

URL-Based Phishing Detection: A Comprehensive Review and New Taxonomy

Rafal Abd-Alkadhim Mohammed^{1,*}, Riyadh Rahef Nuiiaa Alogaili², Ahmed Raad Al-Sudani³

¹Department of Software, College of Computer Science and Information Technology, Wasit University, Al-Kut, Wasit, Iraq. Email: stdm.rmohammed@uowasit.edu.iq

²cybersecurity Department, College of Computer Science and Information Technology, Wasit University, Al-Kut, Wasit, Iraq. Email: riyadh@uowasit.edu.iq

³Department of Software, College of Computer Science and Information Technology, Wasit University, Al-Kut, Wasit, Iraq. Email: araad@uowasit.edu.iq

ARTICLE INFO

Article history:

Received: 26 /03/2026

Revised form: 02 /05/2026

Accepted : 04 /05/2026

Available online: 30/09/2026

Keywords:

Phishing URL Detection
Cybersecurity
Natural Language Processing(NLP)
Metadata analysis
Machine Learning Techniques
Deep Learning Techniques

ABSTRACT

Phishing attacks remain to be classified among one of the most severe and persistent types of cyber threats, introducing a substantial danger to the contemporary online infrastructure. Attackers frequently utilize spoofed URLs to mislead users into disclosing confidential information. Accelerating evolution of phishing strategies, facilitated by AI-driven mechanisms, diminished the reliability of conventional detection approaches, such as blacklist and content-based methods, making these approaches inadequate for real-time deployment. This paper provides a comprehensive review of URL-based phishing detection techniques, with primary focus on Deep Learning, Machine Learning, Natural Language Processing (NLP), and metadata-based approaches. The paper analysis demonstrates the most existing studies rely on isolated feature-sets, restricting their ability to address the sophistication of the contemporary phishing attempts. Consequently, there is a demand for lightweight, real-time phishing detection mechanisms. This paper proposes a lightweight and real-time URL phishing detection method that is based on the concept of integrating Natural Language Processing (NLP) and metadata analysis. In response, this study emphasizes on the integration of NLP-based semantic analysis, accompanied with metadata-based attributes. The integration of these approaches may enhance the capability to discriminate between legitimate and malicious URLs. It may also maintain relatively low computational overhead. Overall, this review highlights the importance of hybrid feature-based, and establishes a foundation for the implementation of computationally-efficient, scalable, and real-time phishing detection frameworks for web applications.

<https://doi.org/10.29304/jqcm.2026.18.32760>

*Corresponding author: Rafal Abd-Alkadhim Mohammed

Email addresses: stdm.rmohammed@uowasit.edu.iq

Communicated by 'sub etitor'

1. Introduction

The digital landscape is experiencing rapid progress due to the growing dependence on various internet-based services. However, such expansion can facilitate the exploitation of adversaries for digital resources, notably personal data, for malicious intents. As a result, online platforms have become more susceptible to a wide range of cybersecurity threats [1]. The ubiquitous nature of the digital environment has significantly amplified the exposure to various cyberattacks conducted by malicious actors. These adversaries pursue diverse objectives and intents; some aim to disrupt services, while others focus on exploiting personal information to gain unauthorized access to user's personal data. This is commonly referred to as phishing attack. A phishing attack is one of the most prevalent and impactful categories of cyber-attack in the contemporary cybersecurity landscape. The consequences of successful phishing incidents encompass, (but are not confined to), financial losses, reputational damage, and identity theft [2]. Phishing attacks have escalated in both number and sophistication, particularly in the recent years with the deployment of diverse AI techniques, including machine learning, deep learning, hybrid learning, scenario-based methods, facilitating attackers to extract sensitive and confidential information via fraudulent emails and counterfeit websites, thus these threats are still growing at an exponential rate [3]. Figure 1 shows the most targeted sectors in the fourth quarter 2025 based on APWG statistics.

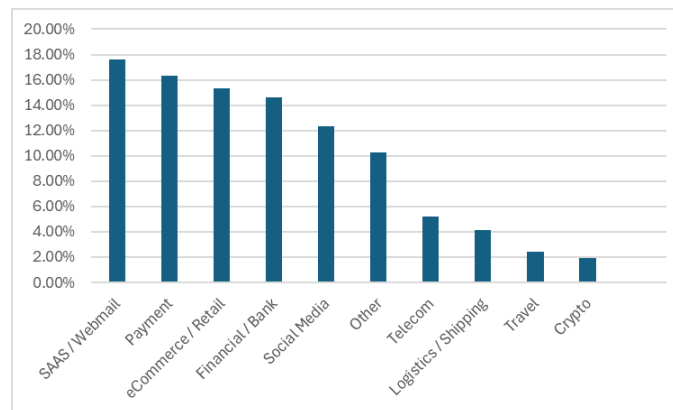


Fig. 1 Most Targeted Sectors in 4th quarter of 2025 [4].

The primary focus of this paper is to detect phishing attempts from a given URL by analyzing specific characteristics that distinguish malicious URLs from benign ones. A URL (Uniform Resource Locator) is recognized as a fundamental entry point for phishing schemes, as it serves as a preliminary point during the interaction between the user and the web-based service. URL structures are manipulated by attackers to mislead victims into perceiving them as legitimate websites. This manipulation may involve inserting arbitrary characters to the original URL and replacing certain letters with homoglyphs from the Greek alphabet; for example, the letter “a” can be replaced with the alpha symbol “α”. In most cases, users may fail in the recognition of these subtle alterations in URL, thus increasing the likelihood of revealing private information to unauthorized parties. As a result, attackers may successfully compromise confidential data obtained from victims. Detection methods based on URL analysis work by checking literal and conceptual inconsistencies. Lexical URL verification can require analyzing embedded URL components, without the need to incorporate additional attributes or features for analysis and validation. This is accomplished the strict

compliance to the precise textual representation of the URL. The term "conceptual consistency" refers to the classification of every letter, number, and symbol in URL into several categories to facilitate comprehensive analysis and prevent false classification [5]. Figure 2 shows the architecture of URL Phishing.

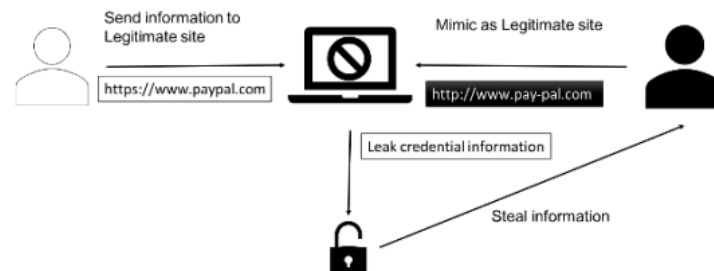


Figure 2: Architecture of URL Phishing [6].

Protecting sensitive data from malicious adversaries requires robust and effective mechanisms. In recent years, attackers started leveraging sophisticated artificial intelligence techniques to bypass traditional phishing detection systems. Therefore, there exists a critical need for advanced approaches that could enhance the quality of detection in order to stay updated with current malicious attacks [7].

Recent studies have integrated Natural Language Processing (NLP) methods to enhance the quality of URL analysis, as well as detecting phishing in real time. NLP is a subfield of Artificial Intelligence that is based on the connection between computers and humans using natural language. NLP is broadly known as robust technique for processing text-based data and recognizing malicious intent conveyed through digital communication systems. NLP techniques enable the extraction of semantic, syntactic, and contextual features that can distinguish phishing URLs from authentic correspondence [8].

Leveraging metadata in URL phishing detection could be quite promising, since it provides a robust real-time solution for analyzing URL structure. Metadata is simply data about data; metadata works by checking the URL's descriptive attributes that can provide information about the URL, including its length, the total number of characters, and other features that could indicate it is a suspicious URL. Namely, metadata measures the entropy of a given URL; entropy is the percentage of URL randomness. It is mainly the haphazard placement of characters and letters. [9].

2. Review Methodology

This review was conducted using a systematic approach ensuring the relevance as well as high-quality of studies in the realm of URL phishing detection, a well-structured and comprehensive literature review was performed across multiple reliable academic databases, including Google Scholar, IEEE Xplore, Science Direct, and Springer Link. The search was performed by using keywords such as phishing URL detection, machine learning, deep learning, NLP, and metadata phishing detection

The inclusion criteria were defined as follows:

- Studies published between 2020 and 2025

- Peer-reviewed journal and conference papers
- Research focusing on URL-based phishing detection

The exclusion criteria:

- Studies not related to phishing URL detection
- Articles published outside the selected time frame (before 2020)
- Studies lacking sufficient technical or methodological detail
- Non-peer-reviewed sources such as blogs, reports, or white papers
- Duplicate or overlapping studies
- Studies not aligned with ML, DL, NLP, or metadata-based approaches
- Studies focusing solely on user behavior or webpage content without URL analysis

3. Background

Phishing attacks are one of the most pervasive types of cyberattacks; their prevalence has made them a significant threat to companies, organizations, and individuals. They can be carried out in different forms. Some types are based on psychological manipulation, where attackers convince their victims to disclose their personal information. Information may encompass their name, email address, residential address, or any other type of sensitive information. In such a form, attackers usually try to act as a legitimate party; for instance, they may pretend to be personnel of a certain company or any authorized individual who has the right to obtain access to their data. People, especially those who lack awareness, might be easily tricked by this form of attack; thus, conducting an awareness program for people is mandatory in this situation to reduce the probability of being deceived by attackers. Figure 3 shows the main phishing detection methods.

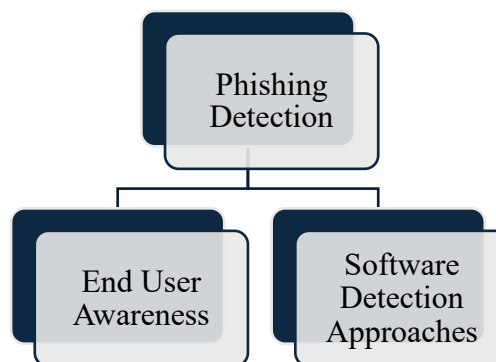


Fig 3: shows the division of phishing detection methods.

The phishing attack rate has witnessed a tremendous exponential growth in recent years, considering the dependence of people on doing most of their work on the internet, as well as shopping and buying various goods through online websites, registering on online applications, and other services. Thus, the chance of being a victim of a phishing attack has increased concurrently; attackers have taken advantage of this heavy reliance on the digital world. Figure 4 shows the growth of phishing attacks in the second quarter of 2025 based on APWG's statistics.

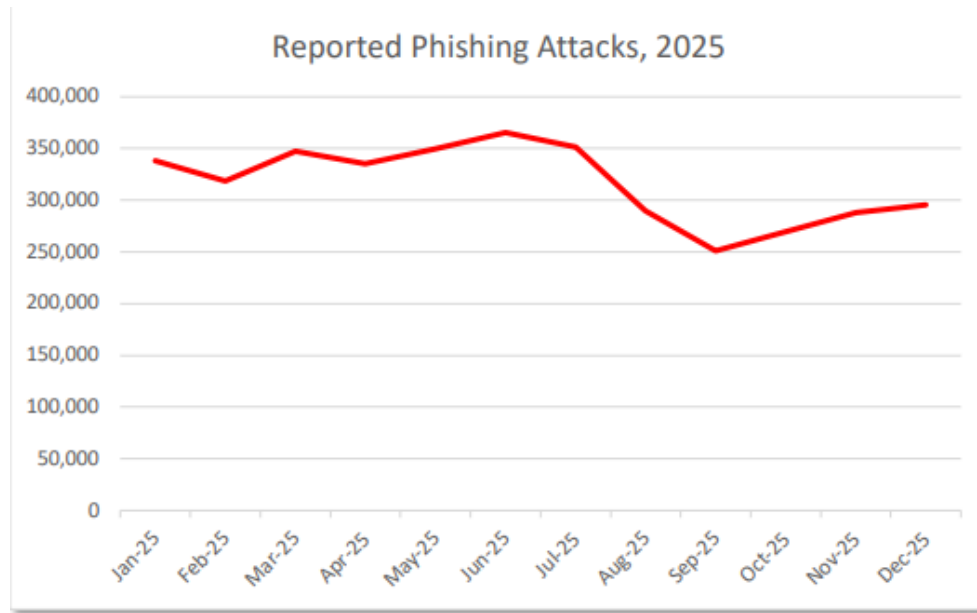


FIG 4: The Growth Rate of Phishing Attack in the Fourth Quarter of 2025 [4].

This research paper is oriented towards detecting phishing in URLs. URL is abbreviated from 'Uniform Resource Locator'. URL serves as a principal gateway to varied cyberattacks; thus, attackers make use of URLs by mimicking the original benign URL and adding minute changes that might be unnoticeable to the victims [10]. Upon accessing the URL, the target is redirected to a counterfeit login page, where they input their credentials as they would on the authentic login page. The phisher thereafter acquires the unsuspecting victim's credentials and access to their user account and private information to perpetrate identity theft, data theft, or financial crime [11].

The main structure of URL can be classified as protocol, sub-domain, second level domain, top level domain, path, file name, as shown in Figure 5.



Fig 5: Structure of URL[12].

Based on [13], there exists five different types of URL obfuscation, obfuscated URLs can be achieved by combining keywords in paths, queries and low-level domain, as listed in the below table 1:

Table 1: URL obfuscation types

Type	Example
Type 1	http://schoo1497.ru/222/www.paypal.com/29370274276105805
Type 2	http://quadrodeofertas.com.br/wwwl.paypal-com/encrypted/ss1218
Type 3	http://cgi-3.paypal-secure.de/info2/verikerdit.html
Type 4	http://69.72.130.98/javaseva/https://paypal.com/uk/onepagepaypal.htm

Phishing attack techniques are continuously changing and improving to escape conventional and standard detection approaches; often, attackers take advantage of advanced artificial intelligence tools. Thus, this points out the need for advanced and intelligent countermeasures to deter phishing attacks [14]. The integration of AI has made the real-time generation of URLs easier and more difficult for instant detection; these real-time generated URLs redirect users to malicious websites where users' data can be easily stolen [15]. While attackers increasingly utilize newly developed and advanced artificial intelligence techniques in launching malicious attacks, AI can also be leveraged in implementing sophisticated defensive mechanisms against phishing attack.

3. Taxonomy of Phishing URL Detection Techniques

Phishing attacks have gained a considerable interest from researchers who are invested in the cybersecurity field, since it is considered a trending attack due to the damage and negative impact it can cause to companies, organizations, and individuals; the damage varies from financial losses to sensitive information theft. Many researchers have attempted to work on phishing detection techniques to protect users from malicious URLs; they usually utilize AI techniques, including Machine Learning (ML), Deep Learning (DL), or employing both of these techniques, which is known as a Hybrid approach. Thus, several researches and studies have been conducted in order to mitigate and control phishing attacks by leveraging various approaches; each research is based on the researcher's perspective and point of view for the sake of discovering the optimal method that could detect phishing attacks instantly before it compromises users' sensitive data. Phishing detection techniques can be categorized into various approaches; the categorization relies on how the URL will be identified. In this research paper, several approaches and methods will be reviewed critically, as well as addressing each research's strength point and limitation. The below Figure 6 shows the taxonomy of phishing URL detection techniques that will be reviewed in this paper.

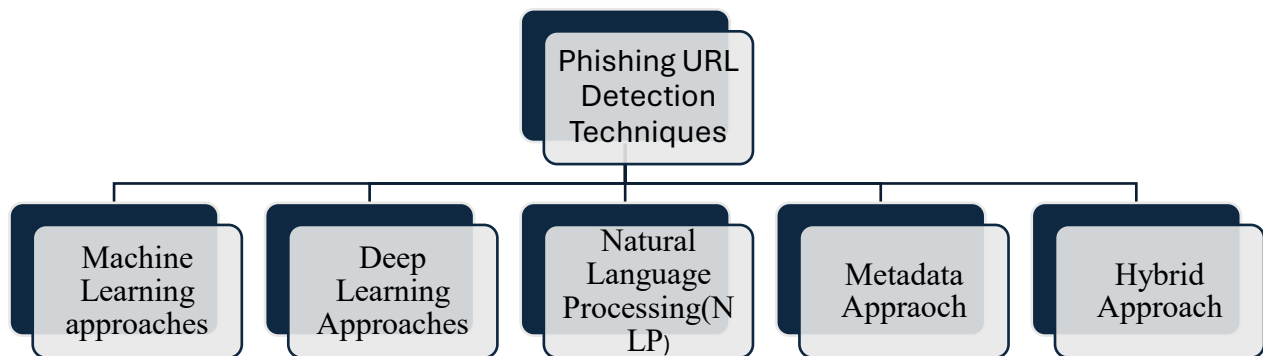


Fig 6: Taxonomy of Phishing URL Detection Techniques.

3.1 Machine Learning-Based URL Detection

Machine Learning techniques have arisen as a promising solution in the realm of detecting phishing URLs because of their capability to learn from previous data. Therefore, the more these models are trained, the more accurate results they will achieve. Once the algorithm is trained, it will proceed to work automatically; thus, ML is leveraged in detecting malicious URLs. Unlike traditional approaches, such as blacklist or heuristic methods, which rely on predefined rules and/or databases of known malicious URLs, these techniques are not well-suited for detecting newly developed attacks, particularly those that integrate AI and evolve beyond known patterns. Therefore, a considerable number of studies have been conducted in the field of Machine Learning. A portion of these studies will be reviewed in this research paper, while addressing each study's strengths and weaknesses.

Rehman et al. (2025) [16] analyzed the use of a real-time detection framework to detect malicious URLs by utilizing multiple Machine Learning (ML) classifiers including Random Forest, Support Vector Machine, and Decision Tree. The results demonstrated that Random Forest achieved the best performance among other models. The accuracy of Random Forest estimated at 99.9% in detecting phishing URLs. However, although this approach achieved a relatively good accuracy rate, the study demonstrated a shortcoming in terms of scalability and adaptability when it is tested against dynamically changing phishing techniques, which consequently minimize its overall effectiveness in future real-time environments.

Similarly, a study by Ahammad et al. (2022) [17] studied the LightGBM model in oppose to conventional machine learning classifiers. The examination of this study was carried out by incorporating 15 engineered features collected from URLs representation, encompassing URL structure-based attributes as well as domain specific features. The final research outcomes showed that LightGBM achieved lower computational overhead and high predictive

accuracy. Nevertheless, the proposed approach demonstrated a significance reliance on external metadata sources such as, WHOIS information, which in turn may introduce latency and restrict deployment in real-time environments.

An analysis to examine the use of Hybrid Machine Learning Frameworks in the classification of malicious URLs was conducted by Karim et al. (2023) [18]. The presented model integrated multiple different classifiers, mixing Random Forest, Extra Trees, and XGBoost. In this study an extensive dataset was leveraged comprising of roughly 455,000 samples, these numerous datasets facilitate a comprehensive evaluation as well as reliable training. The proposed approach is designed to exploits the advantages of individual models while overcoming their limitations via ensemble learning strategies. The empirical results showed a classification accuracy of 98.58%, with the emphasize of minimizing the false positive rate, where benign URLs mistakenly classified as malicious.

Gupta et al. (2021) [19]. Employed Machine Learning technique for Phishing URL detection using lexical-based analysis. The model extracts structural characteristics from a given URL, including URL's length, presence of special characteristics, number of sub-domains and suspicious keywords. The main objective of this study is to address computational overhead by proposing a lightweight and real-time detection framework with reduced computational overhead. The Random Forest classifier achieved an accuracy rate of 99.57%, with response time of 22.5 milliseconds. The dataset used in this study is ISCXURL-2016, supplemented with data sourced form Alexa Top 1 million (legitimate URLs), PhishTank, OpenPhish, and a private dataset from EasyDMARC (phishing URLs). However, the exclusive reliance on lexical features increases the model's limitation in the identification of semantic patterns, which is crucial for the detection of sophisticated phishing attempts.

3.2 Deep Learning-Based URL Detection

Deep Learning is a subset of Machine Learning that is based on the imitation of human brain in processing data in order to make the right decision. It helps to bridge the gap associated with the constant need in traditional machine learning for human intervention to define which features are considered to be suspicious. Deep Learning often utilizes an algorithm called Neural Network. This model is based on interconnected nodes, commonly known as (neurons) which performs data processing and identify patterns. In the contemporary current AI age, Deep Learning techniques have become crucial in detecting sophisticated threats that might remain undetected by traditional defense mechanisms that depend on continuous manual feature engineering.

Aljofey et al. (2020) [20] studied the utilization of a lightweight deep learning model for URL phishing detection based on analyzing sequential patterns captured from URLs. The Character-Level Convolutional Neural Network (CNN) model was utilized in this study, which operates by transforming raw URLs into a matrix representation prior to their processing through convolutional layers. The proposed approach achieved an accuracy of 95.02%, confirming the effective performance of this approach in the URLs phishing detection. The primary advantage of this method is the independence from reliance on external services (e.g. WHOIS, DNS), supporting higher processing speed and more effective real-time detection, achieving a speed of 0.47 milliseconds in processing per URL. The model was evaluated using a dataset consisting of 318,642 instances sourced from Alexa, Open Phish, Spamhaus, Techhelplist, SANS ISC, and PhishTank. Although the proposed model achieved relatively good performance, it may

suffer from the fact that the CNN model requires a long training time, and it may introduce a high chance of false positive rate.

Aldakheel et al. (2023) [21] the researchers of this study explored the leverage of deep learning methodology can be utilized for distinguishing between malicious URLs and benign ones, by utilizing extracted features from URL structure. The detection framework is based on one-dimensional convolutional neural network (1D CNN), which is well-suited for capturing the sequential characteristics of URL data. Particularly, the model processes character-level strings of the uniform resource locator, enabling the model to learn the sequential patterns of URL that are indicative of malicious intents. The model was trained on a balanced dataset comprising 20,000 URLs collected from PhishTank and University of New Brunswick. The model's accuracy achieved 98.77% indicating the effectiveness of the proposed approach. However, the use of relatively small and balanced dataset may limit the model's ability in generalization. Moreover, utilizing character-level features might limit the capability of the proposed model to capture higher-level semantic information. In addition, deep learning models of this type may introduce real-time latency, which can affect their suitability for real-time environments.

Ghalechyan et al. (2024) [22] focuses on resolving the limitations associated with conventional phishing detection approaches that is dependent on third-party services by presenting a framework that is based on deep neural network. The study employed both deterministic and probabilistic neural network (PNN) models for URL classification. In this approach, features are extracted directly from a raw URL by incorporating transfer learning and fine-tuning on a BERT-based architecture. The deterministic model achieved an accuracy of 97%, Probabilistic Model achieved the same level of accuracy of 97%. The model was evaluated using a balanced dataset of 440,000 instances (220,000 phishing and 220,000 non phishing). However, regardless the robust performance achieved by this approach, the dependence on advanced deep learning architecture such as BERT may result in significant computational burden, which can render the model to be less scalable and less suitable for the deployment in real-time environments.

Barik et al (2025) [23] proposed a hybrid detection model that integrates both and Convolutional Neural Network (CNN), Enhanced Grid Search Optimization (EGSO). The proposed framework is designed to minimize errors in the classification while reducing the reliance on manual feature engineering. The study tested 27,534 samples that were collected from multiple repositories including, Kaggle, Alexa, and Phishstorm, to ensure that the model is being trained on newly updated dataset. The experimental results demonstrated accuracy of 99.44%. However, despite the reported accuracy, the incorporation of optimization techniques with deep learning model may increase computational complexity, significantly reducing the effectiveness of the model in real-time deployment. The usage of moderate dataset size may introduce issues regarding the model's capability when applied to more large-scale dataset.

Table 2: Comparison Among Various Studies and their Corresponding Approaches.

Technique	Ref.	year	Dataset	Strength point	Limitation
Machine Learning	Rehman et al.	2025	UCI	Using multiple ML models (RF, OSVM, DT), high accuracy	Not optimal for real time & limited scalability
	Ahammad et al	2022	PhishTank, University of New Brunswick (UNB)	Superior Computational Efficiency Comprehensive Multidimensional Feature Engineering	Highly dependent on WHOIS database, lacks semantic/NLP analysis
	Karim et al.	2023	sourced from PhishTank and University of New Brunswick (UNB)	The hybrid approach reduced the false positive rate at significant amount	High computational complexity
	Gupta et al.	2021	ISCXURL-2016	High response time and speed	Used only lexical analysis and ignored semantic analysis.
Deep Learning	Ghalechyan et al	2024	Alexa Top 1 million (legitimate URLs), PhishTank, OpenPhish, and a private dataset from EasyDMARC (phishing URLs).	The utilization of Novel Probabilistic Framework,	Labeling and Efficiency Constraints
	Aljofey et al	2020	Alexa, Open Phish, Spamhaus, Techhelpist, SANS ISC, and PhishTank.	Low-latency, zero dependency	CNN requires long training time, high chance of false positive
	Aldakheel et al	2023	PhishTank, University of New Brunswick (UNB)	High accuracy, and high efficiency	Dataset bias, real-time latency
	Barik et al	2025	Kaggle, Phishstorm and Alexa	Automated feature extraction, high accuracy rate	Content type restriction, lack of real-time feedback

3.3 Natural Language Processing-Based Detection

Natural Language Processing (NLP) is an advanced subfield of Artificial Intelligence that enables the computer to processes, interpret, and analyze human language, in order to permit machines to extract meanings, as well as identifying certain patterns for the purpose of making informed decisions based on textual input. This field is based on computer science, linguistics and machine learning in order to deal with textual data and convert it to its corresponding computational representation so it can be processed by the machine models. The main core purpose of NLP is to bridge the huge gap between human communication, that can be ambiguous and complex and the machine's

ability of understand, which requires well defined and structured data. To achieve this, multiple techniques are used to transform texts to symbolic and numerical features that could be utilized in the analyzing and classification processes.

E. S. Aung and H. Yamana (2022) [24] proposed a URL tokenization approach, referred to as URL-tokenizer, in order to address the limitation associated with current URL tokenization methods. This approach is based on the employment of both WordSegment and BERT tokenizers to improve the representation of URL text. The model was tested using dataset aggregated from multiple sources, where Phishing URLs were collected from PhishTank, while non-Phishing URLs gathered from IP2Location. Two datasets were considered, the balanced datasets were (10,000 legitimate and 10,000 malicious) and the imbalanced dataset consisted from (5,000 phishing and 50,000 benign). The result demonstrated an improvement in the reduction of unknown tokens, which contributed to improved the overall detection performance, the model's accuracy estimated 95.7% on balanced dataset and 97.7% on imbalanced dataset. However, while this approach enhances tokenization and model performance, it remains dependent on tokenization strategies, which may limit the model's adaptability to newly emerging phishing attempts. In addition, similar to other NLP-based techniques, the proposed technique is strongly focused on text-based representations and may overlook URL structural features, in turn this can significantly reduce the overall framework effectiveness against more sophisticated and evolving phishing attempts.

Kalla and Kuraku (2023) [25] performed a study to examine the use of Machine Learning and Natural Language Processing (NLP) techniques for identifying phishing URLs. The research evaluated seven distinct classifiers on large datasets exceeding 500,000 instances, including 150,000 malicious and 360,000 legitimate URLs. The study findings proved that Support Vector Classifiers (LSVC) and Naive Bayes (NB) achieved high accuracy and precision in the results of detecting lexical pattern within URL, such as frequent occurrence of keywords like "PayPal" or "login". However, while these models showed strong capability in extracting lexical features. Their reliance keyword patterns may limit the model ability the detection of advanced phishing schemes.

Hussan, P. H., & Mangj, S. M. (2025) [26] proposed an approach known as Teacher-Student Learning in which a large RoBERTa model functions as the teacher while lightweight DistilRoBERTa act as the student that acquire necessary knowledge to detect phishing. The main purpose of this framework is to design model with high accuracy and maintains low computational overhead. The model was assessed by using a large dataset sourced from Kaggle repository, the dataset consisted of 500,000 samples, achieving an accuracy of 94.22%. However, despite the efficiency of the proposed approach, this technique is highly reliant on textual URL representations, which in turn can cause limitation in its ability to capture structural and metadata attributes required for robust phishing detection.

In research by Srilachai et al (2025) [27] resilient methodology was formulated for the real-time detection of malicious URLs by integrating, Natural Language Processing (NLP), Deep Learning (DL), and Machine Learning (ML). This study used large dataset consisting of approximately 549,000 URLs, was collected from Kaggle and preprocessed using several techniques including, label encoding, character-level encoding for deep learning models, and TF-IDF vectorization for machine learning models. The result of this study proved that (LSTM) achieved the highest accuracy of 98%. The main motivation of the proposed approach is addressing the dynamic nature of phishing attacks, particularly threats optimized by leveraging various artificial intelligence techniques. In this research five

algorithms were utilized, encompassing Long Short-Term Memory (LSTM), Gated Recurrent Unit (GRU), Random Forest, Support Vector Machine (SVM), and Logistic Regression (LR). However, while the combination of multiple techniques enhances performance of detection process, it may introduce complexity and preprocessing cost, which can lead to a negative impact in real-time performance. In addition, alongside other hybrid methodologies, the integration of diverse feature representations can lead to feature redundancy and demands meticulous feature selection techniques in order to mitigate poor generalization, which is also known as overfitting.

3.4 Metadata-Based URL Detection

Metadata refers to data used to provide information about other data, without the need to capture webpage content, metadata can provide an insight about the structural and descriptive characteristics of URLs. In the context of phishing detection, metadata is comprised of measurable features that can be obtained from URLs, facilitating the process of detecting patterns associated with phishing links. These features deliver valuable insights into the structural composition of Uniform Resource Locator (URL). Metadata attributes can be categorized into several types, including structural, statistical and domain-based features. Figure 7 illustrates the various types of metadata.

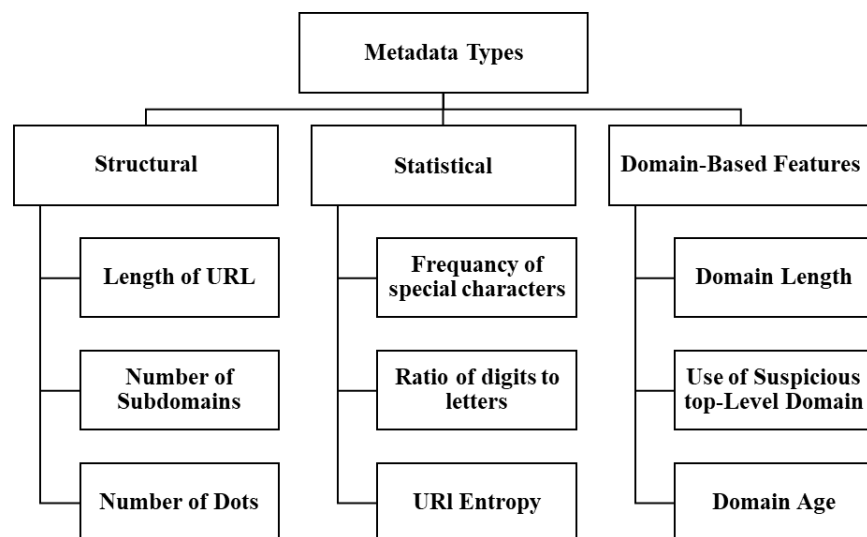


FIG 7: Metadata Types.

These attributes obtained by utilizing metadata, provides a deep insight to a given URL, where anomalies can be effectively detected by utilizing these distinguished features. Metadata plays a critical role in the design of lightweight detection framework, as it can achieve high accuracy rate while reducing the overall computational overhead.

Swarup Panda (2024) [28] proposed a detection framework developed to address the limitations of traditional approaches by designing a model that detect Phishing URLs instantly. The proposed approach is based on extracting metadata features from a certain URLs that disclose potential phishing threats characteristics, including URLs length, number of hyphens, the presence of IP address, and user interaction patterns. In addition, Reinforcement Learning (RL), specifically a contextual multi-armed bandit was utilized to analyze hybrid features. The dataset used in this study was collected from multiple repositories, where Phishing URLs were retrieved from OpenPhish and PhishTank,

while Alexa repository was used to collect legitimate URLs of highly ranked websites that visited by large number of users. The study findings showed a high testing accuracy of 96.85% on 150,000 samples. However, although the model achieved strong performance, the reliance on metadata features may limit the model ability to capture semantic patterns in URLs, precisely in more advanced attacks.

Remya et al (2025) [29] This research study was conducted to integrate three approaches to detect Phishing attack by using three perspectives linguistic content, structural relationships, and metadata. BERT (Bidirectional Encoder Representations from Transformers) model is used to analyze linguistic by analyzing texts within the URL to discover deceptive keywords (e.g., "login," and "secure"). In addition, GNN (Graph Neural Network) is employed to analyze structural features, this model operates on URL components (domain, sub-domain) that are represented as nodes in graph to detect anomalies; for instance, complex sub-domain might refer to a phishing mimicry attempt. The detection framework also utilized LightGBM for metadata analysis, processing structured features such as, domain age, WHOIS registration information, and SSL certificate. The study used a dataset from Kaggle consisted of 549,346, where 40% of the total dataset were Phishing URLs and 60% legitimate. The hybrid model has achieved an accuracy of 97.3. however, despite the promising performance of the proposed framework, the combination linguistic, structural, and metadata features increases the complexity of the detection model and computational overhead.

Despite the significant and substantial advancement in the designing of high-end phishing detection approaches, yet existing approaches remain insufficient in detecting new sophisticated threats, particularly in the last decade with rapid development of diverse Artificial Intelligence techniques. Most studies rely on single-feature, such as, lexical, semantic or metadata-based attributes, which limits the ability to capture the full complexity of phishing URL. Machine learning methods are often lightweight but rely on manual crafted features, restricting its effectiveness against obfuscated type of attacks. In contrast, deep learning and NLP-based techniques tend to achieve higher accuracy when compared to machine learning but introduce computational overhead and increased latency, making the real-time deployment difficult to achieve. Metadata approaches offer efficiency but lack semantic depth and rely on external sources. Although several studies attempted to adopt hybrid approaches, they frequently encounter increased complexity, feature redundancy, and restricted generalization ability. Therefore, there is a critical necessity for light-weight and real-time detection that seamlessly incorporates NLP-based semantic features with metadata-driven characteristics utilizing only URL-level information, maintaining strong accuracy and real-world deployment.

Table 3: Comparison Among Various Studies in the context of the integration of NLP, and Metadata.

Technique	Ref.	year	Dataset	Strength point	Limitation

NLP	E. S. Aung and H. Yamana	2022	PhishTank, IP2Location	This model reduces unknown or meaningless tokens.	The approach depends on an English dictionary for validation only.
	Kalla and Kuraku	2023	Kaggle	High Classification Performance, Large-scale dataset.	Oversimplification of Feature Relationships.
	Hussan, P. H., & Mangj, S. M	2025	Kaggle	High accuracy with low computational cost.	limited feature scope due to the dependence on only URL text.
	Srilachai et al	2025	Kaggle	Robust framework, High Precision.	Difficult for real-time deployment. The proposed framework requires more time for preprocessing.
Metadata	Swarup Panda	2024	Phishtank, OpenPhish and Alexa	Dynamic Adaptation, Hybrid Features integration.	Small dataset, dependency of third part for features.
	Remya et al	2025	Kaggle	High accuracy, low false positive rate.	Computational Intensity, Complexity of Integration.

Conclusion

This paper presents a thorough and detailed overview of various URL-based phishing detection strategies, these approaches include Machine Learning, Deep Learning, NLP, and metadata techniques. The findings revealed a large number of these techniques depend strongly on single feature-set, which leads to a limitation in their capability to address the complexity and the sophistication associated with modern phishing attempts. While some approaches yield a relatively high accuracy rate, they often experience high computational overhead, which can weaken their feasibility for real-time deployment. In order to mitigate these limitations, this study highlights the critical importance of the incorporation of complementary feature set. According to this crucial demand, a hybrid framework that integrates NLP-based semantic features alongside metadata attributes is proposed. By utilizing URL-level data, this approach can support real-time phishing detection while persevering high classification accuracy. Overall, achieving a well-balanced performance between accuracy, efficiency, and real-time remains a primary challenge, and the proposed orientation establishes a baseline standard for more resilient and reliable phishing detection techniques.

References

- [1] S. Garera, N. Provos, M. Chew, and A. D. Rubin, "A framework for detection and measurement of phishing attacks," in *WORM'07 - Proceedings of the 2007 ACM Workshop on Recurring Malcode*, 2007, pp. 1–8. doi: 10.1145/1314389.1314391.

- [2] G. Desolda, L. S. Ferro, A. Marrella, T. Catarci, and M. F. Costabile, “Human Factors in Phishing Attacks: A Systematic Literature Review,” Nov. 30, 2022, *Association for Computing Machinery*. doi: 10.1145/3469886.
- [3] A. Abdallah and A. Sabiel, “Awareness and Preparedness Against AI-Driven Phishing Attacks: A Quantitative Study Among University Students.” [Online]. Available: <https://orcid.org/0009-0000-0172-093X>
- [4] “apwg_trends_report_q4_2025”.
- [5] B. B. Salaudeen, S. Misra, R. Damaševičius, and R. Maskeliūnas, “Identifying phishing attacks in communication networks using URL consistency features,” 2020.
- [6] S. Aung,) Chaw, T. Zan, and H. Yamana, “A Survey of URL-based Phishing Detection.” [Online]. Available: <http://quadrodefertas.com.br/www1>.
- [7] S. Kavva and D. Sumathi, “Staying ahead of phishers: a review of recent advances and emerging methodologies in phishing detection,” *Artif. Intell. Rev.*, vol. 58, no. 2, Feb. 2025, doi: 10.1007/s10462-024-11055-z.
- [8] “Spectrum of Engineering Sciences ISSN (e) 3007-3138 (p) 3007-312X”, doi: 10.5281/zenodo.18030987.
- [9] L. Chen and L. Meng, “Metadata driven malicious URL detection using RoBERTa large and multi source network threat intelligence,” *Sci. Rep.*, vol. 16, no. 1, Dec. 2026, doi: 10.1038/s41598-025-34790-x.
- [10] S. Kailas and R. Roopalakshmi, “‘Think Before You Click’—Malicious URL Detection in Cybersecurity: A Systematic Review and Research Roadmap,” 2025, *Institute of Electrical and Electronics Engineers Inc.* doi: 10.1109/ACCESS.2025.3601387.
- [11] H. Ghalechyan, E. Israyelyan, A. Arakelyan, G. Hovhannisyan, and A. Davtyan, “Phishing URL detection with neural networks: an empirical study,” *Sci. Rep.*, vol. 14, no. 1, Dec. 2024, doi: 10.1038/s41598-024-74725-6.
- [12] M. Nanda, M. Saraswat, and P. K. Sharma, “Enhancing cybersecurity: A review and comparative analysis of convolutional neural network approaches for detecting URL-based phishing attacks,” *e-Prime - Advances in Electrical Engineering, Electronics and Energy*, vol. 8, Jun. 2024, doi: 10.1016/j.prime.2024.100533.
- [13] S. Marchal, J. François, R. State, and T. Engel, “PhishStorm: Detecting Phishing with Streaming Analytics.” [Online]. Available: <http://www.quadrodefertas.com.br/www1.paypal-com/encrypted/ssl218>
- [14] M. Saeed Liaqat, G. Mumtaz, N. Rasheed, and Z. Mubeen, “Exploring Phishing Attacks in the AI Age: A Comprehensive Literature Review”, doi: 10.56979/702/2024.
- [15] M. Sameen, K. Han, and S. O. Hwang, “PhishHaven - An Efficient Real-Time AI Phishing URLs Detection System,” *IEEE Access*, vol. 8, pp. 83425–83443, 2020, doi: 10.1109/ACCESS.2020.2991403.
- [16] A. U. Rehman, I. Imtiaz, S. Javaid, and M. Muslih, “Real-Time Phishing URL Detection Using Machine Learning †,” *Engineering Proceedings*, vol. 107, no. 1, 2025, doi: 10.3390/engproc2025107108.
- [17] S. H. Ahammad *et al.*, “Phishing URL detection using machine learning methods,” *Advances in Engineering Software*, vol. 173, Nov. 2022, doi: 10.1016/j.advengsoft.2022.103288.
- [18] A. Karim, M. Shahroz, K. Mustofa, S. B. Belhaouari, and S. R. K. Joga, “Phishing Detection System Through Hybrid Machine Learning Based on URL,” *IEEE Access*, vol. 11, pp. 36805–36822, 2023, doi: 10.1109/ACCESS.2023.3252366.
- [19] B. B. Gupta, K. Yadav, I. Razzak, K. Psannis, A. Castiglione, and X. Chang, “A novel approach for phishing URLs detection using lexical based machine learning in a real-time environment,” *Comput. Commun.*, vol. 175, pp. 47–57, Jul. 2021, doi: 10.1016/j.comcom.2021.04.023.
- [20] A. Aljofey, Q. Jiang, Q. Qu, M. Huang, and J. P. Niyigena, “An effective phishing detection model based on character level convolutional neural network from URL,” *Electronics (Switzerland)*, vol. 9, no. 9, pp. 1–24, Sep. 2020, doi: 10.3390/electronics9091514.
- [21] E. A. Aldakheel, M. Zakariah, G. A. Gashgari, F. A. Almarshad, and A. I. A. Alzahrani, “A Deep Learning-Based Innovative Technique for Phishing Detection in Modern Security with Uniform Resource Locators,” *Sensors*, vol. 23, no. 9, May 2023, doi: 10.3390/s23094403.

-
- [22] H. Ghalechyan, E. Israyelyan, A. Arakelyan, G. Hovhannisyan, and A. Davtyan, “Phishing URL detection with neural networks: an empirical study,” *Sci. Rep.*, vol. 14, no. 1, Dec. 2024, doi: 10.1038/s41598-024-74725-6.
- [23] K. Barik, S. Misra, and R. Mohan, “Web-based phishing URL detection model using deep learning optimization techniques,” *Int. J. Data Sci. Anal.*, vol. 20, no. 5, pp. 4449–4471, Oct. 2025, doi: 10.1007/s41060-025-00728-9.
- [24] E. S. Aung and H. Yamana, “Segmentation-based Phishing URL Detection,” in *ACM International Conference Proceeding Series*, Association for Computing Machinery, Dec. 2021, pp. 550–556. doi: 10.1145/3486622.3493983.
- [25] D. Kalla and S. Kuraku, “Phishing Website URL’s Detection Using NLP and Machine Learning Techniques,” *Journal on Artificial Intelligence*, vol. 5, no. 0, pp. 145–162, 2023, doi: 10.32604/jai.2023.043366.
- [26] P. H. Hussan and S. M. Mangi, “BERTPHIURL: A Teacher-Student Learning Approach Using DistilRoBERTa and RoBERTa for Detecting Phishing Cyber URLs,” *Journal of Future Artificial Intelligence and Technologies*, vol. 1, no. 4, pp. 417–428, Feb. 2025, doi: 10.62411/faith.3048-3719-71.
- [27] N. Srilachai, P. Viboonsang, and S. Kosolsombat, “Phishing URL Detection Using Machine Learning and Deep Learning in NLP,” in *International Conference on Cybernetics and Innovations, ICCI 2025*, Institute of Electrical and Electronics Engineers Inc., 2025. doi: 10.1109/ICCI64209.2025.10987478.
- [28] “Real-Time Phishing URL Detection Using”.
- [29] S. Remya, M. J. Pillai, B. S. Aparna, S. Rama Subbareddy, and Y. Y. Cho, “BGL-PhishNet: Phishing Website Detection Using Hybrid Model-BERT, GNN, and LightGBM,” *IEEE Access*, vol. 13, pp. 47552–47569, 2025, doi: 10.1109/ACCESS.2025.3551542.