Journal of
Advanced Database
Management & Systems

ISSN: 2393-8730
Volume 12, Issue 3, 2025
September–December
DOI (Journal): 10.37591/JoADMS

https://journals.stmjournals.com/joadms

Review JoADMS

Depression Detection Using Machine Learning: A Comprehensive Review

Sheshang Degadwala[1,*], Shreyas Patel[2], Malini Joshi[3], Dharvi Soni[3]

Abstract

Depression remains one of the most prevalent mental health conditions globally, yet it frequently goes undiagnosed due to the reliance on subjective evaluation methods. With the growing availability of digital behavioral data and significant progress in machine learning (ML), new possibilities have emerged for the automated detection of depression. This review offers a detailed examination of recent advancements in ML-driven approaches to identifying depressive symptoms. It covers a range of topics, including the types of data sources used (such as social media, smartphone usage, and physiological signals), strategies for extracting relevant features, machine learning models employed, and the techniques used to evaluate their effectiveness. Furthermore, the review addresses the limitations and challenges currently faced in the field, such as data privacy concerns, generalizability, and model transparency. It also discusses promising directions for future research, emphasizing the importance of building ethical, explainable, and clinically relevant ML systems. The aim is to deliver a cohesive overview that not only synthesizes the current state of research but also offers guidance to future researchers and practitioners striving to develop responsible and accurate tools for mental health evaluation and support.

Keywords: Depression detection, machine learning, mental health, natural language processing, multimodal analysis, explainable AI

INTRODUCTION

Depression is recognized as one of the most widespread and debilitating mental health conditions worldwide, significantly impacting the quality of life of millions of individuals. According to estimates from the World Health Organization (WHO), more than 264 million people are currently living with depression across the globe. This disorder not only affects emotional well-being but also interferes with daily functioning, productivity, and physical health, making it a critical public health concern. Traditional diagnostic approaches, such as self-reported questionnaires like the Patient Health Questionnaire (PHQ-9) and structured clinical interviews, have long been the cornerstone of assessment. However, these methods often present limitations: they are time-intensive, rely heavily on the honesty and self-awareness of patients, and may be influenced by clinician subjectivity, which can lead to delayed or inconsistent diagnoses.

In response to these challenges, researchers have increasingly turned to machine learning as a promising solution for enhancing depression detection. By leveraging computational techniques, large-scale datasets, and advanced algorithms, machine learning models can identify subtle

*Author for Correspondence
Sheshang Degadwala
E-mail: sheshang13@gmail.com

[1]Professor & Head, Department of Computer Engineering, Sigma University, Vadodara, Gujarat, India
[2]Lecturer, Department of Computer Engineering, Sigma University, Vadodara, Gujarat, India
[3]Assistant Professor, Department of Computer Engineering, Sigma University, Vadodara, Gujarat, India

Received Date: July 24, 2025
Accepted Date: August 06, 2025
Published Date: August 11, 2025

Citation: Sheshang Degadwala, Shreyas Patel, Malini Joshi, Dharvi Soni. Depression Detection Using Machine Learning: A Comprehensive Review. Journal of Advanced Database Management & Systems. 2025; 12(3): 27–32p.

behavioral, linguistic, and physiological patterns that may not be apparent to human evaluators. This growing body of work has the potential to enable earlier intervention, more objective assessments, and the development of personalized treatment strategies.

Aims

This review aims to:

- Provide a comprehensive overview of the diverse data sources currently being utilized in depression detection research.
- Examine and analyze the machine learning techniques that have been most frequently applied in this domain.
- Summarize the evaluation metrics and key results reported in recent studies, offering insight into model performance.
- Identify existing challenges in the field and propose directions for future research to improve accuracy, applicability, and ethical considerations.

LITERATURE REVIEW

Text-based Depression Detection

Social media platforms have become a powerful resource for studying human behavior and mental health because of the openness of their content and the richness of the language people use when interacting online. These platforms allow researchers to observe linguistic cues, posting habits, and engagement patterns that may indicate underlying psychological states. Early research in this area focused on identifying depressive tendencies by examining how individuals express themselves through text. People experiencing depressive symptoms were found to use more self-referential language, such as first-person pronouns, and often relied on words that conveyed sadness, hopelessness, or negativity. They also tended to show changes in their activity levels, such as posting less frequently or engaging less with others [1–3].

Later studies advanced these methods by moving beyond simple linguistic observation and applying computational models to large collections of user-generated content. Approaches such as lexicon-based analysis and traditional machine learning techniques like logistic regression or support vector machines allowed researchers to categorize and predict depression-related patterns with greater accuracy. With the rise of deep learning, neural network models combined with word embeddings became increasingly popular, offering the ability to capture subtle semantic relationships in text. These methods improved early detection by recognizing nuanced differences in language use that simpler models often missed [4–7].

The field has continued to evolve with the introduction of transformer-based models, particularly those using attention mechanisms. Models such as BERT have transformed natural language processing by providing contextualized word representations, making them especially suitable for sensitive tasks like detecting signs of mental distress. When fine-tuned on social media data, these models have shown significant improvements in identifying depressive markers compared to older techniques. This progression illustrates how advancements in computational methods can deepen our understanding of mental health expressions in online environments, paving the way for more accurate, early, and ethical detection systems [8–11].

Audio-based Depression Detection

Speech analysis offers non-invasive and privacy-respecting detection methods.

- A speech-based systems, highlighting features such as pitch, jitter, shimmer, and MFCCs as effective indicators of depression.
- A used a bi-directional LSTM model on speech sequences and transcripts from clinical interviews, achieving robust classification accuracy [12–15].

Multimodal datasets like DAIC-WOZ and AVEC have supported audio-centric research. Deep learning techniques like CNNs and LSTMs have been employed for extracting temporal features, achieving high performance in classifying depressive symptoms based on vocal patterns.

Visual and Multimodal Approaches

Facial expressions, gestures, and head movements are strongly correlated with emotional states. Researchers have analyzed gender-specific facial markers to detect PTSD and depression, emphasizing the importance of demographic factors. By combining facial and text features through deep multimodal fusion, studies demonstrated that integrating multiple modalities significantly improved predictive accuracy. More recent approaches have proposed deep multimodal fusion architectures that integrate audio, visual, and textual data. These models achieved state-of-the-art results in emotion recognition and mental health detection challenges, consistently outperforming systems that rely on a single modality.

Sensor and Behavioral Data

With the proliferation of smartphones and wearable devices, sensor data has become a valuable source for behavioral analysis. Datasets such as those collected from student populations have provided insights into sleep patterns, mobility, and phone usage, offering a clearer picture of how daily habits relate to mental health. Building on this, researchers have proposed deep learning-based recommender systems that leverage smartphone data to predict depressive episodes in real-time, demonstrating the potential of sensor-driven approaches for proactive mental health monitoring and intervention. Such behavioral data, when combined with traditional inputs like text or audio, improves both model performance and user personalization.

Ethical and Interpretability Aspects

Challenges in real-world deployment have been highlighted, including issues related to privacy, algorithmic bias, and explainability. As machine learning models, particularly deep learning ones, become increasingly complex, clinicians require models that are interpretable and trustworthy. This need has driven the emergence of research in Explainable AI (XAI), which aims to bridge the gap between sophisticated machine learning models and healthcare providers, ensuring that predictions and recommendations are transparent, understandable, and actionable in clinical settings.

DATA SOURCES FOR DEPRESSION DETECTION

The performance of ML models heavily depends on the quality and variety of data. The main data sources used are:

Social Media Data

Platforms such as Twitter, Reddit, and Facebook are extensively used. Users often express emotions and behavioral cues that can signal depression. Publicly available datasets include:

- *CLPsych 2015 Dataset* (Reddit).
- *eRisk 2017/2018* (Early risk prediction from social media).

Audio and Speech Data

Speech patterns reflect emotional state. Characteristics such as pitch, tone, energy, and pause duration can reveal depressive symptoms. The DAIC-WOZ dataset is widely used for this purpose.

Facial Video Data

Facial expressions, eye movement, and micro-expressions serve as non-verbal indicators of mental state. Datasets include *AVEC (Audio/Visual Emotion Challenge)* and *RECOLA*.

Textual and Clinical Notes

Diaries, blogs, or psychotherapy session transcripts are sources of structured text data. These datasets are often annotated manually or using standardized depression scales.

Sensor and Wearable Data

Smartphone usage, sleep patterns, physical activity, and location tracking offer behavioral insights. Examples include *Student Life dataset* and *CAMPUS dataset*.

FEATURE EXTRACTION TECHNIQUES

ML models require meaningful features from raw data.

Textual Features

- *Bag-of-Words, TF-IDF,*
- *Sentiment Analysis,*
- *Lexical Features (e.g., LIWC), and*
- *Word Embeddings:* Word2Vec, GloVe, BERT.

Audio Features

- MFCC (Mel Frequency Cepstral Coefficients), and
- Pitch, Energy, Jitter, Shimmer.

Visual Features

- Facial Action Coding System (FACS),
- Facial landmark detection, and
- Head pose, eye gaze.

Behavioral Features

- App usage patterns,
- Call logs and mobility, and
- Sleep and physical activity metrics.

MACHINE LEARNING ALGORITHMS

Table 1 presents widely used machine learning algorithms and their applications in analyzing textual, audio, and multimodal data for mental health research and social media analysis (adapted from multiple studies).

PERFORMANCE METRICS

Evaluating ML models in mental health must balance sensitivity and specificity.

- *Accuracy:* Overall correctness.
- *Precision:* True positive rate.
- *Recall (Sensitivity):* Detection capability.
- *F1 Score:* Harmonic mean of precision and recall.

Table 1. Overview of common machine learning algorithms and their applications in mental health and social media analysis.

Algorithm	Description	Application
Logistic Regression	Simple classifier used in text-based detection	Social media classification
SVM (Support Vector Machine)	Robust in high-dimensional spaces	Text and speech data
Random Forest	Ensemble learning with feature importance	Sensor and multimodal data
Naive Bayes	Probabilistic model for text data	Sentiment-based classification
Neural Networks	Deep learning models for complex features	Audio/video/text fusion
CNN/RNN/LSTM	Temporal and spatial pattern recognition	Speech, video, sequential text
Transformers (BERT, RoBERTa)	State-of-the-art for language modeling	Depression detection from text
Multimodal Fusion Networks	Combine multiple data streams	AVEC/DAIC-WOZ based models

- *ROC-AUC:* Trade-off analysis.
- *Confusion Matrix:* Error distribution.
- *Cross-validation:* Model generalization check.

RECENT TRENDS AND CASE STUDIES

- *BERT-based classifiers* trained on Reddit posts achieved >85% accuracy in early depression detection.
- *CNN-LSTM* models on AVEC datasets outperform traditional SVMs in multimodal fusion.
- *Ensemble approaches* combining sensor, text, and facial features show promise in increasing detection accuracy to over 90% in controlled environments.

CHALLENGES

Despite promising results, several challenges remain:

Data Privacy and Ethics

Sensitive data like mental health indicators demand strict privacy protocols. Federated learning and anonymization are emerging solutions.

Imbalanced and Noisy Data

Depression cases are underrepresented in datasets, leading to bias. Techniques like SMOTE and data augmentation are needed.

Lack of Interpretability

Deep learning models often lack transparency, limiting clinical trust.

Generalizability

Models trained on specific populations (e.g., students) may not perform well across demographics.

Future Directions

- Explainable AI (XAI) for clinical interpretability.
- Federated Learning to preserve user privacy.
- Longitudinal Modeling to monitor depression trends over time.
- Real-time Monitoring via mobile health (mHealth) apps.
- Standardized Benchmarks for fair comparison of models.

CONCLUSION

Machine learning provides a transformative framework for depression detection, offering scalability, early intervention, and personalization. However, for these technologies to transition from research to clinical practice, they must address ethical, technical, and interpretability concerns. The integration of multimodal data, explainable models, and privacy-preserving techniques will shape the future of mental health diagnostics.

REFERENCES

1. Carey JL, Carreiro S, Chapman B, Nader N, Chai PR, Pagoto S, Jake-Schoffman DE. SoMe and Self Harm: The use of social media in depressed and suicidal youth. In Proceedings of the... Annual Hawaii International Conference on System Sciences. Annual Hawaii International Conference on System Sciences. 2018; 2018: 3314.
2. Trotzek M, Koitka S, Friedrich CM. Utilizing neural networks and linguistic metadata for early detection of depression indications in text sequences. IEEE Trans Knowl Data Eng. 2018 Dec 18; 32(3): 588–601.

3. Orabi AH, Buddhitha P, Orabi MH, Inkpen D. Deep learning for depression detection of twitter users. InProceedings of the fifth workshop on computational linguistics and clinical psychology: from keyboard to clinic. 2018 Jun; 88–97.
4. Losada DE, Crestani F, Parapar J. Overview of eRisk: early risk prediction on the internet. In International conference of the cross-language evaluation forum for European languages. Cham: Springer International Publishing; 2018 Aug 15; 343–361.
5. De Choudhury M, Counts S, Horvitz E. Predicting postpartum changes in emotion and behavior via social media. In Proceedings of the SIGCHI conference on human factors in computing systems. 2013 Apr 27; 3267–3276.
6. Cummins N, Scherer S, Krajewski J, Schnieder S, Epps J, Quatieri TF. A review of depression and suicide risk assessment using speech analysis. Speech Commun. 2015 Jul 1; 71: 10–49.
7. Morales M, Scherer S, Levitan R. A linguistically-informed fusion approach for multimodal depression detection. In proceedings of the fifth workshop on computational linguistics and clinical psychology: from keyboard to clinic. 2018 Jun; 13–24.
8. Resnik P, Armstrong W, Claudino L, Nguyen T, Nguyen VA, Boyd-Graber J. Beyond LDA: exploring supervised topic modeling for depression-related language in Twitter. In Proceedings of the 2nd workshop on computational linguistics and clinical psychology: from linguistic signal to clinical reality. 2015; 99–107.
9. Prieto VM, Matos S, Alvarez M, Cacheda F, Oliveira JL. Twitter: a good place to detect health conditions. PloS one. 2014 Jan 29; 9(1): e86191.
10. Al Hanai T, Ghassemi MM, Glass JR. Detecting depression with audio/text sequence modeling of interviews. In Interspeech. 2018 Sep 2; 1716–1720.
11. Stratou G, Scherer S, Gratch J, Morency LP. Automatic nonverbal behavior indicators of depression and PTSD: the effect of gender. J Multimodal User Interfaces. 2015 Mar; 9(1): 17–29.
12. Huang D, Zhou Z, Zhang Z, Dai Q, Lu H, Li Y, Huang Y. Land Use/Land Cover Remote Sensing Classification in Complex Subtropical Karst Environments: Challenges, Methodological Review, and Research Frontiers. Appl Sci. 2025 Sep 2; 15(17): 9641.
13. He L, Niu M, Tiwari P, Marttinen P, Su R, Jiang J, Guo C, Wang H, Ding S, Wang Z, Pan X. Deep learning for depression recognition with audiovisual cues: A review. Inf Fusion. 2022 Apr 1; 80: 56–86.
14. Guntuku SC, Yaden DB, Kern ML, Ungar LH, Eichstaedt JC. Detecting depression and mental illness on social media: an integrative review. Curr Opin Behav Sci. 2017 Dec 1; 18: 43–9.
15. Nedunchezhian P, Mahalingam M. The Improved Depression Recovery Motivation Recommendation System (I-DRMRS) in Online Social Networks. SN Comput Sci. 2022 Mar; 3(2): 166.

