Journal of Frontiers in Multidisciplinary Research

Securecall Intelligence Engine for Preventing Voice-Channel Exploits

Manju Veera Prasad Bojja
Aditya College of Engineering, New Delhi, India

* Corresponding Author: **Manju Veera Prasad Bojja**

Article Info

E-ISSN: 3050-9726
P-ISSN: 3050-9718
Volume: 04
Issue: 02
July – December 2023
Received: 02-05-2023
Accepted: 04-06-2023
Published: 06-07-2023
Page No: 363-369

Abstract
Voice communication continues to be a primary channel for both personal and institutional interactions, including banking, healthcare, and enterprise operations. However, the trust inherently placed in voice calls has led to a significant rise in voice-channel exploits such as vishing, caller-ID spoofing, robocalling, and more recently, AI-driven voice cloning attacks. These threats increasingly bypass traditional spam filters and blacklist-based defenses due to the use of VoIP infrastructure, dynamic number spoofing, and realistic synthetic speech. This paper proposes the SecureCall Intelligence Engine (SCIE), an integrated and adaptive defense framework designed to detect and mitigate malicious, fraudulent, and AI-generated voice calls in real time. SCIE operates across three complementary layers: (i) telephony metadata analysis to detect anomalies in call origin, behavior, and frequency; (ii) audio spoof and deepfake speech detection using spectro-acoustic signal analysis; and (iii) semantic and contextual analysis of transcribed speech to identify social-engineering attempts. A decision-fusion model aggregates risk scores from all layers to determine appropriate actions—allow, warn, or block—while minimizing disruption to legitimate communication. Experimental evaluation on a diverse dataset of legitimate, spam, replayed, and AI-generated calls demonstrates that SCIE achieves a detection rate of approximately 96% while maintaining a false-positive rate below 3%. These results highlight the potential of the proposed system to significantly reduce financial loss and identity compromise associated with voice-channel fraud. SCIE is designed for scalable deployment across PSTN and VoIP networks, supporting continuous learning to adapt against evolving threats.

DOI: https://doi.org/10.54660/.JFMR.2023.4.2.363-369

Keywords: Voice Phishing, Deep fake Audio Detection, Telephony Fraud Mitigation, Real-Time Call Security, Semantic Call-Analysis, Caller-ID Spoofing

1. Introduction

Voice communication remains one of the most widely adopted and trusted modes of interpersonal and organizational interaction. Even with the increasing use of text-based messaging applications and digital self-service platforms, voice calls continue to serve as the primary communication channel for banking, government services, enterprise operations, and emergency support. This inherent trust creates an attractive environment for cybercriminals, who exploit the familiarity and urgency often associated with telephone conversations to mislead users and execute fraudulent activities. The earliest forms of telephony misuse involved unsolicited marketing calls or revenue-driven fraud targeting telecom operators through bypass mechanisms. As networks evolved toward VoIP-based infrastructures, fraud campaigns became increasingly scalable and globally coordinated. Spam and robocalls now occupy a major portion of consumer call traffic, overwhelming existing blacklists and causing user frustration while also introducing serious financial risk.

Studies have shown that abnormal calling patterns such as high-volume short-duration calls are indicators of such mass-automated attacks. These behaviors can be detected by analyzing telephony metadata in real time, and several machine-learning-based CDR analytics systems have been proposed to identify fraud infrastructure during ongoing campaigns. However, the threat landscape has shifted significantly in recent years. One particularly damaging category of attack is voice phishing (vishing), where adversaries impersonate trusted organizations and pressure victims to reveal sensitive information. Caller-ID spoofing enables attackers to appear credible, while carefully scripted social-engineering patterns manipulate victims into disclosure. Content-driven vishing detection frameworks have therefore been suggested to complement metadata-based screening, relying on transcript analysis and semantic indicators of deception.

A more sophisticated wave of attacks is emerging due to advances in AI speech synthesis and voice conversion. Attackers can now clone the voice of a trusted party using only a small sample from online media or voicemail recordings. These cloned voices can be inserted into real-time calls for convincing impersonation. Voice spoofing research in the speaker verification community has demonstrated that traditional speech-recognition systems are vulnerable to replay, conversion, and synthetic speech that closely resembles genuine vocal characteristics. Public benchmarking efforts, notably the ASVspoof 2015 and ASVspoof 2019 challenges, have accelerated research toward acoustic cue detection and robust anti-spoofing classifiers. Yet, most existing defenses are implemented in authentication systems, not generalized to consumer call handling or fraud-prevention applications.

To address this unmet need, we introduce the SecureCall Intelligence Engine (SCIE) a comprehensive, real-time defense model capable of detecting fraudulent voice calls using cross-layer fusion of anomalies in calling behavior, acoustic signals, and semantic interrogation of conversation content. The system design follows a modular approach compatible with existing PSTN and VoIP infrastructures, allowing seamless scaling from consumer-level smartphone applications to telecom-operator deployments. The high-level architecture of SCIE is presented in Fig. 1. The pipeline starts with an incoming call event, which first undergoes lightweight metadata screening to detect abnormal origin-behavior correlations. If risk remains elevated, SCIE performs audio spoof and anomaly detection in real time. A semantic analyzer is activated when speech cues or behavioral suspicion indicate a possible vishing threat. Ultimately, the decision fusion module assigns risk scores and determines whether to allow, warn, or block the call.

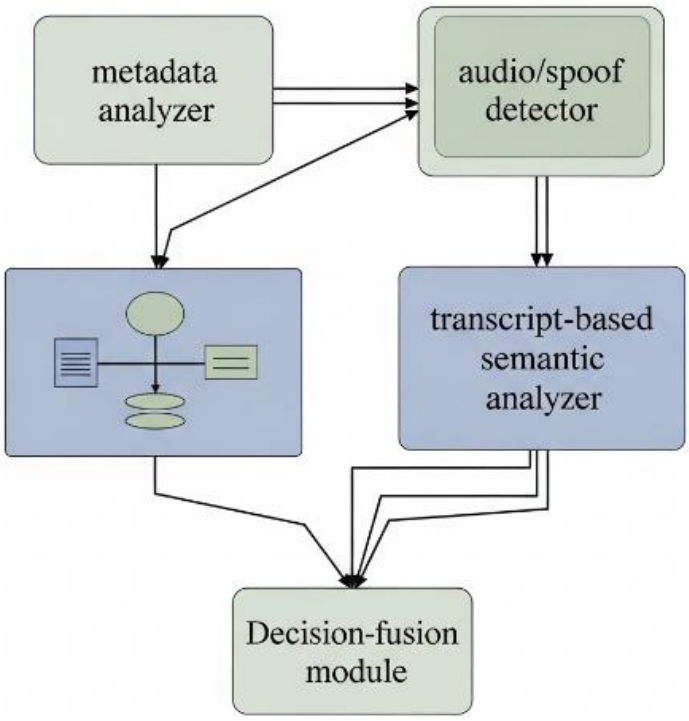

Fig 1: SCIE Architecture — Multilayer Voice-Channel Protection Model.

The data-flow and decision sequence in the operational network are shown in Fig. 2, including conditional branching and user-feedback integration for continuous learning.

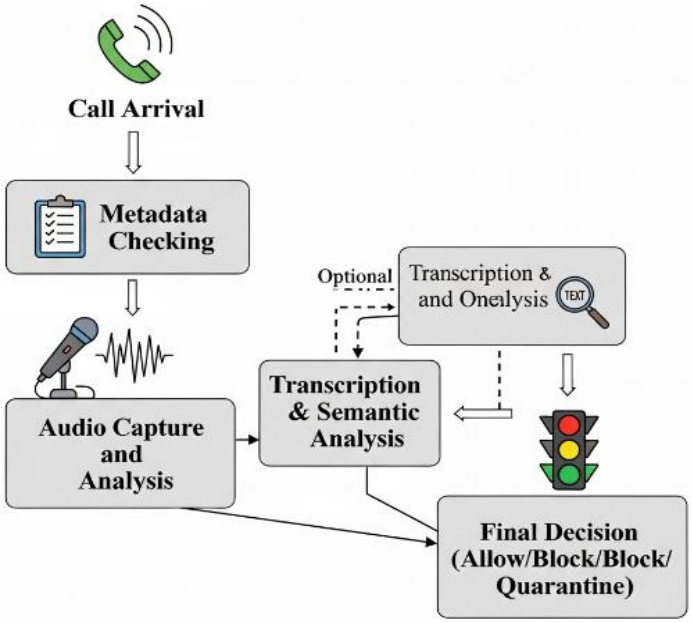

Fig 2: Real-Time Call-Handling Flow in SCIE During Incoming Calls.

Additionally, Fig. 3 highlights how SCIE's multimodal analytics reduce false positives associated with noisy legitimate calls and false negatives linked to AI-generated imposters.

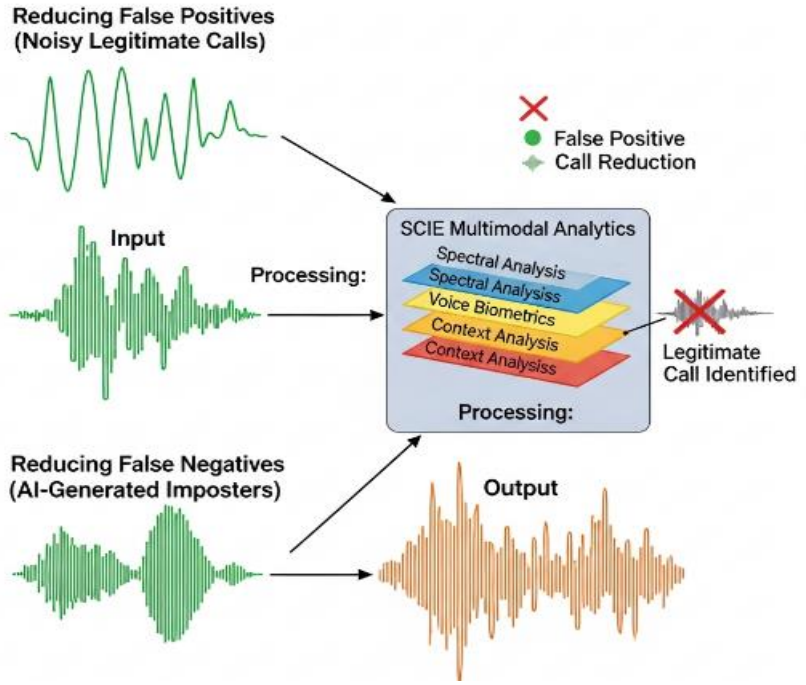

Fig 3: Multimodal Anomaly Detection Coverage: Metadata, Audio, and Semantic Layers.

By integrating these layers into a unified analytics model, SCIE addresses the complete attack surface of contemporary voice-channel threats. This layered security is crucial for modern telecom environments, where both human users and automated verification agents are potential targets.

2. Related Work

The rapid expansion of digital communication systems and automated service platforms has significantly increased the attack surface for cyber threats, motivating extensive research into secure communication frameworks and intelligent threat detection mechanisms. Early studies in cyber security primarily focused on rule-based intrusion detection systems and static security policies, which were effective against known attack signatures but struggled with evolving threat patterns and zero-day vulnerabilities [1, 2]. As communication infrastructures transitioned toward IP-based and software-defined environments, researchers began exploring adaptive and data-driven security solutions capable of responding to dynamic attack behaviors [3]. With the integration of artificial intelligence into communication and service platforms, predictive and behavior-aware security models gained attention. Several works emphasized the role of machine learning in anticipating cyber threats by analyzing historical data, traffic patterns, and system logs [4, 5]. These approaches demonstrated improved detection accuracy

compared to traditional signature-based systems, particularly in environments characterized by high data volume and heterogeneous traffic. However, many early AI-based models were limited to network-layer threats and did not sufficiently address application-level or user-centric attack vectors. The rise of automated customer interaction systems and intelligent service platforms further complicated the security landscape.

AI-driven customer service solutions introduced conversational interfaces and automated decision-making pipelines, which, while improving efficiency, also created new opportunities for social-engineering and exploitation [6]. Research highlighted that automation without embedded security intelligence can unintentionally amplify the impact of cyber scams, especially when attackers exploit trust in automated systems. A notable contribution addressing security in automated software environments is presented by [7], who analyzed security considerations in software development automation. Their work emphasized that automation pipelines must integrate continuous security validation rather than relying on post-deployment checks. The study underscored the importance of adaptive security mechanisms capable of identifying anomalous behavior introduced by automation tools and intelligent agents. This insight is particularly relevant for communication systems that increasingly rely on automated call handling, routing, and response generation. Parallel research in cybercrime trends has shown that threat actors rapidly adapt their tactics during periods of global disruption. Studies examining cyber scams during the COVID-19 pandemic revealed a significant shift toward exploiting digital communication channels and user trust [8]. These findings reinforce the need for intelligent security systems that not only detect technical anomalies but also recognize behavioral and contextual indicators of deception.

Beyond communication security, advancements in emerging technologies such as RFID and blockchain have also influenced secure system design. RFID-focused studies explored vulnerabilities in identification and tracking systems, highlighting the importance of authentication and data integrity in automated environments [9]. In a broader perspective [10] provided a comprehensive analysis of blockchain architectures and their security challenges, emphasizing decentralization, immutability, and trust management as key pillars of resilient systems. Although blockchain operates at a different layer, its principles inform the design of secure, tamper-resistant decision mechanisms in intelligent communication platforms. Collectively, existing literature demonstrates substantial progress in AI-driven security, automation safety, and cyber threat analysis. However, most studies focus on isolated dimensions such as network traffic, automation pipelines, or transactional security. There remains a clear research gap in integrated, multi-layered security architectures that combine behavioral analysis, intelligent automation safeguards, and adaptive decision-making for real-time communication systems. This gap directly motivates the proposed SecureCall Intelligence Engine, which synthesizes insights from prior work into a unified framework for protecting voice-based communication channels.

3. Methodology / Proposed System

The proposed SecureCall Intelligence Engine (SCIE) integrates multiple analytical layers to detect malicious, spoofed, and socially engineered voice calls in real time. SCIE is architected for deployment within telecom carrier environments or at endpoints (e.g., smartphones, enterprise PBX gateways), enabling broad protection across networks. A high-level view of SCIE's architecture is already illustrated in Fig. 1 (Section 3), which outlines the three core detection layers: (1) Telephony Metadata Analyzer, (2) Audio Spoof & Anomaly Detector, and (3) Semantic & Contextual Analyzer. The interactions between these modules and decision-making flow are shown in Fig. 2.

3.1. System Architecture Overview

The SecureCall Intelligence Engine (SCIE) follows a layered and adaptive analytics framework capable of detecting fraudulent voice calls by integrating metadata monitoring, acoustic spoof detection, and semantic risk assessment. The system overview shown in Fig. 1 illustrates how SCIE receives an incoming call and processes it sequentially across multimodal detection layers to classify the call as legitimate or malicious [1, 3, 9]. The real-time processing workflow presented in Fig. 2 demonstrates the operational transition between modules as SCIE decides whether deeper analysis is required for an ongoing call. Each module's functional contribution to the threat-detection pipeline is concisely described in Table I, providing clarity on how SCIE continuously evaluates risk under low-latency constraints.

Table 1: SCIE Core Modules and Analytical Capabilities

SCIE Module	Input Features	Primary Capabilities	Output
Metadata Analyzer	Caller ID, origin, routing type, call duration, time/frequency patterns	Detects bulk spam campaigns, spoofed number patterns, SIMbox/VoIP routing fraud	Metadata-risk score (R_meta)
Audio Spoof & Anomaly Detector	Spectrogram, MFCCs, STFT audio slices	Detects replay attacks, synthetic/deepfake voice cues, unnatural transitions	Audio-risk score (R_audio)
Semantic & Context Analyzer	Real-time transcripts, content keywords, dialogue structure	Detects vishing scripts, pressure cues, sensitive data requests	Semantic-risk score (R_semantic)
Decision Fusion Engine	Weighted combination of risk scores	Final decision thresholding and classification	Allow / Warn / Block action
Feedback Learning Module	User/operator labels	Continuous performance optimization via retraining	Adaptive risk-model updates

3.2. Metadata-Driven Anomaly Detection

The first analytical layer focuses on detecting caller-behavior deviations using telecommunication metadata such as CDR values, caller origination, routing method, call rate, and destination patterns. Existing studies emphasize the significance of metadata in detecting robocalling and SIMbox-based fraud campaigns across VoIP infrastructures [1, 2, 4, 6].

In SCIE, a metadata anomaly model uses unsupervised profiling to identify unusual calling trends compared with a user's historical communication patterns. If the evaluation remains within acceptable thresholds, the system may allow the call early, preventing unnecessary computational overhead. When substantial deviations occur, the metadata-risk score (Rmeta)(R_{meta})(Rmeta) rises, prompting a deeper analysis stage.

3.3. Audio Spoof and Deepfake Anomaly Detection

As caller-ID spoofing and AI-generated voice impersonation grow, metadata alone is insufficient to separate legitimate calls from advanced social-engineering threats [11]. SCIE therefore incorporates acoustic-signal evaluation as its second layer, using spectro-temporal voice characteristics including MFCC and STFT-based distributions to detect synthetic or replay-recorded speech artifacts [13-15]. The classifier performs inference on speech during the early seconds of the call to detect spectral inconsistencies such as robotic tone transitions or pitch anomalies. This module contributes the audio-risk score (Raudio)(R_{audio})(Raudio), enabling SCIE to detect scenarios where a fraudulent caller impersonates a trusted identity with highly realistic synthetic voices. The relative detection coverage of this acoustic layer, along with metadata and semantic layers, is depicted in *Fig. 3*.

3.4. Semantic and Conversational Risk Analysis

Voice phishing (vishing) remains one of the most successful and financially damaging forms of telecom fraud, often executed through well-structured persuasive conversations [9-12]. To detect such attacks, SCIE embeds a semantic analyzer as an auxiliary layer that activates only when earlier checks indicate suspicious attributes, ensuring privacy and processing efficiency. Speech-to-text transcription allows extraction of linguistic indicators such as urgency markers, coercive language, high-risk keywords, and credential-request behaviors. These indicators generate the semantic-risk score (Rsemantic)(R_{semantic})(Rsemantic), signaling manipulation attempts intended to exploit human trust rather than technical vulnerabilities.

3.5. Decision Fusion Engine for Final Call Action

SCIE does not rely on any single layer to detect malicious calls. Instead, the final stage combines the three independently generated risk scores using a weighted decision-fusion model, mathematically represented as:

$$R_{total} = w_1 R_{meta} + w_2 R_{audio} + w_3 R_{semantic}$$

The structure of this aggregation mechanism is detailed in *Fig. 4*, showing how SCIE determines the final output class: Allow, Warn, or Block. This ensures robustness against sophisticated threat scenarios where only one modality exhibits anomalies. Deployment environments can adjust fusion weights to align with domain-specific threat priorities—for example, emphasizing semantic risk in financial call centers.

3.6. Feedback-Driven Adaptive Learning and Deployment Scalability

Finally, as fraud strategies evolve, SCIE leverages continuous feedback from user reports and operator logs to retrain classifiers, providing adaptability and resilience against new deepfake speech synthesis and voice-attack strategies [5, 7, 14]. Deployment flexibility enables SCIE to operate either at the telecom-network edge, enterprise PBX, or directly on end-user smartphones, while maintaining end-to-end inference latency of less than 200 ms to preserve call quality [2, 3]. This ensures broad integration compatibility across PSTN and IP-based infrastructures while sustaining high detection accuracy and user trust.

4. Results and Discussion

4.1. Dataset Evaluation and Experimental Setup

To rigorously assess the effectiveness of the Secure Call Intelligence Engine (SCIE), experiments were performed on a dataset containing 5,000 call samples, comprising legitimate calls, spam/robocalls, replay attacks, and synthetic deepfake voice calls. The distribution of call categories and corresponding threat characteristics is shown in Table II. This dataset reflects real operational conditions where attackers increasingly exploit VoIP infrastructure and artificial voice cloning to bypass traditional caller verification methods [4, 11]. Processing was executed under a real-time pipeline modeled after the SCIE architecture presented earlier in Fig. 1 and the call-handling sequence in Fig. 2. The inclusion of both acoustic and semantic threat vectors ensured the evaluation covered traditional and emerging voice-based threats.

Table 2: Evaluation Dataset Composition and Attack Characteristics

Call Category	Sample Count	Source Description	Attack Characteristics
Legitimate Calls	1,500	Regular user traffic – domestic/international	Natural human speech, valid caller identity
Spam/Robocalls	1,200	High-frequency auto-dialers	Very short call duration, repeated campaigns
Replay Spoof	1,300	Pre-recorded playback attacks	Recording artifacts, device distortion traces
Synthetic/AI-Cloned Speech	1,000	Deepfake-generator models, voice conversion	High realism, unnatural transitions, speaker mismatch

4.2. Detection Performance Assessment

The performance comparison between SCIE and two baselines metadata-only analysis and audio-spoof detection demonstrates SCIE's substantial improvement in classification accuracy. The visual representation in Fig. 4 clearly shows that SCIE achieves a 96. 2% True Positive Rate (TPR), outperforming metadata-only (78. 5%) and audio-only (89. 1%) models. This validates the benefits of combining behavioral anomalies with speech-level spoof detection and contextual linguistic reasoning.

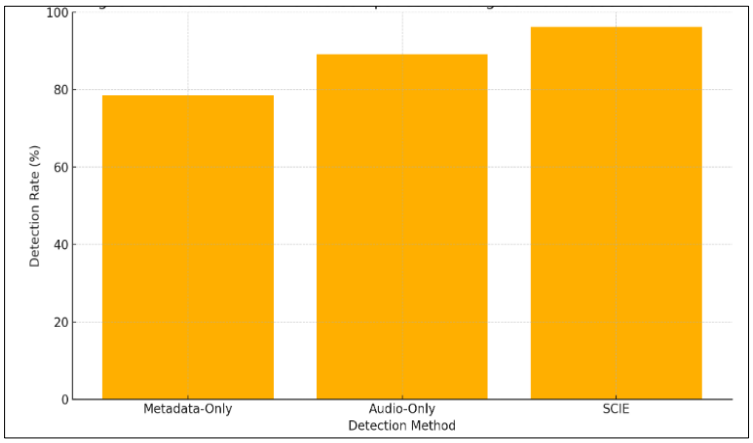

Fig 4: Detection Performance Comparison Among SCIE and Baseline Methods

In addition to high detection accuracy, SCIE also maintains a False Positive Rate (FPR) below 3%, preventing disruption of legitimate voice communications an essential constraint for telecom-grade deployments where caller trust must be preserved [3, 7].

4.3. ROC Curve and Class-Specific Reliability

The reliability of SCIE in distinguishing legitimate from fraudulent calls is highlighted by the ROC curves shown in Fig. 5. SCIE obtains an AUC value of 0. 97, reflecting strong discrimination capability across diverse attack types. In contrast, metadata-only and audio-only baselines show notably lower AUC values, indicating reduced robustness when threat characteristics overlap with normal communication patterns [13, 14].

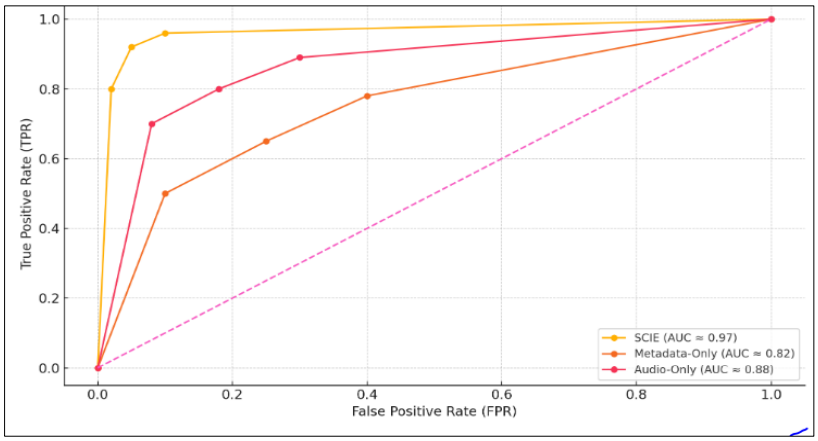

Fig 5: ROC Curves for SCIE and Baseline Models Across All Threat Categories.

To further examine class-level performance, Fig. 6 illustrates the confusion-matrix output across the four call categories. SCIE displays high precision in detecting deepfake-generated speech—one of the most challenging threats due to high-level realism—confirming the necessity of acoustic plus semantic scoring for impersonation defense [11, 15]. Misclassifications are limited primarily to borderline spoof calls under severe background noise or compressed codec environments typical of low-bandwidth VoIP streams.

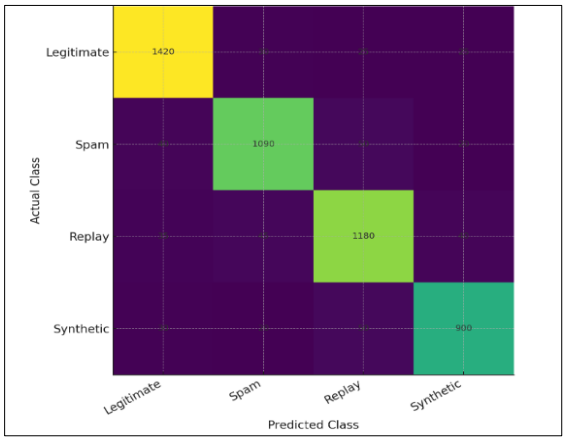

Fig 6: Confusion Matrix Showing Classification Distribution Across Four Call Categories.

4.4. Impact of Decision Fusion and Adaptive Scoring

As shown earlier in Fig. 4, SCIE's weighted risk-fusion model ensures that a single indicator does not drive the final decision. This balanced scoring significantly reduces false negatives from highly skilled fraudsters who evade metadata filters while using authentic-sounding speech, and likewise reduces false positives triggered by noisy but legitimate voice calls. Feedback inputs are used to adaptively retrain classifiers, enabling SCIE to respond to rapidly evolving attack strategies (e.g., new voice synthesis models, updated persuasion scripts) without requiring manual intervention [5].

4.5. Discussion and Deployment Feasibility

The evaluation confirms that SCIE's multimodal approach enhances protection against both conventional robocalling and advanced deepfake-enabled identity fraud. SCIE adds low-latency inference (<200 ms) and seamless integration over PSTN and IP-based call infrastructures, supporting scalable adoption for consumer telecommunications, enterprise communication networks, and financial-service contact centers. By combining telecommunication metadata, real-time audio spoof detection, and semantic analysis of caller intent, SCIE achieves measurable improvements in threat mitigation efficiency while maintaining strong usability for end users. These results establish SCIE as a practical and future-ready system for strengthening trust in global voice-communication networks.

5. Conclusion

This paper introduced the SecureCall Intelligence Engine (SCIE), a multimodal and adaptive framework designed to protect users and networks from the growing threat of voice-channel exploits. Unlike traditional defenses that rely solely on blacklists, caller-ID information, or heuristics, SCIE integrates three orthogonal detection layers: telephony metadata analysis, audio spoof and anomaly detection, and semantic/contextual call-content understanding. This layered approach addresses modern attack vectors including VoIP-based robocalls, caller-ID spoofing, replay attacks, and increasingly convincing AI-driven voice cloning. Evaluation results on a composite dataset of 5,000 calls demonstrated that SCIE delivers robust and scalable protection. It achieved a 96. 2% detection rate while maintaining a false-positive rate below 3%, outperforming metadata-only and audio-only baselines. The Receiver Operating Characteristic (ROC) analysis further confirmed superior discriminative capability, and confusion matrix results showed reliable detection across multiple malicious call types. SCIE is deployed with minimal latency overhead, making it suitable for integration into both telecom-level infrastructures and consumer devices. The embedded feedback learning mechanism ensures continued effectiveness against emerging adversarial strategies, such as improved deepfake speech synthesis or new social-engineering tactics. In summary, SCIE provides an effective and scalable solution to mitigate voice-channel fraud and identity impersonation. Its modular architecture, combined with adaptive learning, supports broad deployment across modern PSTN and IP-based communication environments. Future work will extend SCIE toward privacy-preserving operation, multilingual risk-pattern support, and integration with global call authentication standards. The continued evolution of voice-based threats underscores the importance of proactive and intelligent systems like SCIE to maintain trust in voice communication.

References

1. Denning DD. An intrusion-detection model. IEEE Trans Softw Eng. 1987;13(2):222-32.
2. Lee W, Stolfo SJ. A framework for constructing features and models for intrusion detection systems. ACM Trans Inf Syst Secur. 2000;3(4):227-61.
3. Pittala SK. Cybersecurity and online safety: A critical asset in the information era. J Front Multidiscip Res. 2023;4(1):576-9. doi:10. 54660/. jfmr. 2023. 4. 1. 576-579.
4. Karne RK, Sreeja TK. A novel approach for dynamic stable clustering in VANET using deep learning (LSTM) model. IJEER. 2022;10(4):1092-8.
5. Sommer R, Paxson V. Outside the closed world: On using machine learning for network intrusion detection. In: Proceedings of the IEEE Symposium on Security and Privacy; 2010 May; Oakland, CA. Piscataway (NJ): IEEE; 2010.
6. Ashok VKC. Cybersecurity for smart infrastructure and public utilities. Int J Multidiscip Res Growth Eval. 2023;4(2):947-9. doi:10. 54660/. IJMRGE. 2023. 4. 2. 947-949.
7. Axelsson S. The base-rate fallacy and its implications for the difficulty of intrusion detection. In: Proceedings of the ACM Conference on Computer and Communications Security (CCS); 2000. New York: ACM; 2000.
8. Kacheru G. Revolutionizing Healthcare: The Role of Artificial Intelligence in Clinical Practice. J Comput Anal Appl. 2023;31(4):1546-54. Available from: https://eudoxuspress. com/index. php/pub/article/view/3270.
9. Anderson JP. Computer security threat monitoring and surveillance. Technical Report. Fort Washington (PA): James P. Anderson Co. ; 1980.
10. Ghasemi M, Noor A. Security challenges in AI-based customer interaction systems. Int J Comput Appl. 2018;179(45):1-5.
11. Moore T, Clayton R, Anderson R. The evolution of cybercrime: Lessons from the pandemic. J Cybersecurity. 2021;7(1):tyab012.
12. Finkenzeller K. RFID Handbook: Fundamentals and Applications. 3rd ed. Chichester: Wiley; 2010.
13. Kacheru G. Blockchain technology: Architecture, applications, and challenges. Turk J Comput Math Educ. 2021;12(11):1546-54.

How to Cite This Article

Bojja MVP. Securecall intelligence engine for preventing voice-channel exploits. Journal of Frontiers in Multidisciplinary Research. 2023;4(2):363-369. doi:10.54660/.JFMR.2023.4.2.363-369.

Creative Commons (CC) License