Decision Analytics Journal 7 (2023) 100243

Contents lists available at ScienceDirect

Decision Analytics Journal

journal homepage: www.elsevier.com/locate/dajour

A hybrid deep learning approach for detecting sentiment polarities and knowledge graph representation on monkeypox tweets

Gaurav Meena [a], Krishna Kumar Mohbey [a,*], Sunil Kumar [b], K. Lokesh [a]

[a] Department of Computer Science, Central University of Rajasthan, Ajmer, India
[b] School of Business, Woxsen University, Hyderabad, Telangana, India

ARTICLE INFO

Keywords:
Monkeypox
Knowledge graph
Long Short-Term Memory Networks (LSTM)
Deep learning
Sentiment polarities

ABSTRACT

People have recently begun communicating their thoughts and viewpoints through user-generated multimedia material on social networking websites. This information can be images, text, videos, or audio. With the help of knowledge graphs, it is possible to extract organized knowledge from texts and images to aid in semantic analysis. Recent years have seen a rise in the frequency of occurrence of this pattern. Twitter is one of the most extensively utilized social media sites, and it is also one of the finest locations to get a sense of how people feel about events that are linked to the Monkeypox sickness. This is because tweets on Twitter are shortened and often updated, both of which contribute to the platform's character. The fundamental objective of this study is to get a deeper comprehension of the diverse range of reactions people have in response to the presence of this condition. This study focuses on determining what individuals think about monkeypox illnesses, presenting a hybrid technique based on Convolutional Neural Networks (CNN) and Long Short-Term Memory Networks (LSTM). We have considered all three possible polarities of a user's tweet: positive, negative, and neutral. Knowledge graphs are embedded in various healthcare applications to provide improved data representation and knowledge inference, and they have been shown to be helpful in healthcare analytics. We describe in this study a knowledge graph of related events based on Twitter data, which provides a real-time and eventful source of new information. The recommended model's accuracy was 94% on the monkeypox tweet dataset. Other performance metrics such as accuracy, recall, and F1-score were utilized to test our models and results in the most time and resource-effective manner. The findings are then compared to more traditional approaches to machine learning. In addition, the ability to recognize semantic information has been built into the use of knowledge graphs. The findings of this research contribute to an increased awareness of monkeypox infection in the general population.

1. Introduction

Given the growing reliance on digital media platforms for health-related information, Twitter has been recognized as a potent tool for distributing real-time information to raise public awareness. Past virus outbreaks, like COVID-19, Ebola, Zika virus, and flu, were linked to several studies that examined the multi-modal components of Tweets to deduce the various characteristics of tweets about the corresponding outbreaks [1]. Society has started discussing monkeypox because of the increase in disease cases and the different nations' recommendations, efforts, and measures taken in response. People currently spend more time than ever before online, especially on social media sites, and their daily lives revolve around these activities, which are central to the "Internet of Everything" culture of today [2]. The scientific community has recently shown a great deal of interest in mining social media conversations, such as Tweet mining. Twitter can be used as a data source for a wide range of studies and use-case situations, including research into the dialogue paradigms involved and patterns of information-seeking and sharing on the social media platform. Twitter data analysis has been helpful in public health studies in the past, particularly about infectious diseases [3,4]. One method to manage and prevent epidemics is to track and analyze social media for spreading infectious illnesses. These platforms' feedback mechanisms, such as the comments that are left in response to them, aid in determining public opinion. As a result, social media data can be analyzed for syndrome surveillance to address public health-related issues and shape public perception through web-based information. Twitter data was used for scientific research during COVID-19 to track people's evolving worries, false information propagation, and general attitude. The extensive usage of social media has increased public involvement in emergency response thanks to the social media analysis [5].

* Corresponding author.
E-mail address: kmohbey@gmail.com (K.K. Mohbey).

https://doi.org/10.1016/j.dajour.2023.100243
Received 21 November 2022; Received in revised form 11 March 2023; Accepted 29 April 2023
Available online 5 May 2023

The monkeypox virus is a kind of the double-stranded DNA viral family Poxviridae, belonging to the genus Orthopoxvirus. Viruses of the monkeypox species have been traced back to both the Congo Basin and western Africa, suggesting that these regions may be home to genetically distinct virus subgroups. Conventional wisdom held that the Congo Basin clade was the most infectious and caused the most severe illness. The boundaries of Cameroon serve as a de facto dividing line between the two viral clades since it is the only country in which both clades have been discovered [6]. The monkeypox virus may be transmitted to humans from other animals, most often rats. Even while an epidemic may spread from person to person, it cannot be sustained indefinitely in this way. The signs and symptoms are similar to but not as severe as those of smallpox. The monkeypox virus is sometimes seen in the tropical rainforests of Central and West Africa, even though smallpox was declared extinct worldwide in 1980. The mortality rate during an outbreak of monkeypox has traditionally varied from 1% to 10%; however, most patients may completely recover with the correct treatment.

Social scientists and psychologists striving to understand better the human condition, psychology, and mental health are particularly interested in the exponential expansion of information that has happened in spreading through social media. Social media websites like Twitter have been used as a data collection tool for study in the behavioral science and psychology sectors. It has also been used to ascertain a user's personality type [7,8] and learn about the patterns and histories of internet users. Additionally, it can be applied to promote effective conversation between social media users and on other platforms that can disprove conspiracy theories. Social stigma about developing infectious diseases has led to individuals being stigmatized, stereotyped, subjected to discrimination, treated differently, and losing status, which can harm them.

With the help of Sentiment Analysis (SA), a person's opinion may be automatically placed into one of three categories, depending on whether it is a positive opinion, negative opinion, or neutral about a specific topic, product, movie, or news. Natural language processing (NLP) [9], the analysis of tweets, is used in this system to automatically analyze and quantify the users' feelings based on the information on Twitter. The management of positive, negative, and neutral sentiments communicated in tweets is the primary focus of the work being done on SA. Analyzing sentiments is not difficult to access the opinions spoken in public assessments and survey replies. The recording of text data, the shooting of images, the creation of movies, and the identification of voice are some of its possible applications. Following the data collection, the input is parsed into its parts, which may be individual words or phrases, for SA, which operates on data comparable to tweets. Because of SA on social media, which includes data from Twitter [10], there is a greater need for public perspectives, which are compiled in text. It is also a challenging problem to accurately forecast data from tweets using text analysis to fulfill the requirements of commercial evaluation.

Twitter is a social networking service that has also become a significant outlet for disseminating media via the Internet. Lots of people rely on it and utilize it often. Real-time communication conveys information clearly and succinctly concerning events and records people's opinions and reactions. Twitter is a social networking website where people may share their reactions to the global epidemic regarding their feelings, thoughts, and opinions [11,12]. People continue to use Twitter despite the current epidemic. However, the employment of ML-based algorithms is required due to the challenges in judging the inherent importance of a piece of content using NLP approaches, such as contextual phrases and words and ambiguity in written or spoken language [13–15]. NLP and its applications have significantly impacted text analysis and social media classification. Applications inspired by NLP have significantly affected evaluating and classifying information on social media.

Transferable learning is the knowledge that may be used in other contexts. Transfer learning is often used in image processing to speed up the learning and training processes. Knowledge transfer is fundamental to transferring learning [16] from the source to the target domain. According to the study done by Sv et al. [17] the social media sentiment, more people are posting positively (28.82%) about the monkeypox virus than negatively (23.01%). A closer look at the tweets indicates that most tweets with positive sentiments about monkeypox talk about mild symptoms and lower infection mortality rates. A public opinion survey showed that people are not too worried about the monkeypox virus. The analysis of tweets with negative sentiments about monkeypox revealed that users were discussing issues like the virus's potential to cause death, its severity, the lesions it leaves behind, the availability of vaccines, whether or not monkeypox is the next pandemic after COVID-19, the safety of travel, and the impact of the virus's spread on human health.

Knowledge graphs are helpful in various applications, such as auxiliary diagnostic and knowledge answer systems [18]. Researchers from a wide range of academic fields have studied the relationship between changes in the pandemic and social media use since the novel monkeypox outbreak began. Researchers attempting to stop the spread of the disease are very interested in identifying connections between monkeypox changes and social media activity. Using data from Twitter, we build a knowledge graph to address issues such as the evolution of public opinion around monkeypox and its spread. Researchers want to comprehend how public opinion, policy statements regarding monkeypox, and global events interact. To appropriately capture the content of social media data announcements while retaining the social network's underlying structure and the interactions between entities, a knowledge graph is well suited for the job [19].

In this article, we provide the results of a comparative analysis that we carried out by putting our model up against other models that had been trained using the same dataset [3,20–22]. The primary objective is to create a sentiment analysis model based on traditional and hybrid deep learning forms. As a result of our research of the relevant literature, we discovered that most available models pertain to COVID-19 or other illnesses, but not monkeypox. The recent occurrence of monkeypox has also contributed to a rise in the research and development carried out in this area. However, no effective research done in the past focused on sentiment analysis on the monkeypox epidemic of 2022. In addition, no work has been done in this sector so far that has concentrated on establishing open research directions to improve knowledge, innovation, and discovery in this field. The purpose of this study is to provide solutions to these problems. The primary purpose of this work was to utilize a hybrid method of deep learning to analyze the positive, negative, and neutral opinions expressed on Twitter about monkeypox. In a nutshell, the following are some of the scientific contributions that it contributes to this area:

1. This study uses a hybrid deep learning approach to find user opinion about Monkeypox infection on social media.
2. We consider three possible polarities of a user's tweet: positive, negative, and neutral.
3. The hybrid approach use Convolutional Neural Networks (CNN) and Long Short-Term Memory Networks (LSTM).
4. The hybrid CNN-LSTM approach is proposed to detect the user's sentiments.
5. A knowledge graph was built to help semantic analysis of tweets by extracting the structured knowledge.

After this section, the rest of the paper is organized as follows: first, a review of the relevant prior literature; second, an explanation of the methodology used for the various approaches; third, an account of the experiments conducted and the results obtained; and fourth, a discussion of the implications of the findings.

2. Related work

Due to its global impact on public health, monkeypox is of particular concern to countries in west and central Africa and the rest of the world. In 2003, the United States of America experienced the first country outside of Africa to have an outbreak of monkeypox. In this case, epidemiologists determined that exposure to ill prairie dogs in captivity was the source of the disease. These pets shared a cage with dormice and Gambian pouched rats that had been imported to Ghana. Over 70 cases of monkeypox have been documented in the US due to this outbreak. It has been reported that Nigerian tourists caught monkeypox in Israel in September 2018, the UK in September 2018, December 2019, May 2021, and May 2022, Singapore in May 2019, and the USA in July and November 2021. These nations may be found in Southeast Asia. Many cases of monkeypox were discovered in May of 2022 in countries where it was not ordinarily present. Studies are now being conducted to learn more about the disease's epidemiology, vectors, and transmission dynamics [23].

The monkeypox virus is an orthopoxvirus that may infect humans and produce monkeypox, a viral illness with symptoms including fever and rash similar to smallpox. Since the smallpox virus was eradicated from the human population in 1980, monkeypox has become humans' most severe orthopoxvirus infection. Cases are recorded most often from rural parts of nations located in Central and West Africa, in places near tropical rain forests where humans may have contact with animals afflicted with the disease. Someone can get monkeypox by contacting the respiratory droplets of another person with the disease, either at home or at a medical institution or by touching contaminated objects or materials, such as bedding. Although these are the primary means of person-to-person transmission, monkeypox outbreaks often occur in small clusters of a few cases without progressing to extensive community transmission. This is because monkeypox is a very contagious disease. When swift action is taken in response to an epidemic, it is far simpler to contain the spread of the disease. Other instances of monkeypox are reported in different countries due to importation by travelers or animals afflicted by the disease [6].

The World Health Organization (WHO) convened an "emergency conference" [24] on May 20, 2022, to examine the worldwide worries over the increasing number of cases of the monkeypox virus. The meeting was called to address global concerns. The WHO deliberated over the subsequent few days over whether or not the epidemic should be classified as a "Potential public health emergency of international concern (PHEIC)", as was done in the past for outbreaks of COVID-19 and Ebola [25]. A "Level 2" monkeypox notice was issued by the Center for Disease Control (CDC) in the United States on June 6, 2022. This was in response to the significant rise in reported cases [26]. On July 23, 2022, the WHO proclaimed monkeypox as a global health emergency after another conference.

A generalized prediction model based on an artificial neural network (ANN) was created by Y Kuvvetli et al. [27] to match the distributions of various nations and forecast the future number of daily cases and fatalities by COVID-19. They used data collected from many nations between March 11, 2020, and January 23, 2021. ANN model is another tool the government may use to help hospitals and other medical institutions avert problems.

Opinion mining is a technique for analyzing customers' feelings and may be performed using either Text Mining (TM) or NLP. The advantages of sentiment analysis include upscaling potential, agent monitoring, and real-time data. Service providers may use sentiment analysis to determine whether customers are happy with [28]. Discerning positive and negative sentiments in complex wordplay may be challenging, but sentiment analysis can help [29]. Using the sentiment analysis technique, researchers can determine the overall tone of a piece of writing. Information gathering is improved and made more thorough via sentiment analysis. Choudhary et al. [30] presented a deep learning ensemble to create a recommendation model that considers ratings and reviews. Deep learning, which uses back propagation neural networks with numerous hidden layers and changing nodes, is employed for the recommended approach, which allows for quick learning.

A model was proposed by Hassan et al. [20] to identify polarity by combining Convolutional Neural Networks (CNN) and LSTM with pre-trained word vectors obtained from IMDB movie reviews. The four-layer CNN classifier that was developed using this strategy consisted of two levels of convolutional processing, two tiers of pooling processing, and two layers of output processing. Both the CNN model, which had an accuracy of 87.0%, and the LSTM model, which had an accuracy of 81.8%, fared worse in the trials than the combined CNN + LSTM model, which had an accuracy of 88.3%. Shen et al. [21] developed a unique approach for determining the polarity of movie reviews by integrating CNN with bidirectional LSTM. This method successfully recognized both positive and negative reviews. When combined, the CNN and LSTM classifiers produce a model with an accuracy of 89.7%, much higher than the individual models' accuracy of 83.9 and 78.5%, respectively.

Dash et al. [31] simulate the behavior of this pandemic using Support Vector Regression (SVR) and LSTM algorithms. According to simulation data, LSTM produces more accurate outcomes for the Indian Scenario. Short-term memory is a significant obstacle for most recurrent neural networks (RNNs), making LSTMs an essential machine learning component. Because of this, it becomes difficult to retain data from one process and apply it to another. For instance, if a dataset or numbers are projected to be processed, RNNs may forget some historical information [32,33]. The investigation of human-related occurrences has been significantly aided by the processing and analysis of data obtained from social media platforms, which has transformed the field of infodemiology. In addition, these social networks publish statistical data on social-trend illnesses like monkeypox, such as the number of comments, images, and videos shared. As a result, this makes it possible to anticipate the monkeypox morbidity rates in each region and alerts the decision-makers in charge of health policy to develop educational and preventative initiatives in the areas with the highest risk [4].

Farahat et al. [1] carried out two primary analyses: (1) Sentiment analysis and (2) topic modeling of Monkeypox-related social media data. Due to the recent monkeypox outbreak, opinions and related digital information have flowed on many social media sites, including Twitter. For governments, policymakers, healthcare professionals, and researchers to use the resources available to manage and lessen the burden of the recent outbreak effectively and timely, it is essential to ascertain public trends and perspectives regarding monkeypox. Based on the "model of stigma communication", Dsouza et al. [2] assessed the dynamics of stigma communication. According to this analysis, to enhance the system, authorities must address false information, discrimination against the LGBTQ+ population, and a lack of a thorough risk-communication plan. For sentiment analysis of unlabeled data, they have used the NLP packages VADER (Valence Aware Dictionary for Sentiment Reasoning) and TextBlob, as well as the embedding-based model Flair and Lexicon (or Rule) based models.

The Fuzzy Algorithm for Monitoring, Extraction, and Classification (FAMEC) approach was employed to alert monitoring systems to monkeypox outbreaks promptly [4]. Numerous studies have highlighted the significance of tweets being consistent with reports from the WHO and the Centers for Disease Control and Prevention and acknowledged that data mining from these tweets could be used to identify the geographic location of patients and track and forecast COVID-19 mortality rates. To detect hotspots by clustering the sentiment of these tweets using the point-based location technique, scientists gathered 70,000 vaccine-related geotagged Twitter posts (tweets) in nine African nations [4]. They divided the tweets into three sentiment groups using the VADER pre-trained model for NLP positive, negative, and neutral. Woźniak et al. [34] offer an Internet of Things (IoT) system model that integrates deep learning from BiLSTM with decision tree modeling, data balancing, and automated diagnosis support. The proposed technique consists

Table 1
Comparative study of related works.

S. No	Author	Approach used	Dataset	Accuracy (%)	Year
1	Hassan et al. [20]	CNN + LSTM	IMDB movie reviews	88.3	2017
2	Shen et al. [21]	CNN LSTM	Movie reviews	83.9 78.5	2017
3	Ardakani et al. [36]	Deep learning models	COVID-19	99	2020
4	Wang et al. [37]	Inception-based model	453 CT scan images	73.1	2020
5	Sanddep et al. [22]	CNN	Psoriasis, Melanoma, Lupus, and chickenpox	71	2022
6	Ali et al. [38]	VGG16 ResNet50 InceptionV3 Ensemble	MSLD database	81.48 82.96 74.07 79.26	2022
7	Sahin et al. [39]	MobileNetv2		91.11	2022
8	Ogbuokiri, B et al. [4]	Logistic Regression SVM Decision Tree K-Nearest Neighbor	Tweeter Data	71 65 61 56	2022
9	Villanueva et al. [35]	CNN + LSTM	Monkeypox tweets	83	2023

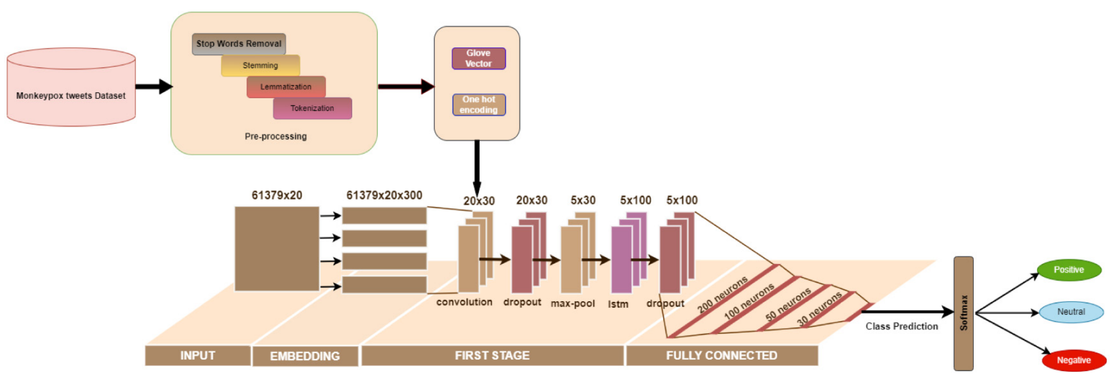


Fig. 1. Outline of sentiment detection.

of an experimental sequence of data preparation using well-known balancing algorithms. These algorithms include ADASYN and SMOTE-Tomek. Between the patient and the appropriate medical team, the system facilitates the secure exchange of documents and the evaluation of questionnaires. A deep learning model's automated diagnosis is visible to patients and doctors at the system level. The model offers a significant advancement to assist patients more quickly.

In addition, Villanueva et al. developed a hybrid-based model architecture that was constructed using CNN and LSTM to assess the accuracy of the prediction of monkeypox tweets. They were able to achieve a forecast that was correct 83% of the time [35]. Table 1 compares the current studies to the proposed study in relative changes.

It is possible to draw the following conclusion from the works that were discussed in Section 2: in the past, datasets on Twitter have been generated on a wide variety of subjects, including but not limited to global concerns, emerging matters, public needs, and viral outbreaks. The exploration of various research directions about advancing timely knowledge, innovation, and discovery in the respective areas has benefited from using such datasets. Examining current works connected to this monkeypox leads one to conclude that there is a paucity of available studies on this particular subject. This further demonstrates the need to develop a hybrid deep-learning model for sentiment identification utilizing tweets about the monkeypox. This study provides a hybrid CNN-LSTM-based model to fulfill the demand mentioned above. Sections 3 and 4 describe the technique used in developing the suggested model and the findings obtained.

3. Proposed methodology

The dataset used in this study was sourced from the open-source hosting site GitHub.[1] All sorts of tweets about monkeypox are included in this collection. After finishing the pre-processing and ensuring there were no duplicate or null values, the data characters were detokenized so that the sentences could be split down into words and labels could be assigned. After that, the performance of the forecasting model was evaluated using deep learning's CNN-LSTM architectures. Fig. 1 is a detailed schematic of the whole system.

3.1. CNN

A Convolutional Neural Network (CNN) is constructed from a series of layers that are added one after the other in sequential order [40]. Convolutional layers, pooling layers, batch normalization layers, fully connected layers, and loss layers are some of the layers that make up these layers. These layers, the parts of a CNN responsible for feature extraction and selection [41], make up the two most significant components of a CNN: the layers themselves.

Consider a CNN with Xk be the input for the kth layer and Wk be the set of parameters that may be trained for each layer. The input 1 is sent through several layers of processing before arriving at the loss layer, where the output yj and the label of the jth tweet $\hat{y}j$ are combined with the loss function's contribution to getting an error z. During the first phase of preparation, this procedure, also known as an advance run, occurs. A second method, backward propagation, is

[1] https://github.com/.

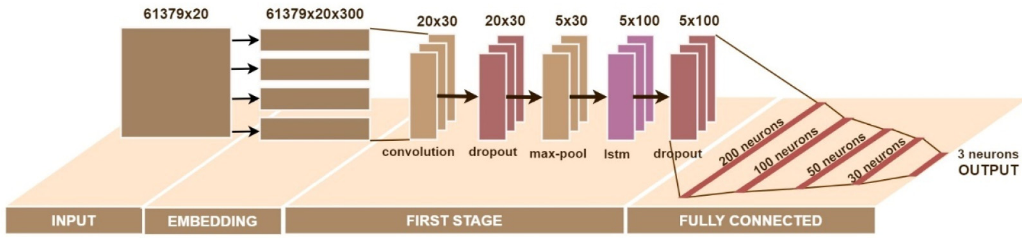


Fig. 2. CNN-LSTM based proposed architecture.

```
Layer (type)                 Output Shape              Param #
=================================================================
embedding (Embedding)        (None, 20, 300)           24475500
conv1d (Conv1D)              (None, 20, 30)            9030
dropout (Dropout)            (None, 20, 30)            0
max_pooling1d(MaxPooling1D)  (None, 5, 30)             0
lstm (LSTM)                  (None, 5, 100)            52400
dropout_1 (Dropout)          (None, 5, 100)            0
flatten (Flatten)            (None, 500)               0
dense (Dense)                (None, 200)               100200
dropout_2 (Dropout)          (None, 200)               0
dense_1 (Dense)              (None, 100)               20100
dropout_3 (Dropout)          (None, 100)               0
dense_2 (Dense)              (None, 50)                5050
dropout_4 (Dropout)          (None, 50)                0
dense_3 (Dense)              (None, 30)                1530
dropout_5 (Dropout)          (None, 30)                0
dense_4 (Dense)              (None, 3)                 93
=================================================================
Total params: 24,663,903
Trainable params: 188,403
Non-trainable params: 24,475,500
```

Fig. 3. Proposed model summary.

utilized during the training process. The error is then used in the process, which adjusts all of the trainable parameters of the CNN using a learning method like stochastic gradient descent (Eq. (1)).

$$(\boldsymbol{w}^k)^{i+1} = (\boldsymbol{w}^k)^i - \eta \frac{\partial z}{\partial (\boldsymbol{w}^k)^{i'}} \tag{1}$$

where η stands for the algorithm's learning speed and i for the ith iteration of training, the learning rate η is a type of hyper parameter, the incorrect choice of which might provide sub optimal outcomes. Convolutional layer (Eq. (2)) is the output of the kth convolutional layer, that is a tensor of order 3 with the notation Y^k $\left(or \;\; X^{k+1}\right) \in \mathbb{R}^{M^k - m+1xN^k - n+1xS}$.

$$y_{i^k,\; j^k,s} = \sum_{i=0}^{m} \sum_{j=0}^{n} \sum_{l=0}^{d^k} F_{i,j,d^k,s} \; X \; x^k_{i^k,j^k,l} \tag{2}$$

The Eq. (2) is solved several times for every $0 \leq s \leq S$ and for every spatial position that satisfies the conditions $0 \leq i^k \leq M^k - m + 1 \; and \; 0 \leq j^k \leq N^k - n + 1$.

Throughout our investigation, we used maximum pooling, resulting in outputs that were generated following Eq. (3).

$$y_{i^k,j^k,d} = \max_{0 \leq i \leq m, 0 \leq j \leq n} x^k_{i^k xm+i,\; j^k xn+j,\; d'} \tag{3}$$

where $0 \leq i^k \leq M^k$, $0 \leq j^k \leq N^k$, $0 \leq d \leq D^k$.

Pooling layers are utilized to lessen the complexity of the output tensors while maintaining the most significant identified patterns [42]. Pooling layers are utilized to achieve this goal, which makes intuitive sense.

A fully connected layer is the last layer that is fully linked. There is always a classification function, such as sigmoid, softmax, tanh, etc., that produces an actual value yi that will be compared with the predicted value $\hat{y}j$ based on the loss function that was set. In the present context, we believe applying the sigmoid function described by Eq. (4) is appropriate for this classification.

$$\hat{y}_j = \frac{e^{x_j}}{1 + e^{x_j}}, \qquad x_j \in \mathbb{R} \tag{4}$$

Finally, ReLU and batch normalization processes serve as significant transition mediums that connect the previously discussed layers. Eq. (5) defines the ReLU function.

$$y_{i,j,k} = \max(0, x^k_{i,j,d}) \tag{5}$$

3.2. LSTM

LSTM, which allows us to describe very long input sequences. The issue of vanishing gradients in RNN prompted the development of LSTM as a solution. It has three gates: an input gate, a forget gate, and an output gate [43,44]. The operation LSTM may be seen in the mathematical equations that are shown below.

$$f_t = \sigma_h(W_f x_t + U_f h_{t-1} + b_f) \tag{6}$$

$$i_t = \sigma_h(W_j x_t + U_j h_{t-1} + b_j) \tag{7}$$

$$o_t = \sigma_h(W_o x_t + U_o h_{t-1} + b_o) \tag{8}$$

$$c_t = f_t \odot c_{t-1} + i_t \odot \sigma_t(W_c x_t + U_c h_{t-1} + b_c) \tag{9}$$

$$h_t = o_t \odot \sigma_h(c_t) \tag{10}$$

where f_t, i_t, o_t, c_t, h_t are the activation vectors for input gate, forget gate, and output gate, respectively.

3.3. CNN-LSTM

The architecture for sentiment categorization is described in this part and will be applied to Twitter data analysis. We used an architecture based on RNN termed CNN-LSTM to test the models. The CNN-LSTM architecture is shown in Fig. 2, and the model summary is depicted in Fig. 3.

Assume that a string represents the word vectors and that $P_i \in R^k$ represents the K-dimensional vector comparable to the ith token in a user review of total size n, where n is the number of tokens in Eq. (11). Zero padding is added if the sentence is smaller than n characters.

$$P_{1:\,n} = P_1 + P_2 + P_3 \ldots\ldots\ldots\ldots + P_n \tag{11}$$

The + operator denotes a concatenation operation in Eq. (11). Similarly, suppose the concatenation of the words $P_i, P_{i+1}, P_{i+2}, \ldots\ldots\ldots P_{i+j}$ is equivalent to $P_{i\,:\,i+j}$. Let $W \in R^{hk}$ denote the convolutional filters

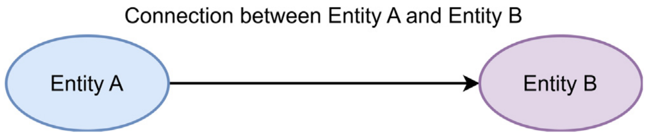


Fig. 4. Entities and relationship graph.

applied in an nxK dimensional matrix of a sentence with a window or gap of h words to produce a new feature matrix. The basic element $P_{i:\ i+j}$ denotes the local feature matrix from the ith to the $(I + J)$th line of the present sentence vector. Eq. (12) can be used to produce a feature C_{if} from a window of words $P_{i:\ i+h-1}$.

$$C_i = f(W\ P_{i\ :\ i+h-1} + b) \tag{12}$$

An activation function, such as a hyperbolic tangent or sigmoid, is used at the place where the bias, denoted by b, is one of the real numbers. The actual numbers include b as one of their members. In order to generate a feature map through Eq. (13), the convolved convolutional filter needs to be applied to every window containing words.

$$C = [C_1, C_2, C_3, C_{n-h+1}] \tag{13}$$

where C belongs to $R^{n-h+1.}$

The Equation for building a single feature map from a single convolutional filter is shown above. It may be thought of as follows: In the same manner, $m(n-h+1)$ features will be generated by a convolutional layer that has multiple m filters. In order to adequately capture long-term dependencies, the features are immediately moved into the LSTM layer, which comes before the fully connected layer.

3.4. Knowledge graphs

The concept of meaningfully interlinking data and representing unstructured information led to the development of Knowledge Graphs (KG). A Knowledge Base, often known as a KG, is a type of technology that maintains detailed knowledge and data in either an unstructured or organized manner. A KG is characterized as a network in which the nodes represent entities, and the edges represent the relationships between those entities [45]. DBpedia,[2] YAGO,[3] and SUMO[4] are three instances of enormous KG that have been produced in the previous decade. These KGs have become a fantastic resource for NLP applications like as Question-Answering (QA) [46,47]. Other examples of huge KG are SUMO and YAGO. The concept of triples is at the core of KG, and its general form can be expressed as follows:

A knowledge graph (KG) is a triple with the notation $KG = (E, R, F)$. *In this notation,* $E = \{e_1, e_2, \dots, e_n\}$ *refers to a collection of entities,* $R = \{r_1, r_2, \dots, r_n\}$ *represents a set of binary relations, and* $F \supseteq ExRxE$ *refers to the relationships that exist between things.*

The KG may facilitate the representation of knowledge in various contexts since it contains data from the actual world in RDF-style triplets[5] such as (head, relation, and tail). Knowledge graphs have been proven helpful in healthcare analytics and are integrated into various healthcare applications to enhance data representation and knowledge inference [48]. As indicated in [49], it can help visualize disorders in the medical field; in [50], it is described in online retail, which maps customers' buy intents to sets of potential items. Entities and their relationships with one another are the building blocks of KG. This network is represented as a graph [51]. Two nodes in Fig. 4 stand for separate entities.

[2] http://aksw.org/Projects/DBpedia.html.
[3] https://yagoknowledge.org.
[4] http://www.adampease.org/OP/index.html.
[5] http://www.w3.org/TR/rdf11-concepts.

Table 2
Hyperparameters detail.

Parameter	Value
Input length	20
learning rate	0.001
Loss function	categorical_crossentropy
optimizer	sgd
No. of epochs	60
Batch size	64
Activation function	Relu
Dropout	0.1
Embedding	300
Conv1D Layer	30
LSTM Layer	100
Pooling	MaxPooling1D

These two nodes are linked together, symbolizing the closeness of their friendship. In this case, Node A is the subject, Node B is the object, and their connection is the predicate. Another name for this is a "semantic triple". To illustrate, Fig. 5 depicts a KG for the triples ("Human Monkeypox", "is a", "Zoonotic Disease"), ("Zoonotic Disease", "caused by", "Monkeypox Virus"), ("Monkeypox Virus", "member of", "Orthpoxvirus").

4. Experimental results

4.1. Data source and description

Tweets are thought to be associated with monkeypox in this particular piece of research. The dataset includes 61379 tweets concerning monkeypox that were published on Twitter on May 7 and June 11, 2022 [52]. This dataset includes a variety of tweets posted by individuals, some of which express the users' positive thoughts, some of which express their negative ideas, and some of which express their neutral thoughts. Before actual tests on the gathered tweets, pre-processing is an essential step that must first be completed. The collected tweets are chaotic, unbalanced, and contain many stop words. Therefore, to do classification and prediction tasks, it is necessary to clean all the tweets. Since Twitter is an unstructured platform that permits publishing in several languages, data pre-processing is paramount. The following procedures were used to prepare tweets for analysis: (1) Duplicate records were removed in two steps. First, duplicates were found by looking at each tweet's Twitter-provided "is retweet" descriptor. Second, duplicate tweets were removed using the tweet id and content. (2) Tweets were cleaned up in the manner described below: (a) eliminated URLs, email, hashtags, mentions, and numbers using TextBlob analyzers, (b) converted tweets to lowercase, and removed punctuation and stopwords. (3) We only allowed English-language tweets. The information gathered for this investigation is openly accessible. User anonymity was preserved by displaying data in an aggregated way [53]. Fig. 6 illustrates the classification of tweets according to the categories to which they belong. In Fig. 6, the 0 class represents those with neutral opinions, the −1 class represents those with negative thoughts, and the 1 class represents those with positive opinions.

4.2. Experimental setting

Experiments are carried out using Python programming on a machine running X86-64 Ubuntu 18.04.4 LTS. The CPU is an Intel(R) Core(TM) i7-8550U operating at 1.80 GHz, with 16 GB of RAM. We use this setup to investigate sentiment emotions with the suggested CNN-LSTM model. To recognize the characteristics, the CNN architecture was given training. Following that, other characteristics from the LSTM architecture were utilized to classify people's feelings. Table 2 contains the optimal values of the proposed model hyperparameters.

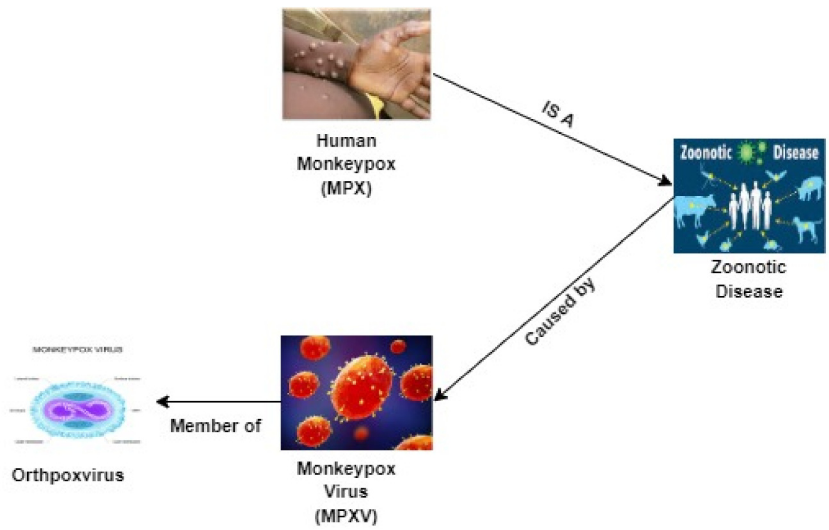


Fig. 5. Knowledge graph with two entities and its relationship.

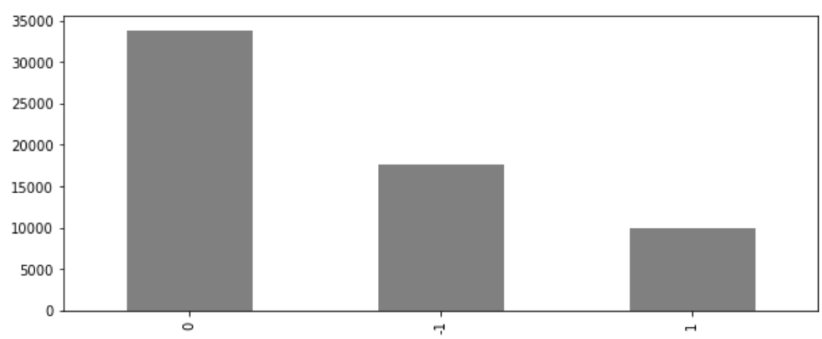


Fig. 6. Histogram of various classes of Monkeypox tweets dataset.

4.3. Performance evaluation

The model's quality was assessed using the fitness function. Based on the validation data, we assessed the proposed model's fitness in terms of loss of classification error. Validation data were considered when calculating the loss error to prevent the over-fitting issue. On training data, the suggested model was trained over several training epochs. In this work, the rectified linear unit (ReLU) was employed as the training activation function, and the sgd optimizer was used to optimize the model. The best model was chosen after evaluating each one using the least loss of error function. Divide the dataset into two parts so the model may be tested on various datasets and then trained on them.

The correct evaluation procedures are required to eliminate the bias in the models' differentiation. Precision (P), recall (R), F1 score, accuracy, and AUC are some of the metrics that are used for the classification standard the most commonly [54,55]. The fraction of properly recognized samples relative to the total number of samples that have been evaluated and identified is used to determine how accurate the results are. The recall is the total number of samples from all positive representations that were successfully identified. This figure may be broken down further into individual samples. The F1 score is arrived at by taking the recall and accuracy scores and averaging them harmonically. The accuracy may be determined by calculating the proportion of correctly recognized samples included within the total number of samples [56]. The AUC is a statistic that may be used to describe the ROC curve [57]. It measures how well a classifier can differentiate between several different data groups. The equations of these measures are shown below.

- **Accuracy:** Percentage of records that have been correctly classified from the total number.

$$Accuracy = \frac{True\_Pos + True\_Neg}{True\_Pos + True\_Neg + False\_Pos + False\_Neg} \quad (14)$$

- **Precision:** The number of records correctly classified as "Positive" as compared to the number predicted as "Positive".

$$Precision = \frac{True\_Pos}{True\_Pos + False\_Pos} \quad (15)$$

- **Recall:** This is calculated by dividing the number of records correctly classified as "Positive" by the total number of records that were "Positive".

$$Recall = \frac{True\_Pos}{True\_Pos + False\_Neg} \quad (16)$$

- **F1-Measure:** It is determined by averaging the Precision (P) and Recall (R) values.

$$F1 - Measure = \frac{2 * P * R}{P + R} \quad (17)$$

- **AUC:** It estimates area under the complete receiver operating characteristic (ROC) curve.

$$AUC = \left(R - \frac{False\_Pos}{False\_Pos + True\_Neg} + 1 \right) /2 \quad (18)$$

Table 3 contains an exhaustive listing of the confusion matrix's parameters. TN refers to the number of negative samples correctly identified, while TP refers to the number of positive samples correctly identified. FP is the number of times unfavorable instances was incorrectly recognized as positive, while an FN predicted value is negative, but the actual value is positive.

Table 3
Confusion matrix.

	Actual value		
		Positive	Negative
Predicted value	Positive	TP	FP
	Negative	FN	TN

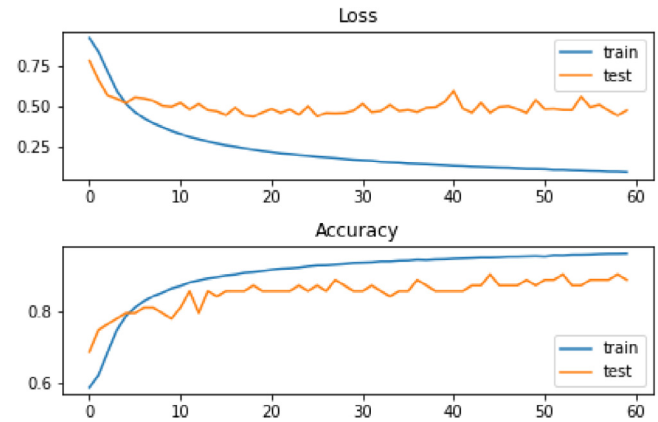


Fig. 7. Accuracy and loss history of the CNN model.

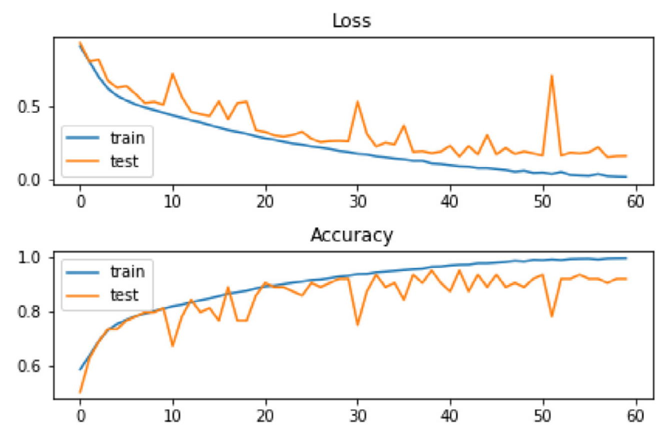


Fig. 8. Accuracy and loss history of the LSTM model.

4.4. Experimental outcomes

To determine whether or not the proposed CNN-LSTM model is valuable, we compared the outcomes of the experiments to the findings obtained from various machine learning models. This allowed us to determine how efficient the proposed model is. During the comparison, each of the following was considered: accuracy, the area under the curve, f1-score, precision, recall, and confusion matrix. The model offered and now being discussed uses various classifications, including neutral, positive, and negative categories. Following each training epoch for the monkeypox tweet dataset, Figs. 7, 8, and 9 demonstrate the accuracy and loss history on several deep learning-based models' test sets.

It is evident from the results that the CNN-LSTM model outperforms more straightforward CNN and LSTM-based deep learning models in terms of performance. In Fig. 9, accuracy increases steadily from epoch ten and remains constant until epoch 50, at about 94%. It will be regarded as the experiment's most acceptable outcome. Eight models' findings from the monkeypox tweets dataset are displayed in Fig. 10. In terms of Accuracy, Precision, Recall, F1-score, and AUC across all five assessment parameters, the suggested technique performs better than existing state-of-the-art methods overall. After epoch 60, the loss for

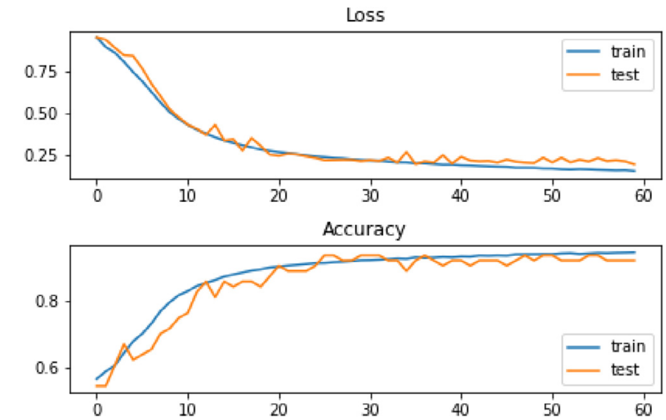


Fig. 9. Accuracy and loss history of the CNN-LSTM model.

the CNN-LSTM is roughly 0.1, and accuracy can approach 94%. This shows that our suggested model has more accuracy, indicating that it can accurately assess different moods.

Figs. 11 and 12 display the confusion matrices generated on the hidden test dataset to help the reader better understand the classification performance of each model. From these figures, it can be concluded that the CNN-LSTM model provides the best classification report compared to other models.

The suggested CNN-LSTM model achieved averages of 94%, 95%, and 95% for accuracy, specificity, and F1 score measures, respectively. Our method yielded superior precision, recall, F1 score, and accuracy compared to the deep learning-based method used by ref [35]. The KG depicted in Fig. 13 is based on a subset of tweets from a dataset about the monkeypox. Multiple entities and their relationships are depicted here in a simple graph.

In addition, we have produced a comprehensive KG of the monkeypox tweets dataset to acquire a deeper comprehension of the concept of knowledge representation. Fig. 14 provides a visual representation of this graph. A knowledge network of tweets about the monkeypox was created and examined. In our experiments, we considered a triplet to imply a relationship if its scoring function value was among the lowest 100 values for its relevant topic. The graph structure has identifiable communities, which enables link prediction. Using such a network structure, we might be able to respond to issues crucial to the actual world, like the relationships between topics and clusters of tweets.

4.5. Discussion

The results produced by the built CNN-LSTM model are compared to the findings produced by various other models that are regarded as being state-of-the-art. In addition to such models, this collection also contains models that use machine learning and deep learning. The research findings indicate that compared to the other models, the CNN-LSTM model yields results of a higher quality than those produced by the other models. When applied to the dataset of tweets on monkeypox, the f1-score measure that the LSTM model generates is correct 94% of the time. In addition to this, the accuracy of the recommended model, and its recall, have both been greatly improved.

Several studies, including those on the prediction of the swine flu pandemic, the detection of influenza epidemics, and the use of surveillance systems for the detection of influenza and cancer, demonstrate the application of tweeter mining in the prediction and detection of various disease outbreaks. The proposed models can be helpful in a variety of circumstances. The findings of this study demonstrated that social media data could be utilized to forecast users' attitudes about vaccination during the monkeypox outbreak or subsequent outbreaks. Vaccination reluctance is a significant obstacle for health officials in

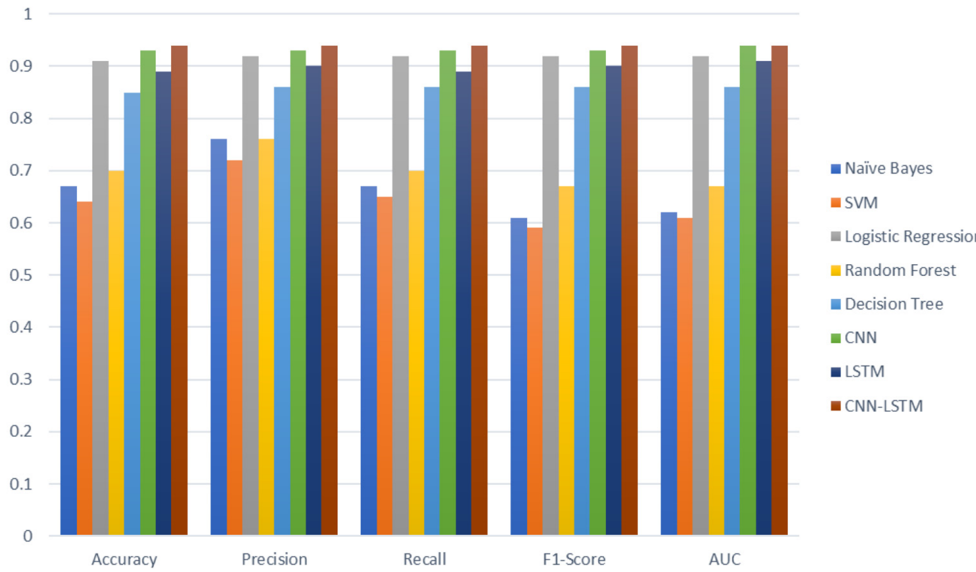
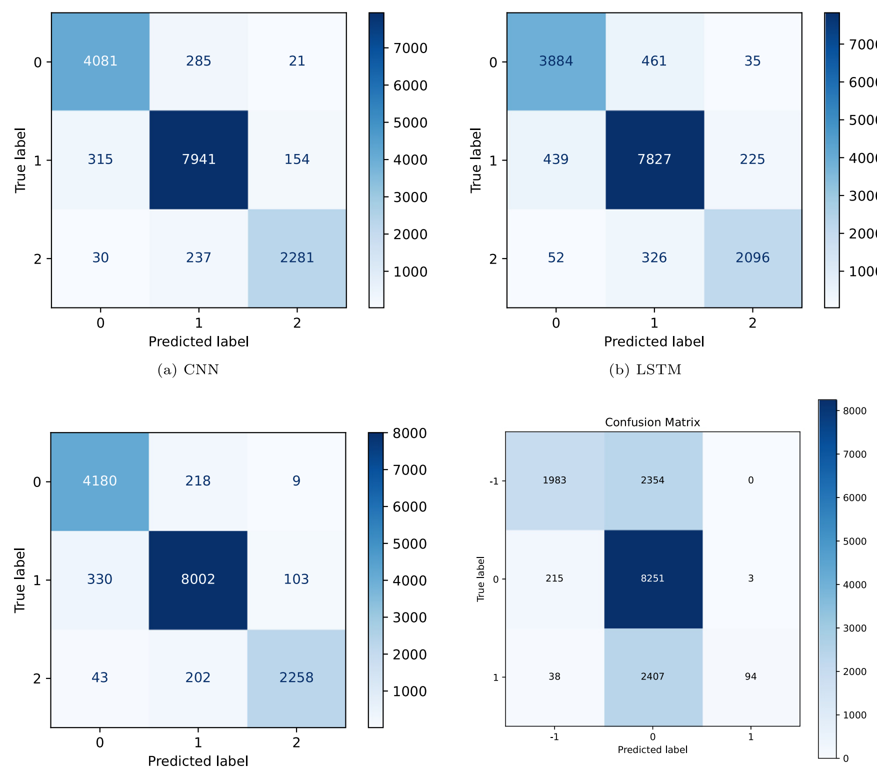


Fig. 10. Performance comparison of different models.



(a) CNN (b) LSTM (c) CNN-LSTM (d) Naive bayes

Fig. 11. Obtained confusion matrix: (a) CNN (b) LSTM (c) CNN-LSTM (d) Naïve bayes.

the fight against infectious diseases that affect communities, such as COVID-19, malaria, monkeypox, and Marburg. Misinformation about diseases on social media can worsen public health problems, which could severely threaten public health. Enhancing existing data and identifying user sentiment hotspots can also help plan and administer health policy. Systems of surveillance can be created to identify outbreaks of monkeypox quickly. Numerous studies have shown how important it is for tweets to match up with reports from the WHO and the Centers for Disease Control and Prevention. They have also shown how useful it is to mine data from tweets to identify patient locations and to track and forecast the monkeypox outbreak. The stigma attached to monkeypox may discourage people from adhering to advice. NLP-based tweets analysis can identify the monkeypox stigma on Twitter using a "model of stigma communication".

5. Conclusion

The findings of our study illustrate how CNN and LSTM techniques may be utilized to analyze tweets to determine the emotional polarity of their contents. The results of this study are based on the user's perception of the monkeypox infectious sickness as either positive, negative, or neutral. The CNN-LSTM model we suggested as part of our study

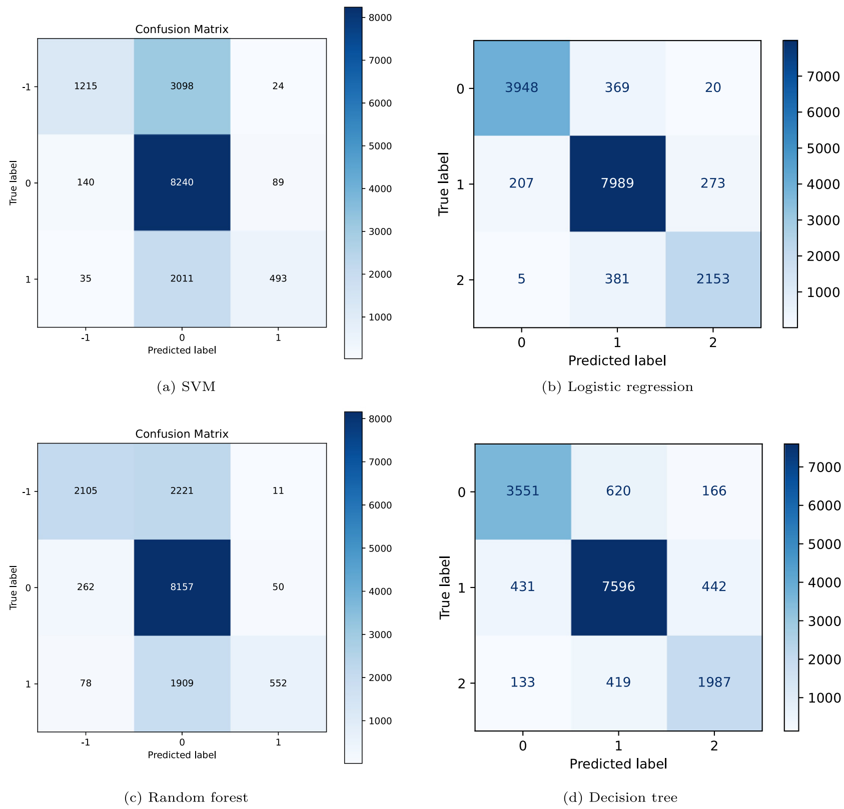


Fig. 12. Obtained confusion matrix: (a) SVM (b) Logistic regression (c) Random forest (d) Decision tree.

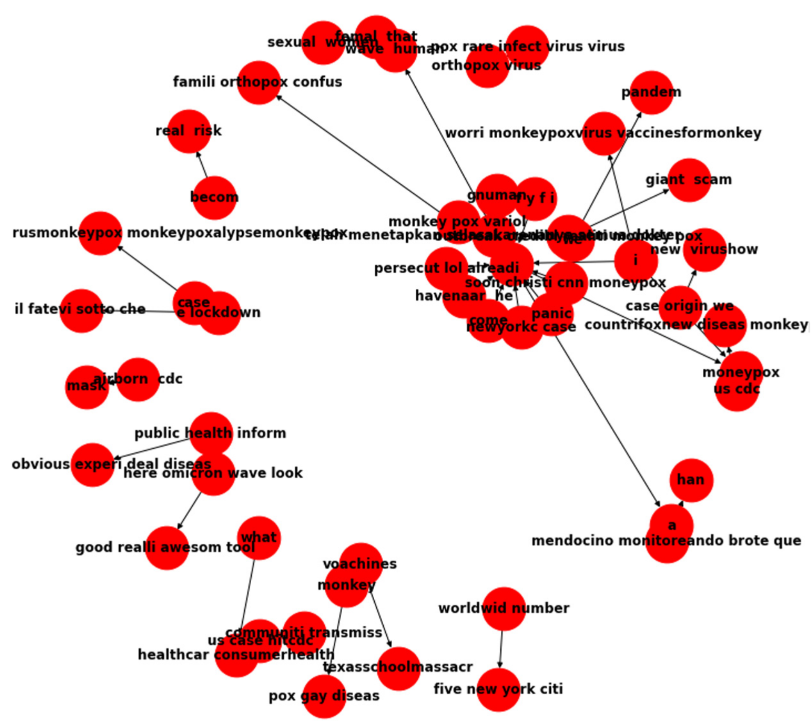


Fig. 13. Knowledge graph of the selected tweets.

helped us identify how, precisely and effectively, we could anticipate and assess the sentiments of people on social media platforms like Twitter. The new model demonstrated superior performance compared to the methods already in use. In the case of the monkeypox datasets, the hybrid CNN-LSTM architecture with hyperparameter tunings achieved a level of accuracy of 94%. According to the findings, the proposed technique is successful and suitable for categorizing the sentiments in tweets about the monkeypox. In addition, the proposed work is helpful

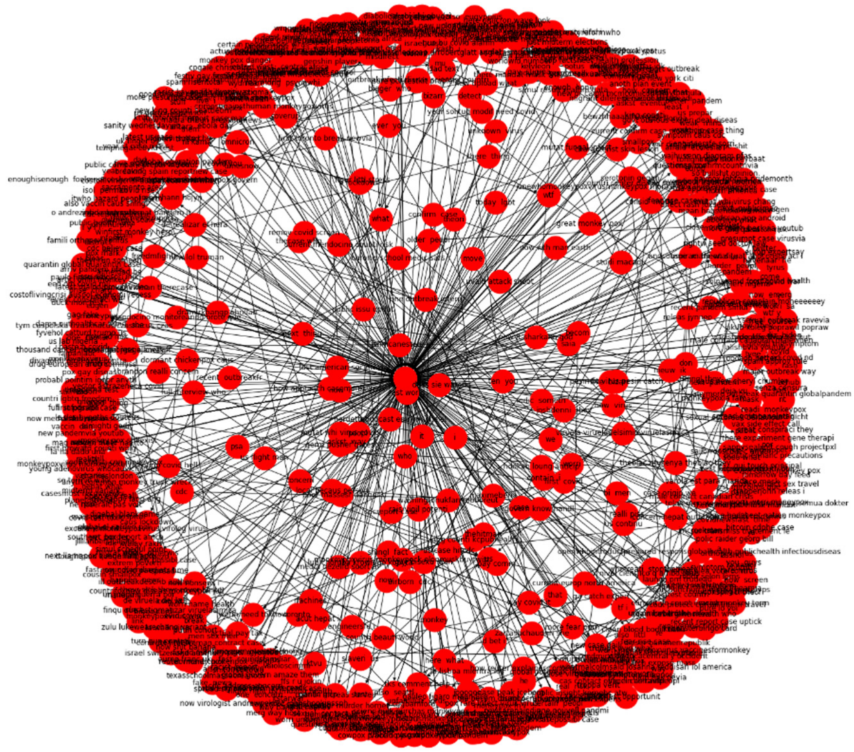


Fig. 14. Knowledge graph of monkeypox tweet dataset.

for society because it raises public awareness of the recently emerged virus monkeypox. An iterative method that employs the more robust techniques, which may be refined further in subsequent research, may be used to optimize the characteristics of a system. This optimization can be accomplished. It is also conceivable to modify it to work so that subject identification and sentiment classification may be carried out simultaneously. It would be helpful in system improvement.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request

References

[1] R.A. Farahat, M.A. Yassin, J.A. Al-Tawfiq, C.A. Bejan, B. Abdelazeem, Public perspectives of monkeypox in Twitter: A social media analysis using machine learning, New Microbes New Infect. 49 (2022) 101053.

[2] V.S. Dsouza, P. Rajkhowa, B.R. Mallya, D.S. Raksha, V. Mrinalini, K. Cauvery …, H. Brand, A sentiment and content analysis of tweets on monkeypox stigma among the lgbtq+ community: A cue to risk communication plan, Dialogues Health 2 (2023) 100095.

[3] B. Ogbuokiri, A. Ahmadi, Z.M. Nia, B. Mellado, J. Wu, J. Orbinski ., J. Kong, Vaccine hesitancy hotspots in africa: An insight from geotagged twitter posts, IEEE Trans. Comput. Soc. Syst. (2023).

[4] K. Jahanbin, M. Jokar, V. Rahmanian, Using twitter and web news mining to predict the monkeypox outbreak, Asian Pac. J. Trop. Med. 15 (5) (2022) 236.

[5] M. Kumar, H.K. Sharma, A GAN-based model of deepfake detection in social media, Procedia Comput. Sci. 218 (2023) 2153–2162.

[6] https://openwho.org/courses/monkeypox-introduction.

[7] J. Golbeck, C. Robles, M. Edmondson, K. Turner, Predicting personality from twitter, in: 2011 IEEE Third International Conference on Privacy, Security, Risk and Trust and 2011 IEEE Third International Conferenceon Social Computing, IEEE, 2011, pp. 149–156.

[8] D. Quercia, M. Kosinski, D. Stillwell, J. Crowcroft, Our twitter profiles, our selves: Predicting personality with twitter, in: 2011 IEEE Third International Conference on Privacy, Security, Risk and Trust and 2011 IEEE Third International Conference on Social Computing, IEEE, 2011, pp. 180–185.

[9] W. Dong, J. Wu, X. Zhang, Z. Bai, P. Wang, M. Woźniak, Improving performance and efficiency of graph neural networks by injective aggregation, Knowl.-Based Syst. 254 (2022) 109616.

[10] S. Lokesh, P.M. Kumar, M.R. Devi, P. Parthasarathy, C. Gokulnath, An automatic tamilspeech recognition system by using bidirectional recurrent neural network with self-organizing map, Neural Comput. Appl. 31 (5) (2019) 1521–1531.

[11] S. Chauhan, R. Banerjee, C. Chakraborty, M. Mittal, A. Shiva, V. Ravi, A self-congruence and impulse buying effect on user's shopping behaviour over social networking sites: an empirical study, Int. J. Pervas. Comput. Commun. (2021) http://dx.doi.org/10.1108/IJPCC-01-2021-0013.

[12] C. Choudhary, I. Singh, M. Kumar, Community detection algorithms for recommendation systems: techniques and metrics, Computing 105 (2) (2023) 417–453.

[13] A. Mondal, S. Mahata, M. Dey, D. Das, Classification of COVID19 tweets using machine learning approaches, in: Proceedings of the Sixth SocialMediaMining for Health (# SMM4H) Workshop and Shared Task, Mexico City, 2021, pp. 135–137.

[14] P.Ashok. kumar, GS. Shankar, S. Gautham, MPK. Reddy, GT. Reddy, A two-stage text feature selection algorithm for improving text classification, in: ACM Trans Asian Low-Resour Lang Inf Process, Association for Computing Machinery, New York, NY, 2021, p. 19, http://dx.doi.org/10.1145/3425781.

[15] S. Hakak, M. Alazab, S. Khan, TR. Gadekallu, PKR. Maddikunta, WZ. Khan, An ensemble machine learning approach through effective feature extraction to classify fake news, Future Gener. Comput. Syst. 117 (2021) 47–58, http://dx.doi.org/10.1016/j.future.2020.11.022.

[16] L. Yang, H. Zhang, D. Li, F. Xiao, S. Yang, Facial expression recognition based on transfer learning and SVM, J. Phys. Conf. Ser. 2025 (1)) (2021) 012015, IOP Publishing.

[17] P. Sv, 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. (2022).

[18] Q. Wang, Z. Mao, B. Wang, L. Guo, Knowledge graph embedding: A survey of approaches and applications, IEEE Trans. Knowl. Data Eng. 29 (12) (2017) 2724–2743.

[19] D. Flocco, B. Palmer-Toy, R. Wang, H. Zhu, R. Sonthalia, J. Lin ., P.J. Brantingham, An analysis of COVID-19 knowledge graph construction and applications, in: 2021 IEEE International Conference on Big Data (Big Data), IEEE, 2021, pp. 2631–2640.

[20] A. Hassan, A. Mahmood, A deep learning approach for sentiment analysis of short texts, in: Proceedings of the 2017 3rd International Conference on Control, Automation and Robotics, ICCAR, Nagoya, Japan, 2017, pp. 705–710.

[21] Q. Shen, Z. Wang, Y. Sun, Sentiment analysis of movie reviews based on CNN-BLSTM, in: International Conference on Intelligence Science, Springer, Berlin/Heidelberg, Germany, 2017, pp. 164–171.

[22] R. Sandeep, K.P. Vishal, M.S. Shamanth, K. Chethan, Diagnosis of visible diseases using CNNs, in: Proceedings of International Conference on Communication and Artificial Intelligence, Springer, Singapore, 2022, pp. 459–468.

[23] https://www.who.int/news-room/fact-sheets/detail/monkeypox.

[24] WHO, Calls emergency meeting as monkeypox cases top 100 in Europe, 2022, Available online: https://www.reuters.com/world/europe/monkeypox-outbreak-europe-largest-ever-region-cases-cross-100-2022-05-20/ (accessed on June 12 2022).

[25] Unlikely monkeypox outbreak will lead to pandemic, WHO says, 2022, Available online: https://www.reuters.com/business/healthcare-pharmaceuticals/unlikely-monkeypox-outbreak-will-lead-pandemic-says-who-2022-05-30/ (accessed on June 12 2022).

[26] S.R. Kelleher, CDC raises monkeypox travel alert to level 2, 2022, Available online: https://www.forbes.com/sites/suzannerowankelleher/2022/06/07/cdc-raises-monkeypox-travel-alert-to-level-2/?sh=269eee1e3f93 (accessed on June 12 2022).

[27] Y. Kuvvetli, M. Deveci, T. Paksoy, H. Garg, A predictive analytics model for COVID-19 pandemic using artificial neural networks, Decisi. Anal. J. 1 (2021) 100007.

[28] Erik. Cambria, Affective computing and sentiment analysis, IEEE Intell. Syst. 31 (2) (2016) 102–107, 2016.

[29] Xiao-Qin Liu, Qiu-Lin Wu, Wen-Tsao Pan, Sentiment classification of micro-blog comments based on randomforest algorithm, Concurr. Comput.: Pract. Exper. 31 (10) (2019).

[30] C. Choudhary, I. Singh, M. Kumar, SARWAS: Deep ensemble learning techniques for sentiment based recommendation system, Expert Syst. Appl. 216 (2023) 119420.

[31] S. Dash, S. Chakravarty, S.N. Mohanty, C.R. Pattanaik, S. Jain, A deep learning method to forecast COVID-19 outbreak, New Gener. Comput. 39 (3) (2021) 515–539.

[32] P. Chhajer, M. Shah, A. Kshirsagar, The applications of artificial neural networks, support vector machines, and long–short term memory for stock market prediction, Decis. Anal. J. (2022).

[33] SV. Praveen, R. Ittamalla, G. Deepak, Analyzing Indian general public's perspective on anxiety, stress and trauma during Covid-19 - a machine learning study of 840, 000 tweets, Diabetes Metab. Syndr. Clin. 15 (2021) 667–671, http://dx.doi.org/10.1016/j.dsx.2021.03.016.

[34] M. Woźniak, M. Wieczorek, J. Siłka, BiLSTM deep neural network model for imbalanced medical data of IoT systems, Future Gener. Comput. Syst. 141 (2023) 489–499.

[35] O. Iparraguirre-Villanueva, A. Alvarez-Risco, J.L. Herrera Salazar, S. Beltozar-Clemente, J. Zapata-Paulini, J.A. Yáñez, M. Cabanillas-Carbonell, The public health contribution of sentiment analysis of monkeypox tweets to detect polarities using the CNN-LSTM model, Vaccines 11 (2) (2023) 312.

[36] A.A. Ardakani, A.R. Kanafi, U.R. Acharya, N. Khadem, A. Mohammadi, Application of deep learning technique to manage COVID-19 in routine clinical practice using CT images: Results of 10 convolutional neural networks, Comput. Biol. Med. 121 (2020) 103795.

[37] L. Wang, Z.Q. Lin, A. Wong, Covid-net: A tailored deep convolutional neural network design for detection of covid-19 cases from chest x-ray images, Sci. Rep. 10 (1) (2020) 1–12.

[38] S.N. Ali, M. Ahmed, J. Paul, T. Jahan, S.M. Sani, N. Noor, T. Hasan, Monkeypox skin lesion detection using deep learning models: A feasibility study, 2022, arXiv preprint arXiv:2207.03342.

[39] V.H. Sahin, I. Oztel, G. Yolcu Oztel, Human monkeypox classification from skin lesion images with deep pre-trained network using mobile application, J. Med. Syst. 46 (11) (2022) 1–10.

[40] M. Chhabra, K.K. Ravulakollu, M. Kumar, A. Sharma, A. Nayyar, Improving automated latent fingerprint detection and segmentation using deep convolutional neural network, Neural Comput. Appl. 35 (9) (2023) 6471–6497.

[41] Y. LeCun, Y. Bengio, G. Hinton, Deep learning, Nature 521 (7553) (2015) 436–444.

[42] D. Scherer, A. Müller, S. Behnke, Evaluation of pooling operations in convolutional architectures for object recognition, in: International Conference on Artificial Neural Networks, Springer, Berlin, Heidelberg, 2010, pp. 92–101.

[43] S. Hochreiter, J. Schmidhuber, Long short-term memory, Neural Comput. 9 (8) (1997) 1735–1780.

[44] Sepp Hochreiter, Jürgen Schmidhuber, Long short-term memory, Neural computation 9 (8) (1997) 1735–1780.

[45] S. Jain, P. Harde, N. Mihindukulasooriya, S. Ghosh, A. Dubey, A. Bisht, Constructing a knowledge graph from Indian legal domain corpus, 2022.

[46] X. Huang, J. Zhang, D. Li, P. Li, Knowledge graph embedding based question answering, in: Proceedings of the Twelfth Association for Computing Machinery International Conference OnWeb Search and Data Mining, Melbourne, VIC, Australia, 2019, pp. 105–113.

[47] Y. Zhang, H. Dai, Z. Kozareva, A.J. Smola, L. Song, Variational reasoning for question answering with knowledge graph, in: Proceedings of the Thirty-Second Association for the Advancement of Artificial Intelligence Conference on Artificial Intelligence, New Orleans, LA, USA, 2018.

[48] B. Abu-Salih, M. AL-Qurishi, M. Alweshah, M. AL-Smadi, R. Alfayez, H. Saadeh, Healthcare knowledge graph construction: State-of-the-art, open issues, and opportunities, 2022, arXiv preprint arXiv:2207.03771.

[49] M. Rotmensch, Y. Halpern, A. Tlimat, S. Horng, D. Sontag, Learning a health knowledge graph from electronic medical records, Sci. Rep. 7 (2017) 5994, [CrossRef] [PubMed].

[50] K. Teru, E. Denis, W. Hamilton, Inductive relation prediction by subgraph reasoning, in: Proceedings of the International Conference on Machine Learning, 2023, pp. 9448–9457.

[51] F.A. Lovera, Y.C. Cardinale, M.N. Homsi, Sentiment analysis in Twitter based on knowledge graph and deep learning classification, Electronics 10 (22) (2021) 2739.

[52] N. Thakur, Monkeypox2022tweets: The first public Twitter dataset on the 2022 MonkeyPox outbreak, 2022, 2022060172, http://dx.doi.org/10.20944/preprints202206.0172.v1, Preprints.

[53] M. Li, E. Ch'ng, A.Y.L. Chong, S. See, Multi-class Twitter sentiment classification with emojis, Ind. Manage. Data Syst. 118 (9) (2018) 1804–1820.

[54] K.K. Mohbey, S. Sharma, S. Kumar, M. Sharma, COVID-19 identification and analysis using CT scan images: Deep transfer learning-based approach, in: Blockchain Applications for Healthcare Informatics, Academic Press, 2022, pp. 447–470.

[55] Y. Xie, F. Xing, X. Kong, H. Su, L. Yang, Beyond classification: structured regression for robust cell detection using convolutional neural network, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, Cham, 2015, pp. 358–365.

[56] D.M. Powers, Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation, 2020, arXiv preprint arXiv:2010.16061.

[57] K.K. Mohbey, Multi-class approach for user behavior prediction using deep learning framework on twitter election dataset, J. Data Inf. Manage. 2 (1) (2020) 1–14.