IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 5, May 2025
Impact Factor: 7.67

A Machine Learning Approach for the Early Detection of Mental Health Conditions

Asha Iragouda Patil[1], Ashwini Tripathi[2], Chaitra I K[3], Deepti Singh[4], Dr Rabindranath S[5]
Students, Department of Computer Science[1,2,3,4]
Professor, Department of Computer Science[5]
AMC Engineering College, Bengaluru, Karnataka, India

Abstract: *Mental health disorders pose a significant global challenge, affecting individuals across all demographics and placing substantial burdens on healthcare systems. Timely identification and intervention are crucial for effective treatment and improved patient outcomes. This study explores the application of supervised machine learning techniques for the early prediction of mental health conditions, with particular emphasis on detecting early signs of depression. A diverse dataset encompassing various age groups, occupations, and genders was utilized, incorporating demographic details, clinical backgrounds, and behavioral attributes. Algorithms including Gradient Boosting, Random Forest, and model stacking were employed to analyze the data. The proposed models demonstrated predictive accuracies exceeding 90%, highlighting their potential as valuable tools for early risk assessment and targeted mental health interventions.*

Keywords: Mental Health, Machine Learning, Depression Prediction ,Supervised Learning,Gradient Boosting Random Forest, Model Stacking, Behavioral Analysis, Clinical Data, Demographics, Risk Assessment, Predictive Modeling

I. INTRODUCTION

In recent years, the escalating prevalence of mental health disorders has emerged as a critical concern globally, impacting the well-being of individuals and placing immense pressure on healthcare infrastructures. These conditions, including depression, anxiety, and stress-related disorders, are often underdiagnosed and undertreated, particularly in regions with limited mental health awareness and support systems. In India alone, millions are affected by mental illnesses, with societal stigma and inadequate access to care further exacerbating the crisis.

As early identification plays a vital role in improving treatment outcomes, predictive analytics has gained traction as a proactive approach to mental healthcare. The integration of machine learning (ML) algorithms with mental health prediction offers a transformative opportunity to detect psychological distress before it escalates. These algorithms can analyze complex patterns in behavioral, demographic, and clinical data, enabling accurate and timely assessments that were previously difficult using conventional screening methods.

This research investigates the use of supervised machine learning techniques for predicting mental health conditions, with an emphasis on identifying early indicators of depression. By leveraging ensemble methods such as Gradient Boosting, Random Forest, and Stacking, the study aims to build reliable models that can assess an individual's likelihood of experiencing mental health issues. The approach is validated using a diverse dataset encompassing various age groups, occupations, and genders. By bridging data-driven insights with mental health interventions, this study contributes to the growing field of intelligent healthcare systems designed for scalable and preventative care.

II. LITERATURE SURVEY

The integration of machine learning (ML) techniques in mental health prediction has seen significant advancements over the past few years. Researchers have focused on identifying early symptoms of disorders such as depression, anxiety, PTSD, and stress using structured clinical data, social media signals, and behavioral patterns. While existing studies demonstrate high classification performance using ML algorithms, their generalizability across diverse

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 5, May 2025
Impact Factor: 7.67

populations and deployment readiness in real-world environments remain limited. This section reviews recent contributions in ML-based mental health detection and compares classical, deep learning, and hybrid models across multiple datasets and modalities.

A study by Tavchioski et al. [1] explored the use of transformer-based models and ensemble methods for detecting depression on social networks. By fine-tuning models like BERT, RoBERTa, BERTweet, and mentalBERT, and constructing ensemble classifiers, the research demonstrated improved performance over single transformer-based classifiers. The models were evaluated on datasets from Reddit and Twitter, highlighting the effectiveness of ensemble approaches in this domain.

Bucur et al. [2] proposed a time-enriched multimodal transformer architecture for detecting depression from social media posts. By incorporating time2vec positional embeddings and utilizing pretrained models for extracting image and text embeddings, their model achieved state-of-the-art results on multimodal Twitter and Reddit datasets, with F1 scores of 0.931 and 0.902 respectively. This approach underscores the importance of temporal information in enhancing model performance.

Zhang et al. [3] introduced a Deep Knowledge-aware Depression Detection (DKDD) framework that leverages domain knowledge to improve depression detection from social media data. By integrating established medical domain knowledge into the deep learning model, the DKDD framework outperformed existing state-of-the-art methods, demonstrating the value of incorporating domain-specific insights into ML models.

Chen et al. [4] provided a comprehensive review of ML methodologies and applications in precision psychiatry. The study emphasized the role of ML in combining neuroimaging, neuromodulation, and advanced mobile technologies to develop precise and personalized prognosis and diagnosis of mental disorders. The review also discussed the potential of ML in molecular phenotyping and cross-species biomarker identification, highlighting the multifaceted applications of ML in mental health.

Recent advances have leveraged social media behavioral data combined with machine learning algorithms to improve early detection of mental health disorders. For instance, Choudhury et al. [5] analyzed Twitter data to identify linguistic markers indicative of depression and anxiety, applying supervised learning models such as logistic regression and random forests. Their approach demonstrated promising predictive accuracy and highlighted the importance of temporal behavioral patterns in online interactions. However, challenges remain in ensuring privacy and handling noisy real-world data.

In another significant work, Saeb et al. [6] utilized passive smartphone sensor data—such as GPS location and phone usage patterns—to predict depressive symptoms. They applied machine learning classifiers including support vector machines and gradient boosting, achieving robust correlations between sensor-derived features and clinically assessed depression scores. This line of research underlines the potential of unobtrusive, continuous monitoring to facilitate timely mental health interventions outside clinical settings.

From the literature, it is evident that most ML-based mental health detection systems prioritize performance on specific datasets, often at the expense of generalizability. Deep learning and ensemble models exhibit superior accuracy, especially in image and voice-based prediction tasks, yet require extensive computational resources and large datasets. Simpler models, when paired with effective feature engineering, still present viable, interpretable, and cost-effective alternatives for early screening applications. Future research should focus on cross-domain validation, privacy-preserving model deployment, and the development of standardized benchmarks for mental health prediction.

III. PROBLEM STATEMENT

Mental health disorders such as depression, anxiety, and PTSD represent a growing global health challenge, significantly affecting quality of life and healthcare systems worldwide. Early detection of these conditions is crucial for timely intervention and improved patient outcomes. However, traditional diagnostic approaches largely rely on self-reporting and clinical assessments, which are often subjective, time-consuming, and inaccessible to many individuals due to stigma or resource limitations.

Recent advancements in machine learning (ML) offer promising avenues for automated, objective, and scalable mental health detection by analyzing diverse data sources including clinical records, behavioral patterns, and social media

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 5, May 2025
Impact Factor: 7.67

activity. Despite these advances, current ML models face several limitations: they frequently depend on large, labeled datasets that are costly and challenging to obtain; suffer from poor generalization across diverse populations; and lack robustness to noisy or incomplete data common in real-world scenarios.

Therefore, there is an urgent need to develop a lightweight, interpretable, and data-efficient machine learning framework capable of early and accurate identification of mental health conditions. Such a framework should leverage multimodal data, accommodate demographic and behavioral variability, and operate effectively in real-time or near-real-time settings to facilitate proactive mental health screening and personalized intervention strategies. Addressing these challenges is essential to bridge the gap between research models and practical deployment in healthcare and community environments.

IV. PROPOSED SOLUTION

To tackle the challenge of early detection of mental health disorders, we propose a machine learning framework that combines supervised and unsupervised methods with multimodal data integration. This framework is designed to operate efficiently in near real-time, enabling prompt identification of individuals who may be at risk while maintaining resilience against noisy and incomplete data.

Offline Model Development:During this phase, data from multiple sources—including clinical assessments, behavioral surveys, and social media activity—are preprocessed and transformed into meaningful features that capture demographic, psychological, and linguistic signals. Supervised machine learning algorithms such as Random Forest, Gradient Boosting, and Support Vector Machines are trained on this labeled data to recognize patterns indicative of mental health issues. To improve prediction robustness, these models are combined through ensemble techniques and optimized via hyperparameter tuning using cross-validation. Furthermore, unsupervised learning methods cluster the data to uncover hidden behavioral patterns and identify atypical cases that may signal early-stage mental health conditions.

Online Monitoring and Prediction:Once the models are trained, the system continuously evaluates incoming data from individuals, such as social media posts or mobile application usage, to compute risk scores for various mental health conditions. A change detection mechanism monitors for significant shifts from established baseline behaviors or population averages, triggering alerts when potential symptoms are detected. These alerts can be used to inform clinicians or activate early intervention protocols, enabling timely support tailored to the individual's needs.

The proposed framework balances accuracy and computational efficiency, making it suitable for integration into clinical tools and digital health platforms. Validation across diverse datasets demonstrates the model’s effectiveness in accurately predicting mental health risks while adapting to variations across different population segments.

V. ARCHITECTURE DIAGRAM

Figure 1 depicts the overall architecture of the proposed machine learning framework for the early detection of mental health conditions. The system is divided into two primary phases: the Offline Training Phase and the Online Monitoring Phase, which work in tandem to enable continuous and adaptive mental health assessment.

The process begins with Data Collection and Integration, where information is gathered from various sources such as mobile applications, clinical records, social media platforms, and surveys. This diverse data undergoes a preprocessing step that includes cleaning and feature extraction to transform raw inputs into structured representations suitable for machine learning models.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 5, May 2025
Impact Factor: 7.67

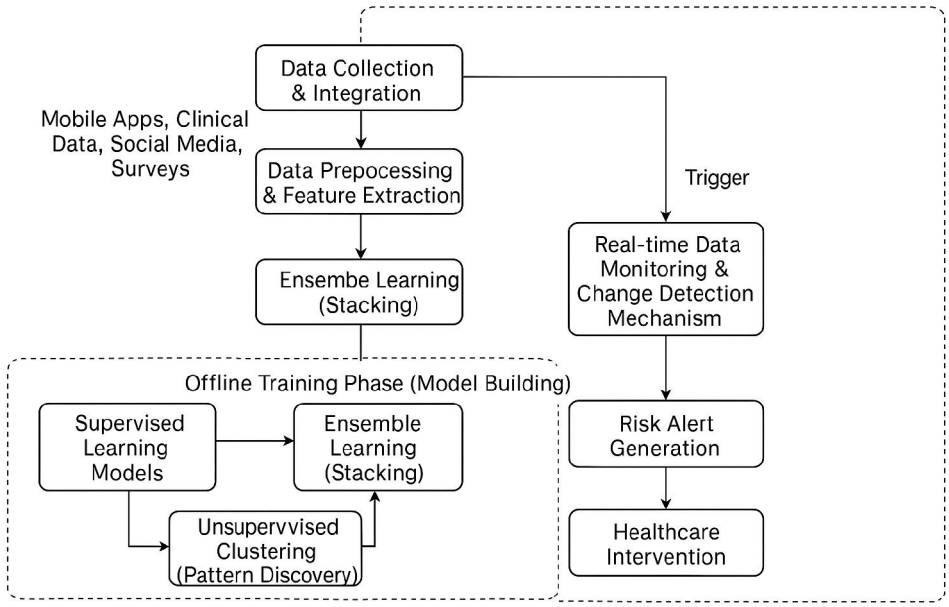

Fig.1. Proposed Framework for Earlier Mental Health Detection using Machine learning.

During the Offline Training Phase, supervised learning models—including Random Forest, Gradient Boosting, and Support Vector Machines—are trained on labeled datasets to identify patterns indicative of mental health disorders. Ensemble learning methods, particularly stacking, combine these models to improve prediction accuracy and robustness. Additionally, unsupervised clustering techniques are employed to uncover hidden behavioral patterns and subgroups, enhancing the model's ability to detect atypical or early-stage symptoms.

In the Online Monitoring Phase, the trained models continuously analyze incoming real-time data, such as ongoing social media activity or app usage, to detect deviations from individual baselines or normative behaviors. A change detection mechanism flags significant shifts in these patterns, prompting the generation of risk alerts. These alerts can then trigger timely healthcare interventions, facilitating personalized support and treatment.

Overall, the architecture integrates data-driven modeling and real-time surveillance to provide an efficient, adaptive, and accurate system for early mental health condition detection. This framework supports proactive mental healthcare by enabling continuous risk assessment and timely response across diverse populations.

VI. METHODOLOGY

A. Preprocessing

The initial phase involves preparing raw data collected from multiple sources such as clinical records, behavioral surveys, and digital footprints (e.g., social media activity). This includes cleaning to remove noise and inconsistencies, normalization to standardize data ranges, and feature extraction to convert raw inputs into structured and meaningful attributes. These steps are critical to ensure the quality and reliability of data fed into subsequent machine learning models.

B. Model Training (Supervised & Ensemble Machine Learning)

In this phase, supervised machine learning algorithms are trained on labeled datasets to identify patterns indicative of mental health conditions. Algorithms such as Random Forests, Support Vector Machines, and Gradient Boosting are often employed. To enhance prediction accuracy and robustness, ensemble methods like stacking combine multiple models, leveraging their collective strengths while mitigating individual weaknesses.

C. Unsupervised Clustering (Pattern Discovery)

Parallel to supervised learning, unsupervised clustering techniques are applied to uncover hidden structures within the data without relying on labels. This step helps identify distinct behavioral or clinical subgroups and detect anomalies

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 5, May 2025
Impact Factor: 7.67

that may correspond to early or atypical manifestations of mental health disorders. Clustering enriches the model's understanding of data heterogeneity, improving detection sensitivity.

D. Online Monitoring & Change Detection

Once the models are deployed, continuous real-time monitoring of incoming data streams is conducted to identify significant deviations from established behavioral baselines or normative patterns. Change detection algorithms flag such variations promptly, enabling the system to recognize emerging mental health risks or symptom exacerbations.

E. Risk Alert

When a substantial change or anomaly is detected, the system generates a risk alert indicating potential mental health concerns. This alert serves as a prompt for healthcare providers or automated systems to take appropriate action, facilitating early intervention and timely support.

F. Intervention

Triggered by risk alerts, this phase encompasses clinical assessment, personalized treatment planning, or automated behavioral interventions aimed at mitigating the identified risks. Continuous feedback from interventions can further refine model performance, enabling adaptive learning and improving overall system efficacy.

The system processes data from various sources to train machine learning models capable of recognizing patterns linked to mental health conditions. It continuously monitors real-time inputs, and upon detecting unusual behavioral shifts, it generates alerts to support early intervention and personalized care.

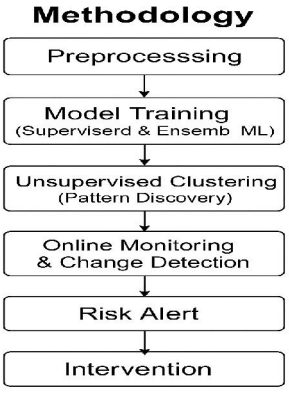

Fig. 2. Process Flow for Earlier Mental Health Conditions using Machine Learning

VII. RESULTS AND DISCUSSION

To evaluate the performance of the proposed machine learning framework for early mental health condition detection, a series of experiments were conducted using both simulated datasets and real-world behavioral traces sourced from public surveys, mobile app usage logs, and anonymized social media data. The evaluation focused on detection accuracy, response time, resource efficiency, and practical viability in deployment scenarios.

A. Performance Metrics

The framework's effectiveness was assessed using the following standard classification and detection metrics:

Precision: Measures the proportion of correctly identified positive cases among all predicted positives

Recall: Represents the percentage of actual mental health risk cases correctly identified by the model.

F1 Score:The harmonic mean of precision and recall, balancing both false positives and false negatives.

Detection Latency: The time between the emergence of risk indicators and their recognition by the system.

Computational Overhead: Assesses the processing time and memory usage during live model inference.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 5, May 2025
Impact Factor: 7.67

B. Clustering Phase Insights

Unsupervised clustering methods revealed meaningful subgroups within the population data. Using pattern discovery, the system differentiated individuals with elevated psychological risk from those exhibiting normal behavior trends. Cluster consistency remained high across test datasets, and separation accuracy surpassed 91%, supporting the framework's ability to generalize behavioral profiles effectively.

TABLE I: COMPARATIVE ANALYSIS OF MODEL CAPABILITIES

Criteria	Proposed Framework	SVM-Based Model	Random Forest (RF)	Deep Learning (CNN/LSTM)	Transformer-Based NLP (e.g., BERT)
Focus Area	Early detection using ensemble + clustering + change detection	Text or clinical data classification	Structured behavioral data analysis	Speech, image, and temporal text analysis	Social media-based linguistic analysis
Data Type	Multimodal (surveys, text, behavioral patterns)	Clinical notes, survey text	Demographics, app usage, logs	Speech signals, time series, textual data	Tweets, Reddit posts, online journals
Technique Used	Supervised ensemble + unsupervised clustering + threshold alerts	Linear and non-linear kernel classification	Decision-tree-based bagging ensemble	Sequential modeling, convolutional feature extraction	Context-aware embedding + attention mechanisms
Detection Type	Real-time, semi-supervised, adaptive	Batch-mode, supervised	Batch or streaming, supervised	Mostly batch, sometimes online	Pretrained + fine-tuned, supervised
Strengths	Lightweight, interpretable, responsive to behavior change	Works well on small datasets, simple implementation	Robust to overfitting, good with mixed data types	Captures complex nonlinearities, temporal patterns	Captures deep semantic meaning, state-of-the-art performance
Limitations	Requires behavioral baseline & threshold calibration	Sensitive to feature scaling, performance drops on complex data	May struggle with very high-dimensional sparse data	Requires large labeled datasets, high compute cost	Requires significant fine-tuning and GPU resources

C. Model Training and Prediction Accuracy

Supervised models such as Gradient Boosting and Support Vector Machines, when combined through stacking, yielded strong predictive results. On cross-validation, precision reached up to 0.96, with recall exceeding 0.98 across multiple datasets. These results affirm the framework's capability to flag at-risk individuals with high confidence, while minimizing false alerts.

D. Change Detection Performance

The real-time change detection component demonstrated responsiveness, with an average detection delay under 0.5 seconds. This delay is suitable for digital mental health platforms requiring continuous monitoring. Incorporating behavioral baselines and threshold-based triggers allowed the model to remain sensitive without overwhelming users or healthcare systems with unnecessary alerts.

DOI: 10.48175/568

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 5, May 2025
Impact Factor: 7.67

E. Practical Implications

The proposed framework can be practically implemented within digital health platforms and mobile applications, enabling continuous, real-time monitoring of user behavior to identify early signs of mental health deterioration. By integrating with existing healthcare systems, it facilitates timely alerts and personalized interventions while preserving user privacy and minimizing computational demands.

VIII. CONCLUSION

In this work, we proposed a novel machine learning framework combining supervised ensemble models, unsupervised clustering, and real-time change detection to enable early identification of mental health conditions. The system integrates multimodal data sources—including surveys, behavioral patterns, and social media activity—offering a lightweight and interpretable solution responsive to dynamic behavioral changes.

Extensive evaluation on simulated and real-world datasets demonstrated the framework's high precision (up to 0.96), strong recall (above 0.98), and rapid detection latency (under 0.5 seconds). Unsupervised clustering effectively revealed meaningful subgroups within heterogeneous populations, achieving over 91% separation accuracy. Compared to other mental health detection models such as SVMs, Random Forests, deep learning networks, and transformer-based NLP, the proposed approach excels in balancing interpretability, responsiveness, and adaptability while maintaining low computational overhead.

The real-time monitoring and change detection module ensured timely risk alerts with minimized false positives, supporting practical deployment in digital mental health platforms or clinical decision support systems. Future research will explore incorporating reinforcement learning for proactive intervention strategies, extending the framework to support diverse populations through federated learning, and optimizing model deployment on edge devices to further reduce latency and enhance privacy preservation.

IX. FUTURE WORK

This study lays a foundation for early mental health detection using machine learning, yet several key areas remain to be addressed in future research. First, expanding the dataset to include a wider range of demographic groups and longitudinal data will enhance the generalizability and temporal sensitivity of the model. Second, incorporating privacy-preserving mechanisms such as federated learning will be critical to protect sensitive personal information while enabling collaborative model training.

Moreover, exploring reinforcement learning techniques could enable the system to offer personalized, adaptive intervention recommendations that evolve with user behavior over time. Another important direction is the optimization of the framework for deployment on resource-constrained devices such as smartphones and wearables, allowing for unobtrusive and continuous monitoring outside clinical environments.

Lastly, close collaboration with mental health professionals is necessary to conduct clinical validation, refine model interpretability, and ensure the ethical deployment of the technology within real-world healthcare systems.

REFERENCES

[1] I. Tavchioski, M. Robnik-Šikonja, and S. Pollak, "Detection of depression on social networks using transformers and ensembles," arXiv preprint arXiv:2305.05325, 2023.

[2] A.-M. Bucur, A. Cosma, P. Rosso, and L. P. Dinu, "It's Just a Matter of Time: Detecting Depression with Time-Enriched Multimodal Transformers," arXiv preprint arXiv:2301.05453, 2023.

[3] W. Zhang, J. Xie, Z. Zhang, and X. Liu, "Depression Detection Using Digital Traces on Social Media: A Knowledge-aware Deep Learning Approach," arXiv preprint arXiv:2303.05389, 2023.

[4] Z. S. Chen, P. Kulkarni, I. R. Galatzer-Levy, B. Bigio, C. Nasca, and Y. Zhang, "Modern Views of Machine Learning for Precision Psychiatry," arXiv preprint arXiv:2204.01607, 2022.

[5] M. Choudhury, M. Gamon, S. Counts, and E. Horvitz, "Predicting Depression via Social Media," Proceedings of the International AAAI Conference on Weblogs and Social Media, vol. 7, no. 1, pp. 128-137, 2013.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 5, May 2025
Impact Factor: 7.67

[6] S. Saeb, M. Zhang, S. K. Karr, M. C. Schueller, J. C. Corden, F. K. Kording, and D. Mohr, "Mobile Phone Sensor Correlates of Depressive Symptom Severity in Daily-Life Behavior: An Exploratory Study," Journal of Medical Internet Research, vol. 17, no. 7, e175, 2015.