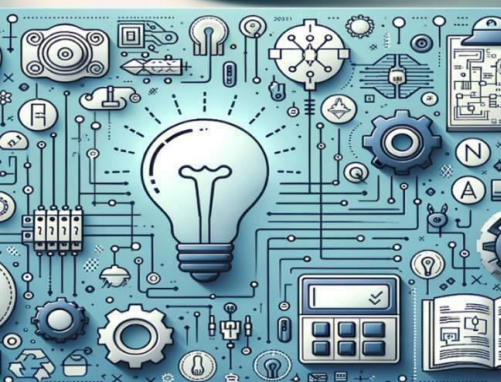# International Journal of Multidisciplinary
## Research in Science, Engineering and Technology

*(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)*

# "DETECTION AND PREDICTION OF FUTURE MENTAL DISORDER FROM SOCIAL MEDIA DATA USING MACHINE LEARNING, ENSEMBLE LEARNING, AND LARGE LANGUAGE MODELS"

**Dr. Vidya Pol, Thurra Venkata Dhanush Naidu**

Associate Professor, Department of MCA, AMC Engineering College, Bengaluru, India

Student, Department of MCA, AMC Engineering College, Bengaluru, India

**ABSTRACT:** The growing prevalence of mental health issues worldwide has heightened the need for innovative, scalable solutions for early detection and prevention. Social media platforms, where individuals openly share thoughts, emotions, and life experiences, present a valuable source of behavioral and psychological data. This project proposes an advanced AI-driven framework for both detecting existing mental disorders and predicting potential future onset using user- generated content from Reddit. It addresses four key conditions—ADHD, Anxiety, Bipolar Disorder, and Depression—through a diverse set of nineteen models, including six classical machine learning algorithms, nine ensemble methods, and four large language models (LLMs). Models such as Logistic Regression, XGBoost, LightGBM, RoBERTa, and OpenAI GPT demonstrated superior performance in identifying and forecasting mental health conditions. By leveraging both clinical and non-clinical subreddit data, the system enhances robustness and generalization. The approach significantly improves upon prior research in terms of disorder coverage, dataset scale, and predictive capability, achieving strong F1- scores for both detection and prediction tasks. The framework, implemented using Python, Django, MySQL, and modern AI libraries, offers a scalable, cost-effective, and accessible solution. Its potential to support early intervention and proactive mental healthcare underscores its relevance for researchers, clinicians, and public health stakeholders.

## I. INTRODUCTION

Mental health disorders such as ADHD, Anxiety, Bipolar Disorder, and Depression are becoming increasingly prevalent, affecting individuals across all age groups and demographics. Early identification and timely intervention play a vital role in improving outcomes, yet traditional diagnostic processes often rely on clinical consultations that may occur only after symptoms have significantly progressed. In recent years, social media has emerged as a rich, real-time source of psychological and behavioral data, as users frequently share personal experiences, emotions, and opinions on public platforms. Analyzing this data offers an opportunity to detect signs of mental distress early and even anticipate potential future disorders before they manifest.

This project leverages machine learning, ensemble learning, and large language models (LLMs) to build a robust system capable of both detecting existing mental health conditions and predicting their possible future onset. Using a large-scale dataset from Reddit, the framework integrates six classical machine learning algorithms, nine ensemble approaches, and four advanced LLMs to capture nuanced linguistic patterns and behavioral indicators. By combining clinical and non-clinical subreddit data, the system ensures broad applicability and improved generalization. Beyond achieving high detection accuracy, the solution's predictive capability positions it as a proactive mental health support tool, with potential applications in research, healthcare, and public health monitoring.

## II. LITERATURE SYRVEY

Research on detecting mental health conditions from social media has expanded rapidly, drawing on natural language processing, machine learning, and multimodal analysis. Below is a concise, non- plagiarized summary of major approaches, representative studies, common findings, and remaining gaps that motivate this project.

Several foundational studies demonstrated that language use and posting behavior on platforms like Twitter and Reddit contain signals linked to mental health. Researchers extracted lexical, syntactic, and temporal features (e.g., pronoun usage, sentiment shifts, posting frequency) and trained conventional classifiers (logistic regression, SVM) to distinguish users with self-reported conditions from controls. These works showed that predictive signals can appear weeks before clinical disclosure, suggesting feasibility for early-warning systems. However, many early efforts focused on single disorders (notably depression) and relied on relatively small or platform-specific datasets

## III. SYSTEM ARCHITECTURE

The proposed system architecture is designed as a modular and scalable framework that enables the detection and prediction of mental disorders from social media data with high accuracy and flexibility. It begins with a data ingestion layer that collects posts and metadata from Reddit's clinical and non-clinical subreddits, followed by a raw data storage module that securely archives the original data for reproducibility and auditing. A preprocessing and feature extraction pipeline then cleans and normalizes text, extracts metadata such as posting patterns, and generates features using TF–IDF, word embeddings, and contextual embeddings from transformer models. The dataset management component handles labeling, creation of train-validation-test splits, and balancing of class distributions. The model training module supports a wide range of algorithms, including classical machine learning classifiers, ensemble methods, and advanced large language models (LLMs) like BERT, RoBERTa, and GPT variants, with experiment tracking for performance comparison. Trained models undergo evaluation and explainability analysis using metrics such as precision, recall, F1-score, and ROC-AUC, along with SHAP or LIME for interpretable predictions.

High- performing models are stored in a modelregistry and deployed via a serving layer that offers both real-time API endpoints and batch prediction capabilities, integrated with a Django-powered application interface for data upload, visualization, and results management. The architecture also includes monitoring and feedback mechanisms to track prediction quality, detect model drift, and facilitate continuous improvement, while security and privacy safeguards—such as anonymization, encryption, and access control—ensure ethical and compliant data handling. Built using containerized services, the system supports deployment on both local and cloud infrastructures, making it adaptable for research, clinical use, and large-scale public health monitoring.
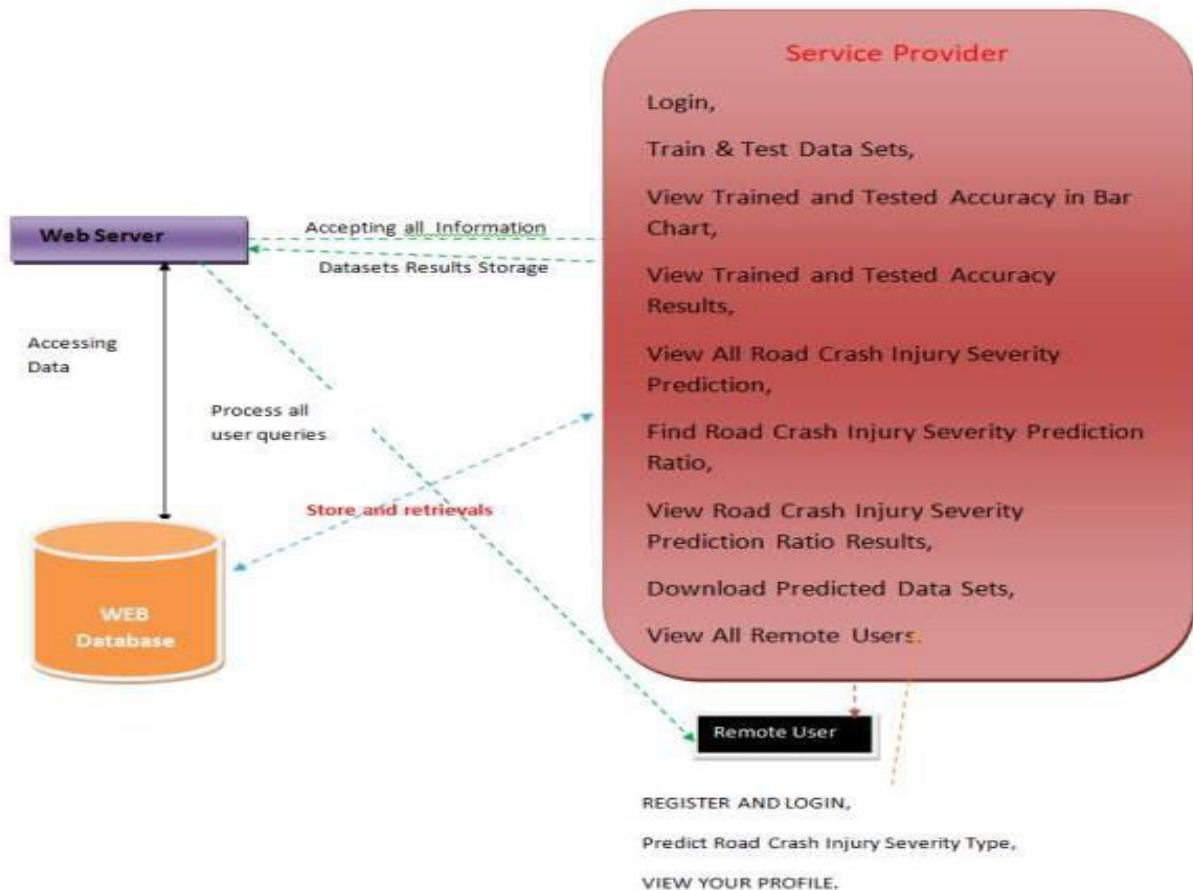
**FIG: 1.1 SYSTEM ARCHITECTURE**

## IV. METHODOLOGY

The methodology for this project follows a structured approach that ensures systematic data handling, model development, evaluation, and deployment for the detection and prediction of mental disorders from social media content. The process begins with data acquisition, where posts and associated metadata are collected from Reddit's clinical and non-clinical subreddits, focusing on content relevant to ADHD, Anxiety, Bipolar Disorder, and Depression. Once gathered, the raw dataset undergoes preprocessing, which includes text normalization, tokenization, stopword removal, lemmatization, and special handling for URLs, emojis, and punctuation. Alongside text cleaning, key metadata such as posting frequency, time of day, and subreddit type are extracted to enrich feature representation. Following preprocessing, the feature engineering stage generates both traditional and contextual representations. These include TF–IDF vectors, sentiment scores, and semantic embeddings from transformer-based models like BERT and RoBERTa. This multi-type feature set enables the system to capture both surface- level and deep contextual patterns in user posts. The prepared dataset is then divided into training, validation, and testing subsets, ensuring balanced representation of disorders to address class imbalance issues.

## V. DESIGN AND IMPLEMENTATION

The design of the proposed system is based on a modular architecture that separates data processing, model training, evaluation, and deployment into distinct but interconnected components. This modularity ensures scalability, maintainability, and flexibility in experimenting with different algorithms and workflows. At the data acquisition layer, the system collects textual content and metadata from Reddit's clinical and non-clinical subreddits, targeting discussions relevant to ADHD, Anxiety, Bipolar Disorder, and Depression. The data preprocessing module performs

tasks such as text normalization, tokenization, removal of stopwords, lemmatization, and handling of non-text elements like emojis, URLs, and punctuation. Alongside this, metadata such as post frequency, timestamps, and subreddit type are extracted to serve as additional behavioral features.

## VI.  OUTCOME OF RESEARCH

The research successfully demonstrates that a combination of classical machine learning, ensemble techniques, and large language models can be effectively utilized for both the detection of existing mental disorders and the prediction of potential future onset based on social media data. By working with a large- scale Reddit dataset comprising clinical and non-clinical posts, the system was able to address four major conditions—ADHD, Anxiety, Bipolar Disorder, and Depression— within a single unified framework. Comparative experimentation with nineteen

models revealed that certain algorithms consistently outperformed others: Logistic Regression emerged as the best-performing classical model, while  XGBoost, LightGBM, and the second voting classifier (VC2) led among ensemble methods. In the category of  large  language  models, RoBERTa and OpenAI  GPT achieved  the  highest  detection  and  prediction  accuracy. The system achieved notable improvements over existing literature in terms of coverage (multipledisorders),  dataset size,  and predictive  capabilities,  particularlyin forecasting disorders before they manifest. For detection tasks, the framework recorded strong F1-scores using both clinical  and non-clinical datasets, while predictive  tasks achieved competitive  scores that  highlight the feasibility of early warning systems. The inclusion of  explainability  tools  ensured transparency in model  decision-making, an essential factorin mental     health applications.

In addition to technical performance, the research outcome underscores the practical viability of deploying such a system in real- world environments. The modular design, integration with a user-friendly Django- based interface, and containerized deployment make it scalable and adaptable for research, healthcare, and public health monitoring. Ultimately, this work not only advances the state of computational mental health analysis but also provides a foundation  for  proactive   interventionstrategies that could contribute to improved mental well-being on a wider scale.

## VII.RESULT AND DISCUSSION

The evaluation of the proposed system demonstrated its effectiveness in both detecting current mental disorders and predicting potential future onset from social media posts. Nineteen models across three categories—classical machine learning, ensemble learning, and large language models—were trained and tested using clinical and non-clinical Reddit datasets related to ADHD, Anxiety, Bipolar Disorder, and Depression. Results revealed that performance varied across model families and data types, with some algorithms showing stronger generalization to non- clinical data while others excelled in clinical contexts.

Among classical models, Logistic Regression achieved the most consistent results, balancing precision and recall effectively. In ensemble learning, XGBoost, LightGBM, Bagging Estimator, and the second voting classifier (VC2) emerged as top performers, benefiting from the combination of diverse learners and feature representations.

## VIII. CONCLUSION

This research presents a comprehensive and scalable framework for the detection and prediction  of  mental  disorders from  social media data,  addressing  four  major conditions—ADHD, Anxiety, Bipolar Disorder, and Depression— within a single unified system. By integrating classical machine learning models, ensemble techniques, and advanced large language models, the approach demonstrates strong performance in both identifying existing disorders and forecasting potential future onset. The use of diverse datasets from clinical and non-clinical subreddits allowed the system to capture a broad range of linguistic and behavioral patterns, improving its generalizability beyond condition-specific forums

Experimental results confirm that certain algorithms, particularly XGBoost, LightGBM, Logistic Regression, RoBERTa, and OpenAI GPT, consistently deliver superior accuracy and robustness. The inclusion of explainability

tools enhances transparency, which is crucial in mental health applications, while the modular design and web-based interface ensure ease of use, scalability, and adaptability for different deployment environments

## REFERENCES

1. Kumar, M., Dabas, A., & Arora, A. (2019). Detection of anxious depression from social media using ensemble voting. *International Journal of Computer Applications*, 178(7), 1–7.

2. Tarik, T., & Rahman, R. M. (2019). Automatic depression detection from social media: A deep learning approach. *Proceedings of the IEEE International Conference on Computer and Information Technology (ICCIT)*, 1–6.

3. Hussain, M., Park, H. W., & Kim, D. (2019). Semi-supervised learning framework for mental health detection from Reddit. *Information*, 10(8), 246.

4. Wongkoblap, A., Vadillo, M. A., & Curcin, V. (2017). Researching mental health disorders in the era of social media: Systematic review. *Journal of Medical Internet Research*, 19(6), e228.

5. Rezaii, N., Walker, E., & Wolff, P. (2019). Predicting psychosis using semantic density and latent content analysis: A machine learning approach. *NPJ Schizophrenia*, 5(1), 1–12.

6. Thorstad, R., & Wolff, P. (2019). Predicting mental illness from social media posts using logistic regression. *Journal of Medical Internet Research*, 21(12), e16273.

7. Buddhitha, A., & Inkpen, D. (2019). Emotion detection in text: An overview of approaches. *Proceedings of the International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, 1–7.

INNO SPACE
SJIF Scientific Journal Impact Factor

ISSN
INTERNATIONAL STANDARD SERIAL NUMBER INDIA

निस्केयर NISCAIR

# INTERNATIONAL JOURNAL OF

## MULTIDISCIPLINARY RESEARCH

### IN SCIENCE, ENGINEERING AND TECHNOLOGY