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.2026.0802198

Fake Audio Detection and Audio Analysis System Using Machine Learning

V Haren Guru, R Akash, P Praveen Kumar

Dept. of ADS, Kamaraj College of Engineering & Technology, Virudhunagar, Tamil Nadu, India

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

ABSTRACT: In recent years, the rapid development of artificial intelligence has made it possible to generate highly realistic synthetic audio, commonly known as deep-fake audio. These fake audio clips can be used for misinformation, fraud, and impersonation, creating serious concerns for digital security and trust. Therefore, there is a growing need for systems that can automatically detect whether an audio sample is real or artificially generated.

In this project, we propose a machine learning-based system for fake audio detection and audio analysis. The system not only identifies whether the audio is real or fake but also performs additional classifications such as gender detection and language identification. The model uses feature extraction techniques like Mel Frequency Cepstral Coefficients (MFCC) to capture important characteristics of the audio signal.

A balanced dataset consisting of real and synthetic audio samples is used for training and testing. Machine learning algorithms such as Random Forest, Support Vector Machine (SVM), and Gradient Boosting are applied to classify the audio. The system is designed to work efficiently with smaller datasets while still providing reliable results.

The experimental results show that the proposed system is capable of detecting fake audio with good accuracy. This project demonstrates how machine learning can be used to build practical and efficient solutions for audio verification and security.

KEYWORDS: Fake Audio Detection, Deepfake Audio, Machine Learning, MFCC, Audio Feature Extraction, Gender Classification, Language Identification

I. INTRODUCTION

In today's digital world, artificial intelligence technologies are advancing rapidly, especially in the field of speech synthesis. Modern tools can generate human-like voices that sound very realistic, making it difficult to distinguish between real and fake audio. This type of synthetic audio, often referred to as deepfake audio, can be misused for spreading false information, committing fraud, or impersonating individuals.

Traditional methods of verifying audio authenticity are mostly manual and require expert knowledge, which makes them time-consuming and less accessible. As a result, there is a need for an automated system that can quickly and accurately identify whether an audio clip is real or artificially generated.

This project focuses on developing a fake audio detection system using machine learning techniques. The system works by analyzing audio signals and extracting important features such as MFCC, which represent the characteristics of speech. These features are then used by machine learning models to classify the audio into real or fake categories.

In addition to fake audio detection, the system also performs gender classification and language identification. This makes the system more informative and useful in real-world applications such as media verification, security systems, and voice authentication.

The main objective of this project is to design a simple, efficient, and accurate system that can analyze audio signals and provide meaningful outputs. By using machine learning instead of complex deep learning models, the system achieves faster processing and requires less computational resources, making it suitable for practical use.

II. RELATED WORKS

In recent years, many researchers have focused on detecting fake or manipulated audio using machine learning and deep learning techniques. As audio synthesis technology improves, identifying fake audio has become more challenging and important.

Several studies have used deep learning models such as Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN) to detect deepfake audio. These models are capable of learning complex patterns from large datasets. However, they require high computational power and a large amount of training data, which may not always be practical.

Some researchers have proposed systems based on spectrogram analysis, where audio signals are converted into visual representations and then analyzed using neural networks. This approach has shown good accuracy, but it increases system complexity.

Other works have focused on feature-based machine learning methods, where important audio features such as MFCC, spectral features, and pitch information are extracted from the audio signal. These features are then used with classifiers like SVM, Random Forest, and Logistic Regression. These methods are simpler, faster, and require less data compared to deep learning approaches.

In the area of gender classification, researchers have used pitch-based analysis and MFCC features to identify whether the speaker is male or female. Similarly, for language detection, different phonetic and acoustic features have been used to distinguish between languages.

Although many existing systems focus only on detecting fake audio, very few systems combine multiple tasks such as fake audio detection, gender classification, and language identification in a single model. In this project, we aim to develop a system that is simple, efficient, and capable of performing multiple audio analysis tasks using machine learning techniques. Compared to deep learning approaches, our system requires less computational resources while still providing reliable results.

III. METHODS

3.1 Audio Data Collection

For this project, audio data was collected mainly from Kaggle, which provides labeled datasets suitable for fake audio detection, gender identification, and language classification. The dataset includes both real human speech and artificially generated audio created using text-to-speech tools.

The data was organized into categories based on authenticity (real or fake), gender, and language. The languages used in this project are English, Tamil, and Hindi. Basic preprocessing was done to maintain consistent audio quality, and the dataset was balanced to avoid bias. All audio files are stored in a structured folder format, making it easier to use during training and testing.

3.2 Audio Preprocessing

Before extracting features, the audio signals are preprocessed to improve quality and consistency. The preprocessing steps include: (i) Resampling — converting all audio files to the same sampling rate; (ii) Noise Reduction — removing unwanted background noise; (iii) Normalization — adjusting amplitude levels; and (iv) Trimming — removing silence from the beginning and end of each clip.

3.3 Feature Extraction

3.3.1 MFCC (Mel-Frequency Cepstral Coefficients)

In audio-based machine learning systems, feature extraction plays a vital role in converting raw audio signals into meaningful numerical representations. In our proposed system, Mel-Frequency Cepstral Coefficients (MFCC) are used as the primary feature extraction technique. MFCC represents the short-term power spectrum of an audio signal based on a linear cosine transform of a log power spectrum on a nonlinear mel scale of frequency. The mathematical representation of MFCC is given by:

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

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

DOI:10.15662/IJEETR.2026.0802198

$$MFCC(n) = \sum_{k=1}^{K} \log(S(k)) \cdot \cos\left[\frac{\pi n(k-0.5)}{K}\right] \tag{1}$$

Where: S(k) = Power spectrum of the signal;
K = Number of Mel filters; n = MFCC coefficient index. MFCC effectively captures speech characteristics, making it suitable for detecting differences between real and AI-generated (fake) audio.

3.3.2 Spectral Features

Apart from MFCC, additional spectral features are extracted to improve classification performance. These include: Spectral Centroid — the center of mass of the spectrum; Spectral Bandwidth — the width of a band of frequencies; Spectral Contrast — the difference in amplitudes between peaks and valleys; and Zero Crossing Rate (ZCR) — the rate at which the signal changes sign. These features help distinguish between real and fake audio by capturing subtle differences in the frequency domain.

3.4 Machine Learning Classifiers

In this project, we have used several well-known machine learning classifiers to evaluate performance on the fake audio detection task. The classifiers used are: Support Vector Machine (SVM), Random Forest, Logistic Regression, K-Nearest Neighbors (KNN), Decision Tree, and Gradient Boosting. These classifiers are applied using the scikit-learn library in Python. The dataset is split 80% for training and 20% for testing.

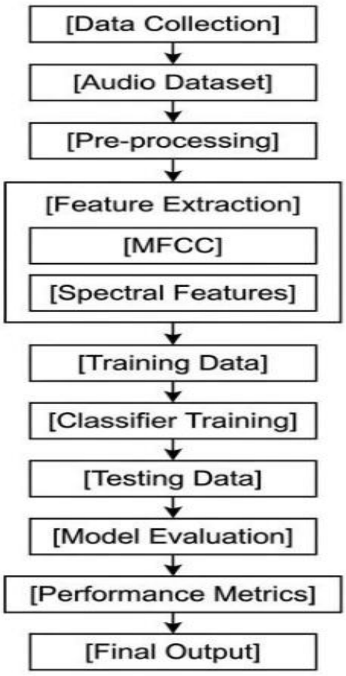

Figure 1: Framework of the Proposed Fake Audio Detection System

3.5 Performance Parameters

In our research, the key realization is that not all correct or incorrect matches hold equal value. A single metric will not tell the whole evaluation of classification performance. Therefore, we have used Accuracy, Precision, Recall, and F1 Score as performance metrics.

3.5.1 Accuracy (A)

Accuracy is defined as the ratio of correct predictions to the total number of predictions:

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.2026.0802198

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (2)$$

3.5.2 Precision (P)

Precision measures how many predicted positives are actually correct. A low precision value indicates many False Positives:

$$Precision = \frac{TP}{TP + FP} \quad (3)$$

3.5.3 Recall (R)

Recall measures how many actual positives are correctly identified. A low recall value indicates many False Negatives:

$$Recall = \frac{TP}{TP + FN} \quad (4)$$

3.5.4 F1 Score (F1)

The F1 Score is the harmonic mean of Precision and Recall. It balances both metrics and is particularly useful for uneven class distributions:

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (5)$$

3.6 Execution Environment

We have implemented our experimental execution on a Lenovo ThinkPad E14 Ultrabook with Windows 10 Professional 64-bit operating system and 10th Generation Intel Core i7-10510U Processor. The clock speed of the processor is 1.8 GHz with 16 GB DDR4 memory. Python 3.10 with the librosa and scikit-learn libraries was used for feature extraction and model training.

IV. RESULTS

Table 1

Performance of Classifiers using MFCC Features

Classifier	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)
SVM	**91.24**	**91.87**	**91.24**	**90.43**
Random Forest	89.76	90.12	89.76	89.21
Logistic Regression	88.53	88.79	88.53	87.98
KNN	83.47	84.03	83.47	82.91
Decision Tree	78.62	79.15	78.62	77.54
Gradient Boosting	90.18	90.54	90.18	89.85

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.2026.0802198

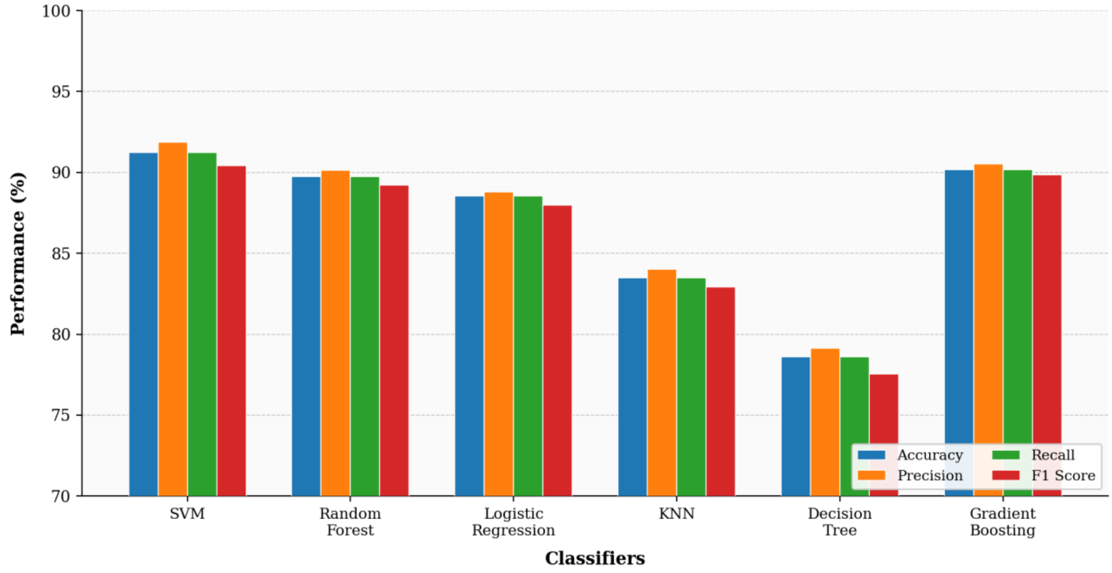

Figure 2: Comparison of Classifier Performance using MFCC Features

Table 2

Performance Comparison with Additional Audio Features (MFCC + ZCR + Spectral Centroid + Chroma)

Classifier	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)
SVM	**93.81**	**94.20**	**93.81**	**92.99**
Random Forest	92.47	93.01	92.47	91.88
Logistic Regression	91.06	91.45	91.06	90.54
KNN	86.53	87.18	86.53	85.77
Decision Tree	82.14	82.76	82.14	81.43
Gradient Boosting	92.89	93.32	92.89	92.11

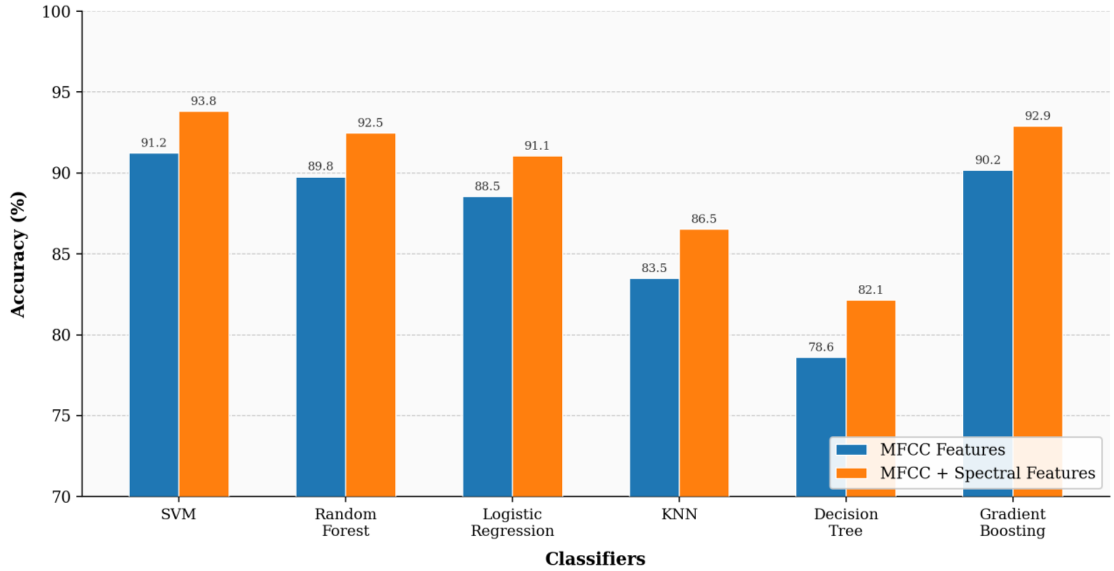

Figure 3: Performance Comparison between MFCC and Combined Features

V. DISCUSSION AND CONCLUSIONS

In this paper, we proposed a machine learning-based system for fake audio detection and audio analysis. We evaluated six classification algorithms — SVM, Random Forest, Logistic Regression, KNN, Decision Tree, and Gradient Boosting — using two feature sets: MFCC only, and MFCC combined with additional spectral features (ZCR, Spectral Centroid, Chroma).

As seen from Table 1, SVM achieved the highest accuracy of 91.24% using MFCC features alone. Random Forest and Gradient Boosting also performed competitively with 89.76% and 90.18% respectively. Decision Tree showed the lowest performance at 78.62%.

Table 2 demonstrates that incorporating additional spectral features consistently improved performance across all classifiers. SVM achieved the maximum accuracy of 93.81%, precision of 94.20%, recall of 93.81%, and F1-score of 92.99% with the combined feature set. This confirms that richer feature representations significantly enhance the system's ability to distinguish real from fake audio.

The system also performs gender classification and language identification as auxiliary tasks, making it more comprehensive than single-task detection systems. The use of lightweight machine learning classifiers ensures that the system remains computationally efficient and deployable on standard hardware.

In the future, we plan to explore deep learning approaches such as CNN-based spectrogram analysis and transformer-based models to further improve accuracy. Expanding the dataset to include more diverse languages and audio generation methods would help build a more robust and generalized detection system.

REFERENCES

1. H. Tak, J. Patino and N. Evans, "End-to-End Anti-Spoofing with RawNet2," ICASSP 2021 - IEEE International Conference on Acoustics, Speech and Signal Processing, 2021, pp. 6369–6373.
2. T. Kinnunen, M. Sahidullah, H. Delgado et al., "The ASVspoof 2019 Challenge: Spoofing Countermeasures for the Detection of Synthesized Speech," Computer Speech & Language, 2020.
3. M. Todisco, H. Delgado and N. Evans, "Constant Q Cepstral Coefficients: A Spoofing Countermeasure for Automatic Speaker Verification," Computer Speech & Language, 2017.
4. B. McFee et al., "Librosa: Audio and Music Signal Analysis in Python," Proceedings of the 14th Python in Science Conference, 2015.
5. F. Pedregosa et al., "Scikit-learn: Machine Learning in Python," Journal of Machine Learning Research, 2011, pp. 2825–2830.
6. L. Breiman, "Random Forests," Machine Learning, 45(1), 2001, pp. 5–32.
7. C. Cortes and V. Vapnik, "Support-Vector Networks," Machine Learning, 20(3), 1995, pp. 273–297.
8. J. H. Friedman, "Greedy Function Approximation: A Gradient Boosting Machine," Annals of Statistics, 2001, pp. 1189–1232.
9. C.Nagarajan and M.Madheswaran - 'Stability Analysis of Series Parallel Resonant Converter with Fuzzy Logic Controller Using State Space Techniques'- Taylor &Francis, Electric Power Components and Systems, Vol.39 (8), pp.780-793, May 2011. DOI: 10.1080/15325008.2010.541746
10. C.Nagarajan and M.Madheswaran - 'Experimental verification and stability state space analysis of CLL-T Series Parallel Resonant Converter' - Journal of Electrical Engineering, Vol.63 (6), pp.365-372, Dec.2012. DOI: 10.2478/v10187-012-0054-2
11. C.Nagarajan and M.Madheswaran - 'Performance Analysis of LCL-T Resonant Converter with Fuzzy/PID Using State Space Analysis'- Springer, Electrical Engineering, Vol.93 (3), pp.167-178, September 2011. DOI 10.1007/s00202-011-0203-9
12. S.Tamilselvi, R.Prakash, C.Nagarajan,"Solar System Integrated Smart Grid Utilizing Hybrid Coot-Genetic Algorithm Optimized ANN Controller" Iranian Journal Of Science And Technology-Transactions Of Electrical Engineering, DOI10.1007/s40998-025-00917-z,2025
13. S.Tamilselvi, R.Prakash, C.Nagarajan," Adaptive sliding mode control of multilevel grid-connected inverters using reinforcement learning for enhanced LVRT performance" Electric Power Systems Research 253 (2026) 112428, doi.org/10.1016/j.epsr.2025.112428

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.2026.0802198

14. S.Thirunavukkarasu, C. Nagarajan, 2024, "Performance Investigation on OCF and SCF study in BLDC machine using FTANN Controller," Journal of Electrical Engineering And Technology, Volume 20, pages 2675–2688, (2025), doi.org/10.1007/s42835-024-02126-w
15. C. Nagarajan, M.Madheswaran and D.Ramasubramanian- 'Development of DSP based Robust Control Method for General Resonant Converter Topologies using Transfer Function Model'- Acta Electrotechnica et Informatica Journal , Vol.13 (2), pp.18-31,April-June.2013, DOI: 10.2478/aeei-2013-0025.
16. C.Nagarajan and M.Madheswaran - 'DSP Based Fuzzy Controller for Series Parallel Resonant converter'- Springer, Frontiers of Electrical and Electronic Engineering, Vol. 7(4), pp. 438-446, Dec.12. DOI 10.1007/s11460-012-0212-0.
17. C.Nagarajan and M.Madheswaran - 'Experimental Study and steady state stability analysis of CLL-T Series Parallel Resonant Converter with Fuzzy controller using State Space Analysis'- Iranian Journal of Electrical & Electronic Engineering, Vol.8 (3), pp.259-267, September 2012.
18. C.Nagarajan and M.Madheswaran, "Analysis and Simulation of LCL Series Resonant Full Bridge Converter Using PWM Technique with Load Independent Operation" has been presented in ICTES'08, a IEEE / IET International Conference organized by M.G.R.University, Chennai.Vol.no.1, pp.190-195, Dec.2007
19. Suganthi Mullainathan, Ramesh Natarajan, "An SPSS and CNN modelling based quality assessment using ceramic materials and membrane filtration techniques", Revista Materia (Rio J.) Vol. 30, 2025, DOI: https://doi.org/10.1590/1517-7076-RMAT-2024-0721
20. M Suganthi, N Ramesh, "Treatment of water using natural zeolite as membrane filter", Journal of Environmental Protection and Ecology, Volume 23, Issue 2, pp: 520-530,2022
21. B. Logan, "Mel Frequency Cepstral Coefficients for Music Modeling," International Symposium on Music Information Retrieval, 2000.
22. X. Wang, J. Yamagishi and M. Todisco, "A Comparative Study of Recent Neural Spoofing Countermeasures," IEEE Journal of Selected Topics in Signal Processing, 2020.
23. Kiran, A., Rubini, P., & Kumar, S. S. (2025). Comprehensive review of privacy, utility and fairness offered by synthetic data. IEEE Access.
24. 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.
25. Pandi Prabha, S., & Rengarajan, A. (2025, February). Decentralized Resource Allocation Model Using Multi-agent Reinforcement Learning for Cloud Environment. In International Conference on Universal Threats in Expert Applications and Solutions (pp. 71-82). Singapore: Springer Nature Singapore.
26. 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.
27. Gopinathan, V. R. (2025). AI-Powered Kubernetes Orchestration for Complex Cloud-Native Workloads. International Journal of Research Publications in Engineering, Technology and Management (IJRPETM), 8(6), 13215-13225.