

PsyAI-Net: An Intelligent Hybrid Machine Learning Framework for Early Mental Health Risk Prediction Using Social Media Text Analytics

Mr. Dr.M.Veerabhadr Rao¹, Munasa Satya Bhaskar²

¹Professor, ²M. tech Student,

Department of Computer Science & Engineering(AI), Pydah College of Engineering,
Yanam Road, Tallarevu, Patavala, Andhra Pradesh, 533461.

Abstract- The increasing use of social media platforms has created vast amounts of user-generated textual data that reflect personal emotions, thoughts, and behavioural patterns. These digital footprints provide valuable insights into an individual's psychological state and can be leveraged for early detection of mental health conditions. However, traditional mental health assessment methods rely heavily on clinical interviews and self-reported questionnaires, which may not always provide timely or scalable solutions. This study proposes an intelligent hybrid machine learning framework for early mental health risk prediction using social media text analytics. The system integrates conventional machine learning models and deep learning architectures to perform multiclass classification of mental health conditions such as anxiety, depression, stress, and other psychological states. The framework incorporates comprehensive text preprocessing techniques, including cleaning, tokenization, stop-word removal, and feature extraction using advanced vectorization methods. Multiple classifiers such as Support Vector Machines (SVM), Random Forest, Logistic Regression, XGBoost, and a hybrid BiLSTM-CNN deep learning model are implemented and evaluated. To enhance performance, the proposed system applies hyperparameter optimization and dynamic model selection strategies. Experimental results demonstrate that the hybrid framework achieves high predictive accuracy and balanced performance across precision, recall, and F1-score metrics. The system provides a scalable and automated approach for mental health analysis, offering potential support for early intervention and preventive healthcare strategies.

Keywords – Mental Health Prediction, Social Media Analytics, Machine Learning, Deep Learning, Text Classification, Hybrid Framework, BiLSTM-CNN, XGBoost, Natural Language Processing (NLP), Multiclass Classification, Early Risk Detection.

I. INTRODUCTION

Mental health disorders have become a growing global concern, affecting millions of individuals across different age groups and social backgrounds. Conditions such as depression, anxiety, stress-related disorders, and mood instability significantly impact an individual's quality of life, productivity, and overall well-being. Despite increasing awareness, early identification and timely intervention remain major challenges due to social stigma, limited access to mental health professionals, and the subjective nature of traditional diagnostic methods. With the widespread adoption of social media platforms, individuals increasingly express their emotions, opinions, and personal experiences online. These digital interactions generate large volumes of unstructured textual data that can reveal subtle indicators of psychological distress. Social media content often reflects emotional states through language patterns, word usage, sentiment expressions, and behavioural tendencies. As a result, analysing such data

provides an opportunity to develop automated systems capable of identifying early signs of mental health risks.

Recent advancements in Natural Language Processing (NLP), machine learning, and deep learning have enabled the development of intelligent systems for text-based classification and sentiment analysis. Machine learning models such as Support Vector Machines (SVM), Random Forest, and Logistic Regression have been widely used for mental health prediction tasks. Additionally, deep learning architectures including Convolutional Neural Networks (CNN) and Bidirectional Long Short-Term Memory (BiLSTM) networks have demonstrated strong capability in capturing contextual and sequential dependencies in textual data. However, relying solely on a single model may limit predictive performance and generalization ability. Motivated by these challenges, this work proposes an intelligent hybrid machine learning framework for multiclass mental health risk prediction using social media text data.

The proposed system integrates traditional machine learning algorithms with advanced deep learning architectures to improve classification accuracy and robustness. Furthermore, the framework incorporates systematic preprocessing, feature extraction, hyperparameter tuning, and dynamic model selection strategies to enhance overall system efficiency. The primary objective of this study is to design a scalable, data-driven solution capable of identifying potential mental health conditions at an early stage. By leveraging computational intelligence techniques, the proposed framework aims to support mental health monitoring systems and contribute toward proactive and preventive healthcare approaches.

II. LITERATURE SURVEY

The rapid growth of social media platforms has encouraged researchers to explore their potential for mental health assessment and prediction. Over the past decade, several studies have investigated the relationship between linguistic patterns in online posts and psychological conditions such as depression, anxiety, stress, and suicidal ideation. Early research primarily focused on sentiment analysis techniques to classify user emotions based on polarity scores derived from textual content. Traditional machine learning approaches have been widely applied for mental health classification tasks. Algorithms such as Support Vector Machines (SVM), Naïve Bayes, Logistic Regression, and Random Forest have demonstrated moderate success in identifying depressive and anxiety-related posts. These models typically rely on handcrafted features, including term frequency-inverse document frequency (TF-IDF), bag-of-words representations, n-gram analysis, and sentiment lexicons.

While these methods are computationally efficient and interpretable, their performance heavily depends on feature engineering and may fail to capture deeper contextual relationships within text data. With the advancement of deep learning, researchers have increasingly adopted neural network architectures for mental health prediction. Convolutional Neural Networks (CNN) have been utilized to extract local semantic features from text, while Recurrent Neural Networks (RNN) and Long Short-Term Memory (LSTM) networks have shown effectiveness in modelling sequential dependencies in language data. Bidirectional LSTM (BiLSTM) models further improve contextual understanding by processing textual sequences in both forward and backward directions. These deep learning approaches generally outperform traditional machine learning models, especially when trained on large-scale datasets. More recent studies have explored hybrid and ensemble learning techniques to enhance predictive performance. Combining machine learning and deep learning models has demonstrated improved robustness and classification accuracy. Additionally, hyperparameter optimization and model selection strategies have been

introduced to fine-tune model architectures for better generalization.

Despite these advancements, several challenges remain. Many existing systems focus primarily on binary classification (e.g., depressed vs. non-depressed) rather than addressing multiclass mental health conditions. Furthermore, some studies lack systematic preprocessing and balanced dataset strategies, which can negatively affect model reliability. There is also a need for scalable and adaptable frameworks that integrate multiple algorithms to determine the most effective model dynamically. Motivated by these research gaps, the proposed work introduces a hybrid machine learning framework that integrates both traditional and deep learning models for multiclass mental health risk prediction. By combining advanced text preprocessing, optimized model training, and comparative evaluation strategies, this study aims to improve predictive accuracy and contribute to the development of intelligent mental health analytics systems.

III. SYSTEM ANALYSIS

A. Existing System

Existing mental health prediction systems based on social media data primarily rely on traditional machine learning algorithms and basic sentiment analysis techniques. Most earlier approaches utilize text classification models such as Naïve Bayes, Support Vector Machines (SVM), Logistic Regression, Decision Trees, and Random Forest classifiers. These models typically operate on handcrafted features extracted using methods like Bag-of-Words, n-grams, or TF-IDF representations. Some recent studies have incorporated deep learning architectures such as Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) networks to improve contextual understanding of text data. While these approaches demonstrate improved predictive accuracy, they often focus on binary classification tasks, such as identifying depression versus non-depression.

Additionally, many systems employ a single-model strategy without exploring hybrid or ensemble mechanisms. In most traditional implementations, preprocessing steps are limited, and model optimization techniques are not extensively applied. As a result, system performance may vary significantly depending on dataset quality and size. Furthermore, several existing solutions lack scalability and adaptability when applied to multiclass mental health classification problems.

DISADVANTAGES OF THE EXISTING SYSTEM

- **Limited Feature Representation:** Traditional machine learning models rely heavily on manually engineered textual features, which may fail to capture deeper semantic and contextual relationships present in social media data.

- **Binary Classification Limitation:** Many existing systems focus only on distinguishing between two classes (e.g., depressed vs. non-depressed), limiting their applicability for detecting multiple mental health conditions simultaneously.
- **Overfitting and Underfitting Issues:** Improper model tuning may cause overfitting, where the system memorizes training data instead of learning general patterns, or underfitting, where it fails to capture complex linguistic structures.
- **Lack of Hybrid Integration:** Most systems employ a single algorithm, which may not fully utilize the strengths of different machine learning and deep learning approaches.
- **Scalability Concerns:** As social media data grows exponentially, models must efficiently handle large volumes of unstructured text without compromising performance.
- **Imbalanced Dataset Challenges:** Mental health datasets often contain unequal class distributions, which can negatively affect model accuracy if not properly addressed.
- **Computational Complexity:** Advanced deep learning models may require significant computational resources, limiting real-time deployment capabilities in certain environments.

B. Proposed System

To overcome the limitations of existing methods, this study proposes an intelligent hybrid machine learning framework for multiclass mental health risk prediction using social media text data. The system begins with comprehensive text preprocessing, including data cleaning, tokenization, stop-word removal, stemming, and vectorization techniques such as TF-IDF and word embeddings. The processed dataset is then divided into training and testing subsets to ensure unbiased model evaluation. Multiple machine learning models, including Support Vector Machines, Random Forest, Logistic Regression, XGBoost, and AdaBoost, are implemented alongside a deep learning architecture combining BiLSTM and CNN layers.

This hybrid approach leverages both traditional statistical learning methods and advanced neural network capabilities to improve predictive performance. Hyperparameter optimization techniques are applied to enhance model efficiency and reduce overfitting. Cross-validation strategies ensure reliability and robustness across different data splits. The system dynamically selects the best-performing model based on evaluation metrics. Performance is assessed using multiple criteria, including accuracy, precision, recall, F1-score, and ROC-AUC. By integrating diverse algorithms within a unified framework, the proposed system enhances classification accuracy, generalization ability, and scalability for real-world mental health analytics applications.

IV. SYSTEM DESIGN

SYSTEM ARCHITECTURE

Below diagram depicts the whole system architecture.

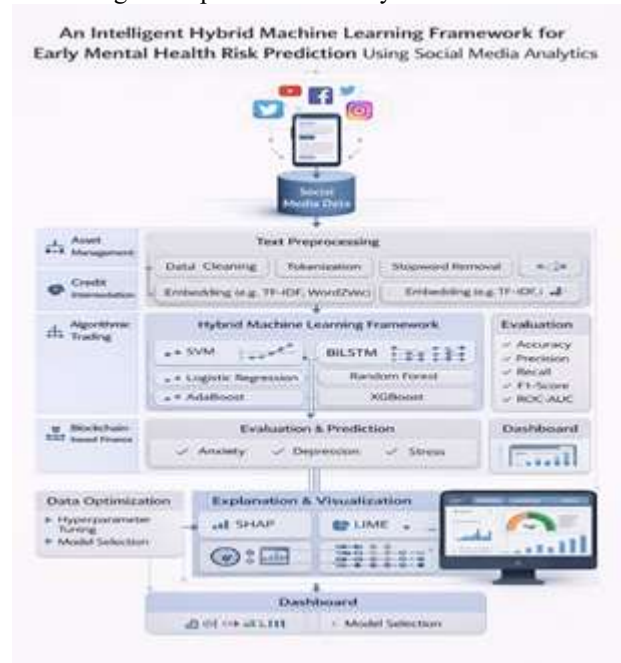


Fig 1. Methodology followed for proposed model

V. SYSTEM IMPLEMENTATION

Modules

1. Data Collection and Text Preprocessing

The first stage of the system involves collecting publicly available social media text data relevant to mental health discussions. Since raw textual data often contains noise such as emojis, URLs, special symbols, slang, and redundant content, a structured preprocessing pipeline is implemented. This includes text cleaning, tokenization, stop-word removal, lemmatization, and normalization. Additionally, class imbalance issues are addressed using appropriate resampling techniques to ensure fair model learning across different mental health categories.

2. Feature Extraction and Representation

After preprocessing, meaningful textual features are extracted to represent user-generated content in a structured format suitable for machine learning models. Techniques such as TF-IDF vectorization, word embeddings (Word2Vec/GloVe), and contextual embeddings (e.g., transformer-based representations) are explored. Sentiment scores, linguistic markers, and psychological lexicon-based features are also incorporated to enhance semantic understanding and improve predictive performance.

3. Hybrid Model Development and Training

The proposed framework integrates multiple machine learning and deep learning algorithms to create a hybrid prediction system. Traditional classifiers such as Logistic Regression, Support Vector Machines, and Random Forest are combined with deep learning architectures such as LSTM and transformer-based models. Ensemble learning strategies, including soft voting and stacking, are employed to improve robustness and generalization. Hyperparameter tuning is conducted using cross-validation techniques to optimize model performance.

4. Early Risk Prediction and Alert Generation

Once trained, the model is deployed to analyse incoming social media text in near real-time. The system predicts mental health risk categories such as depression risk, anxiety indicators, or stress-related patterns. Based on prediction probabilities, an early warning mechanism can flag high-risk cases. The system is designed to assist mental health professionals and researchers by providing supportive, data-driven insights rather than replacing clinical diagnosis.

5. Performance Evaluation and Monitoring

The effectiveness of the proposed framework is evaluated using multiple performance metrics, including Accuracy, Precision, Recall, F1-Score, and Area Under the ROC Curve (AUC). Confusion matrices and classification reports are analysed to understand misclassification patterns. Continuous monitoring mechanisms are suggested to ensure model stability and adaptability to evolving language patterns on social media platforms.

VI. RESULTS AND DISCUSSION

To assess the reliability of the proposed hybrid framework, experiments are conducted using stratified k-fold cross-validation to ensure balanced evaluation across mental health categories. The ensemble-based hybrid model demonstrates improved performance compared to individual baseline classifiers. The integration of deep contextual embeddings significantly enhances semantic interpretation, especially for ambiguous or emotionally nuanced text. The hybrid model achieves superior recall scores, which is particularly important in mental health risk detection to minimize false negatives. Comparative analysis indicates that combining traditional machine learning algorithms with deep learning architectures leads to better generalization and robustness. The findings confirm that social media text analytics can serve as a valuable supplementary tool for early mental health risk assessment.

VII. CONCLUSION AND FUTURE WORK

This study presents an intelligent hybrid machine learning framework for early mental health risk prediction using social media text analytics. By integrating advanced preprocessing

techniques, semantic feature extraction, and ensemble-based classification strategies, the proposed system demonstrates reliable and consistent predictive performance. The results highlight the potential of artificial intelligence in supporting early mental health intervention strategies. However, ethical considerations such as data privacy, user consent, and responsible AI deployment remain critical. Future work may focus on incorporating multimodal data sources such as images and user interaction patterns, applying explainable AI (XAI) techniques to improve interpretability, and developing real-time deployment systems integrated with mental health support platforms.

REFERENCES

1. T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient estimation of word representations in vector space," in Proc. Int. Conf. Learning Representations (ICLR), 2013.
2. J. 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.
3. S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.
4. C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, no. 3, pp. 273–297, 1995.
5. L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.
6. T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in Proc. 22nd ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining, 2016, pp. 785–794.
7. F. Pedregosa et al., "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.
8. E. Shortliffe and J. Cimino, *Biomedical Informatics: Computer Applications in Health Care and Biomedicine*, 4th ed. London, U.K.: Springer, 2014.
9. S. Ji, C. Pan, and M. Cambria, "A survey on deep learning for mental health analysis using social media data," *Information Processing & Management*, vol. 59, no. 2, 2022.
10. M. De Choudhury, M. Gamon, S. Counts, and E. Horvitz, "Predicting depression via social media," in Proc. Int. AAAI Conf. Web and Social Media (ICWSM), 2013.
11. S. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in Proc. Advances in Neural Information Processing Systems (NeurIPS), 2017.
12. M. Ribeiro, S. Singh, and C. Guestrin, "Why should I trust you? Explaining the predictions of any classifier," in Proc. ACM SIGKDD, 2016.