ISSN 2319-5991 www.ijerst.com

Peer Reviewed, Referred & Indexed Journal

Research Paper

# Advancing Phishing Detection with Voting Classifier, Hybrid Deep Models, and Flask-Based Real-Time Prediction Interface

MILHAN AHMED MUSTAFA
Student,
Department of Information Technology , University of the People
mamkk99@gmail.com

**Abstract:** Phishing attacks remain a significant cybersecurity threat, deceiving users into revealing sensitive information through fraudulent websites. This study proposes a machine learning–based phishing detection framework enhanced with advanced feature selection, deep learning, and explainable intelligence. Multiple feature importance techniques, including mutual information, chi-square analysis, and permutation importance, are applied across Feedforward Neural Networks, Deep Neural Networks, TabNet, and Wide and Deep models to identify influential phishing indicators. To address class imbalance and improve robustness, SMOTEENN resampling is combined with an ensemble Voting Classifier integrating Random Forest and Bagging with Decision Trees. Experiments conducted on two public phishing datasets demonstrate superior performance, where the Voting Classifier achieved 98.7% accuracy on the phishing websites dataset and 98.5% accuracy on the web page phishing dataset, outperforming individual deep learning models. Explainable Artificial Intelligence techniques such as LIME and SHAP are incorporated to interpret predictions and highlight feature contributions, ensuring transparency and trust. For real-world deployment, the framework is implemented using the Flask platform, offering an interactive web interface with secure user signup and signin using SQLite. Users submit URLs for analysis, and the system provides real-time predictions as "Phishing website" or "Non Phishing website," supporting reliable and interpretable phishing detection for online security.

***"Index Terms:*** *Phishing detection, cybersecurity, deep learning, neural networks, feature selection, hyperparameter optimization, real-time detection, TabNet, wide and deep model."*

## 1. INTRODUCTION

In the 21st century, cyberspace has become highly interconnected, making cybersecurity a critical concern for individuals, businesses, and organizations worldwide. Among the most prevalent cyber threats, phishing attacks pose significant risks by deceiving users into revealing sensitive information such as passwords, financial details, and personal identifiers [1]. The term "phishing" is derived from "fishing," reflecting the attacker's attempt to lure victims into disclosing confidential information [2]. These attacks often involve impersonation of legitimate institutions, including banks, government agencies, and social media platforms, exploiting users' trust to execute fraudulent activities [3]. Phishing is pervasive across multiple digital channels—email, social media, instant messaging, and other online services—making it a multifaceted threat capable of bypassing even robust security measures [4].

Phishing attacks have evolved beyond simple email scams into sophisticated social engineering schemes that manipulate communication trust and exploit human behavior [5]. Such attacks can lead to severe consequences, including financial loss, identity theft, reputational damage, and legal repercussions [6]. Furthermore, phishing incidents are frequently associated with data breaches, ransomware, and other criminal cyber activities, highlighting the potential expansion of these threats into broader security domains [7]. The growing complexity of phishing has driven extensive research into detection mechanisms that combine heuristic approaches, feature selection, and advanced machine learning techniques [8]. For example, studies emphasize the importance of selecting appropriate features to improve detection precision and combining heuristic algorithms with machine learning models to enhance overall performance [9].

Recent research efforts have explored various machine learning approaches to phishing detection. Techniques such as Feed Forward Neural Networks, Deep Neural Networks (DNN), Wide and Deep models, and the TabNet architecture have been evaluated for their effectiveness in identifying malicious emails and phishing sites [10]. Optimization strategies, including grid search for hyperparameter tuning, further enhance model performance, enabling higher accuracy while reducing false positives and false negatives. Additionally, the development of new evaluation metrics, such as the anti-phishing score, allows for a more comprehensive assessment of detection systems.

This study focuses on systematically evaluating machine learning models for phishing email detection using a dataset from PhishTank comprising 111 features. The feed-forward approach was identified as the most effective model, fine-tuned through grid search optimization to achieve optimal accuracy and efficiency. By integrating classical feature selection, modern machine learning techniques, and holistic evaluation methods, this research aims to advance cybersecurity defenses against evolving phishing threats, providing a robust framework for detecting malicious emails across diverse contexts.

## 2. LITERATURE REVIEW

Phishing attacks have become a significant cybersecurity concern due to the increasing reliance on digital platforms for communication, financial transactions, and personal information management. Researchers have increasingly explored machine learning and heuristic-based approaches to enhance phishing detection across multiple domains. Mosa et al. [11] conducted an extensive study on machine learning techniques for detecting phishing URL attacks, emphasizing the effectiveness of using various supervised learning algorithms to classify phishing URLs. Their work highlighted the importance of combining multiple features to improve detection performance and reduce false positives, laying a foundation for data-driven detection approaches that can adapt to evolving phishing strategies.

Yu et al. [12] proposed a phishing detection system based on a multi-feature neural network, which leveraged a diverse set of URL and web page characteristics to identify malicious sites. Their approach demonstrated that integrating multiple types of features—including domain information, page content attributes, and hyperlink analysis—significantly improved detection accuracy compared to models relying on a single feature set. This study underscored the potential of neural network architectures to capture complex patterns

in phishing data, providing a robust mechanism for automatic identification of phishing attacks. Similarly, Novakovic and Markovic [13] focused on URL-based phishing detection using neural networks, demonstrating that carefully engineered neural models could effectively differentiate between legitimate and malicious URLs. Their approach showed that machine learning-based URL analysis could outperform traditional rule-based systems by adapting to new phishing techniques with minimal manual intervention.

Salihu et al. [14] explored a heuristics-based methodology for detecting phishing URLs, highlighting the importance of domain-specific rules and pattern recognition for early identification of phishing attempts. By analyzing structural features, URL characteristics, and behavioral patterns of phishing websites, their approach enabled rapid classification with low computational complexity. Tanimu and Shiaeles [15] investigated machine learning algorithms for phishing detection, focusing on algorithm selection and comparative performance analysis. They demonstrated that supervised learning models such as decision trees, random forests, and support vector machines could achieve high detection accuracy when trained with well-preprocessed and feature-rich datasets, reinforcing the value of algorithmic evaluation in practical phishing detection systems.

Sánchez-Paniagua et al. [16] introduced a phishing website detection framework using a novel multipurpose dataset and web technology-based features. Their study emphasized the integration of diverse web features—including HTML, JavaScript, and CSS attributes—to improve detection rates. By expanding the feature space to include structural and content-based attributes, they provided a more holistic detection mechanism capable of identifying sophisticated phishing sites that might bypass simpler feature-based methods. Wei and Sekiya [17] addressed the critical role of feature selection in machine learning-based phishing detection, proposing methods to identify the most relevant features that significantly contribute to classification accuracy. Their research highlighted how proper feature selection could enhance model interpretability while reducing computational overhead, making phishing detection more efficient and scalable.

Chinnasamy et al. [18] presented an efficient phishing attack detection system using multiple machine learning algorithms. Their study compared the performance of classifiers such as logistic regression, k-nearest neighbors, and ensemble methods, showing that ensemble approaches could leverage the strengths of individual models to achieve superior performance. The study also emphasized the importance of dataset preprocessing and feature engineering in improving detection robustness. Dangwal and Moldovan [19] further reinforced the importance of feature selection by investigating techniques for machine learning-based phishing website detection. Their work suggested that reducing irrelevant or redundant features not only improved model accuracy but also enhanced training efficiency, which is crucial for deploying detection systems in real-time environments.

Yahya [20] explored various machine learning approaches for detecting phishing websites, highlighting the effectiveness of combining multiple classifiers and feature sets to improve prediction performance. His research also emphasized the challenges associated with imbalanced datasets and proposed strategies to mitigate their impact, including oversampling and feature weighting, to ensure reliable detection results. Collectively, these studies provide a

comprehensive overview of the state-of-the-art in phishing detection research. They demonstrate that integrating feature engineering, feature selection, neural network architectures, heuristic analysis, and ensemble learning can significantly improve detection accuracy, adaptability, and efficiency. Furthermore, the continuous development of novel datasets, advanced evaluation metrics, and optimization techniques has enabled the creation of more reliable and scalable phishing detection systems. These contributions collectively form the foundation for ongoing research aimed at enhancing cybersecurity defenses, particularly against increasingly sophisticated and dynamic phishing threats.

## 3. MATERIALS AND METHODS

The proposed system enhances phishing detection by leveraging advanced feature selection and deep learning techniques. It utilizes two publicly available phishing datasets, which are preprocessed for analysis. [23] Feature selection is performed using mutual information, chi-square, and permutation importance scores derived from models including Feed-Forward Neural Networks (FNN), Deep Neural Networks (DNN), TabNet, and Wide & Deep, identifying the most influential features for classification. Multiple deep learning architectures—CNN, DNN, Wide & Deep, TabNet, and Inception—are employed to capture complex phishing patterns. A Hybrid model combining FNN, DNN, and TabNet, along with a Voting Classifier integrating [21] Random Forest and Bagging with Decision Trees, improves robustness. SMOTEENN addresses class imbalance, while Explainable AI methods, LIME and SHAP, interpret feature contributions. A Flask-based web interface enables real-time URL classification.

Fig.1 Proposed Architecture

Figure 1 outlines a machine learning approach for Web Page Phishing Detection. The process begins with Datasets, followed by Preprocessing to clean the data. After Data Visualization, Feature Selection and Data Splitting occur. Various Models (e.g., CNN, DNN, Hybrid) are built, trained, and subjected to Testing. Model Evaluation uses metrics like Accuracy and F1 Score. The Best Trained Model is saved, deployed via a Flask Framework to take User Input, and predicts a Phishing or Non Phishing Website, with explanations from XAI—LIME & SHAP.

**a) Dataset Collection:**

***i) Phishing Websites Dataset:*** The phishing websites dataset, sourced from Mendeley Data, comprises 57,405 cleaned records with 111 numerical features representing URL, domain, and webpage-based characteristics. [22] Each entry captures attributes such as symbol counts, server configurations, SSL certificates, redirects, and indexing status, with the target label indicating phishing or legitimate websites. This dataset provides a comprehensive foundation for training and evaluating machine learning models in phishing detection, enabling robust analysis of diverse online threat indicators while ensuring data quality through preprocessing.

Fig.2 Phishing Websites Dataset

***ii) Web page phishing detection:*** The web page phishing detection dataset, available on Mendeley Data, contains 11,256 processed records with 88 attributes describing lexical, host-based, and content-based features of URLs and websites. [24] These features include length measures, symbol counts, domain properties, HTTPS usage, page structure, traffic statistics, and ranking information. Each record is labeled as "phishing" or "legitimate," serving as the target variable. This dataset offers a rich and diverse foundation for building and validating machine learning models for phishing website detection.

Fig.3 Web Page Phishing Detection Dataset

**b) Pre-Processing:**

The preprocessing steps ensure data quality, meaningful feature selection, and balanced evaluation by cleaning, visualizing, selecting key attributes, and splitting the dataset for robust phishing detection model development.

***i) Data Processing:*** This stage ensures data quality by removing null values that may cause inconsistencies during analysis and eliminating duplicate entries that could introduce redundancy or bias into the model. By cleaning the dataset, we establish a reliable foundation for further processing and model training, ensuring accuracy and consistency throughout the machine learning pipeline.

***ii) Data Visualization:*** Data visualization helps in understanding the dataset structure and relationships. A correlation matrix is used to examine dependencies and interactions among features, while sample outcomes highlight the class distribution. These visual insights guide further analysis and provide clarity on how variables contribute to phishing detection.

iii) ***Feature Selection:*** Feature selection is conducted using multiple methods to identify the most informative attributes. Mutual information and chi-square tests assess feature relevance statistically, while permutation importance is evaluated through deep learning models such as FNN, DNN, TabNet, and Wide & Deep. The top 20 ranked features are retained to enhance model interpretability and efficiency.

**c) Training and Testing:**

The cleaned and feature-selected dataset is partitioned into training and testing sets. This division ensures that models are trained on one portion of the data and validated on another, enabling unbiased evaluation of predictive performance and generalization capability across unseen phishing and legitimate website samples.

**d) Algorithms:**

**CNN (Convolutional Neural Network):** Processes URL and website features through convolutional and pooling layers, capturing local dependencies and hierarchical patterns. [25] Enhances predictive accuracy by detecting subtle phishing indicators and complex relationships among features, enabling robust classification of websites as legitimate or phishing.

$$S(i,j) = \sum_m \sum_n I(i+m, j+n) \cdot K(m,n) \quad (1)$$

**DNN (Deep Neural Network):** Learns intricate, non-linear relationships among website and URL attributes through multi-layer transformations. [26] Captures high-level abstractions and interactions, allowing accurate classification in high-dimensional data and improving detection of malicious website behaviors.

**Wide and Deep:** Combines linear and deep components to model explicit feature interactions and high-level abstractions simultaneously. Captures complex relationships among categorical and continuous features, improving generalization, accuracy, and detection of subtle phishing patterns in diverse datasets.

**TabNet:** Processes tabular website and URL attributes using sequential attention, dynamically identifying important features. [30] Balances high predictive performance and interpretability, efficiently handling complex interactions while emphasizing influential phishing indicators.

**Inception:** Analyzes website features through parallel convolutions of varying scales to capture local and global patterns. [27] Improves detection accuracy by identifying diverse phishing indicators, including URL structure, content characteristics, and metadata interactions.

$$O = concat\left(F_1(X), F_2(X), F_3(X), F_4(X)\right) \quad (2)$$

**Hybrid Model (FNN + DNN + TabNet):** Integrates Feedforward, Deep Neural Network, and TabNet outputs to capture basic, complex, and attention-based feature patterns. [28] Enhances robustness, feature transparency, and accuracy while handling high-dimensional, imbalanced website data effectively.

**Voting Classifier (RF + Bagging with DT):** Combines [29] Random Forest and Bagging with Decision Trees to stabilize predictions, reduce variance, and highlight feature importance. Achieves higher accuracy and robustness, improving classification of websites as phishing or legitimate compared to individual models.

$$\hat{y} = argmax_c \left( \sum_{i=1}^{n} II\left(\hat{y}_i = c\right) \right) \quad (3)$$

**e) Integration of XAI and Flask Framework:**

The integration of XAI (Explainable Artificial Intelligence) with the Flask framework provides an interactive and interpretable environment for deploying machine learning models. By combining XAI techniques with Flask's lightweight web capabilities, users can not only obtain predictions but also understand the reasoning behind them. This approach enhances transparency, allowing insights into feature contributions, model decisions, and prediction confidence, which is particularly valuable for sensitive applications such as phishing detection or healthcare analytics.

In this setup, Flask serves as the web interface for user interactions, handling input data and displaying outputs, while XAI methods generate explanations for the model's predictions. This integration ensures that complex deep learning or ensemble models remain interpretable, trustworthy, and user-friendly, bridging the gap between high-performance AI and practical real-world deployment.

## 4. EXPERIMENTAL RESULTS

**Accuracy:** The accuracy of a test is its ability to differentiate the patient and healthy cases correctly. To estimate the accuracy of a test, we should calculate the proportion of true positive and true

negative in all evaluated cases. Mathematically, this can be stated as:

$$Accuracy = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{TN} + \text{FN}} (4)$$

**Precision:** Precision evaluates the fraction of correctly classified instances or samples among the ones classified as positives. Thus, the formula to calculate the precision is given by:

$$Precision = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}} (5)$$

**Recall:** Recall is a metric in machine learning that measures the ability of a model to identify all relevant instances of a particular class. It is the ratio of correctly predicted positive observations to the total actual positives, providing insights into a model's completeness in capturing instances of a given class.

$$Recall = \frac{\text{TP}}{\text{TP} + \text{FN}} (6)$$

**F1-Score:** F1 score is a machine learning evaluation metric that measures a model's accuracy. It combines the precision and recall scores of a model. The accuracy metric computes how many times a model made a correct prediction across the entire dataset.

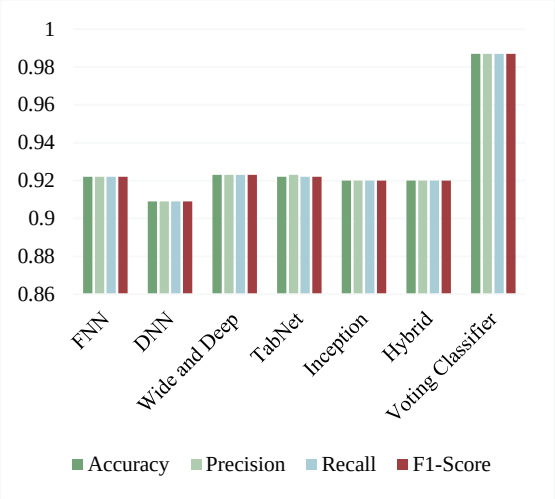
$$F1\ Score = 2 * \frac{Recall\ \text{X}\ Precision}{Recall + Precision} * 100 (7)$$

***Table.1*** Performance Evaluation – Phishing Dataset

| ML Model | Accuracy | Precision | Recall | F1-Score |
|---|---|---|---|---|
| FNN | 0.922 | 0.922 | 0.922 | 0.922 |
| DNN | 0.909 | 0.909 | 0.909 | 0.909 |
| Wide and Deep | 0.923 | 0.923 | 0.923 | 0.923 |
| TabNet | 0.922 | 0.923 | 0.922 | 0.922 |
| Inception | 0.920 | 0.920 | 0.920 | 0.920 |
| Hybrid | 0.920 | 0.920 | 0.920 | 0.920 |
| **Voting Classifier** | **0.987** | **0.987** | **0.987** | **0.987** |

In Table.1, the performance of various machine learning models for phishing detection is compared, highlighting the Voting Classifier's superior overall effectiveness.

Fig.4 Comparison Graph Phishing Dataset




In fig.4, model performance is illustrated with Accuracy is represented in green, Precision in light green, Recall in blue, and F1-Score in red, showing the Voting Classifier clearly outperforming other models.

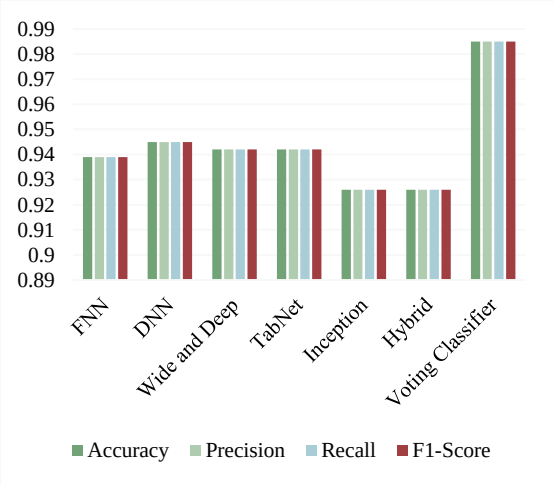
***Table.2*** Performance Evaluation – Web Phishing Dataset

| ML Model | Accuracy | Precision | Recall | F1-Score |
|---|---|---|---|---|
| FNN | 0.939 | 0.939 | 0.939 | 0.939 |
| DNN | 0.945 | 0.945 | 0.945 | 0.945 |
| Wide and Deep | 0.942 | 0.942 | 0.942 | 0.942 |
| TabNet | 0.942 | 0.942 | 0.942 | 0.942 |
| Inception | 0.926 | 0.926 | 0.926 | 0.926 |
| Hybrid | 0.926 | 0.926 | 0.926 | 0.926 |

| | | | | |
|---|---|---|---|---|
| **Voting Classifier** | **0.985** | **0.985** | **0.985** | **0.985** |

In Table.2, the performance of different machine learning models for phishing detection is compared, highlighting the Voting Classifier's highest overall effectiveness.

Fig.5 Comparison Graph Web Phishing Dataset




In fig.5, Accuracy is represented in green, Precision in light green, Recall in blue, and F1-Score in red, clearly showing the Voting Classifier outperforming all other models.


Fig.6 Enter Input URL

In Fig.6, the interface allows users to enter a domain, enabling automated detection of phishing or legitimate websites efficiently.


Fig.7 Predicted Result

In Fig.7, after entering a domain, the system classifies it as a phishing website, demonstrating accurate automated detection capabilities.


Fig.8 Enter Input URL

In Fig.8, users can input a domain, and the system swiftly identifies whether the website is phishing or legitimate.


Fig.9 Predicted Result

In Fig.9, after entering a domain, the system accurately identifies it as a non-phishing website, showcasing reliable automated detection.

## 5. CONCLUSION

This work successfully demonstrates a robust and accurate phishing detection framework by integrating advanced feature selection, deep learning, ensemble learning, and explainable AI techniques. Feature importance methods such as mutual information, chi-square, and permutation

importance applied across FNN, DNN, TabNet, and Wide & Deep models effectively identified influential attributes, reducing redundancy and improving learning efficiency. Deep learning architectures, including CNN, Inception, and hybrid combinations, enhanced representation learning, while SMOTEENN addressed class imbalance to ensure better generalization. Among all evaluated approaches, the ensemble Voting Classifier combining Random Forest and Bagging with Decision Trees achieved superior performance, recording 98.7% accuracy on the phishing websites dataset and 98.5% accuracy on the web page phishing dataset. Model transparency was ensured using LIME and SHAP, enabling clear interpretation of feature contributions and building user trust. For real-world applicability, the optimized model was deployed using the Flask framework, providing a lightweight and interactive web interface. The system supports secure user signup and signin through SQLite, real-time URL input, backend preprocessing, and instant prediction. The application clearly displays outcomes as “Phishing website” or “Non Phishing website,” delivering an accurate, interpretable, and deployable solution for practical phishing detection environments.

Future work for this phishing detection system can focus on expanding datasets with real-time web traffic to improve model adaptability against evolving phishing strategies. Integration of Natural Language Processing (NLP) techniques for analyzing webpage content, domain names, and embedded scripts could further enhance detection accuracy. Advanced ensemble methods and reinforcement learning can be explored for dynamic adaptability. Incorporating graph-based approaches may help identify hidden relationships among URLs, IPs, and domains. Future improvements may also emphasize lightweight models for deployment in mobile and low-resource environments. Additionally, extending explainable AI methods will improve transparency, fostering user trust. Finally, integration with browser extensions and enterprise-level security platforms can ensure broader usability and real-world application.

**REFERENCES**

[1] Zara, U., Ayub, K., Khan, H. U., Daud, A., Alsahfi, T., & Gulzar, S. (2024). Phishing website detection using deep learning models. IEEE Access.

[2] Sahingoz, O. K., BUBE, E., & Kugu, E. (2024). Dephides: Deep learning based phishing detection system. Ieee Access, 12, 8052-8070.

[3] Şengel, Ö. (2024, October). Analysis of Learning Techniques for Phishing Website Detection. In 2024 Innovations in Intelligent Systems and Applications Conference (ASYU) (pp. 1-6). IEEE.

[4] Aldakheel, E. A., Zakariah, M., Gashgari, G. A., Almarshad, F. A., & Alzahrani, A. I. (2023). A deep learning-based innovative technique for phishing detection in modern security with uniform resource locators. Sensors, 23(9), 4403.

[5] Do, N. Q., Selamat, A., Krejcar, O., Yokoi, T., & Fujita, H. (2021). Phishing webpage classification via deep learning-based algorithms: An empirical study. Applied Sciences, 11(19), 9210.

[6] R. Jayaraj, A. Pushpalatha, K. Sangeetha, T. Kamaleshwar, S. U. Shree, and D. Damodaran, ‘‘Intrusion detection based on phishing detection with machine learning,’’ Meas., Sensors, vol. 31, Feb. 2024, Art. no. 101003. [Online]. Available:

https://www.sciencedirect.com/science/article/pii/S2665917423003392

[7] M. R. Chinguwo and R. Dhanalakshmi, ‘‘Detecting cloud based phishing attacks using stacking ensemble machine learning technique,’’ Int. J. Res. Appl. Sci. Eng. Technol., vol. 11, no. 3, pp. 360–367, Mar. 2023, doi: 10.22214/ijraset.2023.49422.

[8] C. Rajeswary and M. Thirumaran, ‘‘The LSTM-based automated phishing detection driven model for detecting multiple attacks on tor hidden services,’’ J. Intell. Fuzzy Syst., vol. 44, no. 6, pp. 8889–8903, Mar. 2023, doi: 10.3233/jifs-224142.

[9] B. Subba, ‘‘A heterogeneous stacking ensemble-based security framework for detecting phishing attacks,’’ in Proc. Nat. Conf. Commun. (NCC), Feb. 2023, pp. 1–6, doi: 10.1109/NCC56989.2023.10068026.

[10] K. R. Nataraj, D. K. Yashaswini, R. Hema, N. S. Pawar, and S. Yashaswi, ‘‘Phishing attack detection using machine learning,’’ in Proc. 4th Int. Conf. Data Sci., Mach. Learn. Appl. Singapore: Springer, 2023, pp. 355–370, doi: 10.1007/978-981-99-2058-7_33.

[11] D. T. Mosa, M. Y. Shams, A. A. Abohany, E.-S. M. El-Kenawy, and M. Thabet, ‘‘Machine learning techniques for detecting phishing URL attacks,’’ Comput., Mater. Continua, vol. 75, no. 1, pp. 1271–1290, 2023, doi: 10.32604/cmc.2023.036422.

[12] S. Yu, C. An, T. Yu, Z. Zhao, T. Li, and J. Wang, ‘‘Phishing detection based on multi-feature neural network,’’ in Proc. IEEE Int. Perform., Comput., Commun. Conf. (IPCCC), Nov. 2022, pp. 73–79, doi: 10.1109/IPCCC55026.2022.9894337.

[13] J. Novakovic and S. Markovic, ‘‘Detection of URL-based phishing attacks using neural networks,’’ in Proc. Int. Conf. Theor. Appl. Comput. Sci. Eng. (ICTASCE), Sep. 2022, pp. 132–136.

[14] S. A. Salihu, I. D. Oladipo, A. A. Wojuade, M. AbdulRaheem, A. O. Babatunde, A. R. Ajiboye, and G. B. Balogun, ‘‘Detec tion of phishing URLs using heuristics-based approach,’’ in Proc. 5th Inf. Technol. Educ. Develop. (ITED), Nov. 2022, pp. 1–7, doi: 10.1109/ITED56637.2022.10051199.

[15] J. Tanimu and S. Shiaeles, ‘‘Phishing detection using machine learning algorithm,’’ in Proc. IEEE Int. Conf. Cyber Secur. Resilience (CSR), Jul. 2022, pp. 317–322, doi: 10.1109/CSR54599.2022.9850316.

[16] M. Sánchez-Paniagua, E. Fidalgo, E. Alegre, and R. Alaiz-Rodríguez, ‘‘Phishing websites detection using a novel multipurpose dataset and web technologies features,’’ Expert Syst. Appl., vol. 207, Nov. 2022, Art. no. 118010, doi: 10.1016/j.eswa.2022.118010.

[17] Y. Wei and Y. Sekiya, ‘‘Feature selection approach for phishing detection based on machine learning,’’ in Proc. Int. Conf. Appl. CyberSecur. (ACS), H.R.HassenandH.Batatia,Eds.,Cham,Switzerland:Springer,Jan.2022, pp. 61–70.

[18] P. Chinnasamy, N. Kumaresan, R. Selvaraj, S. Dhanasekaran, K. Ramprathap, and S. Boddu, ‘‘An efficient phishing attack detection using machine learning algorithms,’’ in Proc. Int. Conf. Advancements Smart, Secure Intell. Comput. (ASSIC), Bhubaneswar, India, Nov. 2022, pp. 1–6.

[19] S.DangwalandA.-N.Moldovan,‘‘Featureselection for

machinelearning based phishing websites detection,'' in Proc. Int. Conf. Cyber Situational Awareness, Data Anal. Assessment (CyberSA), Dublin, Ireland, Jun. 2021, pp. 1–6.

[20] F. Yahya, ''Detection of phishing websites using machine learning approaches,'' in Proc. Int. Conf. Data Sci. Appl. (ICoDSA), Bandung, Indonesia, 2021, pp. 40–47.

[21] F. Salahdine, Z. E. Mrabet, and N. Kaabouch, ''Phishing attacks detection a machine learning-based approach,'' in Proc. IEEE 12th Annu. Ubiquitous Comput., Electron. Mobile Commun. Conf. (UEMCON), New York, NY, USA, Dec. 2021, pp. 250–255.

[22] A. Alswailem, B. Alabdullah, N. Alrumayh, and A. Alsedrani, ''Detecting phishing websites using machine learning,'' in Proc. 2nd Int. Conf. Comput. Appl. Inf. Secur. (ICCAIS), Riyadh, Saudi Arabia, May 2019, pp. 1–6.

[23] M. Almseidin, A. A. Zuraiq, M. Al-Kasassbeh, and N. Alnidami, ''Phishing detection based on machine learning and feature selection methods,'' Int. J. Interact. Mobile Technol. (iJIM), vol. 13, no. 12, p. 171, Dec. 2019. [Online]. Available: https://www.learntechlib.org/p/216410

[24] M. Baykara and Z. Z. Gürel, ''Detection of phishing attacks,'' in Proc. 6th Int. Symp. Digit. Forensic Secur. (ISDFS), Antalya, Turkey, Mar. 2018, pp. 1–5.

[25] K. L. Chiew, K. S. C. Yong, and C. L. Tan, ''A survey of phishing attacks: Their types, vectors and technical approaches,'' Expert Syst. Appl., vol. 106, pp. 1–20, Sep. 2018.

[26] H. Zuhair, A. Selamat, and M. Salleh, ''Feature selection for phishing detection: A review of research,'' Int. J. Intell. Syst. Technol. Appl., vol. 15, no. 2, p. 147, 2016.

[27] Zara, U., Ayub, K., Khan, H. U., Daud, A., Alsahfi, T., & Gulzar, S. (2024). Phishing website detection using deep learning models. IEEE Access.

[28] Bhimavarapu, U. (2025). Phishing Attack Response and Risk Mitigation Using Deep Learning and Machine Learning. In Critical Phishing Defense Strategies and Digital Asset Protection (pp. 109-120). IGI Global Scientific Publishing.

[29] Almousa, M., Zhang, T., Sarrafzadeh, A., & Anwar, M. (2022). Phishing website detection: How effective are deep learning-based models and hyperparameter optimization?. Security and Privacy, 5(6), e256.

[30] Şengel, Ö. (2024, October). Analysis of Learning Techniques for Phishing Website Detection. In 2024 Innovations in Intelligent Systems and Applications Conference (ASYU) (pp. 1-6). IEEE.