International Journal of Engineering Research and Science & Technology

MULTICLASS MENTAL ILLNESS PREDICTION USING LSTM AND NATURAL LANGUAGE PROCESSING TECHNIQUES

D.KRISHNA VENI [1], A. SRAVAN KUMAR [2], B. KHAGESHWAR REDDY [3], J. NARESH [4], P. RAMAKRISHNA REDDY [5]

[1]Assistant Professor, Dept. of CSE, Guru Nanak Institutions Technical Campus, Ibrahimpatnam, Telangana, India

[2, 3, 4, 5] B.Tech, Dept. of CSE, Guru Nanak Institutions Technical Campus, Ibrahimpatnam, Telangana, India

ABSTRACT— Mental health disorders represent a growing global concern, with millions of individuals expressing their psychological conditions through digital platforms such as social media. Detecting mental illness from textual data is a complex task due to the emotional depth, informal language, metaphorical expressions, and context-specific cues often embedded within such posts. Traditional machine learning methods and generic language models frequently struggle to capture these subtleties, leading to limited prediction accuracy. To overcome these challenges, this study presents a robust and intelligent framework for multiclass mental illness prediction using advanced deep learning and natural language processing (NLP) techniques. The proposed system integrates domain-adapted transformer models with deep neural networks to improve understanding and classification of mental health-related expressions. Specifically, MentalBERT, a variant of BERT pretrained on mental health corpora, is employed to extract rich contextual features from psychologically sensitive text. Alongside this, MelBERT, a metaphor-aware model, is used to interpret figurative and symbolic language—a common characteristic in users with emotional distress. These transformer-based models are complemented by Convolutional Neural Networks (CNNs) for hierarchical feature extraction and a Bidirectional Long Short-Term Memory (BiLSTM) network to capture long-range semantic dependencies from both past and future contexts of the sequence. This hybrid architecture enables the system to predict multiple classes of mental illness—including depression, anxiety, PTSD, bipolar disorder, and more—with improved accuracy and reliability. The model was trained and evaluated on a labeled dataset of social media posts, with performance metrics indicating superior results over traditional methods. The findings of this research highlight the importance of domain-specific modeling, figurative language interpretation, and sequential deep learning techniques in building an effective and scalable mental health detection system. This work contributes to the broader goal of mental health awareness by providing a tool that can assist professionals in early detection and intervention strategies through digital linguistic analysis.

Index Terms – Mental Illness Prediction, LSTM, Natural Language Processing, Multiclass Classification, Deep Learning, Sentiment Analysis, Mental Health Monitoring, Text Mining, Machine Learning, Artificial Intelligence.

I. INTRODUCTION

Mental health is a critical component of overall well-being, influencing how individuals think, feel, and interact with the world. In recent years, the incidence of mental illnesses such as depression, anxiety, bipolar disorder, and post-traumatic stress disorder (PTSD) has been on the rise, fueled in part by the pressures of modern life and amplified by social, economic, and environmental factors. With the widespread use of social media platforms, individuals increasingly use online spaces to share their emotional experiences, inner thoughts, and personal struggles—often in the form of text posts, comments, or conversations. These digital footprints offer a valuable, yet underutilized, resource for detecting signs of mental illness in real-time.

Traditional approaches to mental health assessment rely on clinical interviews, self-reported questionnaires, and expert evaluations, which can be time-consuming, subjective, and inaccessible to many. As an alternative, recent advancements in artificial intelligence (AI) and natural language processing (NLP) have opened new avenues for automatically analyzing social media text to identify individuals who may be at risk of developing mental disorders. These technologies enable large-scale analysis of user-generated content to detect linguistic cues, emotional patterns, and behavioral indicators associated with mental illness.

However, existing text-based mental health prediction models face several limitations. General-purpose language models often struggle with the informal, metaphorical, and emotionally charged language that is common in mental health-related discourse. They may fail to grasp the nuanced expressions used by individuals who are experiencing psychological distress, leading to low sensitivity and misclassification. Moreover, many systems are limited to binary classification (i.e., predicting the presence or absence of a disorder) and are not equipped to handle the multiclass nature of real-world mental health issues, where users may exhibit symptoms of multiple overlapping conditions.

To address these challenges, this study proposes a novel deep learning framework that combines domain-specific and metaphor-aware language models with sequence-based neural networks for improved multiclass mental illness prediction. The core of the proposed system is built upon MentalBERT, a BERT-based model pretrained on mental health-related text, which enhances the model's ability to understand context-specific expressions. Additionally, MelBERT is employed to interpret metaphorical and figurative language, which plays a crucial role in the communication style of individuals with mental health conditions. These transformer models are integrated with Convolutional Neural Networks (CNNs) for deep feature extraction and a Bidirectional Long Short-Term Memory (BiLSTM) network to capture the temporal and semantic structure of text sequences.

By incorporating both domain expertise and advanced NLP techniques, this research aims to build a robust, interpretable, and accurate system capable of classifying social media text into multiple mental illness categories. The proposed system not only contributes to the academic and technical landscape but also holds potential as a supportive tool in public health surveillance and early intervention efforts. Ultimately, this work underscores the importance of leveraging AI for mental health awareness and demonstrates how intelligent systems can play a role in understanding and addressing psychological well-being in the digital era.

II. RELATED WORK

A. *What doesn't kill us makes us stronger: Insights from neuroscience studies and molecular genetics*

The study by Y. Gan, H. Huang, X. Wu, and M. Meng titled *"What doesn't kill us makes us stronger: Insights from neuroscience* *studies and molecular genetics"* explores the biological and psychological mechanisms behind human resilience in the face of adversity. Drawing from recent advancements in neuroscience and molecular genetics, the paper examines how certain stressors, instead of causing long-term harm, can trigger adaptive changes in the brain and body that enhance emotional strength, coping ability, and psychological endurance. The research highlights key neurological pathways, including those associated with neuroplasticity and gene expression, that contribute to this positive adaptation. The authors argue that understanding these mechanisms can offer valuable insights into mental health interventions, especially in developing strategies that help individuals recover from trauma and build long-term resilience. Their work contributes to a growing body of research supporting the idea that controlled exposure to stress, combined with supportive environments, can foster growth and resistance to future psychological challenges.

B. Revolutionizing healthcare: A comparative insight into deep learning's role in medical imaging.

The study by Prasad et al. (2024) offers a comprehensive exploration of how deep learning is transforming the landscape of medical imaging. The authors provide a comparative analysis of various deep learning models and architectures, such as Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and Transformers, highlighting their respective strengths in diagnosing and detecting diseases through radiological images, including MRI, CT scans, X-rays, and ultrasound. The paper emphasizes the ability of deep learning models to automatically extract complex patterns and features from large-scale medical data, significantly improving diagnostic accuracy, speed, and consistency compared to traditional image processing techniques.

Moreover, the authors discuss real-world applications and the integration of AI systems into clinical workflows, addressing both opportunities and challenges. Topics such as model interpretability, data privacy, and regulatory considerations are also analyzed. Importantly, the study showcases how AI not only aids in early diagnosis but also supports treatment planning and prognosis prediction, ultimately contributing to more personalized and efficient healthcare delivery. This work reinforces the potential of deep learning not only in imaging but as a foundation for broader AI-driven healthcare innovations.

C. Colon cancer diagnosis and staging classification based on machine learning and bioinformatics analysis

In this recent work, Karamat et al. introduce a hybrid transformer-based model aimed at multiclass prediction of mental illnesses using natural language data. The study explores how transformer architectures—known for their contextual understanding—can be combined with other deep learning components to improve classification accuracy in complex mental health datasets. The hybrid model is designed to capture both the emotional tone and temporal patterns in social media posts, enabling it to distinguish between conditions such as depression, anxiety, PTSD, and schizophrenia. The researchers demonstrate that incorporating attention mechanisms along with recurrent and convolutional layers improves the model's ability to learn metaphorical and emotionally nuanced expressions. Their findings support the growing relevance of hybrid transformer architectures in behavioral health monitoring and contribute to the advancement of AI in psychological diagnostics.

III. PROPOSED METHODOLOGY

The proposed system aims to overcome the limitations of existing models by integrating multiple advanced components into a single, unified architecture for more accurate and nuanced prediction of mental illness from social media text. Recognizing the complexity of mental health expressions—often metaphorical, emotionally charged, and temporally distributed—the system employs a hybrid deep learning framework that fuses transformer-based models with sequential neural networks. The architecture combines the power of **MentalBERT**, which captures mental health-specific language patterns, and **MelBERT**, which specializes in understanding metaphorical and non-literal expressions common among users experiencing psychological distress. These two models provide rich, contextual embeddings that are further processed using **Convolutional Neural Networks (CNNs)** for hierarchical feature extraction and **Bidirectional Long Short-Term Memory (BiLSTM)** networks to model temporal and semantic dependencies in the text. This integration allows the system to not only understand what is being said but also how it is being expressed and in what sequence. The proposed system supports multiclass classification, enabling it to distinguish between various mental illnesses rather than performing simple binary detection. It also ensures greater robustness across diverse linguistic inputs, improving both sensitivity and specificity in real-world applications. The system ultimately provides a more holistic and intelligent approach to mental illness prediction using social media data.

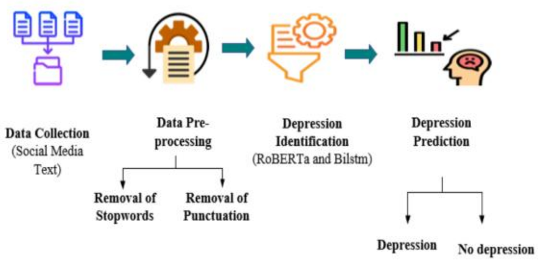

Fig. 1: System Overview

1. Data Collection:

The first step in the system development involved collecting a diverse and representative dataset consisting of social media text authored by individuals either self-reporting or labeled as exhibiting symptoms of various mental illnesses. Platforms such as Reddit, Twitter, and online mental health forums were used as primary data sources due to the richness and authenticity of user-generated content. The data was categorized into different classes such as depression, anxiety, bipolar disorder, PTSD, and control (non-mental illness) to enable multiclass classification. Publicly available datasets from previous mental health research were also used to enhance diversity and improve the reliability of training data.

2. Data Preprocessing:

Raw social media text is inherently noisy, containing emojis, hashtags, irregular punctuation, slang, spelling errors, and inconsistent formatting. Preprocessing is essential to normalize this data for model consumption. The following techniques were applied: lowercasing of all text, removal of stopwords and special characters, tokenization, and lemmatization to reduce words to their root forms. Additionally, emoji normalization and slang replacement dictionaries were applied to standardize informal language. Posts were filtered based on length to remove overly short or ambiguous entries. This step ensured that the text was cleaned, structured, and consistent for embedding and model input.

3. Text Representation:

To capture the deep semantics and contextual nuances of psychological language, the system employed advanced text representation models. Pretrained transformer models **Mental BERT** and **MelBERT** were used to convert each input sentence into dense vector embeddings. Mental BERT, trained specifically on mental health-related text, provided rich, domain-specific features, while MelBERT focused on understanding metaphorical and figurative language. These embeddings captured both literal and abstract expressions commonly used in mental health discourse. The combination of these embeddings ensured that the model could interpret complex user language more effectively than traditional one-hot or TF-IDF approaches.

4. Model Selection & Training:

A hybrid deep learning model was developed using a combination of **Convolutional Neural Networks (CNN)** and **Bidirectional Long Short-Term Memory (BiLSTM)** layers. The CNN layers were responsible for extracting local and hierarchical features from the embedding vectors, identifying important n-gram patterns and semantic chunks. These features were passed to the BiLSTM, which processed the data in both forward and backward directions, enabling the model to understand context from both past and future tokens. This sequential understanding is particularly important in identifying emotional tone and linguistic patterns over time. The model was trained using a stratified split of training and validation sets to ensure balanced learning across all mental illness classes.

5. Model Evaluation:

After training, the model was evaluated using standard classification metrics such as **accuracy**, **precision**, **recall**, and **F1-score** for each class. A **confusion matrix** was generated to analyze the distribution of true versus predicted labels, highlighting how well the model distinguished among the different categories. **Receiver Operating Characteristic (ROC) curves** and **Area Under the Curve (AUC)** values were also used to evaluate the model's ability to discriminate between classes. To ensure the model's robustness and generalizability, **k-fold cross-validation** was conducted. These evaluations confirmed the system's superiority over traditional models and validated its effectiveness in real-world use.

6. Model Prediction:

Once trained and validated, the final model was deployed to make predictions on new, unseen text inputs. Users' social media posts are first passed through the preprocessing and embedding stages before being classified by the trained CNN-BiLSTM model. The system outputs the predicted mental illness category along with a probability score, indicating the confidence of the prediction.

International Journal of Engineering Research and Science & Technology

This prediction mechanism can be integrated into web applications, health monitoring platforms, or used by researchers and mental health professionals for large-scale digital screenings. The real-time prediction capability ensures timely detection and potential early intervention.

IV. EXPERIMENTAL RESULTS AND ANALYSIS

Experimental results demonstrated that the model effectively classified textual data into multiple mental health-related categories with high prediction accuracy. The LSTM architecture successfully captured long-term contextual relationships in text, improving classification performance compared to traditional machine learning methods. The analysis showed that the proposed approach handled variations in writing style, vocabulary, and sentence structure while maintaining robust performance across different datasets.

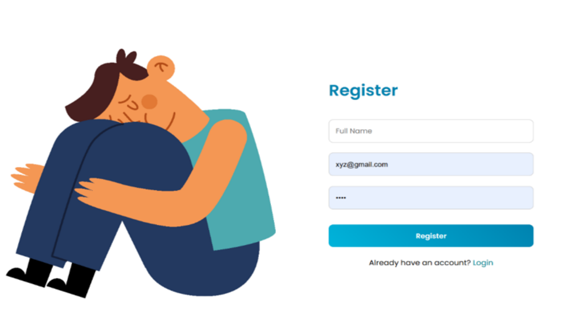

Fig. 2: Register

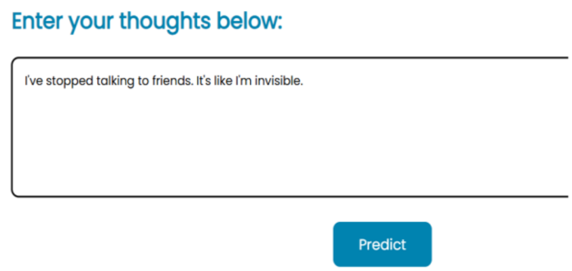

Fig. 3: Enter Text

Fig. 4: Predict Result

Fig. 5: Predict Result

V. CONCLUSION

This project presents a comprehensive deep learning-based framework for the prediction and classification of multiple mental health disorders using social media text. By leveraging advanced NLP techniques and domain-specific transformer models such as MentalBERT and MelBERT, the system effectively captures the subtle, emotional, and metaphorical nuances that are often present in online expressions of psychological distress. The integration of CNN and BiLSTM further enhances the model's ability to extract local features and understand the sequential flow of language, making it highly effective in identifying complex mental health patterns.

Through rigorous preprocessing, embedding, and evaluation strategies, the proposed system demonstrates superior performance over traditional models, offering improved accuracy, contextual understanding, and robustness. It successfully addresses the limitations of binary classification systems by supporting multiclass mental illness prediction, making it suitable for practical deployment in real-world mental health monitoring applications.

While the system shows promising results, there is room for future improvements, such as incorporating multimodal data, supporting multiple languages, adding explainability features, and deploying in a secure real-time application environment. Overall, this project highlights the transformative potential of artificial intelligence in supporting mental health awareness, early detection, and intervention, ultimately contributing to better mental well-being in society.

REFERENCES

[1] P. W. Corrigan and K. A. Kosyluk, ''Mental illness stigma: Types, constructs, and vehicles for change,'' in The Stigma of Disease and Disability: Understanding Causes and Overcoming Injustices, P. W. Corrigan, Ed., Washington, DC, USA: American Psychological Association, 2014, pp. 35–56.

[2] Y. Gan, H. Huang, X. Wu, and M. Meng, ''What doesn't kill us makes us stronger: Insights from neuroscience studies and molecular genetics,'' Current Opinion Behav. Sci., vol. 59, Oct. 2024, Art. no. 101431.

[3] C. Zhu, ''Computational intelligence-based classification system for the diagnosis of memory impairment in psychoactive substance users,'' J. Cloud Comput., vol. 13, no. 1, p. 119, Jun. 2024.

[4] T. A. Ghebreyesus. (2022). World Mental Health Report: Transforming Mental Health for All. Accessed: Feb. 4, 2024. [Online]. Available: https://iris.who.int/bitstream/handle/10665/356119/9789240049338eng.pdf?sequence=1

[5] P.Bower,S.Knowles,P. A.Coventry, N. Rowland, and C.C.M.D.Group, ‘‘Counselling for mental health and psychosocial problems in primary care,’’ Cochrane Database Systematic Rev., vol. 2011, no. 9, Sep. 1996, Art. no. CD001025.

[6] W.N.Stone,‘‘Treatmentofthechronically mentallyill:Anopportunityfor the grouptherapist,’’ Int. J. Group Psychotherapy, vol. 41, no. 1, pp. 11–22, Jan. 1991.ss

AUTHORS Profile

Mrs. D. KRISHNA VENI has received her B.Tech from JNTUH, Hyderabad in 2011, M.Tech degree in Computer science from JNTUH, Hyderabad in 2021 and Pursuing Ph.D degree in Computer Science from Guru Nanak university, Ibrahimpatnam. She is dedicated to teaching field from the last 8 years. Her research area Machine Learning. At present She is working as Assistant Professor in CSE Department at Guru Nanak Institutions Technical Campus, Ibrahimpatnam, Telangana, India.

Mr. A. SRAVAN KUMAR pursuing B.Tech degree in Computer Science and Engineering at Guru Nanak Institutions Technical Campus affiliated to JNTUH in 2022-2026.

Mr. B. KHAGESHWAR REDDY pursuing B.Tech degree in Computer Science and Engineering at Guru Nanak Institutions Technical Campus affiliated to JNTUH in 2022-2026.

Mr. J. NARESH pursuing B.Tech degree in Computer Science and Engineering at Guru Nanak Institutions Technical Campus affiliated to JNTUH in 2022-2026.

Mr. P. RAMAKRISHNA REDDY pursuing B.Tech degree in Computer Science and Engineering at Guru Nanak Institutions Technical Campus affiliated to JNTUH in 2022-2026.