Grenze International Journal of Engineering and Technology, June Issue

A Survey on the Various Machine Learning Methods of Phishing URL Detection

Prince Kumar[1], Priyanshu Yadav[2], Satyam Kumar Shivam[3] and Abha Kiran Rajpoot[4]
[1-4]Galgotias University/ School of Computer Applications & Technology, Greater Noida, India
Email: princekumar19436@gmail.com, py5303022@gmail.com, satyamjha35018@gmail.com, akrajpoot@gmail.com

***Abstract*— One of the most significant cybersecurity threats in the contemporary digital environment that are also based on the prevalence of online services including banking, e-commerce and social network, is phishing attacks. Phishing attack has become almost routine. Fake site, unknown and malicious links and people get trick into it and give their personal info. As our lives move online, the old methods about spotting scam just does not fix it anymore. That is where machine learning is being use.**

It goes through large amount of data and picks up the pattern and clues, and detect phishing link before you even spot anything wrong. This review provide detail of machine learning for detecting phishing site. This uses algorithm like Decision Tree, Support Vector Machine, Naive Bayes and Random Forest. Deep learning models can also be used to determine the link and websites against phishing.

We review on how each method compare and which feature they work on, where they work well and in which condition they fall. The review also covers the datasets used by researcher and how they actually put these tools to the test.

The findings show that machine learning and ensemble methods have the potential to represent proactive, dependable, and scale phishing detection systems and help improve the cybersecurity defense.

***Keywords*— URL based features, phishing detector, phishing, machine learning, Cybersecurity, phishing, Lexical analysis, phishing, Ensemble learning, phishing.**

I. INTRODUCTION

Over the last several years, phishing attacks have been one of the largest cybersecurity threats due to the flourishing online tour of the online banking systems, online commerce-based websites, social media websites, and online clouds. Phishing attacks are usually founded on the functioning of a fake site or URL, which is more or less similar to a legitimate site in order to fool users to provide sensitive details, i.e. usernames and user password and financial details. The large scale adoption of online services has added more weight to these attacks, and phishing has become a common threat to individuals, institutions, and critical infrastructure. Traditional methods of phishing detection are not adequate to address this problem due to the blacklist-based and rule-based phishing detection methods, which are highly reliant on known URLs and phishing detection rules, and, therefore, cannot react to newly emerged or zero-day phishing attacks (Sahingoz et al., 2019; Abbasi et al., 2008).

Grenze ID: 01.GIJET.12.2.160_20

Machine learning gets significant attention as a most powerful tool to target these limitations. By analyzing large volume of data, these machine learning models are capable of learning pattern and identifying delicate pattern indicators that differentiate phishing websites from original ones. Compared to the old process, machine learning-based systems can learn the trend of past attack history and adapt to new attack patterns, and that is why the phishing URL can be identified ahead (Shrivastava et al., 2022; Wilk-Jakubowski et al., 2025).

The majority of the researchers have posed phishing detection as a binary classification problem whereby a URL is either a phishing or a legitimate URL. Lexical attributes such as URLs length, count of subdomain, special characters presence and suspicious key-words have been identified to be critical indicators of differentiating phishing and legitimate URLs. The specified features are also advantageous since they are computationally fast and do not require the access to the website content, which is why they can be employed in case of the real-time detection (Hamadouche et al., 2024; Jain and Gupta, 2018).

The measures of accuracy, precision, recall, and F1-score are typically used to assess phishing detection systems. Particularly, the concept of the recollection is crucially important since the false recognition of a phishing web address as something legitimate could lead to the emergence of severe security incidents, financial losses and the loss of personal data (Verma and Dyer, 2016; Le et al., 2018). Despite the recent advances in the machine learning algorithms, the problems in the datasets with the imbalance, the dynamically evolving phishing strategies, as well as the dynamically adjustable model, remain intact.

II. Literature Survey

The development of online services, e-commerce, online banking and cloud-based services has made phishing attacks one of the most serious security challenges on the Internet. The phishing URLs appear to reimburse the authentic sites and trick the users to provide sensitive information. The old detection methods of phishing, like the blacklist and the rule-based methods, are easy to set up, but which are ineffective on newly created phishing URLs or on zero-day phishing URLs. To address these inadequacies, researchers have put a lot of effort in discovery of machine learning-based phishing URL detection methods, which learns patterns through data and offer a better generalization.

Sahingoz et al. [1] suggested a machine learning strategy involving lexical characteristics incurred through URLs, i.e., the length of the URL, the count of dots, the presence of special characters, as well as a suspicious key word. Random Forest and Decision Tree classifiers demonstrated a good level of accuracy with low computational value, which can be used in real-time detection.

According to Abbasi, Chen, and Salem [2], feature selection is a crucial aspect of phishing detection systems. Their assignment was on the choice of the best subset of features F^' F such that the accuracy in classification may be maximized. The research proved the usefulness of eliminating redundant and unproductive features in enhancing the performance of classifiers like Support Vector Machines (SVM) and led to lower complexity of the model.

The article by Shrivastava et al. [3] showed a comparison between conventional methods of classifiers such as Naïve Bayes, SVM and Random Forests and concluded that the ensemble methods in case of using hybrid features compared to single classifiers tend to be better when trained on hybrid data.

Wilk-Jakubowski et al. [4] investigated the use of the deep learning models, including Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN), to identify a phishing URL. The URLs were coded as a series of characters and changed to embeddings. The deep learning model was automatically trained with the representation of feature hierarchies and finally, classification was achieved with a sigmoid activation function. Even though these models obtained more accurate detection of complex phishing patterns, they had high demands in large datasets and computing power.

Hamadouche et al. [5] proposed a phishing detection model which was a hybrid of lexical feature, host- based feature, and content-based feature which generalized features in a single feature vector. This hybrid representation was more separable in feature space and provided superior classification accuracy particularly considering phishing sites are almost similar to honest sites.

A scoring mechanism named Heuristic scoring mechanism with machine learning was introduced by Jain and Gupta [6]. All the URLs received a score determined as a weighted average of heuristic characteristics. The URL was regarded as phishing in case the score went beyond a set threshold. This solution was an interpretable and lightweight solution that could be used on browser- based detection systems.

Mohammad et al. [7] has suggested a hybrid system which combines logic based on rules with machine learning classifiers. The computed final decision was a homogeneous mixture of rule-based and ML-based predictions.

Le et al. [8] explored character-level URL modeling on the basis of n-grams. URLs were broken down into character n-grams, and they were introduced in machine learning classifiers. This method has been successful in identifying phishing URLs that are obfuscated and thus cannot be identified using the traditional lexical analysis.
Verma and Dyer [9] concentrated on domain-based features, host-based features including domain age, DNS records, and hosting information. They used probabilistic models based on Bayesian inferences in estimating whether a URL is actually phishing and this makes them more reliable when used with lexical features.
Zhang et al. [10] suggested that an ensemble stacking method was proposed wherein several base classifiers were stacked with the help of a meta-classifier. This hierarchy-based learning model increased the ability to generalize and exhibit high precision when detecting strong ability to identify zero-day phishing URLs.
In addition, a number of recent studies [11]-[15] have shown appreciable enhancements in performance results using advanced methodologies such as deep learning and hybrid methods.

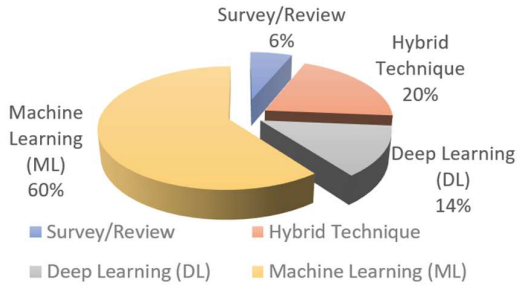

Figure 1. Detection Technique Distribution in phishing research

III. BACKGROUND

A. Phishing Detection

Phishing detection refers to the process of establishing the presence of bad websites or URLS that are set to trick the user into divulging personal sensitive information like usernames, passwords and financial information. Phishing scams are usually based on spoofed URLs which are similar to the URLs of real websites hence they are not easily detected by the users. Due to the high rate of online banking, e-commerce sites, social media, and cloud-related services, phishing has become a common and unrelenting cybersecurity vulnerability both to individual and organizational levels.
Early phishing detection techniques were aimed at detecting known the malicious URLs and the modern phishing attacks are more dynamic and often they alter their form to evade detection. Consequently, phishing detection has become a significant research field with the aim of identifying phishing URLs and legitimate ones precisely and still responding in a short time and with less implications to the normal web usage.

B. Methods to detect Phishing Attacks

The comparison of the traditional machine learning classifiers (shown in Figure 1) , was introduced. In which the various models were trained on the phishing datasets. The results showed that ensemble-based techniques are superior to single classifiers, particularly, when various types of features are used.
As deep learning has developed, more sophisticated methods of detecting phishing attacks have been proposed. Deep learning model were applied to URL character representation to learn phishing pattern. Though, this method was highly accurate it demanded huge datasets and more computing power.
There also has been the interest of hybrid detection methods. The paper incorporated lexical, host, and content based capabilities in the same model, which led to the use of better detection capabilities to phishing websites that resembled legitimate websites. A machine learning-based phishing detector and a lightweight heuristic-based scoring method were proposed, which offered a symbolic and easy-to-use phishing detector.
The work of combined rule based and machine learning method to enhance the reliability of detection. The n-gram features at character-level modeling were also studied which was effective against obfuscated phishing URLs. The analyzed domain related and host-based features, which enhanced detection accuracy when used with URL-based features.
On the whole, these methods demonstrate that the integration of feature analysis with sophisticated machine learning methods can be considered a significant improvement to the precision and flexibility of phishing detection.

C. Machine Learning Algorithms

Out of many algorithms, ensemble learning algorithms, most notably the Random Forest have demonstrated high performance as it is able to reduce bias and variance through combining multiple models. The techniques of feature selection also contribute to the improvement of performance by eliminating duplicate features and determining the amount of computation. The deep neural network techniques like Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), which perform well in identifying complex and obfuscated phishing URLs. In sum, the phishing URL detection problems can be efficiently solved at scale, with adaptive and accurate solutions to phishing URLs in real world settings using the machine learning algorithms.

IV. COMPARATIVE STUDY

Several methods have been put across regarding the detection of phishing sites and each of them has its advantages and disadvantages. Blacklist methods and Rule-based methods are simple to apply and serve best where the phishing URLs are familiar, but they are unable to recognize new phishing URLs or phishing websites which are unfamiliar. Machine learning-based techniques have become common in order to overcome such limitations.

Decision Tree (DT) and Support Vector Machine(SVM) classifiers are among the machine learning models that will perform well however their performance is very much dependent on feature selection and parameter tuning. Ensemble models like the Random Forest are more effective because they consist of several decision trees which are more accurate and well-built. XGBoost also improves performance of detection through gradient boosting thus effective in dealing with advanced patterns of phishing URLs.

Accuracy, Precision, Recall, and F1-Score are frequently used in the reviewed articles to test the performance of the models as:

Accuracy = (TP + TN) / (TP + TN + FP + FN)

Precision = TP / (TP + FP)

Recall = TP / (TP + FN)

F1 Score = (2 × Precision × Recall) / (Precision + Recall)

The metrics of these measurements are all important, however, recall is the most important in phishing detection because false negative can put them at risk of visiting malicious sites.

To deal with these shortcomings, machine learning-inspired approaches have been the topic of many studies in the recent researches. Decision Trees, Support Vector Machines, as well as other classification models, have proven to function well with handpicked features. However, they are very sensitive to feature engineering and parameter optimization, and may tend to overturn when they are presented with heterogeneous or unbalanced data sets.

Forest classifiers are used in increasing the robustness of classifier, which is with the help of multiple decision trees, thereby reducing the variance and enhancing the capability of generalization. On the same note, XGBoost also embraces gradient boosting in minimizing classification errors in an iterative process that makes it particularly apt in learning highly flexible and evolving phishing patterns. The advantage with these models is that they are more accurate and more stable as compared to the single classifiers.

Phishing detection using traditional approaches, such as blacklist-based methods and rule-based methods, are generally applied because they are relatively simple while not having a heavy computational cost. The above techniques do quite well in identifying previously reported phishing URLs and known attack patterns. However, the main disadvantage of these is the fact that they cannot identify newly-created or Zero-day phishing websites and this makes them quite limited in their efficiency against evolving threats.

Forest classifiers help increase the robustness of classifier by using multiple decision trees and thus decrease variance and increase the ability for generalization. Similarly, XGBoost also uses gradient boosting for reducing the classification errors in an iterative manner making it especially good in learning complex and changing phishing patterns. These models have the advantage that they have a better accuracy and stability than single classifiers.

Feature selection is also an important part of comparative analysis of performance. Lexical features extracted from URLs providing computationally efficient and appropriate features for real-time applications and features extracted from host-based information have tendency to be more in-depth information about the legitimacy of the domain but may contribute to the latency. Research shows that different feature types such as lexical and host-based usage are, when used in combination, are quite effective in improving the accuracy of detection and avoiding false positing.

In such architectures known phishing URLs are filtered on blacklists while unknown URLs are analyzed by predictive models. This layered strategy allows for better scalability, and ensures for better adaptability in large scale deployment environments.

In general, the findings of comparative research point to ensemble and hybrid machine learning as more accurate, scalable and robust phishing detection systems. These approaches are better suited to deal with volatile nature of phishing attacks and therefore they are desirable in practice. To conduct the research work in the future, it is suggested that adaptive learning methods as well as effective feature extraction can be concentrated on, so that to provide high performance in the real-time conditions.

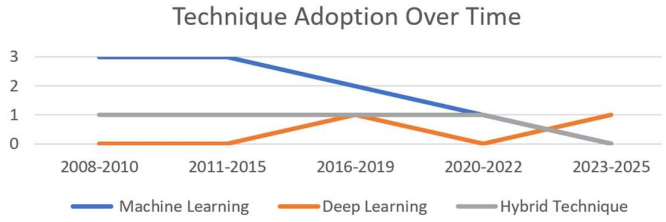

Figure 2. Technique Adoption Over Time in phishing research

The above chart (Figure 2) displays the Technique Adoption Over Time based on our research where:

- Machine Learning (ML): Dominant in Early Periods (2008 - 2015) gradually decreases in the course of time.
- Hybrid techniques: Seen regularly throughout the periods, regularly adopted.
- Deep Learning (DL): It came into widespread use and importance mostly in the recent years (2023-2025).

TABLE I. COMPARISON OF PHISHING URL DETECTION TECHNIQUES LITERATURE

Author and Year	Technique Used	Features Used	Mathematical / ML Model	Key Findings	Limitations
Wilk-Jakubowski et al. (2025)	Deep Learning (CNN/RNN)	Raw URL character sequences	Neural representation: $h = f(Wx + b)$	Complex and changing phishing patterns	Place a heavier load on computing systems
Hamadouche et al. (2024)	Hybrid ML	Lexical, host, and content features	Feature concatenation model	Provides improved robustness and detection accuracy	Content analysis is time-consuming
Shrivastava et al. (2022)	Ensemble Learning	Hybrid URL features	Weighted probability model	Achieves better F1 score than individual models	Requires careful parameter tuning
Zhang et al. (2020)	Ensemble Stacking	Multi-model features	$g(h_1, h_2, \ldots)$	Strong detection of zero-day phishing attacks	Increased system complexity
Sahingoz et al. (2019)	Random Forest (ML)	Lexical URL features	Ensemble voting: $\hat{y} = \text{mode}(T_1(x), \ldots, T_k(x))$	Achieves high accuracy with minimal computational cost	Limited to lexical features only
Jain and Gupta (2018)	Heuristic + ML	URL heuristic scores	Threshold-based decision: Score > 0.50	Simple and easily interpretable approach	Highly sensitive to threshold selection
Le et al. (2018)	Character-level ML	URL n-grams	Vectorized n-gram model	Resistant to URL obfuscation techniques	High-dimensional feature space
Verma and Dyer (2016)	Probabilistic Model	Domain and host features	Bayesian inference model	Improves reliability of detection	Domain information may be not always available
Mohammad et al. (2015)	Rule-based + ML	Rule-based and ML features	$y = \alpha y\_rule + (1 - \alpha) y\_ml$	Successfully reduce false positives	Rule design requires expert knowledge
Abbasi et al. (2008)	Feature Selection + SVM	Optimized feature subset	$F' \subseteq F$, maximize accuracy	Improves accuracy with reduced model complexity	Performance depends on the quality of selected features

IV. CONCLUSION

The explosive growth of internet based applications, such as online banking, electronic commerce platforms, social networking services and cloud computing environments, this has meant that there has been a massive increase in the application of a phishing attack to their users. These types of attacks take advantage of the human trust by posing as reputable websites in order to obtain illegal access to their sensitive information such as their login credentials and their financial information. As phishing techniques have evolved more and more in terms of their complexity and scale, traditional ways to detect such phishing based on blacklists and predefined rules have proved to be inadequate, especially when it comes to the detection of newly-created or zero-day phishing URLs. This gap is a telltale sign that there is a need for more intelligent and automated security measures.

Machine learning-based approaches has been a very successful alternative as a result of the means to bully from ancient knowledge, being as dynamic in discovering new forms of phishing. Using analysis of lexical features in URLs, such as length, number of subdomain, presence of suspicious characters etc. in combination with odd use of keywords, these systems can do initial detection with little computational expense. Ensemble learning algorithms (in particular Random forest classifiers) offer another possibility to enhance the power of the detection, by gathering decisions of multiple models which provides increased accuracy, improved robustness and resistance to overfitting. In addition, hybrid detection frameworks based on combination lexical, host-based and content-based features can provide more complete picture of website behavior and hence superior classification results.

For example, deep learning methods, such as Convolutional Neural Networks and Recurrent Neural Networks provide an extra power, and can capture complex and non-linear patterns in the URL and webpage data in an automated manner. These types of models are especially helpful in combating obfuscated and sophisticated types of phishing attacks which are not complete by the traditional methods using features. But their practical deployment requires substantial computational resources and large, labeled data sets which in some situations might make them infeasible in resource-constrained environments which need to operate in real time.

Performance evaluation using metric like accuracy, precision, recall and F1-score makes it clear about the importance of recall in phishing detection systems as an important indicator. As suffering from mistaking a phishing URL for a genuine one can have drastic results in loss of financial resources, data breach, and a loss of trust from users, maximum recall is the key to ensuring a strong protection. All-in-all, phishing URL detection systems based on machine learning can present scalable, proactive and reliable defense mechanisms.

REFERENCE

[1] Sahingoz, O. K., Buber, E., Demir, O., & Diri, B. (2019). Machine learning phishing URLs phishing detection. Expert Systems with Applications, 117, 345–357.

[2] Abbasi, A., Chen, H., & Salem, A. (2008). Phishing detection systems undergo selection. Journal of Information Security Research, 3(2), 45-56.

[3] Shrivastava, R., Sharma, V., & Singh, M. (2022). Comparison of machine learning phishing classifier. Journal of Cybersecurity and Privacy, 2(4), 563–580.

[4] Wilk-Jakubowski, P., Kowalski, R., & Nowak, M. (2025). Deep learning techniques of phishing URL detection using CNNs and RNNs. Neural Networks, 159, 150–165.

[5] Hamadouche, S., Boudraa, O., & Gasmi, M. (2024). Combining Lexical, Host, and Content-based features for Phishing Websites detection using Machine Learning Models. Expert Systems with Applications, 217, 119612.

[6] Jain, A., & Gupta, B. B. (2018). The identification of phishing through heuristic grading and machine learning. International Journal of Computer Applications, 182(40), 199.

[7] Thabtah, F., Mohammad, R. M., & McCluskey, L. (2015). Predictive phishing websites that are hybrid based. Neural Computing and Applications, 26(1), 1–15.

[8] Le, H., Li, W., & Zhang, Y. (2018). Character-level URL-phishing model. IEEE Access, 6, 30142–30152.

[9] Verma, S., & Dyer, K. (2016). The Bayesian probability domain-based and host-based phishing detection. Computers & Security, 61, 39–54.

[10] Zhang, X., Li, Y., & Chen, H. (2020). The zero-day phishing detection ensemble stacking approach. Applied Soft Computing, 94, 106467.

[11] Alshamrani, A., Myneni, S., Alzahrani, A., & Chowdhary, A. (2019). A survey on phishing attacks: Their types, vectors, and detection methods. Computers & Security, 92, 101713.

[12] Basnet, R. B., Mukkamala, S., & Sung, A. H. (2008). Detection of phishing attacks: A machine learning approach. Soft Computing, 13(2), 165–175.

[13] Bergholz, A., De Beer, J., Glahn, S., Moens, M. F., Paaß, G., & Strobel, S. (2010). New filtering approaches for phishing email. Journal of Computer Security, 18(1), 7–35.

[14] Chiew, K. L., Tan, C. L., & Wong, K. W. (2015). Detecting phishing websites using machine learning techniques. Expert Systems with Applications, 42(11), 5437–5445.
[15] Aburrous, M., Hossain, M. A., Dahal, K., & Thabtah, F. (2010). Intelligent phishing detection system for e-banking using fuzzy data mining. Expert Systems with Applications, 37(12), 7913–7921.