BODHIVRUKSHA JOURNAL OF DIVERSE DISCIPLINE
E-ISSN : 3139-1486
An Open Access, Peer-Reviewed, Multidisciplinary Scholarly Journal
Bi-monthly | Multilingual (English, Hindi, Marathi) | Academic Research
@ Volume 2, Issue 5 (September-October 2026) (Special Issue 1) ISSN: 3139-1486 (Online)

Generative AI for Early Detection of Depression Using Multimodal Data: A Review

Soha Ghaniwale

Student
Department of Artificial Intelligence and Data Science
Ajeenkya DY Patil School of Engineering, Pune , India

Prof.Ashima Mahendra

Guide
Department of Artificial Intelligence and Data Science
Ajeenkya DY Patil School of Engineering, Pune, India

Abstract

Depression is a major mental-health condition that affects emotion, behaviour, communication and daily functioning, and its early identification remains a significant challenge for both clinicians and researchers. Conventional assessment relies mainly on clinical interviews and self-report instruments such as the PHQ and BDI, which are subjective and dependent on the willingness and ability of an individual to accurately report symptoms. Advances in Generative Artificial Intelligence (GenAI), deep learning and multimodal data fusion have opened new possibilities for analysing text, audio, video and behavioural signals together to support early depression-risk detection. This paper presents a structured review of recent literature on multimodal and Generative-AI-based approaches for depression detection, tracing the progression from traditional machine-learning techniques to deep-learning, transformer-based and generative architectures. Eight representative studies are analysed and compared in terms of methodology, key findings and limitations. Based on this synthesis, key research gaps are identified, including limited integration of GenAI with rich multimodal information, scarcity of large and diverse real-world datasets, generalization difficulties, GenAI reliability concerns and privacy and ethical issues associated with sensitive mental-health data. A conceptual multimodal GenAI framework that combines text, audio, video and behavioural indicators while addressing reliability, privacy and explainability is subsequently outlined as a direction for future work.

Keywords: Generative AI; Multimodal Data Fusion; Depression Detection; Mental-Health Informatics; Large Language Models; Deep Learning; Transformer Networks

I. INTRODUCTION

Depression is a major mental-health condition that can affect a person's emotions, behaviour, communication and day-to-day activities, and its prevalence is particularly notable among students and corporate workers who face sustained academic or occupational stress. Left undetected, depression can progressively affect an individual's quality of life, productivity and physical well-being, which makes early and reliable identification an important research and public-health priority.

BODHIVRUKSHA JOURNAL OF DIVERSE DISCIPLINE
E-ISSN : 3139-1486
An Open Access, Peer-Reviewed, Multidisciplinary Scholarly Journal
Bi-monthly | Multilingual (English, Hindi, Marathi) | Academic Research
@ Volume 2, Issue 5 (September-October 2026) (Special Issue 1) ISSN: 3139-1486 (Online)

Traditional depression assessment mainly depends on clinical interviews and self-assessment questionnaires such as the Patient Health Questionnaire (PHQ) and the Beck Depression Inventory (BDI). While clinically validated, these approaches depend heavily on what a person chooses or is able to report, on the availability of a trained professional, and on subjective clinical judgement, all of which can delay detection or introduce inconsistency.

Recent research has therefore explored the use of additional, more objective information sources, including linguistic patterns in text, acoustic characteristics of speech, facial and visual cues, and behavioural indicators such as sleep and activity patterns. Combining these different sources is known as multimodal analysis. Generative AI and modern deep-learning models can further analyse such multimodal information, learning complex cross-modal patterns to support early depression-risk identification, broadly summarised as text, audio, video and behavioural data being processed through AI-based analysis to obtain a depression-risk assessment.

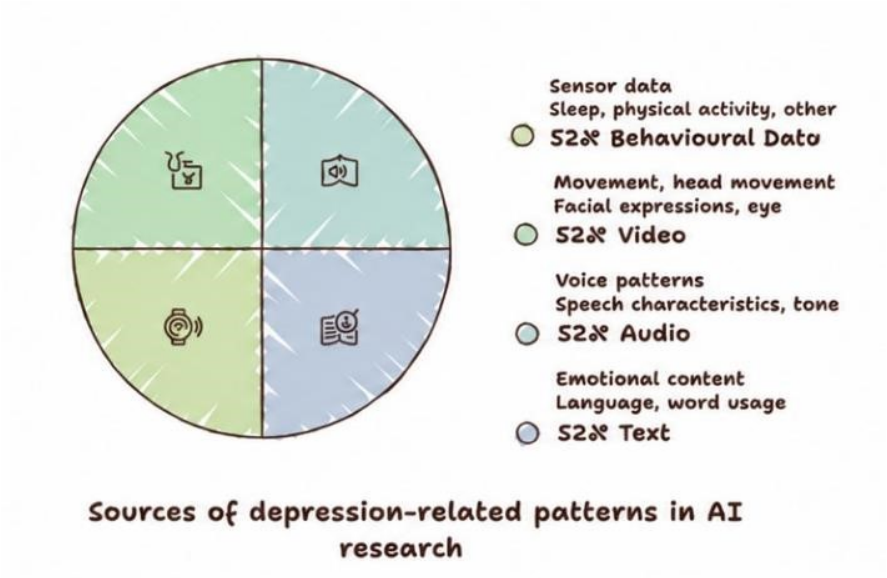


Fig. 1 Modalities commonly used for depression-related pattern analysis in AI research: text, audio, video and behavioural data.

This paper reviews recent literature on multimodal and Generative-AI-based depression detection, identifies the key research gaps that remain, and outlines a conceptual framework that combines multimodal indicators with Generative AI capabilities while accounting for reliability, privacy and ethical considerations.

II. Problem Statement and Motivation

Existing depression-detection methods still face several limitations. Traditional assessment can be subjective and dependent on self-reported information, while automated approaches often analyse only a single type of information, such as text alone or speech alone, limiting the range of symptoms that can be captured.

At the same time, multimodal approaches that attempt to combine several information sources face additional challenges, including limited and small datasets, privacy concerns involving sensitive user data,

BODHIVRUKSHA JOURNAL OF DIVERSE DISCIPLINE
E-ISSN : 3139-1486
An Open Access, Peer-Reviewed, Multidisciplinary Scholarly Journal
Bi-monthly | Multilingual (English, Hindi, Marathi) | Academic Research
@ Volume 2, Issue 5 (September-October 2026) (Special Issue 1) ISSN: 3139-1486 (Online)

difficulty in combining different modalities effectively, limited generalization to real-world environments, and reliability and interpretability concerns in Generative-AI systems used for sensitive mental-health applications.

There is, therefore, a need to investigate how Generative AI combined with multimodal information can contribute to more comprehensive early depression-risk detection while explicitly considering accuracy, privacy, reliability and ethical limitations. Formally, the problem addressed by this review may be stated as follows: to study and explore a Generative-AI-based multimodal approach for early detection of depression by analysing text, audio, visual and behavioural indicators, while addressing limitations related to accuracy, privacy, reliability and ethical use.

III. Objectives of the Review

The objectives of this review are as follows:

1. To study existing AI and Generative-AI approaches used for depression and mental-health risk detection.
2. To identify depression-related indicators available in text, speech, facial behaviour and behavioural data.
3. To study multimodal data-fusion techniques for combining information from different modalities.
4. To understand the role of Generative AI and Large Language Models (LLMs) in mental-health assessment and early detection.
5. To identify limitations and research gaps in existing approaches related to accuracy, datasets, privacy, bias and reliability.
6. To explore a conceptual multimodal GenAI framework for supporting early depression-risk identification.

IV. Relevance to Current Technology Trends

The topic of this review is closely aligned with several rapidly evolving areas of artificial-intelligence research, which together motivate the study of Generative-AI-based multimodal depression detection.

A. Generative AI and Large Language Models

Modern GenAI models can understand and generate contextual natural language, creating opportunities for analysing mental-health-related conversations and text for early indicators of risk.

B. Multimodal Artificial Intelligence

Current AI research is moving beyond text-only systems toward models capable of processing multiple modalities such as text, audio and visual information jointly, which is directly relevant to capturing the diverse symptomatology of depression.

C. Transformer-Based Models

BODHIVRUKSHA JOURNAL OF DIVERSE DISCIPLINE
E-ISSN : 3139-1486
An Open Access, Peer-Reviewed, Multidisciplinary Scholarly Journal
Bi-monthly | Multilingual (English, Hindi, Marathi) | Academic Research
@ Volume 2, Issue 5 (September-October 2026) (Special Issue 1) ISSN: 3139-1486 (Online)

Transformer architectures and attention mechanisms are increasingly used for learning complex patterns and for combining multimodal information in a unified representation, forming the backbone of many recent multimodal depression-detection systems.

D. AI in Healthcare

GenAI is increasingly being investigated for diagnosis and assessment support, therapeutic tools and clinician support, positioning mental-health screening as an important application area for responsible AI deployment.

Fig. 2 Approximate distribution of AI research areas relevant to this review: Generative AI and LLMs, multimodal AI, transformer-based models, AI in healthcare, and NLP and deep learning.

In summary, the seminar and review topic sits at the intersection of Generative AI, multimodal AI, natural-language processing, deep learning and healthcare AI, each of which is currently among the most active areas of AI research.

V. Literature Survey

Research in automated depression detection has progressed through several distinct stages: traditional machine learning, deep learning, multimodal deep learning, transformer-based models, and, most recently, Generative AI. Earlier studies mainly used handcrafted audio, visual and textual features together with conventional machine-learning algorithms. Deep-learning approaches later introduced Convolutional Neural Networks (CNNs), Deep Neural Networks (DNNs) and Long Short-Term Memory (LSTM) networks for automatic feature extraction, reducing the dependence on manual feature engineering.

More recent research combines audio, video and text using attention mechanisms and Transformer-based multimodal architectures, allowing models to learn cross-modal relationships directly from data. In parallel, Generative AI and LLMs are being investigated for mental-health detection, assessment and conversational support, extending the scope of automated systems beyond classification toward interactive and generative capabilities. However, challenges involving dataset availability, multimodal fusion, real-world generalization, privacy, bias and GenAI reliability remain largely unresolved. Tables I and II summarise eight representative studies drawn from this literature.

TABLE I. LITERATURE SURVEY SUMMARY — PART 1

Author / Year	Methodology Used	Key Findings	Limitations
Matthew Purver, 2025	Survey of multimodal ML datasets, models and fusion approaches	Combining modalities provides complementary information for mental-health analysis	Privacy, fairness, explainability, limited longitudinal/representative data
Le Yang, Dongmei Jiang, 2017	DCNN–DNN fusion using audio, video and text	Multimodal framework outperformed baseline depression-measurement approaches	Older work; further text/fusion and end-to-end learning needed

BODHIVRUKSHA JOURNAL OF DIVERSE DISCIPLINE
E-ISSN : 3139-1486
An Open Access, Peer-Reviewed, Multidisciplinary Scholarly Journal
Bi-monthly | Multilingual (English, Hindi, Marathi) | Academic Research
@ Volume 2, Issue 5 (September-October 2026) (Special Issue 1) ISSN: 3139-1486 (Online)

Author / Year	Methodology Used	Key Findings	Limitations
Lang He, Kai Chen, 2024	LMVD dataset with Transformer-based MDDformer using audiovisual features	Real-world vlog data supports depression recognition; MDDformer outperformed compared baselines	Further dataset expansion and multimodal behavioural exploration needed
Tanmay Bhanushali, 2024	Scoping review of GenAI applications in mental health	GenAI used for detection, counselling, therapy and decision support	Detection evidence remains limited; GenAI should supplement professionals

TABLE II. LITERATURE SURVEY SUMMARY — PART 2

Author / Year	Methodology Used	Key Findings	Limitations
Wang, Zhou, 2025	Systematic review of 79 GenAI mental-health studies	GenAI shows potential in assessment, therapeutic tools and clinician support	Privacy, bias, safety, limited real-world validation
Santosh Kumar, 2025	NLP with sentiment analysis and behavioural/wearable/EHR data	Proposes a multimodal GenAI framework for early depression/anxiety detection	Primarily a proposed framework; requires stronger real-world validation
Liying Wang, Tanmay Bhanushali, 2025	Systematic review of GenAI capabilities	GenAI shows strengths in psychoeducation and emotional awareness	Diagnostic accuracy, cultural competence, trust and emotional engagement remain limited
Zhang et al., 2024	AVTF-TBN combining audio, video and text	Multimodal model achieved F1 = 0.78, precision = 0.76, recall = 0.81	Dataset and fusion limitations; further generalization needed

Across the reviewed studies, a consistent pattern emerges: multimodal approaches that fuse two or more of text, audio, video and behavioural signals tend to outperform single-modality baselines, and Transformer-based fusion architectures such as MDDformer and AVTF-TBN report the strongest quantitative results. At the same time, GenAI-focused reviews indicate that, while large language models show promise for psychoeducation, conversational support and clinician assistance, their direct use for diagnostic-grade detection is still comparatively under-validated, and most GenAI-based mental-health frameworks proposed in the literature remain conceptual rather than empirically deployed and validated at scale.

BODHIVRUKSHA JOURNAL OF DIVERSE DISCIPLINE
E-ISSN : 3139-1486
An Open Access, Peer-Reviewed, Multidisciplinary Scholarly Journal
Bi-monthly | Multilingual (English, Hindi, Marathi) | Academic Research
@ Volume 2, Issue 5 (September-October 2026) (Special Issue 1) ISSN: 3139-1486 (Online)

VI. Research Gap

Based on the literature synthesised above, five interrelated research gaps can be identified.

1) Limited multimodal GenAI integration: Many studies focus either on multimodal deep learning or on text-based GenAI in isolation. Integration of modern GenAI with rich multimodal information remains comparatively less explored.

2) Limited real-world datasets: Depression datasets are often small because mental-health information is highly sensitive and difficult to collect at scale, limiting the statistical power and diversity of trained models.

3) Generalization issues: Models developed using controlled clinical interviews or specific social-media populations may not generalize effectively to diverse, real-world users across different demographics and cultures.

4) GenAI reliability: Current GenAI systems can produce inaccurate or inconsistent outputs, making reliability particularly important when applied to mental-health applications where errors carry real consequences.

5) Privacy and ethical concerns: Text, voice, facial information, health records and behavioural information are sensitive data and require strong privacy safeguards, informed consent and careful ethical oversight.

Main gap identified: there is a need for a reliable, privacy-aware and explainable approach that combines multimodal depression indicators with the capabilities of Generative AI for early risk identification, rather than relying on a single modality or on GenAI alone.

VII. Proposed Conceptual Framework

To address the gap identified above, this review outlines a conceptual multimodal GenAI framework for early depression-risk identification. As illustrated in Fig. 3, text, audio, video and behavioural data are first processed independently to extract modality-specific features, which are then combined within a multimodal data-fusion layer. The fused representation is analysed by a Generative-AI or LLM-based component, which supports contextual reasoning across modalities and can generate an interpretable depression-risk output rather than a single opaque score.

BODHIVRUKSHA JOURNAL OF DIVERSE DISCIPLINE
E-ISSN : 3139-1486
An Open Access, Peer-Reviewed, Multidisciplinary Scholarly Journal
Bi-monthly | Multilingual (English, Hindi, Marathi) | Academic Research
@ Volume 2, Issue 5 (September-October 2026) (Special Issue 1) ISSN: 3139-1486 (Online)

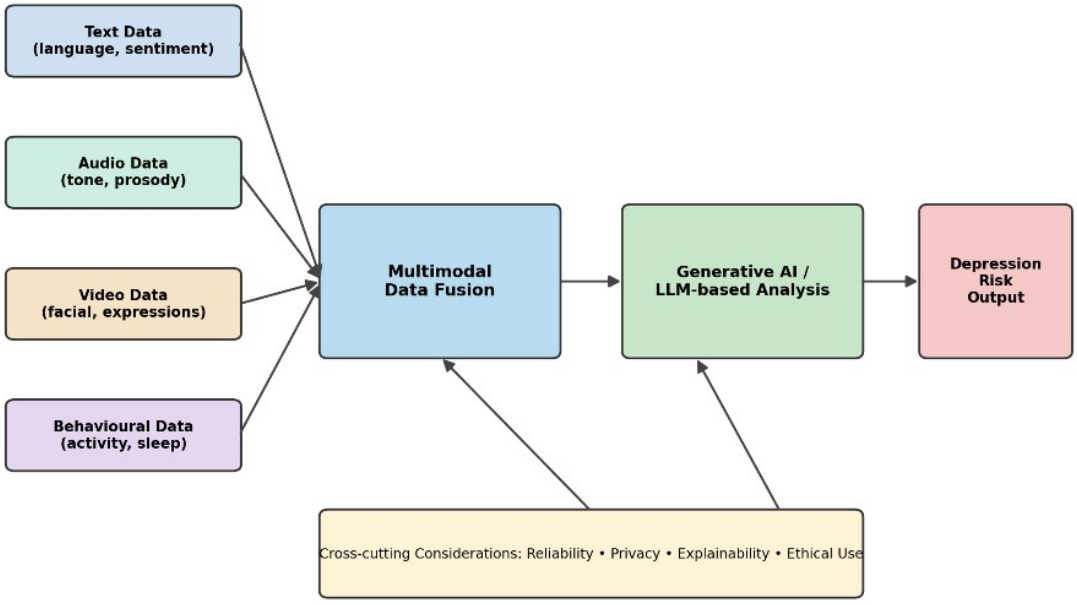


Fig. 3 Conceptual multimodal Generative-AI framework for early depression-risk identification.

Reliability, privacy and explainability are treated as cross-cutting design considerations rather than optional add-ons: privacy-preserving handling of sensitive text, voice, facial and behavioural data is required throughout the pipeline, model outputs should be explainable to support clinician trust, and reliability checks are needed to reduce inconsistent or inaccurate GenAI outputs before any risk indication reaches an end user. This framework is conceptual in nature and is intended to guide subsequent empirical work rather than to represent a validated system.

VIII. Conclusion

This review has examined recent progress in Generative-AI-based and multimodal approaches for early depression detection. The seminar topic and domain were finalized, and research papers related to Generative AI, multimodal depression detection and mental-health AI were studied in detail. Existing approaches using text, audio, video and behavioural information were analysed, and different techniques including CNNs, DNNs, LSTMs, BERT-based models, Transformers and multimodal fusion strategies were compared in terms of their advantages and limitations.

Major research gaps involving multimodal-GenAI integration, dataset limitations, privacy, reliability and real-world generalization were identified. Overall, the literature survey establishes the foundation for exploring a Generative-AI-based multimodal approach for early depression-risk detection, and the conceptual framework proposed in this review provides a direction for future empirical validation that balances detection performance with reliability, privacy and ethical responsibility.

References

[1] Z. Al Sahili, I. Patras, and M. Purver, “A survey of multimodal machine learning for mental health,” *arXiv preprint, 2025.*

BODHIVRUKSHA JOURNAL OF DIVERSE DISCIPLINE
E-ISSN : 3139-1486
An Open Access, Peer-Reviewed, Multidisciplinary Scholarly Journal
Bi-monthly | Multilingual (English, Hindi, Marathi) | Academic Research
@ Volume 2, Issue 5 (September-October 2026) (Special Issue 1) ISSN: 3139-1486 (Online)

[2] L. Yang, D. Jiang, et al., "Multimodal measurement of depression using deep learning models," in *Proc. AVEC*, 2017, pp. 53–59.

[3] L. He, K. Chen, et al., "LMVD: A large-scale multimodal vlog dataset for depression detection in the wild," *ACM Multimedia*, 2024.

[4] X. Xian, T. Bhanushali, et al., "Debate and dilemmas regarding generative AI in mental health care: Scoping review," *Interactive Journal of Medical Research*, 2024.

[5] X. Wang, Y. Zhou, and G. Zhou, "The application and ethical implication of generative AI in mental health: Systematic review," *JMIR Mental Health*, 2025.

[6] S. Kumar, "Early detection of depression and anxiety in the USA using generative AI," *International Journal of Research in Engineering*, 2025.

[7] L. Wang, T. Bhanushali, et al., "Evaluating generative AI in mental health: Systematic review of capabilities and limitations," *JMIR Mental Health*, 2025.

[8] Z. Zhang et al., "Multimodal sensing for depression risk detection: Integrating audio, video, and text data," *Sensors*, vol. 24, art. 3714, 2024.