Deep Learning-Based Sentiment Analysis of COVID-19 Pfizer Vaccine Tweets Using Transformer and Bi-LSTM Architectures

Zilong Xia[1, 2, *], Fanchao Niu[3]
[1]Faculty of Education, Qufu Normal University, Shandong, Jining, Shandong Province, 273165, China
[2]Development and Planning Office, Chengdu Technological University, Chengdu, 611730, China
[3]School of Continuing Education, Chengdu Technological University, Chengdu, 611730, China
E-mail: colexia@163.com
[*]Corresponding author

Keywords: COVID-19, vaccines, twitter tweets, pfizer, deep learning

Received:

Starting in December 2019, the COVID-19 virus has impacted economies globally, infected billions of people worldwide, and created a global health disaster. The discovery of vaccines against SARS-CoV-2, the virus responsible for COVID-19, has proven safe and successful in combating the epidemic. As of July 2021, there were 184 vaccine candidates in preclinical development, 105 in clinical testing, and 18 vaccinations that were authorized for use in emergencies. These efforts represent the hard work of the scientific community in combating the pandemic. Language processing tactics can be used for the guidance of health communication strategies and to reduce misinformation. This investigation focuses on the emotion analysis of Pfizer vaccines using data from the Twitter platform based on different deep-learning methods and transformers. The dataset used in this study includes 11,021 tweets from the Twitter platform, collected from Kaggle, related to the Pfizer and BioNTech vaccines. The survey analyzes recent tweets to find out what people are saying about the Pfizer and BioNTech vaccines. The database was divided based on tweets related to Pfizer vaccines, which were categorized using DL frameworks. The sentiment distribution provides an overview of the opinions on positive, negative, and neutral comments. This can be represented using graphical and chart representations such as word clouds, ROC curves, and precision-recall curves. Deep learning models are employed for sentiment analysis, including Transformers and Bi-LSTM models.Various models, including DistilBERT, Google Electra-base, Bi-LSTM, and other Transformers, were utilized for this analysis. the results are reported using metrics such as accuracy, F1-score, and other evaluation metrics. The sentiment analysis results from the models show that Model DistilBERT outperformed the others in both accuracy (0.92) and F1-score (0.91), as depicted in the bar chart, where DistilBERT had the highest performance across all models. Such analyses will help healthcare providers, policymakers, and the general public understand the overall sentiment of Pfizer vaccines.

Povzetek: Narejena je analiza sentimenta tvitov o Pfizerjevem cepivu z metodami DistilBERT, ELECTRA in Bi-LSTM. DistilBERT doseže najboljše rezultate na 11.021 tvitih.

1 Introduction

Social media can be a very strong channel not only for communication and content sharing but also for relationship building. Employing social media for the monitoring of vaccine adverse effects could extend public health and safety with real-time data and open discussion. [1].

This COVID-19 epidemic originally came from Wuhan, China, and then spread to several 213 countries and territories of the world. On February 17, 2020, the WHO reported that 80% of COVID-19 patients show mild symptoms, such as fever, and recover without special treatment. It has a reported mortality rate of about 2%, which is rather low compared to other coronavirus-related diseases [2].

Among the vaccines developed against COVID-19, some are generated with great effort by many researchers, out of which the development by BioNTech and Pfizer exhibits extraordinary effectiveness and safety. They have been in wide distribution and use in different regions, especially in Europe and Canada, for their indispensable contributions towards the combat against the pandemic. These vaccines have played an indispensable role in the global fight against the pandemic. BioNTech's collaboration with Pfizer in developing and distributing their COVID-19 vaccine is among the most telling on record in medical science and public health. Full FDA approval on August 23, 2021, for individuals aged 16 and older underlines established safety and effectiveness requisite ingredients in instilling confidence in the vaccine's use as a crucial device in the ongoing fight against the COVID-19 pandemic [3].

Comirnaty, developed by Pfizer-BioNTech, is the first mRNA-based vaccine, purposed to enhance immunity against infection with COVID-19. Although highly effective against COVID-19 infection and grave illness, a wide array of side effects has also been reported for this vaccination [4].

Indeed, sentiment analysis does sit at the crossroads of linguistics and computer science, using the latter to make clear distinctions and classify the emotive content within the text. Applications also span many fields, right from customer feedback regarding certain products or services to public sentiments expressed across social media platforms. This is a worthy insight for businesses, marketers, and researchers by automating the analysis of sentiments in texts. Thus, it helps in understanding the overall sentiment landscape of any particular topic or entity by classifying sentiments into positive, negative, or neutral [5].

Sentiment analysis constitutes a considerable basis for decision-making processes in a variety of industries. The combination of NLP and computational tactics in emotion analysis makes possible the classification and extraction of emotion from the pile of text data, whether reviews or social media posts. Therefore, sentiment analysis has become pivotal in developing better decision-making processes and optimizing business outcomes [6].

Recently, deep learning methods for code analysis have been increasingly used in code vulnerability detection or code classification. Normal steps would involve data pre-processing, where the source code is put in a shape to serve the deep learning schemes. Further, this includes steps such as code representation and tokenization, where code gets divided into meaningful components like tokens or sequences of tokens [7].

Tokenization is a very important step in NLP. It helps break down the stream of text into certain manageable pieces or tokens that can later be processed and analyzed by the various schemes developed within the scope of NLP. Sometimes, other tokenization strategies provide more efficient outcomes for one type of task or another. On the other hand, subword tokenization is used with transformer schemes like BERT and GPT that deal efficiently even with rare or unknown words [8].

The transformer model is an extremely powerful DL framework drawing on the attention mechanism for the processing of sequential data, which may be text or time-series information. The transformer model does indeed constitute a unique class of neural network architecture that differs from how conventional recurrent neural networks and LSTMs process the data series. This parallel processing ability makes it efficient for transformers to handle long-range dependencies and capture context much more powerfully [9].[10]

Sentiment analysis for Pfizer's vaccines requires scraping and interpreting data from different channels where public opinions have been expressed. This paper, therefore, focuses on the tweets mentioning the Pfizer and BioNTech vaccines. In this investigation, a DL framework is applied to categorize sentiments expressed by the tweets, whether positive, negative, or neutral, concerning the Pfizer vaccine. The database utilized in this paper has been obtained from Kaggle, titled "Pfizer Vaccines Tweets.". It includes recent tweets mentioning Pfizer vaccines, hence a valid basis for sentiment analysis. Sentiment analysis gives insight into how the public feels towards vaccines by Pfizer. The distribution of sentiments breaks down into positive, negative, and neutral, giving an overview of what the public is saying in a few words. Such information can be depicted using graphs and charts, including word clouds, ROC curves, and precision-recall curves. This can help providers, policymakers, and the general public understand the general sentiment related to Pfizer vaccines.

1.1 Related works

Hasan Dwi Cahyono et al. in 2024 presented Fast Using Naïve Bayes classifiers for real-time fake detection in COVID-19 news on social networks could be quite impactful, given the proliferation of misinformation during the pandemic. The emphasis on the CNB model's efficiency and effectiveness in identifying online misinformation about COVID-19 highlights the significance of leveraging advanced algorithms to combat the spread of false information, particularly during public health crises. This investigation could potentially contribute significantly to the development of more robust tools for combating misinformation online [11].

Janko et al. (2021) utilized machine learning techniques to analyze factors affecting the early spread of COVID-19, focusing on non-countermeasure factors. Their research highlighted the role of socioeconomic factors, geographic location, and public behaviors in the early phases of the pandemic, offering valuable insights into the dynamics of disease spread. By integrating data analytics and machine learning models, they provided an in-depth look into how these factors contribute to pandemic outcomes, which can inform future public health interventions [12].

Bharati Sanjay Ainapure et al. 2023 reported research on emotion analysis of COVID-19 tweets utilizing lexicon-based and DL tactics. This research included outstanding outcomes for vaccination-related tweet classification, with Bi-LSTM and GRU schemes achieving high accuracy values of 92.48% and 93.03%, accordingly. It shows the effectiveness of these DL tactics in handling emotion analysis tasks. Study schemes could indeed provide valuable insights for healthcare workers and policymakers, helping them understand public sentiment and make more informed decisions during future pandemics [13].

Using DL frameworks for sentiment analysis of COVID-19 vaccination reactions collected from Twitter data was presented by Kazi Nabiul Alam et al. in 2023. The study covered tweets collected from December 21 to July 21, capturing a significant timeline of public discourse on COVID-19 vaccines. The analysis included tweets discussing various vaccines available globally. To gauge sentiments, the researchers utilized the Valence Aware Dictionary for Sentiment Reasoner (VADER), an NLP tool. The sentiments were categorized as positive (33.96%), negative (17.55%), and neutral (48.49%) [14].

Saleh Albahli et al. 2023 presented a study utilizing deep learning to evaluate public emotions toward COVID-19 vaccines. Using the Random Multimodal DL (RMDL) classifier, they attained an accuracy rate of 94.81%. Their goal is to increase public knowledge about vaccinations to combat anti-vaccination movements and promote COVID-19 booster doses [15].

Andrew T. Lian et al. in 2022 presented a study that used machine learning to analyze Twitter data and detect COVID-19 vaccine adverse events. It identified tweets discussing personal experiences with COVID-19 vaccinations and extracted information about vaccine doses, types, and symptoms. This is the inaugural study, to our understanding, to detect indications of COVID-19 vaccine side effects from social media, proposing that this method could enhance current vaccine safety monitoring systems [1].

Abdulaziz et al. in 2021 presented an in-depth investigation on sentiment analysis of COVID-19 relevant tweets, focusing on Twitter data. By examining tweets in English during two distinct periods of the pandemic, they identified the most discussed topics and analyzed the sentiments associated with them. Interestingly, conflicting topics emerged during both periods, reflecting the diverse perspectives and concerns of individuals regarding the pandemic. This type of research can provide valuable insights into public opinion and sentiment dynamics during significant events like the COVID-19 pandemic [16].

Etaiwi et al. 2021 explored the use of deep learning techniques for sentiment analysis, focusing on the complexities of social media discourse during the COVID-19 pandemic. They investigated how NLP tools, specifically tailored for sentiment analysis, can help identify public sentiment and enhance health communication strategies. The study examined tweets discussing COVID-19, identifying key topics and emotions associated with the pandemic's progression. This research is important as it helps capture public sentiment, providing actionable insights for policy-makers and health communication strategies during a global crisis [17].

Siru Liu et al. in 2021 shed light on the flexible nature of emotion regarding vaccination efforts. The research looked at classifying tweets into positive, neutral, or negative feelings using the compound scores from SA tools by analyzing English-language tweets from November 1, 2020, to January 31, 2021. The findings indicated that public opinion on COVID-19 vaccines was very intricate, influenced by several reasons such as trial outcomes, administration processes, trust issues in public authorities, effectiveness, and access to information. Positive feeling tweets discussed trial outcomes, administration processes, improvement in life, efficacy, and provision of information. On the negative side, some tweets reflected negative opinions on trial outcomes, conspiracy theories, trust issues, effectiveness, and satisfaction with administrative procedures [18].

G. G. Md. Nawaz Ali et al. presented, in the year 2021, an approach to realizing public opinions on COVID-19 vaccines. They use sentiment analysis in tweets to tap into a real-time source of public opinion. Interestingly, they are not only looking at sentiment but also comparing it to vaccination data from authoritative sources such as the CDC and the Household Pulse Survey. This will be quite useful in pointing out how sentiment can relate to vaccination rates and trends. Overall, if such data are tapped from social media, then this may most probably lead to more updated, integrated insights for policymakers to enable them to formulate more efficient public health strategies, including those that would accelerate vaccination efforts toward the goal of immunity[19][40].

A study by Tanmay Vijay et al., dated 2020, was titled "Sentiment Analysis on COVID-19 Twitter Data"; it focused on tweets about COVID-19 in India, ranging from November 2019 to May 2020. It thus aimed to understand the sentiments of the masses and their shifts over time. This paper provides important insights into how public sentiment evolved at the beginning of the COVID-19 pandemic. This investigation gives a snapshot of the emotional panorama and underlines the great value of sentiment analysis in understanding and tackling public concerns during a crisis [21].

Farah Shahid et al. 2020 presented a study on predicting COVID-19 trends using different DL schemes, namely LSTM, GRU, and Bi-LSTM. The research was focused on the COVID-19 trend forecast for ten major countries and, for comparison, used some traditional schemes like ARIMA and SVR. It serves to underline a finding that deep learning methods have been coming up to deal with intricate situations and non-linear patterns, or in other words, the "sine qua non" of time series data, characterizing pandemics like COVID-19. Further, the study emphasizes the importance of strong forecasts with schemes that can contribute to public health decision-making and resource allocation when considering health crises [2] [41].

In comparison with the state-of-the-art (SOTA) methods, the proposed approach introduces several innovations. While existing solutions predominantly rely on traditional deep learning models like LSTM and CNN, the authors integrate advanced Transformer-based models, such as DistilBERT and Google Electra-base, to better capture the nuances of sentiment in vaccine-related tweets. The proposed method outperforms previous models in terms of accuracy and F1-score, addressing the gaps in existing research, particularly the inability to effectively capture sentiment shifts in rapidly changing public opinions. Furthermore, the approach incorporates multi-modal analysis, allowing for more robust sentiment detection compared to prior works. Table 1 summarizes the key findings from previous research on sentiment analysis of tweets related to COVID-19 vaccines. It includes details on the methods used, datasets, and performance metrics of various deep learning approaches, highlighting the most relevant studies in this area. The comparison demonstrates the range of techniques, including machine learning, deep learning, and lexicon-based methods, and provides an overview of their accuracy and key findings in analyzing public sentiment regarding COVID-19 vaccines.

Table 1: Summary of related works on sentiment analysis

Author(s)	Year	Method Used	Dataset	Key Findings	Accuracy/Performance Metric
A. Lian, J. Du, L. Tang	2022	Machine Learning, NLP	Twitter (Dec 2020–Aug 2021)	Identified vaccine adverse events (VAE) from tweets; common symptoms: sore to touch, fatigue, headache	Not specified
B. S. Ainapure et al.	2023	Deep Learning, Lexicon-based	COVID-19 Tweets	Combined deep learning and lexicon-based approaches for sentiment analysis	Bi-LSTM: 92.48%, GRU: 93.03%
K. N. Alam et al.	2021	Deep Learning (LSTM, Bi-LSTM)	Twitter (Dec 2021–Jul 2021)	Analyzed vaccination responses; sentiment distribution: 33.96% positive, 17.55% negative, 48.49% neutral	LSTM: 90.59%, Bi-LSTM: 90.83%
S. Albahli, M. Nawaz	2023	Deep Learning (CNN, LSTM)	Twitter (COVID-19 Vaccines)	Proposed TSM-CV for sentiment analysis of COVID-19 vaccines	Not specified
M. Abdulaziz et al.	2021	Topic-based Sentiment Analysis	COVID-19 Tweets	Focused on topic modeling and sentiment analysis during and after the first wave of the pandemic	Deep Learning: Up to 81%

2 Materials and method

In this paper, four advanced DL classification algorithms have been developed for sentiment evaluation related to COVID-19 Pfizer vaccination. These include distilBERT, Google-Electra-base, 3 layers Bi-LSTM with 32 units each (model-1), 3 layers Bi-LSTM with 64 units each (model-2), 5 layers Bi-LSTM with 32 units each (model-3), and a hybrid model consisting of 1 CNN layer with kernel size 128 and 1 Bi-LSTM layer with 64 units (model-4). Further, the architecture followed by the implementation of each model will be discussed in detail.

2.1 Database description

This "COVID-19 Vaccines Tweets" database on Kaggle includes the latest tweets related to Pfizer and BioNTech vaccines. Collected via Twitter, the database comprises 11,021 tweets with users' information. This database is aimed at analyzing and performing the NLP tasks of the tweets related to Pfizer and BioNTech vaccines [3].

Some of the common cleaning and preprocessing tasks performed on text data to get them ready for analysis involve URL removal, email removal, unwanted character removal, and finally tokenization of the text [23].

The tokenization process was tailored to the specific architecture of each model employed in this study. For the Transformer-based models (DistilBERT and Electra-base), pre-trained tokenizers provided by the HuggingFace Transformers library were used to ensure compatibility. These tokenizers applied subword tokenization (e.g., WordPiece for BERT variants), handled case normalization, and managed sequence formatting through truncation and padding. Input sequences were truncated to the models' default maximum lengths (e.g., 512 tokens for DistilBERT) and padded with the default [PAD] token where necessary.

In contrast, the Bi-LSTM model relied on the Keras Tokenizer, which converts input text into sequences of word indices based on vocabulary frequency observed in the training data. Stop words were retained to preserve linguistically significant elements that could contribute to sentiment understanding. All sequences were padded to match the maximum sentence length found in the training set, ensuring uniform input dimensions for model training.

2.1.1 Preprocessing

Given the informal and noisy nature of Twitter data, several preprocessing steps were applied using the Natural Language Toolkit (NLTK) to prepare the dataset for analysis. Hashtags, emojis, and special characters were removed to reduce noise and standardize the textual content. Unlike typical text preprocessing pipelines, stemming and lemmatization were deliberately omitted. Preliminary experiments showed that these techniques often disrupted the grammatical structure and altered verb tenses in a way that could distort sentiment interpretation, especially in short, context-dependent tweets. Since the dataset consisted entirely of English-language tweets, no language filtering was required. To maintain consistency with previous studies that utilized the same dataset, no filters were applied based on tweet length, publication date, or tweet quality. Furthermore, an analysis of the sentiment class distribution revealed no significant imbalance among the categories (positive, negative, and neutral); therefore, resampling methods such as SMOTE, oversampling, or class weighting were not deemed necessary.

2.2 DL frameworks

Neural networks (NNs) are utilized in DL, which is essentially an ML technology, to automatically extract and engineer traits. DL frameworks can learn automatically from data by locating pertinent extracted representations using deep nonlinear transformations, in contrast to conventional standard machine learning where domain experts often manually create traits. This capability for autonomous discovery and refinement of traits greatly increases the accuracy and performance of DL schemes.[24], [25].

Various common deep learning methods are vastly used, each for its advantage and purpose, which includes CNNs, RNNs, De-noising Autoencoders, Deep Belief Networks (DBNs), and LSTMs [26].

2.2.1 CNN

The CNN revolutionized image-processing tasks by fundamentally changing how traits are learned and represented from pixel data. Its pooling layers downsample the feature maps to save computation and make the network invariant to slight fluctuations in the input, while its convolutional layers apply filters to input pictures to detect edges, textures, and other characteristics. As the network deepens, these layers capture increasingly abstract features. These features are then passed to the fully connected layers, which interpret the extracted information and perform the final classification task [27].

2.2.2 Recurrent neural networks (RNN)

Because of their ability to remember previous inputs in the sequence, RNNs are excellent for sequence data. This makes them very good for use in tasks where the order of data is important: speech recognition, handwriting recognition, or text analysis. RNNs can learn dependencies and patterns in sequential data and therefore are a valuable instrument for solving many tasks in which sequences occur [28].

2.2.3 DBN

A DBN is a DL tactic consisting of several layers of latent variables, or hidden units, and is used in unsupervised feature extraction. The concept behind it is based on stacking RBMs to build a deep architecture. Neural network problems with deep layers can be resolved by the DBN, where the velocity during learning will be low and overfitting might occur. Thus, initialization in the DBN for every layer is followed, first by unsupervised pre-training and then by supervised fine-tuning [29].

2.2.4 LSTM

RNNs are powerful in handling sequential data because they can retain information from previous steps and leverage this to contextualize the current processing. However, standard RNNs inherently cannot handle long-term dependencies due to the issues of vanishing or exploding gradients during backpropagation, making learning and information conservation over long sequences problematic. This is addressed effectively by LSTMs [2].

The LSTM mechanism comprises three memory gates, specifically, the input gate (i_t), forget (f_t) gate, and output gate (o_t). As a result, the LSTM network effectively handles long-term dependencies in sequential data. The mathematical formulas for LSTM are [30]:

$$i_t = \sigma(W_i x_i + U_i h_{t-1} + b_i) \quad (1)$$

$$f_t = \sigma\left(W_f x_t + U_f h_{t-1} + b_f\right) \quad (2)$$

$$o_t = \sigma(W_o x_t + U_o h_{t-1} + b_o) \quad (3)$$

$$\tilde{c} = f_t \odot c_{t-1} \quad (4)$$

$$h_t = o_t \odot tanhct \quad (5)$$

In the context of an LSTM network, the offered notation describes how different components interact within the unit at time step t. Here's a detailed explanation:

x_t: The input sample at time t.

σ: The sigmoid activation functions.

c_t: The memory unit (cell state) at time t.

b_f, b_i, b_o : The forget gate, input gate, and output gate have different bias terms.

W_f, W_i, W_o: The forget gate, input gate, and output gate weight matrices are as follows.

$\odot$: multiplication of elements.

h_{t-1}: The concealed condition from the preceding time interval (t−1).

c_{t-1}: The condition of each cell at the preceding time step (t−1).

2.2.4.1. Bidirectional LSTM (Bi-LSTM)

Bi-LSTMs are quite successful in enhancing the contextual understanding of sequential data like text. In particular, Bi-LSTMs can better capture the dependencies in sequential data by processing the input sequence in both directions simultaneously. This is particularly useful in tasks where both the past and future contexts are informative, like natural language processing, sentiment analysis, or machine translation. In the context of this study, Bi-LSTM was selected due to its capability to effectively handle the short, informal, and often ambiguous nature of tweets. By considering both preceding and succeeding words in a sentence, the model can better interpret subtle cues in sentiment, such as sarcasm or negation, which are common in social media content. Fig. 1: Architecture of the Bi-LSTM, a composition of two layers - one processing the input from start to end (forward) and the other from the end towards the beginning (backward), both made of LSTM units that capture information in both directions [31].

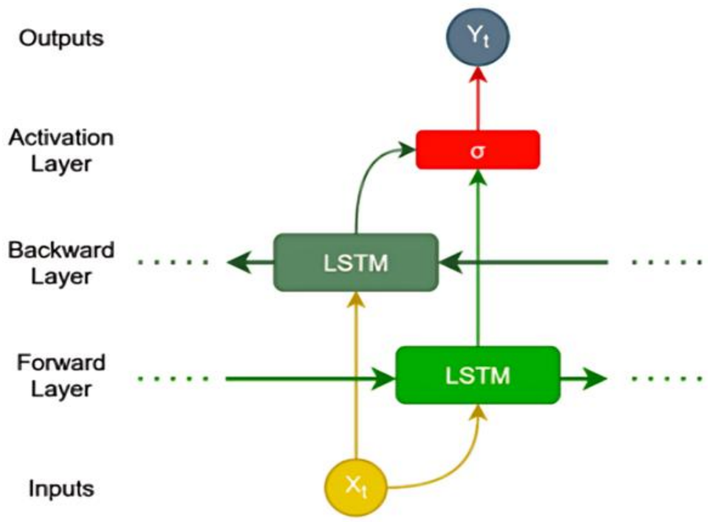

Figure 1: Bidirectional LSTM (Bi-LSTM) architecture

Bidirectional LSTMs are a variation of RNNs where the model takes in an input sequence in two directions: one forward and one backward. As a result, the model may simultaneously gather data from past and future situations. In certain applications, such as those requiring bidirectional contexts—many NLP duties, including emotion analysis or machine translation—this may be quite helpful [32]. This will help the model learn the context in both directions, which can come in handy when performing certain tasks for which understanding context is key. Similar to NLP, it embeds information from both directions and often results in better performance and generalization ability compared to unidirectional LSTMs. The output of the forward layer (h_t^f) and backward layer (h_t^b) of the BiLSTM framework is generated as [33].

$$h_t = \alpha h_t^f + \beta h_t^b \tag{6}$$

$$y_t = \sigma(h_t) \tag{7}$$

Herein, α and σ probably represent weights or factors determining how much each direction-foward and backward accordingly-contributes in the model BiLSTM. The sum of α and σ is equal to 1 because the weights sum up to 1; that is, both directions have been taken into consideration and their contribution is balanced [30].

2.3 DAE

DAE is a variant of the artificial neural network that extends the basic idea of the Autoencoder, enhancing its robustness and offering better feature learning from noise-laden databases [34].

2.4 The transformer framework

In 2017, Vaswani et al. introduced the Transformer model, which transformed NLP and produced innovative language translation outcomes with a significant decrease in training times compared to schemes before the transformer [35].

Transformers are a powerful DL architecture. They have been widely used in recent tasks and provide effective results due to attention mechanism. The Transformer represents a class of topologies for Artificial Neural Networks. In this transformer, an attention mechanism is employed to handle the input of sequential data including text sentences and time-series data. In contrast to conventional RNN and CNN, it captures contextual information and long-range relationships differently. It is intended to operate in conjunction with an encoder-decoder system [7].

Figure 2 illustrates the architecture of the Transformer model, which underpins state-of-the-art language models such as BERT and DistilBERT used in this study. The left block represents the encoder, and the right block shows the decoder. Both components consist of multiple identical layers (denoted by Nx), where each layer includes multi-head attention mechanisms and position-wise feed-forward networks, followed by residual connections and layer normalization. In the encoder, self-attention layers allow the model to weigh the importance of each word relative to others in the input. The decoder includes an additional masked multi-head attention layer that prevents access to future tokens during training, enabling autoregressive generation. Positional encoding is added to the input embeddings to retain sequence order information, as the architecture lacks recurrence. The final output is passed through a linear layer and a softmax function to produce the probability distribution over possible output tokens. This structure allows the Transformer to model complex dependencies in language efficiently and is a foundational component of the pre-trained models used for sentiment classification in this study [9].

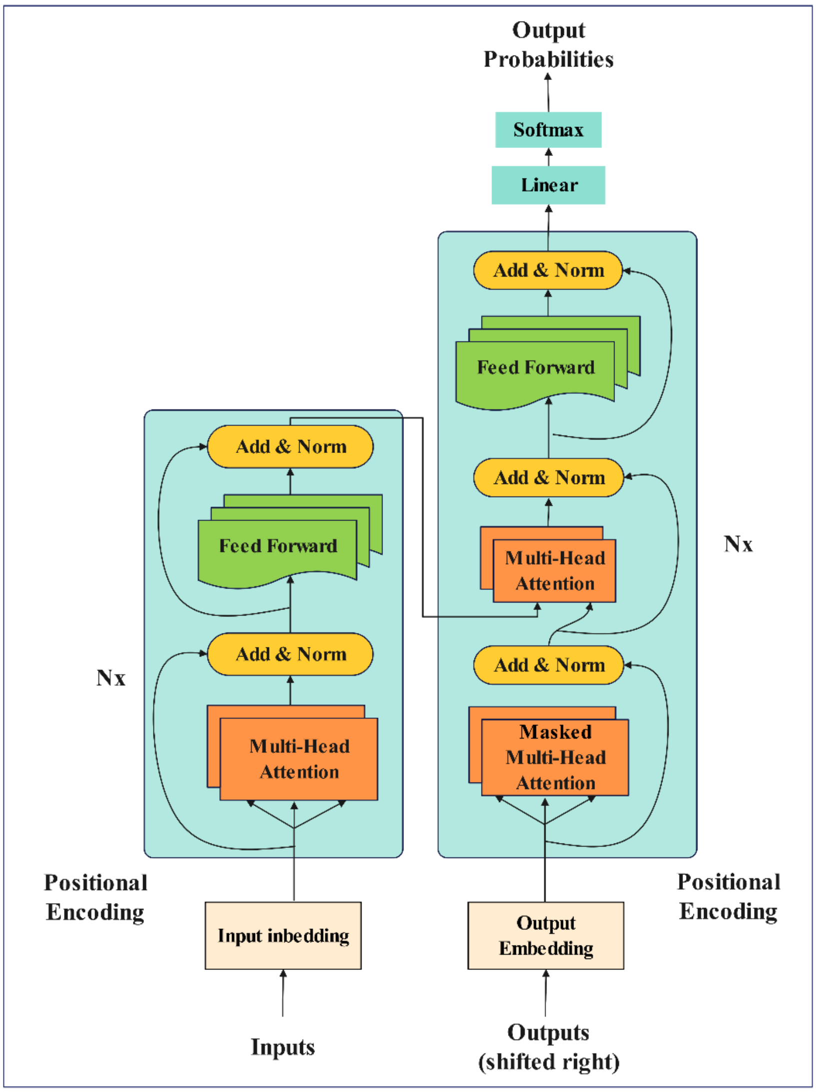

Figure 2: The Transformer model architecture

2.4.1 BERT

BERT was presented by Google in 2018, greatly increasing NLP. By examining the words that come before and after a word in a phrase, the language model BERT can understand its definition. In contrast to previous schemes that sequentially analyze text, BERT's methodology takes into account a word's whole context to maximize its performance for diverse NLP applications, including emotion analysis, linguistic inference, and question answering [8].

2.4.2 DistilBERT

DistilBERT is a compact version of the BERT model that preserves a large portion of BERT's language understanding capabilities while offering increased efficiency. It consists of six transformer layers, compared to BERT's twelve, making it approximately 40% smaller and 60% faster in execution [7] [36].

The model is trained using a method known as knowledge distillation, in which a smaller model, referred to as the *student*, learns to replicate the behavior of a larger, pre-trained *teacher* model. This is achieved by minimizing the difference between the output distributions of the two models. Through this process, the student model acquires much of the performance characteristics of the teacher model, while maintaining a lower computational footprint. In the proposed framework, DistilBERT is utilized for extracting contextual and semantic features from text data efficiently. Its reduced size and computational requirements make it particularly suitable for real-time or resource-constrained NLP applications, without causing significant degradation in performance.

2.4.3 Google Electra-base

Google's ELECTRA is a new pre-training method for NLP tasks employing replaced token detection (RTD) to enhance efficiency. In contrast to BERT, ELECTRA allows the main framework to learn from all input tokens rather than just a subset by substituting certain input tokens with alternatives produced by a tiny masked language model [37].

For the sentiment classification task, three deep learning models were employed: DistilBERT, Google Electra-base, and a Bidirectional Long Short-Term Memory (Bi-LSTM) network. These models were selected due to their proven effectiveness in handling short and informal textual data, such as tweets. All models were trained using the Adam optimizer, which is widely adopted for its efficiency and adaptability in deep learning applications. The learning rate was set to 2e-5 for the Transformer-based models (DistilBERT and Electra-base), while a higher rate of 1e-3 was used for the Bi-LSTM model to facilitate faster convergence. A uniform batch size of 32 was applied across all models to strike a balance between computational efficiency and training stability. Each model was trained for up to five epochs, with early stopping implemented based on validation loss to mitigate the risk of overfitting. Regarding regularization, dropout with a rate of 0.3 was applied in the Bi-LSTM model to prevent overfitting, while Transformer models utilized their internal dropout mechanisms. Additionally, a weight decay of 1e-5 was applied to the Transformer models as an extra layer of regularization.

2.5 Evaluation metrics

This investigation utilized accuracy, precision, recall, F1-score, and a confusion matrix (comprising TP, FP, TN, and FN values) to evaluate the performance of the model across various dimensions[38]**.**

2.5.1 Accuracy

Accuracy is the general metric of performance that is used commonly to tell how well a certain classification model is performing. It provides one complete measure of the performance of the model, the ratio between correct predictions, and the whole count of anticipations. In this case, determining the accuracy will be of paramount importance, but it must be taken into consideration that it might not be appropriate every time since some problems can be with imbalanced classes or different types of errors which may have different consequences [39].

Accuracy is a measure of the overall correctness of the model. It measures the ratio of accurately expected instances (both positive and negative) to the whole count of samples [38].

Here's the formula for accuracy (8):

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{8}$$

2.5.2 Recall

This investigation utilized accuracy, precision, recall, F1-score, and a confusion matrix with values for TP, FP, TN, and FN to evaluate the model's performance using a variety of measures.[40].

$$\text{Recall} = \frac{tp}{tp+tn} \tag{9}$$

2.5.3 Precision

Precision is indeed an essential performance indicator in categorization tasks, particularly in the context of determining specific classes such as attack records [41]. the accuracy metric, which determines the percentage of accurate positive predictions among all the model's positive predictions. The following formula is used to calculate precision (10):

$$\text{Precision} = \frac{TP}{TP + FP} \tag{10}$$

2.5.4 F1score

The F1 score is a helpful metric for considering both precision and recall in a single value. It's especially valuable when you need to strike a balance between these two metrics, as it gives equal weight to both. This is particularly important in tasks where you want to avoid missing too many positive instances (high recall) while ensuring that the ones you label as positive are indeed correct (high precision) [42] [39].

For the F1 score, the following Eq. (11) has been used.

$$\text{F1score} = \frac{2TP}{2.TP + FP + FN} = 2.\frac{Recall.precision}{Recall + precision} \tag{11}$$

3 Result

Sentiment analysis of social media discussions, particularly on platforms like Twitter, provides valuable insights into public opinion and discourse around significant topics such as COVID-19 vaccination. Analyzing the sentiment of tweets can help researchers, policymakers, and healthcare providers understand the public's perception and tailor communication strategies accordingly. In our analysis of the sentiment surrounding the Pfizer COVID-19 vaccine, four deep-learning classification algorithms have been used. These included DistilBERT, Google-Electra-base, three-layer Bi-LSTM with 32 units each (model-1), three-layer Bi-LSTM with 64 units each (model-2), and five-layer Bi-LSTM with 32 units each (model-3). Additionally, a 1D convolutional layer with 128 kernel size and one Bi-LSTM layer with 64 units (model-4) have been used.

DistilBERT: A distilled version of BERT, which retains 97% of BERT's language understanding while being faster and lighter.

Google-Electra-base: A pre-trained transformer model that improves BERT's performance with more efficient training.

Three-layer Bi-LSTM with 32 units each (Model-1): A bidirectional LSTM model with three layers, each consisting of 32 units.

Three-layer Bi-LSTM with 64 units each (Model-2): Similar to Model-1 but with 64 units in each layer.

Five-layer Bi-LSTM with 32 units each (Model-3): A deeper Bi-LSTM model with five layers of 32 units each.

1D Convolutional Layer with 128 Kernel Size and One Bi-LSTM Layer with 64 Units (Model-4): This model combines a convolutional layer to capture local patterns and a Bi-LSTM layer to capture sequential dependencies.

Table 2 presents the architectural details of the deep learning models employed for sentiment classification of tweets related to the Pfizer COVID-19 vaccine. Also Fig. 3 shows Chart of the quantity of tweets in each class, including positive, negative, and neutral.

Table 2: Model architecture descriptions used for sentiment classification

Model Name	Architecture Description
DistilBERT	A lightweight Transformer-based model for efficient sentiment detection.
Google-Electra-base	A Transformer-based model using replaced token detection pretraining.
Model-1	Three-layer Bi-LSTM with 32 units in each layer.
Model-2	Three-layer Bi-LSTM with 64 units in each layer.
Model-3	Five-layer Bi-LSTM with 32 units in each layer.
Model-4	One 1D convolutional layer (kernel size = 128) + one Bi-LSTM layer with 64 units.

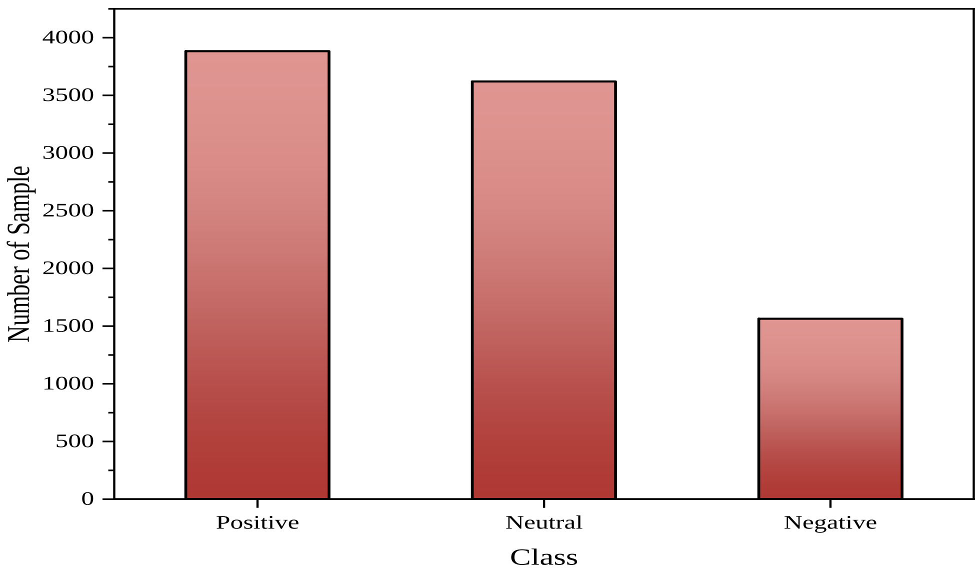

Figure 3: Sentiment analysis of positive, negative, and neutral tweets count

In Fig. 4, the diagram illustrates the distribution of samples in the training and testing data for each class. This demonstrates that the database is randomly separated into training and testing subsets, with 25% allocated for testing, 15% for validation, and 60% for training.

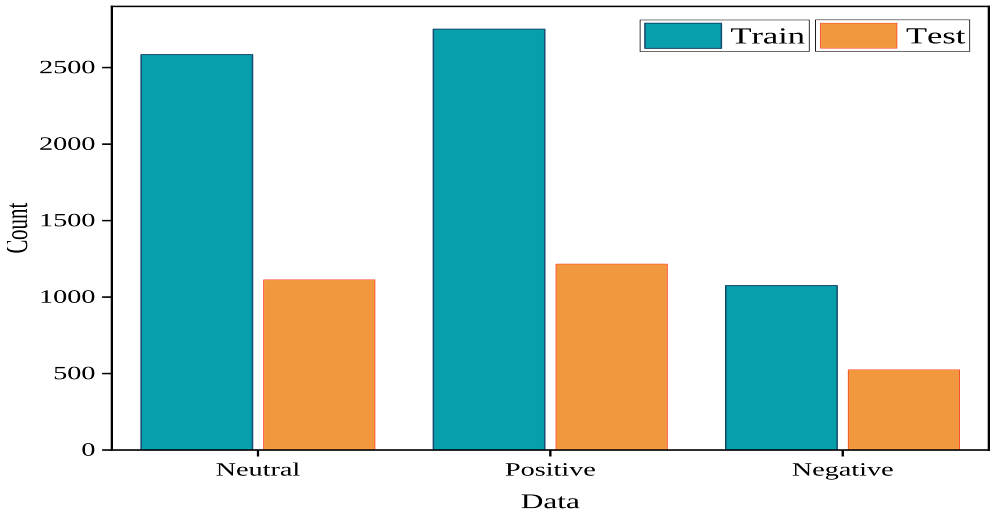

Figure 4: Samples of the training and testing data for each class

Stop words are commonly used words (such as "and", "the", "is", etc.) that are often filtered out of text during the preprocessing phase of natural language processing. These words are regarded as insignificant for tasks like text analysis, as they occur frequently and don't carry much meaning. By removing stop words, the focus can be placed on the more meaningful words in the text.

In Fig. 5, the blue section clearly illustrates the length of the sentences before removing stop words and punctuation, while the orange section distinctly indicates the reduced length after their removal. It is evident that this substantial reduction significantly cuts down computational costs and processing time (about 40%), particularly when dealing with magazine content.

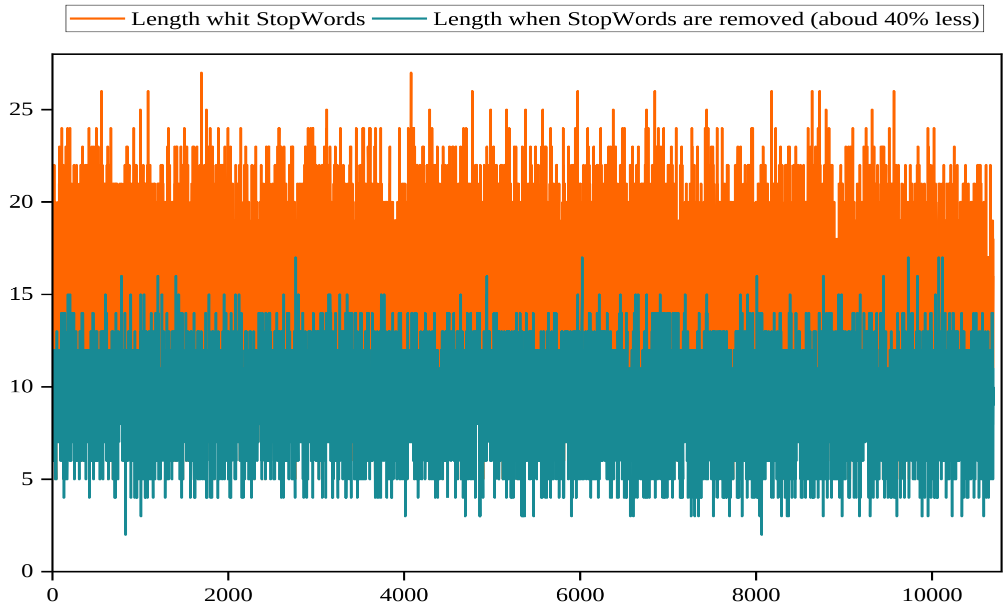

Figure 5: Sentence length with stop words (blue part), sentence length without stops words (orange part)

In this section, Word Clouds, a popular visualization tool that graphically represents the frequency or importance of words within a given text or topic, will be discussed. The size of each word in the Word cloud correlates with how often it appears or its significance in the text. The word clouds are used for sentiment analysis tasks to see tweets classified as good, negative, or neutral. By examining these Word Clouds, insights into the psychological underpinnings of the sentiments expressed in the tweets have been gained. Overall, Word Clouds proved to be a valuable tool for quickly grasping the prevalent emotions and themes within different sentiment categories, providing a visual summary of the psychological foundation behind the tweets [14].

Certain words or phrases were divided into polarity groups, which categorize language based on emotional tone positive, neutral, or negative to reflect the underlying sentiment. This classification helps in identifying the emotional orientation of the content. Using Word Cloud visualization, words with negative sentiment are shown in Figure 6, those with neutral sentiment in Figure 7, and positive sentiment words are displayed in Figure 8.

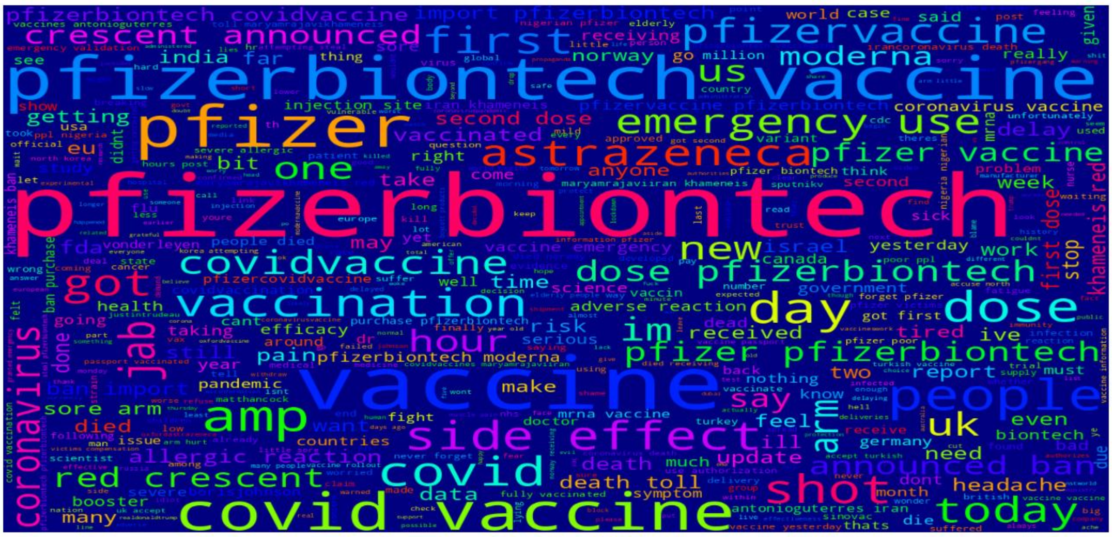

Figure 6: The most frequent words in negative tweets

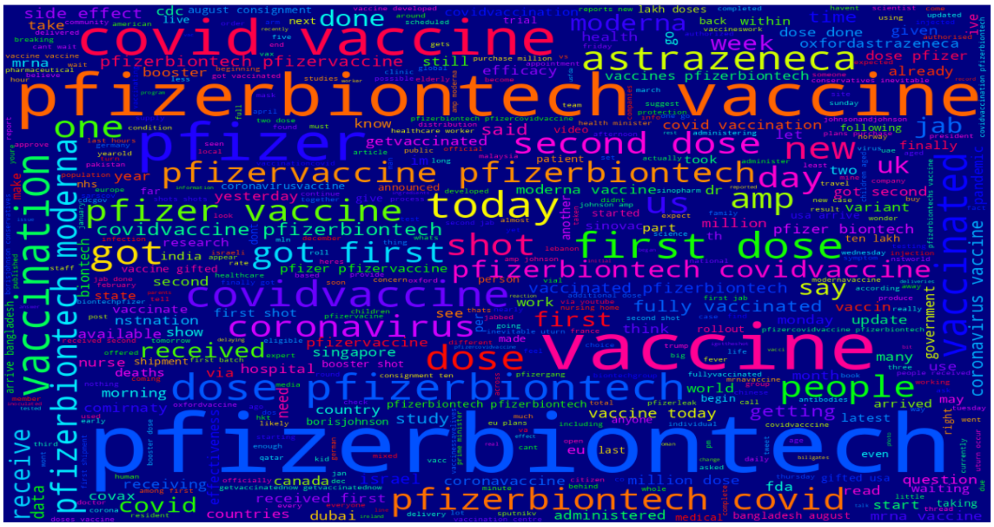

Figure 7: The most frequent words in neutral tweets

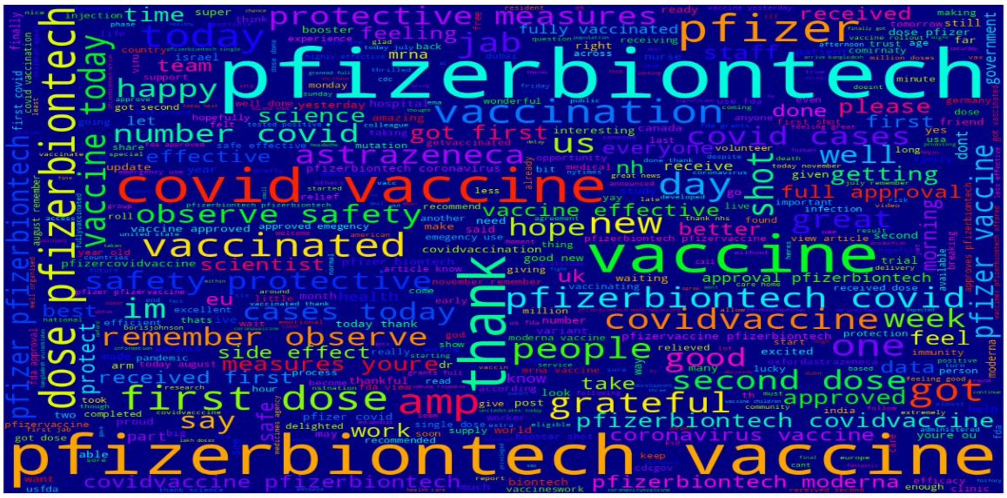

Figure 8: The most frequent words in positive tweets

In this article, several measures of accuracy, namely Accuracy, Precision, Recall, and F1-score have been employed. Fig. 9 presents the outcomes for diverse schemes on the database. It illustrates various schemes, and it's necessary to note that the outcomes have been used from the test data to evaluate an algorithm. The Distilbert and Google Electrabase schemes are transformer schemes, while the other schemes are neural network schemes used for comparison in this paper. According to Fig. 9, the Distilbert model delivered better outcomes than the other schemes presented.

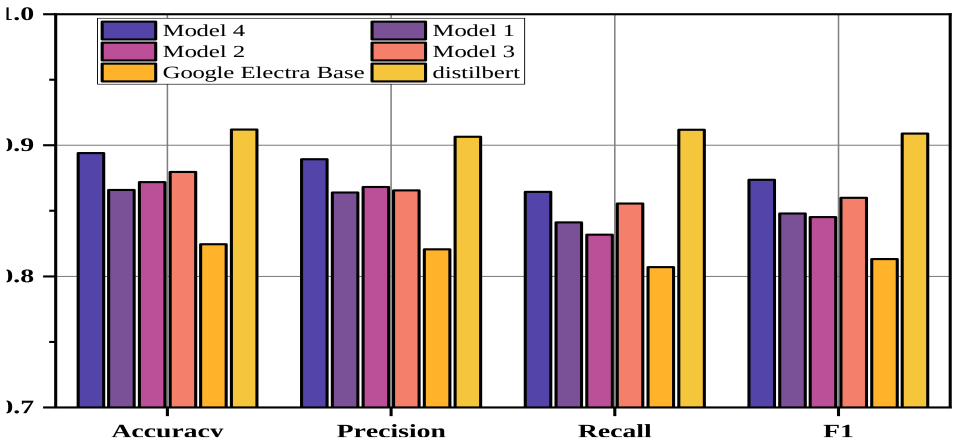

Figure 9: The accuracy of different schemes is based on the dataset

Comparing schemes based on the average time each epoch takes to complete is a common practice in machine learning research. This metric can offer a glimpse into the efficiency of different schemes during training. Fig. 10 depicts that the schemes were trained for a maximum of 70 epochs, after which training was halted to avoid wasting computing resources if performance did not increase over time. This approach is often used to prevent overfitting and to ensure that schemes converge to the best possible solution within a reasonable amount of time.

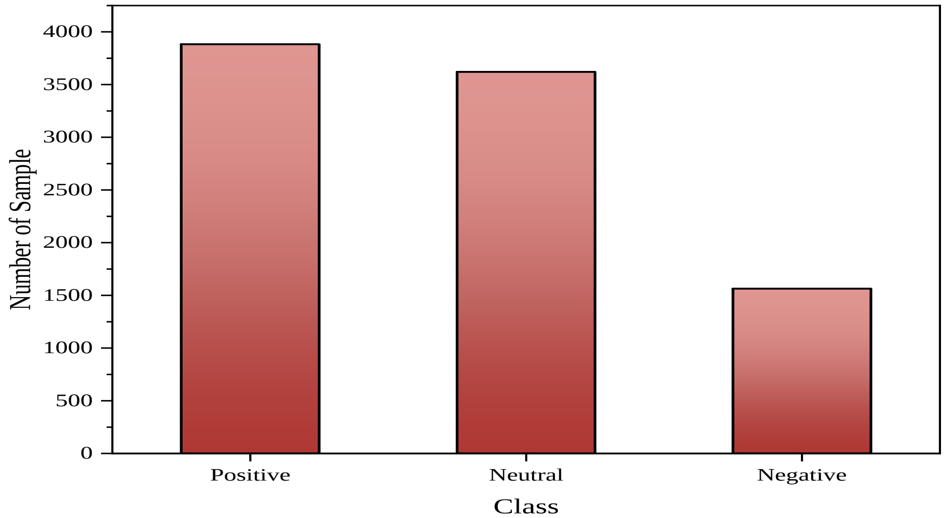

Figure 10: Comparison of schemes in terms of time

To assess the performance of multi-class classification schemes using ROC and AUC to check or visualize the performance. Figure 11 presents the ROC curve for DistilBert, model-1, model-2, model-3, model-4, and Google-Electra-base. In this graph, a larger area under the curve indicates better outcomes from the model. As depicted in Fig. 11, the DistilBert model covers the largest area (0.9859) compared to the other schemes, while the Google-Electra-base model has the smallest area (0.9485).

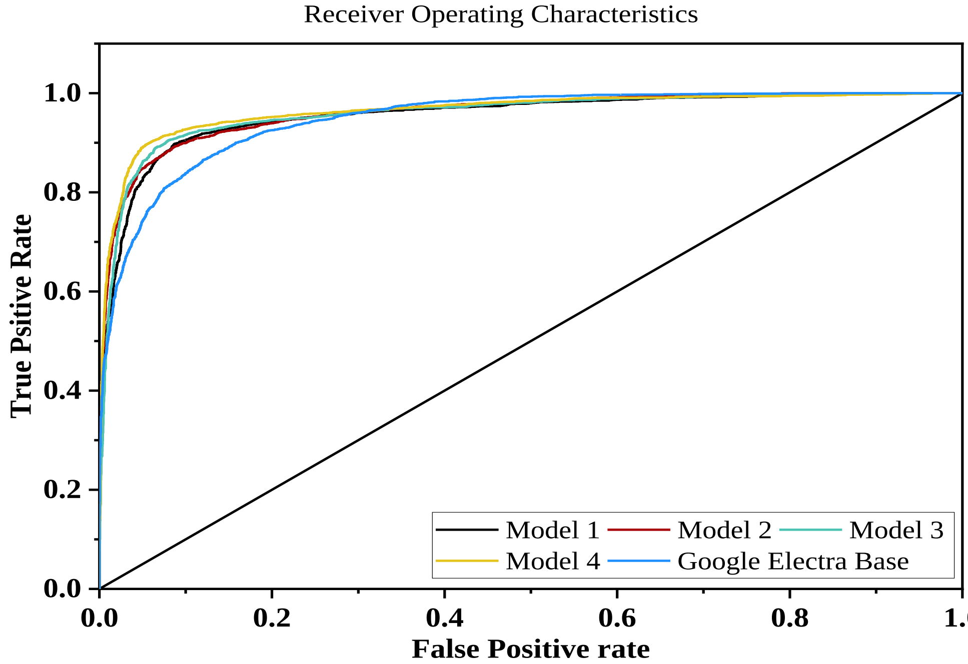

Figure 11: ROC comparison for DistilBert, model-1, model-2, model-3, model-4, and Google-Electra-base

The precision-recall (PR) curve showcases recall on the x-axis and precision on the y-axis over a range of threshold values. As the categorization threshold varies, this graph illustrates the trade-off between accuracy and recall values. This precision-recall curve is shown in Fig. 12. With PR curves, it can be quickly determined which curves represent superior performance for a certain class or model. Figure 12 presents the model's accuracy and recall values. It is clear from Fig. 12 that the DistilBERT model covers a larger area (0.97) than the other schemes, while the Google-Electra-base model has the smallest area (0.91).

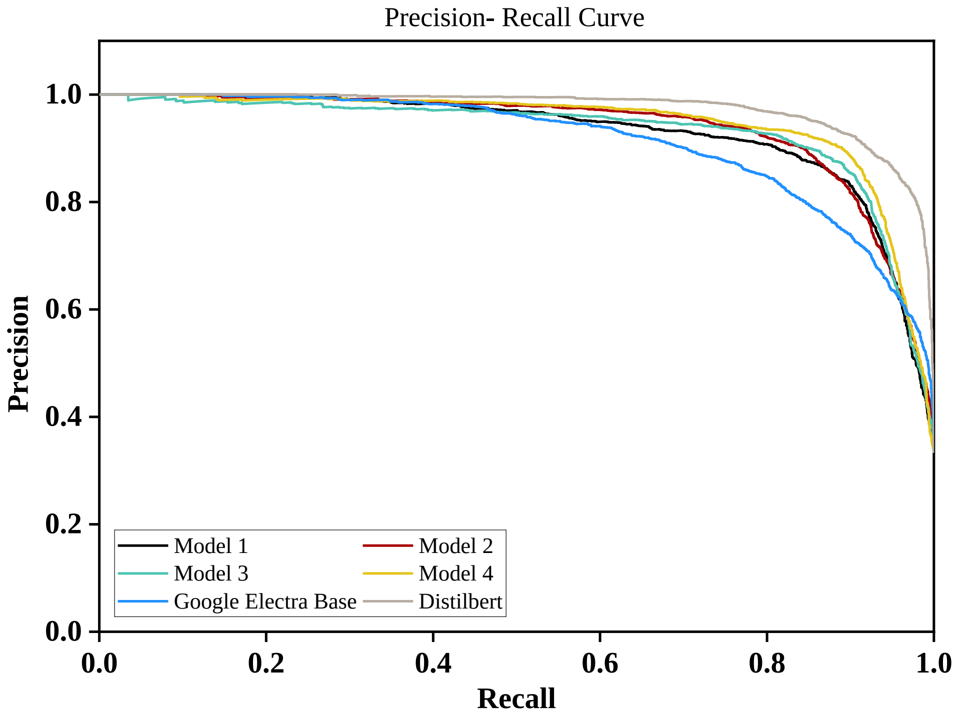

Figure 12: PR comparison for DistilBert, model-1, model-2, model-3, model-4, and Google-Electra-base

According to Fig. 13, most schemes performed effectively in classifying the positive and neutral classes, which had a larger number of samples. However, a notable exception is the DistilBERT model, which demonstrated superior accuracy in classifying the negative class, despite its lower representation in the database. This suggests that DistilBERT's contextual understanding and robust feature extraction capabilities may be well-suited for handling imbalanced class distributions, especially for the negative sentiment class.

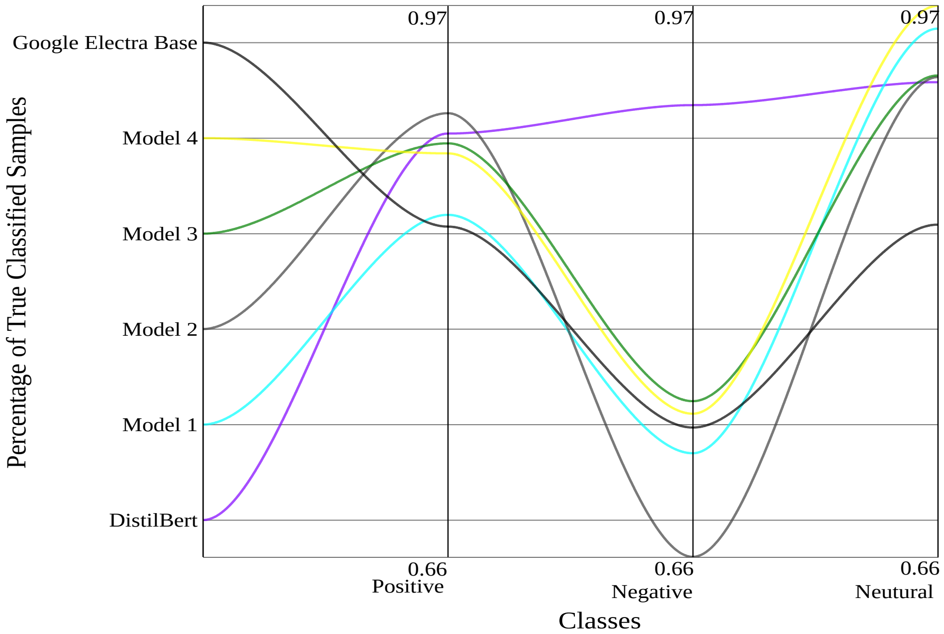

Figure 13: The performance of schemes in different classifications

4 Discussion

This section contrasts the performance of the proposed models with findings from comparable literature. To ensure a robust and consistent evaluation of the sentiment classification models, the dataset comprising 11,021 tweets was randomly split into training (60%), testing (25%), and validation (15%) subsets. The splitting process maintained a balanced distribution of sentiment classes (positive, negative, and neutral) across all sets to avoid class imbalance. Model performance was measured using standard evaluation metrics, including accuracy, precision, recall, F1-score, and ROC-AUC. In addition, a misclassification plot (Fig 13) was generated to visualize class-wise prediction errors and identify potential biases or weaknesses in model performance. All experiments were conducted in a Google Colab Pro environment equipped with an NVIDIA Tesla P100 GPU (16 GB VRAM) and 32 GB of RAM.

The Transformer models, i.e., DistilBERT and Google Electra-base, performed comparatively better than Bi-LSTM and CNN-based models in terms of accuracy and F1-score. Better performance in Transformer models is due to the fact that these models can identify long-range dependencies within textual data and also employ attention mechanisms in order to achieve better contextualization than conventional approaches such as Bi-LSTM. Meanwhile, although Bi-LSTM models perform well, they lack the capability to identify long-range dependencies and context within the dataset and therefore reflect fewer performances in tweet sentiment analysis on vaccines.

The results clearly demonstrate that DistilBERT outperforms the other models across key metrics, including AUC (Area Under the Curve) and ROC (Receiver Operating Characteristics). As shown in Figure 11, DistilBERT (represented by the yellow line) has the highest True Positive Rate (TPR) and the lowest False Positive Rate (FPR) compared to the other models, particularly when analyzing the ROC curve. This highlights DistilBERT's superior ability to correctly classify positive samples while minimizing false positives, making it particularly effective in sentiment analysis tasks with imbalanced datasets. In Figure 12, the Precision-Recall Curve illustrates how DistilBERT maintains its high performance, especially in scenarios with high recall. This superior performance is attributed to DistilBERT's attention mechanisms, which allow it to effectively capture long-range dependencies and contextual nuances in the text, something that other models, like Bi-LSTM or CNN-based approaches, struggle to achieve. Additionally, Figure 13 further strengthens this analysis by showing the time complexity and model performance comparison. DistilBERT consistently delivers superior results but requires higher computational resources, as evidenced by the steep performance curves at the upper right, where it excels in accuracy but also demands more training time. The gap between DistilBERT and models like Google Electra Base and Model 3 is clearly visible in these curves, showing how DistilBERT surpasses them in both training speed and final performance, but with a noticeable computational trade-off.

Regarding computational efficiency, While DistilBERT provides the best performance, its higher time complexity could limit its applicability in real-time systems where efficiency is critical. In contrast, models like Model 1 and Model 2, though slightly less accurate, offer a better balance between performance and computational efficiency, making them more suitable for deployment in environments with constrained resources. Due to the high computational cost of training large-scale Transformer models, cross-validation was not applied. Likewise, statistical significance testing was not included in this version but is planned for future work to further support model comparisons.

One of the key limitations of this research is the prevalence of bias in the dataset, which is predominantly based on tweets by certain populations or geographical locations, and this may result in a biased sentiment distribution. Such bias may potentially influence the overall generalizability of the results because certain opinions or sentiments are disproportionately represented or insufficiently represented.

5 Conclusion

This investigation's main objective is to examine public sentiment regarding Pfizer vaccines by focusing on tweets that mention Pfizer and BioNTech vaccines. The study employs DL frameworks to sort the sentiments expressed in these tweets as positive, negative, or neutral. This exploration tries to create an improved sentiment classification model scheme to test the performance of the new model against methods explored previously.

The material analyzed in this investigation was obtained from Kaggle and represents the "Pfizer Vaccines Tweets" database. The database includes recent tweets about Pfizer vaccines; thus, it was a good basis for sentiment analysis. The research provides useful insights into the public view of Pfizer vaccines through the determination of public perception through sentiment distribution. Such insights are visualized by different graphical tools: word clouds, ROC curves, and precision-recall curves. These visualizations help in viewing the outcomes of sentiment analysis and presenting findings effectively to stakeholders. They also allow for the understanding of the trend of public sentiment and making decisions judiciously based on the analysis.

Data on sentiment about vaccines for healthcare providers, policymakers, and the general public could, therefore, be very instrumental. Furthermore, this insight is bound to make an impact on the way strategies for public communication, vaccine advocacy, and policy-making will be carried out. Research is targeted at classifying tweets into positive, negative, or neutral using ratio analysis. The application of deep learning and transformer methods, including RNN-based LSTM and Bi-LSTM schemes, demonstrates a comprehensive exploration of various deep learning architectures.

Analysis of the ROC curve and its integral indicates that the DistilBERT model outperforms all other schemes. This result is consistent with the findings from the precision-recall curve analysis. Our sentiment analysis provides valuable insights for vaccine producers, governments, health ministries, and organizations like the WHO. Understanding public sentiment can help these entities tailor their communication strategies and address concerns effectively, thereby increasing trust and uptake of vaccines. Analyzing reactions to different vaccines across various countries offers insights into public perceptions and concerns about the vaccination process, aiding health researchers in understanding factors influencing vaccine acceptance and hesitancy. This information can contribute to more effective communication strategies and public health interventions.

Nomenclature

Abbreviation	Explanation	Abbreviation	Explanation
ANN	An Artificial Neural Network	FP	False Positive
ARIMA	An Autoregressive Integrated Moving Average	LSTM	Long Short-Term Memory
AUC	Area Under the Curve	ML	Machine Learning
BERT	Bidirectional Encoder Representations from Transformers	NLP	Natural Language Processing
		PR	precision-recall
BI-LSTM	A Bidirectional LSTM	RNN	A Recurrent Neural Network
CDC	Centers for Disease Control and Prevention	ROC	Receiver Operating Characteristic
CNN	Convolutional neural network	RTD	Replaced Token Detection
DAE	Denoising Autoencoder	SVR	Support Vector Regression
DBN	Deep Belief Network	TN	True Negative
DL	Deep Learning	TP	True Positive
Electra	Efficiently Learning an Encoder that Classifies Token Replacements Accurately	WHO	World Health Organization
FN	False Negative		

Authors' contributions

All scholars contributed to the study's conception and design. Data collection, simulation, and analysis were performed by" Zilong XIA and Fanchao Niu". Also, the first draft of the manuscript was written by Zilong XIA. Ling Ding commented on previous versions of the manuscript.

Data availability

All datasets used during the current study are publicly available at the repository link included in the manuscript.

Conflicts of interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

Author statement

The manuscript has been read and approved by all the authors, the requirements for authorship, as stated earlier in this document, have been met, and each author believes that the manuscript represents honest work.

Ethical approval

All authors have been personally and actively involved in substantial work leading to the paper, and will take public responsibility for its content.

References

[1] A. Lian, J. Du, and L. Tang, “Using a Machine Learning Approach to Monitor COVID-19 Vaccine Adverse Events (VAE) from Twitter Data,” *Vaccines (Basel)*, vol. 10, no. 1, 2022, doi: 10.3390/vaccines10010103.

[2] F. Shahid, A. Zameer, and M. Muneeb, “Predictions for COVID-19 with deep learning models of LSTM, GRU and Bi-LSTM,” *Chaos Solitons Fractals*, vol. 140, 2020, doi: 10.1016/j.chaos.2020.110212.

[3] M. Shah, A. Vasant, K. Gandhi, and S. Panesar, “Sentiment Analysis of Twitter Tweets for COVID- 19 Pfizer Vaccines,” in *Proceedings of the 7th International Conference on Intelligent Computing and Control Systems, ICICCS 2023*, 2023, pp. 929–937. doi: 10.1109/ICICCS56967.2023.10142744.

[4] I. M. Dighriri *et al.*, “Pfizer-BioNTech COVID-19 Vaccine (BNT162b2) Side Effects: A Systematic Review,” *Cureus*, 2022, doi: 10.7759/cureus.23526.

[5] M. Taboada, “Sentiment Analysis: An Overview from Linguistics,” 2016. doi: 10.1146/annurev-linguistics-011415-040518.

[6] D. M. E. D. M. Hussein, “A survey on sentiment analysis challenges,” *Journal of King Saud University - Engineering Sciences*, vol. 30, no. 4, pp. 330–338, 2018, doi: 10.1016/j.jksues.2016.04.002.

[7] A. Bahaa, A. E. R. Kamal, H. Fahmy, and A. S. Ghoneim, “DB-CBIL:A DistilBert-Based Transformer Hybrid Model using CNN and BiLSTM for Software Vulnerability Detection,”

IEEE Access, 2024, doi: 10.1109/ACCESS.2024.3396410.

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

[9] Vasvani A. *et al.*, "Attention is all you need," in *Advances in Neural Information Processing Systems*, 2017, pp. 5998–6008.

[10] J. Xu, Z. Meng, and L. Xu, "Integrated system of health management-oriented reliability prediction for a spacecraft software system with an adaptive genetic algorithm support vector machine," *Eksploatacja i Niezawodność*, vol. 16, no. 4, pp. 571–578, 2014.

[11] H. D. Cahyono, A. Mahadewa, A. Wijayanto, D. W. Wardani, and H. Setiadi, "Fast Naïve Bayes classifiers for COVID-19 news in social networks," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 34, no. 2, p. 1033, 2024, doi: 10.11591/ijeecs.v34.i2.pp1033-1041.

[12] V. Janko *et al.*, "Machine learning for analyzing non-countermeasure factors affecting early spread of covid-19," *Int J Environ Res Public Health*, vol. 18, no. 13, Jul. 2021, doi: 10.3390/ijerph18136750.

[13] B. S. Ainapure *et al.*, "Sentiment Analysis of COVID-19 Tweets Using Deep Learning and Lexicon-Based Approaches," *Sustainability (Switzerland)*, vol. 15, no. 3, pp. 523–535, 2023, doi: 10.3390/su15032573.

[14] K. N. Alam *et al.*, "Deep Learning-Based Sentiment Analysis of COVID-19 Vaccination Responses from Twitter Data," *Comput Math Methods Med*, vol. 2021, 2021, doi: 10.1155/2021/4321131.

[15] S. Albahli and M. Nawaz, "TSM-CV: Twitter Sentiment Analysis for COVID-19 Vaccines Using Deep Learning," 2023. doi: 10.3390/electronics12153372.

[16] M. Abdulaziz, A. Alotaibi, M. Alsolamy, and A. Alabbas, "Topic based Sentiment Analysis for COVID-19 Tweets," *International Journal of Advanced Computer Science and Applications*, vol. 12, no. 1, pp. 626–636, 2021, doi: 10.14569/IJACSA.2021.0120172.

[17] W. Etaiwi, D. Suleiman, and A. Awajan, "Deep Learning Based Techniques for Sentiment Analysis: A Survey," 2021, *Slovene Society Informatika*. doi: 10.31449/inf.v45i7.3674.

[18] S. Liu and J. Liu, "Public attitudes toward COVID-19 vaccines on English-language Twitter: A sentiment analysis," *Vaccine*, vol. 39, no. 39, pp. 5499–5505, 2021, doi: 10.1016/j.vaccine.2021.08.058.

[19] G. G. M. Nawaz Ali *et al.*, "Public perceptions of covid-19 vaccines: Policy implications from us spatiotemporal sentiment analytics," *Healthcare (Switzerland)*, vol. 9, no. 9, 2021, doi: 10.3390/healthcare9091110.

[20] Y. Fang, Y. Wang, J. Sha, T. Gu, and H. Zhang, "A mechanism reliability analysis method considering environmental influence and failure modes' correlation: a case study of rifle automaton," *Eksploatacja i Niezawodność*, vol. 25, no. 2, 2023.

[21] T. Vijay, A. Chawla, B. Dhanka, and P. Karmakar, "Sentiment Analysis on COVID-19 Twitter Data," in *2020 5th IEEE International Conference on Recent Advances and Innovations in Engineering, ICRAIE 2020 - Proceeding*, 2020. doi: 10.1109/ICRAIE51050.2020.9358301.

[22] E. ÖzcAn, T. DAnışAn, R. YumuşAk, and T. EREn, "An artificial neural network model supported with multi criteria decision making approaches for maintenance planning in hydroelectric power plants," *Eksploatacja i Niezawodność*, vol. 22, no. 3, 2020.

[23] "Retracted: Deep Learning-Based Sentiment Analysis of COVID-19 Vaccination Responses from Twitter Data," *Comput Math Methods Med*, vol. 2023, pp. 1–1, 2023, doi: 10.1155/2023/9831021.

[24] S. Wankhade and V. S., "A novel hybrid deep learning method for early detection of lung cancer using neural networks," *Healthcare Analytics*, vol. 3, 2023, doi: 10.1016/j.health.2023.100195.

[25] KADERS. M. R. I. PROTIMA KHANMD. FAZLUL, "Machine Learning and Deep Learning Approaches for Brain Disease Diagnosis: Principles and Recent Advances," *IEEE Access*. doi: 10.1109/ACCESS.2021.3062484.

[26] J. Lu, L. Tan, and H. Jiang, "Review on convolutional neural network (CNN) applied to plant leaf disease classification," 2021, *MDPI*. doi: 10.3390/agriculture11080707.

[27] Md. A. Hossain and Md. S. Alam Sajib, "Classification of Image using Convolutional Neural Network (CNN)," *Global Journal of Computer Science and Technology*, pp. 13–18, 2019, doi: 10.34257/gjcstdvol19is2pg13.

[28] S. Min, B. Lee, and S. Yoon, "Deep learning in bioinformatics," 2017. doi: 10.1093/bib/bbw068.

[29] Y. Hua, J. Guo, and H. Zhao, "Deep Belief Networks and deep learning," in *Proceedings of 2015 International Conference on Intelligent Computing and Internet of Things, ICIT 2015*, 2015, pp. 1–4. doi: 10.1109/ICAIOT.2015.7111524.

[30] M. Massaoudi, S. S. Refaat, H. Abu-Rub, I. Chihi, and F. S. Oueslati, "PLS-CNN-BiLSTM: An end-to-end algorithm-based savitzky-golay smoothing and evolution strategy for load forecasting," *Energies (Basel)*, vol. 13, no. 20, 2020, doi: 10.3390/en13205464.

[31] H. C. de C. Filho *et al.*, "Rice crop detection using LSTM, Bi-LSTM, and machine learning models from Sentinel-1 time series," *Remote Sens (Basel)*, vol. 12, no. 16, 2020, doi: 10.3390/RS12162655.

[32] Y. Imrana, Y. Xiang, L. Ali, and Z. Abdul-Rauf, "A bidirectional LSTM deep learning approach

for intrusion detection," 2021. doi: 10.1016/j.eswa.2021.115524.
[33] Z. xian Liu, D. gan Zhang, G. zhao Luo, M. Lian, and B. Liu, "A new method of emotional analysis based on CNN–BiLSTM hybrid neural network," *Cluster Comput*, vol. 23, no. 4, pp. 2901–2913, 2020, doi: 10.1007/s10586-020-03055-9.
[34] M. M. A. Rahhal, Y. Bazi, H. Alhichri, N. Alajlan, F. Melgani, and R. R. Yager, "Deep learning approach for active classification of electrocardiogram signals," *Inf Sci (N Y)*, vol. 345, pp. 340–354, 2016, doi: 10.1016/j.ins.2016.01.082.
[35] A. Chandra, L. Tünnermann, T. Löfstedt, and R. Gratz, "Transformer-based deep learning for predicting protein properties in the life sciences," 2023. doi: 10.7554/eLife.82819.
[36] A. R. Naira, Rimjhim Padam Singha, and P. K. , Deepa Guptaa, "Evaluating the Impact of Text Data Augmentation on Text Classification Tasks using DistilBERT," 2024, *Elsevier*. doi: 10.1016/j.procs.2024.04.013.
[37] H. M. Roisman, "ELECTRA: PRE-TRAINING TEXT ENCODERS AS DISCRIMINATORS RATHER THAN GENERATORS Kevin," 2016. doi: 10.1002/9781119257530.ch12.
[38] Ž. Vujović, "Classification Model Evaluation Metrics," *International Journal of Advanced Computer Science and Applications*, vol. 12, no. 6, pp. 599–606, 2021, doi: 10.14569/IJACSA.2021.0120670.
[39] H. Dalianis, "Evaluation Metrics and Evaluation," 2018. doi: 10.1007/978-3-319-78503-5_6.
[40] M. N. Hossin, M. and Sulaiman, "A REVIEW ON EVALUATION METRICS FOR DATA CLASSIFICATION EVALUATIONS," 2015, *International Journal of Data Mining & Knowledge Management Process (IJDKP)*. doi: DOI : 10.5121/ijdkp.2015.5201.
[41] M. S. Alsahli, M. M. Almasri, M. Al-Akhras, A. I. Al-Issa, and M. Alawairdhi, "Evaluation of Machine Learning Algorithms for Intrusion Detection System in WSN," *International Journal of Advanced Computer Science and Applications*, vol. 12, no. 5, pp. 617–626, 2021, doi: 10.14569/IJACSA.2021.0120574.
[42] D. Chicco and G. Jurman, "The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation," 2020. doi: 10.1186/s12864-019-6413-7.