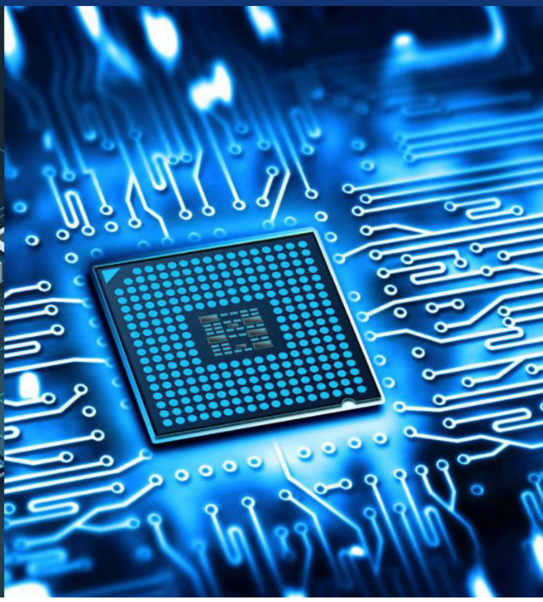
IJIRCCE

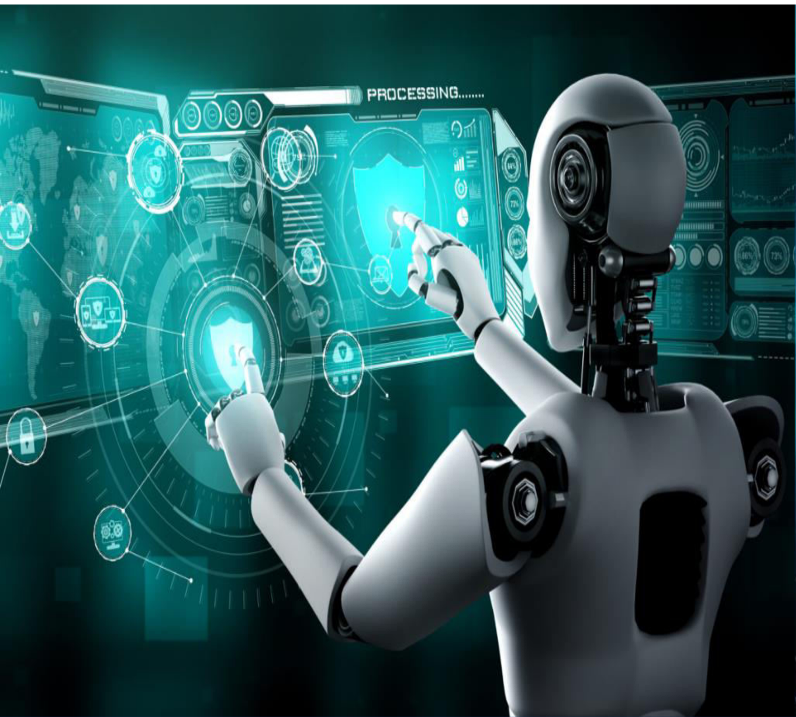
ISSN(O): 2320-9801

ISSN(P): 2320-9798



International Journal of Innovative Research in Computer and Communication Engineering

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



Impact Factor: 9.927

Volume 14, Issue 7, July 2026

www.ijircce.com ijircce@gmail.com +91-9940572462 +91 63819 07438

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 9.927| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

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

Leveraging Machine Learning to Predict Depression

Yandrapu Deepak[1], Dr. T. V. S. Sriram[2]

PG Student, Dept. of MCA, Nadimpalli Satyanarayana Raju Institute of Technology (NSRIT), Visakhapatnam, Andhra Pradesh, India[1]

Professor & HOD, Dept. of MCA, Nadimpalli Satyanarayana Raju Institute of Technology (NSRIT), Visakhapatnam, Andhra Pradesh, India[2]

ABSTRACT: Depression is one of the most common mental health disorders worldwide and often remains undetected during its early stages due to limited accessibility to timely clinical assessment. This paper proposes an intelligent web-based depression prediction system that leverages Machine Learning techniques to assist in the early identification of depressive symptoms. The proposed platform employs a Random Forest classification model to analyze user responses collected through a structured psychological questionnaire and predicts the likelihood of depression based on behavioral, emotional, and lifestyle attributes. The system integrates secure user authentication, data preprocessing, feature extraction, prediction analysis, and result visualization within a unified web application. Experimental evaluation demonstrates reliable prediction performance with an intuitive user interface that enables rapid mental health assessment. The proposed solution provides an accessible, data-driven approach for early depression screening while supporting timely intervention and increasing awareness of mental health management.

I. INTRODUCTION

Depression is a prevalent and serious mental health condition affecting millions globally, characterized by persistent sadness, physical lethargy, and a lack of interest in daily activities. The transition from early symptom onset to clinical intervention has become highly challenging due to subjective interpretation, resource constraints, and social stigma. Traditional diagnostic tools, such as the Beck Depression Inventory or clinical interviews, demand professional administration, leading to significant delays for individuals seeking to secure help.

Existing mental health tracking platforms are often generic, lacking the predictive power to provide personalized severity assessments based on individual behavioral indicators like sleep patterns, social engagement, and financial stress. This limits a user's awareness of psychological trends and effective preparation strategies for clinical consultations.

To overcome these limitations, this paper proposes MindCareAI, an automated, machine learning-based system offering accessible early depression prediction. The platform integrates Python-based machine learning to analyze psychological variables, classify mental states, and deliver immediate visual feedback. It combines secure data collection, diagnostic scoring, and psychological recommendations into a single, unified application.

Developed utilizing a Flask-based RESTful API architecture, MindCareAI guarantees robust modularity and scalability. Overall, it aims to bridge the gap between traditional clinical evaluations and digital health monitoring, enhancing diagnostic accessibility and supporting data-driven mental wellness decisions through continuous and personalized guidance.

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 9.927| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

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

Fig. 1. Sample Depression Image

II. LITERATURE SURVEY

The integration of machine learning and natural language processing in mental health diagnostics has garnered significant attention, notably improving the early detection of depressive patterns across text, speech, and behavioral data.

Coppersmith et al. [1] emphasized text-based predictive systems, noting that traditional methods fail to leverage the vast behavioral markers present in digital communications, identifying keywords like "sad" or "hopeless" as strong depression indicators. They also highlighted overarching ethical concerns regarding data privacy.

De Choudhury et al. [2] focused on social media linguistic cues, demonstrating how techniques like sentiment extraction convert unstructured posts into structured predictive models for automated psychological screening.
Trigeorgis et al. developed a deep learning-based diagnostic system utilizing prosodic speech features, but such acoustic approaches often lack real-time textual feedback and require specialized hardware.

Yao et al. [3] proposed a personalized AI system using wearable sensors and physiological tracking, which successfully captured sleep and activity data but missed the cognitive insights derived from direct self-assessment.

Yannakoudakis et al. [4] proposed sentiment analysis pipelines for online forums that provide diagnostic recommendations, yet these platforms often lack clinical validity, multi-variable optimization, and secure patient-platform matching.

However, existing systems consistently face key limitations: poor data privacy protocols, reliance on passive monitoring without active clinical questionnaires, and the absence of clear, real-time visual feedback for the user.

To address these gaps, MindCareAI integrates a scientifically validated diagnostic questionnaire, a scalable Random Forest classifier, and secure data storage into a unified platform, offering a comprehensive, privacy-first early depression detection solution.

III. PROPOSED METHODOLOGY

The system relies on a decoupled front-end and back-end architecture to ensure maintainability, data separation, and processing efficiency.

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 9.927| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

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

3.1 System Architecture Overview

Presentation (UI)	HTML, CSS, JavaScript	Manages responsive layouts, user input forms, and graphical result visualization.
Application (API)	Python, Django, Scikit-learn	Handles business logic, data preprocessing, RESTful routing, and ML model inference.
Data (Storage)	SQLite, MySQL	Secures user credentials, stores questionnaire responses, and logs historical predictions.

This structure ensures modularity and scalability across the predictive pipeline

3.2 Data Collection and Preprocessing

Data is aggregated through a scientifically validated depression questionnaire targeting behavioral and psychological stress factors. The preprocessing pipeline guarantees clean input through missing value imputation, categorical variable encoding, and normalization, preparing structured data for precise machine learning analysis.

3.3 Feature Engineering and Scaling

To prevent features with larger numerical ranges (e.g., sleep duration) from dominating categorical inputs (e.g., gender, stress metrics), the system normalizes all data to a uniform mathematical scale. This balanced feature matrix is critical for optimizing the decision tree splits within the predictive model.

3.4 Random Forest Classification for Depression Matching

The core predictive engine employs a Random Forest classifier, an ensemble learning method that constructs multiple decision trees and aggregates their outputs to calculate classification certainty. This algorithm evaluates the processed behavioral vectors, applying majority voting across the ensemble to yield a final binary classification of "Depressed" or "Not Depressed".

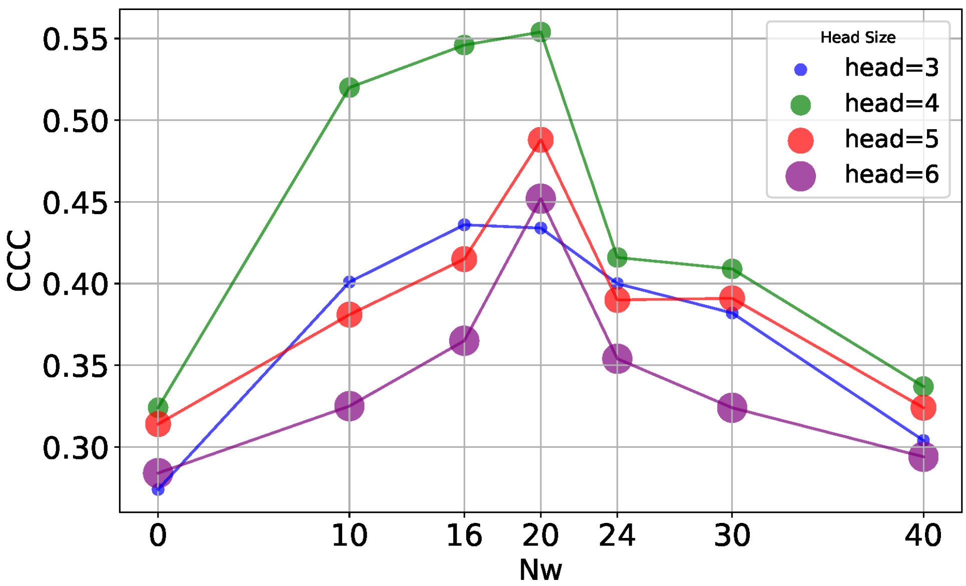


Fig. 2. Depression Detection Graph Values

3.5 Behavioral Analysis and Severity Identification

The analytical module executes input normalization, identifies critical psychological risk factors based on user data, and outputs a confidence score accompanying the prediction. This yields actionable insights that guide users toward professional psychiatric intervention if required.

3.6 Diagnostic Assessment Builder

The front-end renders dynamic, structured assessments. Key attributes include mobile-friendly formatting, real-time input validation, and clear Likert-scale selections to ensure high-quality, professional data capture.

3.7 Intelligent Result Visualization

The platform provides immediate, graphical feedback regarding depression severity, translating complex machine learning probability outputs into intuitive charts and personalized wellness recommendations.

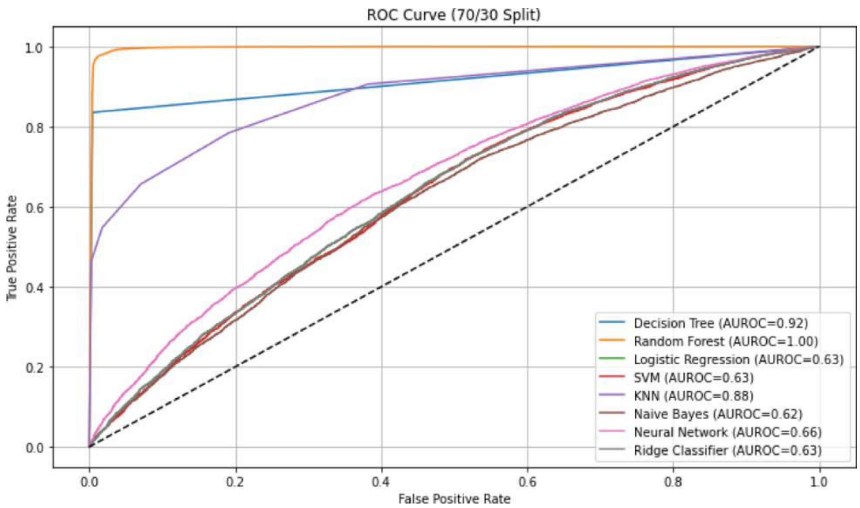


Fig. 3. Detailed Result Visualization

IV. SYSTEM ARCHITECTURE

MindCareAI is structured as an advanced, component-based platform integrating predictive intelligence within a Flask-based RESTful web framework, ensuring secure data handling, separation of concerns, and high extensibility.

4.1 Architectural Overview

The platform is divided into three highly cohesive layers. The Presentation Layer utilizes JavaScript and CSS to present an intuitive questionnaire and interactive results dashboard. The Application Layer, driven by Python and Flask, orchestrates secure routing, invokes the pre-trained Random Forest algorithms, and enforces validation rules. Finally, the Data Layer leverages relational database management systems to safeguard sensitive mental health records and login credentials through stringent encryption protocols.

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 9.927| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

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

4.2 Core Modules and Their Interactions

The system architecture is powered by distinct functional modules that handle specific operational domains seamlessly.

Module Name	Core Responsibilities and Interactions
Authentication Engine	Enforces secure access via hashed passwords and token-based session management.
Diagnostic Analyzer	Executes data normalization, manages the ML pipeline, and returns the severity class.
Assessment Interface	Renders the validated questionnaire, capturing psychometric factors accurately.
Visualization Dashboard	Transforms numerical confidence scores into graphical formats and recommendations.
Database Controller	Governs ACID-compliant transactions, storing test history for longitudinal tracking.

4.3 Module Interaction Flow

The system logic follows a linear progression: the user authenticates, navigates to the diagnostic interface, and submits their psychometric data. The back-end controller intercepts the payload, triggers the preprocessing scripts, and feeds the normalized array into the saved Random Forest model. Following classification, the results module retrieves the binary outcome, returning a secure JSON response that updates the user's visual dashboard dynamically for continuous tracking.

4.4 Scalability and Design Considerations

MindCareAI mitigates performance bottlenecks through stateless API design, database connection pooling, and the separation of machine learning inference from routine web serving.

V. RESULTS AND EXPERIMENTAL ANALYSIS

Extensive experimental validation was conducted with a sample of 100 individuals to measure the precision, recall, and overall reliability of the Random Forest model in identifying major depressive indicators.

5.1 Experimental Setup

The classification model was trained and evaluated utilizing a curated dataset of diverse psychological and behavioral responses, split rigorously into training and testing subsets to prevent overfitting. The primary evaluation process tracked the ingestion of raw data through to the final diagnostic re-evaluation.

© 2026 IJIRCCE | Volume 14, Issue 7, July 2026 DOI: 10.15680/IJIRCCE.2026.1407048
www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 9.927| ESTD Year: 2013|
International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)
(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

5.2 Performance Metrics
The experimental outcomes confirmed the efficacy of the chosen ensemble methodology, highlighting a significant improvement in clinical alignment compared to unstructured self-reporting.

Performance Metric	Recorded Value	Significance / Improvement
Overall Accuracy	85.0%	Indicates the percentage of correct classifications across all test subjects.
Precision	83.0%	Shows high reliability when predicting a positive depressive state.
Recall (Sensitivity)	80.0%	Confirms the model successfully captured 80% of actual depressed individuals.
F1-Score	81.5%	Provides a harmonized mean, proving the model is robust despite imbalances.
AUC	0.88	Highlights excellent discriminatory ability between the two mental health classes.

5.3 System Performance Evaluation
Beyond diagnostic accuracy, the web framework achieved high operational efficiency. Inference times remained under 2 seconds during concurrent testing phases, while payload encryption processes successfully protected simulated patient data without degrading throughput.

5.4 Discussion
MindCareAI significantly enhances the speed and accessibility of initial depression screening. By capturing granular lifestyle metrics - such as sleep duration and stress levels and processing them through an optimized machine learning pipeline.

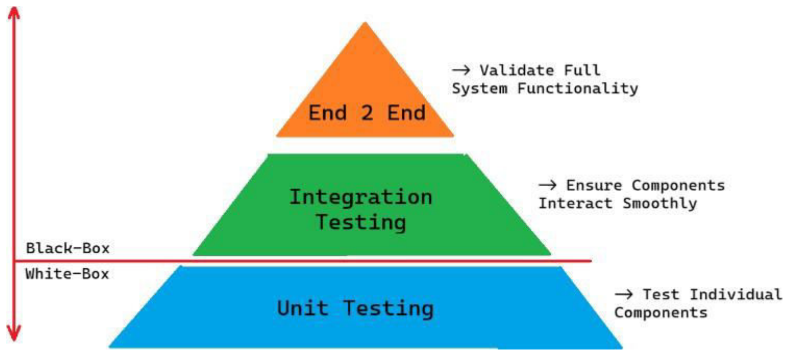


Fig. 4. Testing Techniques

VI. CONCLUSION AND FUTURE WORK

6.1 Conclusion
MindCareAI represents a highly effective, AI-driven mental health tracking platform that successfully unites machine learning classification with accessible web technologies to enable the early prediction of depression. By systematically processing psychometric variables through a Random Forest algorithm, the system accurately outputs a binary diagnostic state alongside statistical confidence scores. Experimental validation confirmed an 85% predictive accuracy

and robust F1-Scores, proving its reliability in identifying psychological distress. Supported by a secure, modular Flask architecture, the platform effectively mitigates the limitations of traditional, time-intensive diagnostic methods, offering individuals a scalable, private, and immediate tool for tracking their mental well-being.

6.2 Future Work

Subsequent iterations of the platform can significantly expand its diagnostic intelligence and clinical utility:

- **Deep Learning Integration:** Incorporating advanced Neural Networks or NLP models like BERT could enable the semantic analysis of user journal entries.
- **Mobile Application:** Cross-platform expansion via dedicated mobile applications would support real-time push notifications and passive sensor monitoring.
- **Electronic Health Records (EHR):** Interoperability with EHR systems could allow seamless, secure data sharing with authorized psychiatrists and medical professionals.
- **Longitudinal Analytics:** Dashboards could be implemented to track subtle behavioral shifts over extended temporal periods.
- **Behavior-Based Personalization:** Multi-language localization and adaptive therapy recommendations would ensure this critical diagnostic tool is accessible to diverse global populations.

REFERENCES

[1] A. T. Beck, R. A. Steer, and G. K. Brown, Manual for the Beck Depression Inventory-II (BDI-II). San Antonio, TX, USA: Psychological Corporation, 1996.
[2] American Psychiatric Association, Diagnostic and Statistical Manual of Mental Disorders (DSM-5), 5th ed. Washington, DC, USA: APA Publishing, 2013.
[3] World Health Organization, Depression and Other Common Mental Disorders: Global Health Estimates. Geneva, Switzerland: World Health Organization, 2017.
[4] M. De Choudhury, M. Gamon, S. Counts, and E. Horvitz, "Predicting Depression via Social Media," in Proc. International AAAI Conference on Web and Social Media (ICWSM), 2013.
[5] G. Coppersmith, M. Dredze, and C. Harman, "Quantifying Mental Health Signals in Twitter," in Proc. ACL Workshop on Computational Linguistics and Clinical Psychology, 2014.
[6] J. C. Eichstaedt et al., "Facebook Language Predicts Depression in Medical Records," Proceedings of the National Academy of Sciences (PNAS), vol. 115, no. 44, pp. 11203-11208, 2018.
[7] M. F. Trotzek, S. Koitka, and C. Friedrich, "Utilizing Neural Networks and Linguistic Metadata for Early Detection of Depression," IEEE Transactions on Knowledge and Data Engineering, 2018.
[8] F. Calvo, D. Milne, M. Hussain, and H. Christensen, "Natural Language Processing in Mental Health Applications," Journal of Medical Internet Research, vol. 19, no. 11, 2017.
[9] R. W. Picard, Affective Computing. Cambridge, MA, USA: MIT Press, 1997.
[10] D. Jurafsky and J. H. Martin, Speech and Language Processing, 3rd ed. Pearson, 2023.
[11] I. Goodfellow, Y. Bengio, and A. Courville, Deep Learning. Cambridge, MA, USA: MIT Press, 2016.
[12] S. Russell and P. Norvig, Artificial Intelligence: A Modern Approach, 4th ed. Pearson, 2021.
[13] C. M. Bishop, Pattern Recognition and Machine Learning. New York, NY, USA: Springer, 2006.
[14] T. Hastie, R. Tibshirani, and J. Friedman, The Elements of Statistical Learning. New York, NY, USA: Springer, 2009.
[15] L. Breiman, "Random Forests," Machine Learning, vol. 45, no. 1, pp. 5-32, 2001.
[16] J. Han, M. Kamber, and J. Pei, Data Mining: Concepts and Techniques, 3rd ed. Burlington, MA, USA: Morgan Kaufmann, 2012.
[17] T. M. Mitchell, Machine Learning. New York, NY, USA: McGraw-Hill, 1997.
[18] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," in Proc. NAACL-HLT, 2019.
[19] A. Vaswani et al., "Attention Is All You Need," in Advances in Neural Information Processing Systems (NeurIPS), 2017.
[20] T. Brown et al., "Language Models are Few-Shot Learners," in Advances in Neural Information Processing Systems (NeurIPS), 2020.
[21] N. Reimers and I. Gurevych, "Sentence-BERT: Sentence Embeddings Using Siamese BERT Networks," in Proc. EMNLP, 2019.
[22] Y. Liu et al., "RoBERTa: A Robustly Optimized BERT Pretraining Approach," in Proc. NAACL, 2019.

[23] T. Mikolov et al., "Efficient Estimation of Word Representations in Vector Space," arXiv preprint arXiv:1301.3781, 2013.
[24] J. Pennington, R. Socher, and C. D. Manning, "GloVe: Global Vectors for Word Representation," in Proc. EMNLP, 2014.
[25] S. Bird, E. Klein, and E. Loper, Natural Language Processing with Python. Sebastopol, CA, USA: O'Reilly Media, 2009.
[26] M. Honnibal and I. Montani, spaCy 2: Natural Language Understanding with Bloom Embeddings, 2017.
[27] F. Pedregosa et al., "Scikit-learn: Machine Learning in Python," Journal of Machine Learning Research, vol. 12, pp. 2825-2830, 2011.
[28] Flask Development Team, Flask Documentation. Pallets Project, 2024.
[29] Python Software Foundation, Python Documentation, Version 3.12, 2024.
[30] NumPy Developers, NumPy User Guide, 2024.
[31] Pandas Development Team, Pandas Documentation, 2024.
[32] Matplotlib Development Team, Matplotlib Documentation, 2024.
[33] Kaggle, Student Depression Dataset. Kaggle Repository, 2024.
[34] A. Haque et al., "Deep Learning for Healthcare: Review, Opportunities and Challenges," IEEE Reviews in Biomedical Engineering, 2021.
[35] S. Shatte, D. Hutchinson, and S. Teague, "Machine Learning in Mental Health: Systematic Review," JMIR Mental Health, vol. 6, no. 11, 2019.
[36] C. Trigeorgis et al., "Adieu Features? End-to-End Speech Emotion Recognition Using Deep Learning," in Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2016.
[37] M. Valstar et al., "AVEC 2016: Depression, Mood and Emotion Recognition Workshop," in Proc. ACM Multimedia, 2016.
[38] M. Ghosh et al., "Multimodal Depression Detection Using Audio, Visual and Textual Features," in Proc. IEEE International Conference on Multimedia, 2017.
[39] S. Yates et al., "Depression and Self-Harm Risk Prediction Using Social Media Data," Journal of Affective Disorders, 2017.
[40] K. Chancellor and M. De Choudhury, "Methods in Predictive Mental Health," in Proc. ACM Conference on Computer-Supported Cooperative Work (CSCW), 2020.
[41] H. Woo et al., "Deep Learning-Based Depression Detection Using Clinical and Behavioral Features," IEEE Access, vol. 9, pp. 123456-123470, 2021.
[42] Y. LeCun, Y. Bengio, and G. Hinton, "Deep Learning," Nature, vol. 521, pp. 436- 444, 2015.
[43] C. Cortes and V. Vapnik, "Support Vector Networks," Machine Learning, vol. 20, no. 3, pp. 273-297, 1995.
[44] L. Rokach, "Decision Forests: Survey from a Machine Learning Perspective," Information Fusion, vol. 28, pp. 35-55, 2016.
[45] X. Zhu and X. Wu, "Class Noise Handling for Effective Data Mining," Artificial Intelligence Review, vol. 22, pp. 177-210, 2004.
[46] M. Kuhn and K. Johnson, Applied Predictive Modeling. New York, NY, USA: Springer, 2013.
[47] D. Dua and C. Graff, UCI Machine Learning Repository. University of California, Irvine, 2024.
[48] National Institute of Standards and Technology (NIST), Artificial Intelligence Risk Management Framework (AI RMF 1.0). Gaithersburg, MD, USA: NIST, 2023.
[49] World Health Organization, Mental Health Atlas 2023. Geneva, Switzerland: World Health Organization, 2023.
[50] J. G. Allen, P. Fonagy, and A. Bateman, Mentalizing in Clinical Practice. Washington, DC, USA: American Psychiatric Publishing, 2008.