International Journal for Multidisciplinary Research (IJFMR)
E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

AI-Powered Depression Detection Using Text and Speech

Durgunala Ranjith[1], Nalamasu Tharun[2], Baswa Shruthika[3], Sama Vaishnavi[4]

[1]Assistant Professor Of Department Of CSE (AL & ML) Of ACE Engineering College, India.
[2,3,4]Students Of Department Of CSE (AL & ML) Of ACE Engineering College, India.

Abstract

Depression is a growing global concern, affecting millions of individuals and often going undiagnosed due to stigma and lack of access to mental health professionals. This project aims to develop an AI-powered system for early detection of depression using text and speech analysis. The model leverages Natural Language Processing (NLP) to analyze textual input for depressive language patterns and Machine Learning (ML) techniques to detect vocal features associated with depression from speech samples. By integrating sentiment analysis, emotion detection, and acoustic feature extraction, the system can assess linguistic and vocal cues indicative of depressive tendencies. The model will be trained on datasets containing both text-based conversations and speech recordings labeled for depression severity. The goal is to create an assistive tool that can help mental health professionals with early screening and intervention, ultimately improving mental health outcomes.

Keywords: AI-powered system, Early detection, Text analysis, Speech analysis, Natural Language Processing (NLP), Machine Learning (ML)

1. INTRODUCTION

Depression is a prevalent and serious mental health condition that affects over 280 million people globally, according to the World Health Organization (WHO). Despite its widespread impact, depression often goes undiagnosed and untreated, largely due to social stigma, limited awareness, and insufficient access to mental health professionals. Early detection plays a critical role in effective treatment and improved patient outcomes. With advancements in artificial intelligence (AI), there is growing potential to develop intelligent systems that can assist in the early screening of depression.

This project aims to design and implement an AI-powered system that analyzes both text and speech to identify early signs of depression. Utilizing Natural Language Processing (NLP), the system examines written inputs for depressive language patterns, while Machine Learning (ML) algorithms analyze speech recordings to detect acoustic markers such as tone, pitch, and speech rate associated with depressive states. By combining sentiment analysis, emotion recognition, and vocal feature extraction, the system provides a comprehensive assessment of a user's mental state.

2. Literature Survey

2.1. Utnal et al. [1] Introduced a method for sentiment analysis and hate speech detection on Twitter using

Natural Language Processing (NLP). The framework extracts semantic, sentiment, and unigram features for both binary and ternary classification. The findings highlight that while effective models exist, no perfect solution has been found, emphasizing the need for richer datasets and integrated features to improve accuracy.

2.2. Sharma et al. [2] Developed the D-SCAN system, which integrates face, audio, and text-based emotion detection to identify depression through multimodal signals. The model provides personalized well-being recommendations and is scalable, offering significant benefits for mental health support.

2.3. Singh [3] Proposed a methodology for early detection and diagnosis of mental health status using NLP-based methods with ML classifiers. The system utilizes a BILSTM with an attention mechanism on both audio and text features, fused through an Additive Attention Cross-Modal Network (ACMA). The study found that Logistic Regression outperformed other models, achieving 80% accuracy, although its performance was lower than expected across two datasets.

2.4. Sethu [4] Introduced a natural language processing approach for depression detection based on acoustic and landmark event-based features in speech. By combining spectral and articulatory speech cues, the model achieved improved classification accuracy, with up to 15% improvement in F1 scores. The study also noted a shift from rule-based and classical ML models to deep learning and multimodal systems. However, a lack of standardized benchmarks hinders its generalizability across different demographics and platforms.

2.5. Khalid [5] Reviewed machine learning and deep learning techniques for depression detection in social media. This comprehensive review explores various methods, discussing their strengths and limitations in the context of mental health detection on digital platforms.

2.6. Utnal et al. [6] Proposed a deep Recurrent Neural Network (RNN) framework utilizing Mel-Frequency Cepstral Coefficients (MFCCs) for detecting depression and assessing its severity from speech signals. The model achieved an accuracy of 76.27% on the DAIC-WOZ dataset, demonstrating the effectiveness of MFCC features in depression recognition. However, the approach primarily focused on speech data, limiting its applicability to multimodal analysis.

2.7. Zhou et al. [7] Proposed TAMFN (Time-Aware Attention Multimodal Fusion Network) for detecting depression based on non-verbal behaviors (acoustic and visual) in vlogs. The model incorporates a Global Temporal Convolutional Network (GTCN), Intermodal Feature Extraction (IFE), and a Time-Aware Attention Multimodal Fusion (TAMF) module. GTCN captures both local and global temporal patterns, IFE enriches early intermodal interactions, and TAMF dynamically fuses multimodal data based on time-aware attention. Tested on the D-Vlog dataset, TAMFN outperformed all benchmark models, confirming its superior performance in multimodal depression detection.

2.8. Skaik and Inkpen [8] Developed a system to automatically fill out the Beck's Depression Inventory (BDI) questionnaire using the eRisk 2021 Task 3 dataset. The method involved training separate models for groups of questionnaire items and consolidating them into a unified model (BDI_Multi_Model). This model achieved state-of-the-art results in automatically detecting depression and was applied to a Canadian population dataset. The predictions showed a strong Pearson correlation (0.90) with official national mental health statistics, demonstrating its reliability and potential for large-scale public health monitoring.

2.9. Penava and Buettner [9] Proposed a novel Convolutional Neural Network (CNN) architecture for detecting early-stage non-severe depression using resting-state EEG data. The model was developed with ethical approval and based on real EEG recordings, offering a promising direction for non-invasive and early detection of depression. Results showed that the CNN model could accurately distinguish between depressed and non-depressed individuals, offering an alternative to subjective questionnaires.

2.10. Iyortsuun et al. [10] Introduced the Additive Cross-Modal Attention (ACMA) Network for depression detection using audio and text data. The model uses BiLSTM backbones and an additive attention mechanism to capture intricate cross-modal interactions without relying on preset questionnaires. It was evaluated on the DAIC-WOZ and EATD-Corpus datasets, showing strong performance in both classification and regression tasks. The study highlights ACMA's potential to detect depression from natural conversations, providing a more authentic and passive diagnostic tool.

3. Comparison Table: Literature Review on AI-Based Depression Detection

S.No	Title	Methodology Used	Key Findings	Limitations
1	Sentiment Analysis & Hate Speech Detection on Twitter	NLP-based sentiment, semantic, and unigram feature extraction for classification	Highlights potential of sentiment analysis models for detection tasks	Lack of perfect model; needs richer datasets and feature integration
2	D-SCAN: Depression Detection	Multimodal (face, audio, text) emotion analysis	Personalized well-being suggestions; effective for scalable mental health support	Scalability tested; performance on diverse datasets not detailed
3	Early Detection using NLP and ML Classifiers	Logistic Regression, BiLSTM + attention, ACMA model for SMS-based classification	Logistic Regression achieved 80% accuracy in SMS-based depression detection	Accuracy lower than expected; limited to textual input
4	Acoustic & Landmark Event-Based Speech Features	Spectral and articulatory speech cues combined with NLP	Achieved 15% F1 score improvement, robust in noisy environments	Lack of standard benchmarks and demographic diversity
5	Review on ML and DL for Depression in Social Media	Literature review of rule-based, ML, DL, and multimodal techniques	Shows shift toward deep learning and explainable AI systems	General review; does not propose or evaluate a specific model
6	RNN with MFCC for Depression Detection	Deep RNN + MFCC features on speech data	Achieved 76.27% accuracy on DAIC-WOZ dataset	Limited to unimodal (speech) analysis
7	TAMFN: Time-Aware Attention Multimodal Fusion	GTCN + IFE + TAMF modules to fuse vlog acoustic/visual signals	Outperformed benchmark models on D-Vlog dataset	Limited to non-verbal behaviors in vlogs
8	Predicting Depression via Auto-Filled BDI Questionnaire	BDI_Multi_Model using NLP on eRisk data	0.90 Pearson correlation with official stats; scalable population-level tool	Dependent on questionnaire-based design; data availability limits

				model generalization
9	CNN-Based Early-Stage Detection from EEG	CNN on resting-state EEG signals	Effective early-stage detection; ethically approved real EEG data	Requires EEG hardware; clinical environment needed
10	ACMA for Audio & Text-Based Depression Detection	BiLSTM + Additive Attention across audio and text on DAIC-WOZ and EATD datasets	High performance in classification and regression; no preset questionnaire needed	Evaluation limited to two datasets; generalizability to real-world scenarios not assessed

4. Research Gaps

Despite notable progress in AI-based depression detection, several research gaps persist. Many models focus on unimodal inputs (e.g., only text or speech), limiting the richness of analysis. Multimodal approaches show promise but face challenges in data fusion, synchronization, and computational efficiency. Additionally, a lack of standardized, diverse, and ethically sourced datasets hinders model generalizability across populations. Most studies evaluate models in controlled settings, leaving their real-world applicability uncertain. Furthermore, explainability and transparency remain underexplored, which affects clinical trust and adoption. Addressing these gaps is crucial for developing robust, scalable, and ethically responsible AI tools for mental health support.

5. Proposed Methodology

This study proposes a **multimodal deep learning framework** that integrates **text, audio, and facial expression data** for accurate depression detection. Textual features will be extracted using **BERT embeddings**, while audio features will be captured using **MFCCs and prosodic cues**. Facial features will be analyzed through a **CNN-based visual emotion recognition model**. These features will be fused using an **attention-based fusion mechanism** to emphasize the most indicative signals across modalities. The model will be trained and evaluated on benchmark datasets such as **DAIC-WOZ** and **EATD-Corpus**, using metrics like accuracy, F1-score, and AUC. This approach aims to improve detection accuracy, robustness, and real-world applicability.

6. Conclusion

The development of an AI-powered system for early depression detection using text and speech analysis holds significant promise for enhancing mental health care. By leveraging advanced techniques in Natural Language Processing (NLP) and Machine Learning (ML), the system can provide timely and accurate assessments of depression severity, aiding mental health professionals in early diagnosis and intervention. This approach combines both textual and vocal cues, offering a more holistic view of an individual's emotional state. While challenges such as data diversity, model generalization, and explainability remain, addressing these gaps will contribute to more effective and inclusive mental health tools. Ultimately, this AI system can serve as a valuable assistive tool, supporting better mental health outcomes through early detection and personalized care, without replacing the need for clinical judgment.

References

1. S. Utnal, S. Yedle, S. Patle, P. Paslakar, H. Kumbhar, and O. Omkar, "Sentiment Analysis Hate Speech Detection on Twitter Using NLP," 2022 International Journal of Creative Research Thoughts (IJCRT).
2. A. Sharma, A. Devi, and E. Ashwini, "D-SCAN: Depression Detection," International Journal of Scientific Research in Engineering and Management (IJSREM), 2021.
3. A. Singh, "The Early Detection and Diagnosis of Mental Health Status Employing NLP-Based Methods With ML Classifiers," International Journal of Creative Research Thoughts (IJCRT), vol. 9, no. 6, pp. 123–129, 2021.
4. N. K. Iyortsuu, S.-H. Kim, H.-J. Yang, S.-W. Kim, and M. Jhon, "Additive Attention Cross-Modal Network (ACMA) for Depression Detection Based on Audio and Textual Features," IEEE Access, vol. 11, pp. 5890–5905, 2023.
5. V. S. Sethu, "Natural Language Processing Methods for Acoustic and Landmark Event-Based Features in Speech-Based Depression Detection," Proceedings of the Annual Conference on Signal Processing, 2022.
6. S. Khalid, "Depression Detection in Social Media: A Comprehensive Review of Machine Learning and Deep Learning Techniques," IEEE Access, vol. 9, pp. 103452–103471, 2021.
7. I. Rejaibi, M. Jaoua, and H. Mahdi, "Depression Severity Estimation from Speech Using Deep Recurrent Neural Networks with MFCC Features," Procedia Computer Science, vol. 192, pp. 3890–3895, 2021.
8. L. Zhou, Z. Liu, Z. Shangguan, X. Yuan, Y. Li, and B. Hu, "TAMFN: Time-Aware Attention Multimodal Fusion Network for Depression Detection," IEEE Transactions on Affective Computing, 2023.
9. R. S. Skaik and D. Inkpen, "Predicting Depression in Canada by Automatic Filling of Beck's Depression Inventory Questionnaire," IEEE Access, vol. 10, pp. 80942–80957, 2022.
10. P. Penava and R. Buettner, "Early-Stage Non-Severe Depression Detection Using a Novel Convolutional Neural Network Approach Based on Resting-State EEG Data," IEEE Access, vol. 11, pp. 10070–10082, 2023