International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 05 , May - 2026

Review on Intelligent Detection Frameworks for Phishing Websites using Machine Learning

Palak Panchal
M. Tech Scholar
Department of Information Technology
Mahakal Institute of Technology
Ujjain, India

Riya Verma
Assistant Professor
Department of Information Technology
Mahakal Institute of Technology
Ujjain, India

Abstract- Phishing attacks are one of the most common and dynamic cybercrimes, which target people, organizations, and important digital systems. The conventional rule-based and blacklist-based detection systems are no longer sufficient to counter the more advanced and constantly evolving phishing schemes. The objective of this review paper is to conduct a systematic study of machine learning (ML)-based phishing web detectors with the emphasis made on critical feature engineering methods, the assessment of popular machine learning and deep learning (DL) models, and the deployment of cybersecurity systems in real-time. The research is presented using the methodology of a structured literature review to summarize modern research findings on phishing detection in 2020-2026. It classifies the existing strategies according to the type of features such as URL/lexical, domain-based, HTML/content, visual, behavioral and hybrid features. In addition, it revises classic supervised learning models, ensemble, deep learning architectures, and hybrid frameworks, as well as typical evaluation metrics and optimization policy. A comparative study of the chosen studies is carried out to outline the tendencies in the methodologies, strengths, and weaknesses of the research and present the gaps.The review shows that there is an increasing trend towards ensemble, multimodal, and explainable AI-based models, which are paying more attention to scalability, adaptability, and real-time applicability. It further highlights the enduring problems of adversarial robustness, high-dimensionality, dataset imbalance, and efficiency. The article creates a unified base on what can be done to advance the research in the future to create the adaptive, interpretable, and scalable phishing detection systems using the ML.

Keywords : Machine Learning, Phishing Website Detection, Feature Engineering, Deep Learning, Cybersecurity Analytics

I. INTRODUCTION

The rapid increase in digital transformation has vastly widened the attack surface in the world, and cybercriminals are able to use vulnerabilities in web-based ecosystems. Phishing is one of the most insidious and harmful types of cybercrime in which, through the impersonation of valid individual, malicious users/hackers can steal sensitive data, login access, and other personal information. Organizations like APWG and Verizon have continuously reported a significant escalation in the quantity and complexity of phishing campaigns, compelling a sense of urgency in regard to the use of smart and reactive detection systems. The old methods of rules and blacklists, which have proven useful in curbing attacks, which were initially known, find it hard to identify newly created or obfuscated phishing sites that keep on changing in format and strategy [1], [2], [3], [4]. This dynamic threat environment makes the implementation of advanced computational methods with the ability to learn intricate patterns and generalize among unknown forms of attacks attractive.Machine learning (ML) has been found as an influential paradigm in cybersecurity, with the capacity to identify patterns automatically, scale, and be flexible. Because as opposed to the traditional detection systems, the ML-based models are capable of identifying various characteristics based on the Uniform Resource Locators (URLs), domain registration metadata, and Hypertext Markup Language (HTML) source code to differentiate between a legitimate and a malicious web page. Using supervised learning methods that are trained on labeled datasets, such systems are capable of identifying delicate correlations and nonlinear connections that would not be readily known with the help of manual inspection or signature-based detection. Moreover, ensemble learning algorithms, kernel-based classifiers, and neural networks are combined to increase resistance to a single perspective of prediction [5], [6].

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 05 , May - 2026

Fig. 1 Phishing Detection[7]

An efficient phishing recognition system can be achieved through the methodical feature engineering, stringent model evaluation, and experimental validation. Experimental environments Benchmark datasets like the UCI Machine Learning Repository Phishing Websites Dataset offer a standard experimental environment in which to assess the performance of an algorithm [8]. By providing extensive investigation into discriminative aspects-such as URL length anomalies, the occurrence of suspicious tokens, atypical domain age and inconsistencies between embedded resources-researchers are able to discover patterns that have a high likelihood of being related to phishing behaviour. Also, hyperparameter optimization, cross-validation, and feature importance analysis as model optimization techniques can be used to achieve better generalization and interpretability [9].

Even with this tremendous progress, some research issues have not been addressed. Detection of the zero-day phishing attacks, low-false positive rates, computational efficiency to implement in real-time are essential issues in an actual cybersecurity setting. Additionally, since attackers are rapidly using automated phishing kits and AI-generated content, the detection systems also need to change to be resistant to adversarial manipulation. A scalable ML-based system should thus be configured to strike a balance between predictive performance and operational effectiveness so as to ensure a smooth blend into browser security controls, enterprise threat management systems and automated email protection systems.

This review article measures the design, comparative analysis and implementation of machine learning-based frameworks to detect phishing websites. The proposed study will synthesize already developed methodologies, feature engineering solutions, and algorithm development capabilities to bring a complete picture of how ML can be used to improve proactive defense against advanced phishing schemes. Finally, the study will help in the creation of adaptable, explainable, and real-time cybersecurity solutions that can alleviate one of the most common digital security risks in the digitalized world.

II. RELATED WORK

Dandotiya et al., 2026[10] provides a prototype of a browser extension to support machine learning that inspects URLs and visual elements in Google Chrome to detect phishing sites on-the-fly. The proposed system collects and analyzes the data of the websites with the support of the Support Vector Machine (SVM), Decision Tree (DT), and Random Forest (RF) algorithms, derives hybrid components such as lexical, structural, and visual layout parameters, and organizes them. The most desirable characteristics, which are used to differentiate between things, are discovered with the help of the Grey Wolf Optimizer (GWO). This minimizes the power usage of the computers and it makes locating items easier. GWO improved the Random Forest model that was effective on the benchmark data set including the Berkeley ML Archives and PhishTank. The extension uses the method to score URLs on visual similarity in real time on the MCC test with a score of 0.96 and a 98.7 accuracy rate, and can address URLs that have been obfuscated.The proposed system is superior to existing anti-phishing solutions as it works in real time, its false-positive rate is lower, and can work with obfuscated URLs. The current project results in a handy, user oriented defense system that is able to defend against phishing attacks that evolve with the times using smart security on the browser-level.

Ari Kustiawan & Imran Ghauth, 2025 [11] steps into this gap and examines recent research on feature engineering in the detection of phishing URLs with machine learning. The review subtypes feature types including URL/lexical, HTML/content, domain, behavioral and emerging derived and hybrid features and offers an analysis of the use of feature selection and dimensionality reduction methods. The results indicate that URL/lexical features continue to be used due to their simplicity and wide applicability even with a recent shift to using content-based and composite methods in an effort to enhance strength. Among the important issues are the issues of keeping up with changing phishing tricks, handling the limitations of datasets, high dimensionality mitigation, and the balance between computational performance and accuracy of detection. Other opportunities that have been underutilized in the previous literature that are noted in the review include the use of explainable AI (e.g., SHAP, LIME) to validate and optimize feature sets, extending cross-lingual and adversarial-resilient features, and creating lightweight solutions in real-time or resource-constrained settings. This paper provides a specific roadmap to the further research in phishing detection by summarizing and examining the purpose of feature engineering. It

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 05 , May - 2026

underscores its significant contribution in designing more precise, adaptive and effective detectors.

Reddy Bonikela & Goel, 2025[12] The traditional machine-learning techniques using hand-crafted features like URLs and domain metadata have been effective, however, they fail to support the dynamic and adaptive characteristics of phishing attacks. The most recent trends in deep learning (convolutional neural networks (CNNs), recurrent neural networks (RNNs), and generative adversarial networks (GANs)) have demonstrated promise of improving phishing detection by learning more complicated patterns using raw data like webpage displays, user actions, etc. Regardless of these developments, however, there are several gaps in the research that are present. Real- time detection is also an acute issue since the existing systems are full of latency and resource constraints especially in mobile and low-resource conditions. Phishing sites also keep evolving, using more advanced obfuscation methods to make them look more like a legitimate site, and it is becoming even harder to keep up with them using standard mechanisms. There are also no skillful models that have the capacity to generalize across multiple areas including social media, e-commerce, and IoT platforms. Another weakness is a shortage of privacy preserving models that are sensitive to strike a balance between detection and security of user data. Such gaps, among others, need to be filled in by future studies, which need to seek to come up with light-weight, real time detection systems, founded on transfer learning within cross-domain applicability, and with privacy- preserving frameworks. Moreover, there is an opportunity to develop the process of multimodal learning, reinforcement learning, and adversarial training, which can increase the chances of gaining robustness and adaptability to adaptive phishing strategies. The paper will fill the abovementioned gaps and provide the possible solutions to the improvement of the effectiveness of web-based phishing and cyberattack detection systems.

Kumar & Kollwitz, 2025[13] examines how machine learning (ML)-based cybersecurity models can be used in protecting the e-commerce ecosystem through real-time detection, prevention, and response to cyber attacks. Supervised learning, unsupervised anomaly detection and deep learning algorithms are machine learning techniques that can be applied to large volumes of data produced by users behavior, transaction trends and network traffic. These models recognize anomalies in normal activities and can learn using the past data to recognize malicious activities with a high level of accuracy and minimum human input. The proactive solution to cybersecurity presented by ML-driven systems is far better than the old-fashioned unchanging defenses by responding to new threats patterns. Moreover, the models provide the ability to detect fraud in payment processing, attempts to take over accounts, and fake listing of products and promote a more secure online purchasing experience both to the consumer and the merchant. The area of the incorporation of ML cybersecurity models with real-time monitoring systems and incident response frameworks has been also covered in this paper. It puts emphasis on data quality, feature engineering and model explainability to obtain successful and reliable threat detection. Furthermore, it discusses the problems of adversarial attacks on ML models and stresses the necessity of strong and secure practices of model training and deployment.

Ali et al., 2025 [14] proposed remedies with the help of ML and DL-based security are summed up in adaptive threat detection, anomaly-based intrusion prevention, and smart risk mitigation. We also compared various solutions dependent on ML and DL in order to detect and stop cyber-attacks as an efficient solution. These ML and DL based research papers are reviewed in the IEEE repository and are published in the years between 2020 and 2024 and are up to date with the literature on IOT security. The findings demonstrate that ML and DL security models enhance resilience to IOT because they provide the ability to detect attacks in real-time, minimize the number of false positives, and respond to emerging threats. More so, this piece of work determines the current obstacles to adoption of ML/DL technologies in IOT security and highlights the possible future research directions that can consolidate the entire security platform of IOT ecosystems.

III. FEATURE ENGINEERING TECHNIQUES FOR PHISHING WEBSITE DETECTION

The machine learning-based phishing detection systems are mainly based on feature engineering because the discriminative capability of the extracted features directly correlates with the classification accuracy and resilience. In this section, the review of the key categories of features used in phishing websites detection frameworks is conducted [15], [16], [17], [18], [19].

A. URL and Lexical Features

URL and lexical features examine the structural patterns of web addresses. The length of the URL, special characters, excessive number of subdomains, use of IP addresses rather than domain names, suspicious keywords, and abnormal redirection patterns are some of the attributes that can be used to determine the intent of phishing. Artificial intelligence is used to calculate these features, which are simple and can be applied in real-time because of their applicability and efficiency in computation [20].

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 05 , May - 2026

B. Domain-Based Features

Domain-based features analyze registration and hosting data that is received in WHOIS records and DNS databases. The parameters such as the age of the domain, credibility of the registrar, DNS stability and location of hosting anomalies are used in the identification of short lived or suspicious domains that are often linked with phishing campaigns [21].

C. HTML and Content-Based Features

The HTML and content-based characteristics pay attention to the source code of webpages. Such signs include irregular form activities, concealed iframes, in-built JavaScript operations, discrepant hyperlinks and loading external resources patterns are indicators of deceptive design tactics employed by attackers[22].

D. Visual and Layout Features

Visual and layout attributes evaluate the look of web pages and structural resemblance to the authentic web pages. They can be logo similarity detection, CSS structure similarity detection, analysis based on screenshot, and matching webpage templates to identify impersonation attempts [9].

E. Behavioral and Hybrid Features

Behavioral and hybrid features combine several data sources, such as user interaction patterns, traffic abnormalities, and inter-feature correlations. Both of these representations enhance resistance to obfuscation and changing phishing-attacks.[23].

F. Feature Selection and Dimensionality Reduction Techniques

In order to increase the efficiency and interpretability of the model, feature selection and dimensionality reduction algorithms, including Grey Wolf Optimizer (GWO), Principal Component Analysis (PCA), SHAP, and LIME are utilized. These methods minimise redundancy, enhance generalisation, and maximise computation time [24].

G. Challenges in Feature Engineering

Nonetheless, there are still a lot of challenges such as high dimensionality, imbalanced data, manipulation of adversarial features, dynamic phishing patterns, and the presence of problems with cross-domain adaptability. These are the challenges to be tackled to construct scalable and resilient phishing detection systems.

IV. MACHINE LEARNING AND DEEP LEARNING MODELS FOR PHISHING DETECTION

In this section, the analysis of machine learning (ML) and deep learning (DL) algorithms in phishing websites classification will be provided thoroughly. It builds on the topic of feature engineering by discussing model development, comparative analysis, and the possibility of its deployment to real-world cybersecurity settings [24].

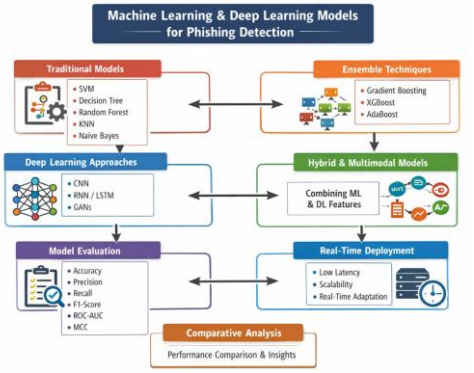

Fig. 2. Machine Learning and Deep Learning Models for Phishing Detection (Compiled by the Author)

A. Traditional Supervised Learning Models

The research of phishing detection is based on traditional supervised learning algorithms. Support Vector machine (SVM), Decision Tree (DT), Random Forest (RF), K-Nearest neighbors (KNN), and Naive Bayes (NB) are some of the methods that have gained popularity because they are easy to interpret and computationally efficient. SVM is successful when dealing with high-dimensional feature space, whereas Decision Trees offer transparency in the form of a rule. Random Forest is also stronger with ensemble aggregation, and KNN is good in classification via similarity. These models are specially applicable whereby structured URL, domain and HTML properties are present [25].

B. Ensemble Learning Techniques

Ensemble learning techniques enhance predictive performance by pooling a group of weak learners to a more powerful model. Gradient Boosting, XGBoost, and AdaBoost are methods of optimization of classification limits, which are iterative and minimize bias and variance. The models are also more effective at benchmark phishing data sets since they can identify nonlinear interactions of features and multifaceted decision patterns [25], [26], [27], [28], [29].

C. Deep Learning Approaches

Deep learning models have become prominent due to the ability to acquire hierarchical representations using raw inputs. Convolutional Neural Networks (CNNs) prove useful in the detection of visual similarities and analysis of screenshots of webpages. The recurrent neural networks (RNNs) and Long short term memory (LSTM) networks are those that learn sequential dependencies in the URL strings and script behaviors. Generative Adversarial Networks (GANs) are also being investigated in the context of

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 05 , May - 2026

adversarial resistiveness and synthetic phishing detection [30], [31], [32].

D. Hybrid and Multimodal Architectures

Hybrid frameworks combine the traditional ML classifiers with the deep neural networks in order to exploit both handcrafted and automatically learned features. Multimodal systems integrate lexical, content-based, behavioral, and visual inputs, which increase their resistance to advanced and obfuscated phishing methods [33], [34], [35].

E. Model Evaluation Metrics

The performance is measured by metrics of Accuracy, Precision, Recall, F1-score, Receiver Operating Characteristic- Area Under Curve (ROC-AUC), and Matthews Correlation Coefficient (MCC). These measures will guarantee proportional evaluation especially in unequal phishing samples [36].

F. Comparative Performance Analysis

Comparative studies indicate that the ensemble and deep learning models are typically more accurate and more robust than the single classifiers. Nevertheless, there are a trade-offs between interpretability and the complexity of the computation, so the careful choice of models is required [37].

G. Real-Time Deployment and Scalability

In order to have a practical cybersecurity integration, there has to be models that guarantee low latency, scalability and that it is adaptive to the emerging threats. Browser extensions, enterprise firewalls, and real-time monitoring systems are highly dependent on lightweight architectures, optimized inference pipelines, and incremental learning mechanisms [20], [38].

V. COMPARATIVE ANALYSIS

This section is a comparison of the recent machine learning and deep learning in the context of phishing detection, dataset analysis, feature strategies, model performance, applicability in the real time, and limitation of research to identify trends and existing gaps.

TABLE I. COMPARATIVE ANALYSIS

Author (s) & Year	Focus Area	Techniques / Models Used	Dataset & Features	Key Results	Major Contribution / Limitation
Jackson (2025) [39]	ML-based phishing email and website detection	Decision Tree, SVM, Neural Networks	Public phishing datasets; email content, sender behavior, user interaction features	High accuracy with reduced false positives; effective real-time integration potential	Demonstrates ML superiority over traditional methods; limited quantitative benchmarking comparison provided
Mahmud et al. (2025) [40]	Malicious URL detection using ML and DL comparison	LR, SVM, DT, KNN, GNB, RF, XGBoost, LightGBM; LSTM, BiLSTM, GRU; Stacking Model	URL-based features; benchmark malicious URL datasets	Traditional ML achieved up to 92%; DL up to 91%; Stacking model reached 99.99% accuracy	Shows ensemble stacking significantly improves performance; computational complexity not deeply analyzed
Zara et al. (2024) [1]	Phishing website detection using ensemble and DL models	Ensemble Learning, Deep Learning; Feature Selection (Information Gain, Gain Ratio, PCA)	11,055 website dataset; selected optimal features	Ensemble model achieved 99% accuracy	Combines feature selection with ensemble learning effectively; scalability discussion limited
Kamrul Hasan Chy & Nana Buadi (2024) [2]	Real-time phishing detection via browser extension	ML-driven real-time risk scoring system	Website behavior, network traffic, user interaction data	Real-time alerts and fraud detection; practical deployment focus	Strong real-time usability focus; lacks detailed performance metrics comparison
Tran & Sovilj (2024) [6]	Multi-class malicious website classification	ML models with 77 engineered features	441,701 samples; 9-class website classification; URL, content, DNS, host features	95.89% accuracy; feature subset impact analyzed	Large-scale dataset and feature importance ranking; computational overhead may be high

VI. CONCLUSION

The process of phishing detection has seen a massive change with the incorporation of machine learning as well as deep learning systems which has provided a revolutionary move of a fixed and rule oriented security systems to adaptive and data oriented defense systems. This has reviewed literature on feature engineering approaches, conventional supervised models, ensemble learning approaches, deep learning frameworks, and hybrid systems used in the detection of phishing websites. The discussion indicates that the basis of

successful detection system is still feature engineering and URL, domain, content, visual, and behavioral features combined to make models more robust. The review also highlights the increasing focus on multimodal and ensemble architectures, multi-feature representations and learning paradigms in order to enhance their flexibility to a variety of advanced and dynamically changing phishing strategies. Meanwhile, practical deployment factors, including computational performance, scaling, usability, and real-time reactiveness continue to be crucial to the successful implementation into browser extensions, enterprise security frameworks, and automatic threat examination systems. Although significant progress has been made, such issues as adversarial manipulation, unequal distribution of data, the detection of zero-day attacks, and cross-domain generalization still hamper complete operational efficiency.

The future of AI integration should thus be directed towards explainable AI integration, lightweight architectures, privacy-preserving frameworks and adaptive learning. Altogether, ML-based phishing detection systems take shape of a rather hopeful and scalable avenue as to boosting active cyber security defenses in an ever-expanding digital landscape.

REFERENCES

[1] U. Zara, K. Ayyub, H. Ullah Khan, A. Daud, T. Alsahfi, and S. Gulzar Ahmad, “Phishing Website Detection Using Deep Learning Models,” *IEEE Access*, vol. 12, no. October, pp. 167072–167087, 2024, doi: 10.1109/ACCESS.2024.3486462.

[2] M. Kamrul Hasan Chy and O. Nana Buadi, “A Machine Learning-Driven Website Platform and Browser Extension for Real-Time Risk Scoring and Fraud Detection for Website Legitimacy Verification and Consumer Protection,” *J. Multidiscip. Eng. Sci. Technol.*, vol. 11, no. 10, pp. 2458–9403, 2024, [Online]. Available: https://www.researchgate.net/publication/385418567_A_Machine_Learning-Driven_Website_Platform_and_Browser_Extension_for_Real-Time_Risk_Scoring_and_Fraud_Detection_for_Website_Legitimacy_Verification_and_Consumer_Protection

[3] S. Li and O. Dib, “Enhancing Online Security: A Novel Machine Learning Framework for Robust Detection of Known and Unknown Malicious URLs,” *J. Theor. Appl. Electron. Commer. Res.* , vol. 19, no. 4, pp. 2919–2960, 2024, doi: 10.3390/jtaer19040141.

[4] R. Alhamyani and M. Alshammari, “Machine Learning-Driven Detection of Cross-Site Scripting Attacks,” *Inf.*, vol. 15, no. 7, 2024, doi: 10.3390/info15070420.

[5] L. Guo, R. Song, J. Wu, Z. Xu, and F. Zhao, “Integrating a machine learning-driven fraud detection system based on a risk management framework,” *Appl. Comput. Eng.*, vol. 87, no. 1, pp. 80–86, 2024, doi: 10.54254/2755-2721/87/20241541.

[6] K. Tran and D. Sovilj, “Advancing Malicious Website Identification: A Machine Learning Approach Using Granular Feature Analysis,” pp. 1–16, 2024, [Online]. Available: http://arxiv.org/abs/2409.07608

[7] “How to detect a phishing URL using Python and Machine Learning.” Accessed: Feb. 13, 2026. [Online]. Available: https://www.activestate.com/blog/phishing-url-detection-with-python-and-ml/

[8] U. Haresh, A. Perla, and R. Kumari, “Advanced Machine Learning Framework for Robust Phishing Website Identification,” vol. 11, no. 3, pp. 1–4, 2025.

[9] D. Goel, H. Ahmad, A. K. Jain, and N. K. Goel, “Machine Learning Driven Smishing Detection Framework for Mobile Security,” 2024, [Online]. Available: http://arxiv.org/abs/2412.09641

[10] M. Dandotiya, N. K. Goyal, A. Khunteta, and B. Tiwari, “Scientific Reports Article in Press Real time identification of phishing attacks through machine learning enhanced browser extensions IN IN,” pp. 0–33, 2026.

[11] Y. Ari Kustiawan and K. Imran Ghauth, “Feature Engineering for Phishing Website Detection Using Machine Learning: A Systematic Review,” *IEEE Access*, vol. 13, no. October, pp. 192080–192104, 2025, doi: 10.1109/ACCESS.2025.3630334.

[12] H. Reddy Bonikela and L. Goel, “Deep Learning for Cybersecurity: Ai-Based Detection of Phishing and Fraudulent Web Pages,” *Int. Res. J. Mod. Eng. Technol. Sci.* , no. 04, pp. 1–10, 2025, [Online]. Available: https://www.researchgate.net/profile/Harish-Reddy-Bonikela/publication/394531434_DEEP_LEARNING_FOR_CYBERSECURITY_AI-BASED_DETECTION_OF_PHISHING_AND_FRAUDULENT_WEB_PAGES/links/68a349de7984e374ace969a6/DEEP-LEARNING-FOR-CYBERSECURITY-AI-BASED-DETECTION-OF-P

[13] L. Kumar and E. Kollwitz, “Machine Learning-Driven Cybersecurity Models for E-Commerce Ecosystem Protection,” 2025, [Online]. Available: https://www.researchgate.net/publication/392658360

[14] M. Ali, Aamir Raza, Malik Arslan Akram, Haroon Arif, and Aamir Ali, “Enhancing IOT Security: A review of Machine Learning-Driven Approaches to Cyber Threat Detection,” *J. Informatics Interact. Technol.*, vol. 2, no. 1, pp. 316–324, 2025, doi: 10.63547/jiite.v2i1.64.

[15] R. Rani, A. Alam, and A. Javed, “Secure UPI: Machine Learning-Driven Fraud Detection System for UPI Transactions,” *2024 2nd Int. Conf. Disruptive Technol. ICDT 2024*, pp. 924–928, 2024, doi: 10.1109/ICDT61202.2024.10489682.

[16] A. Mathur, A. S. Dubey, P. S. Mahara, and U. D. Gandhi, “Unified Approach Integrating Machine Learning Algorithms for Detection of e-Mail Phishing Attacks,” *2024 Int. Conf. Comput. Sci. Commun. ICCSC 2024*, 2024, doi: 10.1109/ICCSC62048.2024.10830374.

[17] S. Das Guptta, K. T. Shahriar, H. Alqahtani, D. Alsalman, and I. H. Sarker, “Modeling Hybrid Feature-Based Phishing Websites Detection Using Machine Learning Techniques,” *Ann. Data Sci.*, vol. 11, no. 1, pp. 217–242, 2024, doi: 10.1007/s40745-022-00379-8.

[18] P. Sharma, “Evaluating Machine Learning and Deep Learning Approaches for Phishing URL Detection: A Systematic Review and Future Directions,” vol. 22, no. 5, pp. 3–6, 2024, [Online]. Available: https://google.academia.edu/JournalofComputerScience

[19] F. Malik, M. Suliman, M. Q. Khan, N. Rahman, K. Khan, and M. Khan, “Optimizing Malicious Website Detection with the XGBoost Machine Learning Approach,” *J. Comput. Biomed. Informatics*, vol. 7, no. 2, 2024.

[20] S. Alrefaai, G. Ozdemir, and A. Mohamed, “Detecting Phishing Websites Using Machine Learning,” *HORA 2022 - 4th Int. Congr. Human-Computer Interact. Optim. Robot. Appl. Proc.*, pp. 1–17, 2022, doi: 10.1109/HORA55278.2022.9799917.

[21] Z. Hu and Z. Yuan, “A Review of Data-Driven Approaches for Malicious Website Detection,” *Proc. 2023 7th Asian Conf. Artif. Intell. Technol. ACAIT 2023*, pp. 75–82, 2023, doi:

10.1109/ACAIT60137.2023.10528600.

[22] J. Kaur, U. Garg, and G. Bathla, *Detection of cross-site scripting (XSS) attacks using machine learning techniques: a review*, vol. 56, no. 11. 2023. doi: 10.1007/s10462-023-10433-3.

[23] K. Thakur, M. L. Ali, M. A. Obaidat, and A. Kamruzzaman, "A Systematic Review on Deep-Learning-Based Phishing Email Detection," *Electron.*, vol. 12, no. 21, pp. 1–26, 2023, doi: 10.3390/electronics12214545.

[24] Z. Alshingiti, R. Alaqel, J. Al-Muhtadi, Q. E. U. Haq, K. Saleem, and M. H. Faheem, "A Deep Learning-Based Phishing Detection System Using CNN, LSTM, and LSTM-CNN," *Electron.*, vol. 12, no. 1, pp. 1–18, 2023, doi: 10.3390/electronics12010232.

[25] R. Maddali, "Enhancing Data Security with Machine Learning-Driven Threat Detection," vol. 9, no. 3, pp. 513–522, 2022, [Online]. Available: https://www.researchgate.net/profile/Raghavender-Maddali-2/publication/390348375_Enhancing_Data_Security_with_Machine_Learning-Driven_Threat_Detection/links/67eaf05803b8d7280e163056/Enhancing-Data-Security-with-Machine-Learning-Driven-Threat-Detection.pdf

[26] A. Almomani *et al.*, "Phishing Website Detection With Semantic Features Based on Machine Learning Classifiers: A Comparative Study," *Int. J. Semant. Web Inf. Syst.*, vol. 18, no. 1, pp. 1–24, 2022, doi: 10.4018/IJSWIS.297032.

[27] R. Shiva, "AI-Driven Identity Theft Prevention: Using Machine Learning for Fraud Detection and Prevention," 2022, [Online]. Available: https://www.researchgate.net/publication/388525379

[28] 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*, vol. 10, no. November, pp. 124420–124428, 2022, doi: 10.1109/ACCESS.2022.3223111.

[29] C. Catal, G. Giray, B. Tekinerdogan, S. Kumar, and S. Shukla, *Applications of deep learning for phishing detection: a systematic literature review*, vol. 64, no. 6. Springer London, 2022. doi: 10.1007/s10115-022-01672-x.

[30] D. C. C. Rupa, G. Srivastava, S. Bhattacharya, P. Reddy, and T. R. R. Gadekallu, "A Machine Learning Driven Threat Intelligence System for Malicious URL Detection," *ACM Int. Conf. Proceeding Ser.*, no. February, 2021, doi: 10.1145/3465481.3470029.

[31] A. Hannousse and S. Yahiouche, "Towards benchmark datasets for machine learning based website phishing detection: An experimental study," *Eng. Appl. Artif. Intell.*, vol. 104, pp. 1–21, 2021, doi: 10.1016/j.engappai.2021.104347.

[32] E. O. Alonge, N. L. Eyo-Udo, B. C. Ubanadu, A. I. Daraojimba, E. D. Balogun, and K. O. Ogunsola, "Enhancing Data Security with Machine Learning: A Study on Fraud Detection Algorithms," *J. Front. Multidiscip. Res.*, vol. 2, no. 1, pp. 19–31, 2021, doi: 10.54660/.ijfmr.2021.2.1.19-31.

[33] T. G. Thirusubramanian Ganesan, "Machine Learning-Driven AI for Financial Fraud Detection in IoT Environments," *Int. J. HRM Organ. Behav.*, vol. 9, no. 4, pp. 9–21, 2021, doi: 10.62650/ijhrmob.2021.v9.i04.pp9-21.

[34] F. O. Catak, K. Sahinbas, and V. Dörtkardeş, "Malicious URL detection using machine learning," *Artif. Intell. Paradig. Smart Cyber-Physical Syst.*, vol. 1, no. 1, pp. 160–180, 2020, doi: 10.4018/978-1-7998-5101-1.ch008.

[35] J. Feng, L. Zou, O. Ye, and J. Han, "Web2Vec: Phishing Webpage Detection Method Based on Multidimensional Features Driven by Deep Learning," *IEEE Access*, vol. 8, no. June, pp. 221214–221224, 2020, doi: 10.1109/ACCESS.2020.3043188.

[36] S. A. Karim, "Smart Shields against Cyber Threats: Machine Learning-Driven Phishing URL Detection," *Int. J. Sci. Res. Eng. Trends*, vol. 10, no. 6, pp. 2379–2386, 2024, doi: 10.61137/ijsret.vol.10.issue6.339.

[37] L. Tang and Q. H. Mahmoud, "A Deep Learning-Based Framework for Phishing Website Detection," *IEEE Access*, vol. 10, pp. 1509–1521, 2022, doi: 10.1109/ACCESS.2021.3137636.

[38] M. Priyanka, N. Kokare, S. P. Mulani, S. A. Pathan, N. A. Kshirsagar, and P. M. Kolekar, "International Journal of Research Publication and Reviews Ransomware Detection Using Machine Learning," *Int. J. Res. Publ. Rev.*, no. 5, pp. 10891–10908, 2024, [Online]. Available: www.ijrpr.com

[39] M. Jackson, "Machine Learning for Phishing Detection : A Data-Driven Approach to Network Security Date : December , 2022," no. December 2022, 2025.

[40] T. Mahmud *et al.*, "A Machine Learning-Based Framework for Malicious URL Detection in Cybersecurity," *Proc. - 2025 8th Int. Conf. Inf. Comput. Technol. ICICT 2025*, pp. 61–65, 2025, doi: 10.1109/ICICT64582.2025.00016.