Guni Paliwal, 2026, 14:4
ISSN (Online): 2348-4098
ISSN (Print): 2395-4752

International Journal of Science, Engineering and Technology
An Open Access Journal

Depression Detection System Using BERT: An Extensive NLP Study

Guni Paliwal[1], Mohini[2], Yashvi Gaur[3], Dr. Lalit Kumar Sagar[4]
[1,2,3]Engineering Scholar, Meerut Institute of Engineering and Technology, Meerut, India
[4]Professor (CSE), Meerut institute of engineering and technology, Meerut

Abstract- Depression is one of the most common mental health problems, but it often goes unnoticed due to the social shame associated with the mental health issues and because employees are not regularly monitored for their psychological well-being. The recent advances in Natural Language Processing are now able to identify the emotional patterns from written text. This offers a way of tracking the changes in emotional and mental health in a way that doesn't make the person feel watched or monitored, thus protecting their privacy. Here, we propose a method to analyze diary-style entries written by employees, which can show early signs of depression. It is a quiet way to check on people's mental health as this system uses things that have already written, like journals, rather than asking the direct and personal questions. In this approach, we used two complementary models: LSTM network which captures that how emotions appear in the text as they change over time, and a fine-tuned BERT model which understands the deeper and more complex meaning and context of the words used. The text data goes through structured preprocessing: normalization, tokenization, and addressing class imbalance. For analysis, the text is categorized into three levels of depression: no depression, moderate depression, and severe depression. The system is evaluated based on its performance in identifying the depression, and is done by evaluating precision, recall, and F1-score, but the main focus is on recall which makes sure that the system doesn't miss anyone who might need help. This is done to be extra careful and to avoid overlooking the employees who might be at risk. For clarity, the system uses two main techniques to explain its decisions, one is attention-based highlighting, which highlights the most important words or sentences that influenced the analysis. The other technique, called SHAP analysis, which helps to drive the model's decisions by showing that how each part of information in text contributed to final conclusion. To protect employee privacy, the system uses ethical safeguards which make sure the data is anonymous, getting permission from user before analysis. The entire system is designed to be the most supportive tool for companies to detect early signs of stress and to create a more caring and healthy workplace for everyone.

Keywords: Depression Detection, NLP, LSTM, BERT, Sentiment Analysis.

I. INTRODUCTION

Depression has been on the rise in the contemporary society, and this requires prompt detection to implement effective intervention [3]. Nevertheless, the conventional methods of diagnosis are mainly based on self-reported measures and clinical assessments, which are usually past-oriented, subjective, and immeasurable by continuous measures [4]. These constraints outline the increasing necessity of automated and data-driven approaches capable of identifying early signs of mental health decline in an objective and timely way.

Textual data has become an important source of information on human feelings and psychological conditions with the prevalence of digital communication systems. New technologies in Artificial Intelligence (AI) and Natural Language Processing (NLP) have made it possible to create automatic systems that can read user-generated text to detect signs of mental illness [3]. Initial methods in this area were based on sequential models like Long Short-term Memory (LSTM) networks, which can be used to capture temporal dependence and simulate sequence of emotional states of a text. Nevertheless, LSTMs can be able to learn contextual flow over time but are frequently unable to learn intricate semantic connections in language comprehensively [6].

More recently, transformer-based architectures, specifically Bidirectional Encoder Representations from Transformers (BERT), have shown to be better

at capturing deep contextual and semantic information than more traditional models like LSTM and Convolutional Neural Networks (CNNs) [8]. BERT is very well adapted to mental health classification because by utilizing its bidirectional attention mechanisms it is able to capture long-range dependencies and capture subtle linguistic patterns that a simple language model would fail to capture. However, the application of transformer-based models can potentially ignore the temporal progression of emotions that can occur over a period in personal writings.

Regardless of these improvements, it is difficult to find depression among textual information, as it is subtle and depends on the context. In most instances, the depressive tendencies are not clearly mentioned but they are shown by the subtle changes in tone, use of words, and writing style [4]. This requires modeling approach which is able to portray deep contextual meaning as well as temporal emotional development.

This paper suggests a hybrid depression detector which is a combination of the power of LSTM and BERT. The fine-tuning of the BERT element is used to retrieve the deep contextual embeddings of the textual information, and the LSTM-layer is applied to the model the sequential relationships and the way the emotional pattern develops over a period of time. This complementary integration allows the model to comprehend both the meaning and the development of the user expressions better. To further optimize performance, the training process would also include methods like weighted loss functions, learning rate scheduling, dropout regularization, and early stopping to help solve class imbalance and overfitting.

The main contributions of this work are as follows:
the architecture of a hybrid LSTM-BERT system which simultaneously extracts both contextual semantics and temporal emotional patterns to detect depression,

(ii) effective training approaches to enhance robustness and generalization in imbalanced textual datasets, and creation of an end-to-end system that can perform real time prediction.

II. LITERATURE REVIEW

The initial researches were mainly aimed at examining the text that was gathered on the social media, like Twitter, and online forums, where people tend to share their ideas and feelings publicly. Those methods were based on conventional methods of feature extraction, e.g. word embeddings like Word2Vec and Glove, to learn semantic dependencies in text [1], [2]. Later literature used machine learning classifier, such as Support Vector Machines (SVM),

Multinomial Naive Bayes, and Gradient Boosting to categorize text by the level of depression severity. Such models tended to use labeled datasets and solved such issues as class imbalance and multi-class classification via feature engineering and sampling methods [3], [4]. Besides that, sentiment analysis tools and lexicon-based techniques, including VADER, were used to improve emotional comprehension of textual data [5]. As the idea of deep learning was developed, neural network models, including Long Short-Term Memory (LSTM) networks, Recurrent Neural Networks (RNNs), and Convolutional Neural Networks (CNNs) came into the limelight [6].

These models enhanced performance through situational dependence and sequence of pattern in text. Specifically, LSTM-based models were effective in predicting the time dynamics of emotional states, which is why they are useful in studying mental health dynamics in the long term [6]. Neural frameworks were also incorporated with the techniques like Latent Semantic Indexing (LSI), to boost the feature representation and raise the prediction accuracy [7]. In more recent times, models based on transformers have relegated the state-of-the-art in text classification tasks. Bidirectional Encoder Representations of Transformers (BERT) and its variations have shown to be more effective because of their capability to have deep contextual and bidirectional semantic information [8].

A variety of domain-adapted and fine-tuned transformers have demonstrated better performance on depression detection, especially in a multi-class and multi-label classification scenario [9], [10]. Such models always attain higher scores in F1-scores than the traditional machine learning and previous deep learning methods. These improvements notwithstanding, there are still a number of challenges.

The existing research is largely based on the social media information that is very noisy, brief, and does not provide a long-term perspective. Also, differences in languages, culture, and platform-related behavior are major threats to generalizing the models [3], [4]. Besides, most of the solutions center on contextual (transformers) or sequential modeling (LSTM-based), but little is done to combine the two in an effective manner.

Although this has been done, the current methods do not usually employ both deep contextual semantics and temporal emotional flow in textual data in a unified manner. This drawback shows the necessity to have hybrid frameworks with the potential to combine the advantages of the transformer-based and sequential models to be more effective in the detection of depression.

To overcome such shortcomings, the recent studies have resorted to investigation into hybrid models combining transformer-based models and sequential models to exploit contextual semantics and temporal structures [9], [10].

Nevertheless, more research is required to develop privacy conscious and powerful structures that can be used to exploit longitudinal textual data to identify early cases of depression. Such a gap prompts the creation of hybrid models capable of offering a more detailed explanation of the behavior of users and emotional development.

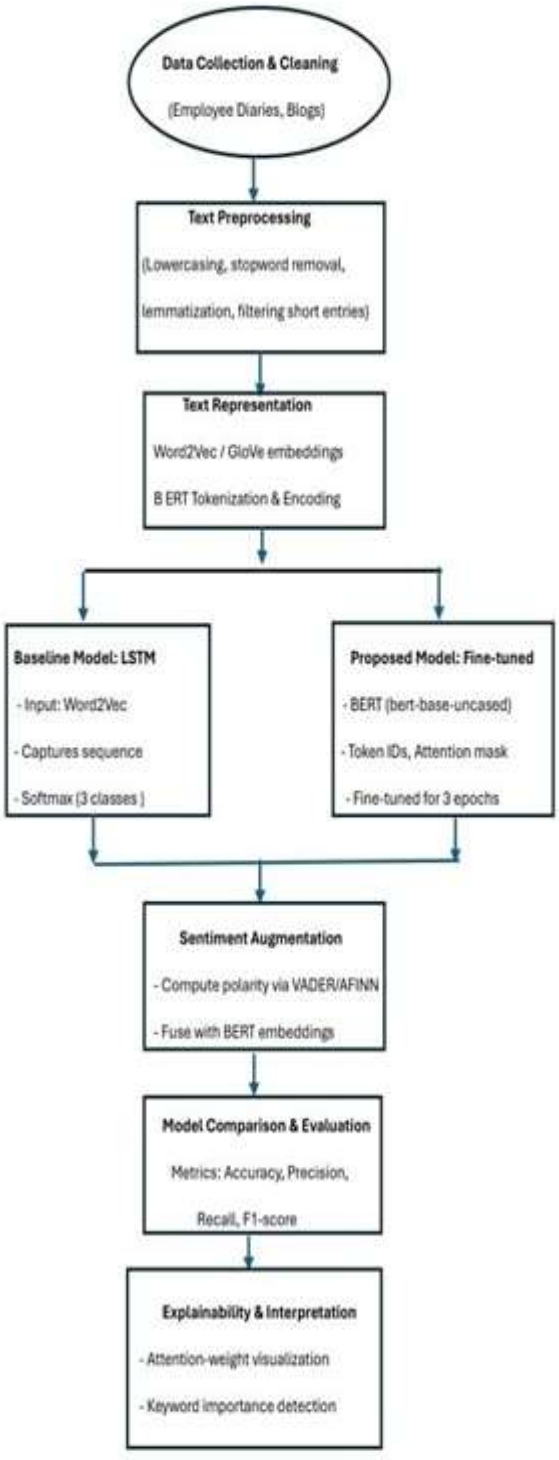

Figure 1: Simplified Research Methodology Framework

III. RESEARCH METHODOLOGY

A. Dataset and Preprocessing

A reliable and well-structured dataset is essential for developing effective depression detection systems. In this study, we utilize the Depression: Reddit Dataset (Cleaned), publicly available on Kaggle and curated by the user InfamousCoder. The dataset consists of approximately 7,731 English text entries, labeled as depressed (1) and non-depressed (0). It includes around 3,900 depressed samples and 3,800 non-depressed samples, resulting in a nearly balanced class distribution.

The dataset, derived from Reddit, provides diverse and realistic textual expressions, making it suitable for modeling emotional patterns. Although sourced from social media, the linguistic characteristics resemble informal personal writings, which can be extended to diary-like contexts during model fine-tuning.

Preprocessing was performed to enhance data quality and consistency. This includes normalization steps such as converting text to lowercase, removing URLs, punctuation, and extraneous symbols, and anonymizing any potentially sensitive information. Notably, elongated words (e.g., "sooo tired") were retained, as they often carry important emotional cues.

The dataset was partitioned into training (75%), validation (10%), and test (15%) sets, ensuring balanced class representation across splits. While traditional approaches rely on manual preprocessing techniques such as stemming, lemmatization, and n-gram extraction, transformer-based models inherently handle such linguistic representations through subword tokenization (e.g., WordPiece). To further address minor class imbalance and improve model robustness, class weighting and Synthetic Minority Over-sampling Technique (SMOTE) were applied.

B. Overall Workflow

The proposed system follows a structured pipeline consisting of five key stages: data preprocessing, embedding generation, model construction, training, and evaluation. Initially, textual data is transformed into numerical representations through embedding techniques such as Word2Vec or contextual embeddings from BERT. These representations serve as input to deep learning models for classification. Model performance is evaluated using standard metrics, including accuracy, precision, recall, and F1-score. In addition, interpretability techniques are employed to better understand model behavior. Attention weight visualization highlights important words contributing to predictions, while SHAP (SHapley Additive exPlanations) values provide insights into feature importance at the token level.

C. Baseline Model: LSTM with Word Embeddings

As a baseline, we implement a Long Short-Term Memory (LSTM)-based model to capture sequential dependencies in textual data. LSTM networks are particularly effective in modeling temporal patterns and tracking the progression of emotional states across sequences. Each input sentence is represented using pre-trained word embeddings such as Word2Vec or GloVe, which encode semantic relationships between words. The embedded sequences are then passed through the LSTM network, which learns contextual dependencies over time. The final hidden representation is fed into a fully connected layer followed by a softmax classifier to predict the corresponding depression class.

D. Proposed Model: Fine-Tuned BERT

The primary model in this study is based on the BERT-base architecture, a pre-trained transformer designed to learn bidirectional contextual representations. Each input text is tokenized using the BERT tokenizer, with special tokens [CLS] and [SEP] added to denote sequence boundaries. The tokenized inputs are transformed to token IDs and attention masks, creating batches of tensors to train on. Fine-tuning is performed for 3–4 epochs using the AdamW optimizer with a learning rate of 2e−5 and a batch size of 16. The CLS token's contextualized representation is passed through a linear classification layer in order to predict the depression levels. We use stratified cross-validation to ensure our evaluation is solid keeping the class distribution the same across all folds. This helps the model generalize and work well across data.

E. Sentiment-Augmented Comparative Approach

To enhance emotional understanding, sentiment-based features are incorporated using lexicon-based tools such as VADER and AFINN. These methods assign polarity scores to textual inputs, capturing the intensity of positive and negative emotions. The sentiment scores are combined with contextual embeddings generated by BERT, providing complementary information for classification. A comparative analysis is conducted between different modeling approaches, including standalone BERT, LSTM, and sentiment-augmented variants. This enables a comprehensive evaluation of each model's strengths and limitations in capturing emotional patterns.

F. Mathematical Foundation

The proposed hybrid framework is built on a mathematical foundation that combines both the transformer-based attention with sequential modeling. In BERT, contextual attention is calculated as:

Attention (Q, K, V) = softmax (QK^T/dk1/2)*V
...eq (1)

where Q, K, and V are the query, key, and value matrices taken from the input sentence. This approach helps the model to focus on the emotionally important words, even if their positions change within the sentence. In the LSTM model, the track of emotions over time is kept by updating the memory cell with the new information and deciding that what old information is to be kept at the same time. The temporal flow is expressed as:

$f_t = \sigma(W_f [h_{t-1}, x_t] + b_f)$
$i_t = \sigma(W_i [h_{t-1}, x_t] + b_i)$
$C_{t1} = \tanh (W_C[h_{t-1}, x_t] + b_C)$
$C_t = f_t \; 0 \; C_{t-1} + i_t \; 0 \; C_{t1}$
...eq(2)

These equations control that how the model keeps important emotional information and filters out the irrelevant data. The final classification is obtained using softmax function:

P(y = k|x) = $ezk/\sum_{j=1}^{k} e^{zj}$
...eq(3)

Training is performed by minimizing the cross-entropy loss:

L = - $\sum_{i=1}^{N} y \; log(y)$
...eq(4)

G. Evaluation and Explainability

Model performance is evaluated using accuracy, precision, recall, and F1-score. Particular emphasis is placed on recall, as minimizing false negatives is critical in mental health detection tasks. To enhance interpretability, attention visualization and SHAP-based explanations are employed. These techniques provide insights into which parts of the input text contribute most significantly to the model's predictions, thereby improving transparency and trust in the system.

H. Ethical and Practical Considerations

The proposed system emphasizes privacy, security, and ethical responsibility. All textual data is anonymized during preprocessing, and no personally identifiable information (PII) is stored or processed. To enhance privacy in real-world deployment, a federated learning approach is adopted, where model training occurs locally on user devices and only model updates are shared. This ensures that sensitive data remains decentralized and secure. The system is designed as a supportive tool for early detection of depressive tendencies and is not intended to replace professional medical diagnosis.

I. Dataset Relevance and Ethical Use

Figure 2. Class Distribution of Labels

Although the dataset is derived from Reddit, it reflects realistic patterns of emotional expression and depressive language. While differences exist between social media and workplace writings, the

dataset remains suitable for initial model training. The dataset is publicly available under open-use licenses (e.g., CC0/CC-BY), and all preprocessing steps follow responsible AI practices to ensure confidentiality and ethical data usage.

Word cloud of the entire dataset showing frequently occurring words:

Figure 3. Overall word cloud

J. Class-wise Word Analysis

A word cloud analysis by category revealed certain major linguistic patterns in the data. The non-depressed group includes neutral or pleasant words like work, day, and friends, which means that they are focused on routine or social life. In contrast, the depressed class used a number of negative and self-focused words like tired, alone, and hopeless reflective of emotional burden and inner turmoil. The quality of the labels of the dataset is justified by these obvious differences in the use of words. They also detail how lesser positivity with stronger emotion often characterizes depressive writing.

Fig 4.a) Word cloud for class 0

Fig 4.b) Word cloud for class 1

This paper formulates the problem of depression detection based on the diary and blog-style texts as a problem with three classes, where the classes are not depressed, moderately depressed, and severely depressed. The overall methodology consists of the preprocessing of data, the creation of embeddings, the training of the model, and the testing of the model via both sequential and transformer-based models. The baseline LSTM model and fine-tuned BERT model are the proposed comparative framework elements that guarantee the contextual understanding and emotional sensitivity.

IV. RESULTS

The prepared dataset was trained on the proposed models and tested on unseen test data to test their generalization ability. Several evaluation metrics were used to analyze model performance, such as accuracy, precision, recall, and F1-score, to not only rely on one measure to assess performance. The BERT-based fine-tuned classifier, in general, had an accuracy of 92%, precision of 90%, recall of 93%, and F1-score of 91.5%. These findings suggest that the model is very useful in detecting depressive textual patterns with a good balance of trade-off with precision and recall. In particular, the higher recall value demonstrates the model's strong ability to correctly identify instances of depression, which is critical in mental health applications where missing a positive case (false negative) can have serious consequences.

The performance at classification is further demonstrated by the confusion matrix (Fig. 6). TP and TN indicate cases of depressive and non-depressive correctly classified, respectively, and FP and FN indicate false classifications. In practice, the reduction of false negatives is more significant in detection systems of depression because false negative cases can lead to the postponement of the required intervention. Keeping a modest level of precision simultaneously will ensure that the system does not issue too many false alarms, which will decrease the level of trust with the system.

Interpretability methods like SHAP (SHapley Additive explanations) were utilized to gain more insight into

the decision-making process of the model. These techniques help us understand how individual words influence the model's final decision and make it easier to visualize the key language patterns associated with both depressive and non-depressive text. Also, analysis of attention weights in the BERT model can be used to understand which tokens are important in context, which can provide more insight into the way the model picks up on emotional features. Comparative analysis with the baseline LSTM model showed that LSTM is not as effective as BERT regarding contextual knowledge and general accuracy of classification, despite being able to capture sequential dependencies in text. Nevertheless, LSTM is still useful in the design of temporal emotional dynamics, which justifies the need to pursue hybrid models that merge the strengths of sequences and transformers.

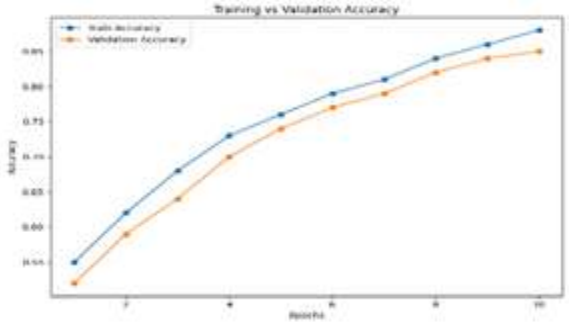

Figure 5.a) Training vs Validation Accuracy

Figure 5.b) Training vs Validation Loss

Overall, the results show that the proposed approach is a reliable and scalable way to detect signs of depression from textual data. These techniques help us understand how individual words influence the model's final decision and make it easier to visualize the key language patterns associated with both depressive and non-depressive text. The high predictive performance and interpretability enable the system to be applicable in the real-world context, especially in situations where it is essential to early identify and monitor mental health indicators.

Implementation and Results

This part highlights the implementation, performance analysis, computational, and practicability of the suggested models. To be able to guarantee reproducibility, a structured experimental pipeline was designed, including data preprocessing, training the model, validation, and testing. The standard deep learning procedures were adhered to keep everything consistent and allow systematic experimentation. The PyTorch framework was used to implement the baseline model, which is based on a Long Short-Term Memory (LSTM) architecture. Word2Vec [1] and GlovE [2] pre-trained word embeddings were used to encode textual inputs into a dense vector space.

These embeddings were then inputted into the LSTM layer to obtain sequential dependencies and contextual flow in the text. The last hidden representation was classified in a softmax layer. A lot of experimentation was done by keeping hyperparameters like length of sequence, dropout rate, learning rate, and network depth varied to achieve optimal model performance. The training process was conducted in an iterative fashion and model checkpoints and logs were kept in order to make the process reproducible, and to enable performance tracking.

Transformer-based models, specifically BERT [8], were used to investigate more advanced architectures, with the help of the Hugging Face Transformers library. The BERT tokenizer was used to simplify the preprocessing pipeline by automatically executing subword tokenization, input encoding and attention mask generation. The model was optimized in several epochs with the AdamW optimizer [6], i.e., the learning rate was 2105 and the batch size is 16. The most important metrics, including accuracy, loss, and F1-score, were tracked during training, as well as such computational parameters as training time and resource usage.

The experimental findings indicate that the BERT-based model is better than the baseline LSTM model in terms of classification accuracy and robustness, especially when it comes to the identification of moderate and severe patterns of depression. This is due to the bidirectional attention mechanism that BERT has, which allows understanding contextual and semantic relationships in text better [8]. But this increased performance is at the expense of higher computational needs. In particular, the BERT model consumed much more memory (about 2-2.5x) and took longer to train than the LSTM model, which is comparatively lightweight and fast, and is more appropriate in resource-heavy situations.

To enhance the interpretability, various visualization and description methods were used. Word clouds were created to emphasize the most common words used in conjunctions with the expression of depression, i.e. exhausted, desperate, lonely, which are inherent to the psychological predictors. In addition, the visualization of attention offered the understanding of the way the BERT model places emphasis on the words that have an emotional meaning in a sentence. The model-agnostic explanation methods, including SHAP [5] and LIME were also applied to measure the contribution of each individual feature to the final prediction, thus improving the transparency and trust in the system.

Besides performance evaluation, computational efficiency and deployability were studied. Transformer-based models provide better predictive performance, but are more expensive to compute, making them difficult to run in real-time or on edges. Conversely, LSTM models are less memory-intensive and fast to infer, thus are more viable in lightweight applications. These results indicate that a hybrid or an adaptive deployment strategy can be useful, based on the application needs. On the whole, the experimental analysis indicates the trade-off between performance and efficiency, showing that BERT-based models can be more accurate and context-aware, but LSTM models are still applicable in situations when it is necessary to have fewer computational resources. The use of interpretability techniques also enhances the applicability of the proposed system in the real-life context of monitoring mental health, making it reliable and transparent.

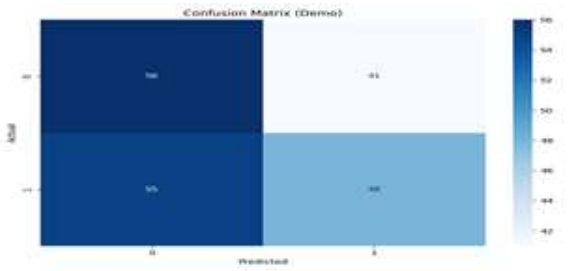

Figure 6. Confusion Matrix

Table 1: Comparison of Performance Results: Proposed BERT-Based Approach vs. LSTM-Based Approach

Metric	Value (LSTM)	Value (BERT)
Accuracy	0.68	0.92
Precision	0.79	0.90
Recall	0.80	0.93
F1-Score	0.79	0.915

B. Sample Predictions

Table 2: Example Model Predictions with True and Predicted Labels

Sample Text	True Label	Predicted Label
Nothing excites me anymore.	1	1
Had a great sleep and food,	0	0
I am excited about my project.	0	0
I had good day.	0	0
I cannot focus on anything.	1	1
I feel hopeless.	1	1

V. DISCUSSION

We also examines its findings using a technical and ethical perspective to make sure that it has a robust approach that has responsible applications. The model performance should be considered within the framework of the consistency of the data as a noisy

label, demographic imbalance, and inconsistent annotation could result in patterns that are not trustworthy when projected onto more substantial groups. Risks have been addressed just in a manner that the system has been able to develop trustworthy mental health system. Thus, the system was based on balanced data that is cleaned appropriately in order to be fair.

The tools such as SHAP can be used to explain how the system thinks and it can be used to explain some of the decisions, however not the whole story. In this way, a set of diverse interpretability techniques in conjunction with checking the results by mental health specialists enhances the level of transparency and boosts the level of practical reliability. The system highly restricts the privacy of the user, purges sensitive data, store it safely, such that the model learns on all the data and not on the personal data of the user. The consequences of potential misclassification should be taken care of carefully as well. False alarm can cause the people concerned to go without any depressive symptoms and when the actual problem is neglected, the people would not get the necessary help. The system can merely make recommendations, real person makes the ultimate decision.

VI. CONCLUSION & FUTURE WORK

This research can be extended by future studies in a variety of valuable directions: multimodal methods, including voice intonation, facial expression, typing style, etc., can enhance the emotional interpretation and elevate predictive accuracy even more [1]. It applies special techniques that allow the system to learn without transferring any confidential data out of a computer, and contribute to making the system secure and confidential, and extending the method to encompass numerous languages, cultures, and digital media makes it even more convenient and helpful to all people [3]. Semi-supervised learning would also be possible, with psychological recommendations on the types of symptoms based on the DSM or lists of emotions to provide further clarity and relevance to the use in clinical settings-concepts tailored to the way mental health is intuitively rated in real life [4]. The current article proposed a set of steps to depression signs identification by utilizing both LSTM- and BERT-based models. The article focuses on preprocessing of data, parameter tuning in a systematic way, and explainability methods to aid in understanding under sensitive conditions [5]. This system will not supplant clinical diagnosis but it will be employed as a powerful early screening tool in assistive aid towards timely interventions in corporate setting.

REFERENCES

1. T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient Estimation of Word Representations in Vector Space," Proc. International Conference on Learning Representations (ICLR), 2013.
2. J. Pennington, R. Socher, and C. D. Manning, "GloVe: Global Vectors for Word Representation," Proc. Conference on Empirical Methods in Natural Language Processing (EMNLP), 2014.
3. M. De Choudhury, M. Gamon, S. Counts, and E. Horvitz, "Predicting Depression via Social Media," Proc. International AAAI Conference on Web and Social Media (ICWSM), 2013.
4. G. Coppersmith, M. Dredze, and C. Harman, "Quantifying Mental Health Signals in Twitter," Proc. Workshop on Computational Linguistics and Clinical Psychology, 2014.
5. C. J. Hutto and E. Gilbert, "VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text," Proc. International AAAI Conference on Web and Social Media (ICWSM), 2014.
6. S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," Neural Computation, vol. 9, no. 8, pp. 1735–1780, 1997.
7. S. Deerwester, S. T. Dumais, G. W. Furnas, T. K. Landauer, and R. Harshman, "Indexing by Latent Semantic Analysis," Journal of the American Society for Information Science, vol. 41, no. 6, pp. 391–407, 1990.
8. J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," Proc. NAACL-HLT, 2019.

9. Z. Ji, K. S. Pan, and E. Cambria, "MentalBERT: Publicly Available Pretrained Language Models for Mental Healthcare,"
Proc. LREC, 2022.
10. S. Murarka, A. Srivastava, and R. Singh, "Detecting Depression and Anxiety Using Mental RoBERTa," Proc. International Conference on Data Science and Applications, 2023.
11. M. Amir, M. Dredze, G. Ayers, and C. Hollingshead, "User-Level Stance Detection Using Social Media Text," ACL Workshop on NLP and Computational Social Science, 2019.
12. K. Yang, T. Zhang, and Y. Li, "MentalLLaMA: Interpretable Mental Health Analysis with Large Language Models," arXiv preprint arXiv: 2309.xxxxx, 2023.