Deep Learning Analysis for Early Mental Health Disorder Detection via Voice Data

Neeta Namdeo Takawale

Department of Computer Science, Dr. D. Y. Patil Arts, Commerce and Science College, Pimpri, Pune-411018, Maharashtra, India

DOI: https://doi.org/10.51583/IJLTEMAS.2025.1413SP016

Received: 26 June 2025; Accepted: 30 June 2025; Published: 23 October 2025

Abstract: Mental health disorders such as depression, anxiety, and bipolar disorder significantly affect the well-being of individuals and often go undiagnosed due to reliance on subjective assessments. Voice data, being non-invasive and widely accessible, provides an excellent medium for detecting emotional and cognitive cues associated with mental health conditions. This research investigates the application of deep learning for analyzing vocal features to detect early signs of mental health disorders. Using publicly available datasets and spectrogram-based preprocessing, we evaluate Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and hybrid models. The results demonstrate the effectiveness of deep learning in identifying subtle vocal biomarkers and provide insights into real-time, scalable mental health screening tools.

Keywords: Mental Health, Deep learning, Voice data, early detection

I. Introduction

Mental health issues are a growing concern globally, with millions suffering from conditions such as depression and anxiety. Early detection and intervention are crucial for effective treatment. However, current diagnostic practices are often subjective and underutilized due to stigma and resource limitations [1][24].

Voice data, which naturally reflects emotions and cognitive states, has emerged as a potential indicator of psychological conditions. This paper focuses on the application of deep learning techniques to analyze voice recordings for the early detection of mental health disorders [2][4][23].

Mental health disorders, such as depression, anxiety, schizophrenia, and autism spectrum disorders (ASD), represent a significant and growing burden globally, impacting nearly 450 million individuals across all age groups. These disorders not only affect psychological well-being but also interfere with physical health, social interactions, and academic or occupational functioning [3][4][16]. According to the Global Burden of Disease reports, mental health conditions account for a substantial percentage of Disability-Adjusted Life Years (DALYs), with no evidence of decline in prevalence over the past decades. Early detection and timely intervention are crucial in mitigating their long-term effects. However, traditional diagnostic methods rely heavily on self-reported data and clinician interpretation, which can be subjective and insufficient for early-stage detection. The World Health Organization (WHO) emphasizes a global strategy to address non-communicable diseases (NCDs) and mental health conditions, advocating for data-driven healthcare policy and personalized care [1][6][17].

In this context, artificial intelligence (AI), particularly machine learning (ML) and deep learning (DL), is emerging as a transformative tool in mental health research and clinical decision-making. Deep learning, a subset of ML, mimics the human brain through artificial neural networks and is capable of learning complex, non-linear relationships from vast and diverse datasets. Its success in domains like image recognition, genomics, and speech analysis has made it a promising candidate for mental health applications, especially when dealing with unstructured data like voice recordings [11][18]. Deep learning methods can automatically extract meaningful features from raw audio data, identifying subtle patterns in tone, pitch, pauses, and speech rhythm—characteristics that often correlate with mental health status. Unlike conventional models, DL techniques offer scalable and objective ways to support early diagnosis, improve prognostic accuracy, and tailor interventions across diverse populations [3][7][12].

Recent studies have demonstrated the effectiveness of DL-based approaches using voice data to detect early symptoms of mental disorders among various populations, including college students—a group particularly vulnerable to mental health challenges [15][19][21]. By integrating voice data with behavioral and physiological data collected from counseling sessions, wearable devices, or mobile applications, researchers have developed predictive models capable of identifying risk factors associated with depression, anxiety, and suicidal tendencies. Despite these advancements, there remains a critical gap in comprehensive reviews that consolidate methods, outcomes, and limitations of DL applications across multiple mental health conditions. Therefore, this study aims to bridge this gap by presenting a systematic review of existing research that employs deep learning techniques for the early detection of mental health disorders through voice data analysis, highlighting current trends, challenges, and future directions in this evolving field [13][22][25].

Problem Statement

Mental health disorders such as depression, anxiety, and bipolar disorder often go undiagnosed or are detected at a late stage due to the subjective nature of existing diagnostic practices, limited access to mental health professionals, and the stigma surrounding psychological conditions. Traditional assessments heavily rely on self-reported symptoms and clinician judgment, which can be inconsistent and fail to capture early, subtle indicators of distress. In this context, there is a pressing need for objective, scalable, and non-invasive screening tools that can assist in early detection and timely intervention. Voice, as a rich and natural form of human expression, carries embedded emotional and cognitive signals that are often overlooked in clinical evaluations. This research aims to address this gap by leveraging deep learning techniques to analyze vocal patterns and identify early markers of mental health disorders. By developing an automated system capable of learning complex, non-linear relationships from voice inputs, the study seeks to enhance early detection capabilities, reducediagnostic delays, and support more accessible mental healthcare solutions.

II. Literature Review

Mental health disorders continue to pose a significant public health challenge across the globe. Non-communicable diseases (NCDs), including mental health conditions, contribute to nearly 50% of healthy life years lost, as quantified by Disability-Adjusted Life Years (DALYs), and are responsible for approximately two-thirds of all deaths worldwide. In the Americas alone, NCDs account for around 80% of total mortality. Mental health disorders specifically represented 14.4% of global disabilities in 2017. Despite various public health initiatives, such as the WHO's Global Action Plan for the Prevention and Control of NCDs, there has been no significant global decline in the burden of mental illnesses since 1990. These alarming figures highlight the critical need for early detection methods, especially considering the large gaps in diagnosis and treatment caused by a lack of timely and objective data to guide resource allocation.

Traditionally, the diagnosis of mental disorders has been largely subjective, relying on self-reported questionnaires and clinician assessments. However, the emergence of Artificial Intelligence (AI), particularly Machine Learning (ML) and Deep Learning (DL) techniques, has opened new avenues for objective, data-driven mental health diagnostics. Deep learning models, which are built upon artificial neural networks, can identify complex, non-linear relationships in large datasets and automatically extract significant features from raw inputs. These methods have demonstrated superior performance in various domains, including healthcare. Unlike conventional algorithms, deep learning approaches allow for end-to-end learning from unstructured data—such as speech, facial expressions, or physiological signals—making them well-suited for mental health applications.

Recent studies have demonstrated the growing utility of DL models in the mental health domain. In 2024, Zhang et al. developed CNN and LSTM-based models to detect early signs of depression among adolescents using neuroimaging data from over 50,000 electronic health records. Their models achieved remarkable results, with 92% F1-score and 97% AUC. Similarly, Satapathy et al. evaluated different models for classifying sleep disorders, including insomnia and sleep apnea, using EEG data, concluding that CNNs and RNNs significantly outperformed traditional methods. Hossain et al. proposed a hybrid deep learning model combining quantum and classical techniques to analyze static, sequential, and video-based facial expressions for emotional tracking—enhancing detection accuracy by aggregating model outputs.

Diwakar and Raj introduced a text classification system using DistilBERT for automated classification of disorders like autism, borderline personality disorder (BPD), and anxiety. Using a balanced dataset of 500 samples per class, their model achieved 96% accuracy and also investigated the gut-brain axis to explore physiological correlations. In another study, Peristeri et al. used gradient boosting (XGBoost) combined with natural language processing (NLP) on storytelling data to differentiate children with Autism Spectrum Disorder (ASD) from typically developing peers. Upadhyay et al. applied a stacking ensemble of SVMs on behavioral datasets to detect Persistent Depressive Disorder (PDD), finding higher incidence rates among nontechnical rural students and those from middle-income groups. Revathy et al. proposed a Dynamically Stabilized Recurrent Neural Network (DSRNN) using the OSMI dataset, focusing on frequency component relationships to distinguish between mentally ill and healthy individuals with improved feature extraction.

While many existing review articles focus on individual mental disorders—such as depression, anxiety, suicide, ASD, or Alzheimer's—there is a notable lack of comprehensive reviews covering a broader spectrum of conditions through deep learning approaches. Furthermore, few studies have explicitly focused on voice data**,** despite its rich potential for reflecting emotional and cognitive states. Vocal attributes such as pitch, tone, pauses, and rhythm offer a non-invasive, continuous, and cost-effective means of detecting mental health conditions. Thus, this review aims to bridge that gap by analyzing existing deep learning methodologies applied to mental health detection using voice data. The review not only highlights current models and their effectiveness but also identifies research gaps, methodological limitations, and future directions for improving diagnostic precision through voice-based DL systems.

III. Methodology

3.1 Dataset

This study utilizes widely recognized dataset for mental health and emotion analysis: the DAIC-WOZ (Distress Analysis

Interview Corpus) The DAIC-WOZ dataset contains audio recordings, transcripts, to detect psychological distress such as depression, and validating deep learning models aimed at early detection of mental health disorders using voice data.

3.2 Preprocessing

The raw audio data undergoes multiple preprocessing steps to ensure optimal model performance. First, voice normalization is applied to minimize variability in loudness and tone across samples. Then, key acoustic features such as Mel-frequency cepstral coefficients (MFCCs), chroma features, and zero-crossing rate are extracted to represent important speech characteristics. Additionally, spectrograms are generated from the audio signals to convert the temporal speech data into visual representations, which serve as input for the convolutional layers of the deep learning model.

3.3 Model Architecture

The proposed deep learning architecture consists of various layers used for effective voice data analysis. Initially, Convolutional Neural Network (CNN) layers are employed to process the spectrogram inputs, capturing spatial and frequency-based patterns. These outputs are then passed through Long Short-Term Memory (LSTM) layers to learn temporal dependencies and sequential features from the audio data. Finally, fully connectedlayers are used to map the extracted features to the target classification labels, enabling the model to predict the presence or absence of mental health disorders with high accuracy.

3.4 Evaluation Metrics

The performance of the model is evaluated using a range of standard classification metrics. Accuracy measures the overall correctness of predictions, while precision indicates the proportion of true positive predictions among all positive predictions. Recall (or sensitivity) assesses the model's ability to identify true positive cases among all actual positives. The F1-score, which is the harmonic mean of precision and recall, provides a balanced evaluation metric especially useful in imbalanced datasets. Finally, ROC-AUC (Receiver Operating Characteristic - Area Under the Curve) is used to measure the model's ability to distinguish between classes, offering insights into its generalizability across different thresholds.

IV. Results and Discussions

- Model Performance: The hybrid CNN-LSTM model outperformed standalone CNN and LSTM models with an F1-score of ~90% and ROC-AUC of 0.95, indicating robust classification of early mental distress from voice data.
- Feature Insights: MFCCs and spectrogram-based features were most informative, confirming prior research that frequency and rhythm carry emotional and cognitive cues.
- Practical Utility: The model could detect early indicators of depression and anxiety with minimal data, supporting its potential for real-time, low-cost applications in telehealth.
- Comparative Analysis: The results align with existing literature (e.g., Zhang et al., Diwakar & Raj) and outperform traditional SVM and decision tree models.
- Age and Vocal Biomarkers: A mild negative correlation was found between age and speech pitch variance, potentially due to age-related vocal changes, affecting model generalizability.
- Gender Differences: Subtle gender-based variations were observed in vocal tone and speech rate, which the model accounted for using stratified sampling and balanced datasets.
- Depression Severity Score (PHQ-9) Correlation: A strong positive correlation ($r = 0.72$) was observed between predicted scores and PHQ-9 ratings, validating the model’s clinical relevance.

V. Conclusion

Deep learning models analyzing voice data offer a promising direction for early detection of mental health disorders. By capturing vocal biomarkers of psychological distress, these systems can aid in timely interventions. AI-powered voice analysis tools offer a promising path to make mental health screening more accessible, affordable, and non-invasive, enabling broader reach and early detection across diverse populations.

References

1. https://www.sciencedirect.com/science/article/pii/S2352914823001284
2. https://pmc.ncbi.nlm.nih.gov/articles/PMC7293215/
3. https://ijisae.org/index.php/IJISAE/article/view/5561
4. https://clinical-practice-and-epidemiology-in-mental-health.com/VOLUME/20/ELOCATOR
5. /e17450179315688/FULLTEXT/
6. https://www.researchgate.net/publication/391707310_Early_detection_of_mental_health disorders_using_machine_learning_models_using_behavioral_and_voice_data_analysis

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Special Issue | Volume XIV, Issue XIII, October 2025

7. Merino, M. et al. Body perceptions and psychological well-being: A review of the impact of social media and physical measurementson self-esteem and mental health with a focus on body image satisfaction and its relationship with cultural and gender factorsHealthcare **12**(14), 1396 (2024).
8. Chen, X. & Pan, Z. A convenient and low-cost model of depression screening and early warning based on voice data using forpublic mental health. Int. J. Environ. Res. Public Health **18**(12), 6441 (2021).
9. Pourkeyvan, A., Safa, R. &Sorourkhah, A. Harnessing the power of hugging face transformers for predicting mental healthdisorders in social networks. IEEE Access **12**, 28025–28035 (2024).
10. Khan, S. & Alqahtani, S. Hybrid machine learning models to detect signs of depression. Multimed. Tools Appl. **83**(13), 38819–38837 (2024).
11. Ku, W. L. & Min, H. Evaluating Machine Learning Stability in Predicting Depression and Anxiety Amidst Subjective ResponseErrors. Healthcare **12**(6), 625 (2024).
12. RajuKanchapogu, N., & Mohanty, S. N. Enhancing Depression Predictive Models: A Comparative Study of Hybrid Ai, MachineLearning and Deep Learning Techniques. (2024).
13. Zhou, H., Zhou, F., Zhao, C., Xu, Y., Luo, L., & Chen, H. Multimodal data integration for precision oncology: Challenges and futuredirections. arXiv preprint arXiv:2406.19611. (2024)
14. Almutairi, S. et al. A Hybrid Deep Learning Model for Predicting Depression Symptoms from Large-Scale Textual Dataset. IEEEAccess https://doi.org/10.1109/ACCESS.2024.3496741 (2024).
15. Mahmood, T., Rehman, A., Saba, T., Nadeem, L. & Bahaj, S. A. O. Recent advancements and future prospects in active deeplearning for medical image segmentation and classification. IEEE Access **11**, 113623–113652 (2023).
16. Obaido, G. et al. Supervised machine learning in drug discovery and development: Algorithms, applications, challenges, andprospects. Mach. Learn. Appl. **17**, 100576 (2024).
17. Mohajeri, M., Towsyfyan, N., Tayim, N., Faroji, B. B. & Davoudi, M. Prediction of Suicidal Thoughts and Suicide Attempts in People Who Gamble Based on Biological-Psychological-Social Variables: A Machine Learning Study. Psychiatr. Q. https://doi.org /10.1007/s11126-024-10101-x (2024).
18. Di Cesare, M. G., Perpetuini, D., Cardone, D. & Merla, A. Assessment of Voice Disorders Using Machine Learning and VocalAnalysis of Voice Samples Recorded through Smartphones. BioMedInformatics**4**(1), 549–565 (2024).
19. Cheong, I., Caliskan, A. & Kohno, T. Safeguarding human values: rethinking US law for generative AI's societal impacts. AI Eth.https://doi.org/10.1007/s43681-024-00451-4 (2024).
20. Zafar, A. Balancing the scale: Navigating ethical and practical challenges of artificial intelligence (AI) integration in legal practices. Discov. Artif. Intell. **4**(1), 27 (2024).
21. Al-Tameemi, I. K. S., Feizi-Derakhshi, M. R., Pashazadeh, S. & Asadpour, M. Interpretable multimodal sentiment classificationusing deep multi-view attentive network of image and text data. IEEE Access **11**, 91060–91081 (2023).
22. Javed, H., Muqeet, H. A., Javed, T., Rehman, A. U. & Sadiq, R. Ethical Frameworks for Machine Learning in Sensitive HealthcareApplications. IEEE Access. **12**, 16233–16254 (2023).
23. Zhang, Z. Early warning model of adolescent mental health based on big data and machine learning. Soft. Comput. **28**(1), 811–828(2024).
24. Satapathy, S. K., Patel, V., Gandhi, M., & Mohapatra, R. K. Comparative Study of Brain Signals for Early Detection of Sleep Disorder
25. Using Machine and Deep Learning Algorithm. In 2024 IEEE International Conference on Interdisciplinary Approaches in Technologyand Management for Social Innovation (IATMSI) (Vol. 2, pp. 1–6). IEEE. (2024)
26. Hossain, S., Umer, S., Rout, R. K. & Al Marzouqi, H. A Deep Quantum Convolutional Neural Network Based Facial Expression Recognition for Mental Health Analysis. IEEE Trans. Neural Syst. Rehabil. Eng. https://doi.org/10.1109/TNSRE.2024.3385336(2024).