IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 13, April 2026
Impact Factor: 8.2

PhishNet: A Lightweight Machine Learning Browser Defense System for Real-Time Phishing Website Detection

Pranay Mahajan[1], Mansi Mhatre[2], Gauri Salvi[3], Omakr Yelgonda[4], Prof. Rashmi Mahajan[5]
Department of Artificial Intelligence & Machine Learning[1,2,3,4,5]
Shivajirao S. Jondhale College of Engineering, Dombivli (E), Maharashtra, India

Abstract: *Phishing attacks continue to be a major cybersecurity threat, causing financial fraud, credential theft, and data breaches [1], [2]. Traditional blacklist and rule-based systems are no longer effective against zero-day and sophisticated phishing attacks [3], [4].*
This paper reviews various phishing detection techniques, including heuristic approaches, machine learning, deep learning, and multimodal methods [5], [6], [7]. It analyzes URL-based detection, content-based features, CNN-based visual detection, and sequence modeling techniques, focusing on performance and scalability.
Key challenges such as dataset imbalance, lack of explainability, multilingual threats, and privacy concerns are also discussed [8], [9]. The study proposes future directions using lightweight models, explainable AI, and real-time browser-based deployment to enhance cybersecurity systems..

Keywords: PhishNet

I. INTRODUCTION

The rapid expansion of digital technologies, including online banking, e-commerce platforms, cloud computing, and social networking services, has significantly increased the global cyber-attack surface. Among the wide range of cyber threats, phishing attacks have emerged as one of the most persistent and damaging forms of social engineering. These attacks exploit human vulnerabilities by impersonating trusted entities such as financial institutions, government agencies, or popular online platforms to deceive users into disclosing sensitive information, including login credentials, credit card details, and personal identification data. As reported in recent studies, phishing attacks continue to grow both in volume and sophistication, causing substantial financial losses and compromising millions of user accounts worldwide [1], [2]. The increasing reliance on digital infrastructure has further amplified the impact of such attacks, making phishing detection a critical area of research in cybersecurity.

Traditionally, phishing detection relied heavily on blacklist-based systems and rule-based filtering techniques. Blacklist systems maintain a database of known malicious URLs and compare incoming web requests against this repository to block access to harmful websites. While these approaches are computationally efficient and easy to deploy, they suffer from significant limitations, particularly in detecting zero-day phishing attacks—newly created malicious websites that are not yet included in the blacklist [3], [4]. Similarly, rule-based systems depend on manually engineered heuristics such as URL length, presence of suspicious keywords, excessive subdomains, or abnormal redirection patterns. Although these techniques are lightweight and suitable for real-time applications, attackers can easily bypass them using obfuscation strategies, domain randomization, and advanced evasion techniques. Consequently, the effectiveness of traditional methods has declined in the face of rapidly evolving phishing tactics.

To address these challenges, researchers have increasingly turned to machine learning (ML) techniques for phishing detection. Machine learning models enable automated decision-making by learning patterns from historical data and identifying previously unseen threats. Common ML algorithms such as Support Vector Machines (SVM), Decision Trees, Random Forests, and Logistic Regression have demonstrated improved accuracy and generalization compared to

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 13, April 2026
Impact Factor: 8.2

traditional methods [5], [6]. These models typically rely on feature extraction from URLs, HTML content, domain registration details, and network traffic patterns. By leveraging these features, ML-based systems can detect suspicious behavior even in previously unseen phishing websites. However, despite their advantages, machine learning approaches still depend heavily on handcrafted feature engineering, which requires domain expertise and may not capture complex patterns inherent in sophisticated phishing attacks. Additionally, issues such as dataset imbalance, overfitting, and limited adaptability to new attack strategies can affect model performance.

With advancements in artificial intelligence, deep learning (DL) techniques have gained significant attention in phishing detection research. Deep learning models, particularly Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), have the ability to automatically learn hierarchical feature representations from raw data, eliminating the need for manual feature extraction. CNN-based approaches are widely used for visual phishing detection, where webpage screenshots are analyzed to identify similarities with legitimate websites, making them effective against visually deceptive attacks such as website cloning [7], [8]. On the other hand, RNN and Long Short-Term Memory (LSTM) models are well-suited for analyzing sequential data, such as URL strings, enabling detection of obfuscated or dynamically generated phishing URLs. Despite achieving high accuracy, often exceeding 95%, deep learning models introduce challenges related to computational complexity, high training time, and resource requirements, which can limit their deployment in real-time systems and resource-constrained environments.

More recently, multimodal and hybrid approaches have been proposed to further enhance phishing detection capabilities. These systems combine multiple types of features, including URL-based, content-based, visual, and behavioral attributes, to provide a more comprehensive analysis of websites. By integrating diverse data sources, multimodal models can improve robustness against different types of phishing attacks and reduce false positives [9], [10]. However, the increased complexity of such systems presents new challenges, including higher computational overhead, difficulty in feature fusion, and lack of standardized frameworks for integrating multiple modalities. Additionally, real-world deployment of these advanced models requires careful consideration of latency, scalability, and privacy concerns, particularly when dealing with sensitive user data.

Despite significant advancements in phishing detection techniques, several critical research gaps remain unresolved. One of the major challenges is the detection of zero-day phishing attacks, where attackers continuously generate new domains and employ novel evasion strategies. Furthermore, adversarial attacks on machine learning and deep learning models pose a serious threat, as attackers can manipulate input data to deceive classifiers without affecting human perception. Dataset-related issues, such as class imbalance, outdated samples, and limited multilingual coverage, also hinder the generalization capability of detection models. Another important concern is the lack of explainability in deep learning systems, which operate as black boxes and provide limited insight into their decision-making process. This lack of transparency can reduce trust and hinder adoption in critical cybersecurity applications.

In addition to technical challenges, practical deployment constraints must also be addressed. High-performing deep learning models often require significant computational resources, including GPU acceleration, making them unsuitable for deployment in browser extensions, mobile devices, or edge computing environments. Privacy concerns also arise when sensitive webpage data is transmitted to cloud-based servers for analysis. Emerging solutions such as federated learning and on-device inference aim to address these issues by enabling decentralized model training and preserving user privacy. However, these approaches are still in the early stages of research and require further exploration to achieve practical implementation.

In this context, this paper aims to provide a comprehensive review of phishing website detection techniques, covering traditional methods, machine learning approaches, deep learning models, and multimodal frameworks. The study systematically analyzes the strengths and limitations of each approach, focusing on key factors such as detection accuracy, computational complexity, scalability, robustness, and deployment feasibility. Furthermore, it identifies critical research gaps and proposes future directions for developing lightweight, explainable, and real-time phishing detection systems. By integrating advanced AI techniques with practical deployment strategies, this research

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 13, April 2026
Impact Factor: 8.2

contributes to the development of next-generation cybersecurity solutions capable of addressing the evolving landscape of phishing attacks.

II. REVIEW METHODOLOGY

This review adopts a systematic approach to analyze existing phishing detection techniques and identify research gaps in cybersecurity. The methodology includes literature selection, classification of approaches, performance evaluation, and gap analysis. Relevant research papers were collected from trusted sources such as IEEE Xplore, Springer, and arXiv using keywords like "phishing detection" and "machine learning in cybersecurity" [1], [3]. Recent studies and survey papers were prioritized to ensure up-to-date insights [13], [5].

The selected studies are categorized into four main groups: (i) blacklist and heuristic-based methods, (ii) machine learning approaches, (iii) deep learning techniques, and (iv) multimodal systems. Traditional blacklist methods provide fast detection but fail against zero-day attacks [6], while machine learning models such as SVM and Random Forest improve generalization and detection accuracy [9], [10]. Deep learning models like CNN and LSTM automatically extract features and achieve high performance [2], [7], and multimodal approaches further enhance robustness by combining multiple data sources [17].

For comparative analysis, studies are evaluated using metrics such as accuracy, precision, recall, F1-score, computational complexity, and real-time applicability [8]. Dataset quality is also considered, as many studies suffer from issues like imbalance and outdated data, affecting model performance [5], [18]. Additionally, the methodology examines challenges such as lack of explainability, adversarial vulnerability, and privacy concerns in existing systems [16].

Overall, this methodology provides a structured framework to evaluate phishing detection techniques and highlights the need for lightweight, explainable, and real-time solutions for future research.

III. LIMITATIONS OF EXISTING TECHNIQUES

Despite significant advancements in phishing detection, existing techniques still face several limitations that affect their real-world effectiveness. Traditional blacklist-based systems, although fast and easy to implement, are ineffective against zero-day phishing attacks, as they rely on previously identified malicious URLs [3], [4]. Since attackers frequently generate new domains and short-lived phishing websites, these systems fail to provide timely protection. Similarly, heuristic and rule-based approaches depend on predefined patterns such as URL length and suspicious keywords, making them vulnerable to evasion techniques like URL obfuscation and domain randomization [6].

Machine learning-based techniques improve detection accuracy by learning patterns from data, but they are not without challenges. These models heavily depend on handcrafted feature extraction, which may not capture complex and evolving phishing strategies [9], [10]. Additionally, issues such as dataset imbalance and overfitting can reduce the generalization capability of models, especially when dealing with real-world data [5]. Another limitation is that most machine learning models require continuous retraining to adapt to new phishing patterns, increasing maintenance complexity.

Deep learning approaches, including CNN and LSTM models, have demonstrated high accuracy in phishing detection by automatically extracting features from raw data. However, these models introduce significant computational complexity and require large labeled datasets for effective training [7], [8]. Their high resource requirements make them difficult to deploy in real-time environments such as browser extensions and mobile devices. Furthermore, deep learning models often act as black-box systems, lacking interpretability and making it difficult to understand the reasoning behind predictions [16].

Multimodal and hybrid techniques, which combine URL, content, and visual features, offer improved robustness but come with increased architectural complexity and higher computational overhead [17]. Integrating multiple data sources requires efficient feature fusion strategies, which are still not standardized. Additionally, these systems may

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 13, April 2026 Impact Factor: 8.2

face scalability issues when deployed in large-scale environments. Privacy concerns also arise in cloud-based detection systems, where sensitive user data must be processed externally, posing risks to data security.

Overall, while existing phishing detection techniques have achieved notable progress, they still struggle with challenges such as zero-day detection, computational cost, lack of explainability, dataset limitations, and real-time deployment constraints. Addressing these limitations is essential for developing more robust, scalable, and practical phishing detection systems in the future.

IV. RESEARCH GAP

Despite the rapid advancement of phishing detection techniques, several critical research gaps remain that limit their effectiveness in real-world applications. One of the major challenges is the detection of zero-day phishing attacks. Most existing systems rely on historical data for training, making them less effective when encountering newly generated phishing websites with unseen patterns [3], [4]. This highlights the need for adaptive and intelligent models capable of identifying anomalies without depending entirely on labeled datasets.

Another significant gap lies in the lack of robustness against adversarial attacks. Modern phishing detection models, particularly deep learning-based systems, are vulnerable to small perturbations in input data, such as slight modifications in URLs or webpage content, which can mislead the classifier without affecting human perception [7], [8]. Current research has limited focus on adversarial training and defense mechanisms, making systems susceptible to manipulation by attackers.

Dataset-related limitations also present a major research gap. Many existing studies use static and imbalanced datasets, where legitimate samples significantly outnumber phishing instances [5], [18]. This imbalance leads to biased models with reduced detection accuracy for phishing cases. Additionally, there is a lack of large-scale, up-to-date, and multilingual datasets, which restricts the ability of models to generalize across different regions and languages. The absence of standardized benchmark datasets further complicates fair comparison between different approaches.

Explainability and interpretability remain underexplored areas in phishing detection research. Most deep learning models operate as black-box systems, providing high accuracy but little insight into their decision-making process [16]. In cybersecurity applications, understanding why a website is classified as phishing is crucial for trust, transparency, and regulatory compliance. The integration of Explainable AI (XAI) techniques is still limited and requires further investigation.

Another important gap is related to real-time deployment and computational efficiency. High-performing deep learning and multimodal models often require significant computational resources, making them difficult to deploy in browser extensions, mobile devices, or edge environments [17]. There is a need for lightweight and optimized models that can maintain high accuracy while ensuring low latency and efficient resource utilization.

Privacy and data security concerns also remain insufficiently addressed. Many cloud-based phishing detection systems require user data to be transmitted to external servers for analysis, raising concerns about data leakage and confidentiality. Emerging approaches such as federated learning and on-device inference show promise but are still in early stages of implementation and require further research.

Overall, these research gaps highlight the need for developing next-generation phishing detection systems that are adaptive, explainable, lightweight, and privacy-preserving. Addressing these challenges will be essential for creating robust and scalable solutions capable of handling the evolving nature of phishing attacks in real-world scenarios.

V. CONCLUSION

Phishing detection has evolved from traditional blacklist and heuristic-based methods to advanced machine learning and deep learning approaches. While modern techniques achieve high accuracy, they still face challenges such as zero-day attack detection, computational complexity, lack of explainability, and dataset limitations [3], [7], [16]. This study highlights the need for lightweight, adaptive, and privacy-preserving models that can operate efficiently in real-time

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 13, April 2026
Impact Factor: 8.2

environments. Future research should focus on improving robustness, scalability, and interpretability to develop more reliable phishing detection systems [17].

VI. SUMMARY

This paper presents a comprehensive review of phishing website detection techniques, highlighting the evolution from traditional methods to advanced artificial intelligence-based approaches. Phishing attacks remain one of the most critical cybersecurity threats, causing financial losses and compromising sensitive user information [1], [2]. Early detection techniques, such as blacklist-based and heuristic approaches, provided fast and lightweight solutions but were limited in their ability to detect zero-day and sophisticated phishing attacks [3], [4].

With the advancement of machine learning, phishing detection systems have become more adaptive and capable of identifying patterns in data. Techniques such as Support Vector Machines, Random Forest, and Logistic Regression have shown improved detection accuracy and generalization compared to traditional methods [9], [10]. However, these approaches rely heavily on handcrafted feature extraction and may struggle to detect complex phishing strategies or visually deceptive websites.

Deep learning techniques have further enhanced phishing detection by automatically learning hierarchical features from raw data. Models such as Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) have demonstrated high performance in detecting phishing websites through visual analysis and sequential pattern recognition [7], [8]. Despite their advantages, deep learning models require large datasets, high computational resources, and often lack interpretability, making their deployment in real-time systems challenging [16].

To overcome individual limitations, multimodal approaches have been introduced, combining URL-based, content-based, and visual features to improve detection robustness and reduce false positives [17]. While these systems achieve better performance, they also introduce increased complexity, higher computational cost, and challenges in feature integration. Additionally, issues such as dataset imbalance, lack of standardized benchmarks, and limited multilingual coverage continue to affect model reliability [5], [18].

The review also identifies several critical research gaps, including zero-day attack detection, adversarial robustness, explainability, and privacy concerns. Existing systems often fail to adapt to evolving phishing techniques and may be vulnerable to adversarial manipulation. Furthermore, real-time deployment remains a challenge due to latency and resource constraints.

In conclusion, phishing detection research has made significant progress, but further advancements are required to develop scalable, efficient, and secure systems. Future work should focus on lightweight models, explainable AI, privacy-preserving techniques, and real-time integration to enhance the effectiveness of phishing detection in dynamic and real-world environments.

VII. ACKNOWLEDGEMENT

We would like to express our sincere gratitude to our guide, Prof. Rashmi Mahajan (Project Guide), for her invaluable guidance, continuous support, and constructive suggestions throughout the course of this research work. Her insights and encouragement played a crucial role in the successful completion of this project. We are deeply thankful to Dr. Renuka Deshpande (Head of Department, AIML), Shivajirao S. Jondhale College of Engineering (SSJCOE), for providing the academic support, resources, and motivation necessary for carrying out this work effectively. We also extend our heartfelt appreciation to Shivajirao S. Jondhale College of Engineering (SSJCOE), Dombivli, for offering a conducive learning environment and the required facilities that enabled us to conduct this research successfully.

REFERENCES

[1] M. R. Ahmed, M. M. Islam, and M. A. Layek, “Phishing URL Detection Using Comprehensive Feature Extraction and Machine Learning Techniques,” in IEEE CS BDC Symposium, 2024.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 13, April 2026 Impact Factor: 8.2

[2] S. Aslam, H. Aslam, A. Manzoor, C. Hui, and A. Rasool, "AntiPhishStack: LSTM-Based Stacked Generalization Model for Optimized Phishing URL Detection," arXiv, Jan. 2024.
[3] A. U. Rehman, I. Imtiaz, S. Javaid, and M. Muslih, "Real-Time Phishing URL Detection Using Machine Learning," Engineering Proceedings, vol. 107, no. 1, 2025.
[4] I. Altan et al., "Dual-Path Phishing Detection: Integrating Transformer-Based NLP with Structural URL Analysis," arXiv, Sep. 2025.
[5] M. A. Tamal, M. K. Islam, T. Bhuiyan, and A. Sattar, "Dataset of Suspicious Phishing URL Detection," Frontiers in Computer Science, 2024.
[6] S. Author, "Detection of Malicious URLs Using Machine Learning," Wireless Networks, vol. 30, pp. 7543–7560, Mar. 2024.
[7] E. A. Aldakheel et al., "A Deep Learning-Based Innovative Technique for Phishing Detection with URLs," Sensors, vol. 23, no. 9, 4403, 2023.
[8] Q. E. u. Haq, M. H. Faheem, and I. Ahmad, "Detecting Phishing URLs Based on a Deep Learning Approach to Prevent Cyber-Attacks," Applied Sciences, vol. 14, no. 22, 10086, 2024.
[9] A. A. Albishri and M. M. Dessouky, "A Comparative Analysis of Machine Learning Techniques for URL Phishing Detection," ETASR, vol. 14, no. 6, Dec. 2024.
[10] D. R. Patil, R. B. Wagh, V. D. Punjabi, and S. M. Pardeshi, "Enhanced Phishing URLs Detection Using Feature Selection and Machine Learning Approaches," IJWMT, vol. 14, no. 6, Dec. 2024.
[11] L. Breiman, "Random Forests," Machine Learning, vol. 45, pp. 5–32, 2001.
[12] G. Salton and C. Buckley, "Term-Weighting Approaches in Automatic Text Retrieval," Information Processing & Management, 1988.
[13] M. S. Gupta, A. Tewari, and A. Jain, "Survey on Phishing Detection Methods: Machine Learning and AI," International Journal of Advanced Research in Computer Science, 2023.
[14] Rahul R. and S. Sundar, "A Comprehensive Survey on Machine Learning Techniques for Phishing Detection," International Journal of Computer Applications, 2023.
[15] T. Narayana et al., "Improving Phishing Website Detection with Machine Learning: Hidden Patterns for Better Accuracy," IJRITCC, 2023.
[16] A. Systematic Review Authors, "A Systematic Review on Deep-Learning-Based Phishing Email Detection," Electronics, 2023.
[17] M. S. I. Ovi, M. H. Rahman, and M. A. Hossain, "PhishGuard: A Multi-Layered Ensemble Model for Optimal Phishing Website Detection," arXiv, 2024.
[18] T. A. Tamal et al., "Dataset of Suspicious Phishing URL Detection," Frontiers in Computer Science, 2024.
[19] R. G. M. Helali, "Phishing Detection Using Hybrid Machine Learning Techniques," Zhongguo Kuangye Daxue Xuebao, 2024.
[20] T. Bishtawi, R. Alzubi, and H. Kassem, "Improving Web Security through Machine Learning: Feature-Based Methodology for Detecting Phishing URLs," ETASR, 2025.
[21] APWG, "Phishing Activity Trends Report," 2023.