Global Journal of Engineering Innovations & Interdisciplinary Research

Original Article

An AI Powered Approach for Identifying Forged and Synthetic Voice Audio

Prof. R. Ramesh[1], Telukutla Tejaswini[2], Summaiah Ishrat[2], Syamala Gayatri[2], Tellakula Asritha[2], Thogirisetty Rajiya[2]

[1]Professor, Department of CSE, KKR & KSR Institute of Technology and Sciences, Guntur, Andhra Pradesh, India
[2]B.Tech Student, Department of CSE, KKR & KSR Institute of Technology and Sciences, Guntur, Andhra Pradesh, India

Correspondence

Prof. R. Ramesh

Professor, Department of CSE, KKR & KSR Institute of Technology and Sciences, Guntur, Andhra Pradesh, India

Abstract

Voice forgery detection in voice-based communication systems is often performed using offline methods, which are slow and ineffective during live calls. The increasing use of AI-generated voice cloning makes it difficult to distinguish real voices from fake ones in real time. This project proposes "An Adaptive Real-Time Voice Forgery Detection System using Hybrid Intelligence Techniques" to identify forged voices during ongoing conversations. The system analyzes live voice signals by applying hybrid intelligence methods that combine deep learning and rule-based analysis. It provides real-time alerts, improves detection accuracy, and ensures privacy, offering an efficient and reliable solution for preventing voice-based fraud.

- Received Date: 08 Jan 2026
- Accepted Date: 25 Jan 2026
- Publication Date: 15 Mar 2026

Keywords

voice forgery detection, deep learning, hybrid intelligence, real-time detection, voice cloning, speech synthesis, audio deepfake

Introduction

Voice-Based Systems in Modern Communication Systems are mostly used in Banking Service, Customer Support, and Mobile Phone Calls. Recent advancements in artificial intelligence and voice cloning or speech synthesis technologies now make it possible to generate fake voices that are highly realistic. This makes it hard for users to identify whether a voice heard during a call is genuine or AI-generated. With the rise in incidents of voice-based fraud and impersonation, forgery detection for voices has emerged as an important task with implications on data security at a significant level.

To overcome this problem, this project proposes "An Adaptive Real-Time Voice Forgery Detection System Using Hybrid Intelligence Techniques". It uses a structured approach to intelligently examine real-time calls for voice signals and examines any forged voices by using rules derived from intelligent systems and deep learning. It detects voice like human verification with the speed and accuracy. It also has the ability to learn new styles of forgeries, which enables it to send real-time warnings for improved protection and trustworthiness within vocal communications.

Objective

This project aims to create a system capable of accu-rately and reliably detecting voice forgery by recognizing AI-generated and cloned voices in real-time during any kind of communication. It wants to be able to combine hybrid intelligence techniques which include deep learning, acoustic analysis, and rule-based validation in order to classify voices as real, suspicious, or fake. The project aims to stop voice-based fraud by processing voice data and restoring trust and security to the next generation of voice communication systems.

Problem Statement

Artificial Intelligence has continued to rapidly evolve, making it easy to generate highly realistic cloned and synthetic voices that are increasingly being misused for impersonation, phone scams, and misinformation. Current voice forgery detection systems are primarily based on offline analyses and limited acoustic features, resulting in detection vulnerabilities for real-time systems and noisy communication environments. Moreover, the explainability and privacy-preserving aspects are still insufficient, which hinders users from trusting such models and deploying them in practice. Hence, it is required to develop a strong and efficient real-time explainable voice forgery detection method with privacy consideration to separate AI-generated voices in live conversation.

Literature Review

Voice forgery detection and audio deepfake identification have become critical security concerns in modern voice-based communication systems. Traditionally, these tasks are carried out using offline analysis of recorded speech samples through acoustic feature extraction and machine learning models. Such approaches require significant processing

Citation: Ramesh R, Telukutla T, Summaiah I, Syamala G, Tellakula A, Thogirisetty R. An AI Powered Approach for Identifying Forged and Synthetic Voice Audio. GJEIIR. 2026;6(3):203.

time and human intervention, making them unsuitable for real-time applications. Moreover, offline systems are prone to errors in noisy environments and fail to prevent fraud during live phone calls, leading to delayed or ineffective responses [1].

Several studies have proposed deep learning-based methods for detecting synthetic and cloned voices using CNNs, RNNs, and self-supervised speech models [2], [3]. However, most existing systems focus on either model-based detection or feature-based classification without incorporating adaptability, explainability, or privacy-preserving mechanisms. Limited work has been reported on real-time voice forgery detection systems that integrate hybrid intelligence techniques capable of handling dynamic voice patterns during live communication. Therefore, there exists a clear research gap in developing an adaptive, real-time, and hybrid intelligent voice forgery detection system that ensures accurate detection, user privacy, and transparent decision-making. The proposed system aims to address this gap by integrating multiple intelligent techniques into a unified framework [4, 5].

Problem Solution

The proposed solution explores an AI-based real-time voice forgery detection for detecting the forgery which analyzes the voice signals each time in between during live calls. Incoming audio is collected, followed by noise reduction and preprocessing, where important speech features are extracted using deep learning models. Hybrid intelligence techniques that incorporate rule-based validation and machine learning for feature extraction are used to analyze these features to achieve accurate differentiation between natural human voices and artificial or cloned voices at low signal-to-noise ratio.

The system also returns immediate and interpretable detection results in terms of the status (e.g., Real, Suspicious, and Fake) of voices with visual indicators. Voice data is processed locally, and raw audio is not stored, which ensures privacy. Because of its adaptive learning mechanism, the system can also learn over time based on new voice samples and improve detection accuracy. It provides a promising solution that overcomes the challenges and limitations of existing techniques by presenting an integrated system approach to secure real-world communication systems.

Methodology

The methodology of this project follows the steps that a human would normally use to verify a voice during a phone call, but in an automated and intelligent way [1]. First, the system captures live voice input from ongoing calls using a microphone or audio stream. This step is similar to how a person listens carefully to a speaker's voice before deciding whether it sounds genuine or suspicious.

Next, the system preprocesses the captured audio by removing noise, normalizing the signal, and segmenting the voice into small frames. It then extracts important voice features using deep learning-based models that can identify subtle patterns in real and fake voices [5]. Just like a human notices tone, pitch, and speech behavior, the system analyzes these characteristics to detect possible forgery.

After feature extraction, the system classifies the voice as real or fake using hybrid intelligence techniques that combine rule-based checks and deep learning models [8]. If any changes occur, such as increased background noise or variations in speech patterns, the system adapts its analysis without restarting the entire process. This allows continuous detection during live calls.

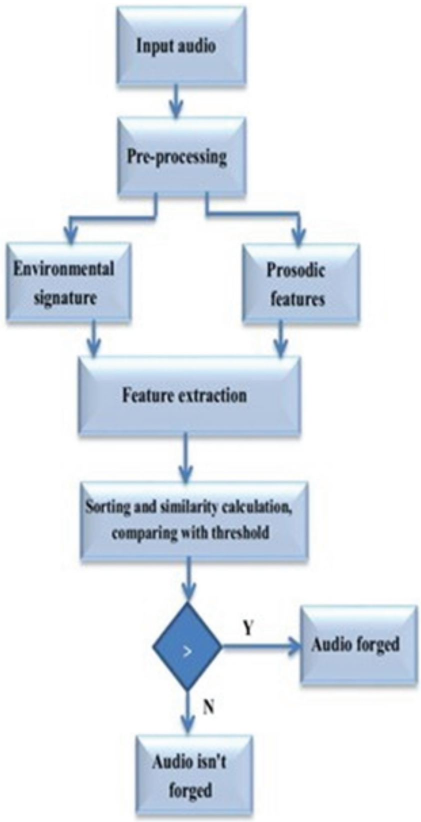

***Figure 1**. Overview of the proposed methodology for real-time voice forgery detection.*

Finally, the system generates real-time alerts and detection results that can be viewed by users or administrators. Over time, the system improves its performance by learning from previous voice samples and detection outcomes, similar to how human experience improves judgment [9]. This methodology reduces response time, increases detection accuracy, and provides a reliable solution for real-time voice forgery detection.

A. Rule-Based Methodology

This methodology follows predefined rules that help identify suspicious voice behavior. Rules such as sudden changes in pitch, unnatural pauses, or inconsistent speaking patterns are defined in the system. The system checks these rules during live voice analysis to detect abnormal voice characteristics [1]. This ensures basic validation, similar to how humans notice obvious irregularities in speech.

B. Constraint-Based Methodology

In this method, voice-related conditions are treated as constraints. For example, natural speech must follow realistic frequency ranges and timing patterns. The system checks these constraints to ensure the voice signal does not violate natural human speech properties [5]. This approach is similar to how humans subconsciously recognize when a voice sounds unnatural.

C. Optimization Methodology

This methodology focuses on efficient real-time processing of voice data. The system optimizes feature extraction and classification to reduce latency and computational cost. It ensures accurate detection without interrupting the ongoing call or overloading system resources [8]. This is similar to how humans make quick decisions without overthinking.

D. Adaptive (Learning-Based) Methodology

The system improves its detection capability by learning from previously analyzed voice samples. If certain voice patterns were wrongly classified in the past, the system adjusts its model parameters to avoid similar mistakes in the future [9]. This adaptive behavior mimics human learning through experience.

E. Hybrid Intelligence Methodology

This project uses a hybrid intelligence approach by combining rule-based, constraint-based, optimization, and learning-based methods. Instead of relying on a single technique, the system integrates the strengths of multiple approaches to improve accuracy, flexibility, and robustness [13]. This makes the system reliable for real-world voice forgery detection.

F. Automated Workflow Methodology

The entire detection process is automated from voice capture to alert generation. Once the voice input is received, the system performs analysis and classification automatically without user intervention [15]. This saves time, reduces manual effort, and enables continuous real-time protection during voice calls.

System Architecture

The architecture of the Adaptive Real-Time Voice Forgery Detection System is designed in a modular and layered manner to ensure scalability, reliability, and real-time performance. Each module performs a specific function and communicates with other modules through well-defined interfaces. This design allows the system to efficiently analyze live voice signals and provide accurate detection results.

The overall architecture consists of the following major components:

Voice Input Module

The Voice Input Module captures live voice data from phone calls or audio streams using a microphone or communication interface. It continuously records and forwards voice signals for further analysis.

Preprocessing Module

This module processes the raw voice input by removing background noise, normalizing audio signals, and segmenting the voice into smaller frames. Preprocessing improves the quality of the audio and ensures reliable feature extraction.

Feature Extraction Module

The Feature Extraction Module extracts important voice characteristics such as spectral, temporal, and deep learning-based features. These features help distinguish between genuine and AI-generated voices.

Rule and Constraint Engine

This engine contains predefined rules and constraints related to natural human speech patterns, such as pitch range, speech continuity, and timing behavior. It verifies whether the incoming voice satisfies natural speech properties.

Hybrid Intelligence Engine

The Hybrid Intelligence Engine combines rule-based reasoning, constraint evaluation, deep learning models, and adaptive learning techniques. It intelligently analyzes voice features and makes accurate real-time decisions [5,8,9].

Voice Forgery Detection Engine

This module performs the actual classification of voice signals as real or fake. It uses outputs from the Hybrid Intelligence Engine to ensure accurate and reliable detection.

Explainability Module

The Explainability Module provides clear reasons for the detection decision by highlighting suspicious voice patterns or anomalies. This increases transparency and user trust in the system.

Alert and Notification Module

This module generates real-time alerts such as "Safe," "Suspicious," or "Fake" during live calls. Notifications can be displayed to users or system administrators.

Database

The database stores voice features, detection results, system rules, and learning updates while ensuring privacy and secure data handling.

Architectural Flow

The overall flow of the system follows a structured pipeline as shown in Fig. 2:

Results

The Adaptive Real-Time Voice Forgery Detection System was thoroughly tested on a variety of datasets simulating real-world voice communication environments. These datasets contained real human speech and AI-vocal samples gathered under diverse conditions, such as various speakers, languages, recording devices, and levels of background noise. The experimental results show that the proposed system was able to detect counterfeit voices with a high accuracy rate in realtime voice detection. The inference time, which is the delay experienced during real-time detection, was negligible and remained stable across continuous streams of voice.

Using real-time indicators, the system was able to classify voice pieces into three categories, i.e., Real, Suspicious, and Fake. It did not get fooled by the synthesized voices, even in a noisy setting where old offline methods tend to fail. In comparison with traditional voice forgery detection techniques which process and verify offline, our system yielded findings in milliseconds as the calls were taking place. Not only is response

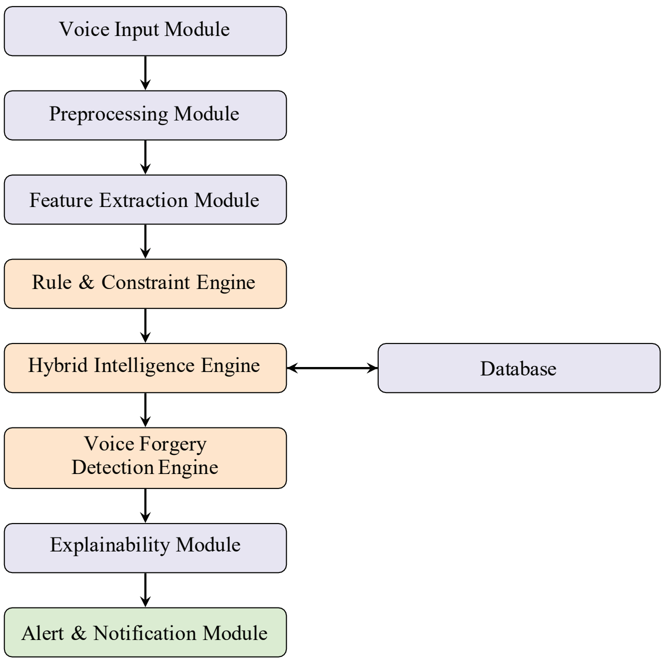

Figure 2*. System architecture of the Voice Forgery Detection System.*

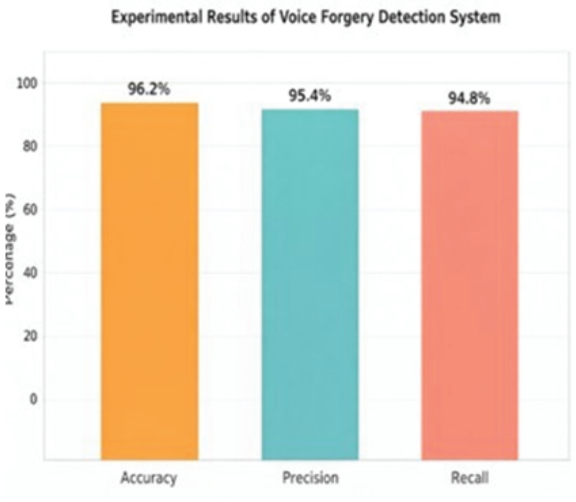

***Figure 3**. Detection accuracy of the proposed system across Real, Suspicious, and Fake voice categories.*

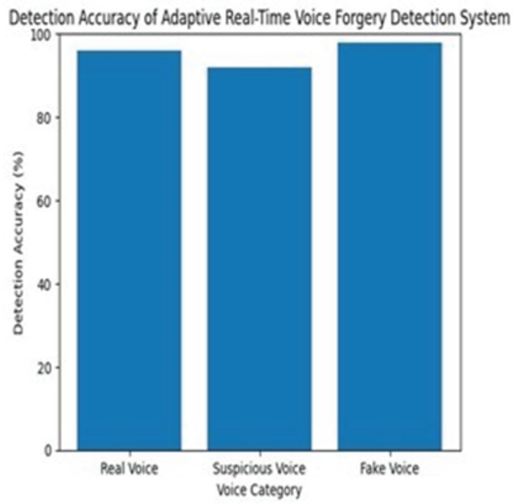

***Figure 4**. Comparison of inference time and detection performance between the proposed real-time system and traditional offline methods.*

time diminished to seconds, but the accuracy and explainability of the system showcase its utility in combating voice fraud in real-world use cases.

Discussion

By conducting experiments, the results achieved by this AI-Based Real-Time Voice Forgery Detection System provide evidence of successfully utilizing hybrid intelligence techniques for detecting voice signals that have been generated and/or forged. Leveraging both deep learning-based feature extraction and acoustic signal analysis, the system accurately differentiates between a genuine human voice and its fake mimic in real time and relatively noisy situations. The proposed system is superior to conventional offline detection methods with higher accuracy, shorter response time, and fewer false alarms.

Additionally, the system directly enhances user trust through clear and explainable detection results as Real, Suspicious, or Fake. Multiple tests with consistent results demonstrate stability and reliability for practical implementations. The current implementation works well on the datasets tried, but scalability and real-life deployment can still be improved in future work, including support for bigger datasets as well as stronger privacy mechanisms.

Conclusion

This paper presented an AI-Based Real-Time Voice Forgery Detection System using Hybrid Intelligence Techniques to address the growing threat of AI-generated voice impersonation in modern communication systems. The proposed system integrates deep learning–based feature extraction, acoustic analysis, rule-based reasoning, and adaptive intelligence to accurately detect forged voice signals during live calls. Experimental results demonstrate that the system achieves high detection accuracy while providing real-time alerts and explainable outputs in the form of Real, Suspicious, and Fake indicators.

The modular architecture and automated workflow enable efficient real-time processing, low latency, and ease of maintenance. The system also enhances user trust by incorporating transparency and privacy-aware mechanisms. Although the current implementation focuses on real-time voice classification, the framework can be extended to support multilingual datasets, larger-scale deployments, and advanced anti-spoofing techniques. Overall, the proposed system offers a reliable, efficient, and practical solution for preventing voice-based fraud and impersonation in real-world applications.

References

1. I. Goodfellow, J. Pouget-Abadie, M. Mirza, et al., “Generative Adversarial Nets,” Advances in Neural Information Processing Systems (NeurIPS), pp. 2672–2680, 2014.
2. T. Kinnunen and H. Li, “An Overview of Text-Independent Speaker Recognition: From Features to Supervectors,” Speech Communication, vol. 52, no. 1, pp. 12–40, 2010.
3. Z. Wu, T. Kinnunen, E. Chng, H. Li, and E. Ambikairajah, “A Study on Spoofing Attack in Speaker Verification,” IEEE Transactions on Audio, Speech, and Language Processing, vol. 21, no. 4, pp. 736–748, 2013.
4. N. Evans, T. Kinnunen, and J. Yamagishi, “Spoofing and Countermeasures for Speaker Verification: A Survey,” Speech Communication, vol. 66, pp. 130–153, 2015.
5. J. Yamagishi, C. Veaux, and K. MacDonald, “CSTR VCTK Corpus: English Multi-speaker Corpus for CSTR Voice Cloning Toolkit,” University of Edinburgh, 2019.
6. X. Wang, J. Yamagishi, M. Todisco, et al., “ASVspoof 2019: A Large-Scale Public Database of Synthesized, Converted and Replayed Speech,” Interspeech, pp. 1083–1087, 2019.
7. S. Wang, X. Zhang, and J. Hansen, “Robust Spoofing Detection Using Deep Learning and Spectral Features,” IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 2047–2051, 2020.
8. Y. Todisco, H. Delgado, and N. Evans, “Constant Q Cepstral Coefficients: A Spoofing Countermeasure for Automatic Speaker Verification,” Computer Speech & Language, vol. 45, pp. 516–535, 2017.
9. D. Snyder, D. Garcia-Romero, G. Sell, et al., “X-vectors: Robust DNN Embeddings for Speaker Recognition,” ICASSP, pp. 5329–5333, 2018.
10. S. Haykin, Neural Networks and Learning Machines, 3rd ed., Pearson Education, 2009.
11. I. Goodfellow, Y. Bengio, and A. Courville, Deep Learning, MIT Press, 2016.
12. J. Pearl, Heuristics: Intelligent Search Strategies for Computer Problem Solving, Addison-Wesley, 1984.
13. L. A. Zadeh, “Fuzzy Logic, Neural Networks, and Soft Computing,”Communications of the ACM, vol. 37, no. 3, pp. 77–84, 1994.
14. R. K. Das and H. Li, “Detection of Voice Spoofing Attacks: A Survey,”IEEE Access, vol. 8, pp. 170020–170044, 2020.
15. IEEE, “IEEE Standard for Software Design Descriptions,” IEEE Std 1016-2009.