International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

A Deep Learning Framework with a Hybrid Model for Automatic Depression Detection in Social Media Posts

[1]T. Srajan Kumar, [2]Dr. M. Narayanan, [3]Dr. Harikrishna Kamatham

Submitted:14/03/2024 **Revised**: 29/04/2024 **Accepted**: 06/05/2024

Abstract: In the contemporary era, there have been increasing incidents of mental health issues due to various reasons, including lifestyle changes. Social media's widespread use has allowed people to openly express their emotions, giving researchers access to much data. In this context, it is possible to mine social media conversations and detect the probability of depression based on the expressions in the underlying text. There are many heuristic approaches for depression detection. With the development of artificial intelligence (AI), learning-based techniques may fare better in identifying depression. However, since there isn't a single solution that works for everyone, deep learning models must be enhanced to perform better in the diagnosis of depression. This study introduced a hybrid deep learning framework that seamlessly integrates bi-directional long short-term memory (biLSTM) and convolutional neural network (CNN) models to extract features and temporal connections from data. Social media posts may be automatically analyzed using the deep learning framework developed to identify depression. Our method, called Learning Based Depression Detection (LBDD), is designed to classify tweets based on their likelihood of containing depression. It works by taking in input from Twitter users. We evaluated our methodology with a benchmark dataset and found that the proposed deep learning model could outperform many existing models with the highest accuracy of 96.32%. Therefore, our deep learning model can be integrated with a clinical decision support system for assessing mental well-being.

Keywords: *Depression Detection, Machine Learning, Deep Learning, Artificial Intelligence, Natural Language Processing*

1. INTRODUCTION

Depression is a serious mental ailment that makes up a significant fraction of all diseases in countries around the globe. More than 350 million individuals worldwide—roughly 4.4% of the total population—struggle from depression. Furthermore, two-thirds of patients don't ask for assistance. The main issue is that sadness inadvertently interferes with a person's social and personal life. When it goes too far, it might contribute to other issues, including mental illnesses, suicide, etc. Roughly one person perishes in their 40s, translating to 8,00,000 suicide deaths annually worldwide.

Teenagers are thought to be particularly vulnerable to depression since suicides account for a significant portion of youth fatalities. Research on depression and how to identify it in a person is essential in the context of India, where suicide rates are astonishing. A 2012 Lancet study states that in India, a student kills himself from despair once per hour. World Health Organization (WHO) data states that in 2015, about 8934 students took their own lives. Over the previous five years, some 39,775 students killed themselves, the vast majority of which went undetected. This figure highlights the need for immediate attention to the problem and implementation of the necessary measures.

Every stage of life, including infancy, adolescence, and adulthood, requires treatment for mental health issues and related problems. Individuals who experience depression typically experience either a transient or persistent low mood that stifles their creativity and passion for daily tasks [1]. Routine tensions and extended periods of low mood might become chronic or recurring, which can result in a major health issue [2]. Depression sufferers often endure a variety of symptoms, including insomnia, loneliness, loss of food and sleep, difficulties concentrating at work or on personal concerns, and possibly an increased risk of suicide [3]. A person may have long-term depression due to several causes, including a difficult upbringing, sexual abuse, alcoholism, physical conditions, strain at work, and the historical legacies of racism, colonialism, and caste [4]. Without therapy, anxiety and despair can worsen over time and can lead to heart difficulties, memory problems, lack of sleep, and other disorders. Under the direction of several nations and renowned organizations like the World Health Organization, some incentives and programs have been launched to treat depression. Most depression sufferers are from middle-class or lower-class backgrounds. Thus, they cannot obtain these therapy [5], [6]. Because of a

[1]*Research Scholar, Department of CSE, School of Engineering, Malla Reddy University, Hyderabad, India.*
Email-ID: 2232CS010020@mallareddyuniversity.ac.in
[2]*Professor, Department of CSE, School of Engineering, Malla Reddy University, Hyderabad, India.*
Email-ID: narayananmecse@gmail.com
[3]*Associate Dean, Department of CSE, School of Engineering, Malla Reddy University, Hyderabad, India.*
Email-ID: kamathamhk@gmail.com

scarcity of finances and resources, poor nations do not have an efficient depression treatment program.

Since there is no useful technique for doing so, it is quite difficult to discriminate between depressed and non-depressed individuals. Furthermore, the availability of trained medical professionals and adequate resources to treat depression is lacking. Due to a dearth of reliable procedures for diagnosing depression, the majority of prediction techniques now in use are inaccurate. Due to the vast user base and activity on these social media behemoths, we can forecast sadness using Twitter, Facebook, Instagram, Snapchat, and other platforms. A public database of social media data is maintained by Twitter and related services, which generate social media data at a rate of 6000 tweets per second, or 200 billion tweets each year [7]. The literature showed that deep learning models could perform well in depression detection from text data. This paper proposes a biLSTM-based deep learning framework to capture temporal relationships. The following are the things we contributed to this paper.

1. Our proposal is a hybrid model-based deep learning framework that effectively integrates long short-term memory (LSTM) and Convolutional Neural Network (CNN) models to extract features and temporal correlations from data.
2. We introduced the Learning Based Depression Detection (LBDD) algorithm, which categorizes tweets based on their likelihood of containing depression.
3. We developed an application to assess and train the suggested hybrid deep learning model.

This is the structure for the rest of the paper. In Section 2, previous research on depression identification is reviewed. Section 3 presents an algorithm and deep learning-based framework for automatically identifying sadness in social media messages. Our experimental results are shown in Section 4. The hybrid deep learning model's performance is compared to that of the state-of-the-art model. We wrap up our work in Section 5 and offer directions for further investigation.

2. RELATED WORK

This section reviews earlier research on the automated identification of sadness using social media chats. Ghosh *et al*. [1] noted that COVID-19-exacerbated depression has spurred efforts to identify and cure depression using data from social media, leading to significant improvements in detection techniques. Malviya *et al*. [2] emphasized that advancements in technology and the broad expression enabled by social media play a crucial role in better depression identification, especially during pandemics, through deep learning. Lin *et al*. [3] effectively diagnosed depression by utilizing Twitter data and multimodal learning in the SenseMood system, enhancing care. Giuntini *et al*. [4] included text analysis, Facebook, and Twitter in their approach. They noted that temporal analysis presents challenges, but sentiment analysis on social networks can identify depression-related mood disorders. Renjith *et al*. [5] emphasized the significance of identifying depression on social media and how deep learning models and natural language processing may benefit from this study area.

Yang *et al*. [6] stressed the necessity of self-reporting for the diagnosis of depression and mentioned that new depression detection frameworks can benefit from the data provided by social networks. They also recognized the essential nature of identifying sadness and stress on social media and the role of KC-Net in leveraging mental state information for enhanced detection. Ghosh *et al*. [8] compared an attention-based BiLSTM-CNN model to previous models and found it performs better in detecting depressed Bangla texts. Rissola *et al*. [9] emphasized language analysis methods aid in early screening by identifying indicators of diseases such as depression. Ziwei *et al*. [10] analyzed social media and highlighted that limited access to therapy has prompted the analysis. They also mentioned that signs of depression include sadness, disinterest, and bodily illnesses.

Tariq *et al*. [11] mentioned that mental health prediction is an area where social media analysis aids decision-making. They suggested potential in co-training with RF, NB, and SVM classifiers. Peng *et al*. [12] opined that Textual Emotion Analysis (TEA) utilizes Deep Learning (DL) techniques to extract emotions from text and aids in sentiment analysis and development. Skaik and Inkpen [13] analyzed social media data using ML and NLP, suggesting that mental health concerns may be identified, with the main challenges lying in making the right feature selection and sampling. Adikari *et al*. [14] suggested an AI framework that uses sophisticated NLP methods and algorithms to analyze emotions in social media, while Chatterjee *et al*. [15] proposed SS-BED. This Deep Learning method outperforms conventional models by combining sentiment and semantic information for effective emotion recognition.

In their study, Yang *et al*. [16] analyzed Chinese microblog entries to investigate depression, an important issue. They used a neural model to address two tasks, outperforming the baseline: predicting the severity and cause of depression. Yao *et al*. [17] looked into an Online Depression Community (ODC) on Sina Weib, utilizing a coding method to classify depression symptoms. They noted that a key indicator of depression is suicidality, which often occurs alongside other symptoms such as anxiety and insomnia. Cao *et al*. [18] discovered that an attention mechanism and a personal suicide-oriented knowledge graph may be used to identify talks on social media reliably. Scepanovic *et al*. [19] found that a deep learning approach consistently extracted medical information from social media across various datasets. Pran *et al*. [20] studied the sentiments of Bangladeshis regarding COVID-19, achieving high accuracy using CNN, with a majority of the sentiments being analytical.

Using a deep learning approach, Blanco *et al*. [21] analyzed optimism and pessimism in COVID-19 Twitter discussions. They evaluated emotional changes and model performance, highlighting social influence. Subramani *et al*. [22] introduced a deep learning method to detect signs of domestic abuse on social media to aid crisis assistance organizations. Farruque *et al*. [23] proposed an algorithm to identify clinical depression in temporal social media posts at the user level. Kumar *et al*. [24] examined the use of social media, particularly Arabic, to detect sadness. They demonstrated cutting-edge performance using a BERT-Bi-LSTM pipeline. Ahmed and Lin [25] explored the use of phrase analysis to identify depression via social media. They suggested text categorization using Graph Attention Networks (GATs) and achieved a 0.91 ROC score using Reddit Depression data.

Reference	Approach	Technique	Algorithm	Data set	Limitation
[2]	ML and DL	NLP, SVM, NB, and Boosting	ML-based algorithm	Reddit	In the future, they intend to exploit more diversified datasets for depression detection.
[5]	ML and DL	NLP, CNN, LSTM	DL and ML algorithm	Reddit	Different attention mechanisms are to be explored in the future.
[8]	ML	CNN and LSTM	ML-based algorithm	Custom dataset	Experiments are to be made in the future with diversified datasets.
[11]	ML	SVM, RF, and NB	C-training and lexicon-based algorithm	Reddit	In the future, user posts in other domains will be considered for completeness.
[13]	DL	Unsupervised learning	GSOM algorithm	Reddit and RSDD	Ambiguous emotional expressions need to be modeled in the future for better performance.
[17]	DL	Text classifier and Annotation scheme	DL algorithm	Custom dataset	More depression symptoms and lexicons are to be explored in the future.
[22]	DL	CNN, RNN, and LSTM	DL algorithms	DV Dataset	In the future, hybrid DL approaches are to be explored.
[24]	DL and AI	LSTM and CNN	DL algorithm	Arabic datasets	Conversation sequences and temporal domain need to be considered in the future. he people.
[25]	DL	DL-based techniques	Cryptographic algorithms	Custom dataset	Text classification at the character level is yet to be considered.
[28]	ML	Text mining and data mining	EMUS algorithm	Custom dataset	Community behavior and impact are yet to be explored.

Table 1: Summary of literature findings

Yuki *et al*. [26] found that analyzing conversations can help in identifying depression at an early stage, as it is often undetected until physical symptoms appear. They achieved the highest accuracy using LSTM. Rissola *et al*. [27] provided information on mental health and highlighted that the lack of datasets hinders innovation in this field. They also presented a dataset and collection technique for a study on depression detection. Giuntini *et al*. [28] developed a technique to assess depressed people's emotional behavior and participation on social media. They integrated network analysis with emotional feature extraction from text data to determine mood levels and interaction patterns. Mendu *et al*. [29] discovered that researchers aim to link private message aspects to mental health through a hierarchical structure, providing insights into individual habits. Govindasamy and Palanichamy [30] investigated and found that depression is a major, often overlooked problem and suggested that machine learning algorithms could potentially signal sadness through social media posts. The literature review revealed that deep learning models perform well in detecting depression from text data. The summary of literature findings is presented in Table 1. This paper proposes a biLSTM-based deep learning framework to capture temporal relationships.

3. PROPOSED METHODOLOGY

This section outlines the suggested approach, which includes the assessment technique, data set information for the suggested algorithm, and deep learning architecture.

3.1 Problem Definition

Given social media posts, the difficult challenge is creating a deep learning framework for automatically detecting the likelihood of depression based on the content in social media posts.

3.2 Our Framework

Our approach for automatically identifying sadness from social media posts is based on deep learning. Figure 1 shows the system model. Data is collected from Twitter tweets and subjected to preprocessing to improve data quality. The system performs feature extraction and generates word embeddings. Subsequently, the suggested hybrid deep learning model is trained to identify depression automatically. The trained model categorizes test data samples presented to it. Different modules make up the suggested system model. Internet user-generated data is extracted. Preparing raw data for cleaning. Feature extraction to deliver data that is machine-readable.

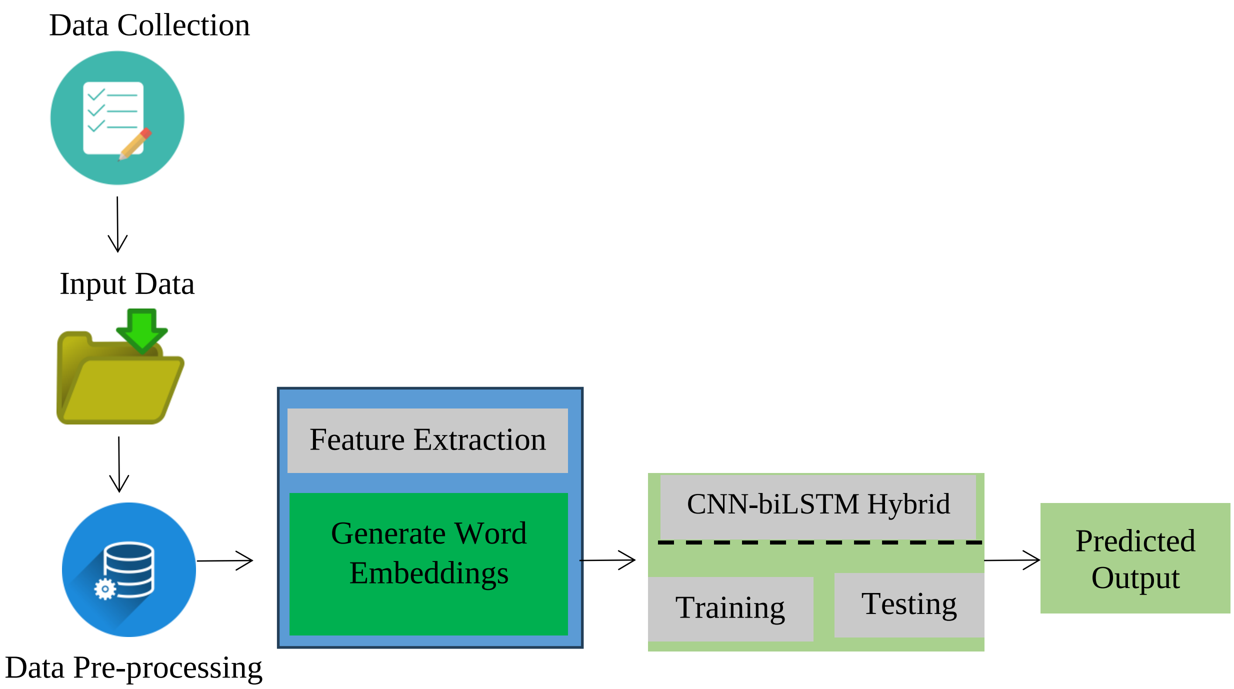


Fig 1: Overview of the proposed system model

Depression categorization distinguishes between tweets that are depressed and those that are not. Parameters are analyzed to compare and evaluate the proposed model with existing classification methods. The hybrid CNN-biLSTM approach is applied in this work to predict depression using Twitter-based depression datasets. Depression prediction improves in precision and accuracy while categorization error decreases.

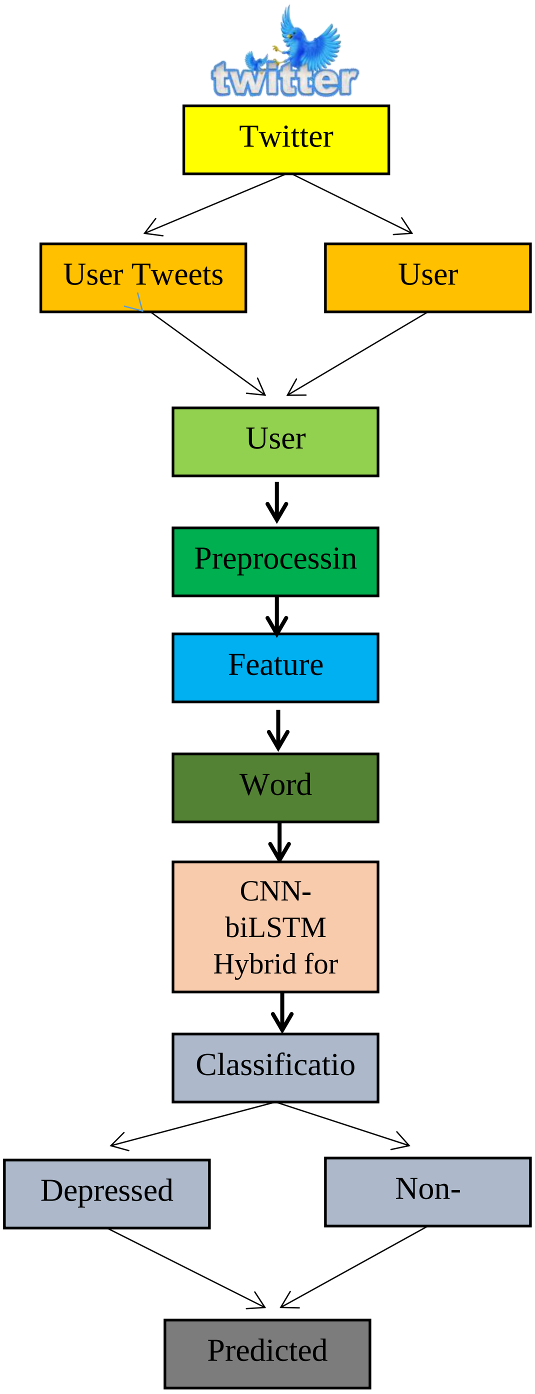


Fig 2: Functional flow of the proposed deep learning framework

As presented in Figure 2, the functional flow of deep learning-based methodology is provided. The framework follows several steps. The framework imports a Twitter list containing normal and depressive symptoms. Remove noise from the data by applying pre-processing procedures. Processing the data by the specifications has a significant favorable impact on the feature extraction quality. The submitted text data is subjected to pre-processing methods such as tokenization, removing stop words, data normalization, and punctuation, among other things. This step performs feature extraction on the collected clean, transparent data. The pre-processed data's feature extraction algorithm is used to identify relevant and noteworthy properties. The important data dimensions are identified by extracted characteristics, which also improve the efficiency of classification algorithms. A suggested hybrid classification technique called CNN-biLSTM is applied to achieve higher accuracy. For training and testing, the classifiers get the optimized features acquired in the feature engineering process. The last step involves validating the system by calculating performance measures using the proposed depression analysis framework.

3.3 Data Pre-processing

One crucial step in data mining is pre-processing the data [41]. Due to its non-domain-specific nature and variety of collection methods, real-world data may be inaccurate, incomplete, and unstructured. When such data is analyzed right away, it results in erroneous forecasts. Our framework uses several methods in its pre-processing phase. The user-defined text patterns are eliminated by the first technique. Eliminating patterns such as "user handles (@username)", "hashtags (#hashtag)", "URLs", "characters, symbols, and numbers other than alphabets", "empty strings", "drop rows with NaN in the column", "duplicate rows", and so on is the aim of this method. This method removes all URLs from every tweet in the dataset and cleans it up. As removing URLs will lower processing complexity and make them useless for prediction, they are not considered. Erasing the time, date, numbers, and hashtags comes next. The time and date are deleted from the tweets since they do not help predict depression. Similarly, while hashtags can be utilized for prediction, numbers are not a suitable factor to consider. It has been noted that using hashtags as the basis for a forecast results in extremely low accuracy. Hashtags have also been eliminated since we don't want to stray from the norm. The next step is eliminating emojis from the phrase and any excess or whitespace.

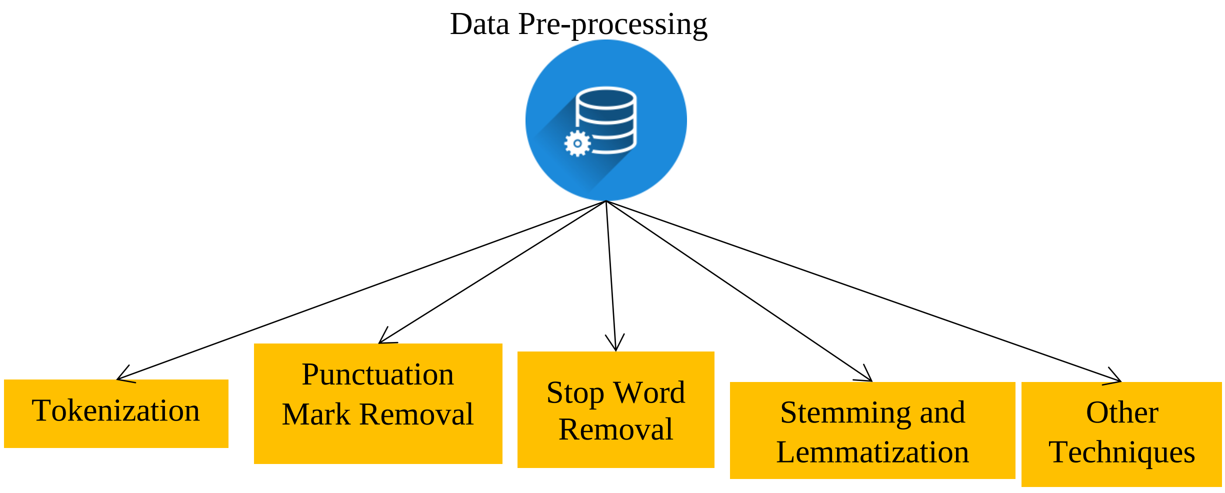


Fig 3: Data pre-processing techniques

The next step is to eliminate stop words and perform stemming. The meaning of a statement is not enhanced by stopwords such as was, at, if, and so on. To eliminate stopwords from our text, we employ a collection of stopwords from the NLTK[10] package. One way to get a phrase back to its original form is by stemming [6]. A stemmer is used to remove a word's prefix or suffix (−ize, −ed, −s, −de, etc.) in order to produce the word's root. Once more, the cleaned tweets are entered into the tokenizer, which is the next step after every tweet has been cleaned up. One important part of the NLP technique pre-processing is tokenizing [54] raw text input. The tools known as tokenizers break a longer text document, divide the given text into smaller words or lines, and then use regular expressions to divide the input into tokens. Figure 3 shows pre-processing techniques.

To utilize the various tokenization functions, import the NLTK package. Providing the tokenizer with cleansed positive and negative tweet datasets is the first step in the tokenization process.the fit_on_texts() method. It uses a list of books to renew the internal vocabulary and constructs the vocabulary index based on word frequency. The lowest index value is given to the term that appears the most frequently. Thus, this function gives an index for each word and the maximum number of words—10,275—in our framework. The method texts_to_sequences() is the next in the tokenization process. Data is obtained from the previous approach, which consists of an index and the maximum number of words. It aims to substitute the corresponding integer value from the word_index for each word in a tweet dictionary by converting each word into a series of numbers. For now, the tweets are transformed into different length integer sequences. Next, length-related tweets less than the Max_Tweet_Length—25 in our frame—are padded with zeroes.

3.4 Word Embeddings

When a word is accessible in extremely sparse vectors, embedding in NLP can help machine learning applications cope with enormous data sets by projecting the word into an embedding vector with modest dimensions. Dense or low-dimensional representations of vectors with large dimensions are called embeddings. Several word vector-based methods that have emerged recently [20, 68] use the provided text corpus as their learning source. This results in large dimensionality in the solution, usually the entire corpus's size. When words with semantically related meanings are grouped in word embeddings [37], they learn from one another and create vectors with almost similar representations; for instance, the words "joyful" and "miserable" have rather distinct meanings. As such, the tokenized numerical vectors will have more distinct properties when they are represented at large distances from one another in the geometric space. An embedded layer modifies these vectors, and recording the semantic link between the corresponding word vectors is important. Even if the embedding vectors in the embedding space do, the first tokenized vector does not have this connection. Instead, they grasp the relationship between words based on how far apart they are. The CNN or RNN networks gain increased prediction capability with each training cycle as additional separable features are retrieved. They are among the most effective methods for encoding a phrase, paragraph, or text to this point and are regarded as one of the innovations in DL that can resolve difficult NLP problems.

We computed a numerical vector for each pre-processed data point using the "Word embeddings" method. We initially converted every word in the sample text into a sequence to construct word indexes. Keras text tokenizer [63] is used to obtain these indexes. We've made sure the tokenizer doesn't assign any word a zero index and that the vocabulary length is appropriately changed. Subsequently, an index unique to each word in the dataset is allocated, and these are used to produce numerical vectors for each text sample. First, the total length of all tweets is gathered to build text sequences. The histogram in Figure 6, which displays the number of tweets increasing in word length, indicates that most tweets in the training set are less than 25 words. Text sequences are therefore converted to integer sequences, and zero padding is used. Since most tweets are less than 25 words, that is the dataset's maximum consecutive length (word count). Since these tweets are incredibly infrequent and add zeros to the vector sequence, slow down model training, and degrade overall performance, five more words are deleted. A matrix of embeddings of the size Max_unique_words * Embedding_dim is made in this investigation. The vector's length, or 300 in our example, is represented by embedding_dim. Zeroes are the first values to fill this matrix. We look for each produced vector from the top 10,275 unique words inside the vector space. The 300-dimensional columns of features in each of the provided vectors, which have a vocabulary of 10,275, are then put in the embedding matrix row by row. With our DNNs, we generated an output embedding vector sized 25 × 50 for every tokenized vector using an embedding layer with a length of 50. The capabilities of our multilayer neural network are restricted to reconstructing the linguistic context of words by utilizing a substantial text corpus as input (the EMBEDDING_FILE). With several hundred dimensions, a vector space is often generated, and a unique vector is made for each unique word in the corpus. The testing and training data sets were divided into positive and negative categories. Testing uses thirty percent of the data, while training uses seventy percent.

3.5 Hybrid Deep Learning Model

To anticipate sad persons on Twitter, we suggested a method combining CNN with bidirectional-LSTM, a kind of RNN (Figure 4), for better classification performance. After conducting several tests, we discovered that CNN effectively removes spatial information and works well when contextual knowledge from the previous sequence is unnecessary. However, RNNs are effective in extracting data when the classification depends on the context of the surrounding objects. Initially, CNN receives multidimensional data directly as a low-level input. Every layer mines critical data throughout the pooling and convolution process. This CNN's hidden layer and output layer are fully linked, in contrast to a typical CNN. Accuracy and detail retrieval of depression tweets are improved using several convolutional layers, pooling layers, and updated convolutional kernels. Overfitting risk exposure might arise in a complex network.

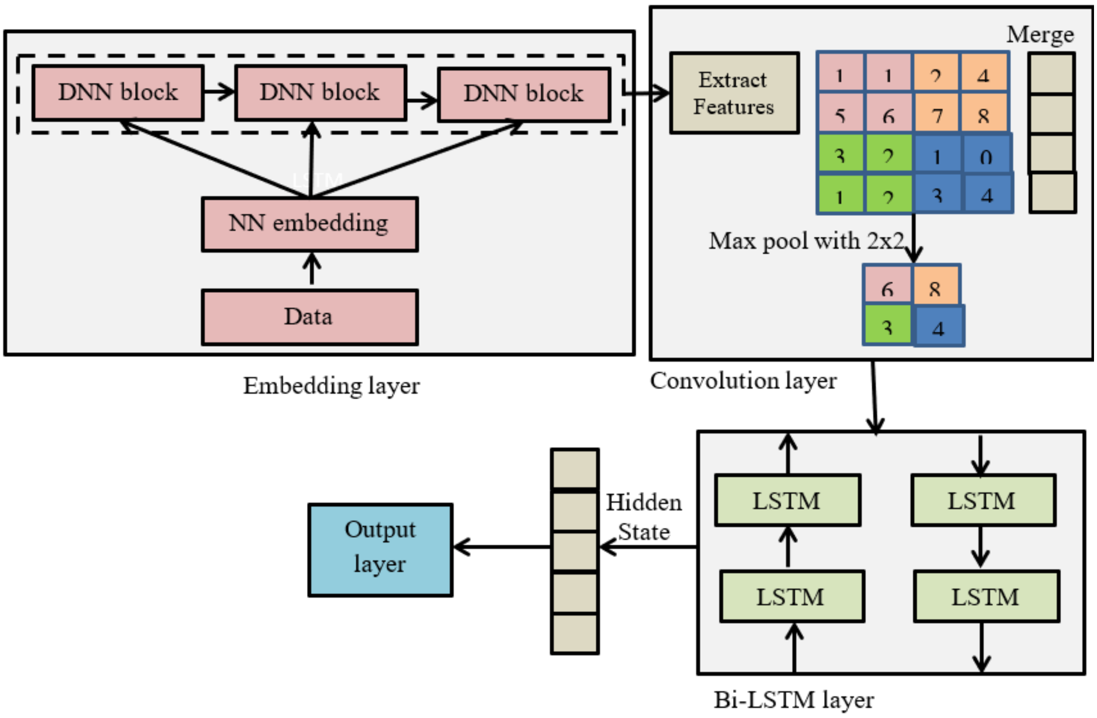


Fig 4: Proposed hybrid deep learning model

The CNN model therefore integrates the Long Short-Term Memory Network (LSTM) system in order to address the sequence problem. Vital information is stored in the state cell for a longer time and is extracted along with comparable and superfluous information using convolution kernels. CNN and biLSTM together produce the majority of the outcomes. Figure 5 presents the biLSTM architecture.

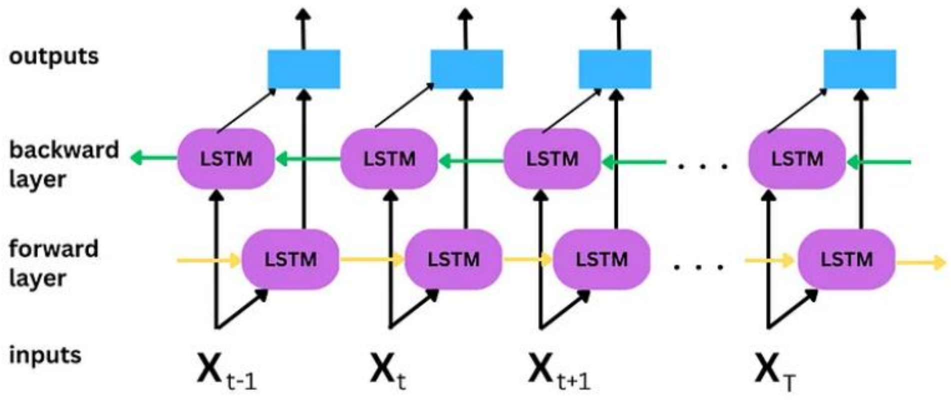


Fig 5: Architectural overview of biLSTM model

When the convolutional layer reduces the number of dimensions and makes it easier to extract low-dimensional semantic characteristics from textual material, biLSTM is utilized. Additionally, the text is processed using biLSTM as an input sequence. To improve performance on the input vectors, multiple 1-dimensional convolutional kernels are combined in this work. According to Eq. 1, the average of each word's embedding vector is how sequential input data is defined. Using the convolutional kernel sizes, 1D CNN is used to apply the X1: T characteristics for Bigram, Trigram, and Unigram. The input comprises the features generated in the t^{th} Convolutional, which takes a window for D words extending from tØ t + d. According to Eq. 2, the convolution procedure produces features for that window.

$$X_1 : T = [x_1, x_2, x_3, x_4, \dots, x_T] \quad (1)$$

$$h_d, t = \tan h(W_d x_{t:t+d-1} + b_d) \quad (2)$$

In the context window, Wd represents the embedding vector of every distinct word, xt Ø t + d-1 are the parameters with the matrix of the learnable weight, and bd is the bias. Each filter also convolves different portions of the text, and the feature map produced by the filter with a size d convolution is given by Eq. 3.

$$h_d = [h_{d1}, h_{d2}, h_{d3}, h_{d4}, \dots x_{T-d+1}] \quad (3)$$

CNN can identify the latent correlation between many nearby words by employing several convolutional kernels of various sizes. Convolution filters reduce the trainable parameters during feature learning, making them useful for feature extraction from textual input. A max-pooling layer is placed after the convolutional layers to do this. [5]. Several convolutional channels, each with a unique set of values, are initially used to process the input. The max-pooling procedure chooses and pools the greatest value from each convolutional layer to produce a new collection of features. To obtain Eq. 4, max pooling is performed to the feature maps of each convolution kernel with a convolutional size of d. Joining pd together every filter size (d = 1, 2, and 3) results in recovering the last characteristics for every window. The hidden features of the bigram, trigram, and unigram are then extracted as indicated by Equation (5).

$$p_d = Maxt(h_{d1}, h_{d2}, h_{d3}, h_{d4}, \dots x_{T-d+1}) \quad (4)$$

$$h_d = [p_1, p_2, p_3] \quad (5)$$

Compared to the traditional LSTM, the main benefit of utilizing a CNN-based feature extraction approach is a considerable decrease in the total amount of features. After feature extraction, a depression prediction model uses these features again. When sequential data is used, gate structures such as input, output, and forget gates paired with the cell states of the LSTM architecture can be used to get around the "vanishing gradient" issue. As additive connections between the states, these cell states serve as the LSTM unit's collective long-term memory. Equation (6, 7, 8, 9, 10, and 11) calculates the output state of an LSTM cell at a given time t, given both the intermediate state (ht) and the input (xt).

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

$$i_t = \sigma(W_i x_t + U_i h_{t-1} + b_i) \quad (7)$$

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

$$g_t = tanh(W_g x_t + U_g h_{t-1} + b_g) \quad (9)$$

$$c_t = f_t o c_{t-1} + i_t o g_t \quad (10)$$

$$h_t = o_t o \tanh(c_t) \quad (11)$$

W, U, and b are the learnable parameters in this instance. Convolution operator (o) and sigmoid function (σ) are identified. The forget, input, and output gates of an LSTM are represented by the symbols ft, it, and ot, respectively, but the memory or cell status is displayed by ct. Cell state is employed to capture any long-term relationships in the data presented in the input sequence, which is the only explanation for LSTMs' applicability to longer sequences. The three convolution layers that make up the CNN portion of the network have varying amounts of filters. Repaired linear activation unit (Relu) and sigmoid activation functions comprise the first two layers, consisting of 128 filters with a 3 x 3 kernel size. A sigmoid activation function and a 3 x 3 kernel size are features of the 64 filters in the third convolution layer. This is followed by a Max-pool layer with a dimension of 4 × 4. Lastly, a biLSTM with marginally distinct hidden computations was employed. We may address the limitation of RNNs: only data from earlier computations is utilized for subsequent calculations by using computations that are performed both forward and backward. We employed a "dropout layer" with 0.1 for the "keep probability" to avoid overfitting on the training set. The optimizer "root mean squared propagation (RMSprop)" and the loss function "binary_crossentropy" are used to train the model. Relu activation was utilized for the output layer. A popular social networking website, Twitter offers simple and open access to data. It takes a long time to design and verify terms that individuals with mental illness utilize as a lexicon when perusing content.

3.6 Proposed Algorithm

We introduced the Learning-Based Depression Detection (LBDD) algorithm, which uses Twitter tweets as input and categorizes them based on the likelihood that they include depression.

Algorithm: Learning Based Depression Detection (LBDD)

Input: Twitter dataset D

Output: Depression detection results R, performance statistics P

1. Begin
2. D←Preprocess(D)
3. features←FeatureExtraction(D)
4. embeddings←WordEmbeddings(features)
5. (T1, T2)←DataSplit(D, embeddings)
6. Configure CNN-biLSTM hybrid model m (as in Figure 4)
7. Compile m
8. Train m using T1
9. Save m for future reuse
10. Load m
11. R←Test(m, T2)
12. P←Evaluate(R, ground truth)
13. Display R
14. Display P
15. End

Algorithm 1: Learning Based Depression Detection (LBDD)

Using the Twitter dataset as input, Algorithm 1 applies several processes to identify depression probability based on posts on social media. The given data is subjected to preprocessing to improve the quality of supervised learning. Feature extraction is carried out to identify the best-contributing features for classification. Deep learning-based generation of word embeddings provides a more robust data representation of data suitable for machine learning. Word embeddings are used to train the hybrid deep learning model, which is stored for further use. The test data comprising tweets are subjected to depression detection by the trained hybrid deep learning model. This process results in the classification of test data. The performance evaluation is made by comparing ground truth with predicted labels.

3.7 Evaluation Methodology

Since our strategy was learning-based (supervised learning), metrics obtained from Figure 6's confusion matrix are employed to assess our approach.

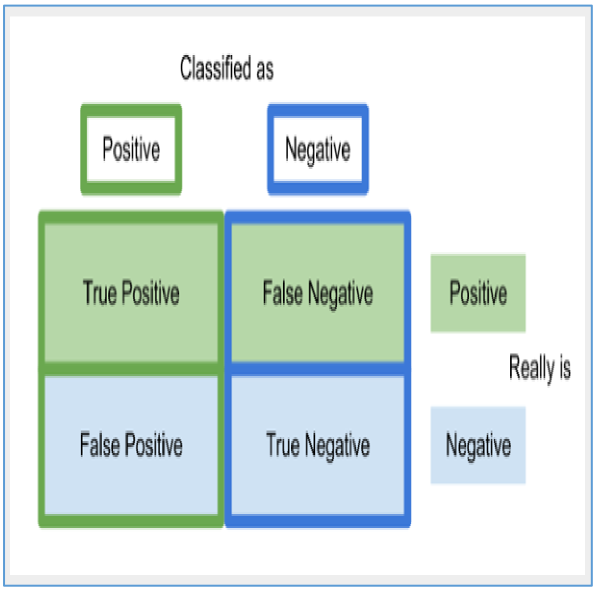


Fig 6: Confusion matrix

Performance statistics are obtained by comparing the predicted labels of our algorithm with the ground truth, based on the confusion matrix. Eq. 12 to Eq. 15 express different metrics used in performance evaluation.

Precision (p) $= \frac{TP}{TP+FP}$ (12)

Recall (r) $= \frac{TP}{TP+FN}$ (13)

F1-score $= 2 * \frac{(p * r)}{(p+r)}$ (14)

Accuracy $= \frac{TP+TN}{TP+TN+FP+FN}$ (15)

A number between 0 and 1 is produced by the performance evaluation metrics. In machine learning research, these measures are often employed.

4. EXPERIMENTAL RESULTS

Here, we report the findings of our study employing a hybrid deep learning model intended to identify signs of sadness in postings on social media automatically. TensorFlow and Keras frameworks in Python 3 are used to create our framework. We used a Twitter dataset for our study. Detecting depression in Twitter posts has great potential for both research and practical applications in mental health. A significant quantity of data from Twitter may be examined to find trends and patterns in mental health. Detecting depression through tweets can assist in creating personalized mental health interventions based on individual expression and needs.

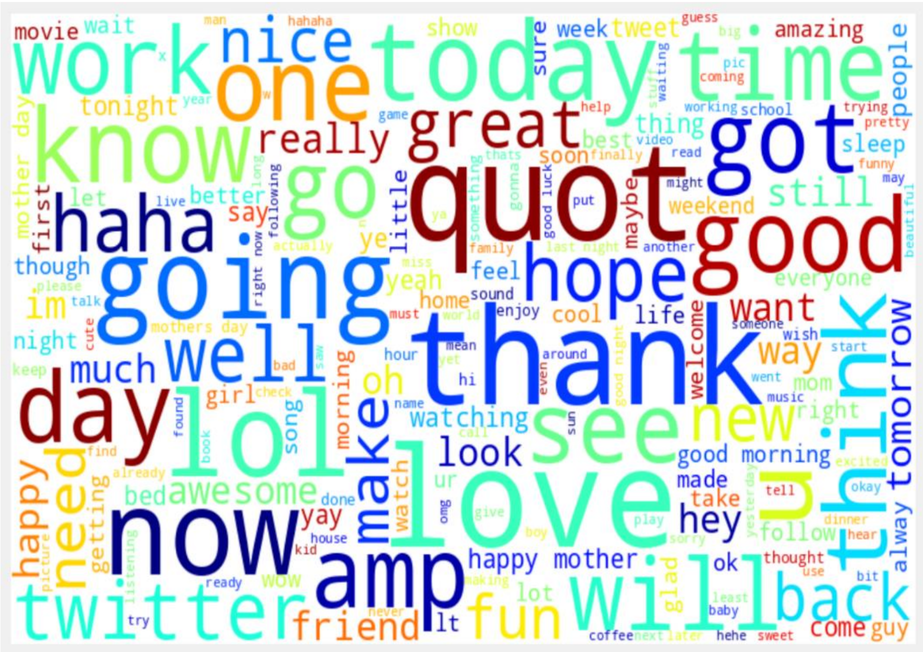


Fig 7: Word cloud generated using random tweets

As presented in Figure 7, it was observed that the tweets were randomly selected to create a word cloud, a mix of different words, including words reflecting human emotions.

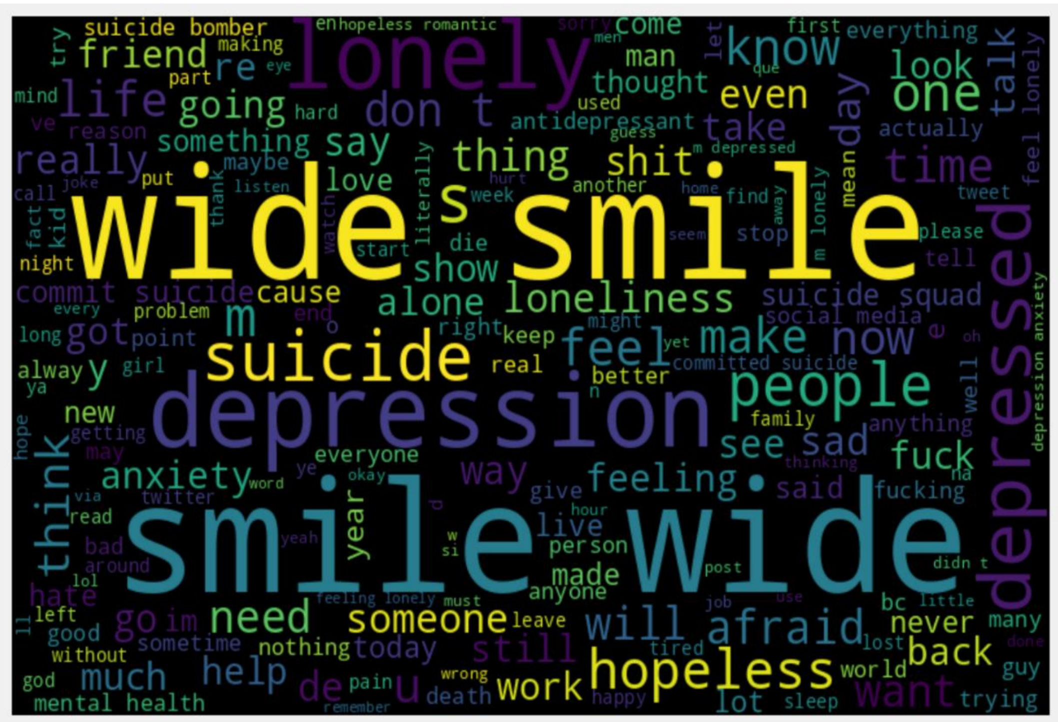
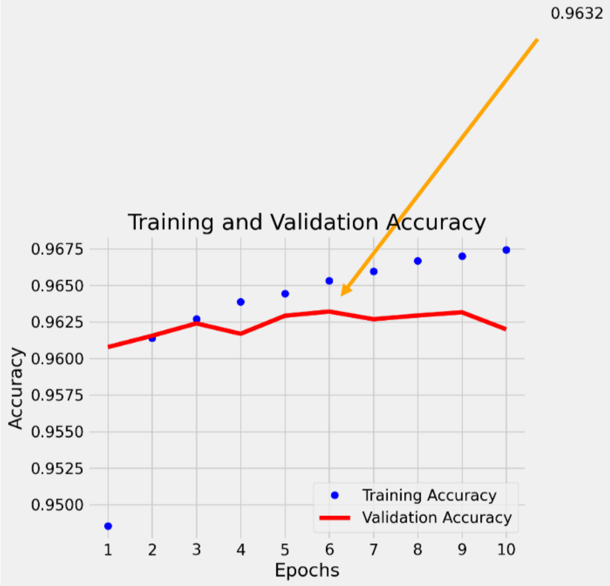


Fig 8: Word cloud generated using depressed tweets

As presented in Figure 8, it was observed that depressed tweets were selected to create a word cloud, words reflecting more depressive emotions.

Training and Validation accuracy	Training and Validation loss



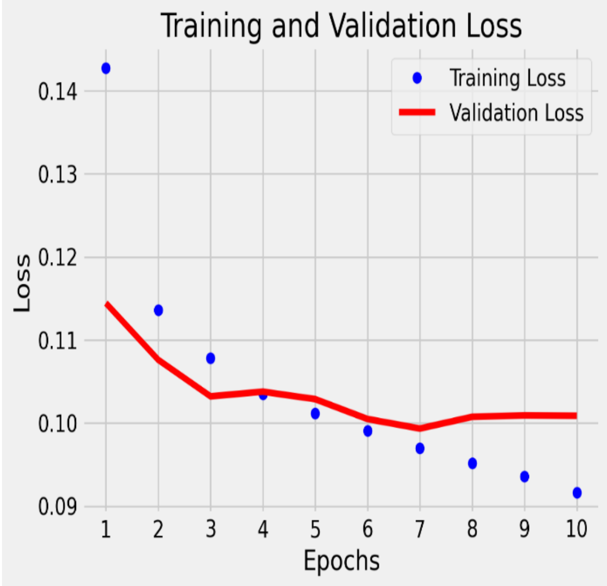
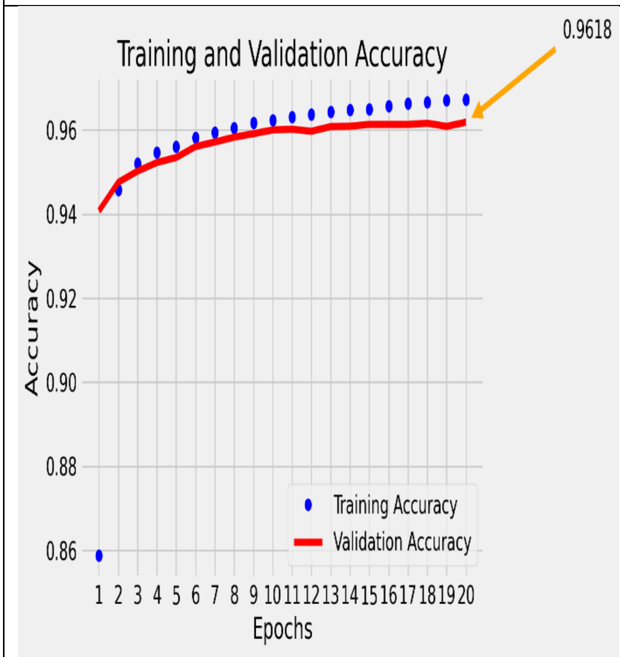


Fig 9: The accuracy and loss analysis of the LSTM model in depression detection

Figure 9 shows two plots that compare training and validation metrics over 10 epochs exhibited by the LSTM model. The plot on the left illustrates "Training and Validation Accuracy." The training accuracy, shown with blue dots, demonstrates a slight upward trend, reaching approximately 0.9632 by the 10th epoch. In contrast, the validation accuracy, represented by a red line, fluctuates around 0.9625 without a clear trend. The plot on the right presents "Training and Validation Loss." The training loss, indicated by blue dots, consistently decreases from about 0.11 to below 0.09 over the epochs. The validation loss, shown by a red line, follows a similar downward trend but stabilizes around 0.1 after the 5th epoch. This suggests that although the model performs better on training data, its validation performance is mostly consistent.

Training and Validation accuracy	Training and Validation loss



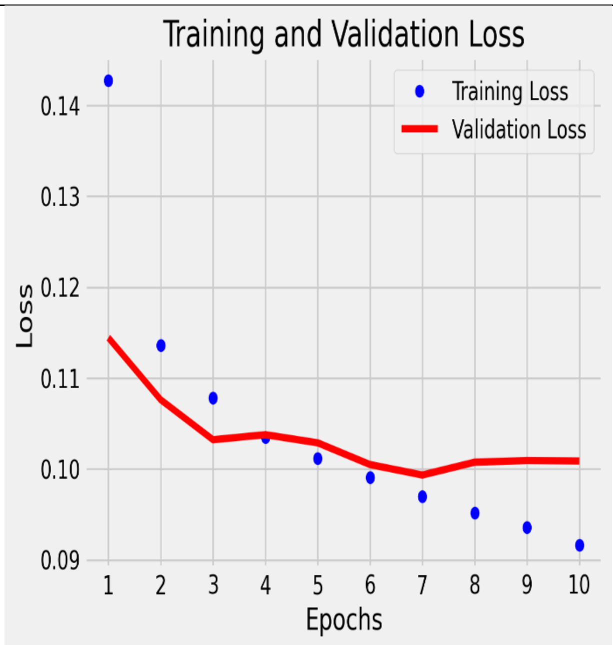


Fig 10: The accuracy and loss analysis of the simple RNN model in depression detection

Two graphs comparing the training and validation performance measures of a basic RNN model across 20 epochs are displayed in Figure 10. The "Training and Validation Accuracy" graph is located on the left." In this graph, the training accuracy, represented by blue dots, steadily increases from about 0.86 to 0.9618. The validation accuracy, shown by a red line, follows a similar upward trend and stabilizes around 0.96 towards the 20th epoch. This is the "Training and Validation Loss" graph on the right. Blue dots represent the training loss, which drops during the epochs from about 0.14 to less than 0.09. The validation loss, represented by a red line, initially decreases from around 0.14 to 0.10 within the first few epochs and then fluctuates slightly, maintaining a generally stable trend around 0.10. Overall, this shows that both the training and validation accuracies improve consistently. The validation loss stabilizes, indicating that the model is learning effectively without overfitting.

Training and Validation accuracy	**Training and Validation loss**

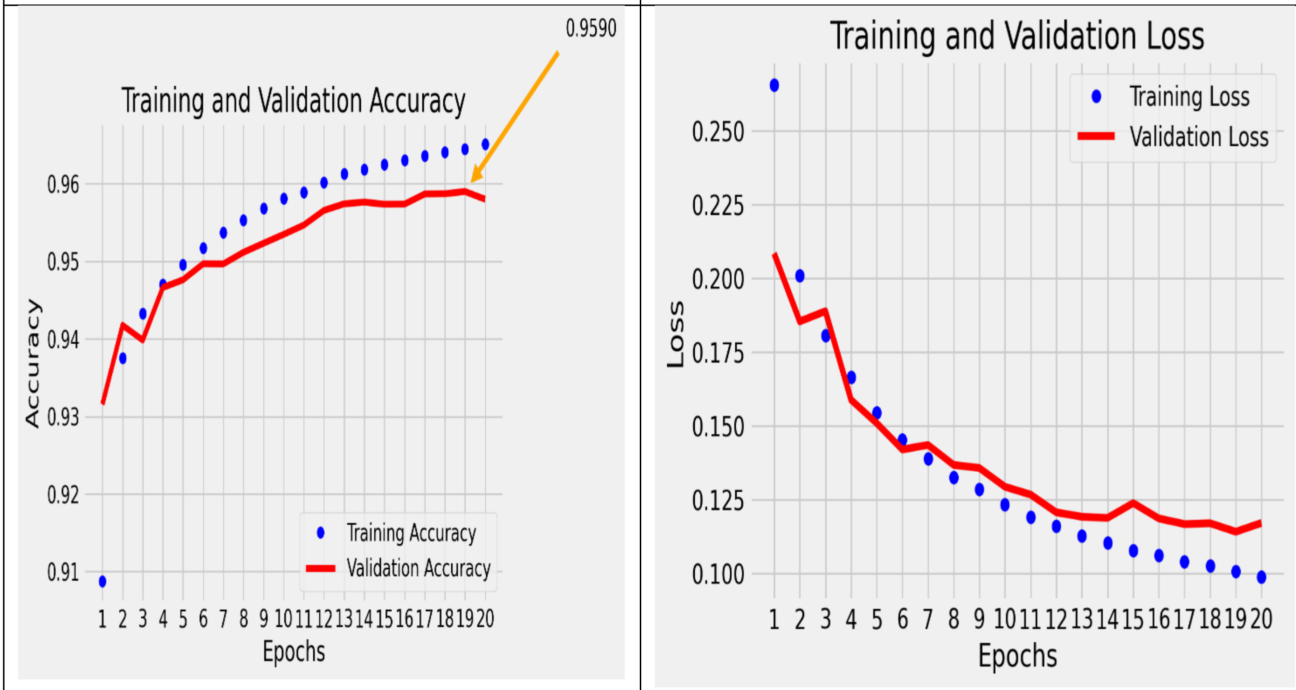


Fig 11: The accuracy and loss analysis of the proposed hybrid model in depression detection

In Figure 11, two graphs illustrate the training and validation performance metrics over 20 epochs. The graph on the left shows "Training and Validation Accuracy." The training accuracy, represented by blue dots, increases from approximately 0.91 to 0.9590. In contrast, the validation accuracy, shown by a red line, rises from about 0.92 to just below 0.96, following a similar trend to the training accuracy. The graph on the right displays "Training and Validation Loss." The training loss, indicated by blue dots, steadily decreases from about 0.25 to below 0.10. Similarly, the validation loss, depicted by a red line, shows a significant decrease from 0.20 to 0.10 within the first 10 epochs and stabilizes with minor fluctuations. These results suggest the model effectively learns and improves its performance on training and validation datasets without overfitting.

Table 2: Performance comparison among depression detection models

Depression Detection Model	**Precision**	**Recall**	**F1-Score**	**Accuracy**
Simple RNN	0.9023	0.895	0.899	0.9304
LSTM	0.925	0.91	0.9023	0.9398
Hybrid Model (Proposed)	0.96	0.96	0.96	0.9632

Table 1 compares the suggested hybrid deep learning model's depression detection performance to that of cutting-edge methods.

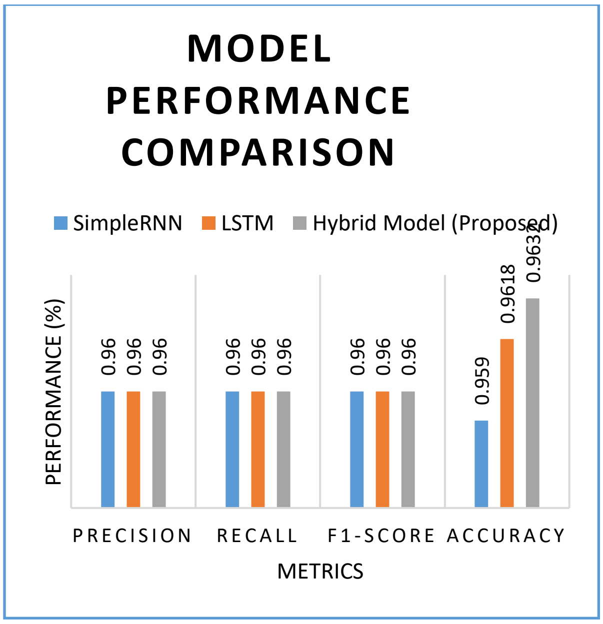


Fig 12: Performance comparison of depletion detection models

Four measures were used to assess the performance of the three models (SimpleRNN, LSTM, and CNN-LSTM) in Figure 12: precision, recall, F1-Score, and accuracy. With the highest Precision, Recall, F1-Score, and Accuracy values, the CNN-LSTM model consistently beat the other two models, according to the findings. In terms of precision and accuracy, the LSTM model outperformed SimpleRNN, whereas SimpleRNN displayed the worst results in terms of all metrics. These results highlight how well the CNN-LSTM model performed in this assessment. According to the findings, the suggested deep learning model had the best accuracy, at 96.32%.

5. Discussion

A deep learning-based approach for automatically identifying sadness in social media messages is presented in this study. According to our literature analysis, tweets are frequently utilized to extract contextual information using deep learning models like CNN. On the other hand, LSTM models perform better when assessing data to identify depression. Inspired by this, we proposed a hybrid deep learning model that enhances detection performance by fusing CNN with biLSTM. Preprocessing and word embeddings are part of our approach to preparing the data for supervised learning. Our dataset was gathered and annotated from Twitter, a popular social media site. According to experimental results, our suggested hybrid deep learning model outperforms state-of-the-art methods. Nonetheless, section 5.1 discusses a few restrictions.

5.1 Limitations

There is room for improvement in the suggested deep learning framework for automatically determining the likelihood of depression in social media messages. However, the framework has limitations. The dataset used for experiments is limited in size and may not be sufficient for generalizing findings. Additionally, the lack of data diversity is a concern as the data is collected from a single source. Another significant issue is the absence of a distributed programming framework and the utilization of big data concepts. Given the large nature of social media data, it is important to consider leveraging big data concepts and distributed programming paradigms.

6. Conclusion and Future Work

We have created a hybrid model that combines bi-directional long short-term memory (biLSTM) and convolutional neural network (CNN) models in a deep learning framework. Our goal was to extract characteristics from the data and temporal correlations efficiently. Our approach combines CNN with a kind of RNN called bidirectional-LSTM to produce better classification performance in predicting depression in Twitter users. Through experimentation, we observed that CNN performs well when contextual knowledge from the preceding sequence is unnecessary and efficiently eliminates spatial features. On the other hand, RNNs excel in data extraction, where categorization relies on the context of surrounding objects. Our suggested deep learning architecture can automatically identify sadness in content shared on social media. Additionally, we have created an algorithm known as Learning Based Depression Detection (LBDD), which categorizes messages from Twitter users according to their likelihood of containing depressive symptoms. Our methodology was evaluated using a benchmark dataset, and we found that our deep learning model outperforms many existing models with the highest accuracy rate of 96.32%. We plan to enhance our framework to analyze human expressions in images or videos and audio content for automatic depression detection.

References

[1] Paffenbarger RS Jr, Lee IM, Leung R (1994) Physical activity and personal characteristics associated with depression and suicide in American college men. Acta Psychiatrica Scandinavica 89:16–22. https://doi.org/ 10.1111/j.1600-0447.1994.tb05796.x.

[2] Zafar A, Chitnis S (2020) Survey of depression detection using social networking sites via datamining.In:202010thInternationalConferenceon CloudComputing,DataScience&Engineering(Confluenc).IEEE, pp 88–93. https://doi.org/10.1109/Confluence47617.2020.9058189.

[3] Kohrt BA, Speckman RA, Kunz RD, Baldwin JL, Upadhaya N, Acharya NR, … Worthman CM (2009) Culture in psychiatric epidemiology: using ethnography and multiple mediator models to assess the relationship of caste with depression and anxiety in Nepal. Annals of Human Biology 36(3):261–280. https://doi.org/10.1080/03014460902839194 .

[4] Oquendo MA, Ellis SP, Greenwald S, Malone KM, Weissman MM, Mann JJ (2001) Ethnic and sex differences in suiciderates relativeto major depression in the UnitedStates. American Journal of Psychiatry 158(10):1652–1658. https://doi.org/10.1176/appi.ajp.158.10.1652

[5] Beard C, Millner AJ, Forgeard MJ, Fried EI, Hsu KJ, Treadway MT, … Björgvinsson T (2016) Network analysis of depression and anxiety symptom relationships in a psychiatric sample. Psychological Medicine 46(16):3359–3369. https://doi.org/10.1017/S0033291716002300.

[6] Biradar A, Totad SG (2018) Detecting Depression in Social Media Posts Using Machine Learning. In: International Conference on Recent Trends in Image Processing and Pattern Recognition. Springer,

Singapore, pp 716–725. https://doi.org/10.1007/978-981-13-9187-3_64

[7] Arora P, Arora P (2019) Mining twitter data for depression detection. In: 2019 International Conference on Signal Processing and Communication (ICSC). IEEE, pp 186–189. https://doi.org/10.1109/ICSC45622. 2019.8938353

[8] Shreya Ghosh and Tarique Anwar; (2021). *Depression Intensity Estimation via Social Media: A Deep Learning Approach . IEEE Transactions on Computational Social Systems.* http://doi:10.1109/TCSS.2021.3084154

[9] Keshu Malviya; Bholanath Roy and SK Saritha; (2021). *A Transformers Approach to Detect Depression in Social Media . 2021 International Conference on Artificial Intelligence and Smart Systems (ICAIS).* http://doi:10.1109/icais50930.2021.9395943

[10] Chenhao Lin; Pengwei Hu; Hui Su; Shaochun Li; Jing Mei; Jie Zhou and Henry Leung; (2020). *SenseMood: Depression Detection on Social Media . Proceedings of the 2020 International Conference on Multimedia Retrieval.* http://doi:10.1145/3372278.3391932

[11] Giuntini, Felipe T.; Cazzolato, Mirela T.; dos Reis, Maria de Jesus Dutra; Campbell, Andrew T.; Traina, Agma J. M. and Ueyama, Jó (2020). *A review on recognizing depression in social networks: challenges and opportunities. Journal of Ambient Intelligence and Humanized Computing.* http://doi:10.1007/s12652-020-01726-4

[12] Shini Renjith, Annie Abraham, Surya B. Jyothi, Lekshmi Chandran and Jincy Thomson. (2022). An ensemble deep learning technique for detecting suicidal ideation from posts in social media platforms. *Elsevier*. 34(10), pp.9564-9575. https://doi.org/10.1016/j.jksuci.2021.11.010

[13] Yang, Xingwei; McEwen, Rhonda; Ong, Liza Robee and Zihayat, Morteza (2020). *A big data analytics framework for detecting user-level depression from social networks. International Journal of Information Management, 54, 102141–* . http://doi:10.1016/j.ijinfomgt.2020.102141

[14] Kailai Yang, Tianlin Zhang and Sophia Ananiadou. (2022). A mental state Knowledge–aware and Contrastive Network for early stress and depression detection on social media. *Elsevier*. 59(4), pp.1-16. https://doi.org/10.1016/j.ipm.2022.102961

[15] Tapotosh Ghosh, Md. Hasan Al Banna, Md. Jaber Al Nahian, Mohammed Nasir Uddin, M. Shamim Kaiser and Mufti Mahmud. (2023). An attention-based hybrid architecture with explainability for depressive social media text detection in Bangla. *Elsevier*. 213(C), pp.1-17. https://doi.org/10.1016/j.eswa.2022.119007

[16] Esteban A. Ríssola; David E. Losada and Fabio Crestani; (2021). *A Survey of Computational Methods for Online Mental State Assessment on Social Media . ACM Transactions on Computing for Healthcare.* http://doi:10.1145/3437259

[17] Ziwei, Bernice Yeow and Chua, Hui Na (2019). *Proceedings of the 2nd International Conference on Computing and Big Data - ICCBD 2019 - An Application for Classifying Depression in Tweets. 37–41.* http://doi:10.1145/3366650.3366653

[18] Tariq, Subhan; Akhtar, Nadeem; Afzal, Humaira; Khalid, Shahzad; Mufti, Muhammad Rafiq; Hussain, Shahid; Habib, Asad and Ahmad, Ghufran (2019). A novel Co-training based approach for the classification of mental illnesses using Social media posts. IEEE Access, 1–1. http://doi:10.1109/ACCESS.2019.2953087

[19] Sancheng Peng, Lihong Cao, Yongmei Zhou, Zhouhao Ouyang, Aimin Yang, Xinguang Li, Weijia Jia and Shui Yu. (2022). A survey on deep learning for textual emotion analysis in social networks. *Elsevier*. 8(5), pp.745-762. https://doi.org/10.1016/j.dcan.2021.10.003

[20] Ruba Skaik and Diana Inkpen; (2021). Using Social Media for Mental Health Surveillance . ACM Computing Surveys. http://doi:10.1145/3422824

[21] Adikari, Achini; Gamage, Gihan; de Silva, Daswin; Mills, Nishan; Wong, Sze-Meng Jojo and Alahakoon, Damminda (2020). A self structuring artificial intelligence framework for deep emotions modelling and analysis on the social web. Future Generation Computer Systems, S0167739X20330053–. http://doi:10.1016/j.future.2020.10.028

[22] Chatterjee, Ankush; Gupta, Umang; Chinnakotla, Manoj Kumar; Srikanth, Radhakrishnan; Galley, Michel and Agrawal, Puneet (2018). Understanding emotions in text using deep learning and big data. Computers in Human Behavior, S0747563218306150–. http://doi:10.1016/j.chb.2018.12.029

[23] Yang, T., Li, F., Ji, D., Liang, X., Xie, T., Tian, S., and Liang, P. (2021). Fine-grained depression analysis based on Chinese micro-blog reviews. Information Processing & Management, 58(6), 102681. http://doi:10.1016/j.ipm.2021.102681

[24] Xiaoxu Yao; Guang Yu; Jingyun Tang and Jialing Zhang; (2021). Extracting depressive symptoms and their associations from an online depression community . Computers in Human Behavior. http://doi:10.1016/j.chb.2021.106734

[25] Lei Cao; Huijun Zhang and Ling Feng; (2022). Building and Using Personal Knowledge Graph to Improve Suicidal Ideation Detection on Social Media . IEEE Transactions on Multimedia. http://doi:10.1109/tmm.2020.3046867

[26] Sanja Scepanovic; Enrique Martin-Lopez; Daniele Quercia and Khan Baykaner; (2020). Extracting medical entities from social media . Proceedings of the ACM Conference on Health, Inference, and Learning. http://doi:10.1145/3368555.3384467

[27] Pran, Md. Sabbir Alam; Bhuiyan, Md. Rafiuzzaman; Hossain, Syed Akhter and Abujar, Sheikh (2020). 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT) - Analysis Of Bangladeshi People's Emotion During Covid-19 In Social Media Using Deep Learning. 1–6. http://doi:10.1109/ICCCNT49239.2020.9225500

[28] Guillermo Blanco and Analia ´ Lourenço. (2022). Optimism and pessimism analysis using deep learning on COVID-19 related twitter conversations. *Elsevier*. 59(3), pp.1-15. https://doi.org/10.1016/j.ipm.2022.102918

[29] Subramani, Sudha; Michalska, Sandra; Wang, Hua; Du, Jiahua; Zhang, Yanchun and Shakeel, Haroon (2019). *Deep Learning for Multi-Class Identification From Domestic Violence Online Posts. IEEE Access, 7, 46210–46224.* http://doi:10.1109/ACCESS.2019.2908827

[30] Nawshad Farruque, Randy Goebel, Sudhakar Sivapalan and Osmar Zaïane. (2024). Deep temporal modelling of clinical depression through social media text. *Elsevier*. 6, pp.1-14. https://doi.org/10.1016/j.nlp.2023.100052

[31] ABHINAV KUMAR, JYOTI KUMARI and JIESTH PRADHAN. (2023). Explainable Deep Learning for Mental Health Detection from English and Arabic Social Media Posts. *ACM*, pp.1-18. https://doi.org/10.1145/3632949

[32] USMAN AHMED, JERRY CHUN-WEI LIN and GAUTAM SRIVASTAVA. (2023). Graph Attention Network for Text Classification and Detection of Mental Disorder. *ACM*. 17(3), pp.1-31. https://doi.org/10.1145/3572406

[33] Jesia Quader Yuki; Md. Mahfil Quader Sakib; Zaisha Zamal; Sabiha Haque Feel and Mohammad Ashrafuzzaman Khan; (2020). *Detecting Depression from Human Conversations . Proceedings of the 8th International Conference on Computer and Communications Management.* http://doi:10.1145/3411174.3411187

[34] Esteban A. Ríssola; Seyed Ali Bahrainian and Fabio Crestani; (2020). A Dataset for Research on Depression in Social Media . Proceedings of the 28th ACM Conference on User Modeling, Adaptation and Personalization. http://doi:10.1145/3340631.3394879

[35] Felipe Taliar Giuntini; Kaue L. de Moraes; Mirela T. Cazzolato; Luziane de Fatima Kirchner; Maria de Jesus D. Dos Reis; Agma J. M. Traina; Andrew T. Campbell and Jo Ueyama; (2021). Modeling and Assessing the Temporal Behavior of Emotional and Depressive User Interactions on Social Networks . IEEE Access. http://doi:10.1109/access.2021.3091801

[36] Sanjana Mendu; Anna Baglione; Sonia Baee; Congyu Wu; Brandon Ng; Adi Shaked; Gerald Clore; Mehdi Boukhechba and Laura Barnes; (2020). A Framework for Understanding the Relationship between Social Media Discourse and Mental Health . Proceedings of the ACM on Human-Computer Interaction. http://doi:10.1145/3415215

[37] Kuhaneswaran AL Govindasamy and Naveen Palanichamy; (2021). Depression Detection Using Machine Learning Techniques on Twitter Data . 2021 5th International Conference on Intelligent Computing and Control Systems (ICICCS). http://doi:10.1109/ICICCS51141.2021.9432203