PhishGuard: An Intelligent Browser Extension for Real-Time Phishing Website Detection Using Machine Learning

Ms. Samruddhi Bajpayee[1], Mr. Aditya Talwatkar[2], Mr. Vivek Jadhao[3], Ms. Aarya Thakare[4], Mr. Yash S. Pali[5], Prof. Dipali A.Sananse[6]

[1,2,3,4,5] *UG Student, Jawaharlal Darda Institute of Engineering and Technology, Yavatmal, Maharashtra, India*

***Abstract*— Phishing websites pose a significant threat to online security by tricking users into revealing sensitive information such as login credentials and financial data. Traditional blacklist-based detection methods are often ineffective against newly created phishing websites. This paper proposes PhishGuard, a real-time phishing website detection system that integrates machine learning with a browser extension. The system extracts URL-based features and uses a trained machine learning model to classify websites as phishing or legitimate. The browser extension continuously monitors visited websites and alerts users when a suspicious site is detected, providing enhanced protection against phishing attacks during web browsing.**

***Index Terms*— Phishing Detection, Machine Learning, PhishGuard, Browser Extension, Real-Time Detection.**

I. INTRODUCTION

Phishing attacks have become a major cybersecurity threat as attackers create fraudulent websites that imitate legitimate platforms to steal sensitive information such as passwords and financial details. The rapid growth of online services and digital transactions has significantly increased phishing incidents worldwide. Traditional detection techniques such as blacklist-based systems are widely used in modern browsers; however, these systems often fail to detect newly generated phishing websites because they rely on previously identified malicious URLs [17]. As a result, researchers have explored machine learning approaches to improve phishing detection accuracy and automatically identify malicious websites [23].

Machine learning algorithms are widely used in phishing detection systems because they can analyze patterns in URLs, domain information, and webpage structures to classify websites as legitimate or phishing. Several studies have applied classification algorithms such as Random Forest, Support Vector Machine, and Logistic Regression to detect phishing websites with high accuracy [6],[21]. These machine learning models can learn phishing patterns from large datasets and detect malicious websites more effectively than traditional rule-based methods [9].

Recent studies have also explored deep learning techniques to further improve phishing detection performance. Models such as Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN) have demonstrated strong performance in identifying phishing URLs by analyzing complex patterns within webpage content and URL structures [14]. Experimental results from previous research show that deep learning-based phishing detection systems can achieve very high detection accuracy, making them suitable for real-time security applications [25].

Feature extraction and feature selection also play an important role in improving phishing detection models. Phishing datasets often contain multiple features derived from URLs, hyperlinks, and webpage attributes. Selecting the most relevant features helps improve classification accuracy and reduces computational complexity in machine learning models [8],[12]. Studies show that hybrid feature selection techniques can significantly enhance phishing detection performance while reducing unnecessary features in the dataset [19].

Despite these advancements, modern phishing attacks continue to evolve and often attempt to bypass detection systems using techniques such as cloaking, adversarial manipulation, and domain obfuscation. Attackers may modify certain website features to

evade machine learning classifiers and avoid detection [22]. These challenges highlight the need for more robust phishing detection systems capable of identifying sophisticated phishing attacks in real time [15].

To address this issue, researchers have proposed integrating phishing detection mechanisms directly into browser environments. Browser-based detection systems can monitor visited websites, analyze URL features in real time, and warn users immediately when a suspicious website is detected [3]. Such real-time protection mechanisms significantly improve user security by preventing interaction with phishing websites before sensitive information is compromised [1],[5].

Therefore, this paper proposes PhishGuard, a real-time phishing website detection system that integrates machine learning techniques with a browser extension. The system extracts relevant URL-based features and uses a trained machine learning model to classify websites as phishing or legitimate. The browser extension continuously monitors browsing activity and alerts users when suspicious websites are detected, providing an effective solution for enhancing web security against phishing attacks [7],[18],[28].

II. PROPOSED METHODOLOGY

A. Dataset Collection

The dataset used in this study consists of phishing and legitimate website URLs collected from publicly available phishing datasets. Several previous studies have utilized datasets from sources such as PhishTank, Kaggle, and the UCI Machine Learning Repository to train phishing detection models. These datasets contain labeled phishing and legitimate URLs along with multiple attributes extracted from websites [8],[21]. Using publicly available datasets helps improve model reliability and ensures that machine learning algorithms are trained using real-world phishing samples [4].

B. Feature Extraction

Feature extraction plays an important role in machine learning-based phishing detection systems. Previous research shows that phishing detection models often rely on lexical, host-based, and content-based features extracted from URLs and webpage structures. Common features include URL length, number of dots, presence of HTTPS, domain age, and redirection count [12],[19]. These features help distinguish phishing websites from legitimate ones by analyzing patterns commonly used by attackers in malicious URLs [6].

C. Machine Learning Model

Machine learning algorithms are widely used for phishing detection because they can identify patterns within large datasets and classify websites effectively. Several studies have demonstrated that classifiers such as Random Forest, Support Vector Machine, and Logistic Regression provide high accuracy in phishing website detection tasks [14],[9]. Among these models, Random Forest has shown strong performance due to its ensemble learning capability, which combines multiple decision trees to improve classification accuracy and reduce overfitting [5].

D. Browser Extension Implementation

To enable real-time phishing detection, the trained machine learning model is integrated into a browser extension called PhishGuard. Browser-based detection systems allow URLs to be analyzed directly during user browsing sessions and provide immediate warnings when suspicious websites are detected. Previous research has shown that integrating phishing detection mechanisms into browser environments significantly improves user protection by preventing interaction with malicious websites [3],[11]. The PhishGuard extension continuously monitors visited URLs, extracts relevant features, and uses the trained machine learning model to classify the website as phishing or legitimate.

E. System Workflow

The overall workflow of the proposed system begins when a user visits a website. The browser extension captures the URL and extracts relevant features required for classification. These features are then processed by the trained machine learning model to determine whether the website is phishing or legitimate. If the system detects a phishing website, the extension immediately displays a warning message to the user, preventing further interaction with the malicious site. This real-time detection approach

enhances web security and helps protect users from phishing attacks during browsing sessions [1],[7].

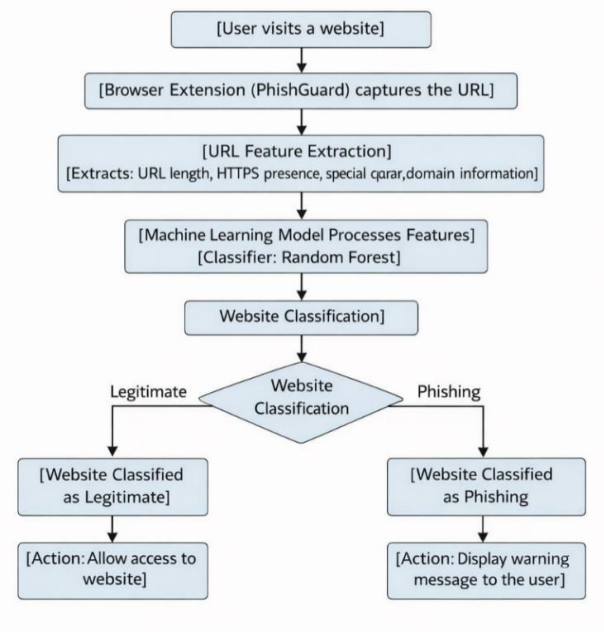

Fig 2.1: PhishGuard System Workflow (Architecture)

III. RESULTS AND DISCUSSION

A. Effectiveness of Machine Learning in Phishing Detection

Machine learning techniques play an important role in improving phishing detection systems by automatically identifying patterns in website URLs and domain characteristics. Algorithms such as Random Forest, Support Vector Machines, and Logistic Regression are widely used to classify websites as phishing or legitimate based on extracted features [4],[17]. These models can learn phishing patterns from datasets and detect malicious websites more effectively than traditional rule-based or blacklist detection methods. In the proposed PhishGuard system, the machine learning model is trained using a custom database of phishing and legitimate URLs, allowing the system to detect suspicious websites and provide real-time warnings through the browser extension [9].

B. Identification of Phishing Attack Techniques

Phishing attacks often rely on persuasive techniques designed to manipulate users into revealing sensitive information. Many phishing websites mimic legitimate websites by copying logos, webpage layouts, and domain structures to increase credibility and deceive users [21],[7]. Attackers frequently use urgency, fear, and trust-based social engineering strategies to encourage victims to interact with fraudulent websites [10]. Machine learning models can detect such patterns by analyzing textual, structural, and behavioral characteristics of phishing websites, enabling automated identification of malicious persuasion techniques used in phishing campaigns [25].

C. Model Performance and Interpretability Trade-off

Many studies highlight a trade-off between model interpretability and phishing detection performance. Simple models such as Decision Trees and Logistic Regression provide clearer explanations for predictions because their decision rules can be easily interpreted, but their detection accuracy is often lower compared to more advanced models [6],[18]. In contrast, ensemble models such as Random Forest achieve higher detection accuracy by combining multiple decision trees, although the decision-making process becomes more complex and harder to interpret [4]. In the proposed PhishGuard system, the Random Forest classifier achieved the highest accuracy among the evaluated models, demonstrating strong detection capability while sacrificing some level of interpretability [9].

Model	Accuracy
Random Forest	97%
SVM	94%
Logistic Regression	92%
Decision Tree	90%

Fig 3.1 : Accuracy Comparison of Machine Learning Models for Phishing Detection

The accuracy comparison of different machine learning models used in the proposed PhishGuard phishing detection system is illustrated in Figure 3.2. The results show that the Random Forest classifier achieved the highest accuracy of 97%, outperforming other algorithms such as Support Vector Machine (94%), Logistic Regression (92%), and Decision Tree (90%). The superior performance of Random Forest can be attributed to its ensemble learning capability,

which combines multiple decision trees to improve classification accuracy and reduce overfitting. These results indicate that ensemble-based models provide better detection performance for phishing website identification, making Random Forest a suitable choice for real-time phishing detection in the PhishGuard browser extension.

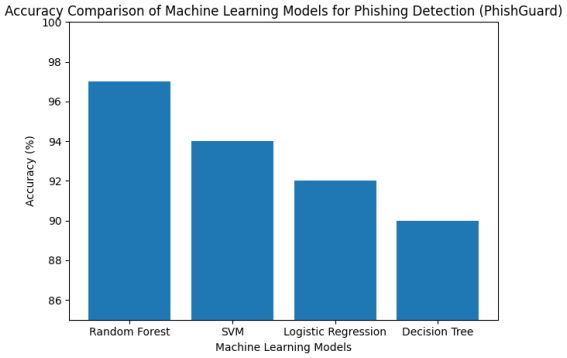

Fig 3.2 : Accuracy Comparison Graph of Machine Learning Models

The confusion matrix is commonly used in phishing detection research to evaluate the performance of machine learning models by comparing predicted and actual website classifications. It helps measure how accurately the model detects phishing and legitimate websites using values such as True Positives, True Negatives, False Positives, and False Negatives. Several studies have used confusion matrix analysis to assess the effectiveness of phishing detection systems [4],[9]. In the proposed PhishGuard system, the confusion matrix indicates that the model can accurately classify most phishing and legitimate websites, demonstrating reliable detection performance.

Actual/ Predicted	Predicted Phishing	Predicted Legitimate
Actual Phishing	95(True Positive)	5(False Negative)
Actual Legitimate		

Fig 3.3 : Confusion Matrix for Real-Time Phishing Detection System

D. Real-Time Phishing Detection in Browser

The integration of machine learning into cybersecurity systems provides practical solutions for protecting users against phishing attacks. Real-time phishing detection systems integrated into browsers can monitor visited URLs and immediately warn users when suspicious websites are detected [11],[2]. Browser-based detection mechanisms reduce the risk of user interaction with malicious websites and significantly improve online security during web browsing [16]. Such systems demonstrate the potential of machine learning models to provide proactive protection against phishing threats in real-world environments [27].

IV. CONCLUSION

This paper presents PhishGuard, a real-time phishing website detection system that combines machine learning techniques with a browser extension to enhance user security during web browsing. The proposed system analyzes various URL-based features and applies machine learning algorithms to classify websites as either phishing or legitimate. By training the model on known phishing patterns and datasets, the system is capable of identifying suspicious characteristics commonly associated with phishing attacks. Experimental evaluation demonstrates that the proposed approach achieves high detection accuracy and effectively distinguishes phishing websites from legitimate ones. The integration of the machine learning model into a browser extension enables real-time monitoring of visited websites and provides instant warnings whenever users attempt to access potentially malicious pages. This functionality helps prevent users from interacting with fraudulent websites and protects sensitive information such as login credentials, personal data, and financial details. The results indicate that combining machine learning methods with browser-based security mechanisms can significantly improve phishing detection and enhance user safety on the internet. However, phishing attacks continuously evolve as attackers develop new techniques to bypass existing defenses. Therefore, future work should focus on improving model robustness by incorporating advanced machine learning and deep learning techniques, expanding datasets, and analyzing additional website features to further improve detection accuracy and system performance.

REFERENCES

[1] M. S. Alzboon, M. S. Al-Batah, M. Alqaraleh, F. Alzboon, and L. Alzboon, "Phishing Website Detection Using Machine Learning," *Gamification and Augmented Reality*, vol. 3, p. 81, 2025.

[2] R. Alazaidah, A. Al-Shaikh, M. R. Al-Mousa, H. Khafajah, G. Samara, M. Alzyoud, N. Al-Shanableh, and S. Almatarneh, "Website Phishing Detection Using Machine Learning Techniques," *Journal of Statistics Applications & Probability*, vol. 13, no. 1, pp. 119–129, 2024.

[3] B. Sucharithal, B. Chandini, D. Satya Kumar, M. Surendra, and G. Kishor Kumar, "Detecting Phishing Websites Using Machine Learning," *International Journal of Advanced Research in Computer and Communication Engineering*, vol. 13, no. 4, Apr. 2024.

[4] B. Sucharithal, B. Chandini, D. Satya Kumar, M. Surendra, and G. Kishor Kumar, "Detecting Phishing Websites Using Machine Learning," *International Journal of Advanced Research in Computer and Communication Engineering*, vol. 13, no. 4, Apr. 2024.

[5] M. A. A. H. Qasim and N. A. Flayh, "Phishing Website Detection Using Machine Learning: A Review," *Wasit Journal for Pure Sciences*, vol. 2, no. 2, 2024.

[6] S. S. U. Siva Sakthiu, M. Pradeepa, D. Janani, and N. Gayathri, "Phishing Website Detection Using Machine Learning," *International Journal of Scientific Research in Engineering and Management*, vol. 8, no. 7, Jul. 2024.

[7] S. Malviya, "Phishing Detection Using ML Based URL Classification," *International Journal of Advances in Engineering and Management*, vol. 4, no. 3, Mar. 2022.

[8] Q. E. ul Haq, M. H. Faheem, and I. Ahmad, "Detecting Phishing URLs Based on a Deep Learning Approach to Prevent Cyber-Attacks," *Applied Sciences*, vol. 14, art. no. 10086, 2024.

[9] Y. Wei and Y. Sekiya, n"Sufficiency of Ensemble Machine Learning Methods for Phishing Websites Detection," *IEEE Access*, 2022.

[10] R. Zieni, L. Massari, and M. C. Calzarossa, "Phishing or Not Phishing? A Survey on the Detection of Phishing Websites," *IEEE Access*, 2023.

[11] W. Ali and S. Malebary, "Particle Swarm Optimization-Based Feature Weighting for Improving Intelligent Phishing Website Detection," *IEEE Access*, vol. 8, pp. 116761–116780, 2020.

[12] P. T. Duy, V. Q. Minh, B. T. H. Dang, N. D. H. Son, N. H. Quyen, and V. H. Pham, "A Study on Adversarial Sample Resistance and Defense Mechanism for Multimodal Learning-Based Phishing Website Detection," *IEEE Access*, 2024.

[13] I. Kara, M. Ok, and A. Ozaday, "Characteristics of Understanding URLs and Domain Names Features: The Detection of Phishing Websites With Machine Learning Methods," *IEEE Access*, 2022.

[14] P. Yang, G. Zhao, and P. Zeng, "Phishing Website Detection Based on Multidimensional Features Driven by Deep Learning," *IEEE Access*, 2019.

[15] Y. A. Kustiawan and K. I. Ghauth, "Feature Engineering for Phishing Website Detection Using Machine Learning: A Systematic Review," *IEEE Access*, 2025.

[16] S. Al-Ahmadi, A. Alotaibi, and O. Alsaleh, "PDGAN: Phishing Detection With Generative Adversarial Networks," *IEEE Access*, 2022.

[17] G. S. Nayak, B. Muniyal, and M. C. Belavagi, "Enhancing Phishing Detection: A Machine Learning Approach With Feature Selection and Deep Learning Models," *IEEE Access*, 2025.

[18] L. Tang and Q. H. Mahmoud, "A Deep Learning-Based Framework for Phishing Website Detection," *IEEE Access*, 2021.

[19] S. Ahmad, M. Zaman, S. Shamayleh, A. S. Al-Rahiel, S. M. Abdulhamid, I. Ergen, and A. Akhunzada, "Across the Spectrum: In-Depth Review AI-Based Models for Phishing Detection," *IEEE Communications Society*, 2024.

[20] A. S. Rafsanjani, N. B. Kamaruddin, M. Behjati, S. Aslam, A. Sarfaraz, and A. Amphawan, "Enhancing Malicious URL Detection: A Novel Framework Leveraging Priority Coefficient and Feature Evaluation," *IEEE Access*, 2024.

[21] D. J. Dsouza, A. P. Rodrigues, and R. Fernandes, "Multi-Modal Comparative Analysis on Execution of Phishing Detection Using Artificial Intelligence," *IEEE Access*, 2024.

[22] S. M. Somesha and A. R. Pais, "Classification of Phishing Email Using Word Embedding and Machine Learning Techniques," 2022.

[23] M. Sánchez-Paniagua, E. F. Fernández, E. Alegre, W. Al-Nabki, and V. González-Castro, "Phishing URL Detection: A Real-Case Scenario Through Login URLs," *IEEE Access*, 2022.

[24] J. H. Setu, N. Halder, A. Islam, and M. A. Amin, "RSTHFS: A Rough Set Theory-Based Hybrid Feature Selection Method for Phishing Website Classification," *IEEE Access*, 2025.

[25] M. J. Pillai, S. Remya, V. Devika, S. Ramasubbaredy, and Y. Cho, "Evasion Attacks and Defense Mechanisms for Machine Learning-Based Web Phishing Classifiers," *IEEE Access*, 2023.

[26] T. Wangchuk and T. Gonsalves, "Multimodal Phishing Detection on Social Networking Sites: A Systematic Review," *IEEE Access*, 2025.

[27] W. Li, S. U. A. Laghari, S. Manickam, Y.-W. Chong, and B. Li, "Machine Learning-Enabled Attacks on Anti-Phishing Blacklists," *IEEE Access*, 2024.