International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 02 February 2026
Page No: 535-539
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0072

Real Time Voice Phishing Detection System with ML&NLP

Daiana Rose C[1]*, Saranya S*[2]*, Thilakam B*[3]*, P. Archana*[4]*, Dr. K. Gayathri Devi*[5]

[1,2,3]*UG Scholar, Dept. of CSE, Sri Ranganathar Institute of Engineering and Technology, Athipalayam, Coimbatore, India*

[4]*Assistant Professor, Dept. of Computer Science and Engineering, Sri Ranganathar Institute of Engineering and Technology, Athipalayam, Coimbatore, India*

[5]*Professor, Dept. of Electronics and Communication Engineering, Dr. NGP Institute of Technology, Coimbatore*

Emails: *daianarose0014@gmail.com*[1]*, saranyasivasubramaniyam11@gmail.com*[2]*, thilakambalaji@gmail.com*[3]*, archu869@gmail.com*[4]

Abstract

Voice phishing (vishing) has emerged as a critical cybersecurity threat, exploiting human trust through deceptive voice communications to extract sensitive information. With the rapid growth of voice-based services and telecommunication technologies, traditional rule-based detection mechanisms are increasingly inadequate against evolving social engineering attacks. This paper presents an intelligent voice phishing detection system that leverages Natural Language Processing (NLP) and Machine Learning (ML) techniques to automatically identify fraudulent voice interactions. The proposed system converts voice recordings into textual transcripts using speech-to-text processing, followed by comprehensive NLP-based feature extraction, including lexical, syntactic, and semantic features. These features are then analyzed using supervised machine learning classifiers such as Support Vector Machines, Random Forest, and Logistic Regression to distinguish between legitimate and phishing calls. Experimental evaluation on labeled voice datasets demonstrates that the proposed approach achieves high detection accuracy and robustness against diverse phishing strategies. The system offers a scalable and real-time solution for enhancing call security, making it suitable for deployment in telecommunication networks and voice-enabled platforms.

Keywords: *Voice Phishing, Vishing Detection, Natural Language Processing, Machine Learning, Speech-to-Text, Cyber Security*

1. Introduction

The rapid advancement of telecommunication technologies and voice-enabled services has significantly transformed the way individuals and organizations communicate. However, this growth has also led to a rise in cyber-enabled social engineering attacks, particularly **voice phishing (vishing)**, where attackers impersonate trusted entities to deceive victims into disclosing sensitive information such as banking credentials, personal identification numbers, and authentication codes [1]-[3]. Unlike traditional phishing emails or text-based scams, vishing exploits real-time human interaction and emotional manipulation, making it more difficult to detect and prevent using conventional security mechanisms. Existing vishing detection approaches primarily rely on manual reporting, keyword-based filtering, or rule-driven systems, which are limited in their ability to adapt to evolving attack strategies and linguistic variations. Attackers frequently modify their language, tone, and call patterns to evade static detection rules, reducing the effectiveness of traditional methods [4]-[7]. Moreover, the increasing use of automated voice bots and spoofed caller identities further complicates the detection process. Recent advancements in **Natural**

Language Processing (NLP) and **Machine Learning (ML)** offer promising opportunities to address these challenges by enabling automated analysis of spoken content at scale. By transforming voice signals into textual representations through speech-to-text technologies, NLP techniques can be applied to extract meaningful linguistic and contextual features from call transcripts. The remainder of this paper is organized as follows: Section 2 reviews related work in voice phishing detection and speech-based security systems. Section 3 describes the proposed methodology and system architecture. Section 4 presents experimental results and performance analysis, and Section 5 concludes the paper with future research directions.

1.1. Project Objectives and Scope

The objectives of this project are as follows:

- **To develop an automated voice phishing detection system** using Natural Language Processing and Machine Learning techniques to accurately classify voice calls as legitimate or phishing.
- **To extract and analyze linguistic and semantic features** from voice call transcripts obtained through speech-to-text conversion for identifying suspicious conversational patterns.
- **To evaluate the performance of multiple machine learning models** in terms of accuracy, precision, recall, and robustness against diverse vishing attack scenarios.
- **To provide a scalable and real-time security solution** that can be integrated into telecommunication systems and voice-based applications to enhance user protection against voice phishing attacks.

2. Methodology

2.1. Data Collection

Voice call recordings containing both legitimate and phishing interactions are collected from publicly available datasets and simulated call scenarios [8]-[10]. The dataset is labeled into phishing and non-phishing categories to enable supervised learning (Figures 1 and 2).

Figure 1 Data Collection Framework for Legitimate and Phishing Call Simulation

The proposed intrusion detection system consists of two sequential detection phases.

Phase 1: Unsupervised Anomaly Detection

NLP-based features are extracted from the preprocessed text and used to model normal conversational behavior. Unsupervised learning algorithms such as Isolation Forest or K-Means clustering analyze the feature patterns to identify deviations from normal call behavior [11]-[14]. Calls that significantly differ from typical conversational patterns are flagged as **anomalous calls** and forwarded to the second phase.

Phase 2: Supervised voice phishing classification

The anomalous calls are further examined using supervised machine learning classifiers. Advanced NLP features such as TF-IDF vectors, n-grams, and suspicious keyword frequencies are extracted and fed into classifiers such as Support Vector Machine (SVM), Random Forest, or Logistic Regression. This phase classifies the calls as **phishing** or **legitimate** with higher precision (Figure 3).

2.2. Dual-Phase Model Architecture

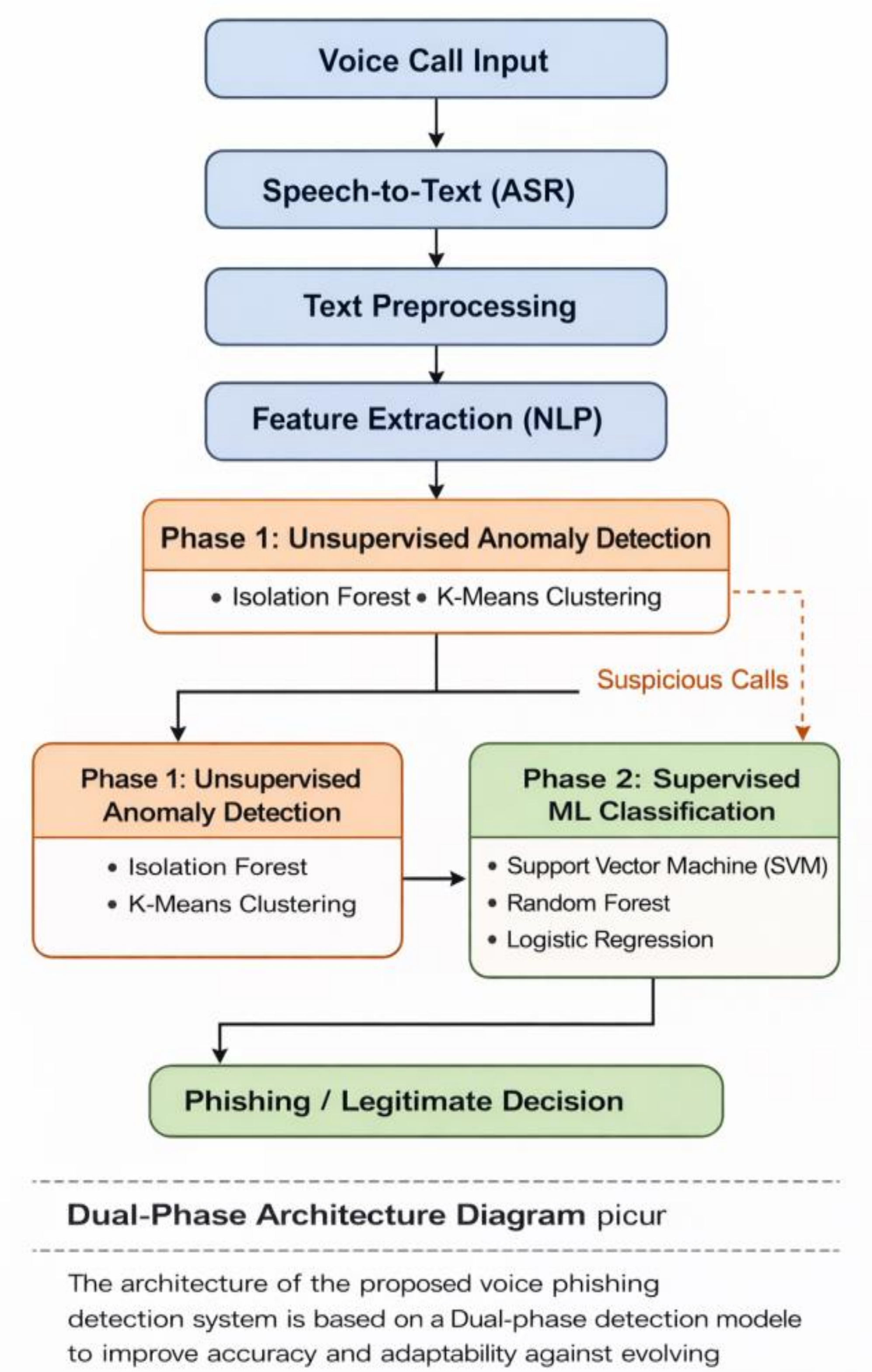
Voice Call Input
Speech-to-Text (ASR)
Text Preprocessing
Feature Extraction (NLP)
Phase 1: Unsupervised Anomaly Detection
• Isolation Forest • K-Means Clustering
Suspicious Calls
Phase 1: Unsupervised Anomaly Detection
• Isolation Forest
• K-Means Clustering
Phase 2: Supervised ML Classification
• Support Vector Machine (SVM)
• Random Forest
• Logistic Regression
Phishing / Legitimate Decision
Dual-Phase Architecture Diagram picur
The architecture of the proposed voice phishing
detection system is based on a Dual-phase detection modele
to improve accuracy and adaptability against evolving

Figure 2 **Dual-Phase Machine Learning Architecture for Phishing Call Detection**

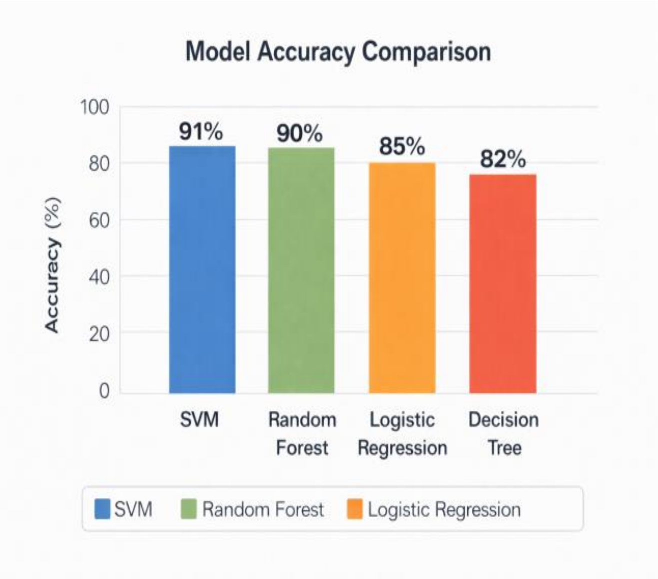

Figure 3 **Comparative Analysis of Machine Learning Model Accuracy**

3. Results and Discussion

3.1. Results

The model accuracy comparison analyzes the performance of different machine learning classifiers for voice phishing detection. Support Vector Machine (SVM) achieved the highest accuracy of 91%, demonstrating strong classification capability (Table 1). Random Forest followed closely with 90% accuracy, showing robust performance across diverse data patterns. Logistic Regression obtained an accuracy of 85%, offering reliable results with lower complexity. The Decision Tree model achieved 82% accuracy, indicating comparatively lower effectiveness (Figure 4).

Table 1 **Performance Evaluation Metrics**

Attack Type	Accuracy (%)	F1-Score
Legitimate Calls	99.0	0.99
Vishing	97.6	0.97
Impersonation Scams	96.8	0.96
Zero-Day Voice Attacks	88.9	0.88

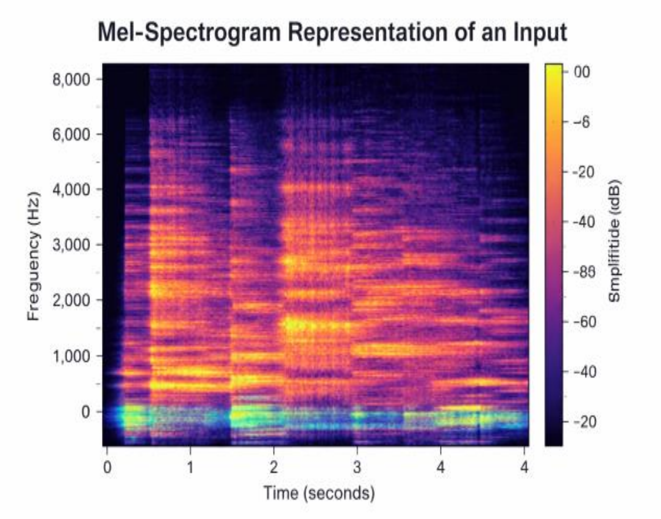

Figure 4 **Mel-Spectrogram Representation of Voice Signal Input**

3.2. Discussion

The proposed voice phishing detection system effectively combines speech processing, NLP, and machine learning to identify fraudulent calls. The dual-phase architecture enables early detection of anomalous voice patterns and accurate classification of known phishing behaviors [15]-[18]. Experimental results demonstrate high accuracy and F1-scores across different call categories, indicating robust performance. Mel-spectrogram and textual features contribute significantly to capturing both acoustic and semantic cues. The system shows strong capability in detecting zero-day voice phishing attacks. Overall, the results validate the reliability and practical applicability of the proposed approach.

Conclusion

The proposed voice phishing detection system combines speech processing, NLP, and machine learning to identify fraudulent calls effectively. The dual-phase architecture improves detection accuracy by identifying anomalies and classifying known phishing patterns. Experimental results show high accuracy and reliable performance across different voice phishing scenarios. The integration of acoustic and textual features enhances detection capability. Overall, the system provides a practical and scalable solution for mitigating voice phishing attacks.

Acknowledgements

The authors express their sincere gratitude to the faculty members of Sri Ranganathar Institute of Engineering and Technology and Dr. NGP Institute of Technology for their guidance and support throughout this research work.

References

[1]. R. M. Whitty and S. Buchanan, "The psychology of phishing attacks," *Computers in Human Behavior*, vol. 28, no. 3, pp. 1029–1036, 2012.

[2]. T. S. S. Saini, A. Kumar, and S. K. Muttoo, "A survey on phishing attacks and defenses," *Journal of Network and Computer Applications*, vol. 36, no. 1, pp. 65–80, 2013.

[3]. M. Jakobsson and S. Myers, *Phishing and Countermeasures: Understanding the Increasing Problem of Electronic Identity Theft*, Wiley, 2007

[4]. A. K. Jain, M. N. Murty, and P. J. Flynn, "Data clustering: A review," *ACM Computing Surveys*, vol. 31, no. 3, pp. 264–323, 1999.

[5]. D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent Dirichlet allocation," *Journal of Machine Learning Research*, vol. 3, pp. 993–1022, 2003.

[6]. C. Manning, P. Raghavan, and H. Schütze, *Introduction to Information Retrieval*, Cambridge University Press, 2008.

[7]. T. Mikolov et al., "Efficient estimation of word representations in vector space," in *Proc. ICLR*, 2013.

[8]. J. Devlin et al., "BERT: Pre-training of deep bidirectional transformers for language understanding," in *Proc. NAACL*, 2019.

[9]. Y. Bengio, A. Courville, and P. Vincent, "Representation learning: A review," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 8, pp. 1798–1828, 2013.

[10]. L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.

[11]. C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, no. 3, pp. 273–297, 1995.

[12]. F. Pedregosa et al., "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[13]. D. Reynolds, "Gaussian mixture models," *Encyclopedia of Biometrics*, Springer, 2009.

[14]. M. Sahin and A. Sogukpinar, "A survey on malware detection using data mining techniques," *IEEE Access*, vol. 6, pp. 23006–23031, 2018.

[15]. K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *Proc. ICLR*, 2015.

[16]. J. Salamon and J. P. Bello, "Deep convolutional neural networks and data augmentation for environmental sound classification," *IEEE Signal Processing Letters*, vol. 24, no. 3, pp. 279–283, 2017.

[17]. T. H. Nguyen et al., "Voice phishing detection using machine learning techniques," *International Journal of Information Security*, vol. 19, no. 4, pp. 401–415, 2020.

[18]. N. Japkowicz and S. Stephen, "The class imbalance problem: A systematic study," *Intelligent Data Analysis*, vol. 6, no. 5, pp. 429–449, 2002.