Depression Classification using Bert Embedding Model on Social Media Posts

[1]Nareshkumar R and [2]Nimala K
[1,2]Department of Networking and Communications, School of Computing, College of Engineering and Technology, SRM Institute of Science and Technology, Kattankulathur, Chennai, Tamil Nadu, India.
[1]nr7061@srmist.edu.in, [2]nimalak@srmist.edu.in

Correspondence should be addressed to Nimala K : nimalak@srmist.edu.in

Article Info
Journal of Machine and Computing (https://anapub.co.ke/journals/jmc/jmc.html)
Doi : https://doi.org/10.53759/7669/jmc202505037
Received 18 September 2024; Revised from 30 October 2024; Accepted 18 November 2024.
Available online 05 January 2025.

Abstract – Developing sophisticated methods to precisely detect health-related concerns on social media, including identifying sadness and anxiety, has become essential due to the expansion of the Internet. These systems focus on using machine learning techniques to determine the meaning and structure of writings shared by users on social media. Social media users' data is confusing and inconsistent. Novel methods using deep learning and social networking platforms data to detect health issues. Provide just a little bit of information and understanding on the various texts individuals provide. This investigation introduces an innovative approach utilizing BERT to accurately and specifically detect posts related to sadness and anxiety. This approach preserves the contextual and semantic importance of words across the collection. The researcher employed word2vec, fasttext, BERT and Enhanced Grey Wolf Optimizer (E-GWO) and Deep Learning technologies to promptly analyze and detect indications of anxiety and melancholy in social media messages. Our solution surpasses previous advanced methods and incorporates the knowledge distillation methodology to achieve an accuracy of 95.9%.

Keywords – Depression, Social Media Posts, Natural Language Processing, LSTM, Deep Learning, RNN, BERT.

I. INTRODUCTION

Currently, social media serves as the primary public opinion analyzer. Over 4.2 billion people are actively using social networking sites throughout the globe. People have acquired a preference for expressing their views and approaches over the Internet, which has led to a growth in user-generated material and self-opinionated knowledge. One of the world's most significant and challenging health issues is the computerized diagnosis of mental health disorders. The state of an individual's mental health in terms of their behavior, thoughts, and emotions has an impact on the way that individual interacts with their surroundings [1]. The clinical experts evaluate the social media posts made by their patients as well as their responses to make a diagnosis regarding their psychological behavior and their mental health status [2]. If the condition is correctly identified, it can be treated pharmacologically and psychologically. However, there are still a limited number of ways to diagnose mental illnesses and a limited number of available treatments. There is a significant correlation between an individual's mental state and language. Lin guistic explanatory variables [3], such as an increased frequency of words that depict more extraordinary negative emotions and references to hopelessness and sadness, are associated with depression [4]. Classifiers may be trained using ML models using a wide variety of characteristics, including the ones shown here [5]. In order to analyze data more effectively, this research incorporates data from many distinct sources that pertain to mental health issues. The outcomes show that our provided technique successfully deals with ambiguous data and improves mental health categorization efficiency. In general, we have made four major contributions to this study:

- In this article, a unique new approach is presented in order to extract from Twitter a large quantity of extremely relevant data linked to mental health conditions such as depression and anxiety. Researchers utilized an approach that combined the circumplex hierarchy of sentiment addressing the necessary information on mental health difficulties.
- Text data about mental health issues can be analyzed using an Enhanced grey wolf optimizer and DL-based BERT model. This model ensures that contextual and semantic implications are maintained. We proposed using a Bi-LSTM sequence computation model as a classifier.

The results of our experiments indicate that our model performs far better than the approaches that we tested, which, after a number of hyper parameters tweaks, achieves an accuracy of 97%. The following is the order in which the remaining parts are presented: The second section provides an explanation of a monitoring system for mental health that makes use of wearable technology and a method based on DL. In Section 3, we will explain the overall conceptual framework that has been presented for this study. In the next section, we will discuss the findings that were acquired from our experiment. A summary of our findings is presented in the last section of this research, which is titled Section 5.

II. RELATED WORK

The use of social media is one of the platforms that has the potential to contribute to the discovery of, and subsequently the proposal of, approaches to diagnose serious depressive illnesses [6]. The research project aims to analyze different models utilizing deep learning and evaluate their performance measures against a baseline model that utilizes TF-IDF vectorization and an SVM linear classifier. In contrast to DL models, shallow models need to have their features engineered before the data can be fitted into the model. Nevertheless, in deep learning, this process is included into the model by learning a series of nonlinear transformations that help to translate features directly to outputs. This allows for more accurate predictions to be made. Using architectures that are based not just on CNNs but also on RNNs seems to be helpful due to this additional benefit. We will use CNNs with a single dimension, several convolutions with filters of varying lengths, and max-pooling layers in our analysis [7]. Due to the fact that RNNs can maintain their 'memory' over a large range of different time periods, they are used extensively in NLP. In addition to this, they enable processing of various lengths while still preserving the sequence order. The model known as LSTM is a popular kind of RNN. These taught modes are going to be put to use for assessment on data that has not yet been seen. These designs will be tested alongside a variety of word embedding that are routinely utilized, such as skip-gram and CBOW, among others [8]. We are interested in broadening the utilization of sentiment and analysis of feelings to determine an individual's stress level by analyzing the posts and comments they have shared on social networking sites. We do sentiment analysis by using large-scale datasets that include tweets, in conjunction with ML techniques and a DL model called BERT, which is used for sentiment classification. We wish to expand the usage of sentiment and emotion analysis so that we can determine an individual's degree of stress by analyzing the posts and comments that person has made on social networking sites. Our goal is to do this by using the data that people have shared on these sites. We do sentiment analysis by making use of large-scale datasets that include tweets, in combination with ML methods and a DL model known as BERT, which is used for the purpose of sentiment categorization [9]. In 2015, 24.6 million years of life were lost due to anxiety disorders over the world. Likewise, rates range throughout WHO Regions, ranging from 267 YLD per 100 000 people in Africa to over 500 YLD per 100 000 people in the Americas [10].

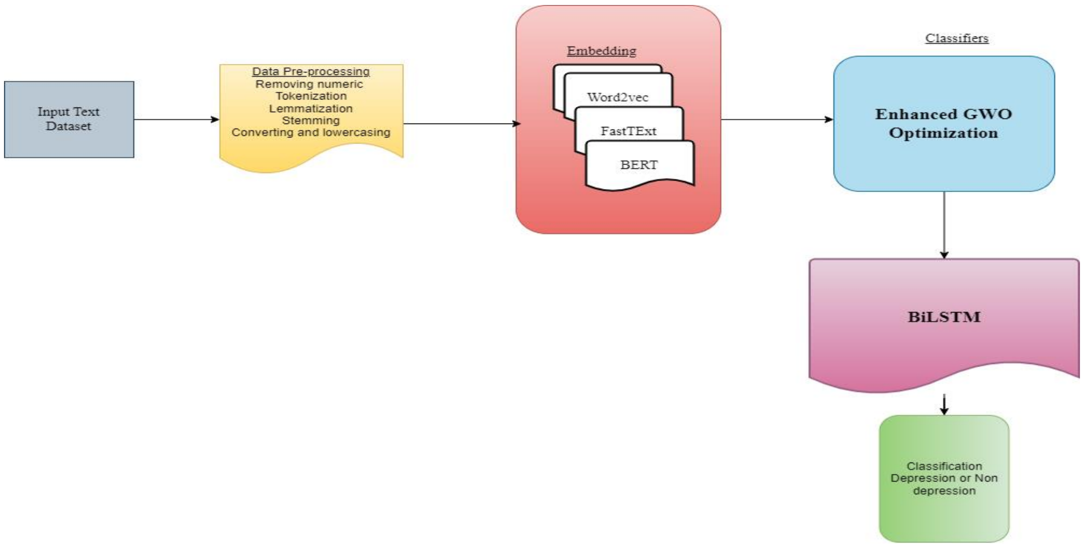

Fig1. Flow Of The BERT+E-GWO Bilstm Proposal Model.

III. METHODOLOGY

This Methodology section provide a concise workflow strategy such as data collection, pre-processing embedding techniques for data collected to validate the prediction of depression using DL. **Fig 1** is an illustration of the structure of the BERT+E-GWO BiLSTM pre trained em bedding model. The imbalance often creates biased findings, which were corrected by employing the objective function in order to cut down on the number of false negatives created, followed by their interpretation, as well as the experimental inferences that may be derived using classification algorithms. In final analysis, a number of different criteria were used to assess the level of success achieved by each classification model. The workflow of concept, the research was segmented into smaller parts, and those parts were worked on many times while keeping an eye out for any potential biases that may have been created throughout the process. This allowed for the creation of a complete categorization model. The inclusion of the bert pre-training language model is the primary distinction between this approach and the conventional text classification model that has been used in the past. The bert pre-training language model obtains its knowledge from a large-scale corpus, and it is able to represent the word via the vector representation of the word.

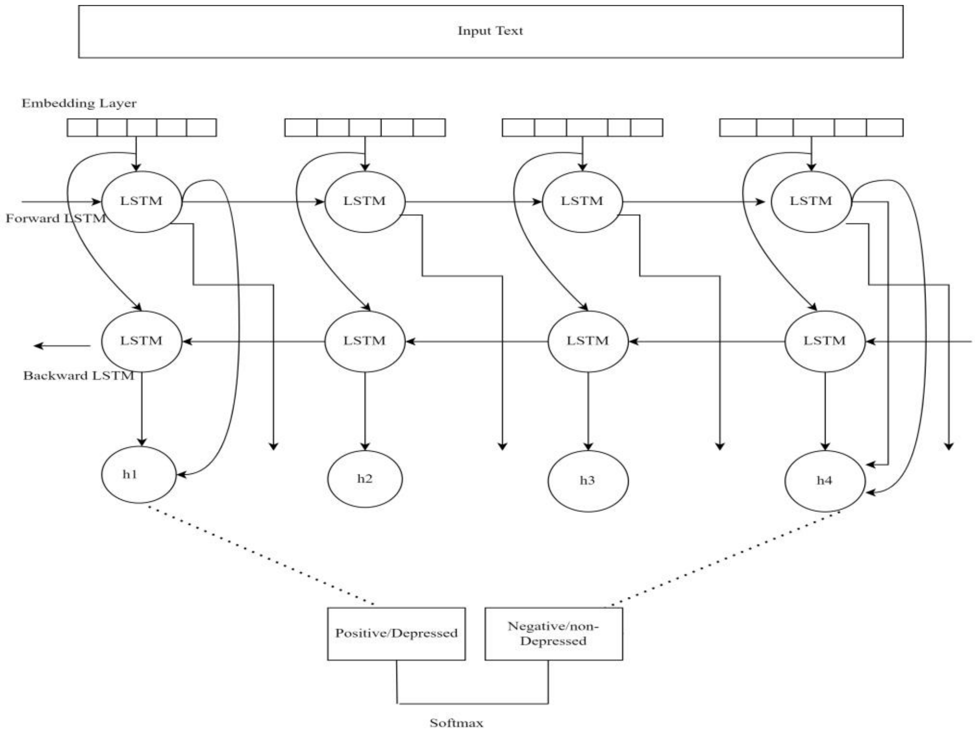

Fig 2. Structure of the BiLSTM Attention Model.

Data Collection

The technique for collecting data from two distinct sources is broken down and discussed in this section. One can observe in **Fig 2** of task 1 the complete process of data collection that is being utilized within the proposed framework. The process for selecting relevant content from Twitter involves identifying appropriate words preceded by a hashtag symbol, which serves as an indicator of the main theme associated with specific topics.

Preprocessing

Data preprocessing refers to the techniques used to clean and filter out noisy and ambiguous data, making it more suitable for feature extraction and other downstream tasks. Social media messages may be mined for context and emotional intensity by extracting significant language [11]. To identify depression and anxiety, we applied many preprocessing techniques to our gathered material, which had numerous technical terms and informal language. These methods were intended to enhance our ability to identify relevant patterns and features within the text. Tokenization is a process used to split a piece of text into smaller units, such as words or subwords, and to remove non-alphanumeric characters. Once the text is tokenized, it can be represented as a bag of words, which is essentially a frequency count of each word in the text. This type of representation is commonly used for further analysis in natural language processing and ML applications. Next, explain the procedure for eliminating terms that do not offer substantial information concerning mental health concerns in a broad context. Common sorts of insignificant phrases include pronouns, prepositions, indicators (such dates and hashtags), conjunctions, and articles (such as ”a,” ”an,” and ”the”). Each word is first converted to its basic form and then changed to lowercase in our situation. This approach aids in preserving uniformity and preventing ambiguity in text analysis. Part-of-speech tagging involves categorizing words according to their grammatical function in a phrase. The main purpose of this phase is to remove any bits of speech that do not aid in identifying sadness and anxiety in the text [12].

Word2vec and FastText Mode

Using a text corpus as input, the word2vec method produces a vector representation of the supplied words [13]. The skip-gram model is a neural network design utilized to acquire dispersed word il lustrations in a high-dimensional vector space. The algorithm aims to forecast the surrounding words when provided with a certain target term. The mathematical equation for the skip-gram model is as follows: Let V be the vocabulary of size |V|, and let wi and wj be two words in the vocabulary. Let c represent the dimensions of the context window, which specifies the quantity of words that belong to the context on either side of the word being targeted that serve as input to the model. Let D represent the dimensionality of the word embeddings, which indicates the number of dimensions in the vector space. The skip-gram model aims to optimize the average log likelihood of the surrounding words given a specific word in a vast body of text. This can be formulated as the following objective function:

$$J = \frac{1}{T}\sum_{t=1}^{T} \sum_{-c\leq j\leq c, j\neq 0} \log\, p\left(w_{t+j} \mid w_t\right) \tag{1}$$

where T is the total number of target words in the corpus, and $(w_{t+j} \mid w_t)$ is the conditional probability of the j-th context word given the target word. The skipgram model defines the conditional probability $(w_{t+j} \mid w_t)$ using the softmax function, which normalizes the dot product of the target word embedding with each context word embedding in the vocabulary. Specifically, we have:

$$p(w_{t+j} \mid w_t) = \frac{\exp(u_j^T v_t)}{\sum_{i=1}^{|V|} \exp(u_i^T v_t)} \tag{2}$$

The conditional probability $(w_{t+j} \mid w_t)$ is defined as follows:

where v_t is the target word embedding, u_i^T is the j-th context word embedding, and the sum in the denominator is over all words in the vocabulary. The skip-gram model underwent training using random gradient descent to minimize the unfavourable log likelihood of achieving the desired function J. The word embeddings are iteratively adjusted using the gradients of the loss function to the model parameters. Following training, the word embedding may be applied to several subsequent natural language processing activities, including sentiment analysis, text categorization, and machine translation. Fast text relies on analyzing the morphological organization of words in text to extract crucial information about their meanings. Fast Text aims to tackle this problem by treating words as though they were made up of smaller words.

Bert

BERT (Bidirectional Encoder Representations from Transformers) is a pre-trained Embedding DL model for natural language processing (NLP) tasks. It was introduced by Google researchers in 2018 and has since become a widely used model for various NLP tasks such as text classification, named entity recognition, and question-answering. Just the encoder mechanism is required for BERT since its primary goal is to generate a language model [14]. In order to learn the context of words in both directions, BERT employs two different methods during training [15].

Classifier Models

As RNN gives more weight to terms that have been lately in a sequence, it may be less effective in capturing the gist of an entire text, making it an inclined model. For sequential data, the LSTM network is a popular recurrent deep neural network design for classification, processing, and prediction [16]. I've detailed a typical LSTM in **Fig 3**. Not all LSTMs are like those above. Nearly every LSTM research employs a somewhat different variation. [17]. Even if the differences are small, they require noting. The forget gate ft, input gate it, and output gate ot are defined as follows:

$$f_t = \sigma(W_f \cdot [C_{t-1}, h_{t-1}, x_t] + b_f) \tag{3}$$

$$i_t = \sigma(W_i \cdot [C_{t-1}, h_{t-1}, x_t] + b_i) \tag{4}$$

$$o_t = \sigma(W_o \cdot [C_t, h_{t-1}, x_t] + b_o) \tag{5}$$

Here, σ represents the sigmoid activation function, and W_f, W_i, W_o are weight matrices, b_f , b_i, b_o are bias vectors. The input to each gate is a concatenation of the cell state C_{t-1}, hidden state ht−1, and input xt. Here, ft stands for point-wise multiplication with C_{t-1} to forget context vector information, and it C˜t stands for new candidate values scaled by the amount of needed updating. This phase entails using point-wise multiplication and addition, respectively, to change the incoming cell state $C_{t-1}$to reflect the choices made by the input gate and forget layers. We make such selections together rather than independently determining what to leave out and what we should update. Only when we're about to input something in its place do we forget. Only when we forget an older value do we add new values to the state.

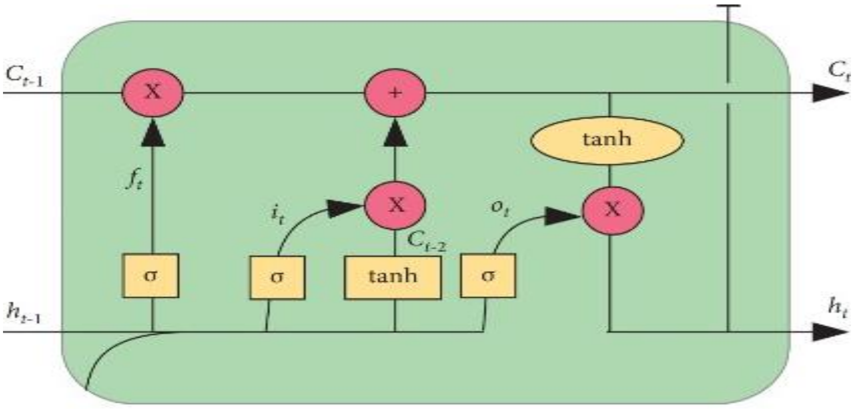

Fig 3. Structure of the BiLSTM Attention Model.

Hyper parameter GWO

The grey wolf optimizer (GWO) approach incorporates four sorts of wolves, drawing inspiration from the hunting behavior and social leadership of gray wolves in nature. The alpha, beta, and delta wolves hold the highest ranks, while the omega wolveslike surrounding, hunting, and attacking the target. Encircling: This step involves surrounding the prey with gray boxes and is represented by Equations (6).

$$\begin{gathered} D = \left| C \times X_p(t) - X(t) \right| \\ X(t+1) = X_p(t) - A \times D \end{gathered} \tag{6}$$

Xp is the prey’s status, D is the amount of distance between the present group of the applicant wolves and X is the position vectorthen and A and C are manipulated from (7) and (8). is:

$$A = 2 \times a(t) \times r_1 - a(t) \tag{7}$$

$$C = 2 \times r_2 \tag{8}$$

The values of C1,C2, and C3 from (9).

$$\begin{gathered} X_{i1} = X_\alpha(t) - A_{i1} \times D_\alpha(t) \\ X_{i2} = X_\beta(t) - A_{i2} \times D_\beta(t) \\ X_{i3} = X_\delta(t) - A_{i3} \times D_\delta(t) \end{gathered} \tag{9}$$

The three optimal candidate solutions, $X\alpha$,$X\beta$, and $X\delta$, from (10).

$$X(t+1) = \frac{X_{i1}(t) + X_{i2}(t) + X_{i3}(t)}{3} \tag{10}$$

The act of hunting wolves concludes when the prey stops moving. The attacking procedure commences at this juncture. The process is formally defined as a reduction in the value of ’a’ from 2 to 0 over the iterations. During the assault stage, the wolves relocate to a position between the prey’s location and their present position.

The method begins by randomly placing the wolf population in the search region and calculating their fitness levels based on their positions. The method stops when it reaches the specified MaxIter value. During each repetition, the surrounding, hunting, and assault processes are repeated. The answer is represented by” a,” which indicates the optimal position of the prey. The E-GWO method was developed to address issues such as premature convergence, exploration-exploitation imbalance, and population diversity in the traditional GWO algorithm.

IV. RESULTS AND DISCUSSION

This section discusses the outcomes of our proposed scheme. In this research, we employ ML and pre-trained DL models to classify depression and anxiety.

Dataset

Researchers created a system to gather data on depression and anxiety from Twitter using the Twitter APIs. They studied several methods for recognizing emotions in text and applied the circumplex approach to the study of affect by using phrases linked to different emotions . In order to characterize a group of emotions, it is usual practice to compile a list of terms that best capture the essence of those feelings [17] [18]. Here included a total of 35,000 tweets in our dataset in **Table 1**. Using an unbalanced data collection, in which data from one class are disproportionately more numerous (or less) than data from other classes, is a typical source of error during data processing. Because of this, the learning algorithm may choose to disregard the under-representation of some groups. So, we made an effort to strike a balance between the amount of upbeat and downbeat courses (52:48).

Table 1. Data Sources And Collected Posts

Data Source	Mental Health Problem	Collected Posts	Description
Twitter	Depression	25,000	Related to Content-based depression
Twitter	Anxiety	10,000	Related to Content-based anxiety

Evaluate Matrices

To evaluate our model, we made use of criteria that are quite standard [19]. A confusion matrix is a kind of performance evaluation matrix that is also known as an error matrix since it provides a tabular comparison of the number of inaccurate guesses to the number of right predictions [20]. Based on the confusion matrix, the following equations (11)-(14) allow us to determine the accuracy, precision(p), and recall(r):

$$\text{Accuracy} = \frac{\mathrm{TP+TN}}{\mathrm{TP+TN+FP+FN}} \tag{11}$$

$$\text{Precision} = \frac{TP}{TP+FP} \tag{12}$$

$$\text{Recall} = \frac{\mathrm{TP}}{\mathrm{TP+FN}} \tag{13}$$

$$\text{F1-Score} - 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \textit{Recall}} \tag{14}$$

The following is a list of the terms that are used in the calculation of the confusion matrix, in order of their frequency of use: P is a positive value, which is depression related. N is a negative value which is non-depression. TP is a true value and the gathered data is true, and our proposed method prediction is depression. TN is true negative which the gathered data is false, and the classification is also false. FP is False positive which the gathered data is false, and the method classification is true. FN is False negative which the gathered data is true, and method prediction is false. **Fig 4** shows the validation accuracy and training loss and **Fig 5** shows the ROC curve for proposed and comparison models.

Comparison of Our Suggested Method with Current State-of-the- Art Algorithms.

We assessed our proposed algorithms with several algorithms such as AdaBoost, RF, SVM, LSTM, CNN, MLP and LG and to identify emotions linked to depression and anxious states during our inquiry in **Table 2**. We made use of a circular basis function that had 150 iterations, 100 estimators, and a support vector machine that was outfitted with a training parameter ridge estimation and, appropriately, a radial basis functional [21] [22]. TF-IDF, word2vec, and fastText among popular word embedding algorithms [24] that we employ in baseline algorithms [23].

AdaBoost's accuracy rose to 83 percentage after adjusting hyperparameters to 50 estimators and 0.8 learning. **Fig 6** also shows the comparison and our proposed model performance.

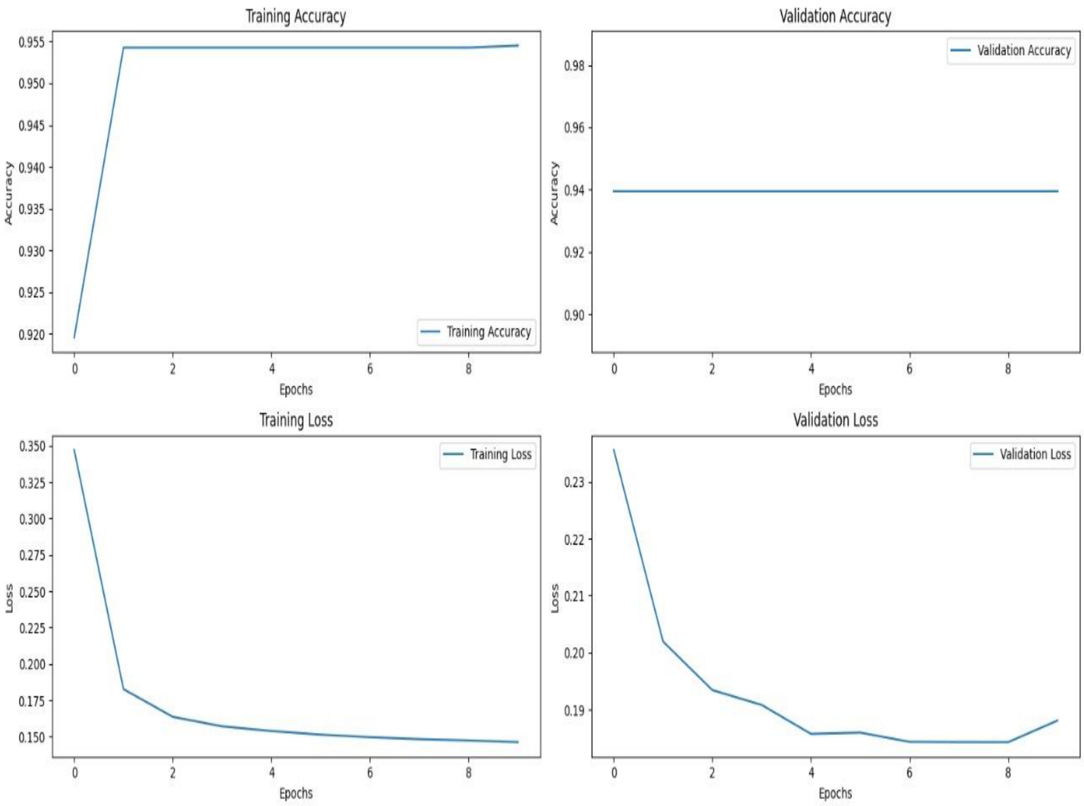

Fig 4. The Relationship Of Validation Accuracy And Epochs (a) Training Accuracy and (b) Training Loss.

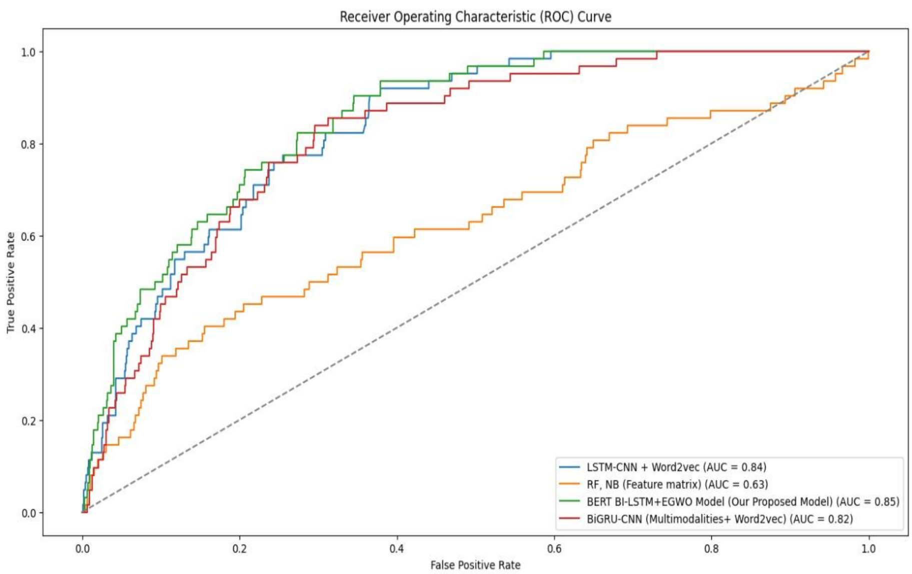

Fig 5. Structure of the BiLSTM Attention Model.

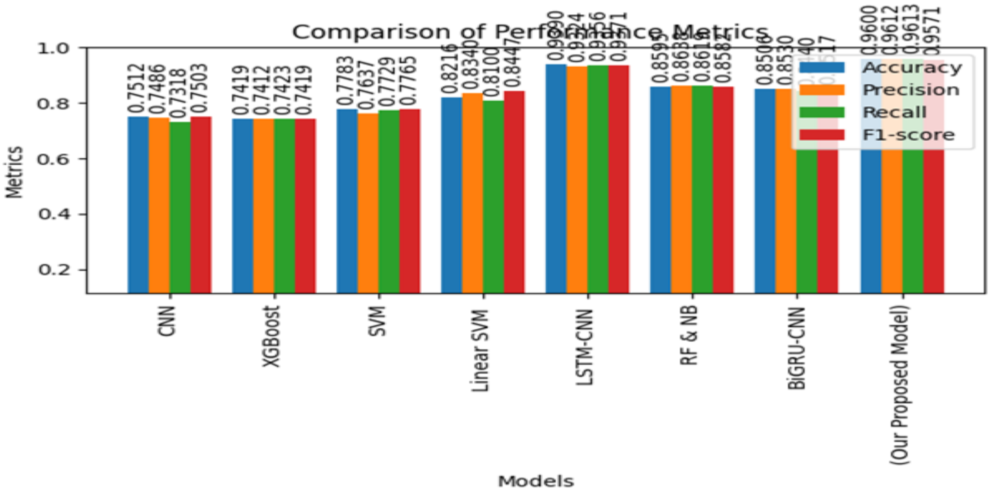

Fig 6. Our Proposed Model Result And Comparsion.

Table 2. Performance Results Of Our Proposed Approach

Methods	*Accuracy*	*Precision*	*Precision*	*F1*
CNN Word2vec	0.751	0.748	0.739	0.750
XGBoost	0.741	0.741	0.742	0.741
SVM (LIWC + LDA)	0.778	0.763	0.772	0.776
CNN + Linear SVM (LIWC)	0.825	0.833	0.810	0.844
LSTM-CNN + Word2vec	0.908	0.903	0.905	0.907
RF, NB (Feature matrix)	0.859	0.863	0.861	0.858
BiGRU- CNN (Multi- modalities+ Word2vec)	0.850	0.852	0.844	0.851
BERT+E- GWO BI- LSTM (Our Proposed Model)	0.959	0.961	0.946	0.957

V. CONCLUSION AND FUTURE WORK

Deep learning techniques are utilized to create a well-structured framework for diagnosing health issues based on user-provided data. This work seeks to use a text tagging approach based on keywords and a depression model to analyze a gathered text corpus and discover important aspects linked to mental health. Our approach to identifying mental health issues utilizes cutting-edge text embed- ding technology driven by deep learning. This guarantees that the exact semantic and contextual significance of terms utilized in user postings may be accurately documented. The model's exceptional performance is due to its integration of the syntactic and contextual comprehension abilities of the pre-trained embedding BERT, E-GWO and Bi-LSTM models, which are widely recognized ML classification methods. Additionally, we utilized BERT embedding to create a response-based knowledge transfer, achieving exceptionally high accuracy in identifying depressive and anxious states and further refining the job.

In the future, expanding the suggested model to incorporate multi-modal data sources or user interactions might offer a comprehensive insight into depression on social media. This allows for thorough and precise detection and categorization.

CRediT Author Statement

The authors confirm contribution to the paper as follows:
Conceptualization: Nareshkumar R and Nimala K; **Methodology:** Nareshkumar R; **Writing- Original Draft Preparation:** Nareshkumar R; **Visualization:** Nimala K; **Investigation:** Nimala K; **Validation:** Nareshkumar R and Nimala K; All authors reviewed the results and approved the final version of the manuscript.

Data Availability

No data was used to support this study.

Conflicts of Interests

The author(s) declare(s) that they have no conflicts of interest.

Funding

No funding agency is associated with this research.

Competing Interests

There are no competing interests

References

[1]. C. Su, Z. Xu, J. Pathak, and F. Wang, "Deep learning in mental health outcome research: a scoping review," Translational Psychiatry, vol. 10, no. 1, Apr. 2020, doi: 10.1038/s41398-020-0780-3.
[2]. S. R. M, A. R. L, and A. A, "Detection of Depression among Social Media Users with Machine Learning," Webology, vol. 19, no. 1, pp. 250–257, Jan. 2022, doi: 10.14704/web/v19i1/web19019.
[3]. Kulkarni, Hrishikesh, Macavaney, Sean, Goharian, Nazli, Frieder, "Ophir. Tbd3: A Thresholding-Based Dynamic Depression Detection from Social Media for Low-Resource Users," Proceedings of the Thirteenth Language Resources and Evaluation Conference, PP. 2157-2165, (2022).
[4]. X. Wang, Y. Sui, K. Zheng, Y. Shi, and S. Cao, "Personality Classification of Social Users Based on Feature Fusion," Sensors, vol. 21, no. 20, p. 6758, Oct. 2021, doi: 10.3390/s21206758.
[5]. T. Zhang, K. Yang, S. Ji, and S. Ananiadou, "Emotion fusion for mental illness detection from social media: A survey," Information Fusion, vol. 92, pp. 231–246, Apr. 2023, doi: 10.1016/j.inffus.2022.11.031.
[6]. H. Zogan, I. Razzak, S. Jameel, and G. Xu, "DepressionNet: Learning Multi-modalities with User Post Summarization for Depression Detection on Social Media," Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 133–142, Jul. 2021, doi: 10.1145/3404835.3462938.
[7]. N. V. Babu and E. G. M. Kanaga, "Sentiment Analysis in Social Media Data for Depression Detection Using Artificial Intelligence: A Review," SN Computer Science, vol. 3, no. 1, Nov. 2021, doi: 10.1007/s42979-021-00958-1.
[8]. J. A. García-Díaz, S. M. Jiménez-Zafra, M. A. García-Cumbreras, and R. Valencia-García, "Evaluating feature combination strategies for hate-speech detection in Spanish using linguistic features and transformers," Complex & Intelligent Systems, vol. 9, no. 3, pp. 2893–2914, Feb. 2022, doi: 10.1007/s40747-022-00693-x.
[9]. T. Nijhawan, G. Attigeri, and T. Ananthakrishna, "Stress detection using natural language processing and machine learning over social interactions," Journal of Big Data, vol. 9, no. 1, Mar. 2022, doi: 10.1186/s40537-022-00575-6.
[10]. K. Jaeschke, F. Hanna, S. Ali, N. Chowdhary, T. Dua, and F. Charlson, "Global estimates of service coverage for severe mental disorders: findings from the WHO Mental Health Atlas 2017 – Addendum," Global Mental Health, vol. 8, 2021, doi: 10.1017/gmh.2021.30.
[11]. B. Guo, C. Zhang, J. Liu, and X. Ma, "Improving text classification with weighted word embeddings via a multi-channel TextCNN model," Neurocomputing, vol. 363, pp. 366–374, Oct. 2019, doi: 10.1016/j.neucom.2019.07.052.
[12]. T. Zhang, A. M. Schoene, S. Ji, and S. Ananiadou, "Natural language processing applied to mental illness detection: a narrative review," npj Digital Medicine, vol. 5, no. 1, Apr. 2022, doi: 10.1038/s41746-022-00589-7.
[13]. Sanh, Victor, Debut, Lysandre, Chaumond, Julien, Wolf, Thomas, "DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter," (2019), arXiv preprint arXiv:1910.01108.
[14]. Un Nisa, Qamar, Muhammad, Rafi, "Towards transfer learning using BERT for early detection of self-harm of social media users, Proceedings of the Working Notes of CLEF," pp. 21-24, (2021).
[15]. 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, Jan. 2022, doi: 10.1007/s11280-021-00992-2.

[16]. T. Chen, R. Xu, Y. He, and X. Wang, "Improving sentiment analysis via sentence type classification using BiLSTM-CRF and CNN," Expert Systems with Applications, vol. 72, pp. 221–230, Apr. 2017, doi: 10.1016/j.eswa.2016.10.065.
[17]. H. Metzler, H. Baginski, T. Niederkrotenthaler, and D. Garcia, "Detecting Potentially Harmful and Protective Suicide-Related Content on Twitter: Machine Learning Approach," Journal of Medical Internet Research, vol. 24, no. 8, p. e34705, Aug. 2022, doi: 10.2196/34705.
[18]. A. Kumar, A. Sharma, and A. Arora, "Anxious Depression Prediction in Real-time Social Data," SSRN Electronic Journal, 2019, doi: 10.2139/ssrn.3383359.
[19]. M. M. Tadesse, H. Lin, B. Xu, and L. Yang, "Detection of Depression-Related Posts in Reddit Social Media Forum," IEEE Access, vol. 7, pp. 44883–44893, 2019, doi: 10.1109/access.2019.2909180.
[20]. H. S. Alsagri And M. Ykhlef, "Machine Learning-Based Approach for Depression Detection in Twitter Using Content and Activity Features," IEICE Transactions on Information and Systems, vol. E103.D, no. 8, pp. 1825–1832, Aug. 2020, doi: 10.1587/transinf.2020edp7023.
[21]. M. Dong, Y. Li, X. Tang, J. Xu, S. Bi, and Y. Cai, "Variable Convolution and Pooling Convolutional Neural Network for Text Sentiment Classification," IEEE Access, vol. 8, pp. 16174–16186, 2020, doi: 10.1109/access.2020.2966726.
[22]. R. Nareshkumar and K. Nimala, "An Exploration of Intelligent Deep Learning Models for Fine Grained Aspect-Based Opinion Mining," 2022 International Conference on Innovative Computing, Intelligent Communication and Smart Electrical Systems (ICSES), pp. 1–7, Jul. 2022, doi: 10.1109/icses55317.2022.9914094.
[23]. J. Kim, J. Lee, E. Park, and J. Han, "A deep learning model for detecting mental illness from user content on social media," Scientific Reports, vol. 10, no. 1, Jul. 2020, doi: 10.1038/s41598-020-68764-y.
[24]. Y. Hou, H. Gao, Z. Wang, and C. Du, "Improved Grey Wolf Optimization Algorithm and Application," Sensors, vol. 22, no. 10, p. 3810, May 2022, doi: 10.3390/s22103810.