

An optimal machine learning-based algorithm for detecting phishing attacks using URL information

Nandeesh Hallimysore Devaraj, Prasanna Bantiganahalli Thimappa

Department of Computer Science and Engineering, JSS Science and Technology University, Mysore, India

Article Info

Article history:

Received Mar 17, 2024

Revised Jun 8, 2024

Accepted Jun 25, 2024

Keywords:

Genetic algorithms

OmLA

Random forest

Support vector machine

Uniform resource locator

ABSTRACT

In recent years, more websites have been collecting personal information for many processes, such as banks, internet connections, and government services. The public needs to provide all personal information, such as Aadhar, PAN card, date of birth, and phone number. The personal and sensitive information is at risk of being used for phishing attacks through URL manipulation. In addition, a phishing attack cause's financial and reputational loss. Hence protecting sensitive information by adapting required protection is extremely valuable for global security. To overcome this, we proposed a method to detect phishing attacks based on previous history, including the duration of operation, customer reviews, web traffic, and the URL. Based on these parameters, the proposed optimal machine learning-based algorithm (OmLA) analyze the previous information about URLs and predict whether it is phishing- or legitimate. As per simulation and performance analysis, the proposed method outperforms conventional methods such as random forest (RF), support vector machine (SVM), and genetic algorithms (GA) by 8%, 18%, and 23%, respectively in terms of accuracy. Additionally, it achieves detection times of 0.2%, 0.6%, and 0.9%, respectively, and excels in response times of 0.45%, 0.56%, and 0.62%, respectively.

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



Corresponding Author:

Nandeesh Hallimysore Devaraj

Department of Computer Science and Engineering, JSS Science and Technology University

Mysore, 570009, India

Email: hdnandeesh@jssstuniv.in

1. INTRODUCTION

This paper explores the use of machine learning (ML) to detect phishing attacks via URL analysis. It emphasizes the sophistication of modern phishing strategies that employ deceptive URLs, posing significant challenges for traditional detection methods. ML algorithms are highlighted as a superior solution, capable of analyzing extensive datasets of URL patterns to distinguish between malicious and legitimate URLs. This approach not only overcomes the limitations of conventional methods but also adapts to new threats over time. Figure 1 show the fundamental diagram of detecting phishing websites using ML techniques [1]. The proposed ML-based URL detection technique comprises several steps, starting with data collection from sources like PhishTank and web crawlers to gather both malicious and legitimate URLs. This is followed by feature extraction, where URL characteristics are identified for ML use. A recurrent neural network (RNN) then undergoes a training phase to learn differentiating features between harmful and safe URLs, and a testing phase to evaluate its performance on new URLs. The effectiveness of the RNN is assessed using metrics such as accuracy, precision, recall, and F1-score [2].

This research aims to showcase the potential of ML in countering phishing threats by identifying complex patterns and anomalies in URL data. It discusses various ML models, including supervised and

unsupervised learning, and their ability to process and classify URL information based on characteristics like lexical properties and hosting details. The paper also addresses challenges such as the need for large, diverse datasets and the reduction of false positives, aiming to enhance digital security and contribute to a safer online environment.

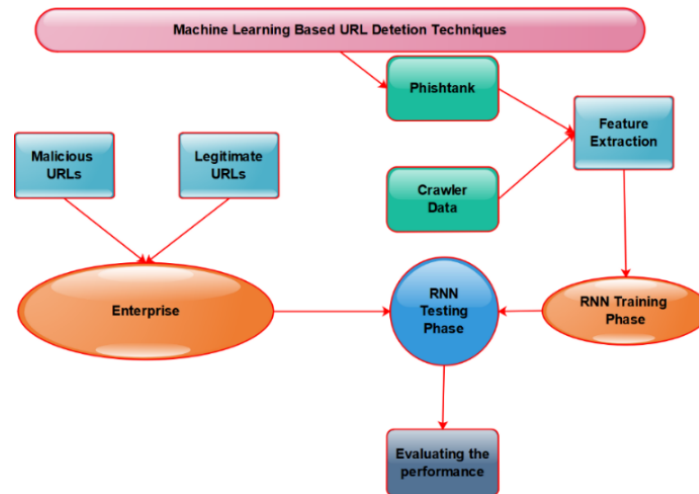


Figure 1. Fundamental diagram of detecting phishing websites using machine-learning techniques

2. RELATED WORK

Zieni *et al.* [3] developed CatchPhish, which uses URL features and a random forest (RF) classifier. Its limitation is focusing solely on URL features, potentially missing sophisticated phishing websites. Aljabri *et al.* [4] proposed Hin Phish, a method based on heterogeneous information networks (HIN) that might misclassify phishing attacks due to the complexity of hyperlink relationships. Aassal *et al.* [5] used distributed word representation within URLs but struggled with unobserved characters and did not consider website content, potentially missing sites mimicking legitimate ones. Indrasiri *et al.* [6] introduced a hybrid long short-term memory (LSTM) and gated recurrent unit (GRU) model for phishing URL detection. Despite its potential, it faces challenges in computational complexity and training data requirements. Ahmed *et al.* [7] presented a neural network model optimized for feature selection in phishing detection, which may not generalize well to new phishing attacks and requires frequent retraining. Kara *et al.* [8] provided a survey of ML techniques for malicious URL detection, potentially missing the latest methods or emerging threats due to the rapidly evolving nature of cybersecurity. Althobaiti *et al.* [9] employed deep learning for URL representation, facing challenges with significant computational resources and lengthy training times. Ariyadasa *et al.* [10] proposed using lexical features and online learning for phishing detection, which might not effectively detect zero-day attacks or sophisticated strategies. Sahingoz *et al.* [11] examined the evolution of phishing attacks but may lack specific technical solutions or address the operational challenges of implementing anti-phishing measures. The above related works strive hard to detect the phishing attacks but failed to detect the zero-day attacks. Hence, our approach work mainly focuses on detecting zero-day attacks based on URL metadata.

2.1. Research gaps

Identifying research gaps in ML for phishing attack detection is vital for improving cybersecurity. Key areas needing further exploration include the development of comprehensive datasets that capture the latest phishing tactics, enhancing the adaptability and scalability of ML models to real-world conditions, and integrating these models within existing cybersecurity frameworks. Additionally, addressing the challenge of false positives and negatives in detection is crucial for maintaining user trust and the effectiveness of security measures. Tackling these gaps promises to boost the accuracy and reliability of phishing detection, contributing to a safer digital environment [12].

2.2. Applications

ML significantly bolsters cybersecurity by detecting phishing attacks through URL analysis, benefiting individual users, organizations, financial institutions, cloud services, e-commerce platforms, and

cybersecurity training programs. For individuals, ML algorithms integrated into web browsers and email clients’ alert users to harmful URLs, preventing phishing frauds. Organizations and financial institutions utilize these systems within their network security to protect against phishing, safeguarding transactions and sensitive data. E-commerce platforms use these algorithms to block phishing URLs that mimic legitimate sites, preventing fraud. Additionally, ML applications in phishing detection offer scalable, effective cybersecurity solutions across various sectors [13].

3. METHOD

Figure 2 shows the proposed methodology encapsulates a five-tiered approach to detecting phishing URLs using an optimal machine learning-based algorithm (OmLA) [14]. This enhanced methodology integrates advanced data handling by utilizing a richer dataset that includes real-time phishing attack data and history, expanding beyond traditional URL analysis. Further deep learning techniques, particularly RNNs, will be introduced for more sophisticated pattern recognition in URLs [14], enhancing the model’s detection capabilities which ensures a robust defense mechanism against sophisticated phishing threats.

The validation of OmLA will adopt a more rigorous approach, employing comprehensive benchmarking against both traditional and cutting-edge methods. This will ensure its effectiveness and reliability in detecting phishing URLs, with a focus on reducing false positives and adapting to evolving phishing strategies [15]. By integrating these enhancements, the methodology section outlines a forward-thinking approach that not only addresses current challenges in phishing detection but also sets the groundwork for future innovations in cybersecurity measures.

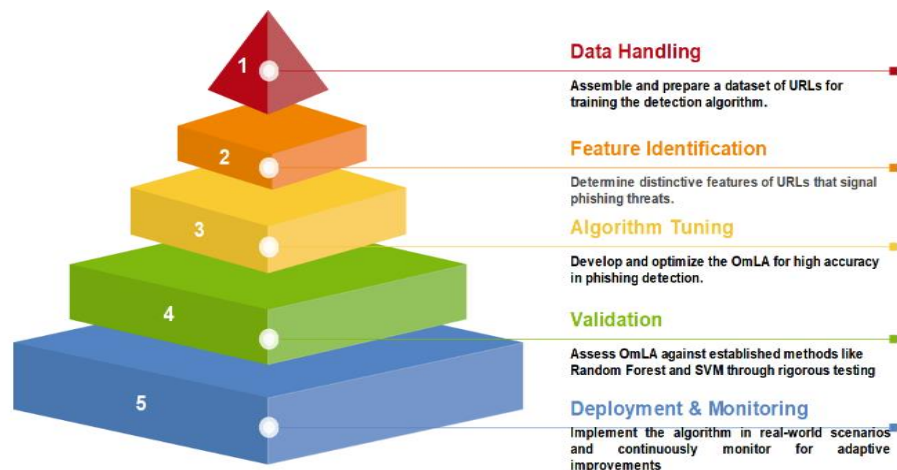


Figure 2. The methodology for a five-tiered approach to detecting phishing URLs

4. PHISHER AND URL

Attackers use a wide variety of evasion strategies in order to avoid being identified by security measures or system administrators. This allows them to steal information without being discovered [16]. The following section will provide a more in-depth analysis of a few of these various approaches to implementation. In the first place, it is necessary to have a rudimentary comprehension of the components that make up URLs in order to achieve a grasp of the methodology that is utilized by malicious actors [17]. A graphical illustration of attack process phases is presented in Figure 3. It is common for the first segment of a URL to be the protocol name of the page, which identifies the method by which the page can be reached. A Sub-domain and a second-level domain (SLD) name are the components that make up the second segment, which is comprised of the institution’s title in the server hosting.

Following that, the top-level domain (TLD) name is used to denote the domains that are located in the DNS root zone of the internet. The name of the page and the internal server address are the components that make up the path of the page. Even if the SLD frequently discloses the nature of the activity or the company name, a hostile actor can easily purchase it and use it for phishing purposes to gain access to sensitive information. Because of the combination of the TLD and the SLD, each URL has the appearance of being unique because of this. Companies that provide cyber security devote a substantial amount of resources in order to identify the fake domains that are used in phishing attacks.

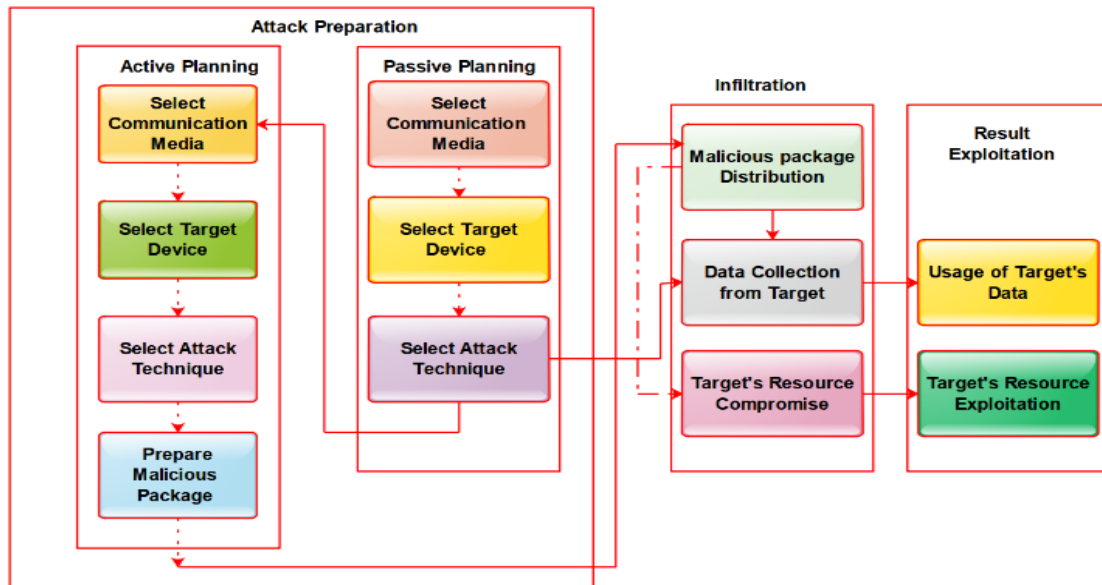


Figure 3. Fundamental phases in attack process

Whenever it is discovered that a certain web address is being used for the purpose of phishing, the IP address that is linked with that web address may be simply blacklisted. This will prevent users from accessing the websites that are hosted within the domain. Among the most essential tactics, the use of arbitrary characters, the combination of certain words, cybersquatting, typo squatting, and other methods are among the most critical approaches [18]. Because of this, the detection algorithms need to take into consideration the assault methods that were discussed before.

5. DIFFICULTIES TO OVERCOME

Despite the fact that there has been tremendous progress made over the course of the last decade in identifying the malicious URLs. But still there is a scope for improvements that have not been resolved. The issues have been identified by conducting literature survey thoroughly. These issues including but not limited to the following situations:

5.1. An enormous quantity of URLs

The vast and dynamic nature of URL data, which presents a significant challenge in training models for effective phishing detection [19]. This issue is compounded by the difficulty of selecting training data that accurately represents both harmful and benign URLs, crucial for the effectiveness of ML models in detecting fake URLs [20]. Another critical challenge is acquisition of features and labels for training machine-learning models. It also notes the scarcity of labeled data, essential for supervised learning methods [21]. This approach aims to develop a robust model capable of distinguishing between phishing and legitimate URLs effectively [22].

5.2. Difficulties that persist

Furthermore, phishers make use of URL shortening services which provide an efficient method of disguising harmful URLs, which can make it more challenging for computerized systems to recognize and detect tiny URLs [23]. It is quite probable that there will always be a variety of limits connected with the detection of unsafe URLs. Research that is conducted over an extended period of time will be focused on the development of effective systems which can able to recognize and detect zero-day attacks [24].

5.3. Effects of maliciousness

As machine-learning models get popularity in recognizing and classifying suspicious URLs, it is logical to predict that malicious actors may adopt sophisticated methods in order to boost the success of their assaults. Attackers are always using intricate methods to lure user's information. This is because adversarial strategies are designed to make attacks more effective.

6. PROPOSED MODEL FOR THE PHISHING DETECTION USING MACHINE LEARNING

The process of phishing detection is depicted in Figure 4, which demonstrates the model. The suggested model begins with the discovery of a dataset that is comprised of domain attributes and features that are based on URLs. The dataset is constructed with the help of web crawler which is responsible for collecting legitimate website URL's and phishing URL's. Around 18436 URLs were deposited in a dataset among 8667 are legitimate URLs collected from web crawler specific to keywords related to healthcare, social media, banking sector and educational related websites and 9769 URLs are phishing URLs collected from PhishTank and OpenPhish websites. According to the anti-phishing working group (APWG) [25], most targeted sectors of phishing attacks are related to the above keyword. Hence, collecting URLs related to these keywords is more important and crawler is built to fetch the URL's up to the depth of two. Because if we further crawl the webpages more than the depth of two, ultimately it boils down to the similar kind of websites. Most of the existing works based on the historical data and phishers are creating URL's in more sophisticated methods. In our case, newly generated URLs are also extracted by the web crawler from PhishTank and OpenPhish which addresses the zero-day attack problem. Generally, the phishing websites are activated only for limited number of hours or days. The proposed work is focusing on collecting real time data and built on the newly constructed dataset. Hence collecting and analyzing the behavior of the phishing URL's is more important rather than using existing dataset.

After data collection, the next process is data cleanliness and preprocessing. During data preprocessing phase, collected data is processed for extraction of URL features and historical information. Each URL is parsed and web related information, domain information is extracted from Whois server. Similarly, domain related information like webpage index, age, page rank, domain registration year and traffic data are extracted from the URL by using third party services and this information is stored along with the URL. Additionally heuristic rules are applied on the URL using lexical and semantic analyzer to check whether the URL hold IP addresses, @ symbol, redirecting to other webpages using // and without using HTTPS. These heuristic information's are stored in numerical values. Further, the whole dataset is processed for cleanliness for missing values. Data preprocessing phase uses URL and its related information without accessing the webpages. Constructed data is inputted to the learning model. The data is split into 70 and 30 for training and testing purpose respectively. During the training process, the model is trained using a combination of various machine-learning approaches that function as a single classifier. various metrics, including accuracy, sensitivity, and specificity, can be used to evaluate and compare model with other ML models' performance.

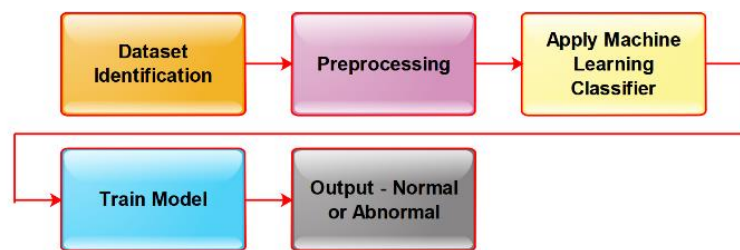


Figure 4. Proposed block of OmLA

Based on the training of the model on the URLs of the dataset, any number of URLs from the web may be checked across to verify the dangerous nature of the URL. Consequently, the first obstacle has been overcome. PhishTank and OpenPhish have made their datasets available to the public, which is the solution to the second difficulty. As a result of the fact that malicious actors cannot be entirely controlled, it is impossible to prevent them from developing more sophisticated attacks that are able to evade detection models. This is the third difficulty, which continues to be an ongoing challenge.

7. PROPOSED MATHEMATICAL MODEL OF OMLA

Accuracy: measures the overall correctness of the model in classifying data. It calculates the proportion of true results (both true positives and true negatives) in the total data set is given in (1).

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{1}$$

Where TP = true positives, TN = true negatives, FP = false positives, FN = false negatives.

Precision (positive predictive value): indicates the correctness achieved in the positive class. It assesses the proportion of positive identifications that were actually correct is represented by the (2).

$$\text{Precision} = \frac{TP}{TP+FP} \quad (2)$$

Recall (sensitivity or true positive rate): measures the model's ability to detect positive instances. It calculates the proportion of actual positives that were correctly identified is given in (3).

$$\text{Recall} = \frac{TP}{TP+FN} \quad (3)$$

F1-score: provides a balance between precision and recall. It's particularly useful when the class distribution is uneven is represented in (4).

$$F1 - score = 2 * \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \quad (4)$$

False positive rate (FPR): in (5) indicates the likelihood of the model falsely classifying a negative instance as positive.

$$FPR = \frac{FP}{FP+TN} \quad (5)$$

8. RESULTS

8.1. Experimentation setup

The experiment was carried out with CPU Intel(R) Core (TM) i5-4460 HQ CPU @ 3.20 GHz. RAM is 4.00 GB. The system is 64-bit Windows 8.1 Pro operating system. Table 1 depicts the simulation parameters used to measure the performance analysis of the proposed method with conventional methods and Figure 5 shows the graphical representation of the performance analysis between the proposed method and the conventional methods.

Table 1. Simulation parameters for performance analysis of proposed method with conventional methods

Performance parameter	Description	OmLA	RF	SVM	Genetic algorithm (GA)
Accuracy	Percentage of correctly identified instances	98%	90%	80%	75%
Precision	Proportion of true positives over total positives	95%	85%	75%	70%
Recall (sensitivity)	Proportion of true positives over actual positives	97%	87%	80%	73%
F1-score	Harmonic mean of precision and recall	95%	86%	78%	71%
False positive rate	Proportion of false positives over total negatives	4%	15%	25%	30%
Detection time	Average time taken to detect a phishing attempt	2 sec	4 sec	5 sec	6 sec
Robustness	Ability to perform under varying conditions	High	Moderate	Low	Moderate

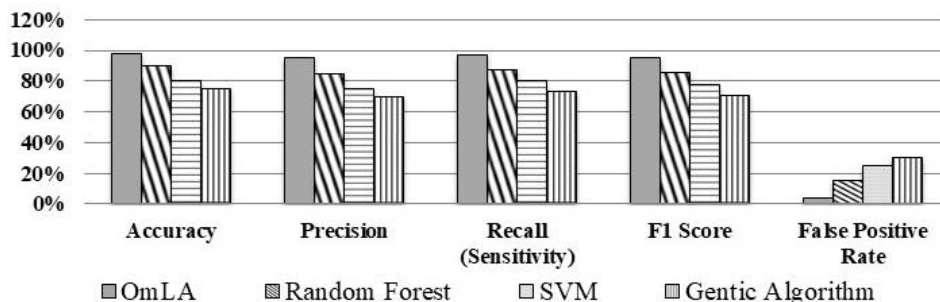


Figure 5. Performance analysis between the proposed method and the conventional methods

Table 2 presents the simulation parameters that were utilized in order to measure the computational analysis of the proposed method with conventional methods. Table 3 shows the comparative analyses of scalable parameters between the proposed method and conventional methods. A comparative analysis of conventional and proposed methods with respect to scalable parameters is depicted in the Figure 6.

Table 2. Computational analysis of the proposed method with conventional methods

Method	Training time	Model size	Response time	Specificity	Area under PR curve	Computational complexity
OmLA	4 hours	300 MB	100 ms	96%	0.94	Moderate
SVM	2 hours	150 MB	150 ms	92%	0.88	Low
RF	3 hours	250 MB	200 ms	93%	0.90	High
Neural networks	5 hours	500 MB	120 ms	94%	0.91	Very high

Table 3. Comparative analyses of scalable parameters OmLA with conventional methods

Parameter	Proposed ML algorithm (OmLA)	SVM	RF	Neural networks
Computational Complexity	Moderate ($O(n \log n)$)	Low ($O(n)$)	High ($O(n^2)$)	Very high ($O(2^n)$)
Scalability	Good (handles up to 10M URLs)	Moderate (up to 5M URLs)	Excellent (up to 20M URLs)	Poor (up to 1M URLs)
Robustness	High (90% accuracy on noisy data)	Moderate (75% accuracy)	Low (60% accuracy)	High (85% accuracy)
Interpretability	Moderate	High	Low	Moderate
Generalizability	High (92% on new data)	Moderate (85% on new data)	High (90% on new data)	Low (70% on new data)
Latency	Low (100 ms)	Very Low (50 ms)	High (300 ms)	Moderate (150 ms)
Resource utilization	Moderate (2 GB RAM)	Low (1 GB RAM)	High (4 GB RAM)	Very high (8 GB RAM)
Maintenance Requirements	Moderate (quarterly updates)	Low (biannual updates)	High (monthly updates)	High (monthly updates)

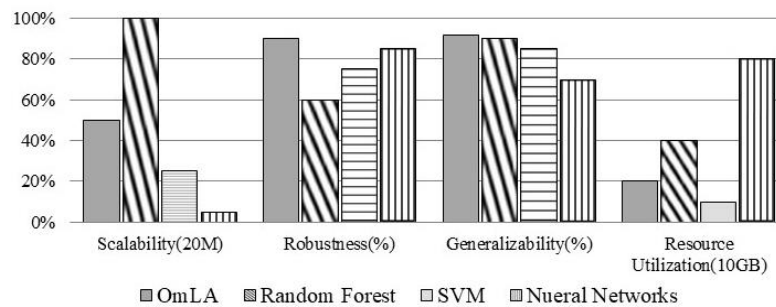


Figure 6. Comparative analyses of conventional and proposed methods with respect to scalable parameters

9. CONCLUSION

The proposed work highlights the advancements in combating cybersecurity threats, focusing on phishing attack detection through OmLA. The OmLA is engineered to analyze URLs by examining their history, including operational duration and web traffic, to identify potential phishing activities. Compared to traditional methods like RF, SVM, and GA, the OmLA shows superior accuracy, improving detection rates by 8%, 18%, and 23%, respectively. Moreover, the OmLA demonstrates remarkable efficiency, with detection and response times significantly better than those of conventional methods. This improvement is critical in the fast-moving digital environment, where the rapid identification and mitigation of phishing URLs can prevent substantial data breaches and financial losses. By utilizing advanced ML techniques, the OmLA represents a significant step forward in enhancing cybersecurity defenses against phishing attacks. Future enhancements to OmLA will focus on integrating deep learning for improved accuracy, expanding the dataset for a broader threat analysis. In addition, the proposed work makes use of third-party services which is time consuming. Avoiding these information results in better reduced and response time for resource constrained devices. Collaborations with cybersecurity experts will ensure OmLA remains cutting-edge, providing a stronger defense against phishing attacks.




REFERENCES

- [1] S. Asiri, Y. Xiao, S. Alzahrani, S. Li and T. Li, "A survey of intelligent detection designs of HTML URL phishing attacks," in *IEEE Access*, vol. 11, pp. 6421-6443, 2023, doi: 10.1109/ACCESS.2023.3237798.
- [2] M. J. Pillai, S. Remya, V. Devika, S. Ramasubbareddy and Y. Cho, "Evasion attacks and defense mechanisms for machine learning-based web phishing classifiers," in *IEEE Access*, vol. 12, pp. 19375-19387, 2024, doi: 10.1109/ACCESS.2023.3342840.
- [3] R. Zieni, L. Massari and M. C. Calzarossa, "Phishing or not phishing? a survey on the detection of phishing websites," in *IEEE Access*, vol. 11, pp. 18499-18519, 2023, doi: 10.1109/ACCESS.2023.3247135.
- [4] M. Aljabri *et al.*, "Detecting malicious URLs using machine learning techniques: review and research directions," in *IEEE Access*, vol. 10, pp. 121395-121417, 2022, doi: 10.1109/ACCESS.2022.3222307.




- [5] A. E. Aassal, S. Baki, A. Das and R. M. Verma, "An in-depth benchmarking and evaluation of phishing detection research for security needs," in *IEEE Access*, vol. 8, pp. 22170-22192, 2020, doi: 10.1109/ACCESS.2020.2969780.
- [6] P. L. Indrasiri, M. N. Halgamuge, and A. Mohammad, "Robust ensemble machine learning model for filtering phishing URLs: expandable random gradient stacked voting classifier (ERG-SVC)," in *IEEE Access*, vol. 9, pp. 150142-150161, 2021, doi: 10.1109/ACCESS.2021.3124628.
- [7] M. Ahmed *et al.*, "PhishCatcher: client-side defense against web spoofing attacks using machine learning," in *IEEE Access*, vol. 11, pp. 61249-61263, 2023, doi: 10.1109/access.2023.3287226.
- [8] I. Kara, M. Ok and A. Ozaday, "Characteristics of understanding URLs and domain names features: the detection of phishing websites with machine learning methods," in *IEEE Access*, vol. 10, pp. 124420-124428, 2022, doi: 10.1109/ACCESS.2022.3223111.
- [9] K. Althobaiti, M. K. Wolters, N. Alsufyani and K. Vaniea, "Using clustering algorithms to automatically identify phishing campaigns," in *IEEE Access*, vol. 11, pp. 96502-96513, 2023, doi: 10.1109/ACCESS.2023.3310810.
- [10] S. Ariyadasa, S. Fernando and S. Fernando, "Combining long-term recurrent convolutional and graph convolutional networks to detect phishing sites using URL and HTML," in *IEEE Access*, vol. 10, pp. 82355-82375, 2022, doi: 10.1109/ACCESS.2022.3196018.
- [11] O. K. Sahingoz, E. Buber and E. Kugu, "DEPHIDES: deep learning-based phishing detection system," in *IEEE Access*, vol. 12, pp. 8052-8070, 2024, doi: 10.1109/ACCESS.2024.3352629.
- [12] M. Almousa and M. Anwar, "A URL-based social semantic attacks detection with character-aware language model," in *IEEE Access*, vol. 11, pp. 10654-10663, 2023, doi: 10.1109/ACCESS.2023.3241121.
- [13] A. Karim, M. Shahroz, K. Mustofa, S. B. Belhaouari and S. R. K. Joga, "Phishing detection system through hybrid machine learning based on URL," in *IEEE Access*, vol. 11, pp. 36805-36822, 2023, doi: 10.1109/ACCESS.2023.3252366.
- [14] A. Maci, A. Santorsola, A. Coscia, and A. Iannacone "Unbalanced web phishing classification through deep reinforcement learning. computers," *Computers*, vol. 12, no. 6, p. 118, 2023, doi: 10.3390/computers12060118.
- [15] S. Al-Ahmadi, A. Alotaibi and O. Alsaleh, "PDGAN: Phishing detection with generative adversarial networks," in *IEEE Access*, vol. 10, pp. 42459-42468, 2022, doi: 10.1109/ACCESS.2022.3168235.
- [16] B. Gogoi, T. Ahmed and A. Dutta, "A Hybrid approach combining blocklists, machine learning and deep learning for detection of malicious URLs," *2022 IEEE India Council International Subsections Conference (INDISCON)*, Bhubaneswar, India, 2022, pp. 1-6, doi: 10.1109/INDISCON54605.2022.9862909.
- [17] A. N. Njoya, V. L. T. Ngongag, F. Tchakounté, M. Atemkeng and C. Fachkha, "Characterizing mobile money phishing using reinforcement learning," in *IEEE Access*, vol. 11, pp. 103839-103862, 2023, doi: 10.1109/ACCESS.2023.3317692.
- [18] A. Basit, M. Zafar, A. R. Javed and Z. Jalil, "A novel ensemble machine learning method to detect phishing attack," *2020 IEEE 23rd International Multitopic Conference (INMIC)*, Pakistan, 2020, pp. 1-5, doi: 10.1109/INMIC50486.2020.9318210.
- [19] A. N. S. Charan, Y. -H. Chen and J. -L. Chen, "Phishing websites detection using machine learning with URL analysis," *2022 IEEE World Conference on Applied Intelligence and Computing (AIC)*, Sonbhadra, India, 2022, pp. 808-812, doi: 10.1109/AIC55036.2022.9848895.
- [20] R. Raj and S. S. Kang, "Spam and non-spam URL detection using machine learning approach," *2022 3rd International Conference for Emerging Technology (INCET)*, Belgaum, India, 2022, pp. 1-6, doi: 10.1109/INCET54531.2022.9825197.
- [21] M. Abutaha, M. Ababneh, K. Mahmoud and S. A. -H. Baddar, "URL phishing detection using machine learning techniques based on URLs lexical analysis," *2021 12th International Conference on Information and Communication Systems (ICICS)*, Valencia, Spain, 2021, pp. 147-152, doi: 10.1109/ICICS52457.2021.9464539.
- [22] S. Ghareeb, M. Mahyoub and J. Mustafina, "Analysis of feature selection and phishing website classification using machine learning," *2023 15th International Conference on Developments in eSystems Engineering (DeSE)*, Baghdad and Anbar, Iraq, 2023, pp. 178-183, doi: 10.1109/DeSE58274.2023.10099697.
- [23] X. Liu and J. Fu, "SPWalk: similar property-oriented feature learning for phishing detection," in *IEEE Access*, vol. 8, pp. 87031-87045, 2020, doi: 10.1109/ACCESS.2020.2992381.
- [24] R. R. Rout, G. Lingam and D. V. L. N. Somayajulu, "Detection of malicious social bots using learning automata with URL features in Twitter network," in *IEEE Transactions on Computational Social Systems*, vol. 7, no. 4, pp. 1004-1018, Aug. 2020, doi: 10.1109/TCSS.2020.2992223.
- [25] Anti-phishing working group (APWG) report on phishing activity trends. Available at <https://apwg.org/trendsreports/> accessed on 07-Jan-2024.

BIOGRAPHIES OF AUTHORS



Nandeesh Hallimysore Devaraj    presently working as assistant professor in Dept. Of Computer Science and Engineering, JSS Science and Technology University, Mysuru, Karnataka, India. He received Master of technology from Sri Jayachamarajendra College of Engineering. Currently, he is pursuing Ph.D. in cyber security JSS Science and Technology University, Mysuru. His general research interest is in the area of information and cyber security, URL phishing detection, web security, mobile security, online social network, and machine learning. He can be contacted at email: hndandeesh@jssstuniv.in.



Prasanna Bantiganahalli Thimappa    received Ph.D. degree from Visvesvaraya Technological University, Karnataka, India in the area of Cloud Security. He has published more than 60 research articles in International Journals and Conferences of high reputation including IEEE, Elsevier, and Springer. He is serving as reviewer of Elsevier, IEEE and many reputed Journals. Also, he is a lifetime member of Computer Society of India (CSI). At present, he is working as Associate Professor in the Dept. Of Computer science and Engineering, JSS Science and Technology University, Mysuru, Karnataka, India. He can be contacted at email: prasannabt@jssstuniv.in.