Vol. 04, No. 01, pp. 54 – 71 (2023)
ISSN: 2708-0757

JOURNAL OF APPLIED SCIENCE AND TECHNOLOGY TRENDS

www.jastt.org

Web Phishing Detection Using Web Crawling, Cloud Infrastructure and Deep Learning Framework

Lozan Mohammed Abdulrahman[1,*], Sarkar Hasan Ahmed[2], Zryan Najat Rashid[3], Yousif Sufyan Jghef[4], Teba Mohammed Ghazi Sami[5], Umed H. Jader[6]

[1]*ITM Dept., Technical College of Administration, Duhok Polytechnic University, Duhok, Iraq, lozan.abdulrahman@dpu.edu.krd*
[2]*Computer Networks Dept., Sulaimani Polytechnic University, Sulaimani, Iraq, sarkar.ahmed@spu.edu.iq*
[3]*Computer Network Dept., Technical College of Informatics, Sulaimani Polytechnic University, Sulaimani, Iraq, zryan.rashid@spu.edu.iq*
[4]*Computer Engineering Dept., College of Engineering, Knowledge University, Erbil, Iraq, yousif.jghef@knu.edu.iq*
[5]*Computer Science Dept., Faculty of Science, University of Zakho, Duhok, Iraq, teba.sami@uoz.edu.krd*
[6]*Information Technology Dept., Erbil Polytechnic University, Erbil, Iraq, omid.jader@epu.edu.iq*

Abstract

The pandemic of COVID-19 obliges citizens to follow the "work from home "scheme. The Internet is also a powerful channel for social connections. The huge dependency of people on digital media opens doors to fraud. Phishing is a form of cybercrime that is used to rob users of passwords from online banking, e-commerce, online schools, digital markets, and others. Phishers create bogus websites like the original and deliver users spam mails. When an online user visits fake web pages via spam, phishers steal their credentials. As a result, it is important to identify these forms of fraudulent websites until they do any harm to victims. Inspired by the ever-changing existence of phishing websites. This paper reviews the work on Phishing attack detection and aims to examine techniques that mainly detect and help in preventing phishing attacks rather than mitigating them. Here we offered a general overview of the most effective phishing attack detection strategies focused on deep learning.

Keywords: *Phishing, Phishing Website, Phishing Attacks, Phishing Detection, Deep Learning.*
Received: January 03, 2023 / Accepted: March 09, 2023 / Online: March 11, 2023

I. INTRODUCTION

Nowadays, too many people are conscious that they use the internet to carry out different things such as online shopping, online charging, and online recharge[1, 2]. Due to the widespread usage of these clients, there are many safety risks such as cybercrime. Many cybercrimes such as spam, theft, cyber terrorism, and phishing are commonly carried out. New cybercrime and very common today are among this phishing [3, 4]. Phishing is an attempt at deception to gain confidential user information. The design site of Phisher is identical to any legal website or spoof user for the purpose of accessing user privacy information, such as username, password, bank information for various reasons [5-8].

Phishing is a form of social engineering assault aimed at using legitimate systems users' naivety and/or gullibility. This style of assault took its name because it uses bait like its homophone "fishing." Bait also seems like a convincing email when a phishing assault is carried out. Attackers can make sure their addresses are as lawful as practicable. These emails also threaten recipients of a malware-controlled website that provides malware or intercepts user identification [9-12]. In simple words, phishing is a deceptive effort to steal sensitive target details using disguises. This is a leading source of data fraud all around the planet. The broad technology known as "blacklist" is not only technologically outdated, but it is also easily deceived. A domain name is needed for the website. A phishing website is flagged as it is blacklisted. To trick the scheme, the phisher could have created a new domain that is completely new to technology [13, 14]. All analysis is then

doi:10.38094/jastt401144

completed again, but it takes a long time. Even a system failure can result by manipulating the website domain. Techniques that are not readily broken by fishermen are modernity prerequisites. The issue above reveals that a computerized approach is required to deter phishing attacks in addition to user training. This way, computers may detect websites that are harmful and discourage users from communicating with them [15-17].

Phishing websites are an unfortunate fact of daily life on the internet, which pose as simple social engineering scams. In such assaults, they build website pages by copying real websites and by sending spam letters, text, or online social networking to suspicious URLs of targeted victims [7, 18]. A falsified version of an original website is dispersed by an intruder by email, phone, or info expecting the intended victims to embrace email instances. You would also aim to include your confidential or extremely sensitive information (e.g., bank details, government savings number, etc.). A phishing attack results in an intruder obtaining bank cards and login details. In either event, there are a few ways to combat phishing. The increased use of Deep Learning (DL) has had a significant impact [19, 20]. Deep learning has improved speed, precision, and the ability to conduct comprehensive investigations as a result of email protection. Deep learning may identify spam, phishing, skewer phishing, and other types of attacks by using prior information in the context of datasets. This method of assault is likely to erode clients' interest in social networks such as online services [21-24].

The performance of detection methods during the classification process can be improved (whether the classifier is human or software). Their classification skills may be improved in the case of end-users through enhancing their awareness of phishing attacks by studying independently through their experience online or through formal training programmers [25-28]. This may be done during the learning process of profound learning or improvement of a rule-based scheme of detection measures in the case of the program classification. Techniques for detection not only help implicitly prevent end-users from falling prey but also help to enhance phishing honeypots in order to distinguish phishing spam from non-phishing spam [29-31].

The web phishing of personal and corporate data is used to capture sensitive knowledge. Some remedies have been identified for online phishing attacks by researchers. A blacklist or a white list is the default way to find out the page is valid [32-35]. We will check the URL and settle on this website to monitor this. The selection of established phishing sites to find unknown phishing sites is one way to identify phishing [36-38]. The study of web functionality is another means of finding out the bogus websites. Content, URL, and DNS functions may be used. A smart approach is to block harmful e-mails and flawed URLs. One constructive approach is to hold the domain and e-mail secure from phishing. Domain and lexical analyzes of malware threats are used to track them [39, 40]. A deep learning method is put into effect via the use of a learning algorithm. In the field of deep learning, some examples of typical learning approaches include a deep neural network, a feed-forward deep neural network, a recurrent neural network, a convolutional neural network, a constrained Boltzmann machine, a deep belief network, and a deep auto-encoder. The reviewers for this post made use of a technique called deep learning in order to recognize phishing attacks.

The remaining sections of the article are organized as shown below: In Section 2, we will talk about the theoretical foundations, and in Section 3, we will analyze the research that is pertinent to this study. In Section 4, we talk about the most recent anti-phishing trend, and in Section 5, we provide our findings and draw a conclusion about this research.

II. BACKGROUND THEORY

Phishing is a technique that fraudsters employ to fool their victims into exposing confidential information or downloading malicious software on their computers, such as ransomware. This is a technique that thieves employ to steal money [41, 42]. In recent years, phishing schemes have gotten so sophisticated that they often look to be clear mirror duplicates of the websites that they are trying to mimic. This is done in order to trick users into providing sensitive information. This provides the offender with comprehensive visibility into their victims' activity while they investigate the bogus site, as well as the opportunity to circumvent any further security measures that the victim may be implementing at this time. Additionally, the offender has access to the victims' personally identifiable information [33, 43].

Phishing attacks were found to be more common than any other kind of cybercrime, according to research that was carried out by the Internet Crime Complaint Center of the FBI in the year 2020. Before it was first mentioned in 1995 in the cracking toolkit AOHell, the term "phishing" was perhaps used for the first time in a publication that was titled Hacker magazine 2600. This was before it was first mentioned in 1995 in the cracking toolkit AOHell. This occurred before to the year 1995. Even though it's not really fishing at all, this activity resembles fishing in that private information is "caught" via unethical methods. However, it's not really fishing at all. It's a kind of fishing you may do [44].

Legislation, user education, public awareness, and technology security measures are some of the ways that phishing attacks may be prevented. These are also some of the ways that the ramifications of these assaults may be mitigated. However, these are not the only methods that this may be accomplished.

Because it is anticipated that the percentage of phishing attempts conducted against firms will develop from 72 percent in 2017 to 86 percent in 2020, it is necessary that both people and corporations acquire a clearer appreciation of the danger [45].

The act of collecting personal information from a user, which is more widely referred to as "phishing," is a severe crime that carries substantial consequences. Phishing is one of the most common terms used to describe this practice. Phishing websites may, in their many methods, aim to steal sensitive information from individuals, companies, cloud-based data storage systems, and even government domains [46]. Because of the reduced cost and simpler operation of software-based anti-phishing solutions, it is suggested to deploy software-based anti-phishing solutions rather than hardware-based ones. In spite of the fact that hardware-based anti-phishing solutions are

becoming more common, software-based anti-phishing solutions continue to be the more popular option. The strategies that are now available for identifying fraudulent websites are entirely ineffective when they are put up against challenges such as zero-day phishing website attacks [47].

Phishing, in which a fake website is made to look like an actual one in order to trick users into disclosing sensitive information, has become one of the most dangerous forms of illegal behavior that can be carried out online today. Phishing takes place when a fake website is made to look like an actual one [48].

It is widely agreed upon that one of the most important steps in the process of improving internet safety is recognizing phishing attempts. One of the current hardware-based methods that may be leveraged to offer further protection against phishing attacks on networks is the installation of an anti-phishing gateway. This can be accomplished by following the instructions provided by the manufacturer [49].

However, since there are so many different ways in which phishing scams may be carried out, the operation of such hardware devices is not only expensive but also ineffective. Virtualization technologies, which are used in fog networks, hold a great deal of promise for the deployment of anti-phishing gateways as software at the network's periphery, where they may include very effective machine learning methods for detecting phishing attacks. This deployment holds a great deal of promise because virtualization technologies are used in fog networks. Because virtualization technologies are used in fog networks, this deployment has a significant amount of potential for success. This deployment has a significant amount of potential that has not yet been used [50].

As a direct consequence of this, phishing has evolved into a major problem for a sizeable portion of the population that utilizes the internet in the modern day. Even though there have been advancements made to the security procedures that are used to prevent phishing, the most current incarnations of the scam are very difficult, if not impossible, to detect. This is in spite of the fact that there have been improvements made to the safety precautions that have been taken [51]. Due to the fact that it may be difficult to determine whether or not a website is real, phishing websites are successful in fooling many naive customers. Because of this, a significant number of consumers provide their personal information to fake websites. The proprietors of websites that engage in phishing will often circumvent anti-phishing systems by locating and banning the Internet Protocol (IP) addresses of individuals who seek to probe their websites. This is how they circumvent the anti-phishing mechanisms that are in place [52].

The process of "blacklisting" URLs is a frequent one in the field of cybersecurity. This is done with the intention of stopping people from accessing websites that contain dangerous content. Before developing a useful blacklist, it is necessary to first carry out a risk assessment. Following the completion of the risk assessment, the list has to be brought up to date so that it includes any potentially harmful websites that have come to light. It is recommended that you begin your investigation of websites that may or may not contain malware by looking through spam emails first since they provide an excellent starting point. On the other side, there is a significant amount of URLs that were acquired via the use of spam emails [53].

The URLs listed below have been compiled. The process of accessing and analyzing the content of such a large number of websites requires a significant investment of both time and resources in order to be properly accomplished. This is because computers and storage devices only have a certain amount of space available to them. The destination of more than half of the URLs we discovered in spam emails was the same location. The destination of these URLs was the same. Because of this, keeping copies of the whole information on each website takes up a substantial amount of storage space that is not necessary in any way [54].

Phishers often use websites that, on the surface, superficially resemble those of legitimate organizations in an effort to dupe unsuspecting victims into giving sensitive information. This is done in the hope that the victims would give over their information. Numerous research have been conducted in an effort to define the characteristics of phishing websites that may be useful in recognizing them. The goal of these studies is to assist users spot fraudulent websites [55].

Because they depend mostly on an out-of-date database that contains known phishing domains, the bulk of the presently available phishing detection systems are unable to correctly identify malicious websites. This is because of the database's poor quality. Because of this, the primary reason for the failure may be attributed to this. However, each year sees the creation of hundreds of brand-new phishing websites, many of which make an effort to pose as trustworthy businesses such as banks, file-hosting providers, and government institutions. Phishing is a form of online fraud in which an individual attempts to obtain sensitive information by impersonating a legitimate business [56]. Phishing is a kind of internet fraud in which a person tries to get sensitive information by impersonating a real company in order to trick their target into providing the information.

Phishing, which is a common technique for gaining the personal information of users, poses a severe threat to the network environments that users interact with on a daily basis. This is because phishing may be used to steal users' login credentials. Phishing websites are very dangerous and should be avoided at all costs; thus, it is essential that their risks be understood. The neural network is able to actively learn from large-scale datasets and is one of the most effective methods for identifying and blocking phishing websites. It is also one of the most beneficial techniques to heuristic machine learning, which is one of the most beneficial techniques to heuristic machine learning. On the other side, the training model may not be able to correctly predict and recognize phishing websites if the machine learning algorithm is pushed to overfit owing to the existence of certain useless properties [57].

Phishing assaults on mobile devices are notoriously difficult to detect using the traditional methods that are often used for desktop PCs. This is because mobile devices lack the screen real estate that desktop PCs do. This is due to the fact that portable electronic devices do not possess the same capabilities as desktop personal computers [58].

When referring to a Phishing Attack, the term "cybercrime" is the one that is the most suitable to use. This is because it includes an ordinary citizen impersonating an official figure via email or some other form of electronic connection. The most acceptable phrase to use is "cybercrime," which describes illegal activities committed online. An individual who has a penchant for engaging in inappropriate activity and who sends phishing emails that contain potentially destructive links or attachments and have the potential to escalate into full-fledged cyber assaults is referred to as an attacker. Because of things like having their identities stolen and losing money, amongst other things, the recipients of these emails are subjected to a variety of different kinds of pain, including financial, emotional, and other forms of hurt. This is because the senders of these emails are perpetrators of this scheme [59].

In recent years, phishing has developed into a significant barrier that individuals encounter in the course of their routine activities that take place online. Phishing is a sort of online con game in which the objective is to get sensitive information from unknowing victims by luring them to malicious websites. This information may then be used to commit fraud. Phishing may be thought of as a kind of online con game. Because of their exceptional classification capabilities across a variety of datasets and their capacity for active learning, BP neural networks have emerged as an essential heuristic machine learning tool in the fight against phishing websites [60]. This is because of the combination of their capacity for active learning and their exceptional classification capabilities across a variety of datasets. This is because kids have the capability of engaging in learning activities on their own. If the initial parameters of the BP neural network, specifically its beginning weight and threshold, are selected incorrectly, the network may be caused to reach a local minimum and experience delayed learning convergence. Both of these issues can be avoided by selecting the initial parameters carefully. This happens as a result of the network's incapacity to effectively gain knowledge from its previous errors. Both of these possibilities should be avoided at all costs if at all possible [61].

Phishing is an intriguing new challenge for everybody who uses the internet, but individuals who participate in online commercial or financial transactions face a very serious risk. Phishing is notoriously difficult to track down since the majority of people whose accounts are hacked do not realize that they are the subject of an attack until after their money has been stolen from their accounts. A client-side solution is unable to collect significant forensic data on phishing attempts, despite the fact that many browsers provide add-ons to protect users from phishing sites. This is the situation in spite of the fact that a large number of browsers provide add-ons to safeguard consumers [62].

Emails that are used for phishing often include links or files that are designed to steal sensitive information, such as the receiver's login credentials. These emails are sent in an attempt to trick the recipient into giving up the information by seeming to be legitimate. As a direct consequence of directly receiving these emails, the victims are exposed to a variety of unpleasant experiences, including emotional agony, financial loss, and the theft of their identities. These experiences are directly caused by the direct receipt of these emails [63].

Phishing is often used as the entry point for an attack when someone is attempting to get into a computer system for the very first time. Phishing is an attack vector. It is probable that its low-risk, high-reward character is to blame for the broad acceptance that it has received. This nature also makes it tougher to detect than it has ever been in the past, which contributes to this difficulty. This is due to the fact that the qualities it has make it far more difficult to identify than it ever has been in the past.

The fast growth of the internet has led to a shift away from traditionally conducted activities that take place offline, such as banking, shopping, and other activities of a similar kind. This shift has prompted a movement away from offline activities. Because of this development, a departure from the traditional ways that have been used for a considerable amount of time has taken place. Phishing is one of the many distinct varieties of cybercrime that have developed as a consequence of this, as a result of the fact that it generated the possibility for their growth. This is due to the fact that it created the opportunity for their development [64].

By engaging in illegal activity on the internet and pretending to be legitimate businesses, the objective of cybercriminals is to collect personally identifiable information such as user names, passwords, and credit card numbers. One tactic that fraudsters use is to impersonate legitimate companies while operating online. Recognizing a website's Uniform Resource Locator (URL) and using that information to ascertain its legitimacy is a difficult and time-consuming operation. This is due to the fact that the procedure makes use of the user's inabilities, which makes completing the assignment a challenging endeavor [65]. Although there are a number of products on the market that make the claim to be able to detect websites that participate in phishing, the majority of these tools make use of a heuristic approach or blacklists, and as a result, they are unable to effectively prevent phishing from happening. Although there are a number of products on the market that make the claim to be able to detect websites that participate in phishing, these tools are unable to detect websites that participate in phishing. This is despite the fact that there are a number of products now available that make this claim about their capabilities [6].

Due of the prevalence of this pattern, internet security has emerged as a primary focus of discussion in recent years. This is a direct consequence of the expanding number of individuals who use the internet in addition to the risks that come along with it. Research on phishing URLs is now placing a significant amount of attention on lexical and host characteristics as the key areas of study in the subject [66].

Customers who make their purchases online face a variety of risks to their privacy and safety that have become notably more complex and dangerous as a direct result of improvements in information technology that have taken place over the course of the last few years. These customers have a much increased possibility of falling prey to the dangers that are there. It is generally agreed that an attack using an Internationalized Domain Name (IDN) homograph is one of the most popular types of cyberattacks [61].

This type of attack takes advantage of the fact that many characters look the same in order to trick internet users into

visiting malicious websites by utilizing domain names that sound similar to those they typically use. This attack also takes advantage of the fact that many characters look the same. This kind of assault also takes use of the fact that a large number of characters have the same general outline. This kind of attack also takes use of the fact that a large number of unique characters have the same appearance. It has proved to be a particularly difficult task to detect IDN homograph attacks owing to the fact that they may send clients to bogus websites or breach their privacy while they are online. Detecting these attacks has been a particularly difficult job. This is due to the fact that there is a chance that they may send customers to the aforementioned other websites [67].

Users of the internet who were unaware of the risks they faced have been found to be responsible for the billions of dollars' worth of damages that have been caused as a result of fraudulent behavior on fake websites. These damages have been caused as a result of fraudulent behavior on fake websites. Users who visit these websites will have a tough time visually recognizing them as fakes owing to the style and structure of the websites that they will visit. This will be the case since these websites will be visited by users. A growing number of consumers increasingly do their shopping and payment for their products online, taking advantage of the plethora of user-friendly payment options that are made available by the internet. This trend is expected to continue in the near future. In the following years, it is anticipated that this pattern will continue to show signs of increasing. In order to complete the authentication procedure on many different websites, users are asked to provide personal information at several separate places. In order for the website to correctly confirm the user's identity, this is an absolutely necessary component [64].

Phishing websites, on the other hand, will use the information they get in a way that is immoral once they have it in their possession. The proliferation of automated detection tools to combat fraudulent websites has resulted in the emergence of fairly simple methods for identifying fraudulent activity and taking appropriate action against it. These instructions are available on a number of different websites that you may visit. The use of these technological tools is very required if one want to battle fraudulent websites in an efficient manner [68].

Websites that encrypt user data using HTTPS are more likely to earn consumers' confidence than those that encrypt user data with SSL/TLS. This is because HTTPS offers a better degree of security than SSL/TLS does. Phishing websites that use this protocol in order to spread their information reap the advantages of an increased degree of trust for themselves as well as their users. In this research, we explored the usefulness of a range of different heuristics for spotting the presence of phishing websites. These heuristics included: These heuristics consisted of the following: These heuristics took into account a variety of parameters, including the cipher-suite that was selected for the server, the version of the SSL/TLS protocol, and the certificate data for the server [69].

The relevance of digital activities, in particular those involving the cloud and mobile technology, greatly rose in importance in the year 2020, when people all over the globe were desperately attempting to stem the spread of the Coronavirus Disease 2019 (COVID-19). This was particularly true at the time when individuals were working to contain the COVID-19 virus's spread. In particular, the measures that were taken to prevent the sickness from spreading were of the utmost relevance with respect to this matter. As a direct result of this tendency, there has been a noticeable increase in the frequency of phishing and other forms of cyberattacks [70]. Cyberattacks aim to steal sensitive information from their victims. Machine learning algorithms have the capacity to detect websites that are used for phishing by classifying websites as either safe or hazardous. This allows the algorithms to identify websites that are used for phishing.

In this day and age, when cloud computing is the norm, each and every email that is sent needs to be encrypted from the very beginning all the way through to the very end. This includes any attachments that may be included. This include any attachments that could be included in the package. In the hyperconnected world of today, several forms of cybercrime, such as phishing via email, EBomb attacks, DNS spoofing, and others, are becoming prevalent. Phishing schemes, which are carried out via email, are responsible for the overwhelming bulk of these assaults. The receivers of these emails are tricked into downloading malware that was not previously familiar to them via the use of these emails [71].

With the current pandemic that is Covid-19, which is sweeping the entire world and forcing people all over the place to stay indoors, inadvertently increasing the online digital footprints of all, there is an increase in SMB (Server message block) port all over the world, which is leading attackers to find their victims easily by unique, dynamic, and various other vulnerabilities that no standard virus, malware detection software, which was previously provided by the IT industry, can detect. There is also an increase in the number Because of this, there has been a general uptick in the amount of violent acts that have been committed. This leads one to believe that there has been a rise in the total number of persons who have already, it is feasible to fingerprint and monitor IoT devices, web browsers, phones, and even automobiles, and to reroute their connections via or to criminal organizations. This is also the case when it comes to the ability to divert their connections [72].

Automobiles also have the potential to be used in this manner. When it comes to rerouting their connections, this location experiences the same thing at the same time as every other location. In the same vein, this is how things stand with respect to redirecting their connections. There is no method available at this time to evaluate the dependability or validity of a remote service that is accessed by a device that is part of the Internet of Things (IoT). This is because there is currently no viable mechanism to confirm that the service has been provided. Due to the fact that they are able to communicate in both directions with remote services, Internet of Things devices and other edge devices are both susceptible to attacks that are known as man-in-the-middle attacks. These attacks allow a malicious third party to intercept and manipulate data transmitted between two parties (MiTM). Phishing and pharming are two methods that may be used to fool a person into connecting to a device that they are not acquainted with, even if there is a possibility that the device might be dangerous. This may be performed by use

words and pictures that are intended to mislead. Use of information that is posted on the peripheral of the network and that indicates the starting and stopping places of a connection is one method that may be utilized to lessen the impact of this risk. Another method that may be utilized is the utilization of information that is provided by the network itself [73].

Internet usage has invaded every facet of contemporary life as a direct consequence of the expansion in both the capacity of computers to store information and the speed at which it can be transferred. This development in capacity and speed has allowed for an increase in the amount of information that can be communicated. The increase in both capabilities has made it possible to send information more quickly, which has, in turn, led to an increase in the number of people who use the internet.

As mobile devices and cloud computing become more vital to people's day-to-day lives, hackers are increasingly concentrating their attention on these technologies. As a direct consequence of this fact, hackers are increasingly focusing their attention on mobile devices in addition to cloud computing. In the field of computer network security, the threats posed by phishing and other forms of malicious websites have grown increasingly pervasive and serious over the course of the last several years. This is especially true with regard to the former. This tendency may have its roots in the rise in the total number of risks that might potentially be encountered [74].

It is required for attackers to first install malicious software on the computers that they are targeting using any one of a huge number of different techniques in order for them to effectively steal important information and do considerable harm. This may be done in a variety of different ways. As a direct consequence of the rapid pace at which it has progressed, the primary objective of malware has moved from one of destruction to one of infiltration. This shift occurred as a direct result of the rapid rate at which it has developed. The lightning-fast rate at which technology has progressed led directly to the occurrence of this change in perspective. The signatures of malicious software have gotten more difficult to decipher over the course of the last several years, which has directly contributed to an increase in the level of difficulty associated with the detection process. Malicious software not only makes an effort to disguise static signatures from antivirus software, but it also makes an effort to conceal dynamic signatures from antivirus software. This is done in an attempt to circumvent detection by antivirus software. This is done in an attempt to avoid being discovered or found out [75].

Rogue domains have become one of the most significant dangers to the Internet's continued existence as a result of the passage of time. The Domain Name System (DNS) is frequently abused by threat actors who want to trick people into visiting malicious domains that house malware, botnets, phishing websites, or spam messages by using drive-by downloads. These threat actors want to achieve their goals by tricking people into visiting malicious domains. Threat actors aim to accomplish this goal by manipulating the DNS to deceive users into accessing malicious sites. In order for these threat actors to achieve their goals, innocent people are going to be duped into accessing harmful websites. Threat actors seek to accomplish this goal by taking advantage of loopholes in the DNS in order to deceive customers into accessing harmful websites. Every year, a large number of well-known businesses end up becoming victims of these threats, and in some cases, even a single assault may result in losses that are so severe that the firm is declared inoperable as a consequence of the incident. Because of this, it is of the utmost importance to identify a website as potentially hazardous as soon as it is humanly feasible and to label it as such. It is of the highest significance to note that a website may include potentially hazardous content [76].

Before, the only method to determine whether or not a domain was harmful was to check to see whether it was already on a blacklist that was already in existence. This was the only way to determine whether or not a domain was malicious. On the other hand, this technique was ineffective since it was unable to identify domains that were created from the ground up from the very beginning. This meant that it was unable to find domains that were built from the ground up. As a direct and immediate result of recent advances in Machine Learning (ML) approaches, the ability of domain vetting systems to detect anomalies in data has significantly increased and is now far more powerful. Every attempt to increase the performance of a machine learning model has to first begin with the building of a trustworthy feature engineering process as its base in order to have any chance of being effective at all. This is the only strategy that will give the attempt even a remote possibility of being successful [77].

Phishing is a common method of social engineering that is used to steal sensitive information such as login credentials, passwords, and credit card numbers. Phishing may also be used to trick people into divulging their personal information. Phishing is a term that may also refer to spear phishing and phishing through email. Phishing is a technique of social engineering that collects sensitive information such as passwords and credit card details via the use of deceptive electronic messages known as "phishing." It takes place when a hostile actor poses as a reliable source in an email, instant messaging, or text message in order to trick a victim into installing harmful software. This may be done for fraudulent purposes. It's possible that this might happen in an attempt to deceive a victim into installing harmful software. This might take place through any one of the several various modes of internet communication that are already accessible.

This could take place via any one of a wide range of distinct routes of communication, such as, but not limited to, but perhaps include. After this, the victim is duped into visiting a malicious website, which may result in the broad distribution of harmful software, the locking down of the system as a consequence of an attack employing ransomware, or the disclosure of sensitive data. It is not completely out of the question for a violent act to have effects that are far more serious than was originally anticipated. This is the category that encompasses the potential results for people, such as participating in illicit activities, having their money stolen, or having their identity stolen [39].

Other possible outcomes include. In addition, phishing is often used as a component of a bigger assault, such as an advanced persistent threat (APT) event, in order to get a foothold in corporate or governmental networks. This is done in order to steal sensitive information such as usernames and passwords.

The purpose of this is to get sensitive information such as login credentials by stealing them. The purpose of this action is to get unauthorized access to sensitive information such as passwords and login credentials. This is done in order to give a way via which access may be achieved to credentials or information that is deemed to be confidential. In the second scenario, employees who have been hacked make an effort to circumvent security measures by, for example, spreading malware inside of a restricted network or gaining access to sensitive information. This occurs because employees who have been hacked attempt to circumvent security measures. This takes place as a result of the fact that they are trying to hide the fact that they have been hacked [78].

Both of these fictitious scenarios are examples of circumstances in which compromised persons could circumvent security systems. In addition to the loss of customers and financial resources, a frequent effect of a business being the target of a cyberattack is a decrease in the company's market share, which occurs in addition to the loss of customers.

One of the other typical repercussions is a drop in customer numbers. In addition to this, it is possible that the conclusion will have a detrimental effect on the company's reputation. It is feasible that an assault using phishing may result in a security breach from which an organization would have a difficult time recovering if the effort was large enough. This would depend on the breadth of the phishing attempt. If anything like this happens, the phishing effort will have been unsuccessful. Because of this, it will be considerably more difficult for the firm to continue operating effectively as it has in the past. Phishing emails that seem to have come from the myuniversity.edu domain have been sent to the vast majority of the university's faculty members, making them potential victims [79].

At least one of these exchanges of information has been delivered to each of these people. These are messages that have been sent through email to members of the university's teaching staff. The user will get an email telling them that their current password will become invalid after a certain date, and that this information is contained in the email. The user will also be informed that the email itself contains this information. In the event that you find yourself in a situation in which you need to change your password, the procedures for doing so may be found at MyUniversity.edu/renewal. A time restriction of twenty-four hours is also included in these instructions, and it is imperative that this time limit be followed to in order to complete the project properly.

Phishing is a kind of identity theft that takes place via the use of electronic mail and consists mostly on a game of numbers. Even if only a very small percentage of recipients fall for the hoax, an attacker who sends out thousands of fraudulent communications has the opportunity to collect a large amount of information as well as large quantities of money. This is true even if the percentage of recipients who fall for the hoax is extremely low. This is still the case even if there are just a few people to get the gift. It makes no difference how few individuals wind up actually getting the gift since this is still the case. As was said earlier, those who carry out assaults depend on a broad array of tactics in order to improve the possibility that their efforts would be effective. For instance, they can go to a lot of time to make it seem as though an email was received from a respectable company when, in reality, the email did not come from that company at all [80]. This is done so that the recipient would be more likely to click on the attachment and read it. This is only one example out of an extremely large number of others.

The credibility of a communication is improved when it uses terminology, fonts, logos, and signatures that are all consistent with one another. This increases the likelihood that the communication is trustworthy. In addition to this, those who prey on others will often attempt to coerce the people they target into engaging in some behavior of their choosing by threatening them with very negative outcomes if they do not comply with their demands. The recipient of an email might, for example, find out that they are on a timer and receive a warning that their account will be terminated if they do not react if the email is not responded to within a certain amount of time. This could happen if the email is not responded to within the allotted amount of time [81].

As a direct consequence of the strain that is caused by the circumstance, the user's actions become more careless and sloppy as a consequence of the circumstances. Last but not least, the links that are included inside the messages give the idea that they are authentic, but in fact, they often contain a misspelled domain name or additional subdomains that have been added. This gives the impression that the messages are not genuine. This gives the impression that the security of the communications has been breached, which is a reasonable conclusion to draw. Traditional phishing efforts, on the other hand, aim their attacks at broad groups of individuals or organizations in general and try to trick them into giving out sensitive information. On the other hand, spear phishing is an attack that is directed against certain persons or companies. On the other hand, spear phishing is directed only at a single target, whether it be an individual or a company. There are two distinct types of phishing, which are known respectively as spear phishing and generic phishing. There are two distinct types of phishing, which are known respectively as spear phishing and generic phishing. Two distinct subsets of phishing are distinguishable from one another: general phishing and general phishing [82].

Phishing attacks come in a wide variety of forms, with email phishing being one of the most common types. However, there are many more conceivable varieties. The attacker in this kind of assault will send out spam emails in an effort to trick the recipient into exposing their personal information or login credentials. This sort of attack is known as phishing. They could inquire for the recipient's name or social security number, for instance. The great majority of assaults are what are known as "bulk attacks," which are defined as attacks that are distributed in quantity to several victims.

The term "bulk attacks" is used to describe the vast majority of assaults. The fact that several victims are often involved in an assault is what's meant to be referred to when using the phrase "bulk attacks." Financial institutions, email and cloud productivity service providers, as well as streaming services, are common targets; nevertheless, the objectives of the attacker might vary from benign to malicious. This security weakness might lead to monetary loss, infections with malicious software,

or targeted email assaults on persons working for the targeted firm. All of these outcomes are possible. It's possible that accounts for streaming services that have been compromised may be put available for sale on anonymous markets [83].

Phishing may be broken down into a number of different subcategories, the most common of which is known as spear phishing. There are various subtypes of phishing that may be differentiated from one another. Spear phishing is a subset of the category of targeted phishing attacks. In this kind of phishing, certain persons or corporations are the targeted targets of a series of emails that seem to come from a reliable source. These emails pose as being sent from that source in an attempt to gain the recipient's trust. In order to boost its chances of being successful and to increase the amount of money it may take, it often makes use of the victim's own personal information. As they are the most likely candidates to be attacked because of this access, the executives or workers in the financial department who have access to secret information are typically the targets of attacks of this kind. This is because they are the most likely candidates to be attacked [84].

Phishers often utilize this strategy. It is possible to instantly determine where clicking a link will take you before actually clicking it in many email applications and online browsers. This is made possible by the fact that the URL of the link is shown in the status bar of many programs. Despite this fact, it is feasible for certain phishers to discover methods to break past the security mechanisms that have been put into place. Attackers may make advantage of internationalized domain names (IDNs) by using IDN spoofing or homograph attacks in order to establish fictitious websites with addresses that are confusingly similar to those of actual websites. This is done with the intention of siphoning visitors away from the genuine websites.

Phishers have been known to employ these kinds of attacks to conceal malicious URLs inside ones that seem to be legal by exploiting weaknesses in open URL redirectors. This is done in order to trick victims into giving out sensitive information. Users are led to websites that provide the impression of being authentic in order to accomplish this goal. Even digital certificates such as SSL may not be enough protection against these kinds of attacks because phishers are able to get legal certificates and modify the content on their websites to make them appear to be authentic ones. This makes it possible for phishers to fool users into believing that their websites are legitimate. In order to avoid being detected by anti-phishing systems, it is standard practice for phishers to replace the text in malicious emails with visuals instead. This is done in order to circumvent detection. This is done in order to avoid being detected by these systems and having one's actions taken against them. Anti-phishing software has progressed to the point that it is now able to make use of optical character recognition in order to decipher text that is cloaked inside images (OCR) [79].

Phishing frequently makes use of social engineering techniques to trick victims into performing the actions that the criminals behind the crime want them to, such as opening a file, clicking on a link, or providing sensitive information. These actions include opening a file, clicking on a link, or providing sensitive information. Con artists who masquerade as law companies commonly employ the strategy of threatening to cancel or seize the victim's bank or insurance account as a way of getting the victim's personal information in order to commit the fraud. This is done in order to develop a false feeling of urgency and create the idea that there is an urgent need for the victim's participation [85].

This is done in order to create the impression that there is an immediate need for the victim's cooperation. Phishing through impersonation is a kind of online deception in which the perpetrator assumes the identity of another person in order to trick potential victims into accessing malicious websites. These are the kind of tales that are considered to be fake news. If you click on any of these links, you will almost always be sent to websites that contain dangerous content. These websites may give the impression that they are trustworthy, but in reality, they are operated by cybercriminals who want to steal your personal information or deceive you into downloading malware. If you visit one of these websites, you should exercise extreme caution before providing any personal information [86].

A. Phishing Components

Three elements consist of phishing methods, the attacking vector and strategic approaches employed during the attack [84]. The tools for phishing include phishing methods. Fig. 1. shows the interlink between phishing mediums, vectors and techniques used.

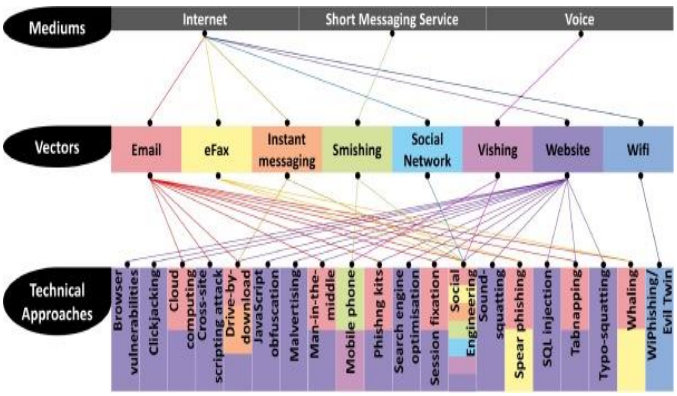

Fig. 1. The interlink between the medium, vector and technical approach of the phishing techniques

1) The phishing medium is the main means by which victims receive phishing assaults. The Website, speech or quick message service are three widely used bases (SMS). Online access has opened a big chance for phishers to quickly reach their victims [84] [87].

2) The channel vector used by medium-based phishers is the vehicle used to initiate phishing attacks. The phishers can access the Internet through emails, eFax, websites, instant messages and social networks. Vishing is the speech phishing vector which is the quick message phishing vector (SMS) [84].

3) Technical methods are the technical means utilized to further improve the potency of phishing in addition to social engineering. Currently, the phisher uses a number of methods such as Drive-by-Download, Man-in-the-Middle (MITM) and XSS assault, tab napping and spear phishing, whaling,

optimization of search engine (SEO), fixing of sessions, ads, social engineering, obfuscation of JavaScript, cell phone vulnerabilities, cloud storage or WiPhishing and Evil Twins. Such techniques that circumvent victim or social engineering needs to be misled include SQL injection, typo-squatting and sound-squatting. Phishing kits are often subject to technological techniques, but are an instrument to help deploy phishing attacks and not a phishing assault by itself [84].

B. *Types of Phishing Attacks*

Phishing threats exist of different kinds. These attacks are primarily designed to extract confidential information from end-users [88]. Fig. 2 shows various types of phishing attacks.

Fig. 2. Types of Phishing Attacks

1) *Email Phishing*

E-Mail phishing attack: an attacker sends users an email to check account credentials, user malfunction scheme, fake account fees, unwanted shift to accounts, new free services needing swift operation, and many other scams are sent to several people. Some e-mails are more difficult to identify as phishing. The language and grammar cannot be used as a phishing e-mail because the email is more carefully designed. It is simpler to detect the source when you search the email source and the actual connection to which you are pointed [89].

2) *Spear Phishing*

In contrast to random device consumers, this intruder targets particular persons or companies. This is a more detailed variant of the phishing industry that needs special expertise, including the power system, about an organization. Emails are sent to those people in this assault unlike phishing [89].

3) *Whaling Phishing*

Attacks on senior executives are also more targeted in whaling. Whaling has the same end result as every other form of phishing assault, but the strategy is far more discreet. Since the perpetrators are attempting to emulate senior personnel, tricks like false links and malicious URLs aren't efficient. Whaling scams involving fictitious tax returns are becoming more popular. Criminals prize tax forms because they include a wealth of details, including names, addresses, Social Security numbers, and bank account information [88, 89]

4) *Smishing Phishing*

Smishing is a phishing scam carried out through SMS (SMS). Crafty phishers deliver text messages, such as banks and internet stores, from trustworthy senders. In standard cases, certain text messages include URLs or connections to websites where receivers download viruses and other malware onto the mobile device of the victim [89].

5) *Vishing Phishing*

It's sometimes called Voice Phishing. It is a type of telephone fraud that utilizes voice messages to collect victims' personal details or assets. In order to attract suspects, Vishing utilizes artificial speech records. An automatic voice call at Vishing indicates that the bank account of the receiver is affected. Then the voice message asks the receiver to dial a certain free number. If you dial the toll-free number, you can collect the user's bank account and other sensitive information using the phone keyboard [90].

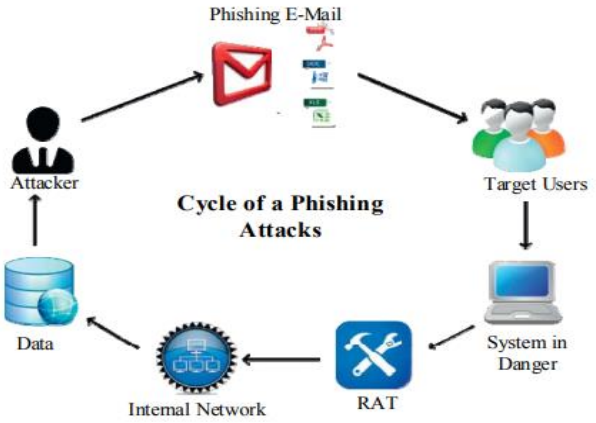

Fig. 3. The processing cycle of phishing attacks

6) *Pharming Phishing*

Pharming is another phishing variation. Contrary to other tactics, people may not be targeted. A vast number of individuals will be victimized without the need to be personally attacked [91]. The processing cycle of phishing attacks is illustrated in Fig. 3.

C. *Phishing Websites and Detection Techniques*

Websites for phishing are replicas of legitimate websites. The whole website is not built for phishing, but just the home page or page where the user will provide inputs, such that some data is sent to the intruder. There are many methods of creating a Phishing website, like accessing the source code of every given website, cloning the website, or using special software. There is a social engineering toolkit named SET in kali-Linux which is primarily used for website cloning[92]. In their URLs, page text, icons, hyperlinks, hosting domains, age of the domain, source code, SSL certificates, etc., the discrepancy between the phishing websites and the legal websites is seen.

Currently, they are shortened since URLs are too big. The URL shorteners are used. These URLs seem to be distinct from the usual URLs, have no domain, sub-domain, and top-level domain (TLD). These URLs just go to the initial URL. Since no details are found on this website, this tool is used by attackers and this kind of URL is forwarded to the end-users. End users pay no attention to URLs; they just click on the URL and enter their passwords, but they remove their sensitive details. These pages must be detected sooner rather than later in order to avoid

data loss. For the detection of phishing websites, there are various methods of detection [93].

1) User training

The education and notification about phishing attacks by consumers and business staff have an effect on phishing attack prevention. For training customers, several approaches were suggested. Several researchers found that immersive instruction is the most effective solution to helping consumers differentiate between phishing and legal websites. While user training is an efficient process, human mistakes remain and people are likely to forget about their training. Training needs considerable time, and non-technical consumers do not like it very much[94].

2) Software detection

While user training can deter phishing attacks, hundreds of websites attack each day, so it's a tedious and often realistic challenge to use our training for any website. The usage of the app is another way to spot phishing websites. The program will assess various variables such as website content, email message, and URL before taking a more accurate final judgment than mankind [95]. Multiple phishing identification program approaches that are classified as following are proposed:

2.1) List-based

This type is categorized into: white-listing and black-listing. Also, they classified as standard methods or approaches to databases. They have very high reaction time and precise identification. By creating a list that blocks the site before loading, the most innovative way is to identify and block phishing websites. We have a whitelist and a matching blacklist including websites permitted to be accessed, and that do not efficiently alert the browser what websites are permitted and barred, respectively. The search will take place until a request for the website is submitted to the server. The URL is then reviewed in the blacklist to see whether it is a phishing website or not. It can then be examined. As detected, the browser prevents website entry. If not, then the consumer may proceed. Many common web browsers already use this way to secure even unknown users [96].

2.2) Heuristic based

The blacklist can be expanded and new threats can be identified by using functions from phishing pages to identify the phishing threat. However, the restriction cannot identify the whole new assault and can easily be circumvented until the assailant learns the algorithm or functionality. Furthermore, the website may or may not have typical characteristics [94].

2.3) Visual Similarity-based

This method takes and stores snapshots of web pages in libraries. If lookalike websites are present, then the snapshots of both websites compare to detect whether or not this website is phishing. The drawbacks of this approach are that it takes longer to perform and is impractical. It needs huge storage area to store website screenshots. The first one that appears on other websites with the same URL is deemed legal. But it is possible that the first one may be a phishing spot [97].

2.4) Machine Learning

In massive datasets, this method operates effectively. The current method is thus removed from pitfalls and zero-day-attacks may be detected. Machine learning classifiers are effective and accurate classifiers. Lei stung depends on training data, function settings, and classification style. This is limited since it struggles to identify when attackers host their sites using infected domains. In this field of phishing identification, a lot of studies have been carried out. Much research has been done on various classifiers to increase the accuracy of phishing website identification. The KNN, SVM, Decision Tree, ANN, Naïve Bayes, and the Random Forest are the different classifiers used. Well before they are developed, these classifiers help predict web pages. Machine learning solves the zero-hour phishing challenge. The accuracy of the classifier varies according to the data set size and the form of characteristics used. Frameworks for phishing attacks are also created[98].

3) Deep Learning Algorithms for Web Phishing Detection

Deep phishing detection learning methods In light of the existing deep learning approaches for phishing detection and the categorization of the previous works, they may also be split into three categories, that is, unsupervised (e.g., autoencoder (AE), deep belief network (DBN), and generative adversarial network (GAN)), supervised (e.g., deep neural network (DNN), convolutional neural network (CNN), and recurrent neural network (RNN)), and other hybrid methods show the details of categorization in Figure 4. deep learning algorithms can execute in an end-to-end manner. Deep learning algorithms are essential compared to shallow models for large datasets. The adoption of various types of algorithms for deep learning could provide various advantages for detecting phishing. Often supervised learning-based algorithms lead to a high degree of accuracy because the amount of knowledge manually labelled samples give. Unattended learning approaches are typically poor in efficiency without adequate awareness of the labelled results. Manual marking however, especially for complex assaults, is a long-term undertaking. Without previously awareness of the attacks phishing, intentioned learning-based algorithms could work well, a clear benefit. Hybrid strategies reduce the amount of workout samples and retain a sufficiently high value to address variable assault scenarios. However, the structure and computer time of the system are typically dynamic and preclude its widespread use [99].

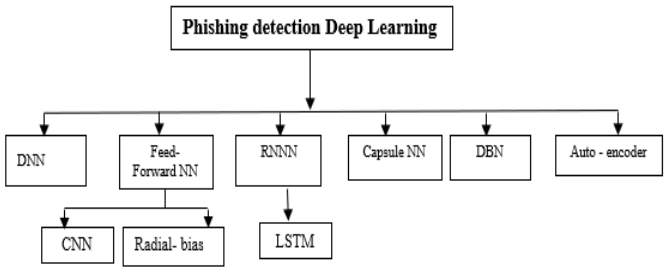

Fig. 4. Categorization of the current deep learning methods for attack detection

3.1) Deep Neural Network (DNN)

The profound neural network represents the kind of machine learning that takes advantage of many layers of nodes to extract high-level functions from data. This implies the data is transformed into a more abstract and artistic part. This multi-layer attribute gives the benefit of expressing complex and lower-parameter functions, making it possible to facilitate function extracting and representation learned by DNN. The deep neural network algorithm for Attack phishing detection In essence, in DNN there are three-layer groups. Generally speaking, we consider the first layer to be the input layer, the last layer to be the output layer and the central layer to be cached. In order to detect the issue of over-fitting, an efficient model for phishing websites based on optimal function selection process and the neural network has been suggested, namely OFS-NN. The algorithm suggested will reduce to a great degree the overfitting issue of the underlying neural network [100].

3.2) Convolutional Neural Network (CNN)

Used to detect the fishing algorithm Assault. CNN has a deep learning computation and framework that are indicative and widely employed deep learning techniques. Specifically, CNN utilizes a multi-layer architecture with limited preprocessing. The CNN basic framework consists of the entry and exit layers and other secret layers, including the convergence, the pooling and the complete link layer. CNN uses a comparatively less processing method compared to other classification algorithms which is unrelated to its primary advantage the feature architecture which contains prior information [101].

3.3) Deep Belief Network (DBN)

Is a generative graphic model consisting of several secret layers with a layer relation and there is no link inside each layer between the units. DBN consists of two stages, uncontrolled training and supervised training. DBN learns to reconstruct the input probabilistically and is called function detectors on the input. DBN acts as a classification and grouping in a regulated manner. DBN derive and learns from this information to produce the best fit for a deep hierarchical representation (knowledge)[102].

3.4) Capsule Neural Network (Caps Net)

The learning machine framework, which is a kind of artificial neural network (ANN) for better modeling of hierarchical relationships (Caps Net). Add structures to a neural network named "capsules" (CNN) and recycle the contribution from some of the capsules to form more secure images for higher capsules. The result is a vector composed of the probability of detection and a pose[99].

3.5) Auto-encoder (AE)

AE is a deep learning process that is not supervised. An entry layer, an output layer, and an overseas layer are part of the simple AE architecture. In this context, the input and input layer are of the same form and the secret layer reflects the possible characteristics and structure of the input, whether the input is identical to the output. The AE objective is to turn inputs into outputs with the minimum variance necessary [99].

3.6) Recurrent Neural Networks (RNN)

The recurring neural network is a type of Artificial Network where links between nodes shape a guided graph via a temporal series (RNN). This enables it to display complex time behavior. RNNs may use their own internal state (memory) to process the sequences of vector data from feedforward neural networks [103].

3.7) Long Term Short Memory (LSTM)

Long-term memory of the artificial recurring neural networks used in the area of deep-seated learning (LSTM). Unlike standard neural feedforward networks, LSTM has feedback connections. It does not only process single datasets (for example images), but also whole data sequences such as voice and video [103].

3.8) Radial Basis Function (RBF)

Is a kind of neural feed-forward network widely used for problems in classification and regression due to its simplest structure and efficient computational data outcomes [104].

III. Literature Review

In this section, we explain the method of detection of phishing websites based on deep learning. Detection of the Phishing website focused on deep learning is a hot point of recent research on the Phishing website. Deep learning outcomes usually rely on the accuracy of the features extracted. The research emphasis is on the way more successful features are extracted and selected before they are processed, and therefore many studies have been done during the years 2018,2019, and 2020 that used algorithms based on deep learning for phishing detection.

Yao et al. [105] suggested relative detection procedure evaluates the URL's validity in the two-dimensional code through the use of a legal logo in the two-dimensional code. The trouble with this approach is that the emblem icon is tiny and by means of the standard recognition method the recognition rate of the small object is limited. For small-scale detection, they used the better Faster R-CNN and measured their effect on the Flickr Logos 32 dataset. The experimental findings demonstrate the efficiency of the logo-detection process, which can be used to identify phishing attacks on two dimensions.

Su et al. [71] proposed a new designated phishing websites detection scheme with LSTM recurring neural networks (RNN). LSTM benefits from data collection and long-term dependence. LSTM has strong learning abilities, can automatically learn data characterization, and has great potential in the face of complex large high-dimensional data without manual extraction of complex features. Experimental findings show that 99.1 percent of this model is better than other neural network algorithms.

Yang et al.[106] proposed a multidimensional phishing method focused on a rapid detection method with profound

learning (MFPD). The first stage involves extracting character sequence features from the provided URL and using them to rapidly classify them in-depth so that third-party assistance or previous information regarding phishing is not needed. The second phase combines URL mathematical functionality, webpage code with webpage text characteristics and quickly classifies the results of profound learning in multidimensional features. The solution will limit the threshold detection period. The precision is 98.99% when tested for a dataset that contains millions of phishing URLs and legal URLs, and false positives are just 0.59%. The experimental findings demonstrate that the detection quality can be increased by changing the threshold fairly.

Singh et al. [107] suggested a new design for a phishing detection device with the use of profound training to avoid such attacks. The machine uses URLs to detect the phishing website using a Convolutional Neural Network (CNN). The device has achieved 98.00% more accuracy than the previous one. No functional engineering is needed because CNN extracts URL features automatically and through hidden layers.

Sindhu et al. [108] applied the current machine learning algorithms used to identify phishing websites. Improved Random Forest Classification System, SVM, and Network Classification Methods for backpacked reproduction. The accuracies obtained by means of the Random Forest, SVM and the Neural Network with background algorithms are 97,369%, 97,451%, and 97,259% respectively. The lexical attribute extraction from the URL was used to develop the algorithms. SVM is the strongest classification of the three since its frequency is higher than neural networks. Although the random forest classification generally provides more precision than the SVM classification, it is not constant that the accuracy rates, therefore, show SVM to be better than that of the random forest classification. The chrome extension will identify phished URLs to 97.451 percent precision.

Saha et al. [109] Presented a data-driven webpage detection system for phishing with a deep learning approach. For predicting phishing web sites, the multilayer perceptron, also known as a feed-forwards neural network. The Kaggle dataset is compiled and includes information on 10,000 web pages. There are 10 characteristics. The results showed 95% accuracy in preparation and 93% precision in testing. The difference between the training and accuracy of the tests was shallow so that the proposed model would effectively identify unknown websites, as well as learning from datasets. The valid precision of website detection is 98.4 percent higher than the phishing method.

Adebowale et al.[110] Proposed exploration of the possible differentiation by CNN and LSTM of unique legitimate URLs from phishing URLs as a joint classifier in a novel method known as Intelligent Phishing Detection System (IPDS). A hybrid data set of one million legit and phishing URLs, both from the PhishTank and Popular Crawl data set, and 10,000 photos captured personally on both phishing and legit Website are used to test a hybrid solution. The proposed IPDS produced an outstanding 93.28 percent grade accuracy.

Digwal et al. [111] Implemented phishing attack detection, focusing on a variety of comprehensive training algorithms CNN-LSTM model that can be used to determine if a website is genuine or phishing. Comprehensive training systems can predict phishing attacks in less than an hour and are well equipped to deal with emerging forms of phishing attacks. As a result, they are favored. The data collection includes millions of phishing and legal URLs, the precision is 98.99 percent, and the false positive rate is just 0.59 percent in our implementation.

Huang et al. [112] Proposed a creative phishing website detection method based on detecting a website's Uniform Resource Locator (URL), which has proven to be an accurate and efficient detection approach To be more precise, their novel capsule-based neural network consists primarily of many parallel branches in which one convolutional layer extracts shallow features from URLs and the following two capsule layers derive valid feature representations of URLs from the shallow features and distinguish the validity of URLs. The approach's final production is achieved by integrating the outputs of all branches. Extensive tests on a verified dataset obtained from the Internet show that their technique will compete with other state-of-the-art detection approaches while retaining a tolerable time overhead.

Wang et al. [113] Proposed the rapidly phishing method of detecting websites named PDRCNN based exclusively on website URLs. PDRCNN shall nor utilize any third-party providers, as has been previously the case, to recover the contents of its aim Web site. First, a two-way LSTM network was used for extracting the global characteristics of the built tensor and for each URL character to be given all the string details. After this, CNN used to dynamically evaluate the main function characters in phishing identification, grab the key components of the URL and compact the extracted attributes into a vector space of defined duration. PDRCNN produces a higher result than utilizing one of them when integrating two forms of networks. We built a dataset of almost 500,000 URLs collected from Alexa and PhishTank. Research has shown that 97% of the PDRCNN identification is accurate and that 99% of the AUC is even higher than the latest approach.

Chatterjee et al. [114] Presented a deep - learning enhancement model for the phishing website detection through examining the URLs given. The model itself is automatically adaptable to URL structure adjustments. This is an instance of the classical grouping issue as phishing websites are identified. To solve this classification challenge, a strengthening learning model has been built using a deep neural network. Our model was used on a balanced and named dataset of valid and malicious URLs in which 14 lexical characteristics were removed for training the model from the URLs in question. Output is calculated by accuracy, reminder, accuracy and F measurements at 86%, 88%, 90% and 87% respectively.

Opara et al. [115] Proposed broad, end-to-end automated phishing classification solution, in HTMLPhish, data-driven deep learning. Specifically, in order to think about semantic dependence in the texts of the HTML, the HTMLPhish receives a webpage HTML manual and hires Convolving Neural Networks (CNNs). Without comprehensive manual feature engineering, the CNN learns adequate feature representation in

HTML documents. In addition, our models will handle new functionality and ensure simple extrapolation to test data with the suggested method of concatenation of term and character integrations. The studies are carried out on more than 50,000 HTML documents that distribute phishing in a real-world manner to innocuous web pages that have an exact rate and a true positive rate of more than 93 percent.

Huang et al. [116] PhishingNet suggested a profound learning method for the timely identification of uniform phishing resources (URLs). The modules are specifically used to derive character spatial representations of URLs from the convolutional neural network (CNN) module. The module is used to derive word-level temporal URL featural representations from a focus-based hierarchic recurrent neural network (RNN). Then the feature representations were fused via a CNN three layer to create correct feature representations of URLs, on which a phishing URL classificatory is trained. Extensive tests with a validated data set obtained on the Internet show that feature representations derived immediately increase the capacity of our approach to generalize newly developed URLs which means that our approach is successful against advanced approaches.

Feng et al.[117] Suggested a modern mode of identification of phishing webpages. The first thing that this model considers as character sequences the URLs, HTML pages and DOMs

(Document Object Model) structures in the webpages is representation training technology, to acquire web pages representation automatically. Then, a hybrid deep learning network consisting of a convergent neural network and a long-term and short-term bidirectional memory network ends up by extracting local and international features from separate channels. Finally, the multi-channel performance is merged into the estimation of classification. The findings reveal that the total classification of the model has a greater impact than the current classic web page identification techniques, that the precision of the model is as high as 99.05% and that the false positive ratio is only 0.25%.

Yerima et al.[118] Proposed a profound learning method for detecting phishing facilities with high precision. In order to differentiate legitimate websites from phishing pages, this technique uses Convolutional Neural Networks (CNNs). Assess the models with data from 6.157 genuine websites and 4.898 photographs. Our CNN models have shown high efficiency in identifying unknown phishing locations, based on detailed experimental findings. The solution focused on CNN was also stronger tested on the same datasets, with a 98.2% phishing identification rate of 0.976 than conventional machine-learning classifications.

Al-Alyan et al.[119] Presented a workaround focused on the phishing identification URL only based on the CNN model. Instead of utilizing pre-determined features like URL duration the proposed CNN takes a URL as its input. More than two million URLs in a huge URL phishing detection (MUPD) dataset is obtained for training and assessment. Divide MUPD into data sets for preparation, evaluation and monitoring. This is accomplished by means of URL schemes (such as HTTP and HTTPS) deleted from the URL, and is achieved with approximately 96% precision on the test data collection. Compared to the current state-of-the-art model URL only on a published dataset, their proposed approach achieves greater precision. The trial findings recommend maintaining the CNN up-to-date in operation for improved results.

Aljofey et al.[120] Suggested a quick learning approach model that uses a phishing identification network (CNN) based on the website URL. The proposed model would not include the recovery or the usage of some third-party resources from the goal website material. It records metadata, sequential URL string patterns without having any previous phishing experience and then uses sequential patterns to quickly classify the individual URL. Comparisons are made with numerous feature sets such as handmade character embedding, TF-IDF character level and vectors at character level, between separate conventional machinery learning models and profound learning models for assessment. According to the experiments, 95% of accuracy and 98.58% of 98.6% of 95.46% of 95.22% of the benchmark data sets, which exceed the current Phishing URL models, is reached by the proposed model.

Ali et.al. [121] Proposal has been made for a DNN-based phishing website that's application of evolutionary feature-selection and biometric attribute weighting techniques According to genetic algorithm (GA) and other hybrid intelligent phishing methods, the most influential characteristics and optimum weights of a historic factors are used to increase the precision of website phishing recognition. The GA has developed a site algorithm that ranks websites by quality which weight and is used to train DNNs to identify phishing sites. This independent testing has shown that the hybrid intelligent phishing website detection techniques outperformed other methods on the training data used for generating the expected outcomes.

Somesha et al. [122] Suggested novel models that we might use the work that we've already done on (a) Deep Short-Term Memory Network (DSMN), (b) Convolutional Neural Network (CNN), and (c) for URL phishing detection only. With suggested techniques, LSTM gives an overall precision of 99.52%, and DNN is somewhat better at 99.57%, and CNN has a better overall than that. One single-third-party solution makes it more secure and faster to catch phishing attacks.

Sahingoz et.al. [123] Phishing attacks presented are a challenge because they are considered a semantic assault that focuses on user vulnerabilities and not the vulnerabilities of network. Most applications for ant phishing use blacklist/whitelist approaches in large measure; however, recent phishing attempts are not caught and there is a high rate of wrongdoing. In order to address this deficiency, it proposed machine-learn algorithms for training the device and to respond to anomalous requests through the URL of web sites, Artificial Neural Networks and Deep Neural Networks (DNNs). Used a dataset which included 37,175 phishing pages and 36,400 machine valid web pages. According to experimental findings, the suggested approaches with the rate of 92% and 96% through the usage of ANN and DNN approaches have the precision of identification of phishing websites.

TABLE 1. SUMMARY OF LITERATURE REVIEW RELATED TO WEB PHISHING DETECTION

RESEARCHER	SIGNIFICANT AIMS	ALGORITHM S	DATASET	ACCURACY AND PERFORMA
Yao et al. 2018 [105]	Proposed a relative detection method for the legitimacy of the URL in the two-dimensional code.	Faster R-CNN	FlickrLogos-32	Good performance in the small-sized object recognition
Su et al. 2020 [71]	designed a new detection system for phishing websites using LSTM Recurrent Neural Networks (RNN)	LSTM - CNN	Phishing websites and legal websites	Accuracy 99.1%
Yang et al. 2019 [106]	Suggested a multidimensional feature phishing detection deep learning (MFPD)	CNN - LSTM	Phishing URLs and Legitime URLs	Accuracy 98.99%
Singh et al. 2020 [107]	The phishing detection system has been implemented using the Deep neural network, (CNN) as a running model	CNN	phishing and legitimate webpage URLs	Accuracy 98.00%
Sindhu et al. 2020 [108]	Phishing detection used machine learning and deep Learning	Random Forest, SVM and Neural Network with backpropagation	UCI Machine learning repository	97.369%, 97.451%, 97.259%
Saha et al. 2020 [109]	Presented a data-driven framework for detecting phishing webpages using a deep learning approach.	Multilayer Perceptron (MLP)	Phishing, suspicious, and legitimate websites	Accuracy 95%
Adebowale et al. 2019[110]	Proposed a new approach called Intelligent Phishing Detection System (IPDS)	CNN - LSTM	Phish Tank and Common Crawl	Accuracy 93.28%
Digwal et al. 2020 [111]	Implemented detect phishing attacks	CNN - LSTM	phishing and legitimate URL	Accuracy 98.99%
Huang et al. 2019 [112]	Proposed an efficient and effective capsule-based neural network	CNN- Capsule NN	legitimate URLs crawled, and phishing URLs	Achieved excellent performance
Wang et al. 2019 [113]	Proposed a fast-phishing website detection approach called PDRCNN	LSTM - CNN	Alexa and PhishTank	Accuracy 97%
Chatterjee et al. 2019 [114]	presented a deep reinforcement learning-based model for detecting phishing website	Deep Reinforcement Learning	phishing and legitimate websites	Accuracy 90%
Opara et al. 2020 [115]	proposed HTMLPhish, deep learning-based data-driven end-to-end	CNN	HTML documents	Accuracy 93%
Huang et al. 2019 [116]	proposed Phishing Net, a deep learning-based approach for timely detection of phishing Uniform Resource Locators (URLs)	CNN - RNN	phishing and legitimate URLs	Accuracy 97%
Feng et al. 2020 [117]	a new phishing webpage detection model is proposed	CNN-BiLSTM	phishing and legitimate webpages	Accuracy 99.05%
Yerima et al. 2020 [118]	presented a deep learning-based approach to enable high accuracy detection of phishing sites	CNN	phishing websites and legitimate websites	Accuracy 98.2%
Al-Alyan et al. 2020 [119]	presented a URL-only phishing detection solution based on a convolutional neural network (CNN) model	CNN	massive URL phishing detection (MUPD)	Accuracy 96%
Aljofey et al. 2020 [120]	uses a character-level (CNN) for phishing detection based on the URL of the website	CNN	phishing sites and legitimate sites	Accuracy 95.02%
Ali et.al. 2019 [121]	proposed a hybrid phishing website approach feature selection and weighting phishing	DNN	UCI phishing websites	Accuracy 89.50%
Somesha et al. 2020 [122]	proposed novel phishing URL detection models using DNN, LSTM, and CNN	DNN, LSTM, and CNN	phishing sites, and legitimate sites	Accuracy's99.52% for DNN, 99.57% for LSTM, and 99.43% for CNN.
Sahingoz et.al. 2018 [123]	proposed machine learning-based algorithms, Artificial Neural Networks (ANNs) and Deep Neural Networks (DNNs)	ANN, and DNN	phishing, and legitimate URLs	Accuracy 92 % for ANN, and 96 % for DNN

IV. DISCUSSION

web phishing attempts may be found from a variety of angles, in terms of the references derived from email addresses or websites, However, phishing detection techniques suffer from poor detection precision and a high rate of false alarms in the majority of cases, particularly where novel phishing techniques are used.

The resulting phishing attacks cannot be detected successfully by blacklist and whitelist dependent methods alone, since they have problems with the upgrade mechanism. Computational and spatial structure problems are often present in visual similarity techniques. Lightweight phishing identification techniques based on similarities are required to maintain success in real-time. In recent online phishing

prevention techniques, deep learning methods have been commonly utilized. Deep learning algorithms, on the other hand, depend entirely on heuristics derived from web addresses or web pages as functions. Both anti-phishing strategies have the same goal: to reduce the effects of a web phishing assault. As a result, several recent studies have merged these various methods to achieve real-time efficiency, a high detection rate, Algorithms, and substantial goals. It is strongly recommended that the lightweight strategies of online phishing detection using hybrid methods should be the best option when dealing with existing phishing scams for those involved in researching these areas. A further potential of researchers is the use of deep learning in the identification of online phishing. Notably, Web page material is recognized as an open invitation for potential work in web phishing identification utilizing deep-learning methodologies.

After several studies, some techniques have been used to detect fishing attack. Each study has used a specific algorithm to protect its website. Most studies have used the LSTM and CNN algorithms. This is because these algorithms have the best accuracy and performance compared to all other algorithms. A number of studies that have used LSTM and CNN are [71, 106, 110, 124]. In conclusion, it can be concluded that the algorithms LSTM and CNN are the most effective algorithms for detecting fishing attack.

V. CONCLUSION

Phishing is a method of obtaining private information from users through email or a website. Because of the widespread use of the internet, nearly anything is now accessible online, whether it's searching for clothing, computer devices, or crockery, or paying for telephone, TV, and power bills. People are becoming conscious of the online method as an alternative to waiting in line for hours. As a result, phishing scams have a lot of space to be implemented. There is no single methodology that can identify all forms of phishing attacks since there has been too much research done in this field. Phishing attacks are becoming more sophisticated as technology advances. This allows us to develop a reliable phishing classifier. In this review study, we performed a detailed literature survey about phishing website detection based on deep learning. According to this, we can say Conventional Neural Network -Long Short-Term Memory networks CNN-LSTM in deep learning approach is best suitable than other.

REFERENCES

[1] R. R. Zebari, S. R. Zeebaree, and K. Jacksi, "Impact analysis of HTTP and SYN flood DDoS attacks on apache 2 and IIS 10.0 Web servers," in 2018 International Conference on Advanced Science and Engineering (ICOASE), 2018, pp. 156-161.

[2] S. R. Zeebaree, K. Jacksi, and R. R. Zebari, "Impact analysis of SYN flood DDoS attack on HAProxy and NLB cluster-based web servers," Indones. J. Electr. Eng. Comput. Sci, vol. 19, pp. 510-517, 2020.

[3] O. H. Jader, S. Zeebaree, and R. R. Zebari, "A state of art survey for web server performance measurement and load balancing mechanisms," International Journal of Scientific & Technology Research, vol. 8, pp. 535-543, 2019.

[4] P. Y. Abdullah, S. Zeebaree, K. Jacksi, and R. R. Zeabri, "An hrm system for small and medium enterprises (sme) s based on cloud computing technology," International Journal of Research-GRANTHAALAYAH, vol. 8, pp. 56-64, 2020.

[5] W. H. Lim, W. F. Liew, C. Y. Lum, and S. F. Lee, "Phishing security: Attack, detection, and prevention mechanisms," ed, 2020.

[6] R. R. Zebari, S. Zeebaree, K. Jacksi, and H. M. Shukur, "E-business requirements for flexibility and implementation enterprise system: A review," International Journal of Scientific & Technology Research, vol. 8, pp. 655-660, 2019.

[7] S. Zeebaree, L. M. Haji, I. Rashid, R. R. Zebari, O. M. Ahmed, K. Jacksi, et al., "Multicomputer multicore system influence on maximum multi-processes execution time," TEST Engineering & Management, vol. 83, pp. 14921-14931, 2020.

[8] S. Zeebaree, R. R. Zebari, K. Jacksi, and D. A. Hasan, "Security Approaches For Integrated Enterprise Systems Performance: A Review," Int. J. Sci. Technol. Res, vol. 8, 2019.

[9] S. Gupta, A. Singhal, and A. Kapoor, "A literature survey on social engineering attacks: Phishing attack," in 2016 international conference on computing, communication and automation (ICCCA), 2016, pp. 537-540.

[10] H. Dino, M. B. Abdulrazzaq, S. Zeebaree, A. B. Sallow, R. R. Zebari, H. M. Shukur, et al., "Facial expression recognition based on hybrid feature extraction techniques with different classifiers," TEST Engineering & Management, vol. 83, pp. 22319-22329, 2020.

[11] K. Jacksi, R. K. Ibrahim, S. R. Zeebaree, R. R. Zebari, and M. A. Sadeeq, "Clustering documents based on semantic similarity using HAC and K-mean algorithms," in 2020 International Conference on Advanced Science and Engineering (ICOASE), 2020, pp. 205-210.

[12] S. R. Zeebaree, H. M. Shukur, L. M. Haji, R. R. Zebari, K. Jacksi, and S. M. Abas, "Characteristics and analysis of hadoop distributed systems," Technology Reports of Kansai University, vol. 62, pp. 1555-1564, 2020.

[13] M. B. Abdulrazaq, M. R. Mahmood, S. R. Zeebaree, M. H. Abdulwahab, R. R. Zebari, and A. B. Sallow, "An analytical appraisal for supervised classifiers' performance on facial expression recognition based on relief-F feature selection," in Journal of Physics: Conference Series, 2021, p. 012055.

[14] H. Malallah, S. Zeebaree, R. R. Zebari, M. Sadeeq, Z. S. Ageed, I. M. Ibrahim, et al., "A comprehensive study of kernel (issues and concepts) in different operating systems," Asian Journal of Research in Computer Science, vol. 8, pp. 16-31, 2021.

[15] D. A. Hasan, S. R. Zeebaree, M. A. Sadeeq, H. M. Shukur, R. R. Zebari, and A. H. Alkhayyat, "Machine Learning-based Diabetic Retinopathy Early Detection and Classification Systems-A Survey," in 2021 1st Babylon International Conference on Information Technology and Science (BICITS), 2021, pp. 16-21.

[16] A. B. Sallow, S. R. Zeebaree, R. R. Zebari, M. R. Mahmood, M. B. Abdulrazzaq, and M. A. Sadeeq, "Vaccine tracker/SMS reminder system: design and implementation," ISSN (Online), pp. 2581-6187, 2020.

[17] H. Shukur, S. Zeebaree, R. Zebari, O. Ahmed, L. Haji, and D. Abdulqader, "Cache coherence protocols in distributed systems," Journal of Applied Science and Technology Trends, vol. 1, pp. 92-97, 2020.

[18] M. R. Mahmood, M. B. Abdulrazzaq, S. Zeebaree, A. K. Ibrahim, R. R. Zebari, and H. I. Dino, "Classification techniques' performance evaluation for facial expression recognition," Indonesian Journal of Electrical Engineering and Computer Science, vol. 21, pp. 176-1184, 2021.

[19] B. R. Ibrahim, F. M. Khalifa, S. R. Zeebaree, N. A. Othman, A. Alkhayyat, R. R. Zebari, et al., "Embedded system for eye blink detection using machine learning technique," in 2021 1st Babylon International Conference on Information Technology and Science (BICITS), 2021, pp. 58-62.

[20] Z. A. Younis, A. M. Abdulazeez, S. R. Zeebaree, R. R. Zebari, and D. Q. Zeebaree, "Mobile Ad Hoc Network in Disaster Area Network Scenario: A Review on Routing Protocols," International Journal of Online & Biomedical Engineering, vol. 17, 2021.

[21] C. Singh, "Phishing website detection based on machine learning: A survey," in 2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS), 2020, pp. 398-404.

[22] H. I. Dino, S. Zeebaree, O. M. Ahmad, H. M. Shukur, R. R. Zebari, and L. M. Haji, "Impact of load sharing on performance of distributed systems computations," International Journal of Multidisciplinary Research and Publications (IJMRAP), vol. 3, pp. 30-37, 2020.

[23] H. M. Shukur, S. R. Zeebaree, R. R. Zebari, B. K. Hussan, O. H. Jader, and L. M. Haji, "Design and implementation of electronic enterprise university human resource management system," in Journal of Physics: Conference Series, 2021, p. 012058.

[24] R. R. Zebari, S. R. Zeebaree, A. B. Sallow, H. M. Shukur, O. M. Ahmad, and K. Jacksi, "Distributed denial of service attack mitigation using high availability proxy and network load balancing," in 2020 International Conference on Advanced Science and Engineering (ICOASE), 2020, pp. 174-179.

[25] R. S. Rao and A. R. Pais, "Detection of phishing websites using an efficient feature-based machine learning framework," Neural Computing and Applications, vol. 31, pp. 3851-3873, 2019.

[26] H. I. Dino, S. Zeebaree, A. A. Salih, R. R. Zebari, Z. S. Ageed, H. M. Shukur, et al., "Impact of Process Execution and Physical Memory-Spaces on OS Performance," Technology Reports of Kansai University, vol. 62, pp. 2391-2401, 2020.

[27] L. Haji, R. Zebari, S. Zeebaree, W. Abduallah, H. Shukur, and O. Ahmed, "GPUs impact on parallel shared memory systems performance," Int. J. Psychosoc. Rehabil, vol. 24, pp. 8030-8038, 2019.

[28] K. H. Sharif, S. R. Zeebaree, L. M. Haji, and R. R. Zebari, "Performance measurement of processes and threads controlling, tracking and monitoring based on shared-memory parallel processing approach," in 2020 3rd International Conference on Engineering Technology and its Applications (IICETA), 2020, pp. 62-67.

[29] G. D. L. T. Parra, P. Rad, K.-K. R. Choo, and N. Beebe, "Detecting Internet of Things attacks using distributed deep learning," Journal of Network and Computer Applications, vol. 163, p. 102662, 2020.

[30] L. Haji, R. Zebari, S. Zeebaree, M. WAFAA, H. Shukur, and O. Alzakholi, "GPUs Impact on Parallel Shared Memory Systems Performance–International Journal of Psychosocial Rehabilitation," ed: May, 2020.

[31] G. M. Zebari, S. Zeebaree, M. M. Sadeeq, and R. Zebari, "Predicting Football Outcomes by Using Poisson Model: Applied to Spanish Primera División," Journal of Applied Science and Technology Trends, vol. 2, pp. 105-112, 2021.

[32] P. RAJKUMAR and S. k. VENKATESH, "A Survey on Data Mining Techniques for Website Phishing Detection," International Journal of Pure and Applied Mathematics, vol. 119, pp. 2127-2133, 2018.

[33] K. P. M. Kumar, J. Mahilraj, D. Swathi, R. Rajavarman, S. R. Zeebaree, R. R. Zebari, et al., "Privacy Preserving Blockchain with Optimal Deep Learning Model for Smart Cities," CMC-COMPUTERS MATERIALS & CONTINUA, vol. 73, pp. 5299-5314, 2022.

[34] Z. N. Rashid, S. R. Zeebaree, R. R. Zebari, S. H. Ahmed, H. M. Shukur, and A. Alkhayyat, "Distributed and Parallel Computing System Using Single-Client Multi-Hash Multi-Server Multi-Thread," in 2021 1st Babylon International Conference on Information Technology and Science (BICITS), 2021, pp. 222-227.

[35] R. R. Zebari, S. R. Zeebaree, Z. N. Rashid, H. M. Shukur, A. Alkhayyat, and M. A. Sadeeq, "A Review on Automation Artificial Neural Networks based on Evolutionary Algorithms," in 2021 14th International Conference on Developments in eSystems Engineering (DeSE), 2021, pp. 235-240.

[36] P. Kalaharsha and B. Mehtre, "Detecting Phishing Sites--An Overview," arXiv preprint arXiv:2103.12739, 2021.

[37] Z. N. Rashid, S. R. Zeebaree, M. A. Sadeeq, R. R. Zebari, H. M. Shukur, and A. Alkhayyat, "Cloud-based Parallel Computing System Via Single-Client Multi-Hash Single-Server Multi-Thread," in 2021 International Conference on Advance of Sustainable Engineering and its Application (ICASEA), 2021, pp. 59-64.

[38] O. H. Jader, S. R. Zeebaree, R. R. Zebari, H. M. Shukur, Z. N. Rashid, M. A. Sadeeq, et al., "Ultra-Dense Request Impact on Cluster-Based Web Server Performance," in 2021 4th International Iraqi Conference on Engineering Technology and Their Applications (IICETA), 2021, pp. 252-257.

[39] M. Khonji, Y. Iraqi, and A. Jones, "Phishing detection: a literature survey," IEEE Communications Surveys & Tutorials, vol. 15, pp. 2091-2121, 2013.

[40] A. S. Abdulraheem, A. I. Abdulla, and S. M. Mohammed, "Enterprise resource planning systems and challenges," Technology Reports of Kansai University, vol. 62, pp. 1885-1894, 2020.

[41] B. W. SALIM and S. R. ZEEBAREE, "ISOLATED AND CONTINUOUS HAND GESTURE RECOGNITION BASED ON DEEP LEARNING: A REVIEW."

[42] F. Abedi, S. R. Zeebaree, Z. S. Ageed, H. M. Ghanimi, A. Alkhayyat, M. A. Sadeeq, et al., "Severity Based Light-Weight Encryption Model for Secure Medical Information System."

[43] V. D. Majety, N. Sharmili, C. R. Pattanaik, E. L. Lydia, S. R. Zeebaree, S. N. Mahmood, et al., "Ensemble of Handcrafted and Deep Learning Model for Histopathological Image Classification," CMC-COMPUTERS MATERIALS & CONTINUA, vol. 73, pp. 4393-4406, 2022.

[44] A. M. Abed, Z. N. Rashid, F. Abedi, S. R. Zeebaree, M. A. Sahib, A. J. a. Mohamad Jawad, et al., "Trajectory tracking of differential drive mobile robots using fractional-order proportional-integral-derivative controller design tuned by an enhanced fruit fly optimization," Measurement and Control, vol. 55, pp. 209-226, 2022.

[45] H. B. Abdalla, A. M. Ahmed, S. R. Zeebaree, A. Alkhayyat, and B. Ihnaini, "Rider weed deep residual network-based incremental model for text classification using multidimensional features and MapReduce," PeerJ Computer Science, vol. 8, p. e937, 2022.

[46] A. S. Aljuboury, S. R. Zeebaree, F. Abedi, Z. S. Hashim, R. Q. Malik, I. K. Ibraheem, et al., "A New Nonlinear Controller Design for a TCP/AQM Network Based on Modified Active Disturbance Rejection Control," Complexity, vol. 2022, 2022.

[47] S. Chavhan, S. R. Zeebaree, A. Alkhayyat, and S. Kumar, "Design of Space Efficient Electric Vehicle Charging Infrastructure Integration Impact on Power Grid Network," Mathematics, vol. 10, p. 3450, 2022.

[48] Y. S. Jghef, M. J. M. Jasim, H. M. Ghanimi, A. D. Algarni, N. F. Soliman, W. El-Shafai, et al., "Bio-Inspired Dynamic Trust and Congestion-Aware Zone-Based Secured Internet of Drone Things (SIoDT)," Drones, vol. 6, p. 337, 2022.

[49] G. M. O. Zebari, K. Faraj, and S. Zeebaree, "Hand writing code-php or wire shark ready application over tier architecture with windows servers operating systems or linux server operating systems," International Journal of Computer Sciences and Engineering, vol. 4, pp. 142-149, 2016.

[50] R. Ibrahim, S. R. Zeebaree, and K. Jacksi, "Semantic Similarity for Document Clustering using TFIDF and K-mean," Master's Thesis, 2020.

[51] B. W. Salim and S. R. Zeebaree, "Design & Analyses of a Novel Real Time Kurdish Sign Language for Kurdish Text and Sound Translation System," in 2020 IEEE International Conference on Problems of Infocommunications. Science and Technology (PIC S&T), 2020, pp. 348-352.

[52] S. M. Saleem, S. R. Zeebaree, and M. B. Abdulrazzaq, "Real-life dynamic facial expression recognition: A review," in Journal of Physics: Conference Series, 2021, p. 012010.

[53] R. K. Ibrahim, S. R. Zeebaree, K. Jacksi, M. A. Sadeeq, H. M. Shukur, and A. Alkhayyat, "Clustering document based semantic similarity system using TFIDF and k-mean," in 2021 International Conference on Advanced Computer Applications (ACA), 2021, pp. 28-33.

[54] S. Zebari, "A new approach for process monitoring," Polytechnic Journal, Technical Education-Erbil, 2011.

[55] K. B. Obaid, S. Zeebaree, and O. M. Ahmed, "Deep learning models based on image classification: a review," International Journal of Science and Business, vol. 4, pp. 75-81, 2020.

[56] S. Zebari and N. O. Yaseen, "Effects of parallel processing implementation on balanced load-division depending on distributed memory systems," J. Univ. Anbar Pure Sci, vol. 5, pp. 50-56, 2011.

[57] K. Jacksi, N. Dimililer, and S. R. Zeebaree, "A survey of exploratory search systems based on LOD resources," 2015.

[58] D. A. Zebari, H. Haron, S. R. Zeebaree, and D. Q. Zeebaree, "Multi-level of DNA encryption technique based on DNA arithmetic and biological operations," in 2018 International Conference on Advanced Science and Engineering (ICOASE), 2018, pp. 312-317.

[59] H. M. Yasin, S. R. Zeebaree, and I. M. Zebari, "Arduino based automatic irrigation system: Monitoring and SMS controlling," in 2019 4th Scientific International Conference Najaf (SICN), 2019, pp. 109-114.

[60] A. Zeebaree, A. Adel, K. Jacksi, and A. Selamat, "Designing an ontology of E-learning system for duhok polytechnic university using protégé OWL tool," J Adv Res Dyn Control Syst Vol, vol. 11, pp. 24-37, 2019.

[61] L. M. Haji, O. M. Ahmad, S. Zeebaree, H. I. Dino, R. R. Zebari, and H. M. Shukur, "Impact of cloud computing and internet of things on the future internet," Technology Reports of Kansai University, vol. 62, pp. 2179-2190, 2020.

[62] A. AL-Zebari, S. Zeebaree, K. Jacksi, and A. Selamat, "ELMS–DPU ontology visualization with Protégé VOWL and Web VOWL," Journal of Advanced Research in Dynamic and Control Systems, vol. 11, pp. 478-85, 2019.

[63] L. M. Abdulrahman, S. Zeebaree, S. F. Kak, M. Sadeeq, A. Adel, B. W. Salim, et al., "A state of art for smart gateways issues and modification," Asian Journal of Research in Computer Science, vol. 63, pp. 1-13, 2021.

[64] H. M. Yasin, S. Zeebaree, M. Sadeeq, S. Y. Ameen, I. M. Ibrahim, R. R. Zebari, et al., "IoT and ICT based smart water management, monitoring and controlling system: A review," Asian Journal of Research in Computer Science, vol. 8, pp. 42-56, 2021.

[65] K. Jacksi, S. R. Zeebaree, and N. Dimililer, "LOD Explorer: Presenting the Web of Data," Int. J. Adv. Comput. Sci. Appl. IJACSA, vol. 9, pp. 1-7, 2018.

[66] F. Q. Kareem, S. Zeebaree, H. I. Dino, M. Sadeeq, Z. N. Rashid, D. A. Hasan, et al., "A survey of optical fiber communications: challenges and processing time influences," Asian Journal of Research in Computer Science, pp. 48-58, 2021.

[67] A. B. Sallow, M. Sadeeq, R. R. Zebari, M. B. Abdulrazzaq, M. R. Mahmood, H. M. Shukur, et al., "An investigation for mobile malware behavioral and detection techniques based on android platform," IOSR Journal of Computer Engineering (IOSR-JCE), vol. 22, pp. 14-20, 2020.

[68] B. T. Jijo, S. Zeebaree, R. R. Zebari, M. Sadeeq, A. B. Sallow, S. Mohsin, et al., "A comprehensive survey of 5G mm-wave technology design challenges," Asian Journal of Research in Computer Science, vol. 8, pp. 1-20, 2021.

[69] L. M. Haji, S. Zeebaree, O. M. Ahmed, A. B. Sallow, K. Jacksi, and R. R. Zeabri, "Dynamic resource allocation for distributed systems and cloud computing," TEST Engineering & Management, vol. 83, pp. 22417-22426, 2020.

[70] H. Shukur, S. R. Zeebaree, A. J. Ahmed, R. R. Zebari, O. Ahmed, B. S. A. Tahir, et al., "A state of art survey for concurrent computation and clustering of parallel computing for distributed systems," Journal of Applied Science and Technology Trends, vol. 1, pp. 148-154, 2020.

[71] H. Shukur, S. Zeebaree, R. Zebari, D. Zeebaree, O. Ahmed, and A. Salih, "Cloud computing virtualization of resources allocation for distributed systems," Journal of Applied Science and Technology Trends, vol. 1, pp. 98-105, 2020.

[72] A. A. Salih, S. Zeebaree, A. S. Abdulraheem, R. R. Zebari, M. Sadeeq, and O. M. Ahmed, "Evolution of mobile wireless communication to 5G revolution," Technology Reports of Kansai University, vol. 62, pp. 2139-2151, 2020.

[73] O. Alzakholi, H. Shukur, R. Zebari, S. Abas, and M. Sadeeq, "Comparison among cloud technologies and cloud performance," Journal of Applied Science and Technology Trends, vol. 1, pp. 40-47, 2020.

[74] M. M. Sadeeq, N. M. Abdulkareem, S. R. Zeebaree, D. M. Ahmed, A. S. Sami, and R. R. Zebari, "IoT and Cloud computing issues, challenges and opportunities: A review," Qubahan Academic Journal, vol. 1, pp. 1-7, 2021.

[75] A. A. Yazdeen, S. R. Zeebaree, M. M. Sadeeq, S. F. Kak, O. M. Ahmed, and R. R. Zebari, "FPGA implementations for data encryption and decryption via concurrent and parallel computation: A review," Qubahan Academic Journal, vol. 1, pp. 8-16, 2021.

[76] Z. N. Rashid, S. R. Zebari, K. H. Sharif, and K. Jacksi, "Distributed cloud computing and distributed parallel computing: A review," in 2018 International Conference on Advanced Science and Engineering (ICOASE), 2018, pp. 167-172.

[77] S. M. S. A. Abdullah, S. Y. A. Ameen, M. A. Sadeeq, and S. Zeebaree, "Multimodal emotion recognition using deep learning," Journal of Applied Science and Technology Trends, vol. 2, pp. 52-58, 2021.

[78] G. A. Akerlof and R. J. Shiller, "Phishing for phools," in Phishing for Phools, ed: Princeton University Press, 2015.

[79] A. Aleroud and L. Zhou, "Phishing environments, techniques, and countermeasures: A survey," Computers & Security, vol. 68, pp. 160-196, 2017.

[80] I. Fette, N. Sadeh, and A. Tomasic, "Learning to detect phishing emails," in Proceedings of the 16th international conference on World Wide Web, 2007, pp. 649-656.

[81] S. Sheng, B. Wardman, G. Warner, L. Cranor, J. Hong, and C. Zhang, "An empirical analysis of phishing blacklists," 2009.

[82] B. Parno, C. Kuo, and A. Perrig, "Phoolproof phishing prevention," in Financial Cryptography, 2006, pp. 1-19.

[83] Y. Zhang, S. Egelman, L. Cranor, and J. Hong, "Phinding phish: Evaluating anti-phishing tools," 2007.

[84] K. L. Chiew, K. S. C. Yong, and C. L. Tan, "A survey of phishing attacks: Their types, vectors and technical approaches," Expert Systems with Applications, vol. 106, pp. 1-20, 2018.

[85] Z. Alkhalil, C. Hewage, L. Nawaf, and I. Khan, "Phishing attacks: A recent comprehensive study and a new anatomy," Frontiers in Computer Science, vol. 3, p. 563060, 2021.

[86] K. Jansson and R. von Solms, "Phishing for phishing awareness," Behaviour & information technology, vol. 32, pp. 584-593, 2013.

[87] J. Milletary and C. C. Center, "Technical trends in phishing attacks," Retrieved December, vol. 1, p. 3.3, 2005.

[88] A. Shankar, R. Shetty, and B. Nath, "A review on phishing attacks," International Journal of Applied Engineering Research, vol. 14, pp. 2171-2175, 2019.

[89] M. Dadkhah, M. D. Jazi, M. S. Mobarakeh, S. Shamshirband, X. Wang, and S. Raste, "An overview of phishing attacks and their detection techniques," International Journal of Internet Protocol Technology, vol. 9, pp. 187-195, 2016.

[90] V. Bhavsar, A. Kadlak, and S. Sharma, "Study on phishing attacks," Int. J. Comput. Appl, vol. 182, pp. 27-29, 2018.

[91] K. Gajera, M. Jangid, P. Mehta, and J. Mittal, "A novel approach to detect phishing attack using artificial neural networks combined with pharming detection," in 2019 3rd International conference on Electronics, Communication and Aerospace Technology (ICECA), 2019, pp. 196-200.

[92] H. Zuhair, A. Selamat, and M. Salleh, "Feature selection for phishing detection: a review of research," International Journal of Intelligent Systems Technologies and Applications, vol. 15, pp. 147-162, 2016.

[93] Z. Futai, G. Yuxiang, P. Bei, P. Li, and L. Linsen, "Web phishing detection based on graph mining," in 2016 2nd IEEE international conference on computer and communications (ICCC), 2016, pp. 1061-1066.

[94] M. Vijayalakshmi, S. M. Shalinie, and M. H. Yang, "Web phishing detection techniques: a survey on the state-of-the-art, taxonomy and future directions," IET Networks, vol. 9, pp. 235-246, 2020.

[95] M. Baykara and Z. Z. Gürel, "Detection of phishing attacks," in 2018 6th International Symposium on Digital Forensic and Security (ISDFS), 2018, pp. 1-5.

[96] T. Chin, K. Xiong, and C. Hu, "Phishlimiter: A phishing detection and mitigation approach using software-defined networking," IEEE Access, vol. 6, pp. 42516-42531, 2018.

[97] A. K. Jain and B. B. Gupta, "Phishing detection: analysis of visual similarity based approaches," Security and Communication Networks, vol. 2017, 2017.

[98] P. Pujara and M. Chaudhari, "Phishing website detection using machine learning: a review," International Journal of Scientific Research in Computer Science, Engineering and Information Technology, vol. 3, pp. 395-399, 2018.

[99] P. Yi, Y. Guan, F. Zou, Y. Yao, W. Wang, and T. Zhu, "Web phishing detection using a deep learning framework," Wireless Communications and Mobile Computing, vol. 2018, 2018.

[100] N. Zhang and Y. Yuan, "Phishing detection using neural network," CS229 lecture notes, 2012.

[101] X. Xiao, D. Zhang, G. Hu, Y. Jiang, and S. Xia, "CNN–MHSA: A Convolutional Neural Network and multi-head self-attention combined approach for detecting phishing websites," Neural Networks, vol. 125, pp. 303-312, 2020.

[102] S. Selvaganapathy, M. Nivaashini, and H. Natarajan, "Deep belief network based detection and categorization of malicious URLs," Information Security Journal: A Global Perspective, vol. 27, pp. 145-161, 2018.

[103] K. H. Park, E. Lee, and H. K. Kim, "Show me your account: detecting MMORPG game bot leveraging financial analysis with LSTM," in International Workshop on Information Security Applications, 2019, pp. 3-13.

[104] S. Priya, S. Selvakumar, and R. L. Velusamy, "Detection of phishing attacks using radial basis function network trained for categorical attributes," in 2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT), 2020, pp. 1-6.

[105] W. Yao, Y. Ding, and X. Li, "Deep learning for phishing detection," in 2018 IEEE Intl Conf on Parallel & Distributed Processing with Applications, Ubiquitous Computing & Communications, Big Data & Cloud Computing, Social Computing & Networking, Sustainable Computing & Communications (ISPA/IUCC/BDCloud/SocialCom/SustainCom), 2018, pp. 645-650.

[106] P. Yang, G. Zhao, and P. Zeng, "Phishing website detection based on multidimensional features driven by deep learning," IEEE access, vol. 7, pp. 15196-15209, 2019.

[107] S. Singh, M. Singh, and R. Pandey, "Phishing Detection from URLs Using Deep Learning Approach," in 2020 5th International Conference on Computing, Communication and Security (ICCCS), 2020, pp. 1-4.

[108] S. Sindhu, S. P. Patil, A. Sreevalsan, F. Rahman, and M. S. AN, "Phishing detection using random forest, SVM and neural network with backpropagation," in 2020 International Conference on Smart Technologies in Computing, Electrical and Electronics (ICSTCEE), 2020, pp. 391-394.

[109] I. Saha, D. Sarma, R. J. Chakma, M. N. Alam, A. Sultana, and S. Hossain, "Phishing attacks detection using deep learning approach," in 2020 Third International Conference on Smart Systems and Inventive Technology (ICSSIT), 2020, pp. 1180-1185.

[110] M. A. Adebowale, K. T. Lwin, and M. A. Hossain, "Deep learning with convolutional neural network and long short-term memory for phishing detection," in 2019 13th International Conference on Software, Knowledge, Information Management and Applications (SKIMA), 2019, pp. 1-8.

[111] H. N. Digwal and N. Kavya, "Detection of Phishing Website Based on Deep Learning," International Journal of Research in Engineering, Science and Management, vol. 3, pp. 331-336, 2020.

[112] Y. Huang, J. Qin, and W. Wen, "Phishing URL Detection Via Capsule-Based Neural Network," in 2019 IEEE 13th International Conference on Anti-counterfeiting, Security, and Identification (ASID), 2019, pp. 22-26.

[113] W. Wang, F. Zhang, X. Luo, and S. Zhang, "Pdrcnn: precise phishing detection with recurrent convolutional neural networks," Security and Communication Networks, vol. 2019, 2019.

[114] M. Chatterjee and A.-S. Namin, "Detecting phishing websites through deep reinforcement learning," in 2019 IEEE 43rd Annual Computer Software and Applications Conference (COMPSAC), 2019, pp. 227-232.

[115] C. Opara, B. Wei, and Y. Chen, "HTMLPhish: enabling phishing web page detection by applying deep learning techniques on HTML analysis," in 2020 International Joint Conference on Neural Networks (IJCNN), 2020, pp. 1-8.

[116] Y. Huang, Q. Yang, J. Qin, and W. Wen, "Phishing URL detection via CNN and attention-based hierarchical RNN," in 2019 18th IEEE International Conference On Trust, Security And Privacy In Computing And Communications/13th IEEE International Conference On Big Data Science And Engineering (TrustCom/BigDataSE), 2019, pp. 112-119.

[117] J. Feng, L. Zou, O. Ye, and J. Han, "Web2vec: Phishing webpage detection method based on multidimensional features driven by deep learning," IEEE Access, vol. 8, pp. 221214-221224, 2020.

[118] S. Y. Yerima and M. K. Alzaylaee, "High accuracy phishing detection based on convolutional neural networks," in 2020 3rd International Conference on Computer Applications & Information Security (ICCAIS), 2020, pp. 1-6.

[119] A. Al-Alyan and S. Al-Ahmadi, "Robust URL phishing detection based on deep learning," KSII Transactions on Internet and Information Systems (TIIS), vol. 14, pp. 2752-2768, 2020.

[120] 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, vol. 9, p. 1514, 2020.

[121] W. Ali and A. A. Ahmed, "Hybrid intelligent phishing website prediction using deep neural networks with genetic algorithm-based feature selection and weighting," IET Information Security, vol. 13, pp. 659-669, 2019.

[122] M. Somesha, A. R. Pais, R. S. Rao, and V. S. Rathour, "Efficient deep learning techniques for the detection of phishing websites," Sādhanā, vol. 45, pp. 1-18, 2020.

[123] O. K. Sahingoz, S. I. Baykal, and D. Bulut, "Phishing detection from urls by using neural networks," Computer Science & Information Technology (CS & IT), pp. 41-54, 2018.

[124] H. Digwal and N. Kavya, "Detection of phishing website based on deep learning," Int. J. Res. Eng. Sci. Manag, vol. 3, pp. 331-336, 2020.