

Evolving strategies in anti-phishing: an in-depth analysis of detection techniques and future research directions

Preeti, Priti Sharma

Department of Computer Science and Applications, Maharshi Dayanand University, Rohtak, India

Article Info

Article history:

Received Apr 5, 2024

Revised Sep 19, 2024

Accepted Oct 7, 2024

Keywords:

Anti-phishing techniques

Cybersecurity

Deep learning

Machine learning

Phishing attacks

URL blacklists

ABSTRACT

Phishing attacks are a major digital threat, impacting individuals and organizations globally. This review paper examines evolving anti-phishing strategies by analyzing five key techniques: URL blacklists, visual similarity detection, heuristic methods, machine learning models, and deep learning techniques. Each technique is evaluated for its mechanisms, unique features, and challenges. A systematic literature survey (SLR) is conducted to compare these methods; effectiveness. The paper highlights significant research challenges and suggests future directions, emphasizing the integration of artificial intelligence and behavioral analytics to combat evolving phishing tactics, this study aims to advance understanding and inspire more effective anti-phishing solutions.

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



Corresponding Author:

Preeti

Department of Computer Science and Applications, Maharshi Dayanand University

Rohtak, India

Email: miskhokhar121@gmail.com

1. INTRODUCTION

Phishing is a prevalent cybercrime where attackers use calls, emails, and texts to steal personal and financial information through deceptive means. Employing social engineering tactics, phishers masquerade as legitimate entities to commit online identity theft [1], [2]. Key studies in this area explore various anti-phishing techniques. URL blacklists are crucial for blocking known phishing sites, preventing users from accessing harmful URLs [3]. Visual similarity detection focuses on analyzing website designs to identify and flag sites that closely resemble legitimate ones [4]. Heuristic methods use predefined rules to detect unusual behaviors and patterns indicative of phishing attempts [5]. Machine learning models apply sophisticated data-driven algorithms to identify and predict phishing attempts by learning from existing patterns and features [6]. Deep learning advances phishing URL detection by using neural networks to analyze and learn from complex patterns in data. These techniques train on large datasets to identify subtle features and anomalies associated with phishing attempts, enabling effective and adaptive detection of new and evolving threats [7].

These techniques collectively advance the ability to detect phishing URL attacks and enhance overall cybersecurity. The key focusing on these anti-phishing techniques are outlined as follows:

- Blacklists based: these techniques detect phishing by comparing URLs against known lists of malicious sites. These lists are compiled from historical data and reported phishing incidents. When a URL matches a blacklist entry, it is flagged as potentially harmful, helping prevent phishing attacks by blocking access to identified fraudulent sites. Table 1 provides analysis of studies on URL blacklist techniques for phishing detection, including various methodologies for maintaining and updating these lists.

Table 1. Key study base on URL blacklists techniques

Reference number	Data sources	Detection accuracy	Limitation	Future improvements
[8]	200 legitimate(10 top login pages, 50 Alexa sites, 140 yahoo), 200 phishing pages (PhishTank).	Evaluates the accuracy of identifying phishing vs. legitimate pages using whitelist and SVM classifier. The approach achieves 98.4% True positive rate, 97.2% Precision, and 97.7% F1-score.	Cannot detect DNS spoofing, reliance on search engine may affect performance.	Incorporate IP addresses, refine classification features, and improve search engine integration.
[9]	PhishTank, SpamScatter, DMOZ, and Yahoo Random URL Generator used for phishing and benign URL analysis.	Effective heuristics and fast approximate matching show low false positives/ negatives, outperforming Google’s Safe Browsing API in speed. The highest accuracy factor is “High Similarity (90-100%): 10,606 URLs from Heuristic H3-Directory.	Trade-off between false positives/ negatives; heuristics not exhaustive; DNS/ content matching dependency; less effective on very short-lived domains.	Expand heuristics, enhance URL generation, refine matching accuracy, integrate more data sources, and implement adaptive learning.
[10]	Utilizes URL-based detection, visual URL extraction, and visual similarity comparisons between suspicious and legitimate pages.	The approach is expected to achieve improved detection accuracy compared to existing methods, with fewer false positives and better coverage.	May not detect all sophisticated phishing attempts; visual similarity might be less effective against highly advanced phishing techniques.	Enhance accuracy by refining URL-based and visual similarity techniques, incorporate additional detection methods, and reduce false positive further.
[11]	Legitimate sources include Alexa, DMOZ; phishing sources include PhishTank, OpenPhish.	Current methods often suffer from imbalanced data, leading to high false positives.	Imbalanced datasets cause bias, and URL features can be manipulated, affecting detection reliability.	Focus on domain name-based features and balanced datasets to enhance detection accuracy and reliability.
[12]	Two datasets from the UCI repository: Phishing Dataset1 with 1353 URLs(548 legitimate, 702 phishing, 103 suspicious) and phishing Dataset2 with 4898 phishing URLs and 6157 legitimate URLs. PhishTank.	The hybrid algorithm achieved accuracies of 0.9453 and 0.9908, surpassing JRip and PART.	The study primarily evaluates performance on specific datasets; generalizability to other phishing scenarios and dataset variations may be limited.	Research will focus on adaptive machine learning techniques to enhance detection of Zero-day phishing threats.
[13]	PhishTank.	The proposed rule-based detection technique achieved competitive accuracy with C4.5 and logistic regression, yielding an accuracy of 99%, with a false positive rate of 0.5% and a false negative rate of 2.5%.	The rule set used in this approach is premature and needs expansion. It may miss phishing webpages designed to minimize rule matches or those hosted on hacked legitimate pages.	Future work will focus on refining the rule set to reduce false positives and negatives, developing a lightweight, real-time phishing detection system, and exploring optimal intervals for retraining the system with new data.
[3]	URLs extracted from spam filters, user reports, and phishing websites identified by heuristics.	High accuracy, with heuristics detecting more phishing attempts initially; blacklists update slower.	Data sourced from a single anti-spam vendor; only email URLs considered; no other vectors.	Speed up blacklist updates, enhance heuristic methods, and improve user phishing awareness.

Table 1 highlights various studies on URL blacklist techniques. One study achieved a 98.4% true positive rate and 97.7% F1-score using whitelist and support vector machines (SVM) classifiers, though it couldn’t detect DNS spoofing and was affected by search engine reliance. Another utilized heuristics and fast matching, outperforming Google’s Safe Browsing API but struggled with very short-lived domains.

A hybrid algorithm showed 94.53% and 99.08% accuracy but needed refinement to handle sophisticated phishing. Future improvements across studies included incorporating IP addresses, refining features, and enhancing heuristic methods.

- Heuristic based techniques: this approach identifies phishing attempts by analyzing URLs and web content for suspicious patterns and characteristics. These methods use predefined rules and algorithms to detect anomalies that might indicate phishing, such as unusual URL structures or content inconsistencies. By assessing various heuristics, these techniques aim to spot phishing sites that might evade simpler detection methods. The detailed analysis of studies on heuristic based techniques, as outlined below, explores their effectiveness in detecting phishing attacks [14].

In this research, PhishShield was developed to detect phishing websites with 96.57% accuracy by analyzing URL and content through heuristics such as footer links and copyright information. It outperformed traditional methods and blacklists, particularly for zero-hour attacks. However, its reliance on heuristics did not cover all sophisticated phishing techniques and faced challenges with evolving threats [15].

This research proposed a novel phishing detection approach using URL features and metrics, combined with page ranking. Evaluated on a dataset of 9,661 phishing and 1,000 legitimate websites, the technique achieved over 97% detection accuracy. However, despite its effectiveness, the approach may face limitations in handling sophisticated phishing methods that mimic legitimate URLs closely and may not fully address evolving phishing tactics [16]. A method for detecting phishing and malware was introduced, analyzing specific strings in URLs and emails messages, used with proxies and anti-spam filters. It achieved detection accuracy between 73.3% and 97.66% with an average more sophisticated phishing techniques [17].

The proposed heuristic-based phishing detection technique, which used URL-based features and machine learning classifiers, achieved 96% accuracy with a low false-positive rate, effectively identifying new and temporary phishing sites. However, the approach faced challenges with emerging phishing tactics due to its reliance on specific URL features. Future work aimed to explore new features, enhance accuracy, and develop a browser plugin for real-time phishing alerts [18].

- Visual similarity based techniques: this section examines the analysis of phishing attack detection through visibility-based techniques. The study focused on evaluating visual features of web pages, such as design elements and layout, to identify phishing threats. By leveraging these visual cues, the approach aimed to distinguish between legitimate and phishing sites. Details of the study are described below, highlighting key findings and limitations of previous work using visual similarity techniques

In this paper, the researchers introduced a novel phishing detection method comparing visual similarity between suspicious and legitimate pages, inspired by existing anti-phishing tools. They analyzed text, images and overall visual appearance. Results with 41 phishing pages showed no false positives and only two missed detections. Key findings included high accuracy and effectiveness, while limitations involved occasional failure to detect highly dissimilar phishing attempts. Future work should enhance detection cases [4].

The study evaluated visual similarity-based phishing detecting models using a dataset of 450k real-world phishing websites, revealing performance gaps between real-world and controlled environments. Key limitations included a lack of user studies, focus work should enhance robustness by integrating text recognition, adversarial data augmentation, and multi-cue ensemble approaches [19].

VisualPhishNet, a robust phishing detection framework, significantly outperformed prior visual similarity approaches, achieving a 56% improvement in matching accuracy and a 30% increase in receiver operating characteristic (ROC) area under the curve (AUC). However, the focus on visual similarity and the limited evaluation scope were key limitations. Future work was recommended to explore additional attack vectors, conduct user studies, and strengthen defense against evolving evasion tactics [20].

The research proposed a visual similarity-based phishing detection method using both local and global web page features. It achieved over 90% true positive and 97% true negative rates on a large dataset. However, its reliance on image-level analysis limits its ability to detect advanced phishing techniques. Future work should focus on enhancing detection capabilities and expanding its application to diverse phishing scenarios [21].

BaitAlarm, an anti-phishing approach that utilized CSS and visual features for similarity comparisons between suspicious and target pages, demonstrated effectiveness through evaluations with numerous phishing pages. Despite its success, it faced limitations due to vulnerability to evasion attacks. Future works aimed to enhance its resilience against such attacks [22].

- Machine learning based techniques: in this section, we describe a study based on machine learning techniques. The study explored various machine learning algorithms to improve phishing detection. It focused on leveraging advanced algorithms to analyze patterns and anomalies in data to identify phishing attempts more accurately. Details of the techniques, models used, and results are provided below.

The study evaluated various machine learning algorithms for phishing URL detection, including decision tree, multilayer perceptron, random forest, XGBoost, Autoencoder neural network, and SVM. Using a dataset from Phishtank and the University of New Brunswick, it found Random Forest and XGBoost outperformed others, achieving an overall accuracy of 98% in phishing detection [23]. A genetic algorithm-based feature selection method improved URL phishing detection, with random forest achieving 99.93% accuracy. Limitations included specific feature reliance and scalability issues. Future work should optimize for diverse datasets and scenarios [24].

A machine learning-based system, PHISH-SAFE, was developed to detect phishing websites using 14 URL features. Trained on a dataset of over 33,000 URLs, the system utilized SVM and Naïve Bayes classifiers. The SVM classifier demonstrated over 90% accuracy in identifying phishing sites, showcasing the method's effectiveness in cybersecurity [25].

The research developed a hybrid ensemble feature selection (HEFS) method using URL features for phishing detection, achieving 97.9% accuracy with a novel CDF-g algorithm. However, the study faced challenges in classifier complexity and data partitioning. Future work should focus on refining these issues to enhance stability and performance further. The highest accuracy was obtained by the proposed model [26].

The study utilized machine learning algorithms, specifically SVM, to detect phishing URLs, achieving improved accuracy by leveraging URL features from a Kaggle dataset. However, the study was limited by its reliance on a specific dataset and the potential need for more comprehensive feature sets. Future work should explore additional datasets and advanced algorithms to enhance phishing detection accuracy and applicability across diverse scenarios [27].

The paper introduced a machine learning-based approach for real-time phishing website detection, utilizing hybrid features from URLs and hyperlinks without relying on third-party systems. By avoiding the limitations of traditional methods like blacklists and heuristics, this approach enhanced detection accuracy. Experiments conducted with newly developed datasets demonstrated the method's effectiveness, achieving a high detection accuracy of 99.17% using XGBoost, surpassing conventional techniques [28].

The evaluation focused on gradient boosting classifier, random forest, and decision tree models for phishing detection, using feature selection methods such as SelectBest and Chi-Square. A comprehensive set of 30 features achieved a baseline accuracy of 97.4%, which slightly decreased to 96.6% after selecting 13 key features. This work emphasized the significance of feature selection in enhancing phishing detection accuracy and maintaining model interpretability, advancing cyber defense capabilities [29].

The study introduced predictive queuing analysis to forecast network performance for SD-UAV network, enhancing security against zero-day cyberattacks. Metrics such as interarrival times and packet count supported machine learning for anomaly detection. Future work aims to integrate this with an IDS for real-time threat mitigation [30]. The research emphasized the growing need for IDS due to rising cyber threats, enhancing SVM with PSO. Results using the KDD-CUP 99 dataset demonstrated improved performance across various cyber-attack types [31].

- Deep learning based techniques: in this section, various deep learning techniques for detecting phishing URL attacks are described. A detailed investigation is presented below, outlining the different methods used to identify and prevent such attacks. The analysis covers multiple approaches, examining their implementation. The findings provide insights into strengths and limitations of each technique, contributing to the ongoing development of robust phishing detection models.

The research proposed a multi-layer adaptive framework that significantly improved the detection rate of phishing attacks by incorporating OCR for image recognition and synthesizing speech from deepfake videos. It overcame the limitations of existing AI-based approaches, which were primarily text or URL based. The study's limitations included reliance on simulated data for image and video-based phishing, and future work suggested reducing computational and server response times [32]. A hybrid convolutional neural network (CNN)-long short-term memory network (LSTM) model achieved 98.9% accuracy for URL spoofing, surpassing individual models. Despite its high performance, real-time application challenges remain. Future work should enhance speed and broaden dataset evaluations [33]. The research attained 98.74% accuracy with a CNN-based model, processing over 5.2 million URLs. Despite its strengths in real-time and language independence, it needs improvements in computational efficiency and dataset handling [34].

Developed an intelligent phishing detection system using deep learning models, including CNN, LSTM, and hybrid models. By applying feature selection and data balancing techniques, the system achieved an accuracy range of 94.12% to 96.88% [35]. Introduced three deep learning approaches: CNN, LSTM, and LSTM+CNN for phishing website detection. The CNN model achieved the highest accuracy at 99.2%, followed by LSTM-CNN at 97.6% and LSTM at 96.8%. While the models demonstrated high accuracy, future work should address potential limitations in generalizability and real-time performance. Enhancement could include integrating additional features and optimizing for diverse datasets [36].

The study proposed an enhanced phishing detection model integrating variational autoencoders (VAN) with deep neural network (DNN), achieving a maximum accuracy of 97.45%. While the model improved detection and response time, its limitation was in potentially handling emerging phishing tactics. Future work should focus on adapting the model to new phishing methods and further optimizing response time [37].

The proposed detection mechanism for malicious URLs using LSTM, Bi-LSTM, and GRU models achieved accuracies of 97.0%, 99.0%, and 97.5%, respectively. However, the research was limited by potential model overfitting and the need for a border dataset to enhance generalizability. Future work could have focused on incorporating real-time detection capabilities and exploring hybrid models to further improve accuracy and robustness against evolving phishing techniques [38].

The study's main finding was that the proposed ODAE-WPDC model effectively detected phishing websites with a maximum accuracy of 99.28%. However, the research was limited by its reliance on pre-processing and specific algorithms, which may not generalize well to all datasets. Future work could explore broader datasets and more adaptable algorithms to improve performance across diverse phishing techniques [39].

Introduced a novel phishing detection technique using BERT for feature extraction and deep learning, achieving 96.66% accuracy. However, the study was limited by its reliance on natural language processing techniques, which may not fully capture all phishing strategies [40]. The study found that deep learning techniques showed promise for phishing detection but struggled with manual parameter-tuning, long training times, and accuracy issues [41].

Developed and optimized deep learning models for phishing website detection, achieving a best accuracy of 97.37%. Hyperparameter optimization using GRID search and genetic algorithm improved accuracy by 0.1%-1%. However, gaps remained in understanding the robustness of these models [42]. Researchers developed a phishing approach using Variational Autoencoders and deep neural networks, achieving 97.45% accuracy and a response time of 1.9 seconds, surpassing traditional blacklist-based methods [39]. The research proposed a model using CNNs to classify webpages as benign or phishing based on URLs and images, achieving 99.67% accuracy [43].

Compared CNN, LSTM-CNN, and LSTM for phishing detection, achieving 99.2%, 97.6%, and 96.8% accuracy, respectively; CNN performed best [44]. The study developed a supervised learning model for Android malware detection, outperforming existing methods in precision, efficiency, and precision [45].

The paper presented a CNN-based phishing detection method with an 86.63% true detection rate and 30% faster execution on a Raspberry Pi. Future work should focus on integrating deep learning models for webpage content analysis [46]. Introduced The ODAE-WPDC for phishing detection, achieving 99.28% accuracy with deep autoencoders and optimized feature selection. Future work should focus on improving real-time application and adaptability [47].

2. SUMMARIZING KEY FINDINGS

Key findings from anti-phishing research reveal that blacklist-based techniques achieved high accuracy (up to 99%) but struggled with DNS spoofing and dataset reliance, suggesting future work should enhance heuristic methods and integrate IP addresses. Heuristic-based methods demonstrated up to 97% accuracy in detecting new phishing sites but faced limitations with sophisticated tactics, recommending further exploration of new features and real-time detection. Visual similarity techniques, achieving up to 98.74% accuracy, encountered difficulties with detecting highly dissimilar phishing sites, indicating a need for improved robustness and border application. Machine learning methods, including random forest and XGBoost, reached up to 99.93% accuracy but had issues with feature dependence and scalability, with future work focusing on dataset diversity and advanced algorithms. Deep learning approaches, such as CNNs and LSTMs, exhibited high accuracy (up to 99.67%) but faced challenges in real-time application and generalizability, highlighting the need for enhanced adaptability and integration of diverse datasets.

3. INTERPRETING RESULTS

The analysis of anti-phishing techniques, as illustrated in Figure 1, reveals substantial differences in detection accuracy across various methods. The deep learning-based approach, particularly the CNN models, achieved the highest accuracy of 99.67%, demonstrating superior performance in detecting phishing threats with minimal false positives and negatives, while still effective, exhibit lower accuracies due to limitations in their scope and adaptability. This comparison highlights the advanced capability of deep learning models and exploring hybrid models that integrate these with heuristic and machine learning approaches to address their individual limitations. Such integration is likely to provide a more comprehensive and resilient defense against evolving phishing tactics.

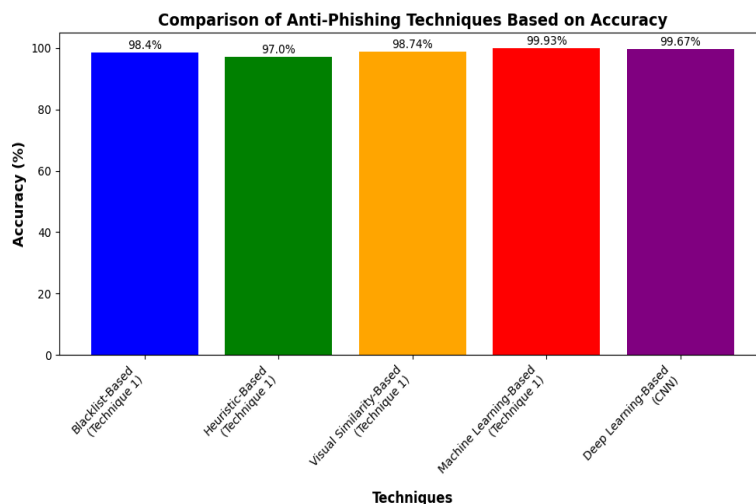


Figure 1. Accuracy comparison of various anti-phishing techniques: blacklist-based, heuristic-based, visual similarity-based, machine learning-based, and deep learning-based approaches

4. ADDRESSING LIMITATION

Blacklist-based techniques face challenges in detecting new or evolving phishing threats due to their reliance on precompiled lists that may not include the latest phishing sites. This method also struggles with issues like DNS spoofing and phishing hosted on legitimate domains. Heuristic-based techniques, while useful for identifying suspicious patterns, may falter against sophisticated phishing attempts that closely mimic legitimate sites and are less adaptable to evolving threats. Visual similarity-based techniques, which assess visual features of web pages, may miss phishing attacks that do not visually resemble legitimate sites, especially those using advanced designs. Additionally, these methods may not handle real time threats effectively. Machine learning and deep learning approaches, though promising, can be limited by dataset biases, computational resource demands, and challenges in real time application, necessitating future optimization and adaptation to enhance overall phishing detection capabilities.

5. IMPLICATIONS FOR FUTURE RESEARCH

Our study demonstrates that deep learning-based techniques, particularly those using CNNs and LSTMs, exhibit superior accuracy and robustness in phishing detection compared to heuristic-based and blacklist-based methods. Future research could explore enhancing these techniques by integrating additional features and developing more efficient models to handle diverse phishing scenarios. Specifically, investigating the combination of deep learning with other detection methods, such as visual similarity and machine learning approaches, could provide a more comprehensive defense against evolving phishing threats. Additionally, focusing on optimizing computational resources and real-time processing capabilities will be crucial for improving the practical deployment of these advanced techniques. Feasible way to produce these improvements include leveraging transfer learning, refining feature extraction methods, and exploring novel hybrid models to achieve higher detection accuracy and reduced false positive in dynamic, real-world environments.

6. CONCLUSION

In reviewing the literature on anti-phishing techniques, it becomes evident that deep learning-based methods, particularly CNNs, offer the highest detection accuracy, reaching up to 99.67%. This surpasses the effectiveness of other techniques, including machine learning approaches that achieve up to 99.93% accuracy and heuristic-based methods. The superiority of deep learning models in accuracy identifying phishing attempts highlights their robustness and effectiveness in minimizing detection errors. To address the evolving nature of phishing attacks, it is crucial for future research to focus on optimizing these advanced models for real-time application and to explore their potential integration with other detection strategies. Enhancing these techniques and their adaptability will be essential in strengthening defenses against increasingly sophisticated phishing threats.





REFERENCES

- [1] V. Bhavsar, A. Kadlak, and S. Sharma, "Study on phishing attacks," *International Journal of Computer Applications*, vol. 182, no. 33, pp. 27–29, Dec. 2018, doi: 10.5120/ijca2018918286.
- [2] Z. Alkhalil, C. Hewage, L. Nawaf, and I. Khan, "Phishing attacks: a recent comprehensive study and a new anatomy," *Frontiers in Computer Science*, vol. 3, Mar. 2021, doi: 10.3389/fcomp.2021.563060.
- [3] S. Sheng, B. Wardman, G. Warner, L. F. Cranor, J. Hong, and C. Zhang, "An empirical analysis of phishing blacklists," in *6th Conference on Email and Anti-Spam, CEAS 2009*, 2009.
- [4] A. V. R. Mayuri, M. Tech, and D. Ph, "Phishing detection based on visual-similarity," *International Journal of Scientific and Engineering Research (IJSER)*, vol. 3, no. 3, pp. 1–5, 2012.
- [5] C. M. R. da Silva, E. L. Feitosa, and V. C. Garcia, "Heuristic-based strategy for Phishing prediction: A survey of URL-based approach," *Computers & Security*, vol. 88, p. 101613, Jan. 2020, doi: 10.1016/j.cose.2019.101613.
- [6] R. Kiruthiga and D. Akila, "Phishing websites detection using machine learning," *International Journal of Recent Technology and Engineering*, vol. 8, no. 2 Special Issue 11, pp. 111–114, Nov. 2019, doi: 10.35940/ijrte.B1018.0982S1119.
- [7] V. Ravi, S. Srinivasan, S. Kp, and M. Alazab, "Malicious URL detection using deep learning," pp. 1–9, Jan. 11, 2020, doi: 10.36227/techrxiv.11492622.v1.
- [8] A. Belabed, E. Aimeur, and A. Chikh, "A personalized whitelist approach for phishing webpage detection," in *Proceedings - 2012 7th International Conference on Availability, Reliability and Security, ARES 2012*, IEEE, Aug. 2012, pp. 249–254, doi: 10.1109/ARES.2012.54.
- [9] P. Prakash, M. Kumar, R. Rao Kompella, and M. Gupta, "PhishNet: predictive blacklisting to detect phishing attacks," in *Proceedings - IEEE INFOCOM*, IEEE, Mar. 2010, pp. 1–5, doi: 10.1109/INFCOM.2010.5462216.
- [10] N. M. Shekokar, C. Shah, M. Mahajan, and S. Rachh, "An ideal approach for detection and prevention of phishing attacks," *Procedia Computer Science*, vol. 49, no. 1, pp. 82–91, 2015, doi: 10.1016/j.procs.2015.04.230.
- [11] E. S. Aung, T. Zan, and H. Yamana, "A survey of URL-based phishing detection," *DEM Forum*, pp. 1–8, 2019. [Online]. Available: <https://db-event.jp.n.org/deim2019/post/papers/201.pdf>
- [12] K. S. Adewole, A. G. Akintola, S. A. Salihu, N. Faruk, and R. G. Jimoh, "Hybrid rule-based model for phishing URLs detection," in *Lecture Notes of the Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering, LNICTS*, vol. 285, 2019, pp. 119–135, doi: 10.1007/978-3-030-23943-5_9.
- [13] Q. L. Ram B. Basnet, Andrew H. Sung, "Rule-based phishing attack detection," in *2011 International Conference on Security and Management-SAM'11*, 2011. [Online]. Available: <https://cs.nmt.edu/~rbasnet/research.html>
- [14] A. Begum and S. Badugu, "A study of malicious URL detection using machine learning and heuristic approaches," in *Learning and Analytics in Intelligent Systems*, vol. 4, 2020, pp. 587–597, doi: 10.1007/978-3-030-24318-0_68.
- [15] R. S. Rao and S. T. Ali, "PhishShield: a desktop application to detect phishing webpages through heuristic approach," *Procedia Computer Science*, vol. 54, pp. 147–156, 2015, doi: 10.1016/j.procs.2015.06.017.
- [16] L. A. T. Nguyen, B. L. To, H. K. Nguyen, and M. H. Nguyen, "Detecting phishing web sites: a heuristic URL-based approach," in *2013 International Conference on Advanced Technologies for Communications (ATC 2013)*, IEEE, Oct. 2013, pp. 597–602, doi: 10.1109/ATC.2013.6698185.
- [17] R. Almeida and C. Westphall, "Heuristic phishing detection and URL checking methodology based on scraping and web crawling," in *2020 IEEE International Conference on Intelligence and Security Informatics (ISI)*, IEEE, Nov. 2020, pp. 1–6, doi: 10.1109/ISI49825.2020.9280549.
- [18] J. Solanki and R. G. Vaishnav, "Website phishing detection using heuristic based approach," *International Research Journal of Engineering and Technology*, pp. 2044–2048, 2016, [Online]. Available: www.irjet.net
- [19] F. Ji *et al.*, "Evaluating the effectiveness and robustness of visual similarity-based phishing detection models." 2024. [Online]. Available: <http://arxiv.org/abs/2405.19598>
- [20] S. Abdelnabi, K. Krombolz, and M. Fritz, "VisualPhishNet: zero-day phishing website detection by visual similarity," in *Proceedings of the ACM Conference on Computer and Communications Security*, New York, NY, USA: ACM, Oct. 2020, pp. 1681–1698, doi: 10.1145/3372297.3417233.
- [21] Y. Zhou, Y. Zhang, J. Xiao, Y. Wang, and W. Lin, "Visual Similarity based anti-phishing with the combination of local and global features," in *2014 IEEE 13th International Conference on Trust, Security and Privacy in Computing and Communications*, IEEE, Sep. 2014, pp. 189–196, doi: 10.1109/TrustCom.2014.28.
- [22] J. Mao, P. Li, K. Li, T. Wei, and Z. Liang, "BaitAlarm: detecting phishing sites using similarity in fundamental visual features," in *Proceedings - 5th International Conference on Intelligent Networking and Collaborative Systems, INCoS 2013*, IEEE, Sep. 2013, pp. 790–795, doi: 10.1109/INCoS.2013.151.
- [23] Vishesh Bharuka, Allan Almeida, and Sharvari Patil, "Phishing detection using machine learning algorithm," *International Journal of Scientific Research in Computer Science, Engineering and Information Technology*, vol. 10, no. 2, pp. 343–349, Mar. 2024, doi: 10.32628/cseit2410228.
- [24] E. Kocyigit, M. Korkmaz, O. K. Sahingoz, and B. Diri, "Enhanced feature selection using genetic algorithm for machine-learning-based phishing URL detection," *Applied Sciences (Switzerland)*, vol. 14, no. 14, p. 6081, Jul. 2024, doi: 10.3390/app14146081.
- [25] A. K. Jain and B. B. Gupta, "PHISH-SAFE: URL features-based phishing detection system using machine learning," in *Advances in Intelligent Systems and Computing*, vol. 729, 2018, pp. 467–474, doi: 10.1007/978-981-10-8536-9_44.
- [26] R. Jayaraj, A. Pushpalatha, K. Sangeetha, T. Kamaleshwar, S. Udhaya Shree, and D. Damodaran, "Intrusion detection based on phishing detection with machine learning," *Measurement: Sensors*, vol. 31, p. 101003, Feb. 2024, doi: 10.1016/j.measen.2023.101003.
- [27] G. R. Kumar, S. Gunasekaran, N. R. S. P. K, S. G, and V. A. S, "Url phishing data analysis and detecting phishing attacks using machine learning in nlp," *International Journal of Engineering Applied Sciences and Technology*, vol. 3, no. 10, pp. 26–31, 2019, doi: 10.33564/ijeast.2019.v03i10.007.
- [28] S. Das Gupta, K. T. Shahriar, H. Alqahtani, D. Alsalmán, and I. H. Sarker, "Modeling hybrid feature-based phishing websites detection using machine learning techniques," *Annals of Data Science*, vol. 11, no. 1, pp. 217–242, Mar. 2024, doi: 10.1007/s40745-022-00379-8.
- [29] S. H. Nallamala, K. Namitha, K. Raviteja, K. S. Sumanth, and J. S. Kota, "Phishing URL detection using machine learning," *International Journal for Research in Applied Science and Engineering Technology*, vol. 12, no. 3, pp. 1984–1995, Mar. 2024, doi: 10.22214/ijraset.2024.59261.





- [30] D. Agnew, A. Del Aguila, and J. McNair, "Enhanced network metric prediction for machine learning-based cyber security of a software-defined UAV relay network," *IEEE Access*, vol. 12, pp. 54202–54219, 2024, doi: 10.1109/ACCESS.2024.3387728.
- [31] N. Omer, A. H. Samak, A. I. Taloba, and R. M. Abd El-Aziz, "Cybersecurity Threats detection using optimized machine learning frameworks," *Computer Systems Science and Engineering*, vol. 48, no. 1, pp. 78–95, 2024, doi: 10.32604/csse.2023.039265.
- [32] T. Ige, C. Kiekintveld, and A. Piplai, "Deep learning-based speech and vision synthesis to improve phishing attack detection through a multi-layer adaptive framework." 2024. [Online]. Available: <http://arxiv.org/abs/2402.17249>
- [33] D. Minh Linh, H. D. Hung, H. Minh Chau, Q. Sy Vu, and T.-N. Tran, "Real-time phishing detection using deep learning methods by extensions," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 14, no. 3, p. 3021, Jun. 2024, doi: 10.11591/ijece.v14i3.pp3021-3035.
- [34] R. Ali Shah *et al.*, "Enhancing phishing detection, leveraging deep learning techniques," *Article in Journal of Computing & Biomedical Informatics*, 2024, [Online]. Available: <https://www.researchgate.net/publication/378964404>
- [35] R. Zaimi, M. Haadi, L. Mahnane, M. Hafidi, and M. Lamia, "A permutation importance based feature selection method and deep learning model to detect phishing websites." Feb. 13, 2024, doi: 10.21203/rs.3.rs-3943049/v1.
- [36] U. G. Chetachi, O. N. Henry, and O. A. Agbugba, "Intelligent phishing website detection model powered by deep learning techniques," *Asian Journal of Research in Computer Science*, vol. 17, no. 1, pp. 71–85, Jan. 2024, doi: 10.9734/ajrcos/2024/v17i1414.
- [37] M. K. Prabakaran, P. Meenakshi Sundaram, and A. D. Chandrasekar, "An enhanced deep learning-based phishing detection mechanism to effectively identify malicious URLs using variational autoencoders," *IET Information Security*, vol. 17, no. 3, pp. 423–440, May 2023, doi: 10.1049/ise2.12106.
- [38] S. S. Roy, A. I. Awad, L. A. Amare, M. T. Erkihun, and M. Anas, "Multimodal phishing URL detection using LSTM, bidirectional LSTM, and GRU models," *Future Internet*, vol. 14, no. 11, p. 340, Nov. 2022, doi: 10.3390/fi14110340.
- [39] H. Alqahtani *et al.*, "Evolutionary algorithm with deep auto encoder network based website phishing detection and classification," *Applied Sciences (Switzerland)*, vol. 12, no. 15, p. 7441, Jul. 2022, doi: 10.3390/app12157441.
- [40] M. Elsadig *et al.*, "Intelligent deep machine learning cyber phishing URL detection based on BERT features extraction," *Electronics (Switzerland)*, vol. 11, no. 22, p. 3647, Nov. 2022, doi: 10.3390/electronics11223647.
- [41] N. Q. Do, A. Selamat, O. Krejcar, E. Herrera-Viedma, and H. Fujita, "Deep learning for phishing detection: taxonomy, current challenges and future directions," *IEEE Access*, vol. 10, pp. 36429–36463, 2022, doi: 10.1109/ACCESS.2022.3151903.
- [42] M. Almousa, T. Zhang, A. Sarrafzadeh, and M. Anwar, "Phishing website detection: how effective are deep learning-based models and hyperparameter optimization?," *Security and Privacy*, vol. 5, no. 6, Nov. 2022, doi: 10.1002/spy2.256.
- [43] S. Al-Ahmadi and Y. Alharbi, "A deep learning technique for web phishing detection combined URL features and visual similarity," *International journal of Computer Networks & Communications*, vol. 12, no. 5, pp. 41–54, Sep. 2020, doi: 10.5121/ijcnc.2020.12503.
- [44] 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," *Electronics (Switzerland)*, vol. 12, no. 1, p. 232, Jan. 2023, doi: 10.3390/electronics12010232.
- [45] A. Gómez and A. Muñoz, "Deep learning-based attack detection and classification in android devices," *Electronics (Switzerland)*, vol. 12, no. 15, p. 3253, Jul. 2023, doi: 10.3390/electronics12153253.
- [46] B. Wei *et al.*, "A deep-learning-driven light-weight phishing detection sensor," *Sensors (Switzerland)*, vol. 19, no. 19, p. 4258, Sep. 2019, doi: 10.3390/s19194258.
- [47] P. Yi, Y. Guan, Y. Zou, Y. Yao, W. Wang, and T. Zhu, "Web phishing detection using a deep learning framework," *Wireless Communications and Mobile Computing*, vol. 2018, no. 1, Jan. 2018, doi: 10.1155/2018/4678746.

BIOGRAPHIES OF AUTHORS



Ms. Preeti     M.Tech., Ph.D. Pursuing from Department of Computer Science and Applications, M.D. University, Rohtak. She has published more than 16 publications in various journals/magazines of national and international repute. Her area of research includes machine learning, deep learning, and cyber security. She can be contacted at email: miskhokhar121@gmail.com.



Dr. Priti Sharma     MCA, Ph.D. (compute science) is working as an assistant professor in the Department of Computer Science and Applications, M.D. University, Rohtak. She has published more than 60 publications in various journals/magazines of national and international repute. She is engaged in teaching and research from the last 15 years. Her area of research includes data mining, big data, software engineering, machine learning, and deep learning. She can be contacted at email: priti@mdurohtak.ac.in.