International Journal of Engineering & Extended Technologies Research (IJEETR)

|ISSN: 2322-0163| www.ijeetr.com | A Bimonthly, Peer Reviewed, Scholarly Indexed Journal |

| Volume 8, Issue 2, March - April 2026 |

DOI:10.15662/IJEETR.2026.0802200

Machine Learning Analysis of Speech Detects Anxiety and Depression in Early Childhood

Dr. J. kirubakaran, S.Sharmitha, G.Vijiyarasan, M.Pravin

Professor, Students, Department of Electronics and Communication Engineering, Muthayammal Engineering College, Rasipuram, Tamil Nadu, India

Publication History: Received: 25.02.2026; Revised: 20.03.2026; Accepted: 25.03. 2026; Published: 28.03.2026.

ABSTRACT: Speech Emotion Recognition (SER) has emerged as a significant research area in machine learning and artificial intelligence, focusing on identifying human emotional states such as anxiety and depression from speech signals. This project presents a deep learning-based approach using hybrid 1D and 2D Convolutional Neural Networks (CNN) to improve the accuracy and robustness of emotion classification systems.

The model is trained and evaluated using benchmark datasets such as RAVDESS and TESS, incorporating feature extraction techniques including Mel-Frequency Cepstral Coefficients (MFCC), chroma features, and Mel-spectrograms. These features enable effective representation of both temporal and spectral characteristics of speech signals.

The integration of both temporal and spatial feature extraction through hybrid CNN architectures enhances the system's ability to capture subtle emotional variations in speech. The proposed model achieves an accuracy of approximately 86–89%, outperforming traditional methods and demonstrating improved generalization on unseen data.

Furthermore, the system is computationally efficient and suitable for real-time implementation. This work highlights the potential of speech-based emotion recognition as a non-invasive and cost-effective solution for early detection of mental health conditions, contributing to advancements in intelligent healthcare systems and human-computer interaction

KEYWORDS: Speech Emotion Recognition, CNN, MFCC, Deep Learning, Anxiety Detection, Depression Analysis.

I. INTRODUCTION

Speech Emotion Recognition (SER) is an important domain in affective computing that aims to detect human emotions from speech signals. Emotions play a crucial role in communication and decision-making, and understanding them can significantly improve human-computer interaction.

With the rapid growth of artificial intelligence and deep learning, SER systems are being widely used in applications such as virtual assistants, call center analytics, healthcare monitoring, and smart devices. Detecting emotions like anxiety and depression at an early stage can assist in mental health assessment and timely intervention.

However, challenges such as variability in speech patterns, background noise, and limited availability of labeled datasets make emotion detection a complex task. This project addresses these challenges by utilizing deep learning models capable of automatic feature extraction and robust classification.

II. LITERATURE SURVEY

Numerous research works have been carried out in the field of speech emotion recognition using both traditional and deep learning techniques. Early approaches relied on machine learning algorithms such as Support Vector Machines (SVM), Hidden Markov Models (HMM), and Random Forests for emotion classification.

Recent advancements have shown that deep learning models, especially Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN), provide superior performance due to their ability to learn complex patterns from

|ISSN: 2322-0163| www.ijeetr.com | A Bimonthly, Peer Reviewed, Scholarly Indexed Journal |

| Volume 8, Issue 2, March - April 2026 |

DOI:10.15662/IJEETR.2026.0802200

raw data. CNN models are particularly effective in extracting spatial features from spectrogram representations of audio signals.

Hybrid approaches combining multiple neural network architectures have also been explored to improve accuracy. Despite these advancements, issues such as overfitting, dataset imbalance, and computational complexity still exist, motivating the need for more efficient and generalized models.

III. EXISTING SYSTEM

Existing systems for speech emotion recognition primarily rely on traditional machine learning techniques or basic deep learning models. These systems often use limited datasets and require manual feature extraction, which reduces efficiency and scalability.

Most existing methods depend on converting speech to text and then performing sentiment analysis, which may lead to loss of emotional information present in audio signals. Additionally, these systems struggle to perform well in real-time environments due to noise and variability in speech input.

The major limitations of existing systems include lower accuracy, lack of robustness, and poor generalization across different datasets. These drawbacks highlight the need for more advanced models capable of handling complex speech patterns effectively.

IV. PROPOSED SYSTEM:

The proposed system introduces a hybrid deep learning approach using 1D and 2D CNN models for improved speech emotion recognition. The system processes raw audio signals and extracts meaningful features such as MFCC, chroma, and Mel-spectrogram representations.

The 1D-CNN model captures temporal features from audio signals, while the 2D-CNN model extracts spatial features from spectrogram images. By combining both models, the system achieves better performance in identifying emotional patterns.

The workflow of the system includes data preprocessing, feature extraction, model training, and classification. The model is trained using RAVDESS and TESS datasets, ensuring diversity and robustness.

Advantages of the proposed system:

- Higher accuracy compared to traditional models
- Automatic feature extraction
- Improved generalization
- Suitable for real-time applications
- Scalable and adaptable to various domains

V. SOFTWARE DESCRIPTION:

The implementation of the proposed system is carried out using various software tools and libraries.
Python: A versatile programming language widely used for machine learning and data analysis.
TensorFlow: An open-source framework for building and training deep learning models.
Keras: A high-level API that simplifies neural network implementation.
Librosa: A powerful library for audio processing and feature extraction.
NumPy: Used for numerical computations and handling arrays.
Matplotlib: Used for data visualization and plotting graphs.
These tools provide flexibility, scalability, and efficiency in developing the SER model.

VI. RESULT AND DISCUSSION

The proposed model is trained and tested using benchmark datasets, and its performance is evaluated using metrics such as accuracy, precision, recall, and F1-score.

|ISSN: 2322-0163| www.ijeetr.com | A Bimonthly, Peer Reviewed, Scholarly Indexed Journal |

| Volume 8, Issue 2, March - April 2026 |

DOI:10.15662/IJEETR.2026.0802200

The system achieves an accuracy of approximately 86–89%, which is higher than many traditional approaches. The use of hybrid CNN architecture improves the model's ability to capture both temporal and spectral features effectively.

Graphs such as training and validation accuracy, loss curves, and confusion matrix are used to analyze model performance. The results indicate that the model is well-trained and does not suffer from significant overfitting or underfitting.Overall, the experimental results validate the effectiveness of the proposed approach in speech emotion recognition.

VII. CONCLUSION

This project presents a deep learning-based approach for speech emotion recognition using hybrid CNN models. The system successfully identifies emotions such as anxiety and depression from speech signals with improved accuracy and efficiency.

The integration of advanced feature extraction techniques and deep learning models enhances the overall performance of the system. The results demonstrate that the proposed method is reliable and suitable for real-world applications.
Future work can focus on improving the model by incorporating larger datasets, advanced neural network architectures, and real-time deployment. The system can also be extended to applications in healthcare, education, and human-computer interaction.

REFERENCES

1. Mustaqueem khan and Soonil Kwon , "TC-Net: A Modest & Lightweight Emotion Recognition System Using Temporal Convolution Network", vol. 26 , 2023.
2. BubaiMaji ,Monorama Swain and Mustaqeem Khan, " Advanced Fusion- Based Emotion Recognition System Using a Dual-Attention Mechanism with Conv-Caps and Bi-GRU Features" , Electronics , vol. 11,pp. 1328, 2022.
3. Yan, Y.; Shen, X., "Research on Speech Emotion Recognition Based on AA-CBGRU Network", Electronics, vol.11, pp.1409,2022
4. Fauzivy Reggiswarashari, Sari Widya Sihw, "Speech emotion recognition using 2Dconvolutional neural network,"vol.12,pp.6594,2022
5. Rakhi Rani Paul, Subrata Kumer Paul, Md. Ekramul Hamid," A 2D Convolution Neural Network Based Method for Human Emotion Classification from Speech Signal," IEEE Acess,vol.10,pp.72-77,2022
6. Mohammad Reza Falahzadeh,Edris Zaman Farsa,AliHarimi,Arash Ahmadi , "3D
7. Convolutional Neural Network for Speech Emotion Recognition With Its Realization on Intel CPU and NVIDIA GPU," IEEE Access,vol.10,pp.112460- 112471,2022
8. E. Lieskovská, M. Jakubec, R. Jarina, and M. Chmulík, "A review on speech emotion recognition using deep learning and attention mechanism," Electronics, vol.10,pp.1163,2021
9. Mustaqueem khan and SoonilKwon , "1D-CNN:Speech Emotion Recognition System Using a Stacked Network with Dilated CNN Features,"vol.67,pp.4039- 4059,materials & continua 2021.
10. Anbazhagan, K. (2024). Trustworthy and Adaptive AI Systems for Enterprise Analytics Cybersecurity and Decision Optimization Using API-First and Cloud-Native Architectures. International Journal of Technology, Management and Humanities, 10(03), 65-74.
11. Padmapriya, V. M., Thenmozhi, K., Hemalatha, M., Thanikaiselvan, V., Lakshmi, C., Chidambaram, N., & Rengarajan, A. (2025). Secured IIoT against trust deficit-A flexi cryptic approach. Multimedia Tools and Applications, 84(9), 5625-5652.
12. Soundappan, S. J. (2020). Big Data Analytics in Healthcare: Applications for Pandemic Forecastin. International Journal of Advanced Research in Computer Science & Technology (IJARCST), 3(1), 2248-2253.
13. Mathew, A., & Alex, H. (2023, January). Hyper automation and augmented intelligence. In 2023 5th International Conference on Smart Systems and Inventive Technology (ICSSIT) (pp. 1230-1234). IEEE.
14. Murugeshwari, B., Sudharson, K., Panimalar, S. P., Shanmugapriya, M., & Abinaya, M. (2020). SAFE–Secure Authentication in Federated Environment using CEG Key code.
15. Anand, L., Tyagi, R., & Mehta, V. (2024, January). Food recognition using deep learning for recipe and restaurant recommendation. In Proceedings of Eighth International Conference on Information System Design and Intelligent Applications (pp. 269-279). Singapore: Springer Nature Singapore.