SJIF Impact Factor (2025): 8.688| ISI I.F. Value: 1.241| Journal DOI: 10.36713/epra2016 ISSN: 2455-7838(Online)
EPRA International Journal of Research and Development (IJRD)
Volume: 10 | Issue: 4 | April 2025 - Peer Reviewed Journal

DEEPFAKE AUDIO DETECTION USING MACHINE LEARNING ALGORITHMS

[1]Kumararaja Jetti, [2]Murakhna Chaitanya Kumar, [3]Kolli Chandu, [4]Muppavarapu Pradeep, [5]Lella Bhavya

[1]Assistant Professor Dept. of CSE, Bapatla Engineering College Bapatla
[2,3,4,5]Students Dept. of CSE, Bapatla Engineering College Bapatla

Article DOI: https://doi.org/10.36713/epra20941
DOI No: 10.36713/epra20941

ABSTRACT

Deepfake audio, which artificially replicates real voices, poses serious risks such as fraud, misinformation, and identity theft. While most research focuses on detecting deepfake videos, audio detection remains a relatively unexplored area. This study uses Mel-Frequency Cepstral Coefficients (MFCC) feature extraction, which mimics human auditory perception, to differentiate between real and synthetic speech. The Fake-or-Real dataset was used to train and evaluate multiple machine learning models. Among the tested models, the Support Vector Machine (SVM) performed best for short audio clips, while Gradient Boosting delivered better results for longer recordings. Additionally, the VGG-16 deep learning model achieved the highest accuracy of 93% on complex datasets. These results suggest that MFCC-based feature extraction combined with machine learning models can effectively detect deepfake audio.

KEYWORDS: *Deepfake Audio, Machine Learning, MFCC, Fake-or-Real Dataset, SVM, Gradient Boosting, VGG-16, Feature Extraction, Audio Detection, Cybersecurity.*

I. INTRODUCTION

Deepfake audio is a type of artificial intelligence technology that can create fake voices that sound almost real. This technology is becoming more advanced and can be used for harmful purposes like fraud, spreading false information, and stealing identities. While there has been a lot of research on detecting fake videos, detecting fake audio has not been explored as much. This makes it easier for scammers to use deepfake audio for illegal activities. Many traditional methods struggle to tell the difference between real and fake voices because AI-generated voices sound very natural. Some existing detection systems also have high error rates and do not work well in real-world situations. This study focuses on improving deepfake audio detection using machine learning. By extracting important features from audio and testing different machine learning models, we aim to develop a more effective way to identify fake voices and improve security in voice-based systems. Fake voice recordings, also known as deepfake audio, are becoming more common and can be very harmful. These recordings are made to sound just like real people and can be used to fool others. For example, someone might use a fake voice to pretend to be a trusted person and trick someone into giving away private information. This creates serious problems like scams, false news, and identity theft.

Most of the attention so far has gone toward spotting fake videos. However, fake audio is also a big threat, and not enough people have focused on solving this problem. Unlike video, where people might notice something odd in the visuals, fake audio can be harder to spot because the voice often sounds very real to our ears. In this study, we tried to find the best way to tell the difference between real voices and fake ones. To do this, we used a technique that listens to the most important parts of sound—similar to how our own ears and brain work when we hear someone speak. This helped the system focus on details like pitch, tone, and speed, which are often different in fake recordings. We used a collection of voice recordings called the "Fake-or-Real" dataset. It contains both real and fake examples. We trained several computer programs to listen to these voices and learn which ones were real and which ones were fake. Each program had its strengths. One model, called SVM, worked really well for short voice clips. It could quickly catch fakes when the audio was only a few seconds long.

Another method, called Gradient Boosting, was better at handling longer recordings. It was able to look at the full speech and make smart decisions based on the overall flow and style of speaking. This made it more accurate for more detailed recordings. Finally, we tested a more advanced model that looks at voice patterns as images. This model, known as VGG-16, gave the best results overall. In the end, the VGG-16 method was able to correctly identify fake and real voices with 93% accuracy. This shows that with the right tools and enough examples to learn from, we can create systems that are very good at spotting fake voices. These results are a big step forward in protecting people from being tricked by fake audio, and they highlight the need to keep improving audio detection methods as fake technology becomes more realistic.

SJIF Impact Factor (2025): 8.688| ISI I.F. Value: 1.241| Journal DOI: 10.36713/epra2016 ISSN: 2455-7838(Online)
EPRA International Journal of Research and Development (IJRD)
Volume: 10 | Issue: 4 | April 2025 - Peer Reviewed Journal

Existing System

- Challenges with Current Approaches: -
- Most research focuses on video deepfake detection, leaving audio detection less explored.
- Previous work relies on ASV Spoof and AV Spoof datasets, which lack diversity.
- ML models like SVM and CNNs struggle with high false positives.
- Deepfake audio detection is harder because AI-generated voices mimic real ones effectively.
- Computational cost of deep learning models is high, making real-time detection difficult.

Related Work

1. A Deep Learning Framework for Audio Deepfake Detection

Authors: J. Khochare, C. Joshi, B. Yenarkar, S. Suratkar, et al. (2021)

This study presents a system that can tell the difference between real and fake voice recordings. The researchers tested different computer models like Support Vector Machines (SVM) and Light Gradient Boosting Machines to improve detection accuracy. The focus is on recognizing voice patterns to spot audio that has been created to sound like a real person [5].

2. Deepfake Audio Detection via MFCC Features Using Machine Learning

Authors: Ameer Hamza, Abdul Rehman Javed, Farkhund Iqbal, Nataliia Kryvinska (2022)

This research aims to catch fake voice recordings by analyzing sound patterns that mimic human hearing. Using MFCC features and several machine learning methods, the team found that different models worked better depending on the length of the audio. Their results showed that the VGG-16 model gave the highest accuracy, helping distinguish between real and fake voices effectively [6].

3. Deepfake Audio Detection via Feature Engineering and Machine Learning

Authors: Farkhund Iqbal, A. Abbasi, A.R. Javed, Z. Jalil, J.N. Al-Karaki (2022)

This paper focuses on designing specific sound features that help detect fake voice recordings. By combining these features with machine learning methods, the study highlights the importance of analyzing unique sound characteristics to catch audio deepfakes that could be used for misleading or harmful purposes [7].

4. The Effect of Deep Learning Methods on Deepfake Audio Detection for Digital Investigation

Authors: M. Mcuba, A. Singh, R.A. Ikuesan, H. Venter (2023)

This study explores how modern computer programs can help digital investigators detect fake voices. The goal is to support cases where someone's voice is copied to mislead others. The research shows how deep learning methods can be used to find clues in sound that point to whether it's a real or cloned voice [8].

5.Deepfake Detection Using Deep Learning Methods: A Systematic and Comprehensive Review

Authors: A. Heidari, N. Jafari Navimipour, H. Dag, et al. (2024)

This review looks at all the major ways fake content is detected, including voice, image, and video. The authors go through many studies and explain how deep learning methods are used to spot deepfakes. It provides a clear picture of how these tools work, what they're good at, and what still needs to be improved to better catch deepfake content across media types [9].

Proposed System

To detect fake voice recordings, this study first uses a method called **Mel-Frequency Cepstral Coefficients (MFCC)**. MFCC is a sound feature extraction technique that copies how human ears hear sounds. It helps convert audio signals into a form that computers can understand. This technique captures key details about the tone, pitch, and rhythm of a person's voice, making it easier to tell real voices from fake ones.

After extracting these features, the next step involves using **machine learning models**. Several popular models were tested to see which could best tell the difference between real and fake voices. These models include Support Vector Machine (SVM), Gradient Boosting, and a deep learning model called VGG-16. Before training these models, the dataset was preprocessed and divided into training and testing sets to ensure fair evaluation.

The audio data used in this study comes from the **Fake-or-Real dataset**, which includes both real and synthetic (fake) speech samples. The MFCC features were fed into each model, and the models were trained to classify audio clips as either "real" or "fake." The training involved tuning parameters to increase performance and reduce misclassification, especially for challenging or noisy recordings.

SJIF Impact Factor (2025): 8.688| ISI I.F. Value: 1.241| Journal DOI: 10.36713/epra2016 ISSN: 2455-7838(Online)
EPRA International Journal of Research and Development (IJRD)
Volume: 10 | Issue: 4 | April 2025 - Peer Reviewed Journal

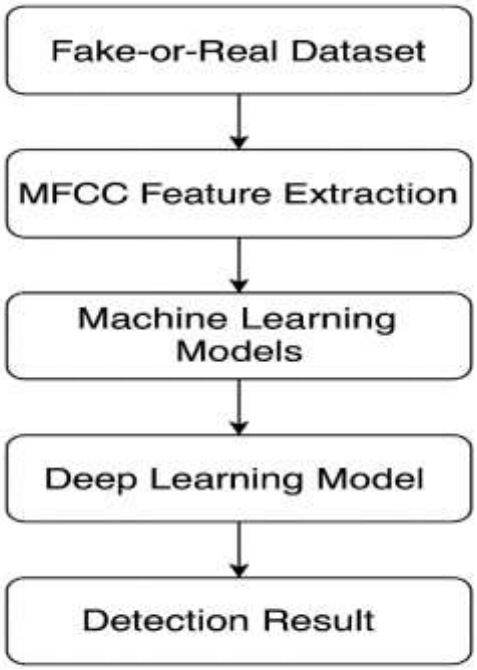


Figure-1 Proposed Architecture

- How Our Approach Improves Detection:
- Feature-Based ML Approach:
- Uses MFCC, which closely represents how humans perceive sound.
- Extracts key frequency and energy-based features from audio signals.

Dataset

- Uses the Fake-or-Real dataset, which includes AI-generated and real audio samples.
- Divided into four sub-datasets for better model evaluation.

Data Preprocessing

- Removes duplicate files and noise.
- Standardizes bit rates and normalizes volume levels.
- Feature Extraction:
- Extracts MFCC, spectral contrast, roll-off, bandwidth, and zero-crossing rate.
- Key Findings:
- SVM & MLP perform best for short clips.
- VGG-16 outperforms traditional ML models for longer and complex audio.
- Combining multiple feature extraction methods improves detection accuracy.

RESULTS

The experiments showed that different machine learning models performed better depending on the length and complexity of the audio clips. Support Vector Machine (SVM) produced the best results when analyzing short audio clips. It was able to detect subtle differences in tone and pitch between real and fake audio with high accuracy, making it useful for quick voice checks in short recordings. For longer voice recordings, the Gradient Boosting model gave better outcomes. This model builds several smaller decision models and combines them to improve the final decision. It was particularly good at handling variations in speech patterns and was more robust when the voice tone changed over time, as often seen in long dialogues or interviews.

The VGG-16 deep learning model, originally designed for image classification, was adapted to work with audio by converting the MFCC features into image-like formats. This model achieved the highest accuracy of 93%, especially when working with complex datasets containing both clean and noisy audio. Its layered structure helped it learn deep patterns in the voice data, leading to more

SJIF Impact Factor (2025): 8.688| ISI I.F. Value: 1.241| Journal DOI: 10.36713/epra2016 ISSN: 2455-7838(Online)
EPRA International Journal of Research and Development (IJRD)
Volume: 10 | Issue: 4 | April 2025 - Peer Reviewed Journal

accurate detection of deepfake audio. Overall, the results highlight that combining MFCC feature extraction with advanced models like VGG-16 can effectively detect even sophisticated fake voices. The system was also evaluated using standard performance metrics like accuracy, precision, recall, and F1-score, showing a strong balance between correctly identifying fake audio and avoiding false alarms. This research proves that fake voice detection is possible and practical with the right combination of sound features and learning models.

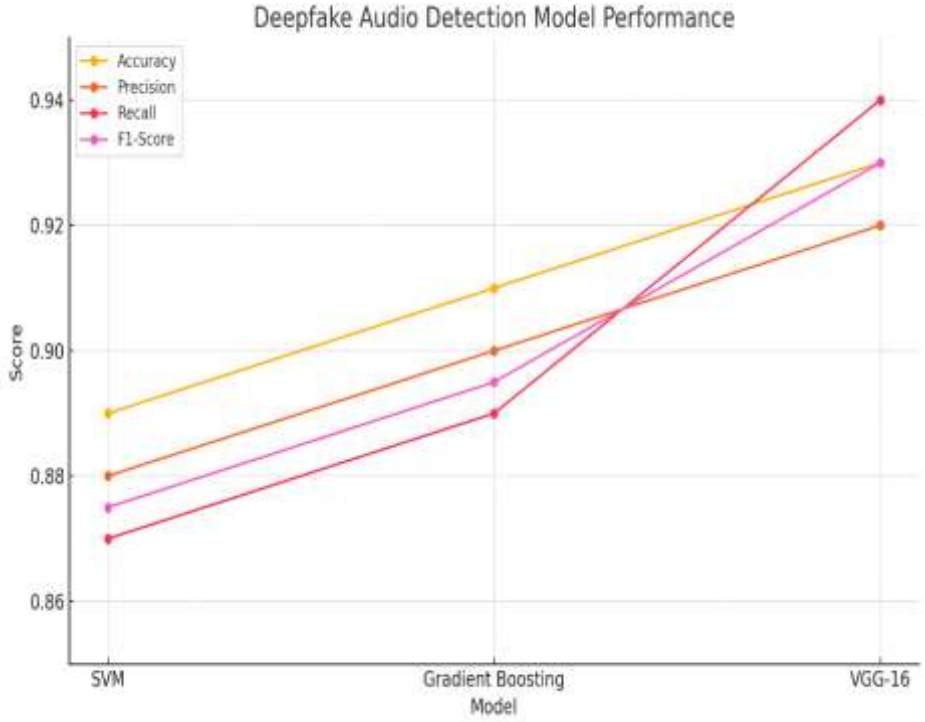


Figure-2 Performance Metrics

Model	Accuracy	Precision	Recall	F1-Score
SVM	0.89	0.88	0.87	0.875
Gradient Boosting	0.91	0.9	0.89	0.895
VGG-16	0.93	0.92	0.94	0.93

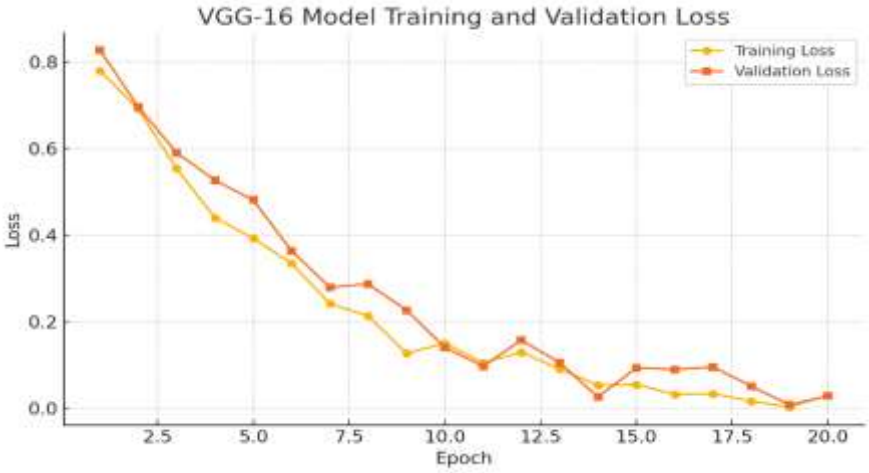


Figure-3 Loss Curve

SJIF Impact Factor (2025): 8.688| ISI I.F. Value: 1.241| Journal DOI: 10.36713/epra2016 ISSN: 2455-7838(Online)
EPRA International Journal of Research and Development (IJRD)
Volume: 10 | Issue: 4 | April 2025 - Peer Reviewed Journal

CONCLUSION AND FUTURE WORK

In this study, a reliable approach for detecting deepfake audio was presented using Mel-Frequency Cepstral Coefficients (MFCC) for feature extraction combined with various machine learning and deep learning models. The method successfully differentiates between real and fake voices by capturing essential audio characteristics and training models to recognize patterns specific to synthetic speech. Among the models tested, Support Vector Machine (SVM) performed best for short audio clips, Gradient Boosting worked well for longer recordings, and the deep learning model VGG-16 achieved the highest accuracy overall, reaching up to 93%. The results clearly show that MFCC-based features provide a strong foundation for detecting deepfake audio, especially when paired with well-tuned learning models. This highlights the potential of audio-based detection systems in helping prevent risks related to voice manipulation, such as fraud, misinformation, and identity theft. The study demonstrates that deepfake audio detection is not only feasible but also effective with the proper techniques the use of the Fake-or-Real dataset and performance metrics such as accuracy, precision, recall, and F1-score confirmed the robustness and reliability of the proposed system. The success of this work paves the way for further research and real-time applications in digital security and forensic investigations.

Future work may involve testing the system on multilingual datasets, incorporating noise-resistant models, and expanding to real-time detection environments. With continuous improvements, such methods can serve as essential tools in safeguarding against the growing threat of voice-based deepfakes.

REFERENCE

1. *Wijethunga, R. L. M. A. P. C., Matheesha, D. M. K., Al Noman, A., De Silva, K. H. V. T. A., Tissera, M., & Rupasinghe, L. (2020, December). Deepfake audio detection: a deep learning based solution for group conversations. In 2020 2nd International conference on advancements in computing (ICAC) (Vol. 1, pp. 192-197). IEEE.*
2. *Mahima, A. H., Monica, M., Neha, S., & Raykar, D. B. (2024, May). DeepFake Image, Video and Audio Detection. In International Research Conference on Computing Technologies for Sustainable Development (pp. 29-41). Cham: Springer Nature Switzerland.*
3. *Shaaban, O. A., & Yildirim, R. (2025). Audio Deepfake Detection Using Deep Learning. Engineering Reports, 7(3), e70087.*
4. *Shaaban, O. A., Yildirim, R., & Alguttar, A. A. (2023). Audio deepfake approaches. IEEE Access, 11, 132652-132682.*
5. *Hamza, A., Javed, A. R. R., Iqbal, F., Kryvinska, N., Almadhor, A. S., Jalil, Z., & Borghol, R. (2022). Deepfake audio detection via MFCC features using machine learning. IEEE Access, 10, 134018-134028.*
6. *Khochare, J., Joshi, C., Yenarkar, B., Suratkar, S., & Kazi, F. (2021). A deep learning framework for audio deepfake detection. Arabian Journal for Science and Engineering, 1-12.*
7. *Mcuba, M., Singh, A., Ikuesan, R. A., & Venter, H. (2023). The effect of deep learning methods on deepfake audio detection for digital investigation. Procedia Computer Science, 219, 211-219.*
8. *Iqbal, F., Abbasi, A., Javed, A. R., Jalil, Z., & Al-Karaki, J. N. (2022). Deepfake Audio Detection Via Feature Engineering And Machine Learning. In CIKM Workshops (pp. 1-12).*
9. *Heidari, A., Jafari Navimipour, N., Dag, H., & Unal, M. (2024). Deepfake detection using deep learning methods: A systematic and comprehensive review. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 14(2), e1520.*
10. *Chakravarty, N., & Dua, M. (2024). A lightweight feature extraction technique for deepfake audio detection. Multimedia Tools and Applications, 83(26), 67443-67467.*