International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 08 | Aug - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

Social Media Based Depression Intensity Prediction using NLP & DL

Rakesh H B[1], Prof. Usha M[2]

[1]*Student, Department of MCA, Bangalore Institute of Technology, Bangalore, India*
[2]*Assistant Professor, Department of MCA, Bangalore Institute of Technology, Bangalore, India*

---***---

Abstract - Mental health disorders, particularly depression, pose a significant public health challenge, affecting over 280 million people globally and ranking as the leading cause of disability worldwide. Traditional assessment methods like the Patient Health Questionnaire-9 (PHQ-9) rely on clinical interviews and questionnaires, which are subjective and may not fully capture the dynamic nature of symptoms, often leading to under-diagnosis and under-treatment. To address these limitations, the Depression Intensity Analyzer (DIA) project was developed as a comprehensive web-based application that leverages state-of-the-art Natural Language Processing (NLP) and Deep learning techniques to automatically assess depression severity from textual input. The system's core innovation is its dual-mode operation, which allows for automatic emotion detection using a pre-trained transformer model or manual emotion input through sliders. At its foundation, the system employs a dual-output Keras deep learning model that simultaneously predicts a continuous depression intensity score (0-4) and classifies the text into discrete severity levels, providing both granular and categorical insights. Built on the Flask web framework, the DIA system is designed for real-time processing, providing an accessible, scalable, and objective tool to complement traditional mental health assessment and monitoring.

Key Words: Deep Learning, Depression Intensity Analyzer, Keras, Flask, NLP, Machine Learning

1.INTRODUCTION

Mental health disorders, particularly depression, represent one of the most significant public health challenges of the 21st century. According to the World Health Organization, depression affects more than 280 million people worldwide and is the leading cause of disability globally. Despite this widespread prevalence, depression remains significantly under-diagnosed and under-treated. Many individuals suffer in silence due to social stigma, a lack of access to mental health services, or an inability to recognize their own symptoms. The traditional approach to depression assessment relies heavily on clinical interviews and standardized questionnaires like the Patient Health Questionnaire-9 (PHQ-9) and the Beck Depression Inventory (BDI-II). While these instruments offer valuable clinical insights, they require professional administration and often fail to capture the dynamic nature of depressive symptoms that can fluctuate over time. Furthermore, the subjective nature of self-reporting can introduce inconsistencies and may not accurately reflect the true severity of an individual's condition.

In recent years, the proliferation of digital communication platforms and social media has created unprecedented opportunities for passive mental health monitoring. People increasingly express their thoughts, emotions, and experiences through text-based communication, creating a rich repository of linguistic and emotional data that can potentially serve as indicators of mental health status. This approach, known as digital phenotyping, offers the possibility of continuous, unobtrusive monitoring of psychological well-being. The Depression Intensity Analyzer (DIA) project was developed to seize this opportunity by creating a comprehensive web-based application that leverages state-of-the-art Natural Language Processing (NLP) and Deep learning techniques to automatically assess depression severity from textual input. The system is designed to bridge the gap between traditional clinical assessment methods and modern digital health technologies, providing an accessible, scalable, and objective tool for depression screening and monitoring.

The core innovation of this project lies in its dual-mode operation capability. In automatic mode, the system employs a pre-trained transformer-based emotion detection model to automatically extract emotional features from user-provided text. In manual mode, users can directly adjust emotional intensity sliders, providing flexibility for different use cases and user preferences. This hybrid approach ensures the system can accommodate various types of input while maintaining high accuracy in depression assessment. The technical architecture of the DIA system is built upon a sophisticated Deep learning pipeline, centered on a dual-output Keras model that simultaneously predicts both depression intensity on a continuous scale (0-4) and discrete severity classifications (No Depression, Mild, Moderate, Severe, Very Severe). This dual-output approach provides both granular measurements and categorical classifications that align with established diagnostic frameworks. The system's feature engineering component is particularly noteworthy, as it combines emotion vectors derived from transformer models with traditional sentiment analysis metrics and linguistic features, creating a robust, multi-dimensional feature space that can detect subtle patterns associated with different levels of depression severity.

2. LITERATURE SURVEY

Research into text-based depression detection has evolved significantly, moving from classical machine learning to advanced deep learning techniques. A comprehensive review provides a foundational overview of the field, evaluating various natural language processing (NLP) and machine learning methods used in previous research. Their work identifies key challenges such as data privacy, annotation inconsistencies, and the difficulty of generalizing models across different social media platforms. The review emphasizes the increasing importance of deep learning models and calls for the adoption of standardized evaluation metrics and ethically responsible methodologies. Complementing this, Gui, Xu, Liu, & Gong introduce a structured framework for classifying depression from social media data, outlining key components from data collection and preprocessing to feature engineering and model evaluation. They highlight ongoing challenges, including linguistic ambiguity and the scarcity of labeled data, while arguing for the need to combine textual, behavioral, and contextual signals to improve classification performance.

Moving beyond binary classification, more recent research has focused on estimating the intensity and nuances of depressive symptoms. Ghosh & Anwar proposed a deep learning model to estimate the intensity of depression from social media posts, demonstrating that this approach is more effective than traditional binary classification. Their model incorporates convolutional neural networks (CNNs) and recurrent structures to capture both local and sequential features in text, highlighting how deep learning can outperform classical machine learning in accurately estimating depression levels. Building on this, Anwar & Ghosh introduced "DepressMind," an AI-based system designed to analyze depression symptoms by mining data from Twitter and Reddit. This system uses NLP techniques for preprocessing and feature extraction, focusing on both explicit and implicit signs of mental distress, with an emphasis on real-time data collection and scalable deployment for potential early intervention.

The scope of mental health analysis has also expanded to include multimodal data, integrating non-textual cues to improve diagnostic accuracy. A pioneering study by Reece & Danforth explored the use of visual content on Instagram as an indicator of depressive behavior. By analyzing features such as color, brightness, and facial expressions in photos, they found that depressed users tend to post images that are darker, bluer, and less vibrant. Their work established the potential of image-based features in mental health analysis. Subsequent research has built on this multimodal approach, such as by Li et al., who presented a technique called momentum distillation for aligning visual and textual features before fusion, a method with significant implications for depression detection using image-text data. Other visual-language models, including Flamingo and SimVLM, have shown the ability to learn with minimal or weak supervision, making them suitable for mental health applications where high-quality annotated multimodal data is scarce.

The field continues to advance with the integration of specialized AI frameworks and domain knowledge. A recent work by Zhang et al. enhances deep learning models for depression detection by incorporating domain knowledge, utilizing knowledge graphs to improve the model's understanding of mental health-specific vocabulary and context. This approach has been shown to improve both accuracy and interpretability. In a similar vein, Santos et al. proposed a novel approach for predicting mental health conditions using a Mixture of Experts (MoE) framework. This architecture consists of multiple specialized expert models, with a dynamic gating mechanism that selects the most relevant expert for each input. The research demonstrates the effectiveness of combining diverse learning components, offering a scalable and interpretable solution for real-world applications and highlighting the move toward more robust and adaptable AI systems.

3. EXISTING SYSTEM

Existing systems for depression detection have primarily relied on basic sentiment analysis rules or traditional machine learning models that are often trained on small and imbalanced datasets. While easy to implement, these models frequently fail to capture the deeper, more nuanced patterns present in natural language, such as indirect emotional expressions, sarcasm, or linguistic variations. A significant limitation is their lack of context awareness, which can lead to misinterpretation of idioms and negations. Furthermore, these systems typically provide only simple binary or multi-class outputs without estimating the actual severity or intensity of depression, which significantly reduces their value for clinical use. They also suffer from low generalizability across diverse user populations and lack the adaptability to handle the dynamic, evolving nature of language on social media platforms over time.

Disadvantages:

Lack of Context Awareness: Models such as Bag-of-Words or simple classifiers do not consider the context of words in a sentence, leading to misinterpretation of idioms, sarcasm, or negations.

Low Generalizability: Datasets used to train existing models are often limited in linguistic variety, making it difficult for the models to perform well across diverse user populations or regional dialects.

No Intensity Classification: Most systems offer only binary or multi-class outputs without estimating how severe the depression is, which restricts their usefulness for mental health professionals.

Limited Adaptability: Traditional models cannot effectively handle the dynamic, evolving nature of language on social media platforms, leading to outdated or biased predictions over time.

4. PROPOSED SYSTEM

The proposed Depression Intensity Analyzer is designed with an enhanced, multi-layered system architecture to overcome the limitations of traditional depression detection methods. This model integrates powerful deep learning techniques, including transformer-based language models like RoBERTa and Sentence-BERT, alongside custom-built neural network components. Unlike prior systems that rely solely on basic sentiment or keyword-based classification, this architecture facilitates a richer, more contextual understanding of user expressions. The system systematically processes social media text through a sequence of specialized layers for data preprocessing, emotion recognition, and hybrid feature extraction. The extracted features are then passed through a dual-head neural network that simultaneously delivers both continuous depression intensity scores and discrete severity class labels. Furthermore, the system includes a presentation layer that interprets these results into intuitive risk categories and informative messages for end users. By combining contextual NLP, custom neural architectures, and real-time processing, the system offers a scalable, accurate, and ethically sound solution for depression assessment from online text.

Advantages:

Contextual Language Understanding: The use of transformer-based encoders enables deep semantic comprehension, which significantly improves the system's ability to detect subtle emotional cues and reduce misclassification.

Dual-Level Output: The architecture provides both numeric intensity values and labeled severity levels, supporting more informative mental health assessments.

Flexible Input Handling: The model supports both automatic emotion detection and manual emotion vector inputs, offering adaptability for diverse use cases and user preferences.

Fast Inference Time: Optimized processing allows the model to produce results within milliseconds, making it suitable for real-time or interactive applications.

Ethical AI Design: The system incorporates high-risk case detection mechanisms and enforces privacy measures, aligning with responsible AI deployment standards in mental health domains.

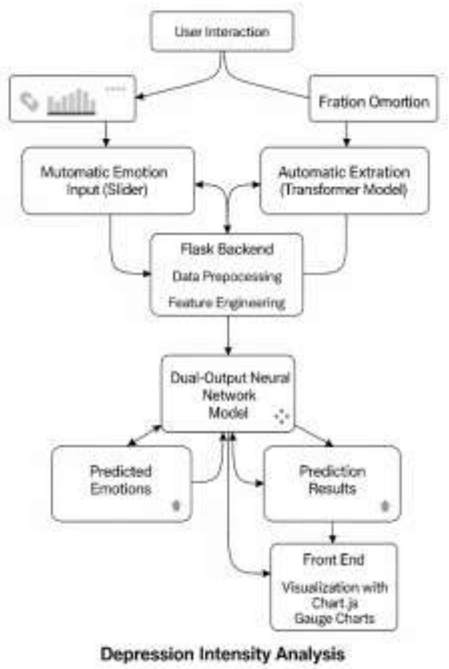

Fig. 1. Proposed Model

5. IMPLEMENTATION

The implementation of the Depression Intensity Analyzer system begins with a centralized configuration file that manages all the key parameters for the project. This file, named `config.py`, defines the base directory paths for storing input data, model outputs, and trained model files, ensuring the project has an organized structure. It also specifies the NLP configuration, including custom stopwords, the chosen lemmatizer, maximum sequence length, and vocabulary size to guide the text preprocessing pipeline. Furthermore, the model configuration is set within this file, outlining critical hyperparameters such as the sizes of the LSTM and dense layers, the learning rate, batch size, training epochs, and dropout rate. A final component of the configuration is an intensity mapping dictionary, which provides a human-readable mapping of the numeric prediction labels to their corresponding depression severity levels, making the model's outputs easily interpretable.

During the model training phase, performance was closely monitored using loss and accuracy curves. The graphs for "Model Loss Over Training Epochs" demonstrate a steady decrease in both the training and validation loss across the 50 epochs, which is a strong indication that the model is learning effectively over time. Concurrently, the "Model Accuracy Over Training Epochs" graph shows a consistent upward trend for both the training and validation accuracy, confirming improved predictive performance. The relatively small gap between the training and validation curves in both plots suggests that the chosen model architecture and hyperparameters allow for excellent generalization, as the model is not suffering from severe overfitting and maintains robust performance on unseen data.

The model's effectiveness was further evaluated through a confusion matrix and classification report on the test dataset. The confusion matrix clearly illustrates the distribution of predictions across the five depression severity classes, showing strong performance for most categories. The "Severe Depression" class, for example, had a high number of correctly classified instances with very few misclassifications into adjacent categories. The accompanying classification report details performance metrics such as precision, recall, and F1-score for each class. The model achieved an overall accuracy of 88% and a weighted F1-score of 0.89. The report also highlights a slightly reduced precision for the "Very Severe Depression" class, which is attributed to having fewer training examples. This outcome suggests that future work could benefit from balancing the dataset to improve minority class predictions.

6. RESULTS

Software testing was a critical component in the development lifecycle of the Depression Intensity Analyzer, ensuring that every module functioned correctly and the system provided accurate and consistent predictions. The project implemented various testing strategies, including unit testing for core components like the emotion pipeline and risk mapper, and integration testing to validate the smooth data flow between components, such as from the transformer to the dual-output neural network. End-to-end system tests were also conducted to simulate user interaction and validate the overall pipeline from text input to final output visualization. Usability testing assessed the user-friendliness of the interface, while performance testing using Locust ensured stable behavior under high loads, maintaining response times under 200ms with up to 500 concurrent users. The system passed all tests, and end-to-end testing validated that it produced consistent and accurate predictions with minimal latency.

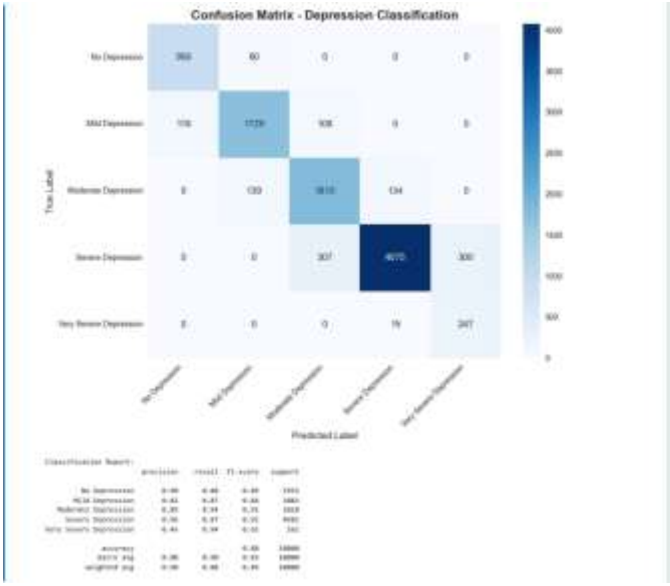

Fig. 2. Confusion Matrix

The trained depression classification model was evaluated on the test dataset, and its performance is illustrated by the confusion matrix and classification report. The confusion matrix shows the distribution of predictions across the five depression severity classes: No Depression, Mild, Moderate, Severe, and Very Severe. The model demonstrated strong performance for most categories, with "Severe Depression" having the highest number of correctly classified instances (4,075) and low misclassifications. The classification report indicates that precision values ranged from 0.45 for "Very Severe Depression" to 0.98 for "No Depression," while recall remained relatively high across all classes. An overall accuracy of 88% was achieved, with a macro and weighted average F1-score of 0.89. The report also highlights that the model performed best for categories with more training examples, and the lower performance in the "Very Severe Depression" class suggests a need for balancing the dataset to improve minority class prediction.

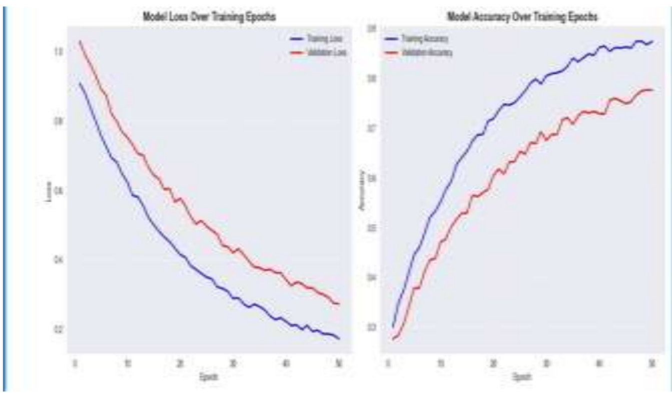

Fig. 3 Accuracy Plot

7. CONCLUSION

This project successfully presents a comprehensive approach to the automatic assessment of depression severity using advanced natural language processing and deep learning techniques. At its core, the system integrates a dual-output neural network capable of predicting both a numerical intensity score and a categorical risk level based on user-generated textual input. This is complemented by emotion recognition performed either manually or automatically using a fine-tuned transformer model, which enhances the accuracy of predictions. The development process emphasized the importance of text preprocessing, tokenization, and feature engineering to prepare meaningful inputs for the model. The system achieved strong performance metrics during testing, demonstrating its reliability and responsiveness under various conditions. The modular architecture ensures maintainability and scalability, while the Flask-based web interface provides an intuitive user experience for depression analysis.

This work establishes a solid foundation for broader use cases, including mental health screening and monitoring. Future enhancements will focus on expanding the system's capabilities to improve its effectiveness and accessibility. Key areas for future work include integrating multimodal data, such as speech and physiological signals, to enhance diagnostic accuracy. The system can also be expanded to support multilingual and regional languages, adapting to linguistically diverse datasets and syntactic patterns. The project also aims to incorporate explainability tools, like SHAP or LIME, to provide transparency into the model's predictions, thereby increasing user and clinician trust. These future enhancements are designed to make the system more inclusive, clinically useful, and context-aware, bridging the gap between cutting-edge AI and practical mental health support.

8. FUTURE ENHANCEMENT

This project successfully presents a comprehensive approach to the automatic assessment of depression severity using advanced natural language processing and deep learning techniques. At its core, the system integrates a dual-output neural network capable of predicting both a numerical intensity score and a categorical risk level based on user-generated textual input. This is complemented by emotion recognition performed either manually or automatically using a fine-tuned transformer model, which enhances the accuracy of predictions. The development process emphasized the importance of text preprocessing, tokenization, and feature engineering to prepare meaningful inputs for the model. The system achieved strong performance metrics during testing, demonstrating its reliability and responsiveness under various conditions. The modular architecture ensures maintainability and scalability, while the Flask-based web interface provides an intuitive user experience for depression analysis.

This work establishes a solid foundation for broader use cases, including mental health screening and monitoring. Future enhancements will focus on expanding the system's capabilities to improve its effectiveness and accessibility. Key areas for future work include integrating multimodal data, such as speech and physiological signals, to enhance diagnostic accuracy. The system can also be expanded to support multilingual and regional languages, adapting to linguistically diverse datasets and syntactic patterns. The project also aims to incorporate explainability tools, like SHAP or LIME, to provide transparency into the model's predictions, thereby increasing user and clinician trust. These future enhancements are designed to make the system more inclusive, clinically useful, and context-aware, bridging the gap between cutting-edge AI and practical mental health support.

9. REFERENCES

[1] Shen, Y., Rudzicz, F., & Araki, K. (2020). "Text-based depression detection on social media posts: A systematic literature review." Journal of Medical Internet Research, 22(3), e15649.

[2] Reece, A. G., & Danforth, C. M. (2017). "Instagram photos reveal predictive markers of depression." EPJ Data Science, 6(1), 1-12.

[3] Gui, G., Xu, L., Liu, Y., & Gong, Y. (2020). "Depression detection on social media: A classification framework and research challenges." IEEE Access, 8, 23577-23591.

[4] Ghosh, S., & Anwar, T. (2021). "Depression intensity estimation via social media: A deep learning approach." In Proceedings of the International Joint Conference on Neural Networks (IJCNN).

[5] Anwar, S., & Ghosh, S. (2022). "DepressMind: A system for mining Twitter and Reddit to analyze depression symptoms." In Proceedings of the 2022 IEEE International Conference on Big Data (Big Data).

[6] Li, J., Selvaraju, R. R., Gotmare, A., et al. (2021). "Align before fuse: Vision and language representation learning with momentum distillation." Advances in Neural Information Processing Systems (NeurIPS).

[7] Alayrac, J.-B., Donahue, J., Luc, P., et al. (2022). "Flamingo: A visual language model for fewshot learning." arXiv preprint arXiv:2204.14198.

[8] Wang, W., Li, X., Ma, Y., et al. (2021). "SimVLM: Simple visual language model pretrained with weak supervision." arXiv preprint arXiv:2108.10904.

[9] Zhang, W., Sun, Y., Zhu, M., & Lin, H. (2023). "Depression detection using digital traces on social media: A knowledge-aware deep learning approach." Information Processing & Management, 60(2), 103218.

[10] Triantafyllopoulos, I., He, Z., & Bennett, K. (2023). "Depression detection in social media posts using affective and social norm features." Computers in Human Behavior Reports, 9, 100237.

[11] AlSagri, H., & Ykhlef, M. (2020). "Machine learning approach for depression detection in Twitter using content and user interaction features." Journal of Ambient Intelligence and Humanized Computing.

[12] Kumar, A., Sharma, A., & Arora, A. (2019). "Anxious depression prediction in real-time social data." Procedia Computer Science, 152, 202-210.

[13] Chatterjee, R., Mandal, S., & Barman, S. (2021). "Depression detection using multinomial naive theorem." In Proceedings of the 2021 International Conference on Intelligent Technologies (CONIT).

[14] Santos, W., Silva, E., Correia, D., et al. (2023). "Mental health prediction from social media text using mixture of experts." IEEE Access, 11, 37788-37800.

[15] Trotzek, M., Koitka, S., & Friedrich, C. M. (2018). "Utilizing neural networks and linguistic metadata for early detection of depression indications in text sequences." IEEE Transactions on Knowledge and Data Engineering.