

**Title:** Mental Health Diagnosis From Voice Data Using Convolutional Neural Networks and Vision Transformers

**Author:** Rafiul Islam, Md Taimur Ahad, Faruk Ahmed, Bo Song, Yan Li

**Abstract:** Integrating Convolutional Neural Networks and Vision Transformers in voice analysis has unveiled a new horizon in mental health identification. Human voice, a powerful indicator of mental health, was the focus of this study. Human voice data representing stable and unstable conditions were gathered from various mental health institutions in Bangladesh. The results of the experiment suggest that the proposed model achieved 91% accuracy, precision of 92% for the "Unstable" category and 90% for the "Stable" category, and recall of 91% for the "Stable" category and 92% for the "Unstable" category. In addition, a high F1 score of 91% was achieved. This study significantly contributes to computer-aided diagnosis in mental health by using deep learning (DL) to diagnose mental well-being. Our research underscores the substantial impact of DL on the advancement of mental health care, instilling hope for a brighter future in mental health care.

**Keywords:** Mental health diagnostics, Convolutional Neural Network, Vision Transformer, Voice analysis, Integrated model, Machine learning, Early detection, Mental stability

**DOI:** <https://doi.org/10.1016/j.jvoice.2024.10.010>

<https://pubmed.ncbi.nlm.nih.gov/39550322>