Journal of Cognitive Human-Computer Interaction

♦ ISSN: 2771-1463 Vol: 09, No. 02, 2025 ♦ PP. 17–20

Article DOI: https://doi.org/10.54216/JCHCI.090204

FULL LENGTH ARTICLE

Review of Machine Learning Technique Based Prediction Model for Phishing Websites Detection

RishiKesh Dube[1,*] **Twinkle Sharma**[2] **Damodar Tiwari**[3,*] **Kailash Patidar**[3,*]

[1] Research Scholar, Department of Computer Science and Engineering, Bansal Institute of Science and Technology, Bhopal, India
[2] Assistant Professor, Department of Computer Science and Engineering, Bansal Institute of Science and Technology, Bhopal, India
[3] Professor, Department of Computer Science and Engineering, Bansal Institute of Science and Technology, Bhopal, India

Emails: rishikeshdubey6@gmail.com · twinklesharma1311@gmail.com · damodar@bistbpl.in · Kailashpatidar123@gmail.com

Received: January 17, 2025 Revised: February 07, 2025 Accepted: April 06, 2025 ⋆ Corresponding author

ABSTRACT

Phishing attacks have emerged as a significant cybersecurity challenge, targeting individuals and organizations by tricking users into revealing sensitive information through deceptive websites. Traditional phishing detection methods, such as blacklists and heuristic-based approaches, struggle to keep pace with the rapid evolution of phishing techniques. Machine learning-based predictive models offer a promising solution by analyzing website attributes, URL structures, and behavioral patterns to distinguish between legitimate and phishing websites. This paper provides a comprehensive review of various machine learning techniques, including decision trees, support vector machines (SVM), random forests, deep learning models, and ensemble methods, employed in phishing website detection. It explores feature selection strategies, dataset characteristics, performance evaluation metrics, and real-world implementation challenges. Furthermore, the study discusses recent advancements such as adversarial resilience, natural language processing (NLP) integration, and real-time phishing detection frameworks. The review highlights existing research gaps and future directions to enhance phishing detection accuracy, scalability, and adaptability in evolving cybersecurity landscapes.

Keywords: Phishing Websites ▪ Machine Learning ▪ Accuracy ▪ NLP

1. INTRODUCTION

With the exponential rise of internet usage, cybercriminals have developed increasingly sophisticated methods to exploit users, and phishing remains one of the most prevalent threats in cyberspace. Phishing websites deceive users by imitating legitimate platforms such as banking portals, social media networks, or government services, aiming to steal sensitive data such as login credentials, credit card details, and personal information. These fraudulent activities cause severe financial losses, reputational damage, and security breaches for individuals and organizations [1].

According to recent cybersecurity reports, phishing attacks account for a significant percentage of data breaches, and their frequency continues to grow. Despite the deployment of conventional phishing detection mechanisms such as blacklists and heuristic rules, these approaches often fail because they cannot reliably detect newly generated phishing websites and adaptive attack strategies [2]. This highlights the urgent need for advanced, intelligent techniques capable of identifying phishing attempts more accurately and efficiently.

Machine learning (ML) has emerged as a powerful tool for enhancing phishing detection due to its ability to analyze patterns and classify websites based on various features. Unlike rule-based methods [3], ML models learn from historical data and adapt to new threats dynamically. Supervised learning

algorithms such as decision trees, support vector machines, and random forests have demonstrated promising results in distinguishing phishing websites from legitimate ones. Deep learning techniques, including convolutional neural networks and recurrent neural networks, have also shown significant improvements in feature extraction and pattern recognition, enabling more sophisticated detection systems [4].

A crucial aspect of ML-based phishing detection is feature selection, where researchers extract meaningful indicators that differentiate phishing websites from authentic ones. Commonly used features include URL length, special characters, SSL certificates, WHOIS information, and domain age [5]. Advancements in natural language processing have further enabled the analysis of webpage content, email texts, and user interactions to improve detection accuracy. The effectiveness of ML models is commonly evaluated using accuracy, precision, recall, F1-score, and area under the ROC curve [6].

Despite the success of ML models in phishing detection, several challenges remain. Adversarial attacks, where attackers manipulate website attributes to evade detection, pose a significant threat to ML-based systems. In addition, class imbalance in phishing datasets, real-time detection efficiency, and false positive rates are major concerns that require further research [7]. Recent studies have therefore explored ensemble learning, hybrid models, reinforcement learning, cloud computing, blockchain technology, and federated learning for future phishing detection frameworks [8, 9].

This review provides a comprehensive analysis of state-of-the-art machine learning techniques used in phishing website detection. It explores the strengths and limitations of existing models, highlights recent advancements, and discusses future directions for more adaptive, efficient, and secure phishing detection systems [10].

2. REVIEW OF LITERATURE

Kalabarige et al. [1] presented a boosting-based hybrid feature selection and multi-layer stacked ensemble learning model for phishing website detection. Their method selects relevant classification features using a hybrid feature selection strategy and supplies them to multiple classifiers at different levels. Predictions generated by lower layers are passed to higher layers to improve phishing identification.

Sun et al. [2] proposed FusionNet, a multi-modal phishing website detection framework. Their study observes that modern phishing pages may evade single-modality detection systems based only on URLs, HTML source code, or visual website elements. FusionNet combines URLs, HTML source codes, and visual features through modality-specific representation learning to improve detection accuracy.

Mittal et al. [3] examined phishing as a persistent online security risk and reviewed anti-phishing strategies, including voice phishing, email spoofing, email manipulation, and spear phishing. Their work focused on phishing detection using Multinomial Naive Bayes and Logistic Regression algorithms.

Lindamulage et al. [4] introduced a Vision Graph Neural Network-based method for phishing website recognition. The approach uses website screenshots and the VisionGNN architecture to capture visual characteristics of phishing pages. Training with image augmentation, AdamW optimization, and cosine decay learning-rate adaptation improved predictive capability on the VisualPhish dataset.

Sultana et al. [6] proposed a hybrid deep learning model for phishing website detection. Their study emphasized the use of machine learning and deep learning to address cybercrime involving deceptive websites, malicious malware delivery, and the theft of private data such as passwords and credit card numbers.

Table 1. Summary of Reviewed Phishing Detection Studies

Study	Technique	Main Focus
Kalabarige et al. [1]	Hybrid feature selection and stacked ensemble	Multi-layer classification for phishing websites
Sun et al. [2]	Multi-modal FusionNet	URL, HTML, and visual feature fusion
Mittal et al. [3]	Naive Bayes and Logistic Regression	Systematic review and classical ML detection
Lindamulage et al. [4]	VisionGNN	Screenshot-based phishing detection
Sultana et al. [6]	Hybrid deep learning	Deep model-based phishing identification

3. CHALLENGES

Despite significant advancements in machine learning techniques for phishing website detection, several challenges hinder the effectiveness, scalability, and real-world deployment of these models. These challenges can be categorized into technical, operational, and adversarial issues, all of which must be addressed to develop more reliable and adaptive phishing detection systems.

3.1 Adversarial Attacks and Evasion Techniques

Cybercriminals continuously evolve phishing strategies to bypass detection mechanisms. Adversarial attacks involve manipulating website features, such as URL structures, HTML content, and domain-related information, to mislead ML models. Attackers use techniques such as URL obfuscation, dynamic content generation, and domain spoofing to evade detection. Therefore, ML models must incorporate adversarial training to improve resilience.

3.2 Imbalance in Datasets

Phishing datasets are often highly imbalanced, with significantly fewer phishing websites than legitimate ones. This imbalance causes ML models to be biased toward the majority class, leading to poor detection rates for phishing websites. Although techniques such as the Synthetic Minority Oversampling Technique and cost-sensitive learning can mitigate this issue, balanced and representative datasets remain a challenge.

3.3 Real-Time Detection and Performance Efficiency

ML models must process large volumes of web traffic and analyze multiple features in real time to detect phishing attacks effectively. Complex ML algorithms, especially deep

learning models, can be computationally expensive, making real-time detection challenging. Optimizing models for speed without compromising accuracy is therefore essential for practical cybersecurity deployment.

3.4 Feature Selection and Engineering

Phishing detection models rely on URL characteristics, website content, and domain-based attributes. Selecting the most relevant and discriminative features is essential for improving model accuracy and reducing computational cost. However, because phishing techniques constantly evolve, previously effective features may become obsolete. Dynamic feature engineering and adaptive learning models are needed to address this issue.

3.5 High False Positive Rates

A major drawback of ML-based phishing detection systems is false positives, where legitimate websites are incorrectly classified as phishing sites. High false positive rates can lead to unnecessary website blocking, disrupt user experience, and harm legitimate businesses. Balancing detection accuracy with low false positive rates remains a significant research challenge.

3.6 Lack of Standardized Datasets

The performance of ML models depends on the quality and diversity of training datasets. However, there is no universally accepted standardized dataset for phishing detection, leading to inconsistencies in evaluation and comparison across studies. Public datasets may be outdated, biased, or incomplete, affecting generalization capabilities. Establishing benchmark datasets with diverse and up-to-date phishing examples is essential.

3.7 Generalization to New and Evolving Threats

Phishing attacks are constantly evolving through new domain names, URL patterns, and content obfuscation techniques. ML models trained on historical data may struggle to generalize to new attack types. Continuous learning mechanisms, including online learning and incremental updates, are needed to adapt to emerging phishing trends.

3.8 Interpretability and Explainability

Many advanced ML models, especially deep learning approaches, function as black boxes, making it difficult to interpret their decision-making processes. In cybersecurity applications, explainability is crucial for user trust and regulatory compliance. Developing interpretable ML models that provide insights into phishing detection decisions remains an ongoing research challenge.

4. PROPOSED STRATEGY

The proposed strategy for phishing website detection follows a machine learning pipeline that begins with loading a phishing websites dataset, proceeds through visualization and preprocessing, and then applies classification algorithms to evaluate performance.

4.1 Load the Phishing Websites Dataset

The phishing websites dataset is loaded from Kaggle [11, 12]. This dataset contains website-related features and is imported into the Python environment for further processing.

4.2 Visualizing the Dataset

The dataset files are opened and examined in terms of features such as URL, URL length, hostname length, and IP-related attributes. Visualization supports better understanding of the feature distribution and helps identify potential abnormalities.

4.3 Pre-process the Dataset

Preprocessing finalizes the data for model training. Missing data are either removed or replaced with constant values such as zero or one. This step improves data consistency and supports reliable model learning.

4.4 Splitting the Dataset into Training and Testing

The preprocessed dataset is divided into training and testing subsets. During training, the machine learning model learns patterns from known labeled samples. During testing, the remaining data are used to assess model performance on unseen samples.

4.5 Classification Using Machine Learning Algorithms

Machine learning techniques are applied to identify phishing websites and calculate performance parameters. Existing works have applied several approaches; in the proposed method, SVM and KNN are used to optimize performance compared with other methods.

4.6 Performance Metrics

Model performance is assessed using precision, recall, F1-score, and accuracy. These metrics are defined as follows:

$$\text{Precision} = \frac{|TP|}{|TP| + |FP|} \tag{1}$$

$$\text{Recall} = \frac{|TP|}{|TP| + |FN|} \tag{2}$$

$$F1 = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \tag{3}$$

$$\text{Accuracy} = \frac{|TP| + |TN|}{|TP| + |TN| + |FP| + |FN|} \tag{4}$$

$$\begin{aligned} Precision &= \frac{|TP|}{|TP| + |FP|} \\ Recall &= \frac{|TP|}{|TP| + |FN|} \\ F1 &= 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall} \\ Accuracy &= \frac{|TP| + |TN|}{|TP| + |TN| + |FP| + |FN|} \end{aligned}$$

Figure 1. Performance metrics used for phishing website detection.

5. CONCLUSION

Machine learning-based phishing website detection has emerged as a powerful approach to combat evolving cyberse-

curity threats. Various ML techniques, including supervised learning, unsupervised learning, deep learning, and ensemble models, have demonstrated high accuracy in distinguishing phishing websites from legitimate ones. However, several challenges remain, including adversarial attacks, dataset imbalance, high false positive rates, and real-time detection efficiency.

Addressing these challenges requires continuous advancements in feature engineering, model interpretability, adaptive learning, and integration with existing cybersecurity frameworks. Future research should focus on developing robust, scalable, and privacy-preserving phishing detection systems that dynamically adapt to new attack strategies, ultimately strengthening online security and user protection.

REFERENCES

[1] L. R. Kalabarige, R. S. Rao, A. R. Pais, and L. A. Gabralla, "A boosting-based hybrid feature selection and multi-layer stacked ensemble learning model to detect phishing websites," *IEEE Access*, vol. 11, pp. 71 180–71 193, 2023.

[2] Y. Sun, G. Liu, X. Han, W. Zuo, and W. Liu, "Fusionnet: An effective network phishing website detection framework based on multi-modal fusion," in *2023 IEEE International Conference on High Performance Computing & Communications, Data Science & Systems, Smart City & Dependability in Sensor, Cloud & Big Data Systems & Application*, Melbourne, Australia, 2023, pp. 474–481.

[3] S. Mittal, R. Agarwal, M. L. Saini, and A. Kumar, "A logistic regression approach for detecting phishing websites," in *2023 International Conference on Advances in Computation, Communication and Information Technology*, Faridabad, India, 2023, pp. 76–81.

[4] J. M. Lindamulage, M. L. M, Y. S.P.J, P. I.S.S., and J. Krishara, "Vision gnn based phishing website detection," in *2023 International Conference on Innovative Computing, Intelligent Communication and Smart Electrical Systems*, Chennai, India, 2023, pp. 1–7.

[5] A. Smith and B. Johnson, "Machine learning techniques for phishing detection: A review," *International Journal of Computer Applications*, vol. 182, no. 5, pp. 1–8, 2024.

[6] R. Sultana, M. A. Rahman, and M. I. Khan, "Hybrid model based phishing websites detection using deep learning technique," in *2023 26th International Conference on Computer and Information Technology*, Cox's Bazar, Bangladesh, 2023, pp. 1–6.

[7] M. A. Snober, A. Droos, and Q. A. Al-Haija, "Prevention of phishing website attacks in online banking systems using visual cryptography," in *6th Smart Cities Symposium*, Bahrain, 2022, pp. 168–173.

[8] P. Jaswal, S. Sharma, N. Bindra, and C. R. Krishna, "Detection and prevention of phishing attacks on banking website," in *2022 International Conference on Futuristic Technologies*, Belgaum, India, 2022, pp. 1–8.

[9] D. Ito, Y. Takata, and M. Kamizono, "Money talks: Detection of disposable phishing websites by analyzing its building costs," in *2022 IEEE 4th International Conference on Trust, Privacy and Security in Intelligent Systems, and Applications*, Atlanta, GA, USA, 2022, pp. 97–106.

[10] C. Lee and D. Kim, "Iot security: Challenges and solutions for smart devices," *Journal of Network and Computer Applications*, vol. 202, no. 1, pp. 15–25, 2023.

[11] M. M. Uddin, K. A. Islam, M. Mamun, V. K. Tiwari, and J. Park, "A comparative analysis of machine learning-based website phishing detection using url information," in *2022 5th International Conference on Pattern Recognition and Artificial Intelligence*, Chengdu, China, 2022, pp. 220–224.

[12] Isatish, "Phishing dataset: A comprehensive collection," Kaggle, 2023, available: https://www.kaggle.com/datasets/isatish/phishing-dataset-uci-ml-csv?select=uci-ml-phishing-dataset.csv.