International Journal of Creative and Open Research in Engineering and Management
ISSN: **3108-1754 (Online)**
Volume 02 Issue 04 April-2026 | Impact Factor: 3.5

Deep Fake Audio Detection Using Deep Learning

Mrs. E. Sushma
Assistant Professor, Dept of CSE(DS) , CMR Technical Campus Hyderabad, Telangana, India
sushma.ds@cmrtc.ac.in

Ms. N. Soujanya
Assistant Professor, Dept of CSE(DS), CMR Technical Campus Hyderabad, Telangana, India
noundlasoujanya516@gmail.com

P. Abhinav
UG Student, Dept of CSE(DS), CMR Technical Campus Hyderabad, Telangana, India,
pabhinav1304@gmail.com

A. Akhil
UG Student, Dept of CSE(DS), CMR Technical Campus Hyderabad, Telangana, India
akhilavula37@gmail.com

N. Rajesh
UG Student, Dept of CSE(DS), CMR Technical Campus Hyderabad, Telangana, India
rajeshrajesh58171@gmail.com

Pankaj Rathod
UG Student, Dept of CSE(DS), CMR Technical Campus Hyderabad, Telangana, India
pankaj707572@gmail.com

How to Cite this Article:
Soujanya, N., Abhinav, P., Akhil, A., Rajesh, N. & Rathod, P. (2026). Deep Fake Audio Detection Using Deep Learning. International Journal of Creative and Open Research in Engineering and Management, 02(04). https://doi.org/10.55041/ijcope.v2i4.325

https://doi.org/10.55041/ijcope.v2i4.325

ABSTRACT— This project is titled as “Deep Fake Audio Detection Using Deep Learning”. The rapid advancement of artificial intelligence and deep learning technologies, deepfake audio has emerged as a significant threat in today’s digital world. It enables the generation of highly realistic synthetic voices that can closely imitate real individuals. Such audio can be misused for malicious purposes such as fraud, impersonation, spreading misinformation, and unauthorized access to voice-based systems. Therefore, detecting deepfake audio has become essential to ensure security, authenticity, and trust in digital communication.

This project proposes a deep learning-based approach for detecting deepfake audio using a hybrid model that combines Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) networks. The system begins with preprocessing the audio data, followed by feature extraction using Mel-Frequency Cepstral Coefficients (MFCC), which effectively capture the important characteristics of speech signals. The CNN model is used to extract spatial features from the audio representation, while the LSTM model analyzes temporal patterns and sequential dependencies in speech.The proposed model is trained and tested on a dataset consisting of both real and fake audio samples. The system is evaluated using performance metrics such as accuracy, precision, recall, and F1-score to ensure its effectiveness and reliability. The system is designed to provide an efficient, scalable, and robust solution for deepfake audio detection.

By leveraging advanced deep learning techniques, the proposed approach enhances the reliability of voice-based systems and contributes to preventing voice-based cyber threats. This project highlights the importance of integrating artificial intelligence with security applications to address emerging challenges in digital media authentication**.**

INTRODUCTION

In recent years, the rapid growth of artificial intelligence and deep learning technologies has led to the development of advanced multimedia generation techniques. One such development is deepfake audio, which allows the creation of highly realistic synthetic voices that can closely mimic real individuals. While this technology has useful applications in areas such as entertainment and virtual assistants, it also poses serious threats when misused for malicious purposes such as impersonation, fraud, and spreading misinformation.

Deepfake audio is generated using sophisticated machine learning models that learn the voice patterns, tone, and speech characteristics of a person. As a result, it becomes extremely difficult for humans to distinguish between real and fake audio. This creates a major challenge in ensuring the authenticity and reliability of voice-based communication systems.To address this issue, deep learning-based detection methods have been developed to identify manipulated audio. These methods analyze audio signals, extract relevant features, and classify them as real or fake. In this project, a hybrid model combining Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) networks is used to effectively detect deepfake audio by analyzing both spatial and temporal features of speech signals. The proposed system aims to provide a reliable and efficient solution for detecting deepfake audio, thereby improving security in digital communication and preventing misuse of synthetic voice technologies.

I. PROJECT PURPOSE

The main purpose of this project is to develop an efficient system for detecting deepfake audio using deep learning techniques. With the increasing use of artificial intelligence in generating synthetic voices, there is a growing need to ensure the authenticity of audio data and prevent its misuse in various applications.This project aims to design and implement a hybrid deep learning model that can accurately distinguish between real and fake audio by analyzing speech features and temporal patterns. The system focuses on improving detection accuracy while maintaining efficiency and scalability for real-time applications.Another important purpose of this project is to enhance security in voice-based systems such as online communication, banking services, and authentication systems. By detecting manipulated audio, the system helps in preventing fraud, impersonation, and other cyber threats.

Overall, the project contributes to the development of a secure and trustworthy digital environment by providing a reliable solution for deepfake audio detection.

1.2 PROJECT FEATURES

The main purpose of this project is to develop an effective system for detecting deepfake audio using deep learning techniques. With the rapid advancement of artificial intelligence, synthetic audio can be generated to imitate real human voices, creating serious risks such as fraud, impersonation, and misinformation.

This project aims to design and implement a hybrid model using Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) networks to accurately classify audio as real or fake. The system focuses on extracting meaningful features from speech signals and analyzing temporal patterns to improve detection accuracy.

Another purpose of this project is to enhance the security and reliability of voice-based systems such as communication platforms and authentication systems. By identifying manipulated audio, the system helps in preventing cyber threats and misuse of voice technology.

Overall, the project aims to provide a reliable, efficient, and scalable solution for deepfake audio detection, contributing to safer and more trustworthy digital communication.

Related Work

Several researchers have explored deepfake audio detection using machine learning and deep learning techniques. Studies show that feature extraction methods like MFCC and spectrogram analysis are effective in representing audio signals. Convolutional Neural Networks (CNN) have been widely used for extracting spatial features, while Long Short-Term Memory (LSTM) networks are used for analyzing temporal patterns in speech. Hybrid models combining CNN and LSTM have shown improved accuracy in detecting fake audioResearch also highlights the use of ensemble models and advanced architectures to enhance performance. Although these methods achieve high accuracy, they often require large datasets and high computational power. Additionally, many existing systems struggle with detecting unseen or highly sophisticated deepfake audio. Therefore, there is a need for efficient and robust models, which is addressed in this project using a hybrid deep learning approach.

II. METHODOLOGY

The proposed system uses a deep learning-based approach for detecting deepfake audio. It combines Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) networks to improve detection accuracy. The system extracts MFCC features from audio signals and processes them through the CNN for feature extraction and LSTM for temporal analysis. A web-based interface is developed where users can upload audio files and receive real-time predictions indicating whether the audio is real or fake. The proposed system provides better performance compared to traditional methods by automatically learning complex patterns in audio data.

1. Data Collection

The first step involves collecting a dataset consisting of both real and deepfake audio samples. The data is gathered from reliable sources and properly labeled to ensure correct classification during training. A diverse dataset helps improve the model's performance and generalization ability.

2. Data Preprocessing

The collected audio dataset is preprocessed to improve audio quality and consistency. The steps include:

• Audio cleaning to remove noise and distortions
• Audio normalization to maintain uniform amplitude
• Conversion of audio into a standard format (e.g., WAV)
• Feature extraction using MFCC (Mel-Frequency Cepstral Coefficients)
• Conversion of audio signals into numerical feature vectors

After preprocessing, the dataset is split into:

• **Training data (80%)**
• **Testing data (20%)**

3. Model Training

The preprocessed audio dataset is used to train the deep learning model. The steps include:

• Input MFCC feature vectors into the model
• Use Convolutional Neural Network (CNN) for feature extraction
• Apply Long Short-Term Memory (LSTM) for temporal analysis
• Train the model using labeled data (real and fake audio)
• Optimize model parameters using Adam optimizer
• Use loss function (binary cross-entropy) for training

After training, the model is ready for:

• Testing on unseen data
• Prediction of real or fake audio

4. Model Evaluation

The performance of model is evaluated using the following metrics:

- Accuracy
- Precision
- Recall
- F1-score

A confusion matrix is also used to analyze correct and incorrect classifications of real and fake audio.

5. Result Comparison

The performance of the proposed model is compared with existing methods using the following metrics:

• Accuracy (Proposed Model: 94%)
• Precision
• Recall
• F1-score

The proposed CNN–LSTM model shows improved accuracy compared to traditional machine learning models such as SVM and basic neural networks.

6.Prediction

Once the model is trained and evaluated, it is used to predict new audio inputs. The system processes the uploaded audio file, extracts features, and feeds them into the trained model to determine whether the audio is real or fake.

7. Output Generation

Finally, the system generates the output based on the prediction. The result is displayed to the user through the web interface as "Real" or "Fake." This step ensures that the user can easily understand the classification result.

III. PROPOSED SYSTEM

III. Proposed System

The proposed system focuses on detecting deepfake audio using a hybrid deep learning approach. The input audio is preprocessed and converted into MFCC features, which capture important speech characteristics.

A combination of Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) networks is used for classification. CNN extracts relevant features, while LSTM analyzes temporal patterns in the audio.

The model is trained using labeled data and is capable of classifying audio as real or fake. A web interface is provided for users to upload audio and receive predictions.

This system improves accuracy and provides an efficient solution for deepfake audio detection.

IV. IMPLEMENTATION DETAILS

The implementation of the Deepfake Audio Detection system involves multiple stages, including data collection, preprocessing, feature extraction, model training, and prediction. The system is developed using Python and various deep learning libraries to ensure efficient processing and accurate detection.

Initially, the audio dataset containing both real and fake samples is collected and preprocessed. Preprocessing includes noise removal, normalization, and conversion of audio signals into a suitable format. After preprocessing, feature extraction is performed using Mel-Frequency Cepstral Coefficients (MFCC), which capture important characteristics of speech signals.

The extracted features are then provided to a hybrid deep learning model consisting of Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) networks. The CNN extracts spatial features from the input data, while the LSTM analyzes temporal patterns in the speech signals. The model is trained using labeled data and evaluated using performance metrics such as accuracy, precision, recall, and F1-score.

After training, the system is integrated with a web interface where users can upload audio files. The trained model processes the input and provides a prediction indicating whether the audio is real or fake. This implementation ensures an efficient, scalable, and user-friendly solution for deepfake audio detection.

4.1 ALGORITHMS USED

4.1.1 CONVOLUTIONAL NEURAL NETWORK(CNN)

Random Forest is a supervised machine learning algorithm used for classification tasks. It works by creating multiple decision trees during training and combining their outputs to improve prediction accuracy. Each tree is trained on a random subset of data and features, which helps reduce overfitting and enhances model performance. In this project, Random Forest is used to classify bone types from X-ray images based on extracted pixel features. It provides high accuracy, handles large datasets efficiently, and delivers stable results, making it the primary algorithm used in the system.

4.1.2 LONG SHORT-TERM MEMORY(LSTM)

LSTM is a type of Recurrent Neural Network (RNN) used for analyzing sequential data. It is capable of learning long-term dependencies in audio signals. In this project, LSTM analyzes temporal patterns such as speech flow, pitch variations, and timing, which are essential for distinguishing between real and fake audio.

4.1.3 MEL-FREQUENCY CEPSTRAL COEFFICIENTS (MFCC)

MFCC is a feature extraction technique used to convert audio signals into numerical representations. It captures important characteristics of speech such as frequency, tone, and pitch in a way similar to human hearing. These features are used as input to the deep learning model.

.4.1.4 SYSTEM MODULES

The system is divided into the following modules:

- Audio Upload Module
- Data Preprocessing Module
- Feature Extraction Module
- Train-Test Split Module
- Model Training Module
- Model Evaluation Module
- Prediction Module
- Output Display Module

4.1.5 BINARY CLASSIFICATION ALGORITHM

The final layer of the model performs binary classification to determine whether the input audio is real or fake. It uses activation functions such as sigmoid to produce output probabilities and classify the audio accordingly.

V. EXPERIMENTAL RESULTS AND DISCUSSION

The proposed Deepfake Audio Detection system achieved high accuracy in classifying audio as real or fake using the CNN–LSTM model. The use of MFCC features helped in capturing important speech characteristics effectively. The model showed good performance in terms of precision, recall, and F1-score. The results indicate that the system can reliably detect manipulated audio. However, performance may vary depending on dataset quality and noise conditions.

System Interface – Home Page:

Home Page Interface of Deep Fake Audio Detection System

Fig. 1.User Login Interface

User Login Page of the System

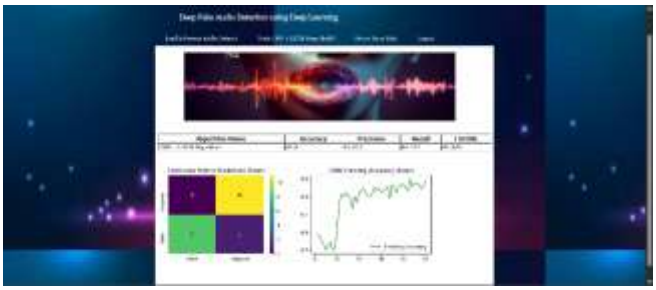
Fig. 2. Successful Login and Dashboard

Admin Dashboard after Succesful Login

Fig. 3. Dataset Loading and Feature Extraction

Audio Dataset Loading and MFCC Feature Extraction Output

Fig. 4. Model Performance and Analysis

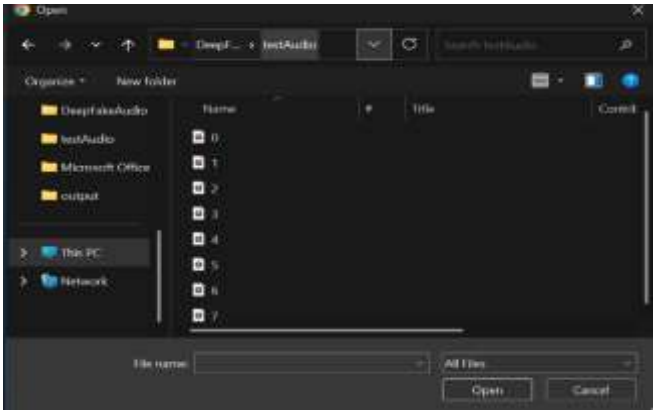


Performance Metrics, Confusion Matrix and Training

Fig. 5. Deep Fake Audio Detection Interface

User Interface for Deep Fake Audio Interface

Fig. 6. Audio File Upload Process



Audio File Selection for Deep Fake Detection

Fig. 6. Deep Fake Audio Prediction Result

Output Showing Predicted Result as Uploaded File

VI. CONCLUSION

In conclusion, The Deepfake Audio Detection system was successfully developed using a hybrid CNN–LSTM model to accurately classify audio as real or fake. The system demonstrated good performance by effectively analyzing both spatial and temporal features of speech signals. It provides a reliable solution for detecting manipulated audio and enhancing security in voice-based systems. In the future, the system can be improved by using larger datasets and more advanced models to handle highly sophisticated deepfake techniques. Additionally, it can be extended for real-time applications and integrated with security systems for better protection against audio-based threats.

VII. FUTURE SCOPE

The Deepfake Audio Detection system can be further enhanced in several ways to improve its performance and applicability. One of the major improvements can be the use of larger and more diverse datasets, which will help the model generalize better and detect a wider variety of deepfake audio samples. The system can also be improved by incorporating advanced deep learning architectures such as Transformer-based models for better feature learning.

Another important enhancement is the implementation of real-time detection systems that can analyze live audio streams. This would make the system more useful in applications such as call monitoring and voice authentication. The model can also be optimized to reduce computational complexity and improve processing speed.

VIII. ACKNOWLEDGMENT

We take this opportunity to express our gratitude to the people who have been instrumental in the successful completion of this project, we take this opportunity to express our profound gratitude and deep regard to our guide **E. Sushma**, Designation for his/her exemplary guidance, monitoring and constant encouragement throughout the project work. The blessing, help and guidance given by him/her shall carry us a long way in the journey of life on which we are about to embark.

We also take this opportunity to express a deep sense of gratitude to the Project Review Committee (PRC) coordinators **N. Soujanya, Shafana Bakshi, Bhookya Ramesh, S. Raghavendra, V. Rajesh, J. Shiva,**

B. Sangamitra, P. Ashwini for their cordial support, valuable information and guidance, which helped us in completing this task through various stages.

We are also thankful to **Dr. K. Murali**, Head, Department of Computer Science and Engineering (Data Science) for providing

encouragement and support for completing this project successfully.

We are deeply grateful to **Dr. A. Raji Reddy**, Director, for his cooperation throughout the course of this project. Additionally, we extend our profound gratitude to **Sri. Ch. Gopal Reddy**, Chairman, **Smt. C. Vasantha Latha**, Secretary and **Sri. C. Abhinav Reddy**, Vice-Chairman, for fostering an excellent infrastructure and a conducive learning environment that greatly contributed to our progress.

The guidance and support received from all the members of CMR Technical Campus who contributed to the completion of the project. We are grateful for their constant support and help.

Finally, we would like to take this opportunity to thank our family for their constant encouragement, without which this assignment would not be completed. We sincerely acknowledge and thank all those who gave support directly and indirectly in the completion of this project.

IX. REFERENCES

1. Yu Gong and X. Li (2025), *Deepfake Voice Detection using Transformer Models*. https://www.mdpi.com/2079-9292
2. Shreyas R., et al. (2024), *Deep Fake Audio Detection using Deep Learning Techniques*. https://www.researchgate.net
3. Z. Almutairi and H. Elgibreen (2022), *A Review of Modern Audio Deepfake Detection Methods*. https://www.mdpi.com
4. Khanjani Z., Watson J., and Sotirakopoulos A. (2023), *Audio Deepfakes: A Survey*. https://pubmed.ncbi.nlm.nih.gov
5. Chandrapalan K. (2025), *Deepfake Speech Detection using Artificial Intelligence*. https://rsisinternational.org
6. Alfalasi R. (2022), *Deepfake Audio Detection using Machine Learning Techniques*. https://repository.rit.edu
7. J. Frank and L. Schönherr (2021), *WaveFake: A Dataset for Audio Deepfake Detection*. https://arxiv.org/abs/2111.02813
8. H. Tak et al. (2021), *End-to-End Anti-Spoofing with RawNet Models*. https://arxiv.org/abs/2107.12710
9. Y. Zhang et al. (2022), *Deepfake Audio Detection using Speaker Verification Techniques*. https://arxiv.org
10. M. Todisco, H. Delgado, and N. Evans (2019), *ASVspoof Challenge: Automatic Speaker Verification Spoofing and Countermeasures*. https://arxiv.org/abs/1904.05441
11. A. Oord et al. (2016), *WaveNet: A Generative Model for Raw Audio*. https://arxiv.org/abs/1609.03499
12. D. Griffin and J. Lim (1984), *Signal Estimation from Modified Short-Time Fourier Transform*. https://ieeexplore.ieee.org