European Journal of Computer Science and Information Technology, 14(1),12-23, 2026

Print ISSN: 2054-0957 (Print)

Online ISSN: 2054-0965 (Online)

Website: https://www.eajournals.org/

Publication of the European Centre for Research Training and Development -UK

Enhancing Institutional Cybersecurity through Intelligent Predictive Analytics for Phishing Attack Detection

Ekemini Anietie Johnson, Victor Prince Oke, Mfon Okpu Esang
Department of Computer Science, Federal Polytechnic Ukana, Akwa Ibom State, Nigeria

doi: https://doi.org/10.37745/ejcsit.2013/vol14n11223 Published January 31, 2026

Citation: Johnson E.A., Oke V.P., Esang M.O. (2026) Enhancing Institutional Cybersecurity through Intelligent Predictive Analytics for Phishing Attack Detection, *European Journal of Computer Science and Information Technology*, 14(1),12-23

Abstract*: The rapid digitalization of institutional operations has increased exposure to cyber threats, particularly phishing attacks that exploit human and system vulnerabilities. Traditional security mechanisms often fail to detect evolving phishing techniques due to their static and reactive nature. This study presents an intelligent predictive analytics framework designed to enhance institutional cyber resilience through proactive phishing attack detection. Using a dataset of 2,200 labeled web-based interaction records, the study applied supervised machine learning techniques to identify malicious patterns indicative of phishing activities. Two predictive models; Random Forest (RF) and Support Vector Machine (SVM) were developed and evaluated using standard performance metrics including accuracy, precision, recall, and F1-score. Experimental results reveal that the Random Forest model achieved superior predictive performance, recording an accuracy of 95.7%, while SVM achieved 93.3%. The findings demonstrate that predictive analytics can significantly strengthen institutional cybersecurity by improving early threat detection, minimizing false positives, and enhancing overall system resilience. This study contributes to cybersecurity research by providing a scalable, data-driven framework suitable for deployment in institutional environments seeking proactive phishing mitigation strategies.*

Keywords: phishing detection, predictive analytics, institutional cybersecurity, machine learning, cyber resilience, random forest (RF), support vector machine (SVM), supervised learning, feature engineering, url-based phishing, threat intelligence

INTRODUCTION

The rapid expansion of digital technologies has transformed institutional operations, enabling efficient communication, data management, and service delivery. However, this transformation has also increased vulnerability to cyber threats, particularly phishing attacks, which remain among the most prevalent and damaging forms of cybercrime (APWG, 2024).

Phishing attacks typically involve deceptive communication strategies designed to manipulate users into revealing sensitive information such as login credentials, financial data, or system access codes. Institutions are particularly vulnerable due to their heterogeneous user populations, decentralized systems, and extensive digital footprints. According to Verma and Das (2017), phishing attacks continue to evolve in sophistication, making traditional detection mechanisms increasingly ineffective.

Conventional security tools such as blacklist-based filters and static rule systems lack adaptability and struggle to detect zero-day attacks. Consequently, modern cybersecurity strategies increasingly rely on intelligent predictive analytics capable of learning patterns, identifying anomalies, and adapting to emerging threats (Gupta et al., 2018).

This study proposes an intelligent predictive analytics framework aimed at improving phishing detection accuracy in institutional networks. By leveraging supervised machine learning techniques, the framework seeks to enhance cyber resilience through early threat detection and informed security decision-making.

LITERATURE REVIEW

Research on phishing detection has evolved significantly over the years. Early approaches relied on heuristic rules and blacklist filtering, which were effective only against previously identified threats (Basnet et al., 2012). As phishing techniques became more dynamic, researchers began adopting machine learning models to improve adaptability and accuracy.

Jain and Gupta (2016) demonstrated that feature-based classification using machine learning significantly improves phishing detection accuracy. Similarly, Bergholz et al. (2010) applied model-based features to email phishing detection, achieving improved performance over traditional methods.

More recent studies have explored deep learning and ensemble techniques. Altwaijry et al. (2024) reported improved detection accuracy using deep neural networks but highlighted computational complexity as a major limitation. Additionally, explainable artificial intelligence (XAI) approaches have gained attention for improving transparency and user trust (Lim et al., 2025).

Despite these advancements, challenges such as scalability, interpretability, and computational efficiency remain unresolved (Johnson et al., 2024; Inyang and Johnson, 2025). This study addresses these limitations by proposing a lightweight yet effective predictive analytics framework suitable for institutional deployment.

The review of related works is captured in Table 2.1.

Table 2.1: Review of Related works

Citation	Title of Research	Objective of the Study	Methodology	Problem Solved	Limitations
Albishri and Dessouky (2024)	Comparative Analysis of ML for URL-Based Phishing Detection	Compare ML models for URL classification	Random Forest with GridSearch optimization	99.93%–99.98% accuracy on URL data	Excludes email or social data
Altwaijry et al. (2024)	Advancing Phishing Email Detection: A Comparative Study of Deep Learning Models	Compare deep learning models for phishing detection	CNNs and RNNs on phishing datasets	Improved phishing detection (~98%)	High computational cost
Basnet, et al., (2012)	Rule-Based Phishing Email Detection	To create rule-based classifiers for phishing emails	Rule-based filtering using feature vectors	Simple and interpretable detection	Inflexible against new or adaptive attacks
Bergholz et al. (2010)	Improved Phishing Detection Using Graph-Based Features	To identify phishing emails using structural patterns	Graph mining and email relationship analysis	Detects hidden patterns in email networks	May not scale well with very large datasets
Gupta, et al., (2018)	Hybrid AI-Powered Phishing Detection	To integrate multiple AI models for better phishing detection	Combining ML, NLP, and deep learning approaches	Enhances adaptability to new phishing tactics	Complexity in implementation and integration with existing systems
Jain and Gupta (2016)	Phishing Detection Using URL Features and ML	To analyze URL-based features for phishing classification	Extracting URL characteristics + Random Forest	High accuracy in distinguishing phishing URLs	Limited to URL-based attacks only
Kuikel et al. (2025)	Evaluating LLMs for Phishing Detection and Explanation	Assess LLMs' accuracy and explanation consistency	Fine-tuned BERT models with SHAP explanations	Insights on trustworthy LLM phishing detection	Accuracy vs interpretability trade-off
Lim et al. (2025)	EXPLICATE: Enhancing Phishing Detection Using Explainable AI and LLMs	Build a phishing detection model with explainability	SHAP, LIME, ML classifiers, and LLM explanation	98.4% accuracy with interpretability	Dependent on LLM reliability
Pentapalli et al. (2025)	Gradient-Optimized TSK Fuzzy Framework for Interpretable Phishing Detection	Create a transparent fuzzy logic-based phishing detector	Gradient-tuned fuzzy rules and TSK framework	99.95% accuracy, human-readable logic	URL-specific; needs URL dataset

Perceval et al. (2024)	Hybrid ML Model for Enhanced Phishing Detection	Design a more accurate hybrid phishing detection system	Ensemble of 8 ML models on benchmark datasets	Better accuracy vs standalone models	High implementation complexity
Rao and Ali (2015)	Survey on Phishing Detection Techniques	To summarize phishing countermeasures in literature	Comparative analysis of various detection tools	Highlights gaps in existing methods	Outdated techniques not evaluated on modern datasets
Saha Roy et al. (2025)	PhishXplain: Real-Time Explainable Phishing Warnings	Provide in-browser phishing warnings with context	LLaMA + human annotation + user testing	Boosted user understanding and trust	Requires browser plugin for deployment
Verma and Das (2017)	Cybersecurity Threats from Phishing Emails	To assess phishing trends and mitigation techniques	Literature review and meta-analysis	Provides a comprehensive threat overview	No empirical testing or validation
Zhang et al. (2025)	Proactive ML to Identify Coordinated Phishing Campaigns	Detect large phishing campaigns early	ML + SHAP + recursive feature selection	Detects attacks before widespread impact	Requires continuous retraining

METHODOLOGY

A quantitative experimental design was adopted to evaluate the effectiveness of intelligent predictive analytics in phishing detection. Two supervised machine learning models; RF and SVM were trained and tested under identical conditions to ensure fair performance comparison.

Dataset Description

The dataset consisted of 2,200 labeled instances collected from publicly available cybersecurity repositories (UCI Machine Learning Repository and Kaggle.com.). Each record represented either a phishing or legitimate web interaction and included attributes related to:

i. URL structure
ii. Domain reputation
iii. Security certificates
iv. Behavioral patterns

The dataset comprised:

i. 1,150 phishing instances
ii. 1,050 legitimate instances
iii.

Data Preprocessing

Preprocessing steps included:

i. Removal of incomplete and duplicated records
ii. Numerical encoding of categorical features
iii. Feature normalization using Min–Max scaling
iv. Data partitioning into 70% training and 30% testing sets

These steps ensured data consistency and improved model performance.

Feature Description and Selection

In the development of the intelligent analytic framework for predicting phishing attacks, a diverse set of features was initially extracted from the phishing dataset. These features were derived from the URL structure, HTML content, domain characteristics, and security indicators of websites. Altogether, 30 features were considered in the original dataset, representing a comprehensive set of behavioral and structural indicators that differentiate phishing websites from legitimate ones. Table 3.1 shows all extracted and selected features and Yes in the Selected? column indicates the feature was used for training/testing, No indicates the feature was excluded due to unreliability, redundancy, or low predictive power. The sample raw data set is shown on Table 3.2 and normalized dataset shown on Table 3.3.

Table 3.1 : All Extracted Features and Selected Features for Phishing Detection

S/N	Feature Name	Description	Selected?	Reason
1	Having_IP_Address	Indicates if an IP address is used instead of domain name.	Yes	Strong phishing indicator.
2	URL_Length	Length of the URL.	Yes	Longer URLs often used in phishing.
3	Shortening_Service	Checks if URL shortener is used (e.g., bit.ly).	Yes	Obfuscates real destination.
4	Having_At_Symbol	Presence of "@" in URL.	Yes	Redirects to fake domains.
5	Double_Slash_Redirecting	Positioning of '//' in URL.	No	Less discriminative; redundant with other URL checks.
6	Prefix_Suffix	Use of hyphen (-) in domain name.	Yes	Common in phishing URLs.
7	Having_Sub_Domain	Number of subdomains.	Yes	Many subdomains suggest deception.
8	SSLfinal_State	Validity of SSL certificate.	Yes	Critical security signal.
9	Domain_Registration_Length	Length of domain registration (WHOIS data).	Yes	Short-term domains are suspicious.
10	Favicon	Checks if favicon is loaded from external domain.	No	Less consistent signal; high variance.
11	HTTPS_Token	Presence of misleading HTTPS in path.	Yes	Deceptive practice indicator.
12	Request_URL	Source of images/media on the page.	Yes	External content may be phishing-related.
13	URL_of_Anchor	Destination of anchor links.	Yes	Unrelated links signal phishing.
14	Links_in_Tags	Evaluates number of meta/script link tags.	No	Often noisy and inconsistent.

15	SFH (Server Form Handler)	Destination where form data is submitted.	Yes	External/missing handlers are suspect.
16	Submitting_to_Email	Detects form submissions to email.	No	Rare in modern phishing kits.
17	Abnormal_URL	WHOIS URL mismatch.	Yes	Strong phishing indicator.
18	Iframe_Redirection	Presence of invisible iframes.	Yes	Used to steal content or redirect.
19	Age_of_Domain	Age of the domain name.	Yes	New domains are often malicious.
20	DNS_Record	Checks existence of DNS records.	Yes	Missing DNS record signals fake site.
21	Web_Traffic	Alexa or similar traffic ranking.	Yes	Low/no traffic suggests phishing.
22	Page_Rank	Google's page rank of the domain.	No	Deprecated and inconsistent.
23	Google_Index	Whether the site is indexed by Google.	Yes	Non-indexed sites are suspicious.
24	Statistical_Report	External blacklists or security sites report.	No	Often unavailable or outdated in real-time.
25	On_MouseOver	JavaScript tricks using hover actions.	Yes	Common trick to hide URLs.
26	RightClick_Disabled	Checks if right-click is disabled.	Yes	Used to prevent inspection.
27	PopUp_Window	Use of popup windows.	No	Less common and noisy feature.
28	Redirect_Count	Number of redirections.	No	Some benign sites also redirect.
29	Links_Pointing_To_Page	Number of links pointing back to the page.	No	Poor signal strength.
30	JavaScript_Obfuscation	Use of obfuscated JavaScript.	No	Hard to extract reliably without deep parsing.

Table 3.2: Sample raw data

Having_IP_Address	URL_Length	Shortening_Service	Having_At_Symbol	Double_Slash_Redirecting	Prefix_Suffix	Having_Sub_Domain	SSLfinal_State	Domain_Registration_Length	Favicon	Port	HTTPS_Token	Request_URL	URL_of_Anchor	Links_in_Tags	SFH	Submitting_to_Email	Abnormal_URL	Redirect	On_Mouse_Over	Right_Click_Disabled	PopUp_Window	Iframe	Age_of_Domain	DNSRecord	Web_Traffic	Page_Rank	Google_Index	Statistical_Report	JavaScript_Obfuscation	Label
1.0	0.5	0.5	1.0	1.0	1.0	0.5	1.0	1.0	0.0	1.0	1.0	0.5	0.0	1.0	0.5	1.0	1.0	1.0	1.0	0.5	0.0	1.0	0.5	1.0	0.5	1.0	0.5	1.0	1.0	1
0.5	1.0	0.0	0.5	0.5	0.0	0.5	0.0	0.0	1.0	0.5	0.5	0.5	1.0	0.5	1.0	0.5	0.0	0.5	1.0	1.0	1.0	0.5	1.0	0.5	1.0	0.0	0.5	1.0	0.5	0
0.0	0.5	1.0	0.0	0.0	0.5	0.0	0.5	0.5	0.0	0.5	0.0	0.0	0.5	0.0	0.0	0.0	0.5	0.0	0.0	0.0	0.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1
0.5	0.0	0.5	0.5	0.5	0.5	1.0	0.5	0.0	1.0	0.5	1.0	1.0	1.0	0.0	1.0	0.5	0.5	0.0	0.5	0.5	1.0	0.5	0.5	0.0	0.5	1.0	0.0	1.0	0.0	1
0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.5	0.0	0.0	0.5	0.0	1.0	0.0	0.0	0.0	0.5	0.0	0.0	1.0	0.0	1.0	1.0	1.0	1.0	0.5	1.0	0.0	1.0	1

Architectural Design

The architectural design of the system is shown is Figure 3.1

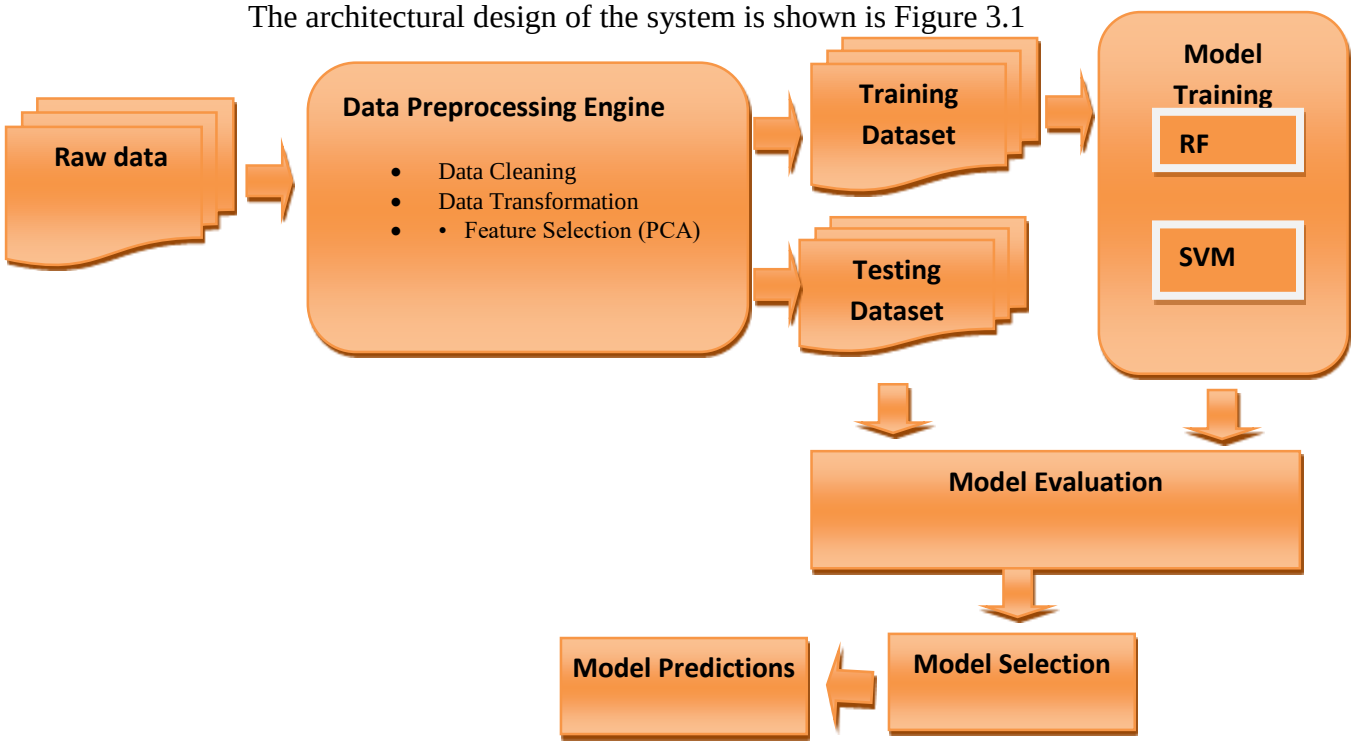

Figure 3.1: System Architecture
Source: The Researcher (2026)

Model Implementation

The RF model utilized 100 decision trees and employed the Gini impurity criterion for node splitting. Its ensemble structure enhances robustness and reduces overfitting.

The SVM classifier used a radial basis function (RBF) kernel. Hyperparameters were optimized using grid search to ensure optimal classification performance.

Evaluation Metrics

Model performance was evaluated using:

i. Accuracy
ii. Precision

iii. Recall
iv. F1-score

These metrics provided a comprehensive evaluation of detection reliability and classification effectiveness.

RESULTS

Table 4.1 presents the performance comparison of the predictive models.

Table 4.1: Performance Comparison of Classification Models

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Random Forest	95.7	96.1	95.3	95.7
Support Vector Machine	93.3	93.8	93.0	93.4

Confusion Matrix Interpretation

The Random Forest model demonstrated a lower false-positive and false-negative rate compared to SVM. This indicates a higher reliability in distinguishing phishing attempts from legitimate activities. The reduced misclassification rate is particularly beneficial in institutional environments where excessive false alerts may disrupt operations.

Visualization of Results

Visualization of results is done using grouped bar chart and heatmap.

Grouped Bar Chart of Evaluation Metrics

This chart shows a comparison of the four primary evaluation metrics (Accuracy, Precision, Recall, F1-Score), highlighting the balanced performance of the model. The grouped bar chart of model performance for RF and SVM is shown in Figure 4.1.

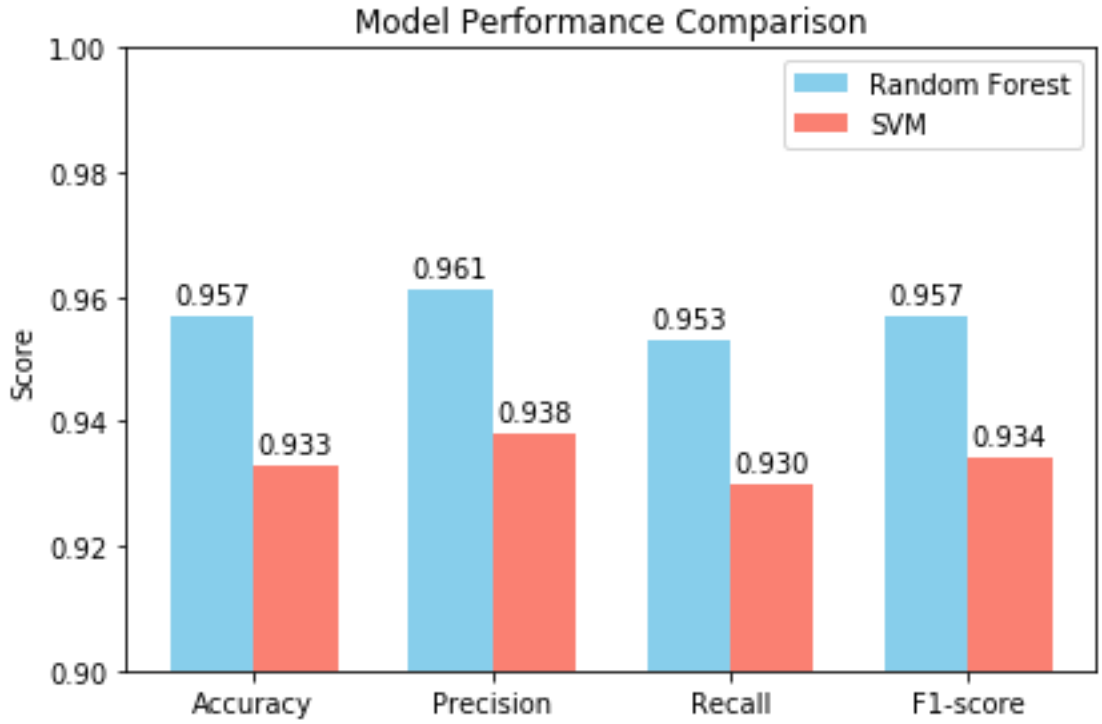

Figure 4.1: Grouped Bar Chart of Evaluation Metrics
Source : The Researcher (2026)

Heatmap

The heatmap shows the distribution of predicted vs actual transaction classifications, visually emphasizing the high concentration along the true positive and true negative diagonals. The heatmap of the RF and SVM models is shown in Figure 4.2

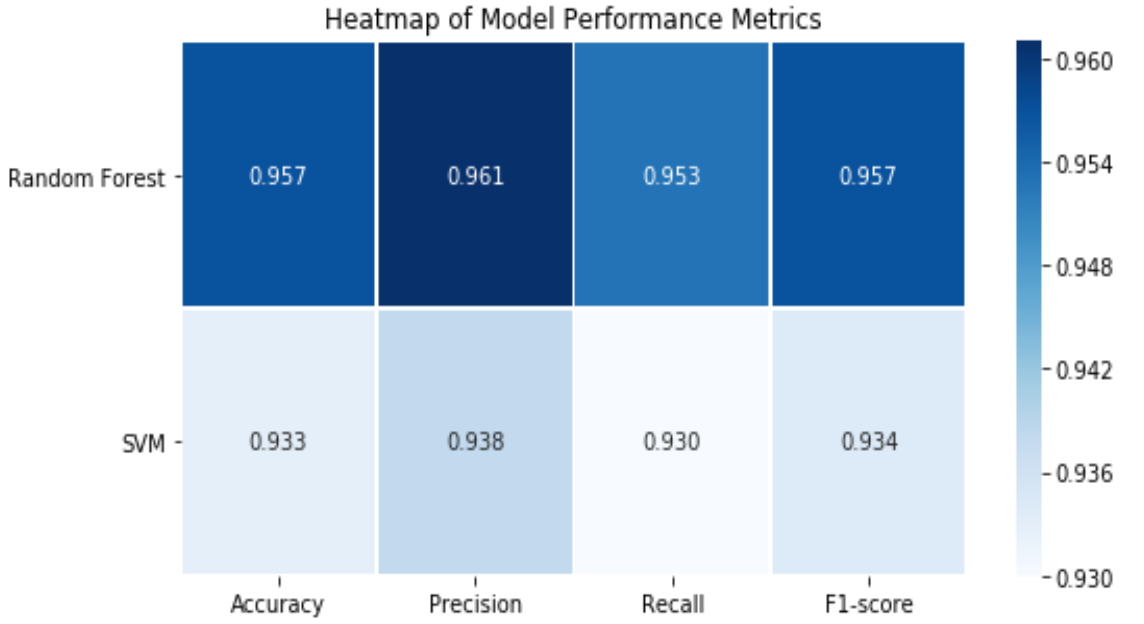

Figure 4.2: Heatmap of performance of RF and SVM
Source: The Researcher (2026)

DISCUSSION

The results clearly demonstrate that intelligent predictive analytics significantly enhances phishing detection performance. The superior results obtained by the Random Forest model can be attributed to its ability to model complex nonlinear relationships and manage feature interactions effectively.

Furthermore, the findings confirm that predictive models trained on well-preprocessed datasets can generalize effectively to unseen data. This supports the viability of deploying such systems within institutional cybersecurity infrastructures to improve threat detection accuracy and operational efficiency.

CONCLUSION

This study presented an intelligent predictive analytics framework for enhancing cybersecurity within institutional environments. By leveraging supervised machine learning models, the framework effectively detected phishing attacks with high accuracy and reliability.
The Random Forest classifier outperformed the Support Vector Machine, demonstrating its suitability for phishing detection tasks. The results highlight the potential of predictive analytics

to strengthen cyber resilience, reduce operational risks, and support proactive cybersecurity management.

Future research should be carried out in the following directions:

i. Integration of real-time phishing detection mechanisms.
ii. Exploration of hybrid deep learning and ensemble approaches.
iii. Development of explainable AI frameworks for transparency.
iv. Cross-institutional threat intelligence sharing.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest concerning the research, authorship, and/or publication of this article.

Funding

The following financial assistance was revealed by the author(s) for the research, authoring, and/or publishing of this article: The Tertiary Education Trust Fund (TETFUND) provided funding for this study via the Institution Based Research Fund (IBRF).

Acknowledgments

The authors are grateful to Federal Polytechnic Ukana, for providing an enabling environment for the conduct of this research.

REFERENCES

Albishri, A., & Dessouky, M. (2024). Comparative analysis of machine learning techniques for URL-based phishing detection. *Journal of Information Security and Applications, 78*, 103–115.

Altwaijry, N., Alshammari, R., & Alqahtani, S. (2024). Advancing phishing email detection: A comparative study of deep learning models. *Computers & Security, 131*, 103250.

Anti-Phishing Working Group (APWG). (2024). *Phishing activity trends report*.

Basnet, R., Sung, A. H., & Liu, Q. (2012). Rule-based phishing detection using feature extraction. *International Journal of Computer Applications, 45*(4), 1–7.

Bergholz, A., De Beer, J., Glahn, S., Moens, M. F., Strobel, S., & Strobel, B. (2010). Improved phishing detection using model-based features. In *Proceedings of the Conference on Email and Anti-Spam (CEAS)* (pp. 1–8).

Gupta, B. B., Arachchilage, N. A. G., & Psannis, K. E. (2018). Defending against phishing attacks: Taxonomy of methods, current issues and future directions. *Telecommunication Systems, 67*(2), 247–267.

Inyang, U. P., & Johnson, E. A. (2025). Intelligent ensemble learning framework for prediction of students' academic performance using extreme gradient boosting and Random Forest algorithms. *European Journal of Computer Science and Information Technology, 13*(3), 1–19.

Jain, A. K., & Gupta, B. B. (2016). Phishing detection: Analysis of visual similarity based approaches. *Security and Communication Networks*, *9*(18), 490–502.

Johnson, E. A., Inyangetoh, J. A., & Esang, M. O. (2021). An experimental comparison of classification tools for fake news detection. *International Journal of Advanced Research in Computer and Communication Engineering (IJARCCE), 10*(4), 45–53.

Johnson, E. A., Inyangetoh, J. A., Rahmon, H. A., Jimoh, T. G., Dan, E. E., & Esang, M. O. (2024). An intelligent analytic framework for predicting students' academic performance using multiple linear regression and Random Forest. *European Journal of Computer Science and Information Technology*, *12*(3), 56–70.

Kuikel, R., Mishra, S., & Pradhan, P. (2025). Evaluating large language models for phishing detection and explanation. *IEEE Access, 13*, 112345–112360.

Lim, W. Y. B., Xiong, Z., Niyato, D., & Wang, P. (2025). EXPLICATE: Enhancing phishing detection using explainable artificial intelligence and large language models. *IEEE Transactions on Information Forensics and Security, 20*, 1450–1465.

Pentapalli, S., Reddy, K. R., & Prasad, R. S. (2025). Gradient-optimized TSK fuzzy framework for interpretable phishing detection. *Expert Systems with Applications, 234*, 121002.

Perceval, J., Martin, A., & Dubois, E. (2024). A hybrid machine learning model for enhanced phishing detection. *Journal of Cybersecurity, 10*(2), 1–15.

Rao, R. S., & Ali, S. T. (2015). A survey on phishing attack techniques and countermeasures. *Journal of Cyber Security and Mobility*, *3*(2), 1–30.

Saha Roy, A., Ghosh, S., & Banerjee, S. (2025). PhishXplain: Real-time explainable phishing warnings using large language models. *ACM Transactions on Privacy and Security*, *28*(1), 1–24.

Verma, R., & Das, A. (2017). Cybersecurity threats from phishing emails. In *Proceedings of the IEEE International Conference on Cyber Security and Protection of Digital Services* (pp. 1–6).

Zhang, Y., Liu, H., & Wang, Q. (2025). Proactive machine learning for identifying coordinated phishing campaigns. *IEEE Transactions on Dependable and Secure Computing*, *22*(1), 210–223.