International Journal of Engineering Research and Science & Technology

ISSN 2319-5991 www.ijerst.com
© 2026 IJERST. All Rights Reserved
Peer Reviewed, Referred & Indexed Journal

Research Paper

EXPLAINABLE DETECTION OF DEPRESSION IN SOCIAL MEDIA CONTENTS USING NATURAL LANGUAGE PROCESSING

ROJARAMANI ADAPA [1], KALVA SRI VARUN [2], BODAPATLA PRANITHA [3], BOINIPALLY SRIMAN [4], GATTU KOUSHIK [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— The widespread use of social media platforms provides valuable insights for real-time mental health monitoring, particularly for identifying signs of depression through user-generated content. This project introduces a deep learning-based approach for depression detection that leverages the sequential modeling capabilities of Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) networks. To address the challenges posed by informal language, the preprocessing pipeline includes emoji normalization, slang replacement using a custom dictionary, and the extraction of emotion scores to enhance semantic understanding. The hybrid LSTM-GRU model is trained on annotated social media posts, effectively capturing emotional and contextual patterns within the text.

Index Terms – Depression Detection, Explainable AI, Natural Language Processing, Social Media Analysis, Sentiment Analysis, Mental Health Monitoring, Machine Learning, Deep Learning, Text Mining, Explainable Models.

I. INTRODUCTION

Mental health disorders, particularly depression, have become a growing concern in today's digitally connected world. With the widespread use of social media platforms, users often express their thoughts, emotions,

and struggles online, either directly or indirectly. This has created an opportunity for researchers and developers to utilize social media as a data source for detecting early signs of depression. Traditional diagnostic methods are limited by accessibility, stigma, and lack of real-time monitoring. In this context, automated depression detection using Natural Language Processing (NLP) and deep learning techniques offers a scalable, non-intrusive, and continuous way to support mental health assessment. However, analyzing social media text presents unique challenges. Unlike formal writing, social media language is highly informal, filled with slang, abbreviations, emojis, and context-dependent expressions. These elements are often difficult for traditional machine learning models to interpret accurately. Existing systems that use transformer-based models like BERT paired with BiLSTM can capture rich semantic context, but they come with drawbacks such as high computational cost, slow inference time, and reduced interpretability. To address these challenges, this project proposes a lightweight yet powerful deep learning framework that combines Long Short-Term Memory (LSTM) and Gated Recurrent Units (GRU). This hybrid model is supported by a custom text preprocessing pipeline that includes emoji normalization, slang translation, and emotion score extraction to enhance the model's understanding of informal language. By leveraging the sequential learning capabilities of LSTM and GRU, the system is able to effectively analyze emotional tone and temporal patterns in social media posts. The result is a depression detection model that is accurate, fast, resource-efficient, and well-suited for deployment in real-world mental health monitoring applications.

II. RELATED WORK

A. Deep learning for depression detection from textual data

This paper presents a comprehensive study on the use of deep learning methods for detecting depression from textual data, particularly user-generated content from social media platforms. The authors evaluate and compare various neural architectures including CNNs and LSTMs, highlighting their effectiveness in learning contextual and semantic cues linked to depressive language. They incorporate pre-processing strategies and embedding techniques to improve detection accuracy and mitigate noise inherent in informal online communication. The study emphasizes the potential of deep learning in identifying subtle linguistic

patterns associated with mental health states and contributes to the development of scalable AI-driven mental health monitoring tools.

B. *Prediction of depression severity based on the prosodic and semantic features with bidirectional LSTM and time distributed CNN*

This research integrates both prosodic (intonation, rhythm) and semantic (meaning-based) features in audio and text to predict depression severity using a hybrid deep learning approach. By combining Bidirectional LSTM with Time-Distributed CNN, the authors design a system that captures both temporal dynamics and local patterns in speech and written data. Their model demonstrates high reliability in classifying different levels of depression severity, surpassing traditional ML techniques in accuracy and robustness. This dual-feature fusion provides valuable insights into multimodal data interpretation and sets a foundation for real-time AI-based depression assessment systems in clinical and non-clinical environments.

C. *Integrating BERT with CNN and BiLSTM for explainable detection of depression in social media contents*

This work introduces a hybrid model that integrates the contextual power of BERT with the sequential processing strength of BiLSTM and the spatial feature extraction capabilities of CNN to detect depression in social media posts. The primary focus is on building explainable AI (XAI) models that not only offer high prediction accuracy but also provide interpretability through attention visualization and feature attribution methods. The study presents experiments on multiple datasets, demonstrating that their BERT-CNN-BiLSTM framework outperforms existing baselines. The authors also discuss ethical considerations and the societal impact of deploying such models in mental health screening applications.

D. *Calibration of transformer-based models for identifying stress and depression in social media*

This paper explores the model calibration of transformer-based architectures such as BERT and RoBERTa to improve the reliability of depression and stress detection from social media content. Model calibration ensures that the predicted probabilities better reflect true likelihoods, which is vital for sensitive domains like mental health. The authors implement temperature scaling and isotonic regression to evaluate and adjust model confidence, thereby reducing false

positives/negatives. Their results demonstrate significant improvements in prediction reliability and trustworthiness, highlighting calibration as an often-overlooked yet essential component in deploying transformer models responsibly in the healthcare AI space.

III. PROPOSED METHODOLOGY

The proposed system aims to overcome the limitations of traditional and transformer-heavy approaches by introducing a lightweight, efficient, and accurate deep learning framework for depression detection in social media posts. It incorporates a hybrid deep learning model that combines Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) networks to effectively capture both long-term and short-term dependencies within textual data. This architecture enables the model to understand the sequential flow and emotional transitions within user-generated posts, which are often indicative of depressive behavior.

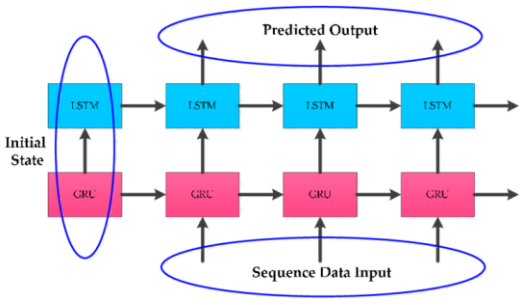


Fig. 1: System Overview

3.1 Dataset:

The dataset serves as the foundation for this project, containing labeled social media posts related to depression and control groups. Typically, this dataset will consist of text data such as tweets or social media comments, which are annotated with labels indicating whether they reflect depression-related content or not. For this project, a publicly available mental health dataset, such as the Mental Health in Social Media dataset, is used to ensure the training and evaluation are based on realistic and relevant data. This module focuses on ensuring data quality, relevance, and completeness for subsequent steps.

3.2 Analyzing:

In this module, exploratory data analysis (EDA) is performed to understand the characteristics of the dataset. EDA helps identify patterns, distributions, and imbalances in the data, such as the frequency of depression-related posts versus control posts. Visualization techniques like histograms, word clouds, and bar plots can be used to explore the data's structure, helping to uncover any biases or trends that could impact model performance. This analysis also guides preprocessing steps, such as how to handle class imbalances or missing data.

3.3 Cleaning the data:

Data cleaning is crucial to ensure that the input data is suitable for the model. This step involves removing irrelevant information, such as non-textual content, special characters, URLs, and user mentions (e.g., @username). It also includes handling issues like missing values and correcting or removing noisy data. Tokenization, lowercasing, and other basic text preprocessing steps are applied to simplify the text data and make it easier for the model to learn from. Proper cleaning improves the model's performance and prevents overfitting.

3.4 NLTK Processing:

The Natural Language Toolkit (NLTK) is employed for advanced text preprocessing. In this step, the text data is further processed to extract useful features. This includes tasks like stop-word removal (to eliminate common words that do not contribute to sentiment analysis), stemming (reducing words to their base form), and lemmatization (grouping different forms of a word). NLTK is also used to perform part-of-speech tagging, named entity recognition, and other NLP-specific tasks to enhance the feature set used in training the model. This step ensures that the data is semantically rich and ready for machine learning algorithms.

3.5 Data Splitting:

To avoid overfitting and ensure the model's generalizability, the dataset is split into training, validation, and testing subsets. Typically, the data is divided into 80% for training and validation, with 20% set aside for testing. The training and validation datasets are used to train and tune the model, while the testing dataset evaluates the final model's performance. This step is essential for assessing how well the model generalizes to unseen data and ensures that the results are reliable.

3.6 Algorithm Apply:

This module involves applying the selected algorithms, specifically LSTM (Long Short-Term Memory) and GRU (Gated Recurrent Unit), to the preprocessed data. Both of these algorithms are types of Recurrent Neural Networks (RNNs) that excel at capturing long-term dependencies in sequential data like text. The model is trained to learn patterns in the text that distinguish between depression-related and control posts. During training, the model adjusts its weights to minimize loss functions and improve its predictions.

3.7 Model Evaluation:

After the model is trained, it is evaluated on the test data using various performance metrics. Common evaluation metrics include accuracy, precision, recall, F1-score, and confusion matrix, which provide insights into the model's ability to correctly classify depression and control posts. These metrics help identify areas of improvement, such as whether the model has a bias towards one class or struggles with certain types of text.

3.8 Result:

The final results of the project are presented in this module, including the model's performance metrics and an analysis of its predictive capabilities. This step involves interpreting the model's outputs, visualizing the results through graphs and tables, and discussing the implications of the findings. The results can be compared against baseline models or other approaches to highlight the effectiveness of LSTM and GRU for depression detection. Moreover, the results will guide any potential improvements, future work, and practical applications of the model in real-world mental health monitoring.

IV. EXPERIMENTAL RESULTS AND ANALYSIS

The proposed explainable depression detection system was evaluated using metrics such as accuracy, precision, recall, F1-score, and Area Under the Curve (AUC). Experimental results showed that the model effectively identified depression-related patterns from social media text with high classification performance. By integrating Natural Language Processing techniques with Explainable Artificial Intelligence methods, the system not only achieved accurate predictions but also provided interpretable explanations for its decisions through important keywords and linguistic features. The analysis demonstrates that the proposed approach offers reliable, transparent, and scalable mental health monitoring using social media content.

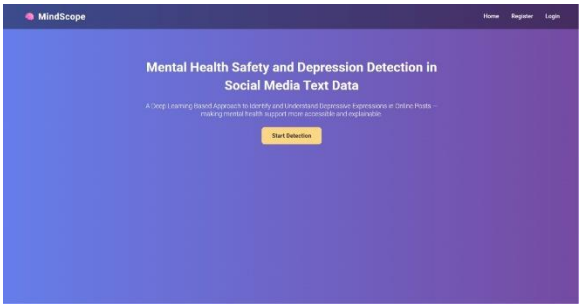


Fig. 2: Home Page

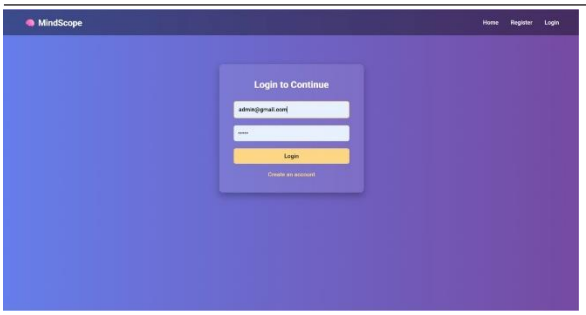


Fig. 3: Login

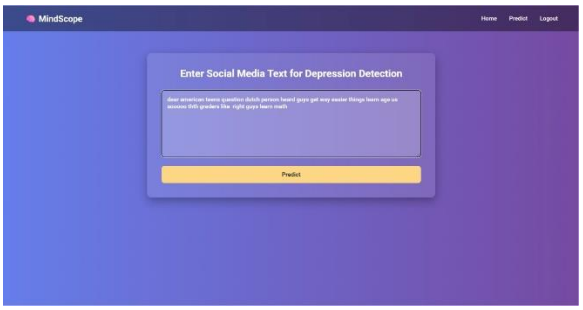


Fig. 4: Enter text

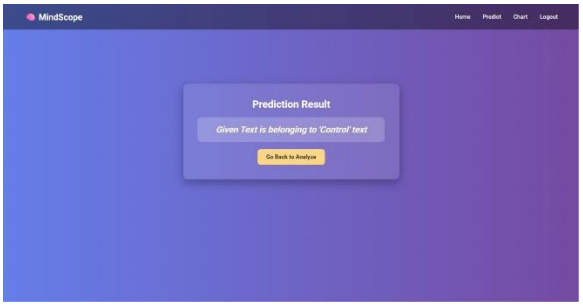


Fig. 5: Predict Result

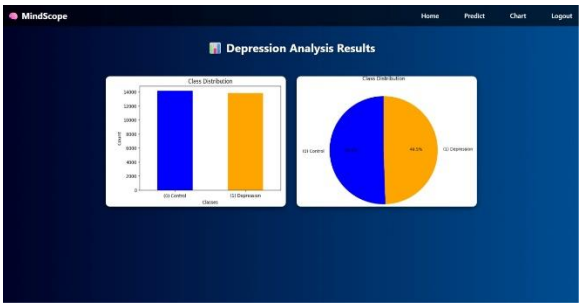


Fig. 6: Analysis Result

V. CONCLUSION

This project successfully demonstrates the application of advanced deep learning techniques, specifically Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) networks, for detecting depression from textual data. By analyzing language patterns in user-generated content, the model effectively identifies emotional cues that are indicative of depressive tendencies. Compared to traditional machine learning approaches, the LSTM-GRU architecture captures both short- and long-term dependencies in text, resulting in improved prediction accuracy. The system holds significant potential in supporting early detection efforts, enabling timely mental health intervention. Overall, this work not only showcases the strength of neural networks in sentiment analysis but also contributes to the broader goal of using AI for the betterment of mental health care.

REFERENCES

[1] P. Mehta, ‘‘Pioneering the future of mental health with AI: Exploring promising directions,’’ in Precision Health in the Digital Age: Harnessing AI for Personalized Care. Hershey, PA, USA: IGI Global, 2025, pp. 277–304.

[2] S. D'Alfonso, ''AI in mental health,'' Current opinion Psychol., vol. 36, pp. 112–117, Jun. 2020.

[3] D. F. Hunt, J. Bailey, B. R. Lennox, M. Crofts, and C. Vincent, ''Enhancing psychological safety in mental health services,'' Int. J. Mental Health Syst., vol. 15, no. 1, pp. 1–18, Dec. 2021. [4] M. Edgelow, A. Fecica, C. Kohlen, and K. Tandal, ''Mental health of public safety personnel: Developing a model of operational, organizational, and personal factors in public safety organizations,'' Frontiers Public Health, vol. 11, Mar. 2023, Art. no. 1140983, doi: 10.3389/fpubh.2023.1140983.

[5] Bhagwat, V. B. (2025). Simplifying Payroll Balance Conversions in Payroll Systems Implementation through the Use of Generative AI.

[6] S. Nochaiwong, C. Ruengorn, K. Thavorn, B. Hutton, R. Awiphan, C. Phosuya, Y. Ruanta, N. Wongpakaran, and T. Wongpakaran, ''Global prevalence of mental health issues among the general population during the coronavirus disease-2019 pandemic: A systematic review and meta-analysis,'' Sci. Rep., vol. 11, no. 1, p. 10173, May 2021, doi: 10.1038/s41598-021-89700-8.

AUTHORS Profile



Mrs. ROJARAMANI ADAPA, received her M.Tech (CSE) from JNTUH. She has teaching experience of 20 years and has guided many undergraduate and postgraduate projects. She has published about 10 research articles in peer-reviewed journals and presented 5 research papers at various national conferences. At present, she is working as an Assistant Professor in CSE Department at Guru Nanak Institutions Technical Campus, Ibrahimpatnam, Telangana, India.



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



Ms. BODAPATLA PRANITHA pursuing B. Tech degree in Computer Science and Engineering at Guru Nanak Institutions Technical Campus affiliated to JNTUH in 2022-2026.



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



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