Depression Detection in Social Media Using NLP and Hybrid Deep Learning Models

Dr. S M Padmaja[1], Dr. Sanjiv Rao Godla[2], Janjhyam Venkata Naga Ramesh[3], Elangovan Muniyandy[4], Pothumarthi Sridevi[5], Prof. Ts. Dr. Yousef A.Baker El-Ebiary[6], Dr. David Neels Ponkumar Devadhas[7]

Professor, Department of Electrical and Electronics Engineering, Shri Vishnu Engineering College for Women, Bhimavaram, India[1]
Professor, Dept of Computer Science and Engineering, Aditya University, Surampalem, Andhra Pradesh, India[2]
Department of CSE, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Andhra Pradesh, India[3]
Adjunct Professor, Department of CSE, Graphic Era Hill University, Dehradun, 248002, India[3]
Adjunct Professor, Department of CSE, Graphic Era Deemed To Be University, Dehradun, 248002, Uttarakhand, India[3]
Department of Biosciences, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Chennai, India[4]
Applied Science Research Center, Applied Science Private University, Amman, Jordan[4]
Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Guntur Dist., Andhra Pradesh - 522302, India[5]
Faculty of Informatics and Computing, UniSZA University, Malaysia[6]
Professor, Department of Electronics and Communication Engineering, Vel Tech Rangarajan Dr.Sagunthala R&D Institute of Science and Technology, Chennai, Tamil Nadu, India[7]

Abstract*—One type of feeling that possesses a detrimental effect on people's day-to-day lives is depression. Globally, the number of persons experiencing long-term sentiments is rising annually. Many psychiatrists find it difficult to recognize mental disease or unpleasant emotions in patients before it's too late to improve treatment. Finding depression in individuals quickest possible time represents one of the most difficult problems. To create tools for diagnosing depression, researchers are employing NLP to examine written content shared on social media sites. Traditional techniques frequently have problems with scalability and poor precision. To overcome the drawbacks of the prior methods, it is suggested to introduce an improved depression detection system based on the RoBERTa (Robustly optimized BERT approach) and BiLSTM (Bidirectional Long Short-Term Memory) approach. This proposed work aims is to take advantage of the contextualized word embeddings from RoBERTa and the sequential learning properties of BiLSTM to determine depression from social media text. The technique is innovative because it combines the use of BiLSTM to accurately describe the temporal patterns of text sequences with RoBERTa to capture subtle linguistic aspects. It removes stopwords and punctuations form the input data to provide clean data to the model for processing. The system illustrates preference over the existing models as they achieve a 99.4 % accuracy, 98. 5% precision, 97. 1% recall, and 97. 3% F1 score. Thus, these results clearly highlight the effectiveness of the combination of the proposed technique with the traditional method in identifying depression with more accuracy and less variance. The proposed method is implemented using python.

Keywords*—*Depression detection; RoBERTa; BiLSTM; social media analysis; deep learning

I. Introduction

Depression, commonly referred as depressed disorder, is a prevalent mental health issue. It is defined by a prolonged depressed incident, joy reduction, or apathy in activities. It is important to distinguish between depression and normal emotional fluctuations or mood swings. Among other sectors of life, it may have an impact on ties with intimate friends, family, and the community. It might be the cause of, or a contributing factor to, issues in the workplace and in educational settings. Depression can hit anybody at any time. People who have experienced horrible tragedies, assault, or other traumatic events are more likely to suffer from depression. Depression is more common in women than in males [1]. Depression is linked to significant psychological, societal, and financial burdens and is a major contribution to the worldwide disease burden. Young people's depression is a developing issue due to two factors: first, it happens during a period of critical life development that includes substantial emotional, social, and intellectual development transitions; second, its frequency has sharply increased in this age group in recent years, particularly among females [2]. Depression is indicated by depressing emotions, emptiness, or irritability as well as changes in cognition and physical functioning that last for a minimum of two weeks and significantly hinder an individual's ability to operate [3].

The most prevalent mental disorder, depression affects 5% of individuals worldwide, with 20% of cases being severe. Adults in their middle and later years are the most at risk. The prevalence of depression is increasing globally, with a constant increase projected between 2005 and 2023. Early professional intervention can treat physical problems and alleviate mental disease, extending individuals' lifespans by reducing their risk of underlying diseases [4]. Early identification of mental illness can significantly improve a person's personal, professional, and social lives, as well as their health. Current methods, based on clinical processes, rely on language and connection between depressed individuals. An automatic depression detection technique is needed [5]. Emotions and primal instincts have a greater ability to control people than reason and argument. Negative feelings are used by social media to encourage interaction and rage. The internet's quick growth provided a means for people to voice their ideas, views, and sentiments in a virtual setting. Individuals dealing with mental health concerns use social media sites like Twitter, Instagram, Reddit, and others to express themselves.

The most common ways to convey emotions are via written words, images, audio, or video. On the other hand, depressed individuals attempt to conceal who they are and are reluctant to share their images. They don't seem as interested in striking up a conversation with strangers. Also, they enjoy engaging in text-based small talk. Text is being employed increasingly and greater as the key criterion for recognizing depression since it is minimal latency, takes less bandwidth, and needs entirely less memory space [6]. Finding signs of mental illness in messages that are published on social media platforms is the subject of recent research. Depression detection systems ought to be created using cutting-edge learning strategies and principles from artificial intelligence. Reliable designs, expensive methods, and costly processing are the foundation of most machine learning algorithms. Deep learning algorithms are suggested by recent research as a means of developing depression detection methods [7]. Finding patterns of speech in the languages that the general public speaks depends on natural language processing methods like word representations [8]. The proposed study on depression diagnosis for social media will help improve depression identification by utilizing the proposed RoBERTa-BiLSTM architecture. Through embedding this approach, the most sophisticated contextual embeddings are combined with sequential analysis to provide a sophisticated view of the textual signs of depression. The value of the approach is in enhancing early identification and subsequent mental health care by means of reliable and centralized text analysis.

The following are the intended study's primary contributions:

- Emphasizes on the features of social media text through RoBERTa-BiLSTM in capturing both the contextual and sequential information for the determination of depression.
- Enables very fine-grained and efficient tokenization as well as embedding of the tokens, which are input for the deep learning models.
- Leverages on RoBERTa's rich contextual obligation to enhance the method of feature extraction of social media posts which helps to consider concealed sentiments and intricate linguistic structures.
- Utilizes BiLSTM in bidirectional analysis of text to capture all the dependencies and contextual relations in the use of social networks by users.
- Facilitates early identification of depression that may result in timely intervention and support for people displaying early signs of mental illness issues.

This is how the paper is organized. Studies connected to Section II are discussed. Section III provides information on the limitations of traditional models. Section IV contains the proposed mode of function. Section V discusses the findings and summary. Section VI has a conclusion and recommendations for more research.

II. Related Works

Lin et al., [9] states that over 300 million individuals worldwide have experienced depression. The majority of them aren't identified in their early stages because of medical supplies and knowledge gaps. Recent research has attempted to employ social media to identify depression since users' mental states can be somewhat reflected in the patterns of thoughts and opinions shown in the text and visuals they upload. In this study, researchers construct a system called SenseMood to show that the suggested system can effectively identify and assess those who are experiencing depression. An approach to revealing users' psychological states on social media platforms has been proposed: deep visual-textual multimodal learning. Images and tweets made by people who have and do not have depression on Twitter are collected and developed in order to recognize depression. Bert and a CNN-based classifier are used for extracting the deep features from user-posted text and images. Subsequently, the textual and graphic components are combined to depict the emotional responses of the customers. Ultimately, the technology uses a neural network to classify individuals as depressed or normal, and an automatic evaluation summary is generated.

According to Sardari et al., [10] depression is a widespread and significant psychiatric condition that has to be diagnosed and treated as soon as possible. Suicidal thoughts may arise from the disorder during severe episodes. The research community has recently become interested in developing an efficient audio-based Automatic Depression Detection system. To automatically determine the extremely essential and compact feature set, a DL techniques implementation is included in an audio-based depression detection system. With the goal to better identify sad individuals, this suggested framework learns particularly pertinent and distinct characteristics from unprocessed sequential audio data using a Convolutional Neural Network-based Autoencoder technique. Further employ a cluster-based sampling strategy to tackle the problem of sample imbalance, which meaningfully lowers the possibility of bias towards the dominant class (non-depressed). The results are compared with feature extraction techniques developed by hand and other notable works in the field. According to the results, the suggested method performs at least 7% better in the F-measure for depression classification compared to different recognized audio-based ADD models.

Zogan et al., [11] research on one major issue, particularly in the medical field, is the model's capacity to provide an explanation for the findings it produced. Since it provides illumination on the model's prediction, model explainability is crucial for fostering trust. It's concerning, though, that the majority of machine learning techniques already in use don't offer explainability. This study uses MDHAN, an explainable method for quickly identifying people who are depressed on social media. More specifically, they compute the importance of each tweet and word using two layers of attention mechanisms that are used at the tweet and word levels; they then obtain thematically sequencing data from user histories (posts) and encode user postings. A hierarchical attention strategy looks for patterns that lead to comprehensible outcomes. The trials demonstrate the advantage of merging multi-aspect data with deep learning, since MDHAN surpasses several well-known and reliable baseline approaches. The method increases the prediction accuracy of identifying melancholy in people who share publicly on social media, according to studies. MDHAN performs quite well and generates enough results.

Figueredo et al., [12] states that depression, which is frequently associated with illness and is one of the factors leading to suicide, poses a threat to public health. Consequently, they propose a preliminary detection of depression technique for social media using convolutional neural networks that blends Early and Late Fusion techniques with context-independent word embeddings. Considering the significance of the deeper feelings contained in the emoticons, these methods are experimentally assessed. The findings demonstrate that the suggested approach might identify users who might be depressed, with a precision of 0.76 and comparable or higher efficacy compared to numerous baselines (F1(0.71)). Furthermore, considerably improved outcomes were obtained through the semantic mapping of emoticons, with improvements of 46.3% and 32.1%, respectively, in recall and precision. The emoticon semantic mapping produced greater recall and precision gains (14.5%) and (48.8%) compared to the baseline word embedding approach. Overall effectiveness-wise, the work improved upon the state-of-the-art taking into account both the fusion-based and individual embeddings. Furthermore, it has been shown that the feelings that depressed individuals express and that are represented by emoticons are significant suggestive proof the issue and a useful tool for early detection.

Niu et al., [13] states that studies on the nervous system have revealed variations in facial expressions and speech patterns between healthy and depressed people. This fact leads us to propose a multimodal attention feature integration technique and an inventive spatio-temporal attention network to acquire the multimodal display of depression markers to predict each person's level of depression. First, fixed-length portions of the spoken word's amplitude spectrum/video are input into the STA network. As a result, the network might incorporate spatial and time-related data through an attention mechanism and highlight audio/video frames associated with depression recognition. The spatio-temporal attention network's last full connection layer produces the audio/video segment-level feature. Second, the Eigen evolution pooling technique is used in this article to build an audio/video level feature by integrating the modifications in each and every dimension of the segment-level features of audio and video. Third, a multimodal representation comprising modal different data is supplied to a support vector regression classifier, which is trained using the MAFF, in order to calculate the extent of depression. But have obstacles on high-quality data, which is computationally intensive, requires many integration challenges.

A novel approach was suggested by Rissola et al.,[14] to help compile a dataset of social media posts that mention depression or not. The author stressed that developing a model that can accurately identify depression is very challenging in the absence of a dataset. As a result, the author's dataset is capable of reliably predicting depression. The BERT model was used by the researchers to train their dataset, and the outcomes were excellent in terms of accuracy. This dataset can be used for more study, which can benefit mental health professionals as well. This novel approach to autonomously gathering large datasets will help present and future academics create instruments and applications that can precisely detect depression.

Chen et al., [15] discussed about how NLP methods had been applied to identify a certain kind of depression. Only a few numbers of research, meanwhile, have used sophisticated sentiment analysis methods to determine an individual's mental health from their social media postings. Chen and his colleagues employed a dynamic sentiment analysis method in this study, which is capable of extracting precise sentiments from people's tweets. Sad, pleased, disgusted, ashamed, surprised, afraid, confused, angry, and an ultimate score are the nine characteristics of an emotive. Thus, a tweet's dominating emotions are indicated by the emotive algorithm, which assigns a tweet's fine-grained emotion scores. Fine-grained emotions were defined as the feelings that people expressed when speaking or writing. These feelings were utilized as characteristics by machine learning algorithms to diagnose individuals who claimed they had mental health issues. Results were better for SVM and RF classifier.

Depression affects over three hundred million people globally, often going undiagnosed early because of gaps in existing sources and understanding. Recent studies have explored modern methods for detecting depression via social media, leveraging multimodal procedures. One gadget, SenseMood, combines visual and textual information using CNNs and BERT to assess emotional states. Another approach focuses on audio-based total detection, employing DL techniques and cluster-based sampling to improve classification accuracy. To address the assignment of version explainability, a hierarchical attention community complements the know-how of predictive outcomes. Convolutional neural networks with fusion methodologies and semantic mapping of emoticons have confirmed stepped-forward precision and recollect. Finally, a spatio-temporal interest network integrates audio and visual facts to predict despair severity, using multimodal representations for extra-accurate detection.

III. Problem Statement

Previous approaches to depression identification on Social Media content lack certain issues corresponding to contextual interpretation, two-way extraction, and integration of DL components. Most models face difficulties in interpreting the subtle signals of emotions and do not handle big data of different social networks [16]. The proposed work, using RoBERTa-BiLSTM models, fills the previous gaps in the following ways to achieve the targeted improvement. This approach improves the multiple layers of textual features and mood representation and increases the model's accuracy and applicability with the aid of hybridity. Further, it eliminates the shortcomings of the models being used today with the incorporation of deep learning approaches that are efficient in handling big data.

IV. Proposed Hybrid Deep Learning Models for Depression Identification

The RoBERTa-BiLSTM model encompasses deep contextualized meanings and sequential dependencies, surpassing conventional models such as SVM and RF, which are insensitive to subtle sentiment understanding. The model avoids current constraints to ensure greater precision in detecting depression. In the proposed research, a hybrid deep learning algorithm combining RoBERTa and BiLSTM is used to detect depression in social media contexts. Tweets and related metadata make up the data that is collected, and it is gathered from social media platforms, primarily Twitter. First of all, the text information is filtered out from the noise, and the format is brought to a single standard. For the feature extraction of the pre-processed text, RoBERTa is employed to obtain the contextual embeddings resulting from the specified complex semantic characteristics. These embeddings are then fed into a BiLSTM network, which is useful to get the sequential dependencies of the text and recreate it both forward as well as backwards directions. A dense layer receives the result from the BiLSTM and evaluates the conditions as being depressed or not. The model is trained using predefined labelled data and assessment measures. For the purpose of to illustrate the ability of the proposed approach to enhance the recognition of depression on social media sites and lay the basis for extensive data analysis on depression, the efficiency of the generated model is contrasted with baseline methods (Fig. 1).

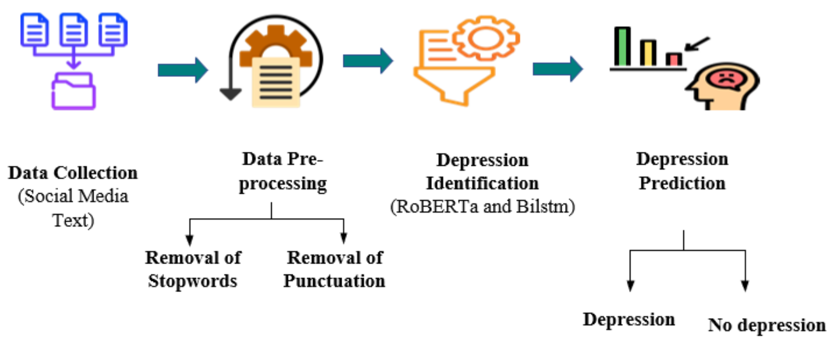

Fig. 1. Proposed hybrid deep learning models for depression identification.

A. Data Collection

The suggested method begins with acquiring data. This dataset, which is in Excel format, includes about 6500 data points from Facebook posts, comments, and other social media platforms. The ultimate depression of the data is represented by a large number of votes. The data collection was gathered from the Kaggle platform [17]. This dataset consists of two columns. There are two types: text and labels. The text columns contain both normal and anxiety/depression content, and the label column shows if the associated text suggests anxiety or depression are shown in Table I.

TABLE I. Data Outline

Text	Label
oh my gosh	1
trouble sleeping, confused mind, restless heart. All out of tune	1
like it, say don't just stay silent when someone takes it, eh, they say cheat	0
morning, have you taken a shower yet?	0
Unfortunately, I got another limit for 3 days	0

B. Data Pre-processing

Numerous unnecessary symbols and other components that complicate the framework of the model are included in the data that was gathered from the sources. Removing unnecessary information from the dataset makes the process of developing the model easier. This work involves a number of preprocessing processes. Texts with stop words, and punctuation that are deemed unnecessary for the analysis are removed.

1) Removal of stopwords: Stopwords are terms like "the," "a," and so on that are often used in a language. They can usually be removed from the text since they don't provide any pertinent information that has to be looked into further.

2) Removal of punctuation: Text preparation techniques also often involve removing punctuation from text data. 'Hurray' and 'hurray!' will be treated similarly because to this text standardization process.

C. Roberta- BiLSTM for Detection Recognition

The RoBERTa model extends BERT. The BERT and RoBERTa, part of the Transformers family, were created for sequence-to-sequence modeling in order to solve the issue of long-range dependencies. Transformers, tokenizers, and heads are the three parts of a transformer model is shown in Fig. 2. Through the tokenizer, sparse index descriptions are created from the unformatted text. The converters then convert the minimal content into meaningful embedding to enable more thorough training. For subsequent operations, the contextual embedding is utilized by wrapping the transformers model with the heads. When it comes to learning the contextual depiction from both sides of the sentences, BERT differs slightly from the language models that are already in use. However, RoBERTa combined a broader vocabulary collection with 50K subword units using byte-level Byte-Pair Encoding. Other than that, by training on larger amounts of data, more training durations, and extended sequences, the RoBERTa model improves upon the BERT model.

The cleaned text is initially tokenized using words and subwords in the suggested RoBERTa-BiLSTM model to facilitate quick encoding into word embeddings. RoBERTa tokenizer is utilized in this study. Some unique tokens are available in the RoBERTa tokenizer, including tokens to denote the start and finish of sentences and tokens to extend the text to the word vector's maximum length. The text in the RoBERTa model is divided into subwords using the level of bytes Tokenizer for Byte-Pair Encoding. There won't be a tokenizer for the commonly used terms. Rare words, yet, will be divided into easier-to-understand concepts. The term "Transformers" will be divided into two parts, namely "Transform" and "ers." The text must be converted from language spoken to an understanding for its structure to comprehend it of numbers. The raw text that the RoBERTa tokenizer encodes using input ids as well as an attention mask. Tokenization is shown in Table II.

TABLE II. TOKENIZATION

	text	label	text_clean	tokens	text_tokens
0	oh my gosh	1	oh my gosh	[oh, gosh]	oh gosh
1	Trouble sleeping, confused mind, restless heart. All out of tune	1	trouble sleeping confused mind restless heart all out of tune	[troubl, sleep, confus, mind, restless, heart, tune]	troubl sleep confus mind restless heart tune
2	All wrong, back off dear, forward doubt. Stay in a restless and restless place	1	all wrong back off dear forward doubt stay in a restless and restless place	[wrong, back, dear, forward, doubt, stay, restless, restless, place]	wrong back dear forward doubt stay restless restless place
3	I've shifted my focus to something else but I'm still worried	1	ive shifted my focus to something else but im still worried	[ive, shift, focu, someth, els, im, still, worri]	ive shift focu someth els im still worri

The input ids are a numerical representation of the token's indices. Conversely, the attention mask was a customization option that is applied when the sequence is assembled in batches. Which tokens should and shouldn't be examined is indicated by the attention mask. The RoBERTa fundamental model receives the input ids along with an attention mask. The RoBERTa basic model consists of 12 RoBERTa foundation layers, 768 concealed state vectors, and 125 million parameters. In order for the subsequent layers to quickly retrieve the important information about the word embedding, the RoBERTa first layer's attempt to provide a pertinent word inclusion as a distinctive depiction.

A more complex kind of RNN that is particularly good at identifying dependencies in sequential data are BiLSTM networks. BiLSTMs process sequences both forward and backward, in contrast to typical LSTMs, It only does one direction of sequence processing. The model might employ historical and prospective data contexts thanks to this bidirectional approach, which is especially helpful for jobs that call for a thorough comprehension of the sequence. An expansion of the LSTM architecture, BiLSTM networks are intended to get around some of the drawbacks of conventional RNNs, such the vanishing gradient issue. Long-range dependencies in the data are captured by LSTMs through the use of a series of gates to control the information flow across the network. The following are the essential parts of an LSTM cell:

- The Input Gate (i_t) regulates the amount of incoming data that is contributed to the cell state.
- The Forget Gate (f_t) regulates in what way much of the previous cell state should be preserved.
- The output gate (o_t) determines the appropriate amount of cell state output.

$$I_t = \sigma(E_i.[D_{t-1}, x_t] + b_i) \tag{1}$$

$$f_t = \sigma(E_f.[D_{t-1}, x_t] + b_f) \tag{2}$$

$$u_t = \sigma(E_o.[D_{t-1}, x_t] + b_o) \tag{3}$$

$$C_t = tanh(E_C.[D_{t-1}, x_t] + b_C) \tag{4}$$

$$C_t = f_t * C_{t-1} + i_t * C_t \tag{5}$$

$$d_t = o_t * \tanh(C_t) \tag{6}$$

where the hyperbolic tangent function in (1), (2), (3), (4), (5) and (6) is represented by $tanh$, the sigmoid function by σ, and element-wise multiplication by $*$. Because of these equations, LSTMs are better equipped to represent complicated temporal connections by maintaining a steady gradient across lengthy durations. Two LSTM layers exist in a BiLSTM network:

1) Forward LSTM layer: The layer moves ahead, processing the given input sequence from beginning to end.

2) Backward LSTM layer: The layer starts to process the input sequence at the beginning and proceeds backward.

The final of the dropout layer is then transferred into the BiLSTM model next. The LSTM network allows information to spread completely in a forward motion that denotes that the information prior to time t is the only source of influence for the state of time t. However, further details are just as useful as earlier ones in characterizing the overall semantics of an input review. Therefore, the BiLSTM model has been employed to represent contextual information more effectively. The two LSTM networks that compose up the BiLSTM model allow it to scan input reviews both forward and backward. The hidden state of the forward LSTM may be represented in Eq. (7).

$$\overrightarrow{hd_a} = LSTM(l_a, \overrightarrow{hd_{a-1}}) \tag{7}$$

Interpreting data from left to right, while the hidden state of the reverse LSTM can be represented in Eq. (8).

$$\overleftarrow{hd_a} = LSTM(l_a, \overleftarrow{hd_{a+1}}) \tag{8}$$

The concatenation of the forward and backward states yields $hd_a = [\overrightarrow{hd_a}, \overleftarrow{hd_a}]$, which is the final overview of the BiLSTM output. Global average pooling and global maximum pooling layers receive the BiLSTM layer's outputs concurrently. The BiLSTM layer's maximum and average values for each feature are retrieved by the former and latter layers, respectively. Instead of using the dense layer(s), only one global pooling layer (each) is used. Before sending it to the last layer, the concatenate layer combines both the global maximum and global average layers into a single layer.

Detecting depression in social media text analysis using a hybrid RoBERTa-BiLSTM version includes leveraging the strengths of each RoBERTa and BiLSTM to capture deep contextual and sequential statistics from the textual content. RoBERTa, a transformer-based totally version, excels in know-how context via its pre-learned on a full-size corpus of statistics, which allows it to generate rich, nuanced embeddings for every token in the textual content. The technique starts with preprocessing the social media texts, which includes casting off punctuation and stopwords to make sure cleanser records input. These texts are then tokenized using RoBERTa's tokenizer, changing them right into a layout that the version can method. RoBERTa strategies these tokens to provide contextual embeddings, shooting the meanings of phrases in the context of surrounding phrases. This is important in social media text evaluation where slang, abbreviations, and casual language are typical.

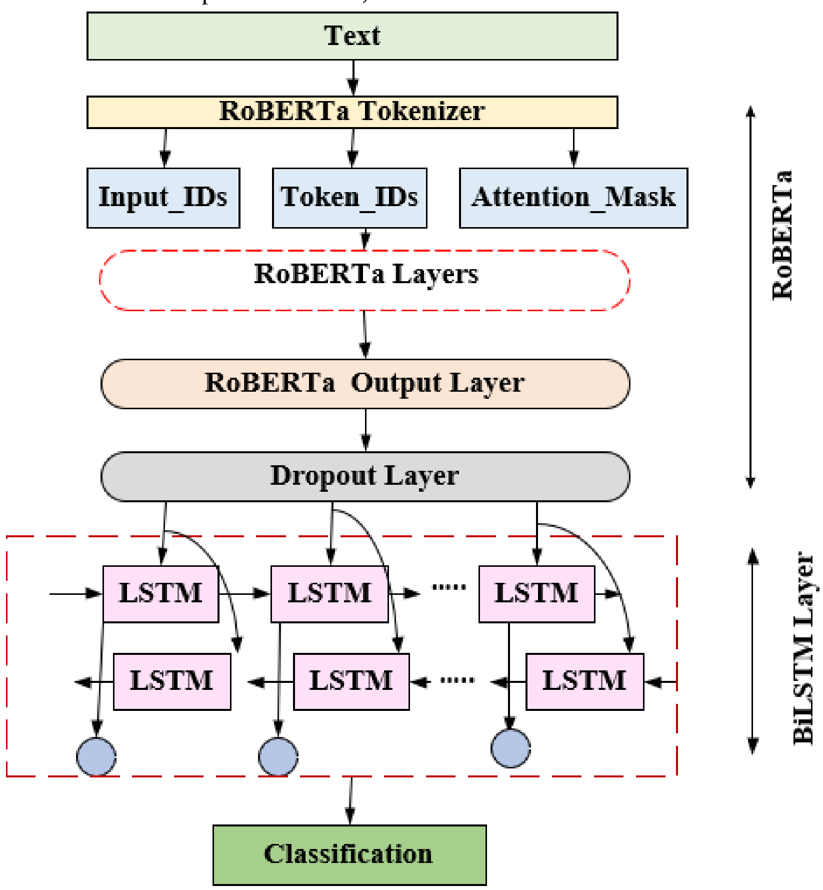

Fig. 2. RoBERTa and BiLSTM method.

The BiLSTM network receives the generated contextual embeddings directly. BiLSTM techniques the series of embeddings in both forward and backward directions, capturing dependencies that span across the entire text. This bidirectional processing is mainly beneficial for knowing the emotional tone and progression inside the textual content, which is important for identifying symptoms of depression. The hidden states generated via the BiLSTM encapsulate records from both beyond and future contexts, imparting a complete know-how of the collection. The model then applies an attention mechanism, which assigns specific weights to exceptional components of the textual content primarily based on their relevance to depression identification. The attention technique enables the version to concentrate on important textual elements which can be more indicative of despair, along with expressions of sadness, hopelessness, or tension. The weighted hidden states are combined to shape a context vector that represents the general emotional and contextual facts of the textual content. This context vector is then used to extract functions that can be fed into a chain of dense layers for category. The dense layers procedure these features to output a prediction indicating whether or not the text shows depressive tendencies. The model is educated on the usage of categorized facts, where the social media texts are annotated with labels indicating the presence or absence of depressive symptoms. During training, the model learns to optimize its parameters to accurately classify new, unseen texts. The pseudocode in Algorithm 1 and flowchart in Fig. 3 outline a technique for detecting depression in social media textual content using RoBERTa-BiLSTM, consisting of facts preprocessing, embedding generation, feature extraction, category, and final predictions.

Algorithm 1: Depression Detection Using RoBERTa-BiLSTM Hybrid Model

```
Input: Social media text samples
Output: Depression prediction for each text sample
Step 1: Data Preparation
Load text samples
Preprocess text:
   - Remove punctuation
   - Remove stopwords
Step 2: Tokenization and Embedding with RoBERTa
tokenized_texts = RoBERTa_tokenizer(preprocessed_texts)
embeddings = RoBERTa_model(tokenized_texts)
Step 3: Hidden State Calculation with BiLSTM
hidden_states = []
for emb in embeddings:
   forward_hidden_state, backward_hidden_state = BiLSTM (emb, previous_forward_hidden_state, previous_backward_hidden_state)
   hidden_states. append ((forward_hidden_state, backward_hidden_state))
Step 4: Feature Extraction
features = extract_features(context_vector)
Step 5: Classification using Dense Layers
depression_prediction = Dense_Layers(features)
Step 6: Training and Evaluation
Train the model using labelled depression data
Evaluate model performance using metrics (accuracy, F1 score, etc.)
Step 7: Prediction
depression_predictions= predict_depression(new_text_samples)
Output: depression_predictions
End
```

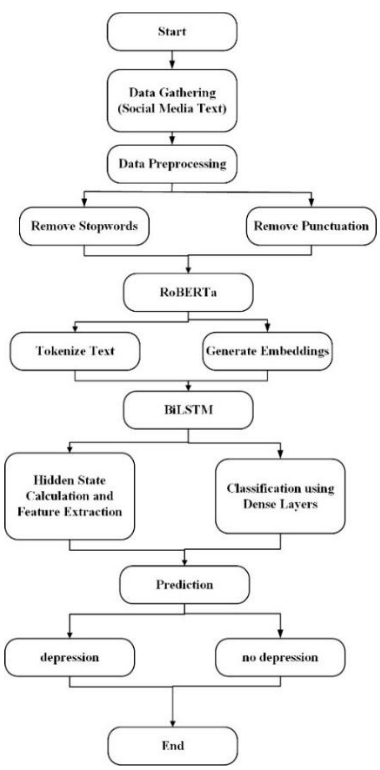

Fig. 3. Flowchart for proposed depression identification method.

V. Results and Discussion

The proposed study using the RoBERTa-BiLSTM hybrid model showed promising results in detecting depression in social media contexts. The model fully processed and analyzed transcript content, using the RoBERTa contextual input and the BiLSTM two-way sequential learning to classify transcripts as "depressed" or "no depressed." indicators Analytical parameters are appeared to perform better than baseline models. The combination of RoBERTa and BiLSTM led to a better understanding of sensory signals and context, resulting in stronger and more reliable predictions, and highlighted the real-world potential of the model in mental health services.

A. Experimental Outcomes

TABLE III. Depression Prediction

text	label	prediction
Hi, I want to tell you... Lately I've been feeling restless, have trouble sleeping, I searched on google it says it's a mild symptom of depression, I used to tell my mom a psychologist friend "don't think too much, it's not important you get depressed easily" then I frequent irregular breathing.	1	0.164200
Yes, the point is that I feel tired, sad, annoyed, restless. It's like the feeling is mixed in my heart and mind, in my brain I'm traveling various things from problems to happiness that I've felt until this moment.	1	0.164194
Every time after sunset, why must this heart be restless as if it can't accept the situation. But with this situation, you can't do anything, if you do it, it can only make things worse	1	0.164191
Anjir looks like he can't watch Dream, I have an appointment at 3 o'clock ”	0	0.095998
w if it doesn't look like it's okay. it's too late for the child w. it's better if it's hard to be alone, so let's be like before	0	0.095998
even though it's just a vanilla latte but why is it strong until the morning, after that it looks like a panda	0	0.095998

The Table III presents the results of a depression detection model using RoBERTa-BiLSTM on social media text samples. The "Text" column contains input text samples, "Label" indicates ground truth (1 for depression, 0 for depression), and

"Prediction" shows the model's confidence score for depression Higher scores (closer to 1) indicate greater confidence in detecting depression. The model demonstrates the ability to distinguish between depressive and no depressive texts, and obtain higher scores for depressive texts.

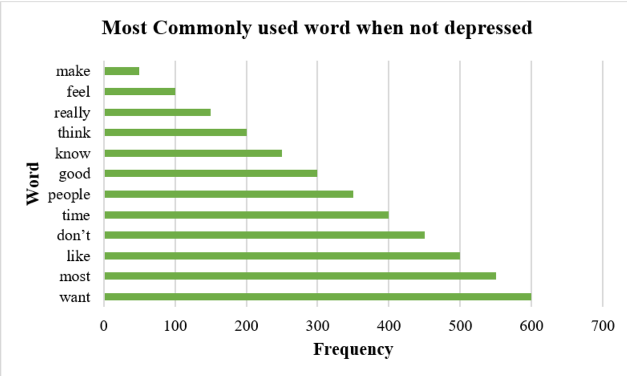

Fig. 4. Not depressed words.

The Fig. 4 indicates the frequency of different words humans use whilst they're no longer feeling depressed. The words are listed from maximum to least frequent: "make," "feel," "certainly," "assume," "know," "humans," "proper," "time," "like," "do not like most," and "want." The maximum frequency is just under seven hundred for "make," and the bottom is around 100 for "need." This figure highlights how regularly those words appear in conversations whilst individuals are not experiencing depression.

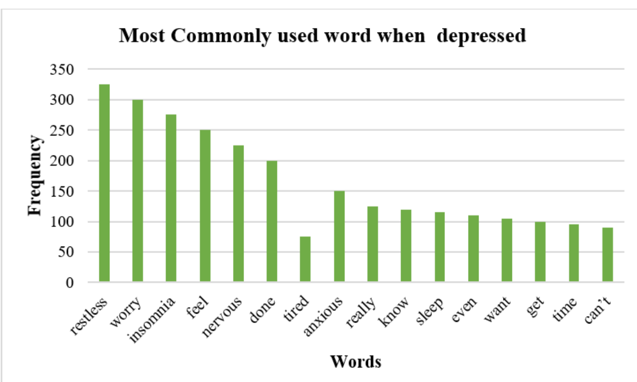

Fig. 5. Depressed words.

The Fig. 5 shows the frequency of particular words used by individuals when feeling depressed. The horizontal axis records terms such as "restlessness," "insomnia," and "fatigue," while the vertical axis measures frequency from 0 to 350 and "restlessness" corresponds to "insomnia," indicating that these terms are mentioned more frequently In the case of depression. This visual representation aids to understand common language patterns associated with depressive states, which can be useful for psychoanalytic or linguistic studies.

B. Training and Testing

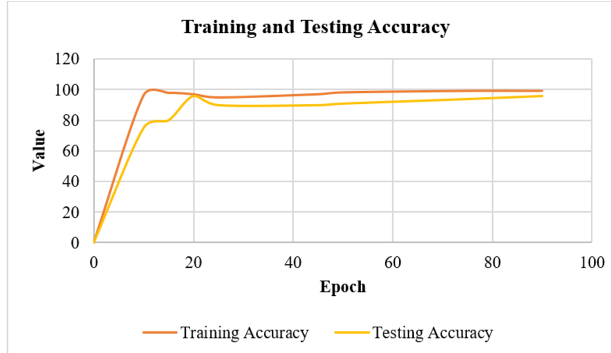

Fig. 6. Training and testing accuracy.

Training and test accuracy measures at different ages demonstrate the learning and generalization capabilities of the model is shown in Fig. 6. Initially, both accuracies start from 0. By time 10, the accuracy of training reaches 96%, the accuracy of testing reaches 75%, indicating initial overfitting and both accuracies improve as training goes on so, with training accuracy standing around 97-99% and testing accuracy of 96 at time 90 peaks at %. This indicates that the model learns the training data well and generalizes well to unseen data, thus confirming the robustness of RoBERTa-BiLSTM hybrid model emphasizes uneasiness found in social media text.

Fig. 7. Training and testing loss.

The training and testing loss show how the error in the model decreases with time is shown in Fig. 7. Initially, at epoch 5, the mean training loss was 2.5 and the mean testing loss was 2.8. By time 10, the loss decreases significantly 0.9 for training and 1.2 for testing, indicating an improvement in model learning. Despite a slight increase at time 20, the loss continues to decrease, with the training loss reaching 0.1 and the testing loss reaching 0.25 at time 60. This steady decrease in loss indicates the accuracy of the model improved and decreased errors, a finding indicative of both effective and general learning for depression identification.

C. Performance Evaluation

When compared to baseline models, the effectiveness evaluation of the suggested hybrid framework for depression detection produced consistent and outstanding results. The equation in (9), (10), (11) and (12) is used to find the F1-score, recall, accuracy, and precision. T_{pos} means true positive, T_{neg} means true negative, F_{pos} means false positive and F_{neg} means false negative.

$$Accuracy = \frac{T_{pos}+T_{neg}}{T_{pos}+T_{neg}+F_{pos}+F_{neg}} \tag{9}$$

$$Precision = \frac{Tpos}{Tpos+Fpos} \tag{10}$$

$$Recall = \frac{Tpos}{Tpos+Fneg} \tag{11}$$

$$F1\ Score = \frac{2\times precision\ \times recall}{precision\times recall} \tag{12}$$

TABLE IV. PERFORMANCE COMPARISON

Approach	Accuracy	Precision	Recall	F1-Score
CNN[9]	0.884	0.903	0.87	0.936
CNN AE+SVM [10]	0.71	0.72	0.71	0.71
MDHAN [11]	0.89	0.9	0.89	0.89
BERT [14]	0.87	0.73	0.96	0.81

The Table IV compares the performance of different models in predicting depression. The corresponding kernel support vector machine achieves an accuracy of 83.1% with precision, recall, and F1 scores of around 80-83%. The RNN model performs slightly lower in accuracy 80.5% but better in F1 score 83.8%. The CNN and BiLSTM models significantly improve the accuracy by 98.1% with higher accuracy, recall, and F1 scores. The proposed RoBERTa and BiLSTM models perform the best among them in terms of 99.4% accuracy, good precision 98.5%, recall 97.1%, and F1 score 97.3%, indicating good performance in depression as they are seen in it.

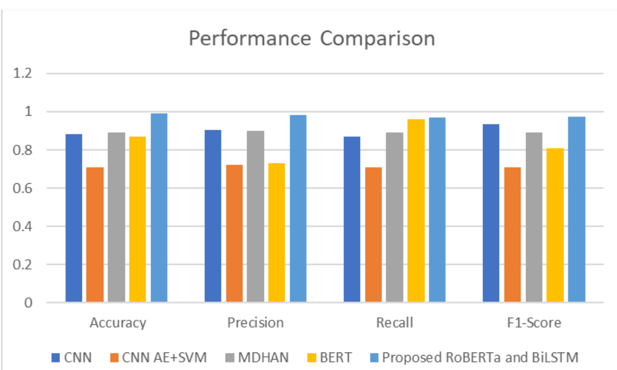

Fig. 8. Performance comparison of various methods.

The Fig. 8 compares four models: The performance comparison graph shows the comparative effectiveness of various models, viz., CNN, CNN AE+SVM, MDHAN, BERT, and the suggested RoBERTa-BiLSTM model, on critical evaluation metrics: accuracy, precision, recall, and F1-score. The suggested RoBERTa-BiLSTM model surpasses all the other models in all metrics, proving to be the best depression detection capability from social media text. Though CNN and CNN AE+SVM exhibit average performance, CNN AE+SVM falls behind because it has less contextual knowledge. MDHAN and BERT are good, but the integration of RoBERTa's contextual embeddings with BiLSTM's sequential learning improves recall and F1-score, and thus the proposed approach is the best.

D. Discussion

The RoBERTa-BiLSTM hybrid model significantly enhances the detection of depression in social media contexts by addressing various limitations of traditional methods. Traditional methods typically involve laborious and subjective manual assessments and clinical interviews. Conventional methods often suffer from limited accuracy and scalability issues [21]. This model uses advanced NLP and deep learning, automates the search process, and provides intuitive and objective solutions. The introduced RoBERTa-BiLSTM model shows considerable strengths compared to the conventional approach by efficiently extracting both contextual and sequential dependencies from social media text. It is superior to common machine learning models such as SVM, RF, and earlier deep learning models including CNN-LSTM. The inclusion of RoBERTa in context captures subtle emotional expressions, while BiLSTM better models' sequential speech dependence, improving sensitivity to depressive symptoms. The results show a notable accuracy of 99.4%, with high accuracy (98.5%), recall (97.1%) and F1 score (97.3%). This implementation demonstrates the model's resilience in differentiating between texts that are sad and those that are not, outperforming both alternative DL techniques and conventional machine learning models. Through guaranteeing accurate and suitable input, preliminary processes that involve the removal of pauses and character input help to develop the model. The accomplishment of this hybrid approach suggests that similar models can be incorporated into mental health care systems, enabling early detection and intervention Future research could extend the model to address or deliver multilingual content other methods such as visual or auditory data will be combined to further enhance the prediction capabilities.

VI. CONCLUSION AND FUTURE SCOPE

For healthcare departments to assist their depressed patients, it is imperative that depression be automatically identified from text. The RoBERTa-BiLSTM hybrid model leads to significant improvements in depression detection in social media contexts. This study achieved 99.4% accuracy, demonstrating the exceptional performance of the model RoBERTa's advanced contextual embeddings with serial capture capabilities dependencies of the BiLSTM Integrating, the model correctly recognizes subtle emotional cues that predict depressive states preliminary steps, including stopword removal and syntactic removal, provide the model's ability to process and analyze contextual information to ensure predictions are accurate and reasonable. The drawbacks of conventional techniques, which usually rely on biased manual analysis, are eliminated by this strategy. Through automating the process of discovery, this model provides a scalable, objective solution that can be incorporated into mental health care systems to facilitate the operation of the early intervention RoBERTa-BiLSTM hybrid model high quality underscores the potential for real-world application in mental health. Subsequent investigations ought to concentrate on several crucial domains to augment the dependability of the framework. Extending the model to handle multilingual contexts could improve its applicability to different language populations. Furthermore, the integration of multiple inputs, such as visual or auditory feedback, leads to a more comprehensive understanding of depressive symptoms, as emotional symptoms are often expressed through multiple mechanisms. The RoBERTa-BiLSTM model has difficulty with non-verbal signals, which restricts accuracy of depression detection. It is challenged in multilingual environments and potentially drops contextually significant words in preprocessing. The need for high computational requirements discourages real-time use, while risks of misclassifications involve sarcasm and doubtful text, necessitating further tuning for enhanced reliability and explainability. Future research encompasses multilingual adaptation, multimodal data integration (text, audio, images), real-time deployment optimization, simplification of computational complexity, improved explainability, fine-tuning sarcasm detection, contextual understanding improvement, and guaranteeing ethical AI deployment in mental health care.

REFERENCES

[1] "Depressive disorder (depression)." Accessed: Aug. 05, 2024. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/depression

[2] A. Thapar, O. Eyre, V. Patel, and D. Brent, "Depression in young people," The Lancet, vol. 400, no. 10352, pp. 617–631, 2022.

[3] M. A. Villarroel and E. P. Terlizzi, "Symptoms of depression among adults: United States, 2019," 2020.

[4] C. B. Nemeroff, "The state of our understanding of the pathophysiology and optimal treatment of depression: glass half full or half empty?," Am. J. Psychiatry, vol. 177, no. 8, pp. 671–685, 2020.

[5] E. A. Pataky and U. Ehlert, "Longitudinal assessment of symptoms of postpartum mood disorder in women with and without a history of depression," Arch. Womens Ment. Health, vol. 23, no. 3, pp. 391–399, 2020.

[6] J. M. Havigerová, J. Haviger, D. Kučera, and P. Hoffmannová, "Text-based detection of the risk of depression," Front. Psychol., vol. 10, p. 513, 2019.

[7] L. Squarcina, F. M. Villa, M. Nobile, E. Grisan, and P. Brambilla, "Deep learning for the prediction of treatment response in depression," J. Affect. Disord., vol. 281, pp. 618–622, 2021.

[8] T. Zhang, A. M. Schoene, S. Ji, and S. Ananiadou, "Natural language processing applied to mental illness detection: a narrative review," NPJ Digit. Med., vol. 5, no. 1, pp. 1–13, 2022.

[9] C. Lin et al., "Sensemood: depression detection on social media," in Proceedings of the 2020 international conference on multimedia retrieval, 2020, pp. 407–411.

[10] S. Sardari, B. Nakisa, M. N. Rastgoo, and P. Eklund, "Audio based depression detection using Convolutional Autoencoder," Expert Syst. Appl., vol. 189, p. 116076, 2022.

[11] H. Zogan, I. Razzak, X. Wang, S. Jameel, and G. Xu, "Explainable depression detection with multi-aspect features using a hybrid deep learning model on social media," World Wide Web, vol. 25, no. 1, pp. 281–304, 2022.

[12] J. S. L. Figuerêdo, A. L. L. Maia, and R. T. Calumby, "Early depression detection in social media based on deep learning and underlying emotions," Online Soc. Netw. Media, vol. 31, p. 100225, 2022.

[13] M. Niu, J. Tao, B. Liu, J. Huang, and Z. Lian, "Multimodal spatiotemporal representation for automatic depression level detection," IEEE Trans. Affect. Comput., vol. 14, no. 1, pp. 294–307, 2020.

[14] E. A. Ríssola, S. A. Bahrainian, and F. Crestani, "A dataset for research on depression in social media," in Proceedings of the 28th ACM conference on user modeling, adaptation and personalization, 2020, pp. 338–342.

[15] X. Chen, M. Sykora, T. Jackson, S. Elayan, and F. Munir, "Tweeting your mental health: An exploration of different classifiers and features with emotional signals in identifying mental health conditions," 2018.

[16] M. O. Nusrat, W. Shahzad, and S. A. Jamal, "Multi Class Depression Detection Through Tweets using Artificial Intelligence," ArXiv Prepr. ArXiv240413104, 2024.

[17] S. Saha, "Students anxiety and depression dataset." Accessed: Aug. 06, 2024. [Online]. Available: https://www.kaggle.com/datasets/sahasourav17/students-anxiety-and-depression-dataset

[18] C.-T. Wu, D. G. Dillon, H.-C. Hsu, S. Huang, E. Barrick, and Y.-H. Liu, "Depression detection using relative EEG power induced by emotionally positive images and a conformal kernel support vector machine," Appl. Sci., vol. 8, no. 8, p. 1244, 2018.

[19] A. H. Uddin, D. Bapery, and A. S. M. Arif, "Depression analysis of bangla social media data using gated recurrent neural network," in 2019 1st International conference on advances in science, engineering and robotics technology (ICASERT), IEEE, 2019, pp. 1–6.

[20] N. Marriwala, D. Chaudhary, and others, "A hybrid model for depression detection using deep learning," Meas. Sens., vol. 25, p. 100587, 2023.

[21] D. Liu, X. L. Feng, F. Ahmed, M. Shahid, J. Guo, and others, "Detecting and measuring depression on social media using a machine learning approach: systematic review," JMIR Ment. Health, vol. 9, no. 3, p. e27244, 2022.