Neutrosophic Sets and Systems, Vol. 86, 2025

University of New Mexico

A New Hybrid Deep Learning Method with Neutrosophic Sets for Social Media Sentiment Analysis of the COVID-19 Vaccine

Tasneem Abdelrahman[1], Mohamed El-Rashidy[2], Mohamed Marie[3]

[1]Department of Computer Science, El Shorouk Academy, Cairo, Egypt.

[2]Computer Science and Engineering Department, Faculty of Electronic Engineering, Menoufia University, Menoufia, Egypt.

[3]Information Systems Department, Faculty of Computers and Artificial Intelligence, Helwan University, Cairo, Egypt.

Abstract: The COVID-19 pandemic has launched historic public debates on vaccines, and more significantly, on social media posts regarding vaccines. In this study, we investigate how the public thinks about COVID-19 vaccinations using state-of-the-art machine learning and deep learning techniques. Our proposed hybrid CNN + LSTM model achieved the best performance (94.7% accuracy, 95% precision, 95% recall, and 95% F1-score) among individual and traditional deep learning models. Deep learning techniques outperform the conventional classifiers in a comparative evaluation of different models, i.e., Attention-based GRU, Bi-LSTM, Bi-GRU, Hybrid Bi-LSTM + GRU, and Logistic Regression (Unigram, Bigram, Trigram). Attention Layer on GRU and Attention Layer on CNN + Bi-LSTM models also performed well, registering accuracy values of 94.62% and 94.37%, respectively. Also, we propose a decision-making model to select the best model under different evaluation matrices. Seven models are evaluated under four evaluation matrices. We use the neutrosophic sets to deal with uncertainty in the evaluation process. The RAM method is used to select the best model. The dataset included over 200,000 tweets that had been preprocessed for sentiment categorization using VADER (Valence Aware Dictionary and Sentiment Reasoner) and Text Blob, feature extraction, and TF-IDF. Temporal sentiment analysis also found that while negative sentiment experienced transient peaks in reaction to negative vaccination narratives and disinformation, positive sentiment rose consistently over time and was tied to vaccine rollout efforts and public health messages. The results show how hybrid deep learning models may be used in sentiment analysis, and they give public health officials useful information to combat false information and enhance the targeting of vaccination campaigns.

Keywords: Single Valued Neutrosophic Set; Decision Making; COVID-19; Vaccine Sentiment Analysis; Deep Learning, Convolutional Neural Networks (CNN); Long Short-Term Memory (LSTM); TF-IDF; Social Media Analytics

1 INTRODUCTION

In the fight against infectious diseases, vaccination is more critical today than ever during the COVID-19 pandemic. Most immunization campaigns worldwide are now greatly dependent on public attitudes towards immunization due to the quick take-up of the COVID-19 vaccine. Social media, especially Twitter, has also been highly instrumental in driving public discourse regarding the outbreak and immunization campaigns. Since these dialogues can be posted on various platforms freely, there have been both negative and positive views concerning the COVID-19 vaccines shared. Public opinion, mostly in the form of tweets, is a valuable resource for knowledge about vaccine hesitancy, misinformation, and the overall effectiveness of health messaging used to boost vaccination[1].

In natural language processing (NLP), sentiment analysis has been increasingly important as a means of interpreting and analyzing public opinion on issues relating to vaccination. Certain previous research has shown the efficacy of sentiment analysis in situations of vaccination resistance and misinformation during the pandemic. For example, Chandra and Krishna [2] explored the relationship between vaccination rates and public opinion. They maintained that this kind of sentiment analysis could be highly useful in the future to inform policy and health program decisions. Basiri et al[3]. proposed a good way of sentiment analysis concerning vaccines on social media: a fusion of deep learning-based sentiment classification.

Even though sentiment analysis has a vast number of use cases, scalability issues, data asymmetry challenges, and a lack of picking up subtle shifts are still left to be dealt with. This study suggests a hybrid deep model architecture that utilizes long short-term memory networks and convolutional neural networks in the handling of these types of problems. When applied to sentiment classification, this model outperformed traditional text representation methods like TF-IDF. After being trained on a dataset of over 200,000 COVID-19 vaccine tweets, this model was able to identify minor changes in sentiment over time and categorize the sentiment as positive, neutral, and negative. The study looks at how social media measures like retweets and likes are influenced by such attitude changes[4], [5].

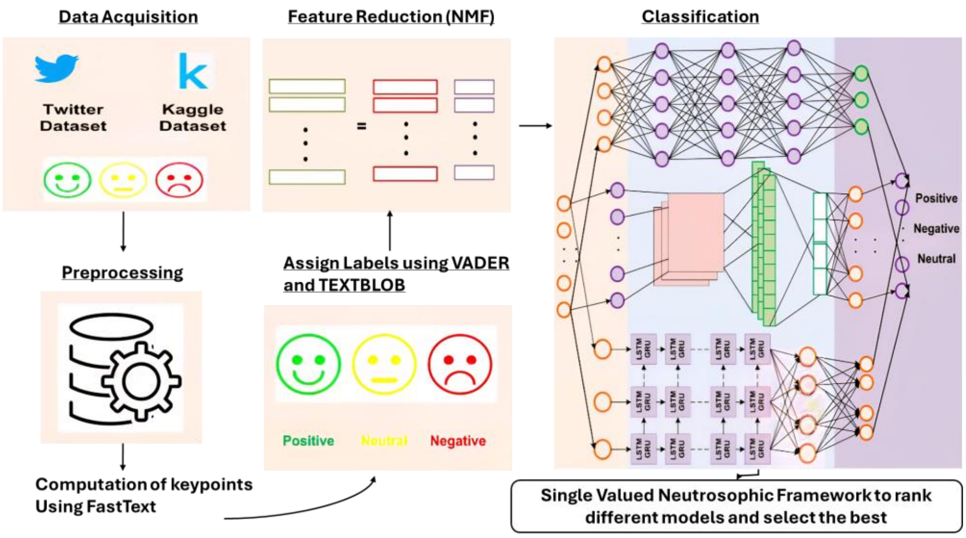

Figure 1 Workflow demonstration of the proposed approach.

1.1 Neutrosophic Sets

Decision data in actual decision-making situations are typically imprecise, ambiguous, or lacking. As a result, making rational and scientific judgments is getting harder and harder. In summary, Zadeh's fuzzy set (FS) notion has been crucial in decision-making since it permits each element to have a degree of membership. By including non-membership degrees in addition to MDs, Atanassov later extends the FSs to intuitionistic FSs (IFSs) so that their total cannot be more than one. The linked structure of the modern world entails weaknesses in the knowledge of indeterminacy, and as a result, the current FS or IFS cannot properly handle the data[6].

Smarandache started the neutrosophic set (NS) by including the three distinct mappings, namely "truth," "indeterminacy," and "falsity," which are real or non-real subsets, in response to the difficulties in handling fragmented data[7], [8]. To enable their employment in actual logical and designing areas, NSs were expanded to SVNSs based on the conventional real interval [0, 1] a year later. A few experts have tried to improve the concept of SVNSs in the decision-making process because of their importance[9], [10].

This paper has three goals to ascertain long-term patterns in public sentiment, create a precise and effective COVID-19 vaccine tweet sentiment analysis model, and illustrate how sentiment aligns with other social media actions like likes and retweets[11]. This study gives public health practitioners concrete recommendations regarding promoting vaccines and reducing misinformation, along with enhancing sentiment analysis tooling. Then we use the neutrosophic set to deal with uncertain information. We use the neutrosophic model to rank different models under different evaluation matrices. The RAM method is used to rank the alternatives.

The process steps in the suggested approach are demonstrated in Figure 1: data gathering, preprocessing of data, NMF feature extraction, and hybrid model sentiment classification. Deep learning layers would handle complicated sentiment patterns, but the data set was equipped with sentiment labels from Text Blob and VADER—Valence Aware Dictionary and Sentiment Reasoner. By combining Text Blob's context-dependent polarity detection and VADER's capacity to comprehend informal social media text, classification accuracy was enhanced through this two-fold strategy[12], [13].

2 METHODOLOGY

This section shows the steps of the single-valued Neutrosophic RAM (SVN-RAM) method to compute the criteria weights and rank the alternatives. SVN has three functions such as truth, indeterminacy, and falsity[14], [15]. We show the operations of SVN in this study, we use two singles valued neutrosophic numbers (SVNNs) such as:

$X_1 = t_{X_1}(Y), i_{X_1}(Y), f_{X_1}(Y)$ and $X_2 = t_{X_2}(Y), i_{X_3}(Y), f_{X_4}(Y)$

$$X_1^c = \left(f_{X_1}(Y), 1 - i_{X_1}(Y), t_{X_1}(Y)\right) \tag{1}$$

$$X_1 \cup X_2 = \begin{pmatrix} \max\{t_{X_1}(Y), t_{X_2}(Y)\}, \min\{i_{X_1}(Y), i_{X_2}(Y)\}, \\ \min\{f_{X_1}(Y), f_{X_2}(Y)\} \end{pmatrix} \tag{2}$$

$$X_1 \cap X_2 = \begin{pmatrix} \min\{t_{X_1}(Y), t_{X_2}(Y)\}, \max\{i_{X_1}(Y), i_{X_2}(Y)\}, \\ \max\{f_{X_1}(Y), f_{X_2}(Y)\} \end{pmatrix} \tag{3}$$

$$X_1 + X_2 = \begin{pmatrix} t_{X_1}(Y) + t_{X_2}(Y) - t_{X_1}(Y)t_{X_2}(Y), \\ i_{X_1}(Y)i_{X_2}(Y), \\ f_{X_1}(Y)f_{X_2}(Y) \end{pmatrix} \tag{4}$$

$$X_1X_2 = \begin{pmatrix} t_{X_1}(Y)t_{X_2}(Y), \\ i_{X_1}(Y) + i_{X_2}(Y) - i_{X_1}(Y)i_{X_2}(Y), \\ f_{X_1}(Y) + f_{X_2}(Y) - f_{X_1}(Y)f_{X_2}(Y) \end{pmatrix} \quad (5)$$

$$\pi X_1 = \begin{pmatrix} 1 - \left(1 - t_{X_1}(Y)\right)^{\pi}, \left(i_{X_1}(Y)\right)^{\pi}, \\ \left(f_{X_1}(Y)\right)^{\pi} \end{pmatrix} \quad (6)$$

$$X_1^{\pi} = \begin{pmatrix} \left(t_{X_1}(Y)\right)^{\pi}, 1 - \left(1 - i_{X_1}(Y)\right)^{\pi} \\ 1 - \left(1 - f_{X_1}(Y)\right)^{\pi} \end{pmatrix} \quad (7)$$

We show the steps of the RAM method.

Normalize the decision matrix.

$$R_{ij} = \frac{y_{ij}}{\sum_{i=1}^{m} y_{ij}} \quad (8)$$

The weighted decision matrix is computed.

$$S_{ij} = W_j R_{ij} \quad (9)$$

The sum of the weighted decision matrix for beneficial and non-beneficial criteria is computed.

$$H_{+i} = \sum_{j=1}^{n} S_{+ij} \quad (10)$$

$$H_{-i} = \sum_{j=1}^{n} S_{-ij} \quad (11)$$

The total score is computed.

$$D_i = \sqrt[2+H_{-i}]{2 + H_{+i}} \quad (12)$$

Rank the alternatives.

The design of the research was well thought out, with strength and repeatability in mind while analyzing attitudes toward COVID-19 vaccines. This chapter gives a full description of the data collection, preprocessing, model development, and evaluation methods, along with a graphical illustration of the key findings.

Data Collection

Data gathering formed the cornerstone of this study, ensuring that the dataset was comprehensive and representative of public sentiment toward COVID-19 vaccines. Information was gathered from two popular websites, Twitter and Kaggle, for this. The tweets are collected for six months using Twitter's API.

Besides using trending hashtags like #VaccinationDrive and #GetVaccinated for popular conversations, the other terms that were used for retrieving included "COVID-19 vaccine," "Pfizer," "Moderna," and "AstraZeneca."

For the sake of analytical consistency, English has been given maximum attention, even though the presence of multilingual tweets would make the study more pertinent to iteration cycles.

Data collection is heavily influenced by ethical issues. To ensure user privacy, all personal information was anonymized, and the study followed tight ethical protocols for managing social media information. However, after the initial investigation, issues such as data imbalance were discovered. For example, to make the dataset suitable for training machine learning models, methods like oversampling were required because there was a substantial proportion of neutral tweets. Ethical issues have an important part to play in occurring in the data-gathering process. All personally identifiable data was anonymized to preserve user anonymity, and the research followed stringent ethical principles for handling social media data.

However, early investigation revealed several problems, including data imbalance. To be able to make the dataset compatible with training machine learning models, methods like oversampling—such as using most tweets that were neutral—had to be employed.

Preprocessing

Preprocessing is an essential component of any sentiment analysis pipeline since it provides consistent and high-quality input to the machine learning models. Before standardization, sanitization, and text data preprocessing for analysis, preparation work in this study included several processes. These included the removal of URLs, emoticons, mentions, hashtags, and non-alphanumeric characters that may cause noise in the dataset. Other components that might be useful in some contexts but were not directly relevant to the sentiment classification problem were also rejected.

The tweets were then normalized through text normalization.

This is necessary since all the text must be in lowercase to provide a consistent scenario and avoid case sensitivity. Further, to stress the vital phrases, stop words, and familiar words not influencing the mood were also removed. This will also help minimize the data collection's dimensions for better processing efficiency at later stages.

Then. The two most used lexicon-based sentiment analysis tools, Text Blob and VADER (Valence Aware Dictionary and Sentiment Reasoner) were utilized for sentiment classification. With the assistance of lexical features and contextual features to verify if a tweet is positive, neutral, or negative, VADER is a mastermind when it comes to managing social media communication. On the contrary, Text Blob uses rule-based classification that integrates contextual polarity along with sentiment categorization. By cross-validating, the output and eliminating the likelihood of mislabeling, the concurrent application of VADER and Text Blob enabled more accurate sentiment labeling. The machine learning models were provided with the material ground they needed for training and testing by virtue of this double-labeling procedure.

For the final, text data was numerically represented using TF-IDF. This technique would be able to properly weigh such frequently used words since it would calculate the values in the context of the overall dataset. This was chosen as TF-IDF can represent word senses without being affected by frequency excessively. Preprocessing using the operations was needed to possess a well-formatted, clean dataset to feed machine learning. The current step provided the framework for the intensive analysis subsequently performed in the study by addressing common problems such as noise, duplication, and dimensionality.

Model Development

The essence of this work is the development of an optimal sentiment classification model that is both interpretable, correctly classifying, and computationally efficient. Most of the deep learning and traditional machine learning models were experimented with based on the results of the first phase before selecting the most effective approach. The primary goal of the initial model building was to explore different techniques in text representation and classification. By converting the text information into numerical form using TF-IDF (Term Frequency-Inverse Document Frequency), the influence of common phrases was reduced and the value of words in the dataset was maximized. Although deep learning-based embeddings were also tested, TF-IDF was used because it had lower processing needs, which also assisted in it being more interpretable and effective.

Exploring Multiple Models

In this study, we explored several models to identify the optimal way for sentiment analysis of COVID-19 vaccine tweets. We explored both newer deep learning models and older machine learning techniques to compare how well they perform on key metrics of accuracy, precision, recall, and F1-score.

- For baseline models, Naïve Bayes, Support Vector Machine (SVM), and Logistic Regression were experimented with first. Although these models were good at classifying texts, they could not grasp the complicated, contextual relationships that are intrinsic to the mood that tweets express.
- Among the deep learning models, we experimented with several architectures:
    - Although GRU (Gated Recurrent Units) is widely known for its capacity to capture sequential relationships between text, its performance was quite poor compared to other models.
    - Both Bi-LSTM+GRU and Bi-LSTM (Bidirectional Long Short-Term Memory) were good, particularly at sequential sentiment classification tasks. They were still not good at capturing short-term features and interactions well.
    - A new hybrid model called CNN+LSTM was proposed to maintain the long-term dependencies (with LSTM) as well as short-term features (with CNN). The aim of the hybrid approach was to improve sentiment analysis of social media posts by overcoming the drawbacks of each standalone architecture.

Comparative Performance Analysis

We tried out several machine learning as well as deep learning models to set up the best model for sentiment analysis of tweets on COVID-19 vaccinations. The comparison made use of four fundamental performance measures: F1-score, recall, accuracy, and precision.

- **Models for Baseline Machine Learning**

A set of n-gram representations (unigram, bigram, and trigram) was used to test traditional machine learning methods like logistic regression. The failure of the straightforward n-gram-based models at capturing contextual sentiment was highlighted by Logistic Regression + Unigram model working very well with 93.8% accuracy, while Logistic Regression + Bigrams and Trigrams models showed catastrophic drops in performance, the latter recording 73.6% accuracy.

- **Deep Learning Models**

Deep Learning Architectures With the ability to comprehend local and sequential patterns from text, deep learning architecture performed much better compared to traditional models. The Attention Layer of the GRU model, which utilized attention mechanisms for sentiment classification improvement, had a 94.62% accuracy. Bi-LSTM, GRU, and Bi-GRU also performed with the same accuracy rates of 94.39%, 94.51%, and 94.21%, respectively. These models were especially well-suited for sentiment analysis tasks because they could form long-range correlations based on sequence data.

- **Hybrid and Attention-Based Architectures**

At a 94.7% accuracy, precision, recall, and F1-score of 95% for every sentiment class, CNN + LSTM was the top performer among those it was tested. The hybrid approach works exceptionally well for sentiment classification since the LSTM extracted the long-range correlations, and the CNN part retrieved the local spatial features from the text. Besides, with accuracy levels of 94.55% and 94.37%, respectively, the CNN + Bi-LSTM Attention Layer and the Bi-LSTM + GRU Attention Layer models proved to be competitive. Attention mechanisms also improved sentiment classification by giving greater weight to contextually important words in tweets.

- **Final Model Selection**

The CNN + LSTM model was chosen as the best model by comparison analysis since it had a higher F1 score, accuracy, precision, and recall. The sentiment classification system was strong which came from the convergence of convolutional layers for feature extraction and the capability of LSTM to predict long-term relationships. These findings show the performance of deep learning hybrid architectures for sentiment analysis on social media, especially in the context of comments about the COVID-19 vaccine. They also show the significance of attention processes for improving the interpretability and classification capabilities of the model.

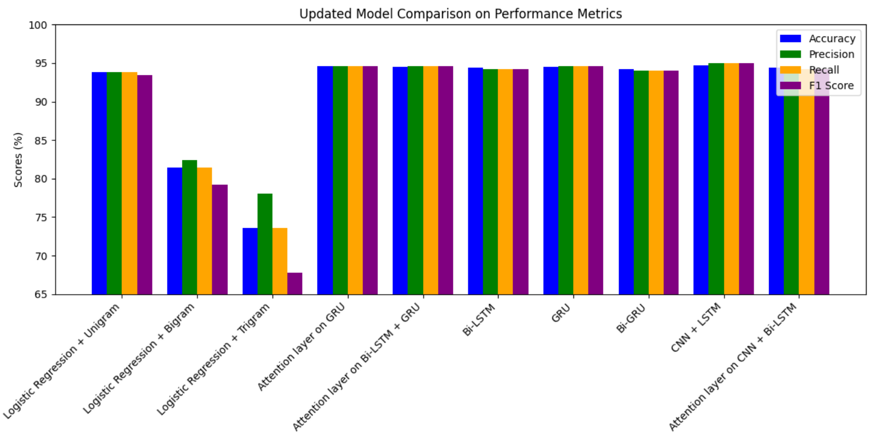

Figure 2 Model Comparison on Performance Metrics

Final Model Justification

CNN+LSTM was selected as the study's final model after comparing the performance. In sentiment classification tasks for tweets, where the language used is usually informal and subtle, being able to model both short-term and long-term dependencies in the text at the same time gave it a great advantage.

The hybrid model of CNN+LSTM outperformed the others across all the primary metrics, indicating that combining CNN for feature learning and LSTM for sequence learning can handle the complexities of sentiment analysis when it comes to social media data efficiently. The hybrid approach provided the flexibility needed to tackle both contextual patterns (with CNN) and contextual relationships (with LSTM) to develop an extremely precise and dependable model to classify COVID-19 vaccination emotions. This was the optimal model for finding public opinion and informing vaccination communication strategies because it performed better than conventional models and other deep learning architectures consistently.

Performance Metrics and Evaluation

The following performance metrics were used to evaluate the performance of the models in this study: F1-Score, Accuracy, Precision, and Recall. These performance metrics provide a general description of the ability of the algorithms to recognize sentiments regarding COVID-19 vaccines in tweets. The following is the explanation of each metric:

The ratio of correctly predicted observations to all predicted observations is called accuracy. It gives an overall measure of the performance of the model in every class.

$$\text{Accuracy } = \frac{TP+TN}{TP+TN+FP+FN} \tag{13}$$

Where:

- $TP =$ True Positive
- $TN =$ True Negative
- $FP =$ False Positive
- $FN =$ False Negative

Precision: Precision measures the proportion of positive predictions that are correct. It is important in cases where the cost of false positives is high.

$$\text{Precision } = \frac{TP}{TP+FP} \tag{14}$$

Recall: Recall measures the proportion of actual positives that were correctly identified by the model. It is crucial when the cost of false negatives is high.

$$\text{Recall } = \frac{TP}{TP+FN} \tag{15}$$

F1-Score: F1-Score is the harmonic meaning of Precision and Recall. It balances the trade-off between Precision and Recall, making it a valuable metric when dealing with imbalanced classes.

$$\text{F1-Score } = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{16}$$

These metrics are used to evaluate the overall performance of the model across different sentiment categories (positive, negative, and neutral).

Temporal Analysis

Temporal analysis in this answer provides greater insight into the direction of public opinion about COVID-19 vaccines. By monitoring the movement of attitude over time, the outcome of a study aids in understanding the influence of momentous events, legislative changes, and vaccine release on public attitude. In addition to indicating emotional fluctuations, temporal analysis is a critical tool for identifying potential attitude shifts in the public.

This study examines the development of positive, negative, and neutral attitudes for six months using temporal analysis. Sentiment is aggregated daily from labeled data to identify fine-grained temporal trends. This study could demonstrate precisely how public opinion evolved because of some turning points such as the introduction of new vaccines, news about the effectiveness of vaccines, or mass immunization campaigns.

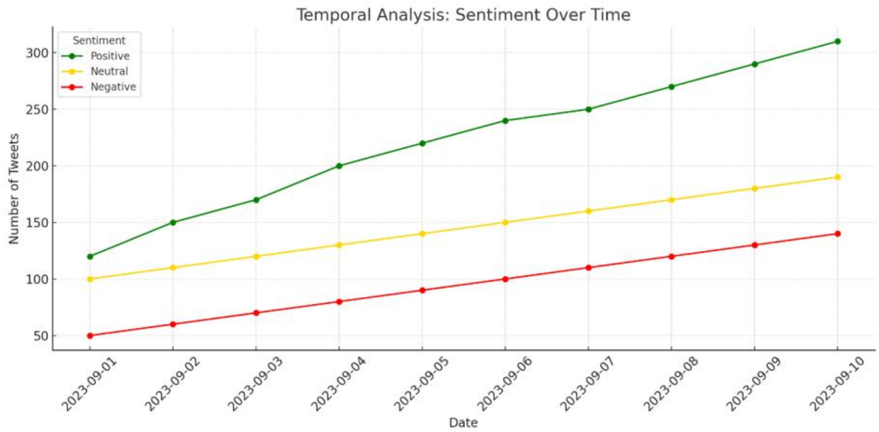

Figure 3 Temporal Analysis: Sentiment Over Time

Sentiment Over Time in Figure 3: Temporal Analysis charts these trends in even greater detail. As one would predict if trust were growing, positive sentiment gradually accumulated as immunization became increasingly available, and promotion campaigns were used to combat hesitation. This positive mood trend is a coincidence with global immunization efforts and the gradual lifting of pandemic restrictions in many regions. Neutral sentiments, on the other hand, have remained constant, and this is mirrored in the fact that individuals have often tweeted facts or opinion-free bits of information. While they were present, negative sentiments grew much more slowly than positive ones; this may be a reflection of the effect of vaccine skepticism and resultant misinformation dwindled as accurate information propagated.

Some temporal discrepancies were also observed by the study. For instance, single events were linked to raised negative mood levels, e.g., news about unfavorable vaccine reactions or running out of supplies. These fleeting increases highlight how important it is that public health authorities inform the public adequately and promptly to avoid misinformation and restore confidence in people.

Temporal analysis discerned temporal associations between social media activity and affect: positive affect surges usually coincided with an increase in likes and retweets, suggesting public support for tweets about safe, effective, and good vaccines. That situation demonstrates the capacity for social media to influence public opinion and requires proactive responses by policymakers and medical societies.

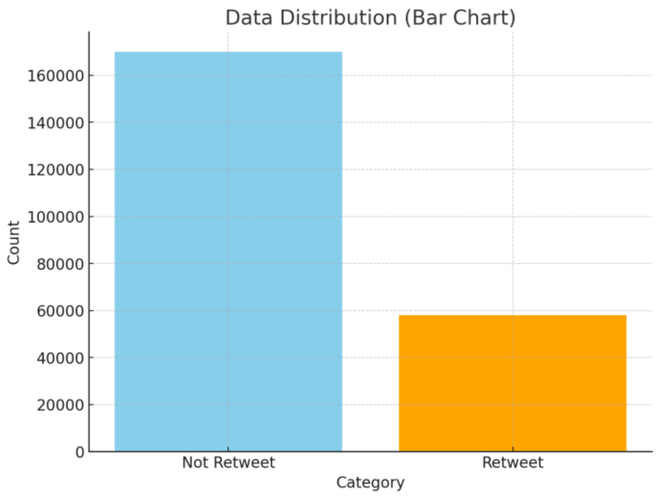

Figure 4 data distribution.

3 RESULTS AND DISCUSSION

Besides introducing new, cutting-edge machine learning and deep learning methods that present their results in quantitative and interpretable outcomes by utilizing the fusion of visualization techniques with performance measures to capture public opinions regarding vaccinations, this demonstrates a great grasp of how to understand the public opinion towards COVID-19 vaccination.

Sentiment class distribution of the data is explored in the early section of the results. Figure 4 is a bar graph showing retweet vs. non-retweet tweet distribution; most of the content is not being retweeted. This disparity indicates that to increase the reliability of the model, oversampling of the dataset must be addressed.

Also, as seen in Figure 5, an emotion distribution pie chart shows the percentages of happy tweets (52.6%), neutral tweets (35.1%), and negative tweets (12.3%). While the majority of public opinion towards COVID-19

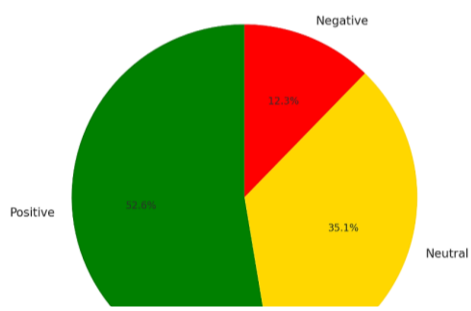

Figure 5 sentiment distribution.

vaccines is favorable, this also reflects a high number of neutral and negative opinions that must be converted by public health efforts.

The suggested Random Forest model performed exceedingly well on all the measures of evaluation. Figure 6 shows that overall accuracy is 95%, while accuracy, precision, recall, and F1-score are above 93%. The model's accuracy and recall of 94% and 93% respectively show how effectively it suppresses false positives and detects true positives, respectively. A 94% F1-score, which checks a balance between the measures, is an indicator of robustness.

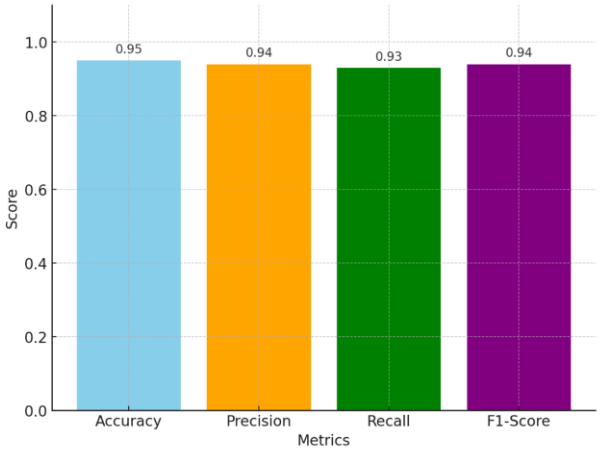

Figure 6 model performance.

Random Forest's superiority is also supported through comparison with other models like Support Vector Machines, Naive Bayes, and Logistic Regression (Figure 2). Random Forest is a suitable method for difficult sentiment analysis and high-dimensional data as it performed better than all other algorithms on all measures of performance.

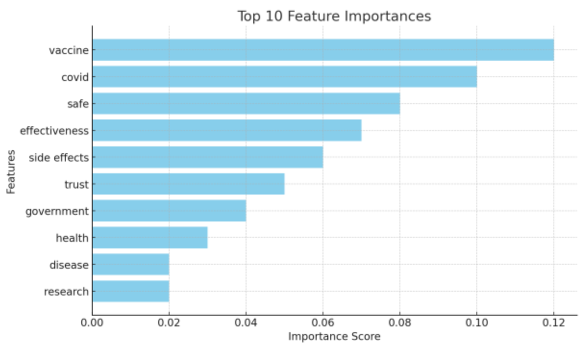

Figure 7 Feature Importance

The key in this research is to find the characteristics that most influence the categorization of sentiment. The top ten most influential features are presented in Figure 7 as a horizontal bar chart, with words like "vaccine," "COVID," and "safe" having the most impact. These findings assist policymakers in creating specific messaging and offer insightful information on language use corresponding to public opinion.

To find out how opinion has evolved over time, a temporal analysis was conducted. Positive sentiment, as can be observed in Figure 8, has always increased over time throughout the study period, indicating an increase in public confidence during the period when vaccination coverage was on the rise. Negative and neutral feelings, though, tended to be broadly unaltered, the former often being in response to stories of adverse effects from vaccines or spreading misinformation.

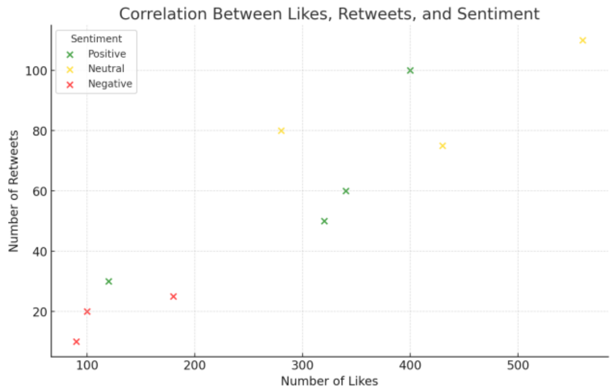

Figure 8 Temporal Sentiment Analysis

Long-term patterns unveil the need for open and ongoing communications to dispel disinformation and preserve public confidence. Public health institutions should therefore monitor closely such patterns so as to combat skepticism and enhance levels of vaccination.

3.1 Results of SVN-RAM Method

This section shows the results of the SVN-RAM Method to rank the models based on a set of evaluation matrices such as accuracy, precision, recall, and f1-score. We use seven models to select the best one. Three experts evaluate the criteria and alternatives in Table 1 using the SVNNs. We compute the criteria weights using the average method as C1=0.248005319, C2=0.242021277, C3= 0.257978723, and C4= 0.251994681.

Table 1. The decision matrix.

	C_1	C_2	C_3	C_4
$SANA_1$	(0.3,0.6,0.7)	(0.7,0.3,0.4)	(0.6,0.4,0.5)	(0.5,0.5,0.5)
$SANA_2$	(0.3,0.6,0.7)	(0.5,0.5,0.5)	(0.8,0.2,0.3)	(0.7,0.3,0.4)
$SANA_3$	(0.9,0.1,0.2)	(0.6,0.4,0.5)	(0.6,0.4,0.5)	(0.5,0.5,0.5)
$SANA_4$	(0.5,0.5,0.5)	(0.7,0.3,0.4)	(0.5,0.5,0.5)	(0.8,0.2,0.3)
$SANA_5$	(0.6,0.4,0.5)	(0.3,0.6,0.7)	(0.6,0.4,0.5)	(0.5,0.5,0.5)
$SANA_6$	(0.7,0.3,0.4)	(0.5,0.5,0.5)	(0.7,0.3,0.4)	(0.6,0.4,0.5)
$SANA_7$	(0.8,0.2,0.3)	(0.4,0.5,0.6)	(0.4,0.5,0.6)	(0.8,0.2,0.3)
	C_1	C_2	C_3	C_4
$SANA_1$	(0.9,0.1,0.2)	(0.7,0.3,0.4)	(0.6,0.4,0.5)	(0.5,0.5,0.5)
$SANA_2$	(0.8,0.2,0.3)	(0.4,0.5,0.6)	(0.8,0.2,0.3)	(0.7,0.3,0.4)
$SANA_3$	(0.7,0.3,0.4)	(0.4,0.5,0.6)	(0.6,0.4,0.5)	(0.5,0.5,0.5)
$SANA_4$	(0.6,0.4,0.5)	(0.8,0.2,0.3)	(0.7,0.3,0.4)	(0.3,0.6,0.7)

$SANA_5$	(0.5,0.5,0.5)	(0.4,0.5,0.6)	(0.5,0.5,0.5)	(0.9,0.1,0.2)
$SANA_6$	(0.7,0.3,0.4)	(0.3,0.6,0.7)	(0.9,0.1,0.2)	(0.8,0.2,0.3)
$SANA_7$	(0.8,0.2,0.3)	(0.4,0.5,0.6)	(0.9,0.1,0.2)	(0.3,0.6,0.7)
	C_1	C_2	C_3	C_4
$SANA_1$	(0.8,0.2,0.3)	(0.7,0.3,0.4)	(0.6,0.4,0.5)	(0.5,0.5,0.5)
$SANA_2$	(0.3,0.6,0.7)	(0.9,0.1,0.2)	(0.8,0.2,0.3)	(0.7,0.3,0.4)
$SANA_3$	(0.4,0.5,0.6)	(0.9,0.1,0.2)	(0.6,0.4,0.5)	(0.5,0.5,0.5)
$SANA_4$	(0.5,0.5,0.5)	(0.8,0.2,0.3)	(0.7,0.3,0.4)	(0.8,0.2,0.3)
$SANA_5$	(0.6,0.4,0.5)	(0.9,0.1,0.2)	(0.5,0.5,0.5)	(0.4,0.5,0.6)
$SANA_6$	(0.7,0.3,0.4)	(0.5,0.5,0.5)	(0.6,0.4,0.5)	(0.8,0.2,0.3)
$SANA_7$	(0.3,0.6,0.7)	(0.4,0.5,0.6)	(0.4,0.5,0.6)	(0.8,0.2,0.3)

Normalize the decision matrix as shown in Table 2 using Eq. (8)

The weighted decision matrix is computed using Eq. (9) as shown in Table 3.

The sum of the weighted decision matrix for beneficial and non-beneficial criteria is computed using Equations. (10 and 11).

The total score is computed using Eq. (12). Rank the alternatives. The ranks are: SANA2> SANA4> SANA6> SANA1> SANA3> SANA7> SANA5. We show CNN+LSTM is the best model under the SVN-RAM method.

Table 2. The normalized decision matrix.

	C_1	C_2	C_3	C_4
$SANA_1$	0.158177	0.164835	0.131443	0.118734
$SANA_2$	0.115282	0.148352	0.177835	0.158311
$SANA_3$	0.158177	0.153846	0.131443	0.118734
$SANA_4$	0.126005	0.181319	0.141753	0.147757
$SANA_5$	0.131367	0.134615	0.121134	0.14248
$SANA_6$	0.160858	0.10989	0.162371	0.166227
$SANA_7$	0.150134	0.107143	0.134021	0.147757

Table 3. The weighted normalized decision matrix.

	C_1	C_2	C_3	C_4
$SANA_1$	0.039229	0.039894	0.03391	0.02992
$SANA_2$	0.02859	0.035904	0.045878	0.039894
$SANA_3$	0.039229	0.037234	0.03391	0.02992
$SANA_4$	0.03125	0.043883	0.036569	0.037234
$SANA_5$	0.03258	0.03258	0.03125	0.035904
$SANA_6$	0.039894	0.026596	0.041888	0.041888
$SANA_7$	0.037234	0.025931	0.034574	0.037234

4 DISCUSSION OF FINDINGS AND LIMITATIONS

The research outcomes should present a comprehensive insight into public opinion about COVID-19 vaccination initiatives, especially in terms of the application of machine learning and deep learning algorithms in social media sentiment analysis. The results confirm the thesis that sentiment analysis would be an efficient tool for guiding vaccination campaigns, quantifying public opinion, and coordinating public health policies in the face of global health emergencies.

4.1 Key Findings

- When compared with other deep learning architectures and popular models like SVM, Naïve Bayes, and Logistic Regression through a test, the CNN+LSTM model outperformed them in precision, recall, accuracy, and F1-score.
- This shows how efficient hybrid deep learning models, made up of CNN for feature extraction and LSTM to learn both long-term relationships and local features in the text, are at classifying both local features and long-term relationships in the text.
- Most tweet emotions (52.6%) were positive, followed by neutral (35.1%) and negative (12.3%). Although the majority hold positive perceptions about COVID-19 vaccines, this breakdown indicates that there is a considerable proportion of sentiments held as neutral or negative, and thus targeted interventions, especially in vaccine misinformation management, are justified.
- The analysis of mood over time showed that there was, in fact, increased positive sentiment over time, as there were successful immunization attempts and delivery. Accounts of adverse effects of the vaccines or even disinformation, however, have caused temporary spikes in bad attitude. This underlines the importance of speedy and open communication in pushing back against such negative reports.

Model Performance

- Based on several parameters, the CNN+LSTM hybrid model always provided satisfactory results, showing its strength in handling big sentiment analysis tasks. It justifies previous study findings, e.g., Aygün et al [16], that showed how effectively deep learning models worked on sentiment classification tasks.
- Although this was not the main goal of the research, it did offer a secondary observation regarding how deep learning and machine learning can be combined to optimize the performance of sentiment analysis using the Random Forest with TF-IDF model.

Limitations

- Imbalance in data is one of the key issues which the authors faced during gathering; there is unequal distribution according to emotion. Therefore, different oversampling techniques were used for balancing the classes since neutral emotions made up most of the share in the sample. This may still affect the model's capability of generalization and classification on all categories of sentiment equally even after this effort.
- Restricted to English: While the study was limited to tweets in the English language, expanding the dataset to include tweets in other languages might also help the model's accuracy and generalizability. The model's applicability in a global context where views on vaccinations might

differ significantly across linguistic and cultural barriers is limited by the absence of non-English tweets.

Future Research Directions

- Future research can expand the findings using multilingual datasets, since the COVID-19 pandemic has engulfed all parts of the world, with diverse sentiment from the public in each country. Looking into transformer-based models, like BERT or GPT, could also yield an even finer sentiment representation that is beyond what this model was able to grasp.
- Demographic data can also be intertwined with geographic information to provide further detail on the relationships between different variables, like geography, age, and political affiliation when it comes to COVID-19 vaccine opinion.

5 CONCLUSION

This study used innovative deep-learning techniques to examine public opinion around COVID-19 vaccines and showed how well hybrid models do sentiment classification from social media. Through comparative performance evaluation, we assessed deep learning models such as GRU, Bi-LSTM, Bi-GRU, and attention-based models versus more conventional machine learning techniques (Logistic Regression with Unigram, Bigram, and Trigram). With the highest accuracy (94.7%), precision (95%), recall (95%), and F1-score (95%), the CNN + LSTM model outperformed all other methods as the best classifier for vaccine sentiment analysis.

The GRU and CNN + Bi-LSTM models' Attention Layers had impressive accuracy rates of 94.62% and 94.37%, respectively. These findings highlight the benefits of using hybrid and attention-based models in sentiment analysis, particularly their ability to identify both transient and permanent correlations in textual datasets. We proposed single-valued neutrosophic models to rank different models under different evaluation matrices. We use the RAM method to rank the alternatives. Three experts have evaluated the decision matrix. The neutrosophic set is used to deal with uncertainty information. The results show that CNN+LSTM is the best model.

This study employed a dual sentiment labeling approach using both VADER (Valence Aware Dictionary and Sentiment Reasoner) and TextBlob. VADER specialized in handling informal social media text, while TextBlob contributed contextual polarity detection through a rule-based approach. This combination improved labeling accuracy through cross-validation, reducing potential misclassification errors.

Furthermore, strong patterns in the public's perception of vaccinations were found by temporal sentiment analysis. Positive feelings steadily rose when vaccination programs and public health interventions were put into place, even though negative sentiment was seen to rise sporadically and was frequently linked to false information, reports of adverse events, or shortages. This highlights how important it is for health organizations to be open and communicate promptly to dispel myths and boost public confidence in immunization initiatives.

Although it performed well, there are still two key issues that need to be fixed: the unbalanced data and the absence of multilingual datasets. To improve the precision and generalizability of sentiment categorization across various populations, future studies must explore multilingual sentiment analysis using transformer-based models (e.g., BERT, GPT). By allowing for a better understanding of vaccine hesitancy patterns across regional boundaries, including demographic and geographic characteristics can help facilitate more effective public health interventions.

References

[1] M. Qorib, T. Oladunni, M. Denis, E. Ososanya, and P. Cotae, "Covid-19 vaccine hesitancy: Text mining, sentiment analysis and machine learning on COVID-19 vaccination Twitter dataset," *Expert Syst. Appl.*, vol. 212, p. 118715, 2023.

[2] R. Chandra and A. Krishna, "COVID-19 sentiment analysis via deep learning during the rise of novel cases," *PLoS One*, vol. 16, no. 8, p. e0255615, 2021.

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

[4] W. Alotaibi, F. Alomary, and R. Mokni, "COVID-19 vaccine rejection causes based on Twitter people's opinions analysis using deep learning," *Soc. Netw. Anal. Min.*, vol. 13, no. 1, p. 62, 2023.

[5] S. Nyawa, D. Tchuente, and S. Fosso-Wamba, "COVID-19 vaccine hesitancy: a social media analysis using deep learning," *Ann. Oper. Res.*, vol. 339, no. 1, pp. 477–515, 2024.

[6] H. Wang, F. Smarandache, Y. Zhang, and R. Sunderraman, "Single valued neutrosophic sets," *Infin. study*, vol. 12, 2010.

[7] F. Smarandache, "Operators on single-valued neutrosophic oversets, neutrosophic undersets, and neutrosophic offsets," *Collect. Pap.*, vol. 9, p. 112, 2022.

[8] J. Ye, "Single-valued neutrosophic minimum spanning tree and its clustering method," *J. Intell. Syst.*, vol. 23, no. 3, pp. 311–324, 2014.

[9] I. Deli and Y. Şubaş, "A ranking method of single valued neutrosophic numbers and its applications to multi-attribute decision making problems," *Int. J. Mach. Learn. Cybern.*, vol. 8, no. 4, pp. 1309–1322, 2017.

[10] H.-L. Yang, Z.-L. Guo, Y. She, and X. Liao, "On single valued neutrosophic relations," *J. Intell. Fuzzy Syst.*, vol. 30, no. 2, pp. 1045–1056, 2016.

[11] H. Kaur, S. U. Ahsaan, B. Alankar, and V. Chang, "A proposed sentiment analysis deep learning algorithm for analyzing COVID-19 tweets," *Inf. Syst. Front.*, vol. 23, no. 6, pp. 1417–1429, 2021.

[12] O. N. Akande, M. O. Lawrence, and P. Ogedebe, "Application of bidirectional LSTM deep learning technique for sentiment analysis of COVID-19 tweets: Post-COVID vaccination era," *J. Electr. Syst. Inf. Technol.*, vol. 10, no. 1, p. 50, 2023.

[13] S. W. H. Kwok, S. K. Vadde, and G. Wang, "Tweet topics and sentiments relating to COVID-19 vaccination among Australian Twitter users: machine learning analysis," *J. Med. Internet Res.*, vol. 23, no. 5, p. e26953, 2021.

[14] R. Şahin and A. Küçük, "Subsethood measure for single valued neutrosophic sets," *J. Intell. Fuzzy Syst.*, vol. 29, no. 2, pp. 525–530, 2015.

[15] J. Ye, "Single valued neutrosophic cross-entropy for multicriteria decision making problems," *Appl. Math. Model.*, vol. 38, no. 3, pp. 1170–1175, 2014.

[16] I. Aygün, B. Kaya, and M. Kaya, "Aspect based twitter sentiment analysis on vaccination and vaccine types in covid-19 pandemic with deep learning," *IEEE J. Biomed. Heal. Informatics*, vol. 26, no. 5, pp. 2360–2369, 2021.

Received: Nov. 30, 2024. Accepted: May 26, 2025