentiment analysis. In particular, the successful extraction of word representations using FastText, the CNN model's ability to automatically discover complex features from these representations, and the BiGRU model's ability to effectively capture bidirectional relationships in data patterns help explain why the proposed model outperforms other models.

Table 7. Comparison of the hybrid CNN-Bi-GRU model with existing models in the literature

Model	Accuracy	Reported by
FastText + CNN-Bi-GRU	93.82%	This study
SVM, Lemmatization, CountVectorizer, and TextBlob	93.48%	[43]
DNN	92%	[61]
CNN-LSTM	91.00%	[44]
CNN-LSTM	83.00%	[42]

5.1. Limitations and Directions for Future Research

This study employed a hybrid model to assess societal emotional responses to the recent monkeypox outbreak, using FastText word embeddings and TF-IDF. The suggested model was evaluated against prominent deep learning and machine learning techniques. Despite the acquisition of encouraging data, the study possesses several limitations.

- This study employed the VADER approach for sentiment labeling of the posts. Future investigations may include manual annotation by subject matter experts and a comparison of automated and human-labeled data.
- The research employed a dataset from Twitter. To assess the generalizability of the suggested approach, data from additional social media platforms may be collected and analyzed.
- This study utilized FastText and TF-IDF vectorization techniques. Future research may evaluate the efficacy of the proposed model by employing various embedding methodologies, such as Word2Vec and GloVe.

6. CONCLUSION

Epidemics such as plague, cholera, SARS, avian flu, swine flu, and Ebola are health disasters that occur periodically throughout human history and deeply shake societies. The most recent of these epidemics, COVID-19, rapidly evolved into a pandemic after its emergence and spread worldwide. While the effects of the COVID-19 pandemic, which shook the world in a multidimensional way, are continuing, monkeypox has emerged as a new epidemic risk. An essential pillar of taking the necessary precautions against this new infectious disease risk is understanding society's views on the disease. This study proposes a CNN-BiGRU method, building on successful deep learning methods, to learn the community's opinions, emotions, and views regarding monkeypox. This hybrid model, which combines the strengths of CNNs and BiGRUs, was applied to this problem, yielding strong results. The proposed model was trained using word embedding techniques, including FastText and TF-IDF, and the experimental results were compared. In the comparisons with BiGRU, CNN, RF, GB, and XGBoost. The FastText-CNN model achieved 0.9303 accuracy, 0.9268 macro precision, 0.9280 macro recall, 0.9271 macro F-score, and 0.9862 macro ROC-AUC while the FastText-BiGRU model achieved 0.9262 accuracy, 0.9224 macro precision, 0.9255 macro recall, 0.9236 macro F-score, and 0.9838 macro ROC-AUC. The CNN-BiGRU model, which combines these two models and uses FastText embeddings, was the most successful method across all metrics, achieving 0.9382 accuracy, 0.9347 macro precision, 0.9367 macro recall, 0.9355 macro F-score, and 0.9875 macro ROC-AUC. RF, XGBoost and GB models were less successful than deep learning models. As a result of this study, a hybrid model was developed that successfully detects public sentiment regarding monkeypox. With the help of this model, the public's opinions about the disease can be determined, and thus, the relevant organizations can produce much more effective health policies.

Statement of Research and Publication Ethics

The study is complied with research and publication ethics.

Artificial Intelligence (AI) Contribution Statement

This manuscript was entirely written, edited, analyzed, and prepared without the assistance of any artificial intelligence (AI) tools. All content, including text, data analysis, and figures, was solely generated by the authors.

REFERENCES

[1] E. Alakunle, U. Moens, G. Nchinda, and M. I. Okeke, "Monkeypox virus in Nigeria: infection biology, epidemiology, and evolution," Viruses, vol. 12, no. 11, p. 1257, 2020.

[2] P. V. Magnus, E. K. Andersen, K. B. Petersen, and A. Birch-Andersen, "A pox-like disease in cynomolgus monkeys," Acta Pathologica Microbiologica Scandinavica, vol. 46, no. 2, pp. 156–176, 1959.

[3] E. M. Bunge et al., "The changing epidemiology of human monkeypox—A potential threat? A systematic review," PLoS Neglected Tropical Diseases, vol. 16, no. 2, p. e0010141, 2022.

[4] J. Quarleri, M. V. Delpino, and V. Galvan, "Monkeypox: considerations for the understanding and containment of the current outbreak in non-endemic countries," Geroscience, vol. 44, no. 4, pp. 2095–2103, 2022.

[5] "Mpox Outbreak Global Map," Centers for Disease Control and Prevention (CDC), 2022. [Online]. Available: https://www.cdc.gov/poxvirus/monkeypox/response/2022/world-map.html. [Accessed: Feb. 1, 2023].

[6] J. A. Roberts et al., "The study of infectious intestinal disease in England: socio-economic impact," Epidemiology and Infection, vol. 130, no. 1, p. 1, 2003.

[7] X. Li and H. Lyu, "Epidemic risk perception, perceived stress, and mental health during COVID-19 pandemic: A moderated mediating model," Frontiers in Psychology, vol. 11, p. 563741, 2021.

[8] A. Montoyo, P. Martínez-Barco, and A. Balahur, "Subjectivity and sentiment analysis: An overview of the current state of the area and envisaged developments," Decision Support Systems, vol. 53, no. 4, pp. 675–679, 2012.

[9] W. Medhat, A. Hassan, and H. Korashy, "Sentiment analysis algorithms and applications: A survey," Ain Shams Engineering Journal, vol. 5, no. 4, pp. 1093–1113, 2014.

[10] A. Giachanou and F. Crestani, "Like it or not: A survey of twitter sentiment analysis methods," ACM Computing Surveys, vol. 49, no. 2, pp. 1–41, 2016.

[11] U. Naseem, I. Razzak, M. Khushi, P. W. Eklund, and J. Kim, "COVIDSenti: A large-scale benchmark Twitter data set for COVID-19 sentiment analysis," IEEE Trans. Comput. Social Syst., vol. 8, no. 4, pp. 1003–1015, 2021.

[12] S. A. Morshed, S. S. Khan, R. B. Tanvir, and S. Nur, "Impact of COVID-19 pandemic on ride-hailing services based on large-scale Twitter data analysis," Journal of Urban Management, vol. 10, no. 2, pp. 155–165, 2021.

[13] 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 Medicine and Infectious Disease, 2022.

[14] H. Tekin, Ş. Kılıç, and Y. Doğan, "DiagNeXt: A Two-Stage Attention-Guided ConvNeXt Framework for Kidney Pathology Segmentation and Classification," Journal of Imaging, 11(12), 433, 2025.

[15] Y. Çelik, "Application of Deep Learning for Voice Command Classification in Turkish Language," Bitlis Eren Üniversitesi Fen Bilimleri Dergisi, 13(3), 701-708, 2024.

[16] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient estimation of word representations in vector space," in Proc. Workshop at ICLR, Scottsdale, AZ, USA, 2013.

[17] J. Brown, A. J. Broderick, and N. Lee, "Word of mouth communication within online communities: Conceptualizing the online social network," Journal of Interactive Marketing, vol. 21, no. 3, pp. 2–20, 2007.

[18] J. Heidemann, M. Klier, and F. Probst, "Online social networks: A survey of a global phenomenon," Computer Networks, vol. 56, no. 18, pp. 3866–3878, 2012.

[19] Statista, "Global social networks ranked by number of users," 2025. [Online]. Available: https://www.statista.com/statistics/272014/global-social-networks-ranked-by-number-of-users. [Accessed: May 30, 2025].

[20] A. Keenan and A. Shiri, "Sociability and social interaction on social networking websites," Library Review, vol. 58, no. 6, pp. 438–450, 2009.

[21] M. L. Kasavana, K. Nusair, and K. Teodosic, "Online social networking: redefining the human web," Journal of Hospitality and Tourism Technology, vol. 1, no. 1, pp. 68–82, 2010.

[22] D. M. E. D. M. Hussein, "A survey on sentiment analysis challenges," Journal of King Saud University – Engineering Sciences, vol. 30, no. 4, pp. 330–338, 2018.

[23] M. Wankhade, A. C. S. Rao, and C. Kulkarni, "A survey on sentiment analysis methods, applications, and challenges," Artificial Intelligence Review, vol. 55, no. 7, pp. 5731–5780, 2022.

[24] L. Zhang, S. Wang, and B. Liu, "Deep learning for sentiment analysis: A survey," Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, vol. 8, no. 4, p. e1253, 2018.

[25] S. Bhuta, A. Doshi, U. Doshi, and M. Narvekar, "A review of techniques for sentiment analysis of twitter data," in Proc. 2014 Int. Conf. Issues and Challenges in Intelligent Computing Techniques (ICICT), 2014, pp. 583–591.

[26] T. Vijay, A. Chawla, B. Dhanka, and P. Karmakar, "Sentiment analysis on covid-19 twitter data," in Proc. 2020 5th IEEE Int. Conf. Recent Advances and Innovations in Engineering (ICRAIE), 2020, pp. 1–7.

[27] S. Boon-Itt and Y. Skunkan, "Public perception of the COVID-19 pandemic on Twitter: sentiment analysis and topic modeling study," JMIR Public Health and Surveillance, vol. 6, no. 4, p. e21978, 2020.

[28] F. Rustam et al., "A performance comparison of supervised machine learning models for Covid-19 tweets sentiment analysis," PLOS ONE, vol. 16, no. 2, p. e0245909, 2021.

[29] Z. B. Nezhad and M. A. Deihimi, "Twitter sentiment analysis from Iran about COVID 19 vaccine," Diabetes & Metabolic Syndrome: Clinical Research & Reviews, vol. 16, no. 1, p. 102367, 2022.

[30] S. De Rosis, M. Lopreite, M. Puliga, and M. Vainieri, "The early weeks of the Italian Covid-19 outbreak: sentiment insights from a Twitter analysis," Health Policy, vol. 125, no. 8, pp. 987–994, 2021.

[31] M. E. Basiri, S. Nemati, M. Abdar, S. Asadi, and U. R. Acharya, "A novel fusion-based deep learning model for sentiment analysis of COVID-19 tweets," Knowledge-Based Systems, vol. 228, p. 107242, 2021.

[32] C. Villavicencio et al., "Twitter sentiment analysis towards covid-19 vaccines in the Philippines using naïve bayes," Information, vol. 12, no. 5, p. 204, 2021.

[33] H. Kaur, S. U. Ahsaan, B. Alankar, and V. Chang, "A proposed sentiment analysis deep learning algorithm for analyzing COVID-19 tweets," Information Systems Frontiers, pp. 1–13, 2021.

[34] K. N. Alam et al., "Deep learning-based sentiment analysis of COVID-19 vaccination responses from Twitter data," Computational and Mathematical Methods in Medicine, vol. 2021, 2021.

[35] D. Sunitha, R. K. Patra, N. V. Babu, A. Suresh, and S. C. Gupta, "Twitter sentiment analysis using ensemble based deep learning model towards COVID-19 in India and European countries," Pattern Recognition Letters, vol. 158, pp. 164–170, 2022.

[36] G. Chandrasekaran and J. Hemanth, "Deep learning and TextBlob based sentiment analysis for coronavirus (COVID-19) using twitter data," International Journal on Artificial Intelligence Tools, vol. 31, no. 01, p. 2250011, 2022.

[37] C. Singh, T. Imam, S. Wibowo, and S. Grandhi, "A deep learning approach for sentiment analysis of COVID-19 reviews," Applied Sciences, vol. 12, no. 8, p. 3709, 2022.

[38] S. Singh, H. Kaur, R. Kanozia, and G. Kaur, "Empirical analysis of supervised and unsupervised machine learning algorithms with aspect-based sentiment analysis," Applied Computer Systems, vol. 28, no. 1, pp. 125–136, 2023.

[39] N. AlShammari and A. AlMansour, "Aspect-based sentiment analysis and location detection for Arabic language tweets," Applied Computer Systems, vol. 27, no. 2, pp. 119–127, 2022.

[40] 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, 2022.

[41] 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, p. 100095, 2023.

[42] O. Iparraguirre-Villanueva et al., "The public health contribution of sentiment analysis of Monkeypox tweets to detect polarities using the CNN-LSTM model," Vaccines, vol. 11, no. 2, p. 312, 2023.

[43] 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.

[44] 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 Generation Computing, pp. 1–19, 2023.

[45] "Monkeypox Tweets," Kaggle dataset. [Online]. Available: https://www.kaggle.com/datasets/vencerlanz09/monkeypox-tweets. [Accessed: Dec. 5, 2025].

[46] S. Wang, W. Zhou, and C. Jiang, "A survey of word embeddings based on deep learning," Computing, vol. 102, pp. 717–740, 2020.

[47] P. Bojanowski, E. Grave, A. Joulin, and T. Mikolov, "Enriching word vectors with subword information," Transactions of the Association for Computational Linguistics, vol. 5, pp. 135–146, 2017.

[48] H. C. Wu, R. W. P. Luk, K. F. Wong, and K. L. Kwok, "Interpreting TF-IDF term weights as making relevance decisions," ACM Trans. Inf. Syst., vol. 26, no. 3, pp. 1–37, 2008.

[49] L. Breiman, "Random forests," Machine Learning, vol. 45, pp. 5–32, 2001.

[50] G. Biau and E. Scornet, "A random forest guided tour," Test, vol. 25, pp. 197–227, 2016.

[51] B. Xu, X. Guo, Y. Ye, and J. Cheng, "An improved random forest classifier for text categorization," Journal of Computers, vol. 7, no. 12, pp. 2913–2920, 2012.

[52] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in Proc. 22nd ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining (KDD), San Francisco, CA, USA, 2016, pp. 785–794.

[53] L. Brown and C. Mues, "An experimental comparison of classification algorithms for imbalanced credit scoring data sets," Expert Systems with Applications, vol. 39, pp. 3446–3453, 2012.

[54] S. L, Ramaswamy, and J. Chinnappan, "Review on positional significance of LSTM and CNN in the multilayer deep neural architecture for efficient sentiment classification". Journal of Intelligent & Fuzzy Systems, 45(4), 6077-6105, 2023.

[55] K. Fukushima, "Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position," Biological Cybernetics, vol. 36, no. 4, pp. 193–202, 1980.

[56] Y. Dogan, "A novel pruning-enhanced hybrid approach for efficient and accurate brain tumor diagnosis". Biomedical Signal Processing and Control, 112, 108466, (2026).

[57] L. Alzubaidi et al., "Review of deep learning: Concepts, CNN architectures, challenges, applications, future directions," Journal of Big Data, vol. 8, pp. 1–74, 2021.

[58] Y. Dogan, "An innovative approach for parkinson's disease diagnosis using CNN, NCA, and SVM". Neural Computing and Applications, 36(32), 20089-20110, 2024.

[59] X. Zhang, F. Guo, T. Chen, L. Pan, G. Beliakov, and J. Wu, "A brief survey of machine learning and deep learning techniques for e-commerce research," Journal of Theoretical and Applied Electronic Commerce Research, vol. 18, no. 4, pp. 2188–2216, 2023.

[60] M. S. Islam, M. N. Kabir, N. A. Ghani, K. Z. Zamli, N. S. A. Zulkifli, M. M. Rahman, and M. A. Moni, "Challenges and future in deep learning for sentiment analysis: a comprehensive review and a proposed novel hybrid approach," Artificial Intelligence Review, vol. 57, no. 3, p. 62, 2024.

[61] H., Gharaibeh et al. "Arabic sentiment analysis of Monkeypox using deep neural network and optimized hyperparameters of machine learning algorithms," Social Network Analysis and Mining, 14(1), 30, 2024.

[62] S. M. Lundberg and S. I. Lee, "A unified approach to interpreting model predictions," Advances in Neural Information Processing Systems, vol. 30, pp. 4765–4774, 2017.

[63] P. Rajkhowa, V. S. Dsouza, R. Kharel, K. Cauvery, B. R. Mallya, D. S. Raksha, et al., "Factors influencing monkeypox vaccination: a cue to policy implementation," Journal of Epidemiology and Global Health, vol. 13, no. 2, pp. 226–238, 2023.

[64] E. T. Overton, S. J. Lawrence, J. T. Stapleton, H. Weidenthaler, D. Schmidt, B. Koenen, et al., "A randomized phase II trial to compare safety and immunogenicity of the MVA-BN smallpox vaccine at various doses in adults with a history of AIDS," Vaccine, vol. 38, no. 11, pp. 2600–2607, 2020.