

Multifusion Multifeature Deepfake Audio detection

Yash S.Kalebag¹, Tejas D.Mhetre², Mubin A. Mulla³, Alsad

J.Mujawar⁴, Priyadarshini P.Yadav⁵

^{1,2,3,4} AIML Student, Adarsh Institute of Technology & Research Centre ,Vita.

⁵AIML Faculty, Adarsh Institute of Technology & Research Centre ,Vita

Abstract - The fast development of artificial intelligence systems and speech synthesis technology has made it easier to create deepfake audio, which produces realistic synthetic audio. The situation creates major risks, which endanger both digital security systems, authentication platforms, and the trustworthiness of information sources. The field of Deepfake Audio Detection has therefore gained significant importance in recent years. The research presents a Multifusion Multifeature Deepfake Audio Detection system, which performs exact detection of genuine speech against synthetic speech content.

The proposed approach uses various audio features, which include speech signal spectral information, temporal data, and frequency-based elements. The system combines these features through a multiple fusion system, which produces better detection results. It also produces more robust performance. The system uses machine learning and deep learning methods to study the extracted features, which determine if audio samples belong to genuine or fake categories. The system detects minimal artificial audio defects, which human hearing fails to detect.

Key Words: Deepfake Audio Detection, Artificial Intelligence, Machine Learning, Deep Learning, Speech Processing, Audio Signal Processing

1. INTRODUCTION

Lately, AI and deep learning stuff has gotten really good at making fake voices that sound just like the real thing, its called deepfake audio. I mean, there are some cool ways to use this tech, but it also opens up all these problems like people scamming with voices or spreading fake news, even stealing identities. That makes figuring out if audio is real or messed with a big deal in research right now.

This project I am working on is about a system called Multifusion Multifeature for detecting deepfake audio. It looks at various parts of speech signals, you know, different

characteristics from the audio and mixes them together to decide if its genuine or not.

1.2 project overview

The project is called Multifusion Multifeature Deepfake Audio Detection. It basically tries to figure out if an audio clip is genuine or made by some AI. These days, with all the progress in artificial intelligence and deep learning, people can whip up fake voices that sound super real. That is kind of scary, I think, because it opens the door to stuff like scams or spreading false info everywhere.

- 1.To develop a system for detecting fake and real audio in the field of Deepfake Audio Detection.
- 2.To extract multiple audio features using techniques from Audio Signal Processing.
- 3.To apply models based on Machine Learning and Artificial Intelligence for accurate audio classification.
- 4.To improve the reliability and accuracy of detecting AI generated speech using a multifusion approach.

1.4 proposed solution

The proposed system detects fake and real audio using techniques from Artificial Intelligence and Machine Learning in the field of Deepfake Audio Detection. First, the input audio is preprocessed to remove noise. Then multiple audio features are extracted using methods from Audio Signal Processing. These features are combined using a multifusion approach and analyzed by a classification model to determine whether the audio is real or fake.

2. METHODOLOGY

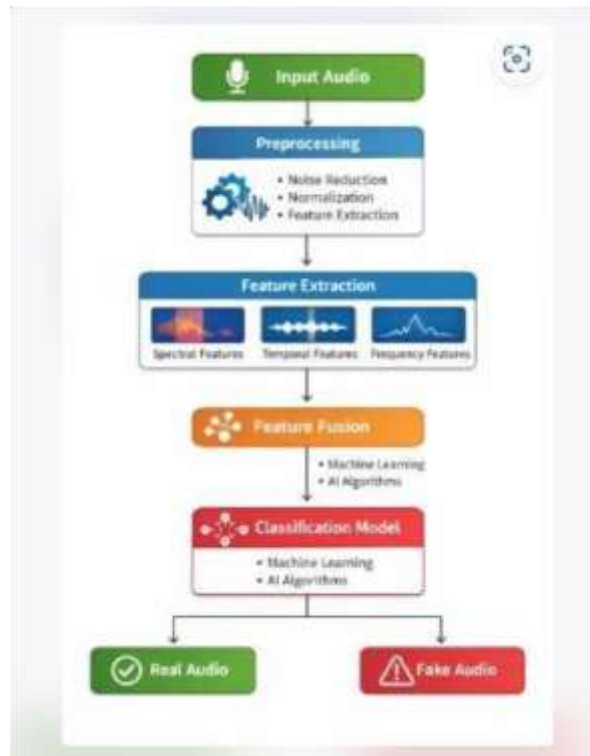


Fig. Project flow

2.1 Input Audio

This is the first stage where the system receives the speech sample. The audio can be recorded from a microphone or uploaded as a file. The system then processes this input to determine whether it is real or fake in Deepfake Audio Detection.

2.2 Preprocessing

In this step, the audio signal is cleaned and prepared for analysis using techniques from Audio Signal Processing. Main tasks include:

- **Noise Reduction** – removes background noise.
- **Normalization** – adjusts audio levels for consistency.
- **Signal preparation** for feature extraction.

2.3 feature extraction

Important characteristics of the audio signal are extracted. These features help the system understand patterns in speech. Types of features:

- **Spectral Features** – represent frequency distribution of sound.
- **Temporal Features** – analyze time-based patterns of speech.
- **Frequency Features** – examine pitch and frequency variations.

2.4. Feature fusion

In this stage, all extracted features are combined into a single dataset. This **multifusion approach** improves detection accuracy by using multiple audio characteristics instead of relying on just one.

2.5 classification model

The fused features are analyzed using models from Machine Learning and Artificial Intelligence. The model learns patterns from training data and predicts whether the audio is genuine or manipulated.

2.6 Output Result

Finally, the system provides the result:

- **Real Audio** – authentic human speech.
- **Fake Audio** – AI-generated or manipulated speech.

3. CONCLUSION

The Multifusion Multifeature Deepfake Audio Detection project aims to identify whether an audio sample is real or fake. Using techniques from Artificial Intelligence and Machine Learning, the system extracts multiple audio features and analyzes them to detect manipulated speech in Deepfake Audio Detection.

The results show that combining multiple features improves the accuracy and reliability of detecting fake audio, helping enhance security in voice-based systems

ACKNOWLEDGEMENT

We would like to express our sincere gratitude to our project guide for their valuable guidance and support throughout the development of this project. Their suggestions and encouragement helped us successfully complete our work.

We also thank the Head of the Department, faculty members, and our institution for providing the necessary resources and support. Finally, we are grateful to everyone who directly or indirectly contributed to the successful completion of this project on Deepfake Audio Detection.

REFERENCE

1. N. Jain, "Detecting Deepfakes: Exploring Machine Learning Models for Audio, Video, and Image Analysis," *International Journal of Intelligent Systems and Applications in Engineering*, vol. 12, no. 4, 2024.
2. S. Dilbar, M. A. Qureshi, S. K. Noon, and A. Mannan, "AudioFakeNet: A Model for Reliable Speaker Verification in Deepfake Audio," *Algorithms*, vol. 18, no. 11, 2025.
3. L. Yu Gong and X. J. Li, "Deepfake Voice Detection Using End-to-End Transformer with Acoustic Feature Fusion," *Electronics*, vol. 14, no. 10, 2025.
4. M. P. Laureta, J. M. Atienza, and J. L. Tapel, "Deepfake Speech Detection Using Hybrid CNN-LSTM Model," *Journal of Engineering, Computing and Technology*, 2024.
5. S. Borade et al., "Improving Deepfake Audio Detection Using SVM with Mel-Frequency Cepstral Coefficients," *International Journal of Intelligent Systems and Applications in Engineering*, 2024.