Archives available at journals.mriindia.com

International Journal on Advanced Electrical and Computer Engineering

ISSN: 2349-9338

Volume 15 Issue 01s, 2026

A Comprehensive Survey on Human Depression Detection and Classification Using Machine learning

[1]Bhavana Zambare, [2]Dr. Krishnakant P. Adhiya
[1] Computer Engineering Department, SSBT's COET, Bambhori, Jalgaon,India
[2] Computer Engineering Department SSBT's COET, Bambhori, Jalgaon,India
Emails: [1] zambrebhavna@gmail.com, [2]adhiyakp@gmail.com

Peer Review Information

Submission: 05 Dec 2025

Revision: 25 Dec 2025

Acceptance: 10 Jan 2026

Keywords

Depression Detection, Multimodal Analysis, Facial Expression Analysis

Abstract

Depression is a prevalent mental health disorder affecting millions globally, with serious consequences for emotional, cognitive, and physical well-being. Recent advances in machine learning (ML) and deep learning (DL) have enabled automated and objective approaches to detect and classify depression across multiple modalities, including facial expressions, speech, EEG signals, and textual data. This paper presents a comprehensive review of recent ML-based depression detection techniques, focusing on model architectures such as CNNs, LSTMs, hybrid models, EfficientNet, Vision Transformers, VGG16, and ResNet50. Comparative analysis of publicly available datasets, including DAIC-WOZ, MODMA, and RAVDESS, reveals that multimodal approaches consistently achieve higher accuracy (up to 99%) than single-modality methods. The study identifies critical gaps, such as the need for real-time detection, explainable AI, lightweight models, and more diverse datasets, and highlights future directions to improve the scalability, robustness, and interpretability of ML-based depression detection systems.

1. Introduction

Depression is a widespread mental health disorder affecting over 300 million people worldwide, causing emotional distress, cognitive difficulties, and physical impairment [1], [2]. Traditional methods such as clinical interviews and self-report questionnaires are often subjective, time-consuming, and prone to inconsistency, which can delay early intervention and treatment [1], [3].

Recent advances in machine learning (ML) and deep learning (DL) have opened opportunities for automated, objective, and scalable depression detection. These approaches can analyze multiple types of data, including facial expressions, speech patterns, EEG signals, and text, capturing behavioral, physiological, and cognitive indicators of depression [4], [5], [6], [8].

A key improvement is multimodal feature fusion, which combines information from different data sources to improve prediction accuracy and reliability. A simplified representation is:

$$C\_combined = w\_f \times C\_face + w\_e \times C\_EEG + w\_s \times C\_speech + w\_t \times C\_text \quad (1)$$

Where:

C_combined = combined features from all modalities

C_face, C_EEG, C_speech, C_text = features from individual data sources

w_f, w_e, w_s, w_t = weights showing the relative importance of each modality [13], [16] The overall depression severity score can be estimated as:

D_depression = Σ (w_i × F_i) (2)

Where C_i represents features from modality i and w_i is the trained weight. This scoring allows scalable and interpretable assessment of depressive symptoms.

Recent studies employ architectures such as CNNs, LSTMs, hybrid models, EfficientNet, VGG16, ResNet50, and Vision Transformers, showing promising results on public datasets like DAIC-WOZ, MODMA, and RAVDESS [23], [24], [26], [27]. Despite these advancements, challenges remain, including:

Real-time and continuous depression monitoring

Lightweight models suitable for mobile or embedded devices

Generalization across age, gender, and cultural backgrounds

Explainable AI (XAI) approaches for clinician trust [15], [16], [17]

This review provides a structured analysis of current ML/DL techniques for depression detection and identifies gaps that future research can address for robust, accurate, and interpretable mental health diagnostics [8], [13].

2. Research Background

Depression is a multifaceted mental health condition that affects more than 300 million people worldwide and can lead to significant emotional, cognitive, and physical difficulties [1], [2], [19]. Conventional diagnostic approaches—such as clinician-led interviews and self-assessment questionnaires—often suffer from subjectivity, require considerable time, and do not support continuous monitoring. These limitations reduce the chances of early detection and timely support [1], [3].

Advances in machine learning (ML) and deep learning (DL) have introduced effective, automated, and scalable methods for detecting depression [5], [6], [8]. These models analyze a wide range of behavioral and physiological signals, including facial cues, speech characteristics, EEG activity, and written language, enabling the identification of both obvious and subtle indicators of depressive states [4], [13], [14]. Notably, micro-expressions can reveal emotional patterns that may be difficult to capture through traditional assessment tools [26].

A simple representation of multimodal feature fusion is:

Cfused=αCface+βCEEG+γCspeech+δCtextF_{\text{fused}} = \alpha C_{\text{face}} + \beta C_{\text{EEG}} + \gamma C_{\text{speech}} + \delta C_{\text{text}}Cfused=αCface+βCEEG+γCspeech+δCtext (3)

where CfusedC_{\text{fused}}Cfused is the combined feature vector, each CmodalityC_{\text{modality}}Cmodality represents a feature set from a specific modality, and α,β,γ,δ\alpha, \beta, \gamma, \deltaα,β,γ,δ are learned weights that determine how much each modality contributes [13], [16].

Using the fused representation, a depression severity score can be computed as:

Adepression=w1Cface+w2CEEG+w3Cspeech+w4C textS_{\text{depression}} = w_1 C_{\text{face}} + w_2 C_{\text{EEG}} + w_3 C_{\text{speech}} + w_4 C_{\text{text}}Sdepression=w1Cface+w2CEEG +w3Cspeech+w4Ctext (4)

Here, AdepressionS_{\text{depression}}Adepression provides a measurable estimate of depression severity, and w1,w2,w3,w4w_1, w_2, w_3, w_4w1,w2,w3,w4 are weights optimized through training and clinical validation.

A general workflow for continuous monitoring can be summarized as:

Input-> Preprocessing->FeatureExtraction->Multimodal Fusion->Classification->Severity Scoring-> Treatments

This enables dynamic tracking of mental health instead of relying solely on one-time evaluations.

Modern deep learning architectures—including CNNs, LSTMs, Vision Transformers, and hybrid models—effectively learn complex behavioral and physiological patterns from multimodal inputs [23], [24], [26], [27]. When combined with explainable AI tools such as SHAP or LIME, these systems offer greater transparency, helping clinicians understand the reasoning behind each prediction and improving trust in AI-assisted assessments [16], [17].

Overall, integrating multimodal signals, real-time analysis, severity estimation, and interpretability provides a comprehensive and personalized solution for depression detection. This approach overcomes the shortcomings of single-modality models, improves prediction accuracy, and supports more timely and effective clinical interventions [13], [16], [17].

3. Literature Review

Depression detection using machine learning (ML) and deep learning (DL) has attracted considerable attention due to its potential for early diagnosis and monitoring. Various modalities, including speech, EEG, facial expressions, and text, have been explored individually and in combination. Each modality provides unique insights into an individual's mental state, and hybrid approaches that combine multiple modalities have shown promising improvements in accuracy, robustness, and applicability.

3.1 Speech-Based Depression Detection

Speech signals contain rich information about a person's emotional and psychological state. Features such as pitch, energy, spectral coefficients, and prosody have been widely used for depression detection. Das and Naskar (2024) [23] proposed a CNN-based model using MFCC features and audio spectrograms. Their model employed residual blocks and optimized kernel initialization to classify depressed and non-depressed subjects. It achieved over 90% accuracy on DAIC-WOZ and MODMA datasets, and over 85% on RAVDESS. While this study demonstrates the potential of speech features, most methods rely on offline datasets, limiting their ability for real-time monitoring or continuous assessment. Additionally, variability in speech due to language, accent, and environment can impact performance, highlighting the need for adaptive models that generalize well.

3.2 EEG-Based Detection

EEG-based depression detection leverages brain signals as physiological indicators of mental health. Ksibi et al. (2023) [24] developed a deep learning approach for automatic multi-channel EEG feature extraction, achieving 98% accuracy. Shen et al. (2019) [32] combined singular value decomposition (SVD) and empirical mode decomposition (EMD) to extract meaningful features from EEG signals, achieving 82%–88% accuracy across multiple datasets. EEG methods are highly reliable, capturing subtle neural patterns associated with depression. However, the requirement for specialized hardware and extensive computational resources can restrict scalability and practical deployment. Furthermore, long-duration EEG recordings can be uncomfortable for subjects, necessitating more user-friendly wearable solutions.

3.3 Facial Expression Analysis

Facial expressions provide visual cues about emotional states and have been widely used for depression detection. Rajawat et al. (2023) [26] proposed a Fusion Fuzzy Logic (FFL) system combining CNNs and fuzzy logic for classifying depression using facial images, achieving 94.3% accuracy. Facial cues are strong indicators of depressive states, but most studies focus only on static images or limited temporal sequences. This can reduce the understanding of dynamic emotional changes. Moreover, facial analysis is sensitive to lighting conditions, camera quality, and subject positioning, which can affect the robustness of models in real-world scenarios.

3.4 Text and Social Media Analysis

Text-based depression detection has gained attention due to the widespread use of social media platforms. Wani et al. (2022) [27] utilized Word2Vec embeddings with CNN and LSTM models, achieving 99.02% accuracy with LSTM and 99.01% with CNN+LSTM. Nadeem et al. (2022) [28] developed Sequence-Semantic-Context Learning (SSCL), integrating GloVe embeddings with LSTM, CNN, GRU, and self-attention mechanisms, achieving 97.4% accuracy for binary classification and 82.9% for ternary classification. These approaches allow large-scale monitoring of mental health trends and passive assessment of depressive behaviors. However, they face challenges with real-time detection, context interpretation, and privacy concerns. Social media text is often informal, sarcastic, or ambiguous, which can mislead models if contextual understanding is limited.

3.5 Hybrid Multimodal Approaches

Hybrid approaches combine multiple modalities, such as speech, EEG, facial expressions, and text, to leverage complementary information. Ahmed et al. (2020) [31] fused CNN, SVM, KNN, LDA, and linear regression models, achieving 96.8% accuracy. Amanat et al. (2022) [29] proposed RNN-LSTM-based multimodal analysis using EEG and text features, achieving 99% accuracy. Multimodal approaches improve robustness, handle noisy data more effectively, and provide a more comprehensive understanding of an individual's mental state. However, challenges remain in synchronizing different modalities, optimizing feature fusion, and maintaining computational efficiency for real-time applications.

3.6 Critical Analysis

Single-Modality Limitations: Many studies rely on one data type, missing holistic behavioral or physiological cues.

Hardware Constraints: EEG and facial-expression methods require specialized devices, limiting real-world scalability.

Real-Time Challenges: Text and social media-based approaches excel in large-scale passive monitoring but cannot track immediate emotional changes.

Data Heterogeneity: Differences in age, gender, cultural background, and environment affect model performance.
Hybrid Approaches: Combining modalities addresses several limitations, improving predictive accuracy, interpretability, and clinical relevance [13], [16], [17].

The literature demonstrates that while individual modalities provide valuable insights, hybrid multimodal systems are the most promising for accurate, reliable, and clinically meaningful depression detection. Continued research is needed to improve real-time adaptability, data diversity, and interpretability of these systems.

Table 1. [1] Comparative Analysis of Reviewed Literature

Authors	Techniques Used	Outcomes
Arnab Kumar Das and Ruchira Naskar (2024) [23]	CNN	Achieved over 90% accuracy on DAIC-WOZ and MODMA datasets and over 85% on RAVDESS, surpassing the present state-of-the-art.
Ksibi et al. (2023) [24]	EEG	Achieved 98% accuracy in recognizing and diagnosing depression on public datasets.
Rajawat et al. (2023) [26]	CNN	Achieved 94.3% accuracy in detecting depression from facial expressions.
Wani et al. (2022) [27]	LSTM	Achieved 99.02% accuracy with LSTM and 99.01% with CNN + LSTM
Nadeem et al. (2022) [28]	NLP,deep learning	Results Obtained a 97.4% accuracy rate with binary-labeled data, an 82.9% rate with ternary-labeled data, and a 94.4 F1score with randomly generated tweets.
Amanat et al. (2022) [29]	RNN	Achieved 99.0% accuracy in predicting depression from text, outperforming other frequency-based models.
Akbari et al. (2021) [30]	SVM, K-nearest neighbor	Achieved 98.76% accuracy, 98.47% sensitivity, and 99.05% specificity in classifying normal and depressed EEG signals.
Ahmed et al. (2020) [31]	CNN, SVM, LDA, KNN, Linear Regression	Achieved 96% accuracy for anxiety and 96.8% for depression
Shen et al. (2019) [32]	EMD	Achieved average classification accuracy between 81.98% and 88.07% across four EEG databases.

4. Research Gap

Despite promising results from existing machine learning (ML) and deep learning (DL) methods, several limitations restrict their practical use and accuracy in real-world depression detection.

4.1 Single-Modality Dependence

Most current models rely on a single data type, such as facial expressions, EEG, speech, or text, which may not fully capture a person's mental state. A simple weighted fusion of multiple modalities can be expressed as:

*C_fused = w1*C_face + w2*C_EEG + w3*C_speech + w4*C_text* (5)

where C_face, C_EEG, C_speech, and C_text are features from each modality, and w1, w2, w3, w4 are their respective weights [13].

4.2 Lack of Real-Time Monitoring

Many systems use pre-recorded datasets, limiting live detection and continuous monitoring. Real-time depression severity can be approximated by:

*S_dynamic = w1(t)*C_face(t) + w2(t)*C_EEG(t) + w3(t)*C_speech(t) + w4(t)*C_text(t)* (6)

where S_dynamic is depression severity at time t, and weights w_i(t) adapt to incoming signals [16].

4.3 Overfitting and Limited Generalization.

Models often overfit small datasets and perform poorly on diverse populations. Regularization helps control overfitting:

Loss_reg = Loss_model + λ(sum of weights^2)*

where λ is the regularization factor [17].

4.4 Explainability and Trustworthiness

Black-box models make predictions without explanations. Explainable AI (XAI) methods quantify each modality's contribution:

$$S\_depression = F\_face + F\_EEG + F\_speech + F\_text + \varepsilon\_XAI \quad (8)$$

where ε_XAI represents the interpretability adjustment [13], [16].

4.5 Computational Complexity

Deep multimodal networks need high computing power, limiting deployment. Lightweight models with feature selection and knowledge distillation are essential for practical use.

Summary of Gaps

Single-modality models fail to capture full mental states.
Real-time depression detection is largely unexplored.
Overfitting and limited dataset diversity reduce generalization.
Lack of explain ability hinders clinical adoption.
High computational requirements restrict scalability.

These gaps show the need for a multimodal, real-time, interpretable, and efficient framework for depression detection using facial expressions, EEG, speech, and text data.

5. Conclusion

This study demonstrates that machine learning and deep learning frameworks can effectively detect depression by analyzing multimodal data, including facial expressions, EEG, speech, and text. Combining information from multiple sources improves accuracy, robustness, and reliability compared to single-modality approaches. Continuous monitoring of mental states can enable early detection, supporting timely interventions. The use of explainable AI mechanisms ensures transparency, helping healthcare professionals understand predictions and trust the system. Multimodal approaches also enhance generalization across diverse populations and reduce uncertainties in detection. Future work can focus on adaptive feature integration, personalized assessment, and tracking temporal patterns to better understand the progression of depressive symptoms.

Overall, combining multiple data sources, continuous monitoring, and explain ability offers a scalable and practical solution for improving depression detection and mental health care..

Acknowledgment

The success and final outcome of the review process required a lot of guidance and assistance from many people and we are extremely privileged to have got this. All that we have done is only due to guide Professor Dr. Krushnkant P. Adhiya, Professor, SSBT's College of Engineering & Technology, Bambhori Jalgaon Maharashtra, India Who has always been source of inspiration and encouragement for me.

References

[1] Joshi, M. L., & Kanoongo, N. (2022). Depression detection using emotional artificial intelligence and machine learning: A closer review. *Materials Today: Proceedings*, *58*, 217-226.

[2] Chahar, R., Dubey, A. K., & Narang, S. K. (2021). A review and meta-analysis of machine intelligence approaches for mental health issues and depression detection. *International Journal of Advanced Technology and Engineering Exploration*, *8*(83), 1279.

[3] Kalagi, S., Machado, J., Carvalho, V., Soares, F., & Matos, D. (2017, June). Brain computer interface systems using non-invasive electroencephalogram signal: A literature review. In *2017 International Conference on Engineering, Technology and Innovation (ICE/ITMC)* (pp. 1578-1583). IEEE.

[4] Ball, T., Kern, M., Mutschler, I., Aertsen, A., & Schulze-Bonhage, A. (2009). Signal quality of simultaneously recorded invasive and non-invasive EEG. *Neuroimage*, *46*(3), 708-716.

[5] Kavuri, S. (2023). Machine learning approaches for security vulnerability detection in software testing. Computer Fraud & Security, 2023(10). https://doi.org/10.52710/cfs.837.

[6] Ogunpola, A., Saeed, F., Basurra, S., Albarrak, A. M., & Qasem, S. N. (2024). Machine learning-based predictive models for detection of cardiovascular diseases. *Diagnostics*, *14*(2), 144.

[7] Belov, V., Erwin-Grabner, T., Aghajani, M., Aleman, A., Amod, A. R., Basgoze, Z., ... & ENIGMA Major Depressive Disorder working group https://enigma. ini. usc. edu/ongoing/enigma-mdd-working-group/. (2024). Multi-site benchmark classification of major depressive disorder

using machine learning on cortical and subcortical measures. *Scientific reports*, *14*(1), 1084.

[8] Khan, S., & Alqahtani, S. (2024). Hybrid machine learning models to detect signs of depression. *Multimedia Tools and Applications*, *83*(13), 38819-38837.

[9] Lorenzoni, G., Tavares, C., Nascimento, N., Alencar, P., & Cowan, D. (2024). Assessing ML Classification Algorithms and NLP Techniques for Depression Detection: An Experimental Case Study. *arXiv preprint arXiv:2404.04284*.

[10] Ullah, K., Ahsan, M., Hasanat, S. M., Haris, M., Yousaf, H., Raza, S. F., ... & Ullah, Z. (2024). Short-Term Load Forecasting: A Comprehensive Review and Simulation Study with CNN-LSTM Hybrids Approach. *IEEE Access*.

[11] Hanifi, S., Cammarono, A., & Zare-Behtash, H. (2024). Advanced hyperparameter optimization of deep learning models for wind power prediction. *Renewable Energy*, *221*, 119700.

[12] Mehta, N., Lorraine, J., Masson, S., Arunachalam, R., Bhat, Z. P., Lucas, J., & Zachariah, A. G. (2024). Improving hyperparameter optimization with checkpointed model weights. *arXiv preprint arXiv:2406.18630*.

[13] Khoo, L. S., Lim, M. K., Chong, C. Y., & McNaney, R. (2024). Machine learning for multimodal mental health detection: a systematic review of passive sensing approaches. *Sensors*, *24*(2), 348.

[14] Nusrat, M. O., Shahzad, W., & Jamal, S. A. (2024). Multi Class Depression Detection Through Tweets using Artificial Intelligence. *arXiv preprint arXiv:2404.13104*.

[15] Pande, S. D., Hasane Ahammad, S. K., Gurav, M. N., Faragallah, O. S., Eid, M. M., & Rashed, A. N. Z. (2024). Depression detection based on social networking sites using data mining. *Multimedia tools and applications*, *83*(9), 25951-25967.

[16] Meshram, P., & Rambola, R. K. (2023). Diagnosis of depression level using multimodal approaches using deep learning techniques with multiple selective features. *Expert Systems*, *40*(4), e12933.

[17] Iyortsuun, N. K., Kim, S. H., Jhon, M., Yang, H. J., & Pant, S. (2023, January). A review of machine learning and deep learning approaches on mental health diagnosis. In *Healthcare* (Vol. 11, No. 3, p. 285). MDPI.

[18] Gedam, S., & Paul, S. (2021). A review on mental stress detection using wearable sensors and machine learning techniques. *IEEE Access*, *9*, 84045-84066.

[19] Limenih, G., MacDougall, A., Wedlake, M., & Nouvet, E. (2024). Depression and Global Mental Health in the Global South: A Critical Analysis of Policy and Discourse. International Journal of Social Determinants of Health and Health Services, 54(2), 95-107.

[20] Cosgrove, L., Patterson, E. H., & Bursztajn, H. J. (2024). Industry influence on mental health research: depression as a case example. Frontiers in Medicine, 10, 1320304.

[21] Tejaswini, V., Sathya Babu, K., & Sahoo, B. (2024). Depression detection from social media text analysis using natural language processing techniques and hybrid deep learning model. ACM Transactions on Asian and Low-Resource Language Information Processing, 23(1), 1-20.

[22] Ahmed, T., Poly, I. H., Mim, M. Z., & Faruk, M. O. (2024). Suicide & Depression Detection Using Machine Learning (Doctoral dissertation, East West University).

[23] Das, A. K., & Naskar, R. (2024). A deep learning model for depression detection based on MFCC and CNN generated spectrogram features. *Biomedical Signal Processing and Control*, *90*, 105898.

[24] Ksibi, A., Zakariah, M., Menzli, L. J., Saidani, O., Almuqren, L., & Hanafieh, R. A. M. (2023). Electroencephalography-based depression detection using multiple machine learning techniques. *Diagnostics*, *13*(10), 1779.

[25] Marriwala, N., & Chaudhary, D. (2023). A hybrid model for depression detection using deep learning. *Measurement: Sensors*, *25*, 100587.

[26] Rajawat, A. S., Bedi, P., Goyal, S. B., Bhaladhare, P., Aggarwal, A., & Singhal, R. S. (2023). Fusion fuzzy logic and deep

learning for depression detection using facial expressions. *Procedia Computer Science*, *218*, 2795-2805.

[27] Wani, M. A., ELAffendi, M. A., Shakil, K. A., Imran, A. S., & Abd El-Latif, A. A. (2022). Depression screening in humans with AI and deep learning techniques. *IEEE transactions on computational social systems*, *10*(4), 2074-2089.

[28] Nadeem, A., Naveed, M., Islam Satti, M., Afzal, H., Ahmad, T., & Kim, K. I. (2022). Depression detection based on hybrid deep learning SSCL framework using self-attention mechanism: An application to social networking data. *Sensors*, *22*(24), 9775.

[29] Amanat, A., Rizwan, M., Javed, A. R., Abdelhaq, M., Alsaqour, R., Pandya, S., & Uddin, M. (2022). Deep learning for depression detection from textual data. *Electronics*, *11*(5), 676.

[30] Akbari, H., Sadiq, M. T., & Rehman, A. U. (2021). Classification of normal and depressed EEG signals based on centered correntropy of rhythms in empirical wavelet transform domain. *Health Information Science and Systems*, *9*, 1-15.

[31] Ahmed, A., Sultana, R., Ullas, M. T. R., Begom, M., Rahi, M. M. I., & Alam, M. A. (2020, December). A machine learning approach to detect depression and anxiety using supervised learning. In *2020 IEEE Asia-Pacific Conference on Computer Science and Data Engineering (CSDE)* (pp. 1-6). IEEE.

[32] Shen, J., Zhang, X., Wang, G., Ding, Z., & Hu, B. (2019). An improved empirical mode decomposition of electroencephalogram signals for depression detection. *IEEE transactions on affective computing*, *13*(1), 262-271.