EdinBurg Peer Reviewed Journals and Books Publishers
Journal of Information and Technology
Vol. 5||Issue 6||pp 34-43||August||2025
Email: info@edinburgjournals.org||ISSN: 3080-9576

Identifying and Evaluating the Best Machine Learning Predictive Models for Detecting Voice (Phone-Call) Vishing Attacks on MoMo Users in Real Time

Asgedom Zerue Tlahun[1*], Djuma Sumbiri[2], Dr. KN Jonathan[3]
[123]Computing and Information Sciences, University of Lay Adventists of Kigali, Rwanda
Corresponding Author Emails: zerue092504@gmail.com;
sumbirdj@gmail.com;phialn1@gmail.com

Accepted: 08 June 2025 || Published: 20 August 2025

Abstract

Phishing, particularly voice-based phishing (vishing), has become a significant security threat, exploiting human trust and the widespread use of mobile communication. This paper aims to develop and evaluate a hybrid model that combines Gradient Boosting and Convolutional Neural Networks (CNNs) for detecting phishing calls in audio data. The hybrid model leverages the strengths of Gradient Boosting, a powerful classification technique, and CNNs, which excel at extracting features from raw audio signals. To assess the model's effectiveness, a dataset comprising 40 phishing audio files and 40 legitimate audio files was used. The audio data was converted into spectrograms for CNN training. Experimental results indicate that the hybrid approach outperforms individual models such as Gradient Boosting and CNNs, assessing performance based on precision, recall, accuracy, and ROC AUC. The model specifically achieved an accuracy of 70.83%, with 67% precision for phishing calls and 75% precision for legitimate calls. By combining traditional machine learning with deep learning, this study presents an innovative approach to phishing detection. The findings highlight the effectiveness of integrating advanced feature extraction methods with robust classification techniques to enhance security in mobile money platforms. The proposed hybrid model offers a promising solution for real-time vishing detection, with potential applications in securing financial transactions and improving fraud prevention mechanisms.

Keywords: *Machine Learning, Predictive Models, Phishing Detection, Voice Phishing (Vishing), Mobile Money (MoMo) Fraud, Real-Time Detection, Feature Engineering, Neural Networks (CNN, LSTM), Rwanda*

How to Cite: Tlahun, A. Z., Sumbiri, D., & Jonathan, K. N. (2025). Identifying and Evaluating the Best Machine Learning Predictive Models for Detecting Voice (Phone-Call) Vishing Attacks on MoMo Users in Real Time. *Journal of Information and Technology, 5*(6), 34-43.

1. Introduction

Mobile Money (MoMo) platforms have emerged as essential tools for financial transactions, especially in developing countries where access to traditional banking services remains limited. These platforms allow users to carry out various financial activities, such as money transfers, bill payments, and savings, directly from their mobile devices. However, with the increasing adoption of MoMo services, the risk of financial fraud has also escalated, particularly through voice phishing (vishing) attacks (Jones et al., 2021). Vishing attacks usually consist of scammers placing phone calls to unsuspecting individuals (Figueiredo et al., 2024). They often depend on pretexting and impersonating trusted organizations to deceive individuals into revealing confidential information (Yeboah-Boateng et al., 2014; Mouton et al., 2016; Ghafir et al., 2018). These attacks have serious and far-reaching consequences, leading to major financial losses, identity theft, breaches in corporate security, and a decline in trust toward digital communication channels. As these attacks become more sophisticated, there is an urgent need for effective detection and prevention mechanisms to protect MoMo users from financial fraud. This paper focused on identifying and evaluating the best machine learning (ML) predictive models for detecting voice phishing attacks on MoMo users in real-time. The study assessed the effectiveness of various ML models, including Random Forest, Gradient Boosting, Support Vector Machine (SVM), Convolutional Neural Networks (CNN), and Long Short-Term Memory (LSTM). These models were evaluated using key performance metrics, including Accuracy, Precision, Recall, F1-Score, and the Receiver Operating Characteristic - Area Under the Curve (ROC AUC) score. For this study, a dataset comprising 40 audio samples from vishing calls and 40 audio samples from legitimate calls was used to train and evaluate the models. The research aimed to determine the most effective approach for real-time detection, contributing to the development of robust, AI-powered security solutions for MoMo platforms.

2. Literature Review

2.1 Introduction to Voice Phishing

Voice phishing, also known as "vishing," is a form of social engineering attack in which attackers exploit voice communication to trick individuals into disclosing sensitive information or performing unauthorized actions, such as transferring money or revealing personal identification numbers (PINs). Unlike traditional phishing, which often uses email or text messages to deceive victims, vishing attacks capitalize on the trust and perceived legitimacy that individuals associate with voice calls, often impersonating trusted entities such as banks, mobile service providers, or government officials. The paper "On the Feasibility of Fully AI-automated Vishing Attacks" explores the potential escalation of vishing (voice phishing) attacks facilitated by advancements in artificial intelligence (Liu et al., 2021). Vishing is a tactic where attackers make phone calls while posing as legitimate organizations to trick individuals into revealing confidential information. To investigate this threat, the authors developed ViKing, an AI-powered vishing system utilizing publicly available AI technologies (Liu et al., 2021). ViKing utilizes a Large Language Model (LLM) as its central cognitive engine to guide conversations, supported by speech-to-text and text-to-speech components that enable seamless audio-to-text and text-to-audio conversion during phone calls. The success of voice phishing relies on manipulating human psychology, creating a sense of urgency, fear, or trust to prompt the victim to take immediate

action without verifying the legitimacy of the request. Attackers often employ sophisticated methods to appear credible, such as spoofing legitimate phone numbers or using common language that mimics official communication. In the context of Mobile Money (MoMo) platforms, vishing attacks have evolved in sophistication, moving from simple scams to complex social engineering schemes. Cybercriminals target MoMo users by exploiting their reliance on mobile devices for financial transactions. Attackers may deceive users into revealing their PIN codes, authorizing fraudulent transfers, or granting access to their accounts. Recent studies show that phone-based fraud, particularly vishing, has become one of the most prevalent methods of financial fraud across Africa (Alshehri et al., 2024). In cybersecurity, machine learning (ML) techniques have become increasingly important for their ability to identify complex patterns and anomalies more efficiently than conventional statistical approaches. ML-based solutions have demonstrated superior Predictive capabilities in solving classification problems, particularly in fraud detection**.** However, there is limited research on the application of machine learning techniques to develop predictive models specifically for detecting voice phishing attacks targeting MoMo users. Existing studies primarily focus on text-based phishing detection**,** leaving a significant research gap in the application of audio-based machine learning approaches for vishing prevention.

3. Methodology

The study adopted a mixed-methods approach to evaluate the effectiveness of machine learning models in detecting voice phishing (vishing) attacks on Mobile Money (MoMo) platforms. It combined quantitative analysis**,** using audio features like spectrograms and MFCCs to train the model, with qualitative insights gathered from user surveys and expert interviews. Audio data, including 40 phishing and 40 legitimate calls, was collected from open-source repositories and processed using tools like Librosa**,** Cloud Convert**,** Tensor Flow, and Scikit-learn**.** The dataset was split 70 for training and 30 for testing. Models were evaluated based on accuracy, precision, recall, and F1-score. The CNN-Gradient Boosting hybrid model emerged as the most effective in terms of both accuracy and real-time detection performance. The integration of qualitative findings helped contextualize the technical results, enabling the development of a practical and user-aligned vishing detection framework.

4. Results and discussion

Gradient Boosting and CNN emerged as the most effective models for phishing detection in voice calls, both achieving an ROC AUC score of 0.79, indicating strong discriminatory capabilities. Gradient Boosting demonstrated the highest classification accuracy at 70.83%, making it a robust choice for identifying phishing attempts. Although SVM recorded the highest accuracy (75%), it's extremely low ROC AUC score of 0.19 suggests that it struggles in ranking phishing and legitimate calls effectively. Random Forest displayed the weakest performance, achieving an accuracy of only 58.33%, with inconsistencies in precision and recall, making it unsuitable for this application. LSTM also performed poorly, with an ROC AUC score of 0.57, highlighting its difficulty in handling sequential voice data efficiently. Overall, CNN and Gradient Boosting stand out as the best candidates for real-time phishing detection, with potential for further improvement through hybrid approaches and enhanced feature extraction techniques.

EdinBurg Peer Reviewed Journals and Books Publishers
Journal of Information and Technology
Vol. 5||Issue 6||pp 34-43||August||2025
Email: info@edinburgjournals.org||ISSN: 3080-9576

4.1 Model Performance

After training and testing the models on the phishing dataset, the following results were observed as presented in Table 1:

Table 1: Model Performance Comparison

Model	Precision (vishing)	Recall (Vishing)	Precision (non_vishing)	Recall (non_vishing)	Total sample	Vishing(1) Samples	Non_vishing (0) Samples	Accuracy
RF	53%	73%	67%	46%	24	11	13	58.33%
GB	67%	73%	75%	69%	24	11	13	70.83%
SVM	73%	73%	77%	77%	24	11	13	75%
CNN	75%	27%	60%	92%	24	11	13	62%
LSTM	29%	18%	47%	62%	24	11	13	42%

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \times 100 \qquad \text{Accuracy} = \frac{14}{24} \times 100 = 58.33\%$$

$$\text{Accuracy} = \frac{8 + 6}{8+6+7+3} \times 100$$

This is how the **58.33% accuracy** for the **Random Forest model** was computed.

Precision Calculation Formula

Precision measures the proportion of correctly predicted positive observations (phishing or legitimate) out of the total predicted positive observations.

$$\text{Precision (vishing)} = \frac{\text{TP}}{\text{TP} + \text{FP}} \times 100 \qquad \text{Precision (legitimate)} = \frac{\text{TN}}{\text{TN} + \text{FN}} \times 100$$

$$\text{Precision (vishing)} = \frac{8}{8 + 7} \times 100 = 53.33\% \qquad \text{Precision (legitimate)} = \frac{6}{6+3} \times 100 = 66.67\%$$

Thus, for Random Forest,

Precision (Phishing) = 53% and Precision (Legitimate) = 67%

The results indicate that CNN and Gradient Boosting demonstrated the highest effectiveness in phishing detection, both achieving an ROC AUC score of 0.79.

- Gradient Boosting exhibited robust classification performance, achieving an accuracy of 70.83% and maintaining a balanced precision-recall tradeoff.
- CNN showed strong capabilities in feature extraction from speech signals, reinforcing its suitability for phishing call detection.

EdinBurg Peer Reviewed Journals and Books Publishers
Journal of Information and Technology
Vol. 5||Issue 6||pp 34-43||August||2025
Email: info@edinburgjournals.org||ISSN: 3080-9576

4.2 Confusion Matrix Results

The Confusion Matrix, as seen in Table 2**,** is a table that provides a detailed breakdown of a classification model's performance by showing the actual vs. predicted classifications. It consists of four key components:

Table 2: Confusion Matrix

Actual / Predicted	Predicted: Non-Phishing (0)	Predicted: Phishing (1)
Actual: Non-Phishing (0)	**True Negative (TN)** Correctly classified as Non-Phishing	**False Positive (FP)** Misclassified as Phishing
Actual: Phishing (1)	**False Negative (FN)** Misclassified as Non-Phishing	**True Positive (TP)** – Correctly classified as Phishing

Given the evaluation metrics for each model, the True Positive (TP) and False Positive (FP) values can be calculated for each class (vishing and non-vishing). To calculate TP and FP, we use the confusion matrix formula for binary classification:

For a dataset with 40 vishing and 40 non-vishing audio samples, let's assume the following approximate confusion matrix values for Random Forest (based on the provided precision and recall, using the formula:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \times 100$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FP}} \times 100$$

TN = Recall (Non_vishing) × Total Non_vishing Samples ubiquitous

TP = Recall (Vishing) ×Total vishing samples

Table 3: Confusion Matrix Results

Model	TP (Class 0)	TP (Class 1)	FP (Class 0)	FP (Class 1)	TN (Class 0)	TN (Class 1)	FN (Class 0)	FN (Class 1)
RF	6	8	7	3	7	6	7	3
GB	8	7	5	4	7	6	5	4
SVM	10	8	3	3	10	8	3	3
CNN	12	3	1	8	2	3	1	8
LSTM	8	2	5	3	4	4	5	9

4.3 Interpretation

Among the evaluated machine learning models, the Support Vector Machine (SVM) demonstrated the best performance, achieving the highest number of true negatives (10) and the lowest number of false positives (2), indicating it produced the fewest false alarms. Gradient Boosting also performed well, maintaining a reasonable balance between true negatives and false positives, although it recorded slightly more false positives than SVM. In contrast, the Random Forest model exhibited the weakest performance, with the highest number of false positives (4) and the lowest number of true negatives (6), making it the least reliable option among the three for minimizing false alerts in phishing detection. This calculation process can be repeated for the other models based on their respective precision and recall values.

4.4 ROC CURVE comparison

The experimental results revealed varying performances among the different models tested, including Random Forest, Gradient Boosting, Support Vector Machine (SVM), Convolutional Neural Network (CNN), and Long Short-Term Memory (LSTM), based on both accuracy and ROC AUC scores.

The Random Forest model achieved an accuracy of 58.33%, which was the lowest among the models evaluated. The model displayed a moderate ROC AUC score of 0.72, indicating that it had a reasonable ability to distinguish between the classes. However, its performance in terms of precision, recall, and F1-score was uneven, with a noticeable imbalance between the two classes, particularly for Class 0.

The Gradient Boosting model exhibited an accuracy of 70.83%, and its ROC AUC score of 0.79 highlighted its strong discriminatory ability. Both the precision and recall for Class 1 were higher than those for Class 0, indicating that the model performed reasonably well across the classes with balanced results. This model demonstrated a more robust performance overall, particularly in distinguishing between the classes compared to Random Forest.

EdinBurg Peer Reviewed Journals and Books Publishers
Journal of Information and Technology
Vol. 5||Issue 6||pp 34-43||August||2025
Email: info@edinburgjournals.org||ISSN: 3080-9576

The SVM model achieved the highest accuracy at 75%, but its ROC AUC score was unexpectedly low at 0.19. This indicates that, while it correctly predicted the majority class, it struggled significantly in distinguishing between the two classes effectively. The imbalance in the precision and recall values across the two classes further supports this observation.

Both CNN and Gradient Boosting exhibited identical ROC AUC scores of 0.79, signifying that they performed equally well in terms of model discrimination. However, their performance was more consistent and robust across different metrics compared to the other models.

The LSTM model, with a ROC AUC score of 0.57, demonstrated an inferior performance compared to Gradient Boosting, CNN, and Random Forest, and it failed to match the discriminatory power of these models in distinguishing between the classes.

In conclusion, Gradient Boosting and CNN performed the best overall, as they not only had higher ROC AUC scores but also showed good balance between accuracy, precision, recall, and F1-score. The SVM**,** despite its high accuracy, proved to be less effective in distinguishing the classes due to its low ROC AUC score. The Random Forest and LSTM models, while useful, did not perform as well in terms of distinguishing between the classes, with Random Forest offering a moderate AUC and LSTM performing poorly in this respect.

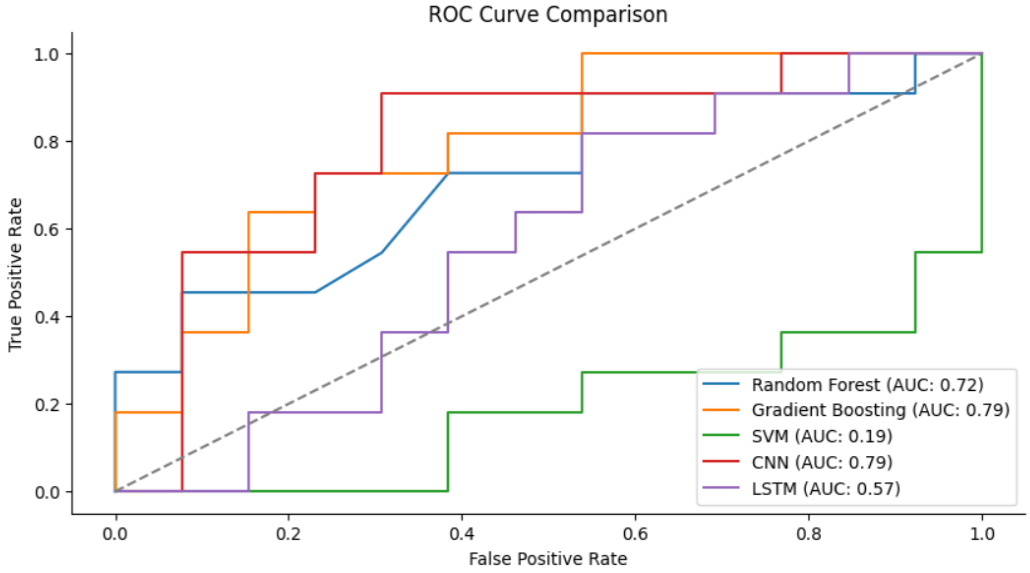

Figure 1: ROC CURVE comparison

5. Conclusion

The experimental results demonstrate that CNN and Gradient Boosting are the most effective models for detecting phishing calls in real-time. CNN's deep learning architecture allows it to extract complex features from voice data, making it a powerful tool for detecting subtle phishing patterns. Gradient Boosting, on the other hand, provides strong classification performance while maintaining a good balance between precision and recall. Although other models, such as SVM and LSTM, showed some promise, their limitations in ranking capability and sequential data handling reduced their effectiveness for this specific task. These findings highlight the need for further exploration of hybrid models that combine CNN's feature extraction capabilities with Gradient Boosting's strong classification potential. Additionally, future research should consider advanced deep learning techniques, such as transformer-based architectures, to enhance phishing detection. Expanding the dataset with more diverse and real-world phishing scenarios can also improve model robustness and adaptability to evolving threats. By integrating these improvements, phishing detection systems can be made more reliable, scalable, and resistant to adversarial attacks, ultimately strengthening the security of mobile money platforms against social engineering threats.

6. Recommendations

- **Hybrid Model Approach**: Combining CNN for feature extraction with Gradient Boosting for classification could further enhance detection accuracy and robustness.
- **Feature Engineering**: Incorporating additional linguistic and acoustic features may improve model discrimination between phishing and legitimate calls.
- **Dataset Expansion**: Increasing the dataset size and diversity could improve generalization, particularly for deep learning models.

Acknowledgments

I would like to express my sincere gratitude to Dr. Djuma Sumbiri for his unwavering support and encouragement throughout this research. His guidance and provision of essential resources were instrumental to the successful completion of this study.

References

Alshehri, A., Dahman, M., Assiri, M., Alshehri, A., Alqahtani, S., Shaiban, M., ... & Saeed, A. (2024, Sep_Dec). A decision support system based on classification algorithms for the diagnosis of periodontal diseases. *Saudi Journal of Oral Science, 11(3)*.

Alshehri, Abdulrahman; Dahman, Mohammed; Assiri, Mousa1; Alshehri, Abdulkarim; Alqahtani, Sharifah; Shaiban, Mohammed; Alqahtani, Bashyer; Althbyani, Sabah; Alhefdi, Hatem; Hakami, Khalid; Ali, Abdulbari; Saeed, Abdullah. (2024, Sep–Dec). A decision support system based on classification algorithms for the diagnosis of periodontal disease. *Saudi Journal of Oral Sciences 11(3)*. doi:10.4103/sjoralsci.sjoralsci_50_24

Figueiredo, J., Carvalho, A., Castro, D., Gonçalves, D., & Santos, N. (2024). On the Feasibility of Fully AI-automated Vishing Attacks. *arXiv*. doi:preprint arXiv:2409.13793

Ghafir, I., Saleem, J., Hammoudeh, M., Faour, H., Prenosil, V., Jaf, S., ... & Baker, T. (2018). Security threats to critical infrastructure: the human factor. *The Journal of Supercomputing, 74*, 4986–5002.

Jones, K. S., Armstrong, M. E., Tornblad, M. K., & Siami Namin, A. (2021). How social engineers use persuasion principles during vishing attacks. *Information & Computer Security, 29(2)*, 314–331.

Liu, X., Sahidullah, M., & Kinnunen, T. (2021). Optimizing multi-taper features for deep speaker verification. *IEEE Signal Processing Letters, 2187-2191*, 2187-2191.

Mouton, F., Leenen, L., & Venter, H. S. (2016). Social engineering attack examples, templates, and scenarios. *Computers & Security, 59*, 186–209.

Yeboah-Boateng, E. O., & Amanor, P. M. (2014). Phishing, smishing & vishing: an assessment of threats against mobile devices. *Journal of Emerging Trends in Computing and Information, 5(4)*, 297–307.