# Introduction

In modern times, the internet has become an indispensable mode of communication, providing individuals with the opportunity to establish virtual environments that complement their offline business activities or create entirely online businesses. Although the internet offers numerous benefits, it is not without its drawbacks. Online fraud, through phishing attacks, is a significant risk associated with the online environment. Phishing is a sort of online attack in which an invader lures unsuspecting individuals to visit fake or malicious websites, aiming to get sensitive information like usernames, authentication codes/passwords, and bank card details (Solanki & Dogiwal, 2015). Regrettably, phishing attempts are now pervasive and account for more than half of all internet fraud instances involving regular people (Orman, 2013). As a result, Google now sends approximately ten million notifications per day to people who visit websites notorious for their phishing scams. Also, a little over 10,000 new sites are added to the list of risky websites every day. This underscores the need for online users to exercise caution and vigilance while browsing the internet.

Phishing crimes have seen a surge in recent times, resulting in billions of dollars in losses. Phishers deploy various methods, including fake sites, search engines, advertisements, spam emails, direct messaging, or phone calls, to launch attacks (G. Liu et al., 2010). However, URL obfuscation is the most widely used technique by phishers (Rao & Ali, 2015). This method is used to lure unsuspecting users by misleading them into visiting fake websites through URLs or genuine websites that the victim is familiar with (Cova et al., n.d.). In these situations, visitors enter personal data onto what looks to be a trustworthy website, including passwords and credit card information. Webmail and Software-as-a-Service (SaaS) websites are the most often targeted sites by phishing assaults (*APWG | Phishing Activity Trends Reports*, n.d.). Millions of online users are sent the phishing webpage link via emails and other modes of communication after phishers develop websites that almost exactly resemble trustworthy ones. Emails, instant chats, or phone calls are frequently used to start these kinds of cyber-attacks (Aljofey et al., 2020). Phishing attacks can be used to transmit other software, such as ransomware, exploit security flaws, or even make money (Dhamija et al., n.d.). They are not merely intended to steal the victim's identity.

Phishing assaults significantly increased in the 3rd quarter of 2022 (*APWG | Phishing Activity Trends Reports*, n.d.) as opposed to earlier during the year. The Anti-Phishing Working Group (APWG) study states that between July and September, 1,270,883 distinct phishing URLs were found (*APWG | Phishing Activity Trends Reports*, n.d.). Also, the average cost of a wire transfer Business E-mail Compromise (BEC) assault was $48,000, down from $80,000 in the preceding quarter and $54,000 in the initial quarter. it was found that financial firms were the target of most phishing scams in the final quarter of 2022. Attacks against Software-as-a-Service (SaaS) and webmail sites decreased, while attacks against E-commerce sites increased. However, attacks against media companies declined slightly from 12.6% to 11.8% (*APWG | Phishing Activity Trends Reports*, n.d.). The prevailing pandemic situation has also seen a significant increase in phishing attacks. Cybercriminals are exploiting the global focus on Covid-19, and according to the World Health Organization (WHO), several hackers and cyber scammers are launching fraudulent emails and chat messages to people (*Cyber Security*, n.d.). These attacks come in various forms, including fake job offers, fabricated messages from health organizations, Covid-19 vaccine-themed phishing, and brand impersonation.

Phishing is a mode of cyber-attack that is difficult to detect and prevent because phishers are constantly coming up with new methods to evade present anti-phishing techniques. Even well-educated and competent users can fall victim to these attacks. To combat phishing, software-based detection methods are preferred. Some common methods for detecting phishing include blacklists/whitelists (Jain & Gupta, 2016), visual similarity (Haruta et al., 2017), rules (Cook et al., 2008), natural language processing (Sahingoz et al., 2019), and machine learning (Jain & Gupta, 2019; Li et al., 2019). Yet, machine learning techniques rely on heuristic criteria including URL, webpage material, web traffic, search engine, WHOIS record, and Page Rank, whereas blacklists/whitelists can fail to recognize new or previously unreported phishing sites. Although these heuristic elements can increase detection effectiveness, they are not always present on phishing websites and occasionally appear on legitimate websites as well, which can cause classification errors.

Some of the heuristic elements that are used to detect phishing are hard to obtain and rely on outside services. Page rank, search engine indexing, and WHOIS are a few examples of services that might not be sufficient to correctly identify malicious URLs that are located on hijacked servers. These websites may be mistakenly classed as benign since they show in search results, or even as phishing websites because of their young domains. New websites are compared to pre-stored signatures that contain screenshots, font styles, photos, screen layouts, logos, and other visual features using visual similarity-based heuristic algorithms. For brand-new phishing websites, these strategies have a significant false-negative rate, nevertheless. There are also restrictions on URL-based approaches (Rao et al., 2020; Xiang et al., 2011; Zhang et al., 2017), which extract details such as the quantity of dots, the existence of special symbols, the length of the URL, and company names. These features are often manually extracted, which requires time and extra maintenance costs. As a result, a hybrid approach that combines multiple methods is recommended for phishing detection, rather than relying on a single technique.

Phishing attacks pose a significant threat as they can be hard to identify. Cybercriminals use social engineering tactics to create convincing phishing emails or messages that deceive victims. To make matters worse, the fake websites used in these attacks are often designed to look like legitimate sites, making it challenging for victims to differentiate between them. Therefore, there is a pressing need for effective phishing detection methods that can help individuals and organizations safeguard against such attacks. One approach to detecting phishing is URL-based detection techniques, which involve examining website URLs to determine their authenticity.

URL-based phishing detection techniques are designed to scrutinize website URLs to differentiate between legitimate and fake sites. These techniques are particularly useful in identifying phishing websites that are created to resemble legitimate sites. The most widely used method for URL-based phishing recognition is the analysis of the domain name. Phishing websites often use domain names that are like legitimate websites but with minor changes. For instance, a phishing website may replace one letter in the domain name with another similar-looking letter. The primary objective of this study is to give an extensive overview of phishing attacks, highlight the importance of detecting such attacks, and explore different URL-based detection techniques that can be utilized to identify phishing websites.

## AIMS AND OBJECTIVES OF THE PROJECT

Phishing refers to a fraudulent technique used to trick individuals into revealing their sensitive information, such as login credentials, credit card details, and other personal information, to fake websites that appear authentic. The attackers then use this information to gain unauthorized access to important accounts, leading to identity theft and financial harm. Among the various types of cyber-attacks, phishing is the most frequently used method. Therefore, it is crucial for computer and internet users to safeguard their personal information, ensure its security, and minimize the risk of falling victim to fraudulent activities while browsing different websites.

Phishing has been recognized as a persistent issue since the early days of the internet, and it is one of the most challenging issues to address and control. The goal of this research is to examine the dangers of phishing and to create a detection system to mitigate the issue. In related studies, deep learning (a branch of machine learning) and natural language processing have been utilized and have exhibited significant progress in detecting phishing. Despite this, several cybersecurity experts have advocated for hybrid or mixed methods rather than a single strategy. To this end, this research proposes an automated phishing responder (APR) that employs a mixed machine learning approach that combines natural language processing to identify phishing attempts. Furthermore, the effectiveness of the proposed approach will be evaluated by comparing it to other techniques.

Thus, the research objective of this project encompasses, but is not restricted to, the following:

* To build a cutting-edge method for identifying dangerous URLs and warning users.
* To use machine learning (ML) techniques in the suggested strategy to evaluate real-time URLs and generate useful results.
* To put into practice the idea of various ML techniques that can manage massive amounts of data.
* Creating a phishing detection system and researching earlier studies on the suggested subject to see how they might be improved.

# Literature Review

## Overview of Phishing

Phishing is a kind of cyber-attack where criminals use fraudulent communication methods, such as emails, text messages, or social media, to trick individuals into revealing sensitive data, including login details, bank card details, and personal identification numbers. Phishing attacks are often disguised as legitimate communications from trusted sources and can take many forms, including spear phishing, clone phishing, and whaling. Phishing attacks cause a serious threat to individuals and organizations, as they can result in the loss of sensitive information, financial losses, and reputational damage. However, there are steps that can be taken to protect against phishing attacks, including educating oneself and employees about the risks of phishing, using strong passwords and two-factor authentication, and using anti-phishing software.

### Phishing Techniques

Phishing techniques involve the use of fraudulent communication methods, such as email, text message, or social media, to trick individuals into revealing sensitive information or taking an action that benefits the attacker. Some common phishing techniques include spear phishing, clone phishing, and whaling. Spear phishing involves sending targeted, personalized messages to individuals or small groups in order to trick them into disclosing sensitive information. Clone phishing involves creating a fake website that appears like a legitimate one and using it to steal login credentials. Whaling targets high-level executives in organizations in order to gain access to sensitive company information. Phishing methods are growing increasingly advanced and might be challenging to spot. To protect against phishing attacks, individuals and organizations should educate themselves and their employees about the risks of phishing, use strong passwords and two-factor authentication, and use anti-phishing software. Individuals and organizations can lessen their chance of becoming phished by following these procedures.

### Phishing Websites

Phishing websites are malicious websites that impersonate trustworthy websites to deceive users into disclosing vital information. These websites can be created to resemble websites of banks, e-commerce sites, social media platforms, or other trusted sources. Phishing websites are often part of a larger phishing scheme that involves sending fraudulent emails or messages that include a link to a fake website. Once a victim visits the website and enters their login credentials or other sensitive information, the attackers can use that information for their own gain. Phishing websites can be difficult to detect, as they often use similar branding, logos, and design to legitimate websites. However, individuals and organizations can take steps to protect themselves, including using anti-phishing software, avoiding clicking on links in suspicious emails or messages, and checking the URL of a website before entering any sensitive information.

### The URL Components

A URL (Uniform Resource Locator) is a string of characters that identifies a resource on the internet. The components of a URL are as follows:

1. Scheme or Protocol: It indicates the protocol used to access the resource. Examples of schemes include HTTP, HTTPS, FTP, and mailto.
2. Host: It is the domain name or IP address of the server where the resource is located.
3. Port: It is an optional component that specifies the port number to which the client should connect on the server.
4. Path: It is the location of the resource on the server's file system.
5. Query String: It is an optional component that specifies parameters to be passed to the server along with the request. The query string is appended to the end of the URL and starts with a question mark (?).
6. Fragment or Anchor: It is an optional component that specifies a particular section within the resource. The fragment is appended to the end of the URL and starts with a hash symbol (#).

For example, in the URL "<https://www.example.com/search?q=url+components#fragment>", the scheme is "https", the host is "[www.example.com](http://www.example.com)", the path is "/search", the query string is "?q=url+components", and the fragment is "fragment".Bottom of Form

## Types of Phishing

Phishing is a type of cybercrime in which an attacker poses as a reliable institution in an effort to fool a user into disclosing sensitive information like login passwords, banking details, or private details. Here are a few prevalent phishing attack types:

1. Email Phishing: This kind of phishing assault is the most prevalent. An attacker sends an email that impersonates a trustworthy source, like a bank, social networking platform, or online merchant. Often, the email contains a link that takes the recipient to a false website that imitates the actual one and asks them to enter personal information like their credit card number or login information.
2. Spear Phishing: Compared to email phishing, this kind of phishing attempt is more individualized and targeted. The attacker researches the victim's personal information from various sources and sends a deceptive email that seems to be from a trusted source such as a colleague, friend, or business associate. The email contains a malicious link or attachment that can infect the victim's computer with malware or steal their sensitive data.
3. Smishing: This is a form of phishing that uses text messages or SMS to lure victims into clicking on malicious links or disclosing sensitive information. The message usually appears to be from a legitimate source such as a bank and asks the victim to provide their account information or personal details.
4. Vishing: Voice messages or phone calls are used in this kind of phishing attempt to deceive victims into divulging their sensitive information. The caller pretends to be a representative from a legitimate company such as a bank or credit card issuer, and asks the victim to provide their personal information such as their account number or password.
5. Malware-based Phishing: In this type of attack, the attacker sends an email or message containing a malicious file or link that, when clicked, installs malware on the victim's computer or device. The malware can then steal sensitive information such as login credentials, credit card details, or other personal data.

It's important to be aware of these types of phishing attacks and to take necessary precautions to protect your sensitive information.

## Current Phishing Detection Techniques

Phishing is a constantly evolving threat, and as a result, there are several practices that are being used to identify and prevent phishing attacks. Some of the current phishing detection techniques are:

1. Content-based Filtering: This technique involves analyzing the content of an email or message to detect phishing attempts. This includes scanning for specific keywords, known phishing patterns, and URLs that are known to be malicious.
2. URL Reputation Filtering: This technique involves analyzing the URL in an email or message to determine its reputation. If the URL is flagged as malicious or suspicious, it can be blocked or redirected to a warning page.
3. Sender Authentication: This technique involves verifying the authenticity of the sender's email address. This can be done using various authentication protocols such as DKIM (DomainKeys Identified Mail), SPF (Sender Policy Framework), and DMARC (Domain-based Message Authentication, Reporting, and Conformance).
4. Machine Learning: This technique involves using machine learning algorithms to analyze large amounts of data to identify patterns and detect phishing attempts. Machine learning can be used to analyze email content, sender information, and other metadata to identify suspicious activity.
5. User Education: One of the most effective ways to prevent phishing attacks is to educate users about the risks and how to identify suspicious emails and messages. This includes providing training on how to identify phishing attempts, how to report suspicious activity, and how to take appropriate action if they receive a suspicious email or message.

It's important to use a combination of these techniques to provide the best possible protection against phishing attacks.

## Limitations of Existing Techniques

While there are several methods for detecting and preventing phishing attempts, they are not perfect and have significant drawbacks. Some of the limitations of existing techniques are:

1. Evolving Tactics: Phishing tactics are constantly evolving, and attackers are finding new ways to bypass existing detection techniques. As a result, some phishing attacks may go undetected.
2. False Positives: False positives can result from some phishing detection algorithms, which mark genuine emails or communications as phishing efforts. Users may experience discomfort as a result, which could lower productivity.
3. User Errors: Even with the best detection techniques, users can still fall victim to phishing attacks due to human error. For example, they may click on a link or download an attachment without realizing the risk.
4. Zero-day Attacks: As zero-day attacks take advantage of previously undiscovered vulnerabilities, current detection methods might not be able to catch them.
5. Encryption: Encryption can make it difficult to detect phishing attempts, as it can hide the content of emails and messages from analysis. Attackers may use encryption to hide their phishing attempts from detection.
6. Mobile Devices: With the rise of mobile devices, phishing attacks are increasingly targeting mobile users. However, existing detection techniques may not be as effective on mobile devices.

It's important to keep these limitations in mind and continue to develop and improve phishing detection techniques to stay ahead of the evolving threat landscape. Additionally, user education and awareness should be a key part of any phishing prevention strategy.

## Review of Related Work

An overview of the suggested phishing recognition techniques in the literature is given in this section. The two main types of phishing techniques are using additional software and raising consumer awareness of the differences between phishing and legitimate websites. List-based identification and machine learning-based identification are further categories for software-based approaches. Yet, because the phishing issue is so complex, there isn't a single way to effectively avoid all threats; instead, a variety of approaches are frequently used to stop specific phishing crimes.

### List-Based Phishing Site Detection

Methods for detecting phishing using lists can either utilize a whitelist or a blacklist. While determining whether a URL is bogus, a blacklist of suspected domain names, URLs, and IP addresses is used. In addition, the whitelist is a collection of trustworthy domain names, site URLs, and IP addresses that can be used to verify a suspected URL. Whitelist-based techniques were used by Jain and Gupta (Jain & Gupta, 2016), Wang et al. (Wang et al., 2008), and Han et al. (Han et al., 2012) to find questionable URLs. On the other hand, blacklist-based practices are frequently employed in openly available anti-phishing toolbars. One illustration is Google Secure Browsing, which maintains a blacklist of URLs and notifies visitors whenever a URL may be fraudulent. Prakash et al.(Prakash et al., 2010) described the Phishnet technique as a way to predict phishing Websites. Phishing Sites are identified from blacklisted URLs using the repository, associated Ip, and firm logo in this approach. Felegyhazi et al.(Felegyhazi et al., n.d.) algorithms matched the domain suffix and name server information with those of blacklisted URLs in order to categorize new suspicious URLs. Between 50% and 80% of the fake names were established after the assault was initiated, according to Sheng et al. (Sheng et al., 2009), and a fake domain was added to the blacklist after a lengthy period of time. Due to the daily establishment of hundreds of malicious websites, the blacklist must be refreshed from its origin on a frequent basis. As a result, phishing offences can be dealt with more effectively using machine learning-based detection algorithms.

### Machine learning-based detection

Several applications, including information security and privacy (Qi et al., 2021), game theory (Y. Liu, Yao, et al., 2022), blockchain systems (Muzammal et al., 2019), healthcare (Y. Liu, Song, et al., 2022), etc., have benefited greatly from the use of data mining algorithms. Many machine-learning-based strategies have been used to investigate the legitimacy of websites due to the latest advancement in phishing recognition methods (Jain & Gupta, 2019; Li et al., 2019; Rao et al., 2020; Sahingoz et al., 2019). The efficiency of these techniques depends on training sets, classification models, and extraction of features. The feature collection comes from a number of sources, such as URLs, web content, third-party services, etc. Unfortunately, certain heuristic feature methods are very time-consuming and difficult to extract, therefore machine learning methods require a lot of computations to retrieve features.

Jain and Gupta (Jain & Gupta, 2018a) presented an anti-phishing strategy that isn't based on any tertiary-party services and derives the characteristics from the URL and source code of the site. Despite applying a small dataset, the proposed technique had good accuracy in identifying phishing websites (1918 legitimate and 2141 phishing web pages). Similar authors (Jain & Gupta, 2019) provide a phishing discovery mechanism which analyses the hyperlinks collected from the web page's HTML to identify phishing assaults. The proposed approach is a client-side and language-neutral solution to identify phishing web pages. However, if an attacker alters all the webpage resource references, the approach may not be able to accurately categorize phishing webpages as it only focuses on the HTML of the website and not on other resources such as CSS, JavaScript, images, etc. To address this limitation, Rao and Pais (Rao & Pais, 2020) proposed a two-phase anti-phishing approach known as BlackPhish. In the first phase, a blacklist of signatures is built based on visual similarities such as file titles, directories, and screenshots, rather than a list of URLs. To find malicious URLs that get past the first phase filtering, heuristic features are retrieved from both the URL and HTML at the second tier. Despite this, there is always two-level filtering applied to authorized websites. Authors of several studies (Jain & Gupta, 2018b) employed search engine-based mechanisms to first level validate the site. Different hyperlinks found in the website's HTML are analyzed in the second stage of validation in order to identify phishing websites. The usage of search engine-based practices can increase the accuracy of identifying genuine websites as legitimate. However, this approach can also lead to misclassifying genuine websites as malware, especially when newly created genuine web pages are not yet visible in the top results of search engines. Search-based methods assume that trustworthy websites will always appear in the top search results, which is not always the case.

A new scamming sites detection technique using embeddings retrieved from pure text and domain-specific content of the HTML source script was proposed by Rao et al (Rao et al., 2022) in a recent study. To validate their model utilizing ensemble and multimodal approaches, they used several word embeddings. The proposed approach, however, depends solely on plain text and language that is exclusive to a given domain, and it may not work if the text is substituted with pictures. By extracting various hyperlink associations from web pages, several academics have attempted to recognize phishing schemes. A phishing websites detection method known as HinPhish was proposed by Guo et al (Guo et al., 2021). The suggested approach entails loading source components and building a heterogeneous information network (HIN) depending on the domain. This network consists of four types of embedded links: external, empty, internal, and relative links, which are connected in three different ways. To analyze the impact of these relationships, the researchers used an authority ranking algorithm to assign a numerical score to each node in the network.

In a study by Sahingoz et al. (Sahingoz et al., 2019), seven different machine-learning models were tested to verify if a URL is a phishing website. The researchers used a technique called the distributed representation of phrases to analyze certain components of the URL. In a separate study, Rao et al (Rao et al., 2020) proposed an anti-phishing method called CatchPhish. This approach involved extracting manually created and Term Frequency-Inverse Document Frequency (TF-IDF) characteristics from URLs. These features were then used to train a classification algorithm using the random forest approach. While working satisfactorily, the aforementioned techniques have the following drawbacks: Due to the fact that URLs frequently contain cryptic and meaningless terms that are not in the training data set, (1) they fail to account for unseen characters, and (2) they neglect to take the web page into consideration. Because of this, some URLs that are familiar to other people yet mimic the official websites might not be recognized based on URL string. Their work, which is primarily concerned with URL properties, is insufficient to detect fraudulent URLs.

Phishing detection has utilized various deep learning techniques such as Convolutional Neural Networks (CNN), and Recurrent Neural Networks (RNN), due to their success in Natural Language Processing (NLP). However, the lengthy training time of deep learning methods has limited their use in phishing detection. Aljofey et al. (Aljofey et al., 2020) presented a content-based- based filtering method that uses a character-level CNN based on URL to address this problem. Using several forms of features, such as TF-IDF symbols, count vectorizer, and manually generated features, the suggested approach was assessed against a number of machine learning and deep learning approaches. Le et al. (Le et al., 2018) proposed URLNet, a technique that employs URLs to identify phishing websites, as an alternative strategy. In order to train and evaluate their model, they use neural networks to extract character- and word-level features from URL sequences. Chatterjee and Namin (Chatterjee & Namin, 2019) reported a deep reinforcement learning (RL) based phishing detection technique. It trains the model using 14 hand-crafted characteristics that were taken from a balanced, annotated dataset of benign and phishing URLs. A fraudulent website detection technique called CNN-MHSA that uses CNN models to extract character characteristics from URLs was introduced by Xiao et al. (Xiao et al., 2020) in a recent study. A multi-head self-attention (MHSA) mechanism is also used in the method to determine the appropriate weights for the CNN-trained features. The Highway Deep Pyramid Neural Network (HDP-CNN), a new deep convolutional network to determine whether a specific URL is safe or malicious, was introduced by Zheng et al (Zheng et al., 2022). Character-level and word-level embeddings were used in this network. Although the technologies previously mentioned have shown outstanding results, because they simply extract attributes from the website's URL, they may misclassify phishing websites located on hacked servers.

Previous studies have extracted features manually, which requires additional effort as the characteristics need to be changed according to the dataset, potentially leading to overfitting in anti-phishing solutions. Taking inspiration from these studies, we propose our approach, which extracts character sequence features from URLs.

## Need for Composite Machine Learning Approach

The need for composite machine learning model approaches in detecting and preventing phishing attacks arises from the limitations of using a single detection technique. As mentioned earlier, no detection technique is foolproof, and attackers are constantly finding new ways to bypass existing techniques. Therefore, using a composite approach that combines multiple techniques can increase the chances of detecting and preventing phishing attacks. A composite machine-learning model approach involves combining multiple machine-learning models to create a more robust and accurate detection system. For example, a composite approach could combine content-based filtering, URL reputation filtering, and machine learning algorithms to analyze email content, sender information, and other metadata to identify suspicious activity.

By combining multiple techniques, a composite approach can overcome the limitations of individual techniques, and increase the accuracy and reliability of the detection system. Additionally, a composite approach can provide a more comprehensive analysis of the data, reducing the number of false positives and false negatives, and improving the overall effectiveness of the system. In summary, a composite machine learning model approach can provide a more comprehensive and accurate phishing detection system, by combining multiple detection techniques to overcome the limitations of individual techniques.

# References

Aljofey, A., Jiang, Q., Qu, Q., Huang, M., & Niyigena, J. P. (2020). An Effective Phishing Detection Model Based on Character Level Convolutional Neural Network from URL. *Electronics 2020, Vol. 9, Page 1514*, *9*(9), 1514. https://doi.org/10.3390/ELECTRONICS9091514

*APWG | Phishing Activity Trends Reports*. (n.d.). Retrieved February 22, 2023, from https://apwg.org/trendsreports/

Chatterjee, M., & Namin, A. S. (2019). Detecting phishing websites through deep reinforcement learning. *Proceedings - International Computer Software and Applications Conference*, *2*, 227–232. https://doi.org/10.1109/COMPSAC.2019.10211

Cook, D. L., Gurbani, V. K., & Daniluk, M. (2008). Phishwish: A stateless phishing filter using minimal rules. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *5143 LNCS*, 182–186. https://doi.org/10.1007/978-3-540-85230-8\_15/COVER

Cova, M., Kruegel, C., & Vigna, G. (n.d.). *There is No Free Phish: An Analysis of “Free” and Live Phishing Kits*.

*Cyber security*. (n.d.). Retrieved February 22, 2023, from https://www.who.int/about/cyber-security

Dhamija, R., Tygar, J. D., & Hearst, M. (n.d.). *Why Phishing Works*. Retrieved February 22, 2023, from www.paypa1.com

Felegyhazi, M., Kreibich, C., & Paxson, V. (n.d.). *On the Potential of Proactive Domain Blacklisting*. Retrieved February 22, 2023, from http://www.uribl.com/gold.shtml

Guo, B., Zhang, Y., Xu, C., Shi, F., Li, Y., & Zhang, M. (2021). HinPhish: An Effective Phishing Detection Approach Based on Heterogeneous Information Networks. *Applied Sciences 2021, Vol. 11, Page 9733*, *11*(20), 9733. https://doi.org/10.3390/APP11209733

Han, W., Cao, Y., Bertino, E., & Yong, J. (2012). Using automated individual white-list to protect web digital identities. *Expert Systems with Applications*, *39*(15), 11861–11869. https://doi.org/10.1016/J.ESWA.2012.02.020

Haruta, S., Asahina, H., & Sasase, I. (2017). Visual Similarity-Based Phishing Detection Scheme Using Image and CSS with Target Website Finder. *2017 IEEE Global Communications Conference, GLOBECOM 2017 - Proceedings*, *2018-January*, 1–6. https://doi.org/10.1109/GLOCOM.2017.8254506

Jain, A. K., & Gupta, B. B. (2016). A novel approach to protect against phishing attacks at client side using auto-updated white-list. *Eurasip Journal on Information Security*, *2016*(1), 1–11. https://doi.org/10.1186/S13635-016-0034-3/FIGURES/5

Jain, A. K., & Gupta, B. B. (2018a). Towards detection of phishing websites on client-side using machine learning based approach. *Telecommunication Systems*, *68*(4), 687–700. https://doi.org/10.1007/S11235-017-0414-0/TABLES/7

Jain, A. K., & Gupta, B. B. (2018b). Two-level authentication approach to protect from phishing attacks in real time. *Journal of Ambient Intelligence and Humanized Computing*, *9*(6), 1783–1796. https://doi.org/10.1007/S12652-017-0616-Z/TABLES/5

Jain, A. K., & Gupta, B. B. (2019). A machine learning based approach for phishing detection using hyperlinks information. *Journal of Ambient Intelligence and Humanized Computing*, *10*(5), 2015–2028. https://doi.org/10.1007/S12652-018-0798-Z/TABLES/6

Le, H., Pham, Q., Sahoo, D., & Hoi, S. C. H. (2018). *URLNet: Learning a URL Representation with Deep Learning for Malicious URL Detection*. https://doi.org/10.48550/arxiv.1802.03162

Li, Y., Yang, Z., Chen, X., Yuan, H., & Liu, W. (2019). A stacking model using URL and HTML features for phishing webpage detection. *Future Generation Computer Systems*, *94*, 27–39. https://doi.org/10.1016/J.FUTURE.2018.11.004

Liu, G., Qiu, B., & Wenyin, L. (2010). Automatic detection of phishing target from phishing webpage. *Proceedings - International Conference on Pattern Recognition*, 4153–4156. https://doi.org/10.1109/ICPR.2010.1010

Liu, Y., Song, Z., Xu, X., Rafique, W., Zhang, X., Shen, J., Khosravi, M. R., & Qi, L. (2022). Bidirectional GRU networks-based next POI category prediction for healthcare. *International Journal of Intelligent Systems*, *37*(7), 4020–4040. https://doi.org/10.1002/INT.22710

Liu, Y., Yao, R., Jia, S., Wang, F., Wang, R., Ma, R., & Qi, L. (2022). A label noise filtering and label missing supplement framework based on game theory. *Digital Communications and Networks*. https://doi.org/10.1016/J.DCAN.2021.12.008

Muzammal, M., Qu, Q., & Nasrulin, B. (2019). Renovating blockchain with distributed databases: An open source system. *Future Generation Computer Systems*, *90*, 105–117. https://doi.org/10.1016/J.FUTURE.2018.07.042

Orman, H. (2013). The compleat story of phish. *IEEE Internet Computing*, *17*(1), 87–91. https://doi.org/10.1109/MIC.2013.16

Prakash, P., Kumar, M., Rao Kompella, R., & Gupta, M. (2010). PhishNet: Predictive blacklisting to detect phishing attacks. *Proceedings - IEEE INFOCOM*. https://doi.org/10.1109/INFCOM.2010.5462216

Qi, L., Hu, C., Zhang, X., Khosravi, M. R., Sharma, S., Pang, S., & Wang, T. (2021). Privacy-Aware Data Fusion and Prediction with Spatial-Temporal Context for Smart City Industrial Environment. *IEEE Transactions on Industrial Informatics*, *17*(6), 4159–4167. https://doi.org/10.1109/TII.2020.3012157

Rao, R. S., & Ali, S. T. (2015). PhishShield: A Desktop Application to Detect Phishing Webpages through Heuristic Approach. *Procedia Computer Science*, *54*, 147–156. https://doi.org/10.1016/J.PROCS.2015.06.017

Rao, R. S., & Pais, A. R. (2020). Two level filtering mechanism to detect phishing sites using lightweight visual similarity approach. *Journal of Ambient Intelligence and Humanized Computing*, *11*(9), 3853–3872. https://doi.org/10.1007/S12652-019-01637-Z/FIGURES/6

Rao, R. S., Umarekar, A., & Pais, A. R. (2022). Application of word embedding and machine learning in detecting phishing websites. *Telecommunication Systems*, *79*(1), 33–45. https://doi.org/10.1007/S11235-021-00850-6/FIGURES/2

Rao, R. S., Vaishnavi, T., & Pais, A. R. (2020). CatchPhish: detection of phishing websites by inspecting URLs. *Journal of Ambient Intelligence and Humanized Computing*, *11*(2), 813–825. https://doi.org/10.1007/S12652-019-01311-4/FIGURES/4

Sahingoz, O. K., Buber, E., Demir, O., & Diri, B. (2019). Machine learning based phishing detection from URLs. *Expert Systems with Applications*, *117*, 345–357. https://doi.org/10.1016/J.ESWA.2018.09.029

Sheng, S., Wardman, B., Warner, G., Cranor, L., Hong, J., & Zhang, C. (2009). *An Empirical Analysis of Phishing Blacklists*. https://doi.org/10.1184/R1/6469805.V1

Solanki, A., & Dogiwal, S. R. (2015). Implementation of an anti-phishing technique for secure login using USB (IATSLU). *Smart Innovation, Systems and Technologies*, *31*, 221–231. https://doi.org/10.1007/978-81-322-2205-7\_21/COVER

Wang, Y., Agrawal, R., & Choi, B. Y. (2008). Light weight anti-phishing with user whitelisting in a web browser. *2008 IEEE Region 5 Conference*. https://doi.org/10.1109/TPSD.2008.4562720

Xiang, G., Hong, J., Rose, C. P., & Cranor, L. (2011). CANTINA+. *ACM Transactions on Information and System Security (TISSEC)*, *14*(2). https://doi.org/10.1145/2019599.2019606

Xiao, X., Zhang, D., Hu, G., Jiang, Y., & Xia, S. (2020). CNN–MHSA: A Convolutional Neural Network and multi-head self-attention combined approach for detecting phishing websites. *Neural Networks*, *125*, 303–312. https://doi.org/10.1016/J.NEUNET.2020.02.013

Zhang, W., Jiang, Q., Chen, L., & Li, C. (2017). Two-stage ELM for phishing Web pages detection using hybrid features. *World Wide Web*, *20*(4), 797–813. https://doi.org/10.1007/S11280-016-0418-9/TABLES/4

Zheng, F., Yan, Q., Leung, V. C. M., Richard Yu, F., & Ming, Z. (2022). HDP-CNN: Highway deep pyramid convolution neural network combining word-level and character-level representations for phishing website detection. *Computers & Security*, *114*, 102584. https://doi.org/10.1016/J.COSE.2021.102584