SURVEY ON VISHGUARD-A VOICE PHISHING DETECTOR

E. RAJKUMAR
Department of Cyber Security
SRM Valliammai Engineering College
Tamilnadu,India
rajkumar.cys@srmvalliammai.ac.in

R. ABIRAMI
Department of cybersecurity
SRM Valliammai Engineering College
Tamilnadu,India
abiramiravi4200@gmail.com

S. BHARATH KRISHNA
Department of Cyber Security
SRM Valliammai Engineering College
Tamilnadu,India
bharathkrishna1404@gmail.com

K. KRISHNA PRASADH
Department of Cyber Security
SRM Valliammai Engineering College
Tamilnadu,India
krishnaprasadha16@gmail.com

 DOA:01.12.2025 / DOP: 10.12.2025 ISSN: 3048-619X
Volume 2 Issue 3 [Sep – Dec 2025] DOI:

ABSTRACT **—Vish Guard is a real-time system for detecting voice phishing that protects users from vishing attacks. It captures audio from live microphone streams or uploaded files and turns them into text. The system then checks the text with a trained machine learning classifier to see if the interaction is safe or suspicious. By setting a confidence threshold, it reduces false positives and sends alerts through visual cues and repeated voice prompts. Built on Streamlit, it has a lightweight and modular design with local data processing, which ensures privacy, easy integration, and improved awareness of voice-based scams.**

Keywords: real-time call monitoring, vishing detection, speech-based threat analysis, voice phishing detection

1. INTRODUCTION

Vishing is one of the most common type of cyber attack in today's attack dominating world.Due to the growth of digital communication, voice-based authentication solutions are heavily used in banking, customer support, and smart gadgets. However, increased usage has also given rise to concerns about security threats, such as vishing (voice phishing), spoofing attack, and speech-based fraud. The flaws in voice recognition technology are used by cybercriminals to impersonate the users, to trick victims, and also to bypass authentication procedures. Thus, it is necessary to ensure the security and dependability of voice authentication systems. Vishing, voice spoofing, and speech-based fraud have become more prevalent issues with the growing adoption of voice-based authentication methods in banking, customer support, and smart devices. Cybercriminals exploit these flaws in a speech recognition technology to trick an victim and overcome the authentication procedures using a social engineering and AI-generated voices. High-end threats such as replay attacks, deep fake voice impersonation, and fake speech usually bypass the traditional call security mechanisms like caller ID validation and blacklisting too. Modern security solutions and AI services use ML-based and AI-based speech analysis and recognition techniques to identify the fraudulent voice activity instantly.

Ml models trained on the extreme large datasets employing voiceprints, speech patterns and anomaly detection for identifying fake voice interactionsreal time monitoring of calls as they're incoming, together with keyphrase-based language detection ("scam", "lotto", "urgent payment") and liveness verification would go a long way in heightening security and prevent fraud-deals. Literature survey of 20 IEEE research papers focusing on voice-authentication security, spoofing detection & anti-vishing strategy. Key advancements and emerging trends as well as future research direction in securing voice-communications against cyber-threats. Traditional authentication methods relying solely on speaker recognition are now insufficient due to the emergence of adversarial techniques that will impersonate the legitimate users.

2. LITERATURE SURVEY

Vishing is a form of social engineering where attackers use fraud phone calls to trick individuals into revealing their sensitive information, such as passwords, personal details, or banking credentials. Unlike normal phishing, which relies on emails, vishing exploits human knowledge over phone conversations. This paper reviews various approaches to detecting and preventing vishing, focusing on speech analysis, AI-driven detection, and existing anti-vishing solutions.

2.1 Detection Mechanisms for Voice Spoofing

Detecting voice spoofing attacks is crucial to the security of voice authentication systems as attackers are increasingly using dangerous attacks like replay, synthesized speech and voice convesion. Different researches have proposed different novel approaches for spoof detection improvement. This includes Time-frequency spectrum difference analysis [12] is a lightweight and effective replay attack detection approach based on the difference in energy distributions of replay or tampered voice signals; Integral knowledge amalgamation [19] to generalize AI generated speech detection by fusing multiple detection models for spoofing resilience against evolving adversarial attacks. To counter spoofing attack and speech synthesis, embedding-based speaker verification models are fused with spoof awareness to identify genuine/synthesised talk [9]. Self-supervised learning is adopted to increase spoofing resilience of speaker verification models to generalize unseen spoofing techniques [11]. Deep learning models are fused by training spoofing models with different attack scenarios to develop robust spoofing countermeasures to increase generalisation capability for different spoofing conditions [7]. Since attackers are using more advanced techniques including replay attacks, synthesised speech, and voice conversion to get around security measures, detecting voice spoofing attacks is essential for protecting voice authentication systems [1]. Strong countermeasures that enhance generalisation in a range of spoofing circumstances have also been built using deep learning-based fusion models that have been trained with a variety of assault scenarios [3].

Another mmWave based voice sensing introduced using RF and acoustic signal together as a voice verification mechanism, thus offering a safeguarded layer against voice spoofing [20]. Speech movement synchronization analysis with accelerometer-based verification is exploited as an anti-spoofing technique in wearables to support the fact that the detected speech is produced physically from the user and not a recorded speech [13].Graph neural network (GNN)- based fraud detection is deployed to detect spoofing pattern on mobile social network by mapping suspicious interaction to improve anomaly detection in voice-based communication [14].

Research demonstrates that silence regions in speech is also used for spoofing detection where synthetic voices often contain unnatural silent regions and can be used to detect spoofing [17]. However, these implementations are hindered by real-world challenges of low-latency real-time detection with minimal false positives. Future research should concentrate on hybrid AI-driven frameworks, multi-modal biometric authentication and adversarial training methods for building robust voice authentication systems against this evolving spoofing threat.

2.2 Machine Learning in Voice Authentication

Machine learning (ML) has empowered voice authentication to detect spoofing attacks, impersonation attempts, and AI-generated voice with high accuracy. Rule-based approaches had difficulties keeping up with the rapidly changing attack types but ML-based approaches trained models using deep learning to capture voice patterns and detect anomalies. Self-supervised learning is a popular technique used to train models on large amounts of data to verify speakers and distinguish them from the legitimate users and attackers in different unseen spoofing conditions [11]. One of the most important ML-based voice authentication is embedding-based speaker verification model which takes the voice input, extract high dimensional features and compare it against stored profiles to verify the users [9]. Spoofing detection and voice authentication accuracy are greatly enhanced by machine learning models. Using RF and acoustic signals, mmWave-based voice

sensing has become a powerful defence against voice spoofing [5]. Furthermore, accelerometer-based verification guarantees that the speech being identified is from a live speaker instead of a recording of playback [8].

Spoof alertness training, a security feature, was inserted into these models making them robust to voice conversion, text-to-speech attacks [7], and feature decomposition learning has been proposed to separate real and fake voice features for better detection [7].Graph neural networks (GNN) to study voice samples relating fraud detection in mobile

networks, which helps to track large voice spoofing operations [14]. GAN was also used to detect faked voice training, where real vs. fake voices were learned by the model [6]. The open-world could be better identified with soft-contrastive pseudo learning so that authentication systems are still effective against new attacks. [4]. Apart from ML-based approaches, confidence of wearable devices was improved with the help of accelerometer-based anti-spoofing technique in a hybrid model with voice authentication [13].A novel RF-based technique, mmWave-based vocal sensing, combined with ML, made it tough for attackers to easily bypass authentication [20].While these ML-based techniques have increased the security of voice-based authentication, issues of adversarial attacks and privacy concerns along with existing limitations on real-time processing are still to be addressed. Future research needs to focus on multi-modal biometric fusion, few-shot learning, and adaptive adversarial training to further enhance ML-based voice authentication system.

The best model for spoof detection is embedding-based speaker verification model which projects the voice into high dimensional space and cluster them for comparison [9]. Also, self-supervised learning of the authentication models allows generalisation so that it can detect known as well as unknown spoofing techniques [11]. ML models based on feature decomposition learning and synthesiser feature augmentation helps in distinguishing real and synthesised voice and increases the robustness of detection [7]. GNNs have also been used to figure out the relations between fraudulent voice patterns and to detect large scale fraud in mobile networks [14]. GANs are further used for many-to-many voice conversion detection to make the authentication models robust to voice spoofing attacks [6]. Detection models can be combined further with accelerometer- based anti-spoofing and radio frequency (RF) sensing for hybrid models [13] or biometric fusion with accelerometer and RF sensing [20]. These techniques support liveness detection to avoid attackers to use the recorded or synthesised voice samples to bypass the authentication. Still, adversarial AI attacks, real-time processing, and user privacy concerns are challenges for voice authentication security. Future research can focus on lightweight deep learning architectures,

2.3 Liveness Detection in Voice Authentication

Liveness detection is necessary in voice authentication to make sure the voice is really coming from a living speaker, not a replay, generation or modification of a voice. Initially, challenge-response is used whereby the user is asked to repeat some phrases to verify. However, this is easy to be attacked by the more advanced attacker who uses AI-synthesized or deepfake voice to tamper the response. Currently there are different liveness detection techniques such as biometric signal analysis[12], accelerometer based authentication[13], and RF voice sensing[20] in preventing replay, generation or modification of voice. Accelerometer based anti-spoofing has some advantages in wearable application where the motion pattern during speaking is used to distinguish real vs spoofed voice. RF voice sensing can capture very delicate vocal cord vibration and airflow pattern and is not susceptible to replay and deepfake attempts [20]. Synthesizer feature augmentation and feature decomposition learning are also proposed in assisting AI-synthesized voice detection which makes it less possible for the attacker to replay the generated voice in bypassing the authentication[7].

Deep learning models trained on spoof databases like PartialSpoof improve detection by detecting the presence of short fake speech segments embedded in the real utterances [16]. One way is self-supervised learning for generalized spoofing detection where models learn latent asymmetries between real and fake speech without needing large amounts of labelled data [19]. Adaptive knowledge amalgamation

frameworks further help improve the system's ability to recognize and nullify newly unseen attacking patterns that strengthen voice authentication's robustness. Despite these contributions, real-time implementation, adversarial attacks and privacy issues necessitate challenges for liveness detection. Future research should focus on lightweight neural architectures, multimodal fusions and privacy-preserving models of AI, to ensure secure and user-friendly voice authentication systems. One of the best liveness detection techniques is accelerometer-based anti- spoofing where motion sensors in wearables can detect natural vocal tract speech movements to confirm the head and vocal cords vibration pattern matches that of the speaker. This makes replay attacks and deepfake generated voice harder to bypass authentication [13]. Similarly, RF-based voice authentication leverages millimeter-wave sensing to capture fine-grained speech-induced vibrations and airflow pattern, preventing voice cloning and recorded playback attacks [20].

2.4 Speech-based threat analysis

Speech-based threat recognition is critical to protect voice communication systems from various vishing attacks, voice spoofing, impersonation fraud, etc.

Several works explored new techniques for detecting these threats using machine learning, signal processing, biometric authentication, etc. Such a technique is spoof-aware speaker verification based on embedding-based learning models which help in recognizing threats by analyzing high-dimensional voice representation that provides robustness against adversarial attacks[9]. Self-supervised learning for speaker verification has been implemented in authentication systems to allow real-time adaptive fraud voice detection[11]. Another major area of interest for threat recognition is voice spoofing that employs techniques like time-frequency spectrum difference analysis, etc. for the detection of replay and synthesis attacks[12]. Integral knowledge amalgamation for generalized spoof detection has also been proposed to improve the detection of AI-generated speech to ensure security from deepfake voice attacks[19]. In addition to this, mmWave-based vocal sensing has been utilized for a multimodal approach towards human voice analysis that makes use of radio frequency (RF) sensing coupled with acoustic sensing to enable accurate authentication[20]. Techniques for liveness detection guarantee that authentication systems are able to distinguish between spoof and authentic speech. To detect suspicious interactions and enhance anomaly identification in voice-based communications, fraud detection based on Graph Neural Networks (GNNs) has been proposed [15]. Given that synthetic voices frequently have artificial silent sections, studies have demonstrated that these speech silences can be accurate markers of spoofing [18].

Emotion and purpose recognition are added to speech-based threat analysis in order to identify fraud. Speech synthesis using high-generalization emotion representation models can identify coercion, stress, or urgency in a phone call, which facilitates the detection of dishonest speech [1]. A method for tracking down AI-generated or altered speech segments and identifying deepfake or cloned voices is soft- contrastive pseudo learning for forged speech attribution [4]. It is also suggested to use wearable anti-spoofing techniques, such as accelerometer-based anti-spoofing authentication and speech-movement synchronisation analysis, to guard against playback attacks. These techniques provide more real-time authentication layers by analysing the relationship between movement and voice activity using accelerometer signals [13]. Detecting voice spoofing attacks is crucial for securing voice authentication systems, as attackers increasingly employ sophisticated methods such as replay attacks, synthesized speech, and voice conversion to bypass security mechanisms [2].

Fraud detection for mobile social networks using graph neural networks (GNNs) is another exciting advancement. By mapping questionable phone exchanges and identifying unusual caller behaviours, these models identify social engineering attacks, improving security beyond voice analysis [14]. Despite tremendous advancements, there are still obstacles in the way of scalable, adaptable, and real-time speech-based threat analysis solutions. Future research will keep using adversarial training, multimodal verification, and self-supervised learning to fortify speech security systems against ever-more- advanced social

engineering and deepfake threats.

2.5 Framework for Integrating Spoofing Detection with Vishing Prevention

The danger of fraudulent calls can be considerably reduced with an integrated framework for spoofing detection and vishing prevention that incorporates adaptive countermeasures, AI-based threat identification, and real-time voice authentication. detecting voice replay, synthetic speech, or deepfake-based impersonation attempts before to processing an incoming call by employing spoofing detection algorithms such feature decomposition learning, time-frequency spectrum analysis, and other techniques [7]. By seeing harmful patterns in mobile networks, graph neural networks for fraud detection can be utilised to identify repeated vishing attempts [14].

To stop attackers from employing prerecorded messages or AI cloned voices, accelerometer-based anti-spoofing technologies that identify vocal vibrations to verify the speaker's authenticity can also be included [13]. The amount of false positives can be decreased by using radio frequency (RF) sensor technologies to distinguish genuine human voices from recorded or synthesised speech [20]. This system can stop vishing efforts before they can trick their victims by using liveness detection, anomaly detection, and real-time call monitoring. Future developments can concentrate on adversarial training to increase resistance to emerging attacks, multitasking authentication, and privacy-preserving AI models.

To guard against fraudulent calls, a robust integrated spoofing detection vishing prevention system integrates network threat detection, biometric authentication methods, and several AI security features. To detect deepfake or synthetic speech, the

system makes use of speech feature anomaly and deep learning-based acoustic pattern recognition [7]. It analyses speech patterns using feature decomposition learning and compares them to a database of real and fake sounds for further detection [12]. Accelerometer-based anti-spoofing authentication is introduced for replay and impersonation attempts to identify speakerneck-induced vibrations and head movement patterns to determine whether the speaker is physically present [13].

To distinguish human speech from AI-generated speech, RF-based human speech sensing is additionally included [20]. In order to counteract text- to-speech or pre-recorded audio attacks, the system additionally uses liveness detection to identify spontaneous speech characteristics [19].The system uses automatic keyword identification using natural language processing (NLP) techniques to highlight calls that contain suspicious words and phrases (such "lottery," "urgent money," and "bank verification") that are frequently used in vishing scams for real-time call monitoring. When a call is flagged as high-risk, the user receives an immediate voice alert about a possible scam attempt [16]. Infrequent but extremely significant attacks are nonetheless detected without overfitting too many false positives thanks to a cost- sensitive fraud detection model [14].This platform improves defences against voice-based social engineering assaults by integrating AI-powered spoofing detection, real-time call monitoring, and adaptive fraud detection. To further improve security without compromising user privacy and accessibility, future research could use adversarial training, multimodal authentication, and privacy-preserving AI models.

3. CONCLUSION

In conclusion, it is crucial to make sure that speech authentication systems are resilient to deepfake manipulation, voice spoofing, and vishing attacks as they grow more common in banking, smart devices, and secure voice communications. Advanced detection techniques currently in use may successfully differentiate between real and false voices to stop security breaches by utilising machine learning, deep learning, and multi-modal authentication. The ability to recognise artificial, replayed, or mimicked speech is further strengthened by liveness detection, adversarial training, and real-time speech-based threat analysis. Furthermore, decentralised verification methods like blockchain-based authentication and privacy-preserving AI models help create a more reliable and tamper-resistant voice security system. Sustaining trust, security, and integrity in speech- enabled apps will require

ongoing research and development in adaptive AI-driven voice security solutions as threats change. In order to increase performance and privacy, future developments in voice authentication security will concentrate on improved real-time detection, flexible AI models, and low-power on-device processing. In order to ensure compliance with data protection requirements, federated learning and privacy-preserving AI will enable systems to learn from new spoofing efforts without disclosing private user information.

4. REFERENCES

J. Zheng, J. Zhou, W. Zheng, L. Tao and H. K. Kwan, "Controllable Multi-Speaker Emotional Speech Synthesis With an Emotion Representation of High Generalization Capability," in *IEEE Transactions on Affective Computing*, vol. 16, no. 1, pp. 68-82,2025.

[2] Y. Ahn, J. Chae and J. W. Shin, "Text-to-Speech With Lip Synchronization Based on Speech- Assisted Text-to-Video Alignment and Masked Unit Prediction," in *IEEE Signal Processing Letters*, vol. 32, pp. 961-965, 2025.

[3] S. Chang, L. Zhou, W. Liu, H. Zhu, X. Hu and L. Yang, "Combating Voice Spoofing Attacks on Wearables via Speech Movement Sequences," in *IEEE Transactions on Dependable and Secure Computing*, vol. 22, no. 1, pp. 819-832, 2025.

[4] Q. Zhang, X. Zhang, M. Sun and J. Yang, "A Soft-Contrastive Pseudo Learning Approach Toward Open-World Forged Speech Attribution," in *IEEE Transactions on Information Forensics and Security*, vol. 20, pp. 1135-1148, 2025.

[5] S. Dosho, L. Minati, K. Maari, S. Ohkubo and H. Ito, "A Compact 0.9µ W Direct-Conversion Frequency Analyzer for Speech Recognition With Wide- Range Q-Controllable Bandpass Rectifier," in *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 33, no. 2, pp. 315-325, 2025.

[6] S. Dhar, N. D. Jana and S. Das, "GLGAN-VC: A Guided Loss-Based Generative Adversarial Network for Many-to-Many Voice Conversion," in *IEEE Transactions on Neural Networks and Learning Systems*, vol. 36, no. 1, pp. 1813-1826, 2025.

[7] K. Zhang, Z. Hua, Y. Zhang, Y. Guo and T. Xiang, "Robust AI-Synthesized Speech Detection Using Feature Decomposition Learning and Synthesizer Feature Augmentation," in *IEEE Transactions on Information Forensics and Security*, vol. 20, pp. 871-885, 2025.

[8] W. Huang, W. Tang, H. Jiang and Y. Zhang, "Recognizing Voice Spoofing Attacks via Acoustic Nonlinearity Dissection for Mobile Devices," in *IEEE Transactions on Mobile Computing*, vol. 23, no. 12, pp. 12080-12096, 2024.

[9] X. Liu, M. Sahidullah, K. A. Lee and T. Kinnunen, "Generalizing Speaker Verification for Spoof Awareness in the Embedding Space," in *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 32, pp. 1261-1273, 2024.

[10] K. Li, C. Baird and D. Lin, "Defend Data Poisoning Attacks on Voice Authentication," in *IEEE Transactions on Dependable and Secure Computing*, vol. 21, no. 4, pp. 1754-1769, 2024.

[11] Susee, S. K., M. Senthil Kumar, and B. Chidambararajan. "A novel homomorphic encryption-based optimization framework for wireless sensor networks." Microsystem Technologies (2025): 1-21

[12] R. He, Y. Cheng, Z. Zheng, X. Ji and W. Xu, "Fast and Lightweight Voice Replay Attack Detection via Time-Frequency Spectrum Difference," in *IEEE Internet of Things Journal*, vol. 11, no. 18, pp. 29798-29810, 2024.

[13] F. Han, P. Yang, H. Du and X. -Y. Li, "Accuth++: Accelerometer-Based Anti-Spoofing Voice Authentication on Wrist-Worn Wearables," in *IEEE Transactions on Mobile Computing*, vol. 23, no. 5, pp. 5571-5588, 2024.

[14] Xinxin Hu,Haotian Chen,Shuxin Liu,Xing Li,Shibo Zhang.., "Cost-Sensitive GNN-Based Imbalanced Learning for Mobile Social Network Fraud Detection," in *IEEE Transactions on Computational Social Systems*, vol. 11, no. 2, pp. 2675-2690, 2024.

[15] Michele Panariello,Natalia Tomashenko,Xin Wang, XiaoXiao Miao,Pierre Champion,Hubert Nourte., "The VoicePrivacy 2022 Challenge: Progress and Perspectives in Voice Anonymisation," in *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 32, pp. 3477-3491, 2024.

[16] L. Zhang, X. Wang, E. Cooper, N. Evans and J. Yamagishi, "The PartialSpoof Database and Countermeasures for the Detection of Short Fake Speech Segments Embedded in an Utterance," in *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 31, pp. 813-825, 2023.

[17] Y. Zhang, Z. Li, J. Lu, H. Hua, W. Wang and P. Zhang, "The Impact of Silence on Speech Anti-Spoofing," in *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 31, pp. 3374-3389, 2023.

[18] S. -I. Ng, C. W. -Y. Ng, J. Wang and T. Lee, "Automatic Detection of Speech Sound Disorder in Cantonese-Speaking Pre-School Children," in *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 32, pp. 4355-4368, 2024.

[19] Y. Ren, H. Peng, L. Li, X. Xue, Y. Lan and Y. Yang, "Generalized Voice Spoofing Detection via Integral Knowledge Amalgamation," in *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 31, pp. 2461-2475, 2023.