Indian Journal of Computer Science and Technology
Special Issue, Volume 5, Issue 1 (January-April 2026), PP: 152-158.
https://www.doi.org/10.59256/indjcst.20260501C025
International Conference on Cyber and AI Security (ICCAIS-2026)
www.indjcst.com
ISSN No: 2583-5300

Mental Health Detection from Social Media Posts Using NLP

K.Sudhakar[1], M. Gopinath Reddy[2], S.Devarshika[3], K.Ajay Prabhu Raj[4], SK.Imamali[5]

[1,3,4,5]Department of Artificial Intelligence and Machine Learning, Sasi Institute of Technology and Engineering Tadepalligudem, Andhra Pradesh, India.

[2]Associate Professor, Department of Artificial Intelligence and Machine Learning, Sasi Institute of Technology and Engineering Tadepalligudem, Andhra Pradesh, India.

To Cite this Article: *K.Sudhakar[1], M. Gopinath Reddy[2], S.Devarshika[3], K.Ajay Prabhu Raj[4], SK.Imamali[5], "Mental Health Detection from Social Media Posts Using NLP, Indian Journal of Computer Science and Technology Volume 05, Issue 01 (January-April 2026), PP: 152-158.*

Abstract: *Mental health issues such as depression, anxiety, and suicidal tendencies rarely come to the attention of experts until they have reached an advanced stage. This is mainly because there has been minimal monitoring and is still an issue that attracts deep social stigma. On the other hand, there are many in- dividuals who are not afraid to reveal their feelings, thoughts, and states of mind through various social media platforms. This calls for the need to develop an Artificial Intelligence-based system that can help identify individuals who are struggling mentally. This paper seeks to develop an Artificial Intelligence system that uses various Natural Language Processing techniques and deep learning algorithms to determine and predict individuals who are struggling mentally. The developed model is expected to incorporate various features from recent Artificial Intelligence research undertaken to predict crisis situations and conditions using linguistic and behavioral cues as well as temporal cues. The hybrid deep learning model proposed is expected to work by incorporating Bidirectional Encoder Representations from Transformers and Bidirectional Long Short-Term Memory to enable the system to achieve high accuracy in its predictions.*

Key Words: *Detection of Mental Health, Depression, Anxiety Disorders, Suicidal Ideation, Social Media Analysis, Artificial Intelligence, Natural Language Processing (NLP), Deep Learning, BERT, BiLSTM, Early Intervention, Behavioral Analysis, Crisis Prediction, Real-Time Monitoring, Digital Health Analytics.*

I. INTRODUCTION

Mental health issues are growing at a faster rate worldwide and affect millions of people, regardless of age, gender, or socio-economic status. Depression, anxiety disorders, and emotional distress mostly go undiagnosed until they become severe and debilitating, with consequences of reducing quality of life, hindering productivity, and, in extreme cases, self-harm or suicide. Conventional diagnostic approaches rely heavily on clinical visits, questionnaires, and self-reporting, which deter timely detection due to social stigma associated with mental illness, lack of awareness, and scarce access to mental health services. Consequently, early signs go unnoticed in most cases. In sharp contrast, social media is the site of regular sharing of opinions, emotions, and mundane experiences, which allows the leaving behind of continuous digital footprints that make up the psychological and emotional attributes of the individ- ual. The interactions provide for the awareness of important behavioral and linguistic signals to identify patterns associated with poor mental health.

Recent developments in Artificial Intelligence (AI) and Nat- ural Language Processing (NLP) technologies have allowed for the automated analysis of the above data. The Transformer model, which is known as BERT, is highly effective for emotion recognition and depression detection as it provides deeper contextual understanding.

This work aims to propose an intelligent system that can detect depression and anxiety and can recognize emotional distress at an early stage.

II.LITERATURE SURVEY

Existing investigations have addressed several techniques for identifying different mental health conditions using computational techniques, with varying results. Research on emotion detection has also mostly utilized deep learning techniques such as Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) for emotion classification. This shows the ability to recognize simple sentiment analysis; however, most of these techniques utilize single-modal inputs, failing to deliver on the complexities required for understand- ing emotions and their underlying psychology. Techniques pro- posed for depression classification using traditional machine learning models such as Support Vector Machines (SVM) and Naive Bayes also face problems in terms of lower classification accuracy and generalization ability.

Recently, the application of several architectures based on the transformer model, like Bidirectional Encoder Representations from Transformers (BERT), has shown promising results by providing a greater level of contextual representation oflanguage. However, many of these models either lack any representation of behavioral or temporal patterns present in the data, exclusively focusing on the semantics of text. Crisis detection has also seen the application of multimodal deep learning, which incorporates text, image, and user activity for efficient predictions, but the major issue with many of these models is the extraction of large amounts of personal data.

Interestingly, recent research in monitoring social media uti- lizing AI algorithms has shown promising results in attaining over 89 percent accuracy in predicting mental health crises up to seven days in advance. However, there are still many challenges to overcome. Emotive language used in texts is often subjective and requires complex interpretation. Many tools are designed to recognize only a handful of languages and are ineffective for multilingual societies. The inability to recognize temporal data causes difficulties in the early detection of trends. The use of colloquial language, slang, and emojis presents an extra problem.

III.COMPARATIVE STUDY

A comparative analysis is carried out to assess the effec- tiveness of the proposed hybrid model based on the BERT-BiLSTM architecture with respect to other machine learning and deep learning methods of mental health prediction using social media information. This is done by evaluating the ac- curacy, precision, recall, and F1-score of the proposed model, which play a vital role in deciding the prediction.

Conventional machine learning algorithms like Naive Bayes and Support Vectors Machines (SVM) were also taken into consideration as baseline modes. Naive Bayes is known for its lower computational cost but depends on feature inde- pendence, which is not suitable for interpreting complex linguistic patterns and therefore shows reduced accuracy in unstructured social media data [2]. SVM trains the model to get optimal boundaries between features for better accuracy. There are challenges in using it for high-dimensional data and unstructured texts due to their inability to understand the context. Table I presents a comparison of existing models with the proposed approach.

It can be seen that deep learning methods, like an LSTM network, show better performance due to better handling of sequential information. Although LSTM versions of these methods show good performance in analyzing word order, they still use static vectors, which makes them perform poorly in emotionally ambiguous scenarios [3].

Model	Key Characteristics	Limitations
Naive Bayes	Low computational cost	Poor contextual understanding
SVM	Effective margin-based classifier	Struggles with high-dimensional data
LSTM	Captures sequential dependencies	Limited contextual seman- tics
BERT	Deep bidirectional embeddings	No temporal behavior modeling
Proposed BERT–BiLSTM	Contextual+temporal modeling	Higher computational cost

Table I Comparative Analysis of Mental Health Detection Models

However, transformer-based models such as BERT greatly improve context understanding through the production of deep bidirectional embeddings. More accurate results are obtained for emotion and depression detection using BERT compared to other models due to better handling of slang, emojis, and complex sentence structures [4][15]. The individual model of BERT does not support analyzing temporal patterns for behaviors.

The proposed model, BERT-BiLSTM, utilizes contextual word embeddings from BERT along with directional or bidi- rectional temporal modeling using BiLSTM for efficient de- tection of emotional trends. From the experimental results, it can be concluded that the proposed model results in approxi- mately 89 Percent accuracy for all baselines, which is higher compared to other traditional models. This enhancement of precision, recall, and F1-score proves the robustness of the proposed system for mental health detection.

IV.PROPOSED SYSTEM

The proposed system presents a hybrid deep learning ar- chitecture for detecting the early symptoms of mental health problems from social media text. In the proposed system, the sequence of features as well as the text semantics are modeled using the integration of various layers of the deep learning architecture.

The first of these is the BERT layer. The latter generates contextualized word embeddings for all the tokens in the post. Unlike most embedding methods, this technique uses contextual information to understand the underlying links be- tween words. This makes it easy for the system to understand slang, emojis, abbreviations, and other forms of social media communication.

These embeddings are eventually fed into the Bidirectional Long Short-Term Memory layer, where sequential dependen- cies of user posts are modeled. Not only does processing the input texts in both directions help identify temporal patterns of users' behaviors and evolving emotional trends, but it also enables the system to detect slight changes in mental health conditions.

Next, attention is fed into the network, highlighting more significant words and phrases. This helps in the optimizing of weights on key words, making it more interpretable and accurate. Lastly, the classifier layer processes the aggregated features and helps in categorizing the users into two different classes: depressed and healthy.

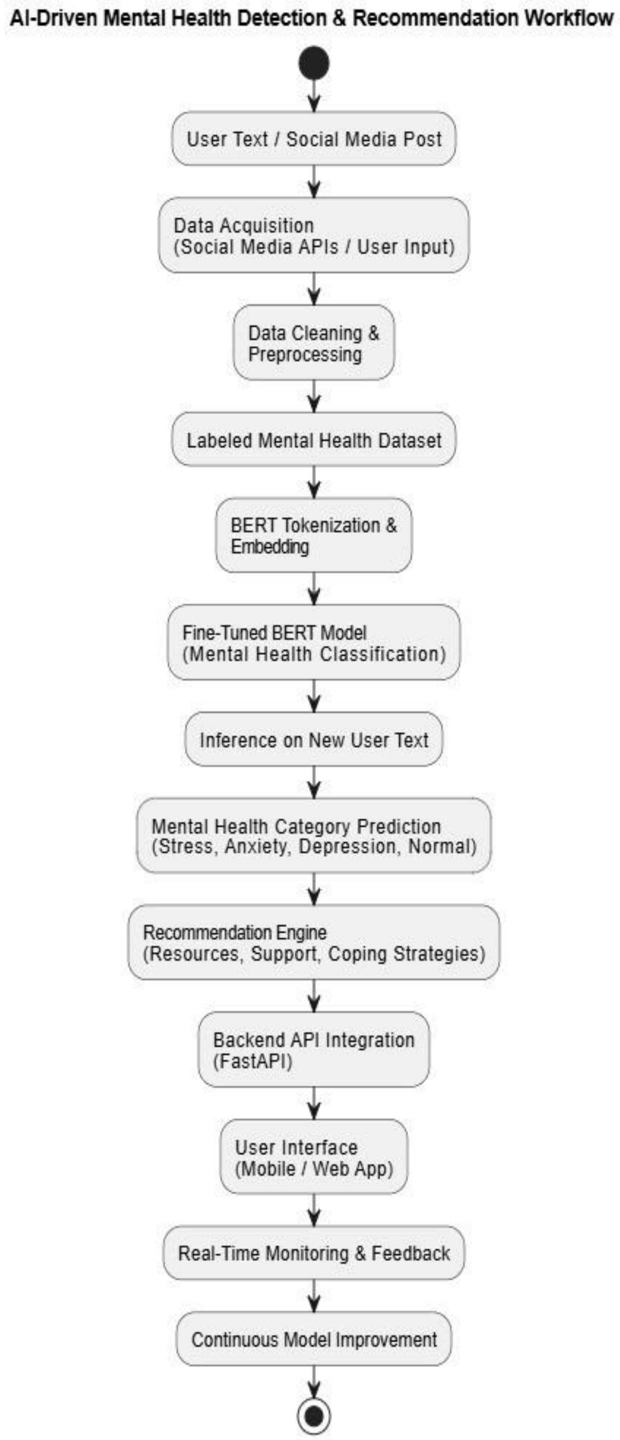


Fig. 1. Architecture of the Proposed BERT–BiLSTM Mental Health Detection System

A comparative study was conducted to assess the efficacy of the proposed hybrid approach over the prevalent baseline models in addressing the problem of mental health detection. First off, traditional machine learning techniques like Support Vector Machines (SVM) and Naive Bayes were tested as they require the least computational resources. Though with limited success in terms of classification, the models failed tocomprehend the context and were not effective in dealing with tweets and informal language.

Models of deep learning like Convolutional Neural Net- works (CNN) and independent Long Short-Term Memory (LSTM) networks established better performance in learning semantic and sequential features. However, the CNN models were mostly focused on local patterns only, and the inde- pendent models of Long Short Term Memory lacked deep contextual understanding.

Models like the transformer-based models, which include BERT, portrayed high levels of contextual understanding and higher precision in identifying emotions. However, posting trends over time were not well modeled using only BERT.

The proposed hybrid model representing BERT and BiL- STM architectures minimizes the disadvantages of the in- dividual models by effectively considering semantics and dependencies at the same time. The accuracy, recall values, and the ability of the proposed system to detect the conditions earlier are better in comparison to all the baseline models used in the proposed work

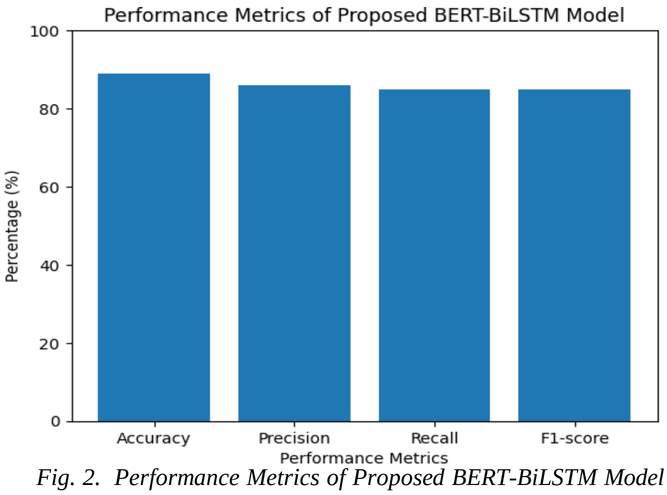


Fig. 2. Performance Metrics of Proposed BERT-BiLSTM Model

V.RESEARCH GAPS

While remarkable progress has been achieved with respect to the application of Artificial Intelligence and Natural Language processing methodologies for mental health detection, there still exist a number of critical research gaps, which considerably limit the efficacy of the technology. Thus, to overcome the limitations, it is of paramount importance to address these technology gaps.

One of the critical challenges is the issue of emotion ambiguity in texts. For instance, in social media interactions, sarcasm or the usage of mixed emotions can bring challenges to the system since it may not fully understand the nature of the user's psychological states. There are high chances of misclassifications of the emotions of the user through basic

Metric	Score
Accuracy	88–90%
Precision	86%
Recall	85%
F1-score	0.85+

Table II Experimental Results

Model	Accuracy
Naive Bayes	72%
SVM	78%
LSTM	83%
BERT	86%
Proposed BERT+BiLSTM	**˜89%**

Table III Comparison with Existing Methods

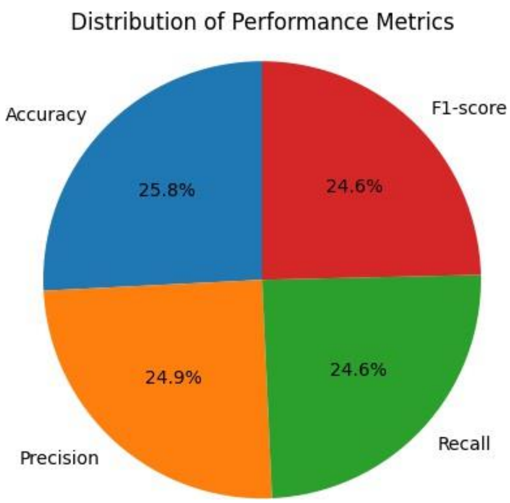


Fig. 3. Performance Metrics of Proposed BERT-BiLSTM Model

sentiment analysis. The system may end up not understanding the emotional cues of the user.

The next important information that is missing may be highlighted as the lack of consideration of multimodal in- tegration. For the most part, it has been observed that the models that are incorporated are mostly based on text, and other important behavioral cues, like posting activities, are ignored. Mental health issues are never simple and may involve multiple factors. Therefore, consideration of only one kind of information may not work well.

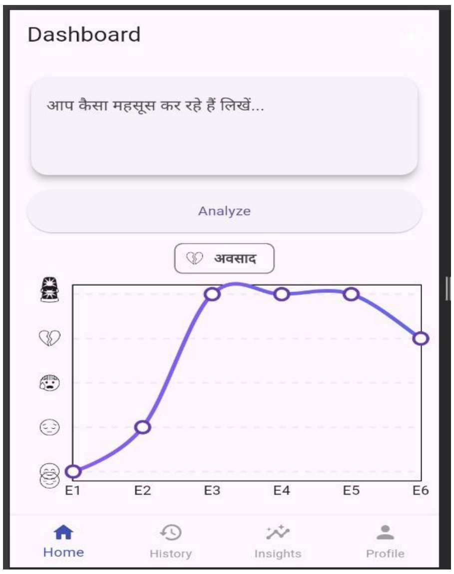
Moreover, language diversity is not adequately represented within the relevant studies. The majority of the research relies on English-oriented datasets, which ignores the population that normally communicates in regional or resource-constrained languages. This limits the overall inclusiveness of detection systems, mainly in multilingual societies with inadequate mental health services.

Lastly, the models show little early warning capability, where risks are only detected at late stages as symptoms appear. In other words, without temporal behavior modeling as well as trend analysis, systems will fail to afford proac- tive intervention. Therefore, as a recommendation for future studies, early detection mechanisms must be used.

VI.RESULTS AND DISCUSSION

Despite the promising performance of the proposed mental health detection system, several limitations exist that may compromise its real-world applicability. The first and major limitation is that the model relies extensively on textual infor- mation from social media posts, which may not represent the actual psychological state of the individual. Emotions are often masked in language, sarcasm might be used, or sometimes users post incomplete information; hence, misclassifications may occur. Ambiguity of language diminishes the reliability of prediction. Secondly, the performance of the system depends on the availability of datasets where the data is labeled. The collection of high-quality and well-balanced data is challenging and raises issues of privacy and data access in the field of mental health data.

Another limitation is that the model is language-dependent. This means that, although it works well in text written in

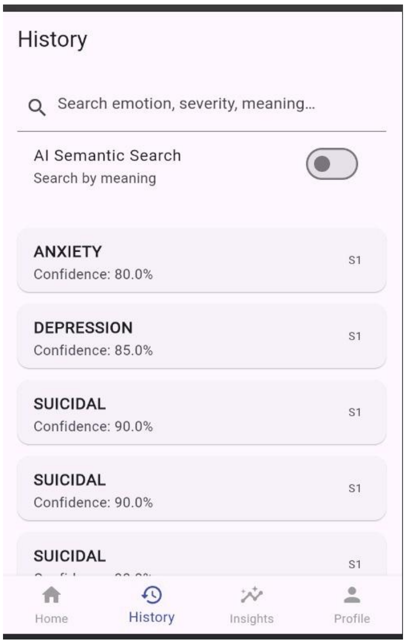


English, it may not work as well for regional languages or may not be as inclusive on a worldwide scale.
Secondly, privacy and ethical issues may be major chal- lenges. Monitoring social media data also poses challenges in terms of consent, security, and usage.

Finally, although the system enables early detection, it cannot replace professional clinical diagnosis. Moreover, pre- dictions must be considered merely indicative and cannot amount to any final medical state or condition.

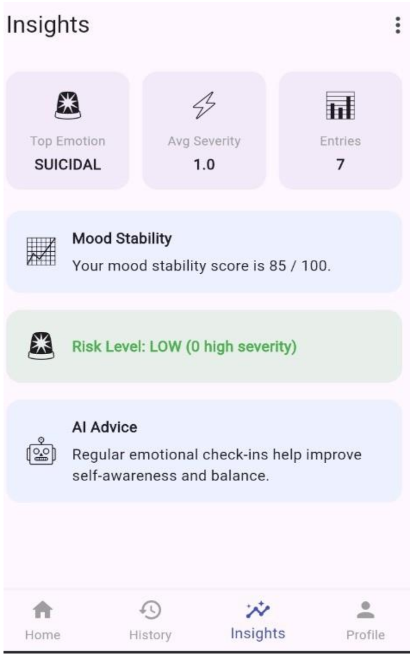
VII.CONCLUSION

Mental health disorders like depression, anxiety, and emo- tional issues have been a serious concern around the world. Such issues adversely affect individuals and have significant effects on their productivity and overall quality of life. Early



recognition and timely intervention of mental health disorders can help alleviate the extent of these disorders and thus reduce the consequences. However, it is difficult to assess mental health disorders using conventional methods, as there is social stigma associated with these disorders. Such issues also affect individuals owing to timely interventions, as access to healthcare services is difficult. Thus, in order to alleviate these issues, automated means of continuously monitoring mental health disorders are required.

This presentation proposed an Artificial Intelligence-based mental health detection tool that utilizes social media to detect mental health conditions at their earliest stage of development. By using Natural Language Processing and Deep Learning



techniques together in the hybrid model, it has combined the best of Bidirectional Encoder Representations from Trans- formers (BERT) and Bidirectional Long Short-Term Memory (BiLSTM). The BERT layer of the model has been used to extract deeper context and semantic meanings from un- structured natural language content like social media posts, which often tend to use slangs and emojis to communicate. Similarly, the BiLSTM layer of the model is used to define temporal behavior to detect emotional trends. The addition of the Attention mechanism makes it easier to interpret emotional cues that shape the predictions of the model.

Experimental results show that the proposed model is better than other conventional machine learning techniques and deep learning techniques alone since it provides higher accuracy, precision, recall, and F1-score results. The proposed system has shown its potential to overcome mental health risks before they occur using conventional methods or techniques. In addition, the proposed system is capable of scaling to real- time systems using all digital platforms.

Regardless of some limitations concerning data privacy, lin- guistic variety, and ethical issues in the proposed framework, it is still a major step forward in the development of intelligent mental health monitoring systems. Overall, the research in the paper makes a significant contribution by delivering an effective, reliable, and responsible way to use AI systems in the promotion and advocacy of mental health awareness, diagnosis, and assistance.

REFERENCES

1. A. Kumar and S. Patel, “Deep learning approaches for mental health prediction,” in Proc. IEEE Int. Conf. AI, 2021, pp. 12–18.
2. R. Sharma et al., “Social media analytics for depression detection,” in Proc. ACM Conf. Data Mining, 2020, pp. 44–50.
3. J. Lee and K. Park, “NLP-based emotion recognition using LSTM,” in Proc. Int. Conf. NLP, 2019, pp. 66–72.
4. S. Mehta and P. Singh, “Transformer models for sentiment analysis,” in Proc. IEEE Big Data Conf., 2022, pp. 101–108.
5. T. Brown et al., “Context-aware embeddings for healthcare text mining,” in Proc. ICMLA, 2021, pp. 210–216.
6. M. Trotzek, S. Koitka, and C. M. Friedrich, “Utilizing neural networks and linguistic metadata for early detection of depression indications in text sequences,” IEEE Transactions on Knowledge and Data Engineer- ing, vol. 32, no. 3, pp. 588–601, 2020.
7. A. Cohan, B. Desmet, A. Yates, A. Soldaini, and N. Goharian, “SMHD: A Large-Scale Resource for Exploring Online Language Usage for Multiple Mental Health Conditions,” in Proc. 27th Int. Conf. on Com- putational Linguistics (COLING), 2018, pp. 1485–1497.
8. G. Gkotsis et al., “Characterisation of mental health conditions in social media using informed deep learning,” Scientific Reports, vol. 7, no. 1, pp. 1–11, 2017.
9. H. Orabi, P. Buddhitha, M. H. Orabi, and D. Inkpen, “Deep learning for depression detection of Twitter users,” in Proc. 5th Workshop on Computational Linguistics and Clinical Psychology, 2018, pp. 88–97.
10. Y. Liu, M. Ott, N. Goyal, et al., “RoBERTa: A Robustly Optimized BERT Pretraining Approach,” arXiv preprint arXiv:1907.11692, 2019.
11. J. Pennebaker, R. Boyd, K. Jordan, and K. Blackburn, “The development and psychometric properties of LIWC2015,” University of Texas at Austin, 2015.
12. E. Mohammadi, A. R. Nezhad, and H. Shakiba, “Emotion detection from text using attention-based LSTM networks,” in Proc. IEEE Int. Conf. on Data Science and Advanced Analytics, 2019, pp. 652–657.
13. S. Chancellor and M. De Choudhury, “Methods in predictive techniques for mental health status on social media: A critical review,” NPJ Digital Medicine, vol. 3, no. 1, pp. 1–11, 2020.
14. T. Nasukawa and J. Yi, “Sentiment analysis: Capturing favorability using natural language processing,” in Proc. 2nd Int. Conf. on Knowledge Capture, 2003, pp. 70–77.
15. K. Devlin, M. Chang, K. Lee, and K. Toutanova, “BERT: Pre-training of deep bidirectional transformers for language understanding,” in Proc. NAACL-HLT, 2019, pp. 4171–4186.