Review of Computer Engineering Research
2022 Vol. 9, No. 1, pp. 13-29.
ISSN(e): 2410-9142
ISSN(p): 2412-4281
DOI: 10.18488/76.v9i1.2983

MACHINE LEARNING AND DEEP LEARNING BASED PHISHING WEBSITES DETECTION: THE CURRENT GAPS AND NEXT DIRECTIONS

Kibreab Adane[1+]
Berhanu Beyene[2]

[1]*Faculty of Computing and Software Engineering, Arba Minch University, Ethiopia.*
Email: *kibreab.adane@amu.edu.et* *Tel +251-(0)924408342*
[2]*Department of Computer Science, Ethiopian Civil Service University, Ethiopia.*
Email: *berhane.beyene@ethernet.edu.et* *Tel: +251-(0)939666347*

(+ Corresponding author)

Article History
Received: 14 February 2022
Revised: 28 March 2022
Accepted: 18 April 2022
Published: 6 May 2022

Keywords
Deep learning
Machine learning
Phishing website attack
Phishing website detection
Anti-phishing website
Legitimate website
Phishing website datasets
Phishing website features.

ABSTRACT

There are many phishing websites detection techniques in literature, namely white-listing, black-listing, visual-similarity, heuristic-based, and others. However, detecting zero-hour or newly designed phishing website attacks is an inherent property of machine learning and deep learning techniques. By considering a promising solution of machine learning and deep learning techniques, researchers have made a great deal of effort to tackle the this problem, which persists due to attackers constantly devising novel strategies to exploit vulnerability or gaps in existing anti-phishing measures. In this study, an extensive effort has been made to rigorously review recent studies focusing on Machine Learning and Deep Learning Based Phishing Websites Detection to excavate the root cause of the aforementioned problems and offer suitable solutions. The study followed the significant criterion to search, download, and screen relevant studies, then to evaluate criterion-based selected studies. The findings show that significant research gaps are available in the rigorously reviewed studies. These gaps are mainly related to imbalanced dataset usage, improper selection of dataset source(s), the unjustified reason for using specific train-test dataset split ratio, scientific disputes on website features inclusion and exclusion, lack of universal consensus on phishing website lifespans and on what is defining a small dataset size, and run-time analysis issues. The study clearly presented a summary of the comparative analysis performed on each reviewed research work so that future researchers could use it as a structured guideline to develop a novel solution for anti-phishing website attacks.

Contribution/Originality: This study took significant steps to find, screen out, and evaluate 30 criterion-based selected recent studies on Machine Learning and Deep Learning Based Phishing Websites Detection to extract core research gaps and propose appropriate solutions that could assist future researchers as structured guidelines to develop novel anti-phishing website attacks.

1. INTRODUCTION

To compete with the rest of the world, every country is relying on the internet for cashless transactions, online commerce, paperless tickets, and other productivity methods. Phishing, on the other hand, is becoming a modern-day threat and an obstacle to this progress, and people no longer believe that the internet is trustworthy [1]. Phishing website attacks are a web-based criminal act, in which phishers create a replica of a legitimate website in order to harvest confidential data from online users by taking advantage of human behavior and by exploiting the existing technical defense [2]. From cybersecurity experts' viewpoints, the website is legitimate when a URL uses the HTTPS encryption protocol. However, to mimic authentic websites, 74% of all phishing websites now use HTTPS and 78% of them use SSL protection [3, 4]. The attacker also uses a redirector to avoid detection [5].

Nearly 50 to 80% of phishing websites were blacklisted following some form of financial loss [6]. Despite the fact that blacklisting fails to detect newly designed phishing website attacks, existing Internet applications such as Chrome, Internet Explorer, Safari, Firefox, Gmail, Google Search, and several web browser extensions use blacklisting to detect phishing websites and display warnings when online users visit them [7, 8]. APWG phishing activity trend reports for the 1st to 3rd quarters of 2021 shows an increase in the number of unique phishing website attacks, APWG [4]; APWG [9]; APWG [10] as shown in Figure 1. In July 2021 alone, APWG detected 260,642 distinct phishing websites (attacks), making it the worst monthly phishing website attack in APWG reporting history [11] as shown in Figure 2.

Researchers have made a great deal of effort to address the problem of phishing website attacks using machine learning and deep learning approaches. However, the problems persist due to attackers continually devising novel strategies to exploit the existing anti-phishing measures. Because the security of online users' and organizations' information cannot be overlooked, and the number of unique website attacks continues to rise at an alarming rate, this study proposes to investigate the key gaps in recent research works focusing on machine learning and deep learning-based phishing website detection so that future researchers could use the identified research gaps and suggested solutions to develop further anti-phishing solutions. The research gaps analysis methods used in this study mainly focus on the best-performed phishing website detection model, website feature selection techniques, dataset source, dataset size, phish-legitimate dataset ratios, percentage of Dataset Train-Test split ratio, the number of website features used, and run-time analysis issues.

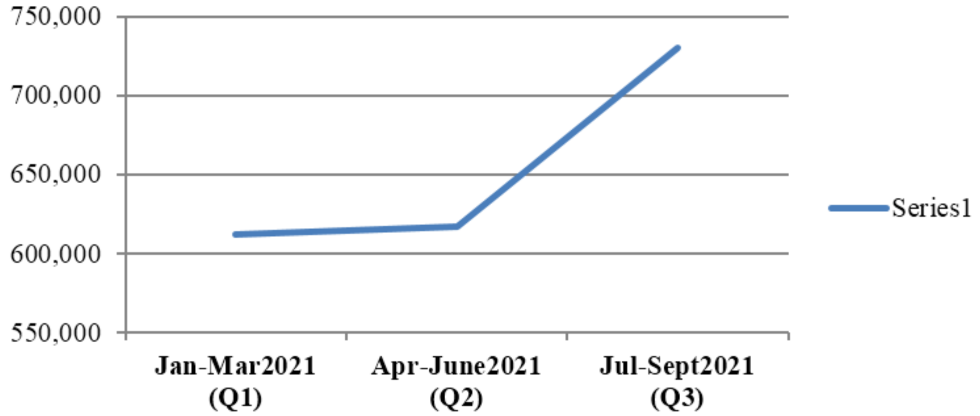

Figure 1. The 2021 quarterly unique website attacks report by APWG [4]; APWG [9]; APWG [10].

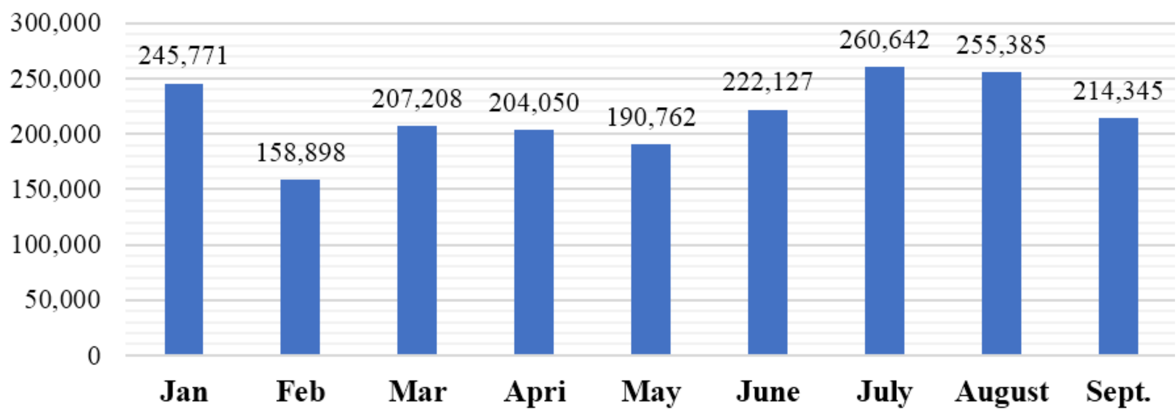

Figure. 2. The 2021 unique website attacks monthly report by APWG [4]; APWG [9]; APWG [10].

2. METHODOLOGY

There are numerous internet databases where scientists can keep and share their research findings with the rest of the world. In this study we purposely chose indexed research publications in Scopus and Web of Science for rigorous reviews. This is mainly due to the reputability, quality, and global acceptance of the aforementioned indexing databases.

To gain access to relevant studies for rigorous review, we first formulated a search strategy that included "Website" AND "Phishing Detection" AND ("Machine Learning" OR "Deep Learning"), "Phishing website detection using Machine Learning approach", and "Phishing website detection using Deep Learning approach". Following that, we sent the query(s) specified in the first step to the Scopus and Web of Science databases, where we were able to access numerous research works focusing on machine learning and deep learning-based phishing website detection, and we limited our search to research published between 2017 and 2021 to look for recent state-of-the-art techniques. Based on the aforementioned criterion, we were able to download a total of 135 studies from both indexing databases, i.e. 84 studies from Scopus and 51 studies from the Web of Science Database.

After downloading 135 studies, we have formulated the criteria for screening relevant studies for rigorous review. We started with reading abstracts and full documents to confirm that the studies focusing on machine learning and deep learning-based phishing website detection: were published between 2017 and 2021, included the lists of website features used, and that the model performance evaluation metrics includes accuracy, precision, recall, and F1-measure. We preferred the studies that contained lists of website features used as domain expertise is required to define features that separate a legitimate website from phishing websites. The same website feature may also be defined differently by different studies. For example, according to the study [12], a website is phishing if the domain age record in the WHOIS database is less than a year, whereas a website is phishing if the domain age record in the WHOIS database is less than six months according to the study [13-16]. Based on the aforementioned criterion, 30 out of 135 research works were qualified for rigorous review, as shown in Figure 3.

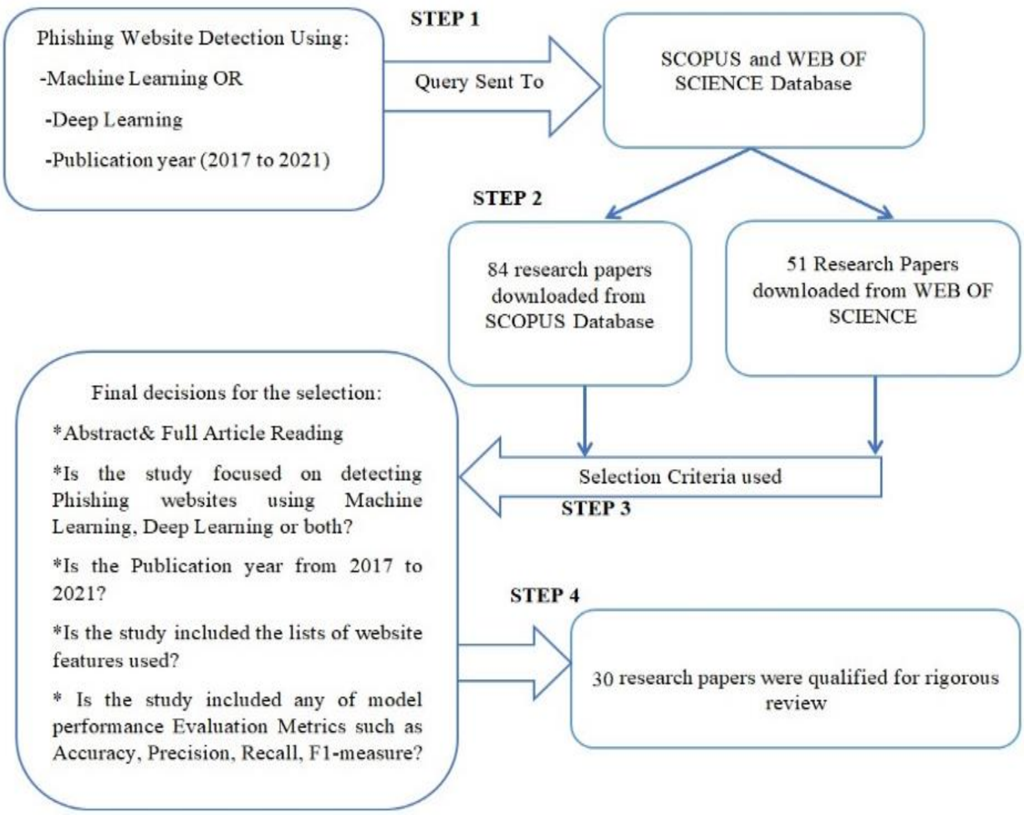

Figure 3. Relevant studies selection criterion.

3. RESEARCH FINDINGS

In this study, 10 parameters were used to excavate the key research gaps from criterion-based selected studies, as shown in Table 1. These include: a) Type of Machine Learning or Deep Learning algorithm used, b) Type of Relevant Feature Selection Techniques used, c) Dataset Source, d) Dataset Size, e) Phish-Legitimate Dataset Ratios, f) Percentage of Train-Test Dataset Split Ratios, g) Number of Website Features Used, h) Best Performed Detection Model, i) Accuracy Rate, and Run-time Analysis.

Table 1. Comparative analysis on machine learning and deep learning-based phishing websites detection.

Author(s) and publ. year	Evaluation Criteria	Evaluation result	Major comments/ research gaps
Hannousse and Yahiouche [3], 2021	ML/DL algorithm used	Decision-Tree Random-Forest Logistic-Regression Naïve-Bayes SVM	No deep learning algorithms. Unsuitability of some content-based features for runtime analysis. No Hybrid-ensemble Feature Selection technique. No percentage of Train-Test dataset split ratio.
	Feature selection techniques used	Pearson Correlation Information Gain Relief rank Chi-Square Wrapper based	
	Dataset source	Phish-tank Open-phish Alexa Yardex	
	Dataset size	11,430	
	Phish-legitimate dataset ratio	Balanced	
	% of Train-Test dataset split ratio	Unknown	
	Number of website features used	87	
	Best Performed detection model	Random-Forest	
	Accuracy rate	96.83%	
Pavan, et al. [17], 2021.	ML/DL algorithm used	CNN	Used single dataset source. Imbalanced Dataset usage. No comparative analysis was made with other ML or DL algorithms. No run-time analysis. No percentage of Train-Test dataset split ratio.
	Feature selection techniques used	Swarm Intelligence Binary Bat Algorithm	
	Dataset source	Kaggle	
	Dataset size	11,055	
	Phish-legitimate dataset ratio	Imbalanced ratio 56%: 44%	
	% of Train-Test dataset split ratio	Unknown	
	Number of website features used	30	
	Best Performed detection model	CNN	
	Accuracy rate	94.8%	
Sabahno and Safara [18], 2021.	ML/DL algorithm used	SVM	Used single dataset source. No Phish-Legitimate ratio of the datasets. ISHO algorithm was not experimented with other ML or DL algorithms.
	Feature selection techniques used	Improved Spotted Hyena Optimization (ISHO) Algorithm	
	Dataset source	UCI Machine Learning repository	
	Dataset size	11,055	
	Phish-Legitimate dataset ratio	Unknown	
	% of Train-Test dataset split ratio	75%:25%	
	Number of website features used	30	
	Best Performed detection model	SVM+ISHO	
	Accuracy rate	98.64%	
Gupta, et al. [19], 2021.	ML/DL algorithm used	K-NN Random-Forest Logistic-Regression SVM	Used single dataset source. No DNS and web-content based features. Small number of website features. No deep learning algorithms.
	Feature selection techniques used	Spearman correlation K best score Random-Forest score	
	Dataset source	"ISCXURL-2016" Dataset	
	Dataset size	19,964	

Author(s) and publ. year	Evaluation Criteria	Evaluation result	Major comments/ research gaps
	Phish-Legitimate dataset ratio	Nearly Balanced ratio 49.9% : 50.1%	
	% of Train-Test dataset split ratio	80%: 20%	
	Number of website features used	9	
	Best Performed detection model	Random-Forest	
	Accuracy rate	99.57%	
Lakshmi, et al. [16], 2021.	ML/DL algorithm used	DNN	Used single dataset source. No the Phish-Legitimate ratio of the datasets. No relevant feature selection techniques. No comparative analysis was made with other ML or DL algorithms. No run-time analysis.
	Feature selection techniques used	Unknown	
	Dataset source	UCI Machine Learning repository	
	Dataset size	11,000	
	Phish-Legitimate dataset ratio	Unknown	
	% of Train-Test dataset split ratio	67% :33%	
	Number of website features used	30	
	Best Performed detection model	DNN	
	Accuracy rate	96.25%	
Mourtaji, et al. [20], 2021.	ML/DL algorithm used	CNN MLP K-NN SVM Classification and Regression Tree	Imbalanced Dataset usage. The Alexa only comprises top-ranked legitimate domains, with sub-domain and URL path details excluded. No Hybrid-ensemble Feature Selection technique.
	Feature selection techniques used	Principal Component Analysis Recursive Feature Elimination Uni-variate Feature Selection	
	Dataset source	Phish-tank Alexa	
	Dataset size	40,000	
	Phish-Legitimate dataset ratio	Highly Imbalanced ratio 26% : 74%	
	% of Train-Test dataset split ratio	80%:20%	
	Number of website features used	37	
	Best Performed detection model	CNN	
	Accuracy rate	97.94%	
Odeh, et al. [8], 2020.	ML/DL algorithm used	MLP	Small datasets usage. No Phish-Legitimate ratio of the datasets. No comparative analysis with other ML or DL algorithms. No run-time analysis.
	Feature selection techniques used	Single attribute evaluator	
	Dataset source	Phish-Tank Google search Miller-Smiles	
	Dataset size	2456	
	Phish-Legitimate dataset ratio	Unknown	
	% of Train-Test dataset split ratio	70%:30%	
	Number of website features used	30	
	Best Performed detection model	MLP	
	Accuracy rate	98.5%	
Zhu, et al. [21], 2020.	ML/DL algorithm used	ANN Naïve-Bayes Logistic-Regression Decision-Tree SVM Random-Forest	Imbalanced Dataset usage. No run-time analysis.
	Feature selection techniques used	Gini coefficient K-medoid	
	Dataset source	UCI Machine Learning repository Phish-tank Alexa	
	Dataset size	25,637	
	Phish-Legitimate dataset ratio	Highly Imbalanced dataset ratio 30% :70%	

Author(s) and publ. year	Evaluation Criteria	Evaluation result	Major comments/ research gaps
	% of Train-Test dataset split ratio	70%:30%	
	Number of website features used	30	
	Best Performed detection model	ANN	
	Accuracy rate	97.8%	
Alam, et al. [22], 2020	ML/DL algorithm used	Decision-Tree Random-Forest	Small dataset usage. Used a single source dataset. No Phish-Legitimate ratio of the datasets. No Percentage of Train-Test dataset split ratio. No Hybrid-ensemble Feature Selection technique. No run-time analysis. No deep learning algorithms.
	Feature selection techniques used	Gain Ratio Relief-F Recursive Feature Elimination Principal Component Analysis	
	Dataset source	Kaggle	
	Dataset size	2,211	
	Phish-Legitimate dataset ratio	Unknown	
	% of Train-Test dataset split ratio	Unknown	
	Number of website features used	32	
	Best Performed detection model	Random-Forest	
	Accuracy rate	96.96%	
Saha, et al. [23], 2020.	ML/DL algorithm used	MLP	No Phish-Legitimate ratio of the datasets. No Percentage of Train-Test dataset split ratio. Small number of website Features. No Hybrid-ensemble Feature Selection technique. No run-time analysis. No comparative analysis with other ML or DL algorithms.
	Feature selection techniques used	Gain Ratio Relief-F Recursive Feature Elimination Principal Component Analysis	
	Dataset source	Kaggle	
	Dataset size	10,000	
	Phish-Legitimate dataset ratio	Unknown	
	% of Train-Test dataset split ratio	Unknown	
	Number of website features used	9	
	Best Performed detection model	MLP	
	Accuracy rate	93%	
Subasi and Kremic [24], 2019.	ML/DL algorithm used	K-NN Random-Forest Decision-Tree SVM ANN Adaboost Multiboost	Did not mentioned the actual dataset size. No Phish-Legitimate ratio of the datasets. No Percentage of Train-Test dataset split ratio. Used a single source datasets. No relevant feature selection techniques. High computational time requirement.
	Feature selection techniques used	Unknown	
	Dataset source	UCI Machine Learning repository	
	Dataset size	Unknown	
	Phish-Legitimate dataset ratio	Unknown	
	% of Train-Test dataset split ratio	Unknown	
	Number of website features used	29	
	Best Performed detection model	SVM with Adaboost	
	Accuracy rate	97.61%	
Abedin, et al. [25], 2020	ML/DL algorithm used	K-NN Random-Forest Logistic-Regression	Used a Single dataset source. No Phish-Legitimate ratio of the datasets. No relevant feature selection techniques. No deep learning algorithms. No run-time analysis.
	Feature selection techniques used	Unknown	
	Dataset source	Kaggle	
	Dataset size	11,504	
	Phish-Legitimate dataset ratio	Unknown	
	% of Train-Test dataset split ratio	80%:20%	
	Number of website features used	32	
	Best Performed detection model	Random-Forest	
	Accuracy rate	97%	
Zamir, et al. [26],	ML/DL algorithm used	Neural Network + Random Forest +Bagging K-NN+ Random Forest + Bagging	Used a single source dataset. No the Phish-Legitimate ratio of the datasets. No Percentage of Train-Test dataset split ratio.

Author(s) and publ. year	Evaluation Criteria	Evaluation result	Major comments/ research gaps
2020.			High computational time requirement. No Hybrid-ensemble Feature Selection technique.
	Feature selection techniques used	Information Gain Relief-F, Recursive Feature Elimination Gain Ratio	
	Dataset source	Kaggle	
	Dataset size	11,055	
	Phish-Legitimate dataset ratio	Unknown	
	% of Train-Test dataset split ratio	Unknown	
	Number of website features used	32	
	Best Performed detection model	Neural Network + Random Forest +Bagging	
	Accuracy rate	97.4%	
Hossain, et al. [27], 2020.	ML/DL algorithm used	K-NN Random-Forest Decision-Tree SVM Logistic-Regression	Used a single source dataset. No percentage of Train-Test dataset split ratio. No deep learning algorithms. No DNS and Page based features. No run-time analysis.
	Feature selection techniques used	Principal Component Analysis	
	Dataset source	Mendeley online repository	
	Dataset size	10,000	
	Phish-Legitimate dataset ratio	Balanced	
	% of Train-Test dataset split ratio	Unknown	
	Number of website features used	48	
	Best Performed detection model	Random-Forest	
	Accuracy rate	99% F1-score	
Suryan, et al. [28], 2020.	ML/DL algorithm used	Random-Forest SVM Generalized Linear Model Generalized Additive Model Recursive Partitioning and Regression Trees	Used a single source dataset. Imbalanced dataset usage No deep learning algorithms. No run-time analysis.
	Feature selection techniques used	Principal Component Analysis	
	Dataset source	UCI Machine Learning repository	
	Dataset size	11,055	
	Phish-Legitimate dataset ratio	Imbalanced dataset ratio 44%: 56%	
	% of Train-Test dataset split ratio	70%:30%	
	Number of website features used	31	
	Best Performed detection model	Random-Forest	
	Accuracy rate	98.34%	
Gandotra and Gupta [29], 2020.	ML/DL algorithm used	K-NN Random-Forest Decision-Tree SVM Naïve-Bayes Adaboost	No relevant feature selection techniques. No deep learning algorithms. No run-time analysis.
	Feature selection techniques used	Unknown	
	Dataset source	Alexa Payment gateway Phish-tank Open-phish	
	Dataset size	5223	
	Phish-Legitimate dataset ratio	Nearly balanced dataset ratio 48% : 52%	
	% of Train-Test dataset split ratio	Unknown	
	Number of website features used	20	
	Best Performed detection model	Random-Forest	
	Accuracy rate	99.5%	

Author(s) and publ. year	Evaluation Criteria	Evaluation result	Major comments/ research gaps
Singhal, et al. [30], 2020.	ML/DL algorithm used	Random-Forest Gradient-Boost Neural-Network	Small number of website features. No DNS and page rank based features. No the percentage of Train-Test dataset split ratio. No relevant feature selection techniques. No run-time analysis.
	Feature selection techniques used	Unknown	
	Dataset source	Majestic repository Phish-tank	
	Dataset size	80,000	
	Phish-Legitimate dataset ratio	Balanced	
	% of Train-Test dataset split ratio	Unknown	
	Number of website features used	14	
	Best Performed detection model	Gradient-Boost	
	Accuracy rate	96.4%	
Zaini, et al. [31], 2019.	ML/DL algorithm used	Random-Forest J48 MLP K-NN	Small dataset usage. No percentage of Train-Test dataset split ratio. No Phish-Legitimate ratio of the datasets. No run-time analysis.
	Feature selection techniques used	Unknown	
	Dataset source	UCI Machine Learning repository	
	Dataset size	2,456	
	Phish-Legitimate dataset ratio	Unknown	
	% of Train-Test dataset split ratio	Unknown	
	Number of website features used	30	
	Best Performed detection model	Random-Forest	
	Accuracy rate	94.79%	
Shabudin, et al. [32], 2020.	ML/DL algorithm used	Random-Forest Naïve-Bayes MLP	Used a single source dataset. Imbalanced dataset usage. No percentage of Train-Test dataset split ratio. No Hybrid-ensemble Feature Selection technique.
	Feature selection techniques used	Relief-ranking, Information Gain	
	Dataset source	UCI Machine Learning repository.	
	Dataset size	11,055	
	Phish-Legitimate dataset ratio	Imbalanced Dataset ratio 44% : 56%	
	% of Train-Test dataset split ratio	Unknown	
	Number of website features used	30	
	Best Performed detection model	Random-Forest	
	Accuracy rate	97.18%	
Kumar, et al. [33], 2020.	ML/DL algorithm used	Random-Forest Naïve-Bayes K-NN Logistic-Regression Decision Tree	Used a single source dataset. No web-content features. No relevant feature selection techniques. No deep learning algorithms. No run-time analysis.
	Feature selection techniques used	Unknown	
	Dataset source	Github.com repository	
	Dataset size	100,000	
	Phish-Legitimate dataset ratio	Balanced	
	% of Train-Test dataset split ratio	70%:30	
	Number of website features used	26	
	Best Performed detection model	Random Forest	
	Accuracy rate	98.03%	
Harinahalli Lokesh and BoreGowda [34], 2021.	ML/DL algorithm used	Random-Forest Decision-Tree K-NN SVM	Did not mentioned the actual dataset size. No Phish-Legitimate ratio of the datasets. No run-time analysis.
	Feature selection techniques used	Wrapper-based	
	Dataset source	Miller-Smiles Phish-Tank	
	Dataset size	Unknown	
	Phish-Legitimate dataset ratio	Unknown	
	% of Train-Test dataset split ratio	80%:20%	
	Number of website features used	30	
	Best Performed detection model	Random-Forest	
	Accuracy rate	96.87%	
		Naïve-Bayes	No run-time analysis was

Author(s) and publ. year	Evaluation Criteria	Evaluation result	Major comments/ research gaps
Chiew, et al. [35], 2019.	ML/DL algorithm used	Random-Forest JRiP C4.5 PART	made using all (48) website features. No deep learning algorithms. No DNS and page rank based features.
	Feature selection techniques used	Hybrid-Ensemble Features Selection technique.	
	Dataset source	Alexa Common-Crawl Open-Phish Phish-tank	
	Dataset size	10,000	
	Phish-Legitimate dataset ratio	Balanced	
	% of Train-Test dataset split ratio	70%:30%	
	Number of website features used	48	
	Best Performed detection model	Random-Forest	
	Accuracy rate	94.6%	
Alswailem, et al. [36], 2019.	ML/DL algorithm used	Random-Forest	Imbalanced dataset usage. No DNS and page rank based features. No comparative analysis was made with other ML or DL algorithms. No run-time analysis.
	Feature selection techniques used	Combinatory (Mini-max) approach	
	Dataset source	Phish-tank 10 experts	
	Dataset size	16,000	
	Phish-Legitimate dataset ratio	Highly Imbalanced dataset ratio 75% : 25%	
	% of Train-Test dataset split ratio	80%:20%	
	Number of website features used	36	
	Best Performed detection model	Random-Forest	
	Accuracy rate	98.8%	
Tumuluru and Jonnalagadda [37], 2019.	ML/DL algorithm used	Extreme Learning Machine Random Forest Naive Bayes SVM	Used a single source dataset. No Phish-Legitimate ratio of the datasets. No percentage of Train-Test dataset split ratio. No relevant feature selection techniques. No run-time analysis.
	Feature selection techniques used	Unknown	
	Dataset source	UCI Machine Learning repository	
	Dataset size	11,000	
	Phish-Legitimate dataset ratio	Unknown	
	% of Train-Test dataset split ratio	Unknown	
	Number of website features used	30	
	Best Performed detection model	Random-Forest	
	Accuracy rate	98.5%	
Sönmez, et al. [13], 2018.	ML/DL algorithm used	Extreme Learning Machine Naive Bayes SVM	Used a single source dataset. No Phish-Legitimate ratio of the datasets. No percentage of Train-Test dataset split ratio. No relevant feature selection techniques. No run-time analysis.
	Feature selection techniques used	N/A	
	Dataset source	UCI Machine Learning repository	
	Dataset size	11,000	
	Phish-Legitimate dataset ratio	Unknown	
	% of Train-Test dataset split ratio	Unknown	
	Number of website features used	30	
	Best Performed detection model	Extreme Learning Machine	
	Accuracy rate	95.34%	
Rao and Pais [12], 2019.	ML/DL algorithm used	Random-Forest Logistic-Regression J48 SVM MLP Adaboost Naïve-Bayes Sequential Minimal	Small dataset usage. Imbalanced dataset usage. Alexa is only comprises top-ranked legitimate domains, with sub-domain and URL path details excluded. No web content-based features such as JavaScript

Author(s) and publ. year	Evaluation Criteria	Evaluation result	Major comments/ research gaps
		Optimization	files, Iframes HTML files. No run-time analysis.
	Feature selection techniques used	Principal Component Analysis.	
	Dataset source	Alexa Phish-tank	
	Dataset size	3526	
	Phish-Legitimate dataset ratio	Imbalanced Dataset ratio 59% :41% .	
	% of Train-Test dataset split ratio	75%:25%	
	Number of website features used	16	
	Best Performed detection model	Random-Forest	
	Accuracy rate	99.31%	
Jain and Gupta [38], 2019.	ML/DL algorithm used	SVM Random-Forest Logistic-Regression C4.5 Sequential Minimal Optimization Neural Network, Adaboost Naïve-Bayes	Small dataset usage. Imbalanced dataset usage. Small number of website features. No DNS and URL based features. No relevant feature selection techniques. No run-time analysis.
	Feature selection techniques used	Unknown	
	Dataset source	Alexa Stuffgate Phish-tank	
	Dataset size	2544	
	Phish-Legitimate dataset ratio	Imbalanced dataset ratio 56% : 44%	
	% of Train-Test dataset split ratio	90%:10%	
	Number of website features used	12	
	Best Performed detection model	Logistic-Regression	
	Accuracy rate	98.42%	
Pratiwi, et al. [39], 2018.	ML/DL algorithm used	ANN	Used a single source datasets. Small dataset usage. No Phish-Legitimate ratio of the datasets. No relevant feature selection techniques. No comparative analysis with other ML or DL algorithms.
	Feature selection techniques used	Unknown	
	Dataset source	UCI Machine Learning repository	
	Dataset size	2,455	
	Phish-Legitimate dataset ratio	Unknown	
	% of Train-Test dataset split ratio	80%:20%	
	Number of website features used	18	
	Best Performed detection model	ANN	
	Accuracy rate	83.38%	
Shirazi, et al. [40], 2017.	ML/DL algorithm used	DNN SVM	Small dataset usage. No Phish-Legitimate ratio of the datasets. Alexa is only comprises top-ranked legitimate domains, with sub-domain and URL path details excluded.
	Feature selection techniques used	Recursive Feature Elimination	
	Dataset source	*Alexa *Phish-tank	
	Dataset size	5,000	
	Phish-Legitimate dataset ratio	Balanced	
	% of Train-Test dataset split ratio	Unknown	
	Number of website features used	28	
	Best Performed detection model	SVM	
	Accuracy rate	93% for Binary Dataset 96% for Non-binary datasets	
Jain and Gupta [6], 2018.	ML/DL algorithm used	SVM Naïve-Bayes Logistic-Regression Neural-Network Random-Forest	Imbalanced dataset usage. Small datasets usage. No DNS and page rank based features.
	Feature selection techniques used	Pearson Correlation Coefficient	
	Dataset source	Alexa Payment gateway	

Author(s) and publ. year	Evaluation Criteria	Evaluation result	Major comments/ research gaps
		Phishtank Open-phish	
	Dataset size	4,059	
	Phish-Legitimate dataset ratio	Imbalanced dataset ratio 53% : 47%	
	% of Train-Test dataset split ratio	90%: 10%	
	Number of website features used	19	
	Best Performed detection model	Random-Forest	
	Accuracy rate	99.09	

4. DISCUSSION ON KEY RESEARCH FINDINGS

The study findings are organized into eight important areas, as shown in Figure 4, to make discussion easier.

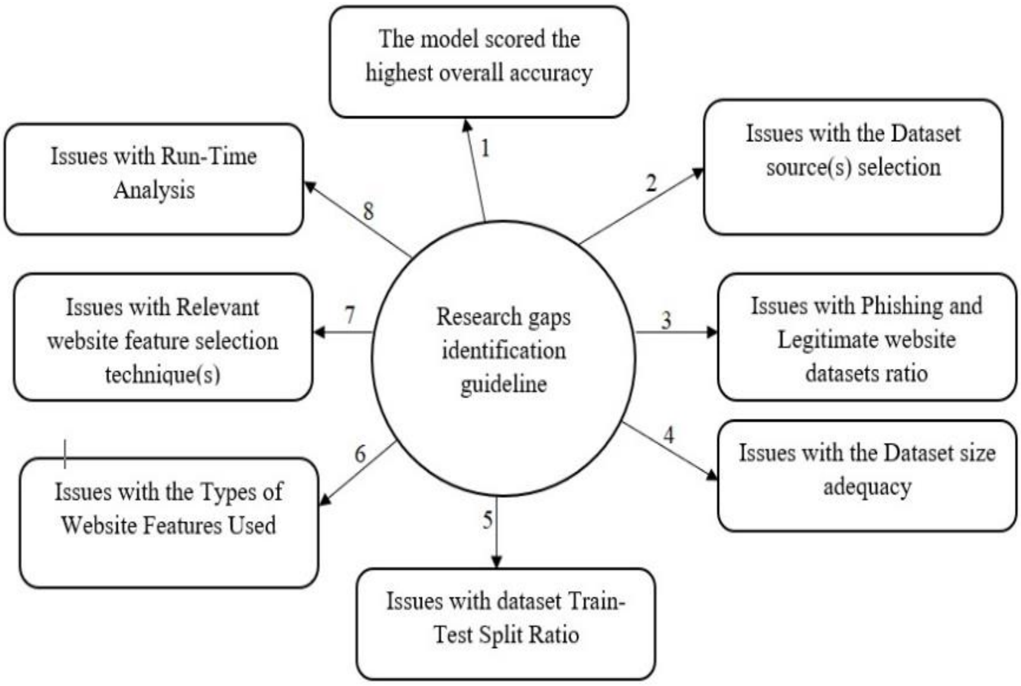

Figure 3. Research gaps analysis and discussion guideline.

4.1. The Model Scored that the Highest Overall Accuracy in Phishing Website Detection

To address phishing attacks effectively, two sorts of misclassifications are expected to be reduced by the phishing website detection model: i) False Positive rate and ii) False Negative rate. The first one is blocking online users from accessing legitimate websites due to incorrectly labeling legitimate websites as phishing, while the second one is allowing online users to visit fraudulent websites due to incorrectly labeling phishing websites as legitimate.

In the 30 reviewed studies, many machine learning and/or deep learning algorithms were applied for tackling the problems in phishing websites detection. These algorithms, however, did not perform equally well in detecting phishing websites. The study findings reveal that Random Forest has the best overall performance in the majority (17) of reviewed research papers, with accuracy results between 94.6% and 99.57%. In the remaining 13 different studies, algorithms such as SVM, MLP, Logistic Regression, Extreme Learning Machine (ELM), Gradient Boost, ANN, CNN, and DNN performed the highest overall accuracy. Because the significant part of phishing defense is detecting phishing websites accurately and timely manner, this study would suggest future researchers choose the aforementioned machine learning and deep learning algorithms as a priority, along with cleaned representative dataset usage and relevant feature selection techniques.

4.2. Issues with the Dataset Source(s) Selection

Constructing a cleaned representative dataset is more important than selecting a specific machine learning model, regardless of datasets size [41]. In the real world scenario, multiple books about a particular subject or topic could be written by various authors. However, due to the differences in scopes, each book would not include the same content. Readers are expected to read multiple books in order to gain a broad variety of knowledge on a certain topic. The datasets used to train and test both machine learning and deep learning algorithms are the same. This means that machine learning and deep learning algorithms are likely to be taught using datasets from a variety of reliable sources. There are numerous dataset sources that scientists can collect to train and test machine learning and deep learning algorithms. Dataset sources such as phish-tank, Kaggle, Alexa, UCI machine learning repository, payment gateway, GitHub, Majestic and open-phish were among the widely used dataset sources in the reviewed studies.

As shown in Table 1 of the finding section, 14 of the reviewed studies used datasets from a single source, either Kaggle or the UCI machine learning repository. The UCI machine learning repository did not have any raw URL datasets, meaning that extracting new additional features from URLs for scalability is impossible [3]. Alexa's repository only included top-ranked legitimate main domains, eliminating sub-domains and URL path features [3]. 3 of the reviewed studies collected legitimate websites dataset from only the Alexa repository [12, 20, 40]. This shows that the learning model used in the aforementioned study is unable to detect phishing websites based on sub-domain and URL path features, and can be viewed as a drawback of using a single dataset source. Since phishing website attacks are a global issue, detecting phishing/fraudulent websites is not expected to be independent of specific sectors such as financial, health, education, agriculture, e-commerce, and more.

4.3. Issues with the Dataset Size Adequacy

There is still no universal consensus reached on what defines a small dataset size [41]. As it was presented in Table 1 of research finding section, different numbers of datasets were used by different studies to train the learning models. To evaluate each criterion-based selected study, we defined a small dataset size as "a dataset contained less than 5000 phishing and 5000 legitimate websites". According to Prusa, et al. [42], a machine learning model trained with huge datasets can outperform a model trained with small datasets in terms of accuracy. This is mainly due to the model trained with small datasets failing to generalize patterns, resulting in unreliable and biased outputs [15]. According to this study criterion, 19 reviewed research papers used at least 5000 legitimate and 5000 phishing website datasets to train the model(s), while 9 studies used small datasets, and as a result the model may wrongly generalize what it was taught or inaccurate results may be displayed to online users.

4.4. Issues with Train-Test Dataset Split Ratio

A dataset train-test split is needed at the data preprocessing stage. This is mainly because it is not recommended to use the same datasets for both training and testing the model. As shown in Table 2, the majority of the reviewed studies used the dataset split ratios of 80:20 and 70:30 %. However, there are no clearly established rules for what dataset train-split ratio to use for how much dataset size. More study is needed here.

4.5. Issues with Phishing and Legitimate Website Datasets Proportion

When learning models are trained on unbalanced datasets, their accuracy is misleading. The highest accuracy does not always imply that the model is the best, as the model's accuracy can decline if the classifiers fail to consider all classes in an equal ratio [7, 43]. Furthermore, in binary classification tasks, using a balanced data set is often needed particularly when accuracy is utilized as the model evaluation metric [44]. According to Kumar, et al. [33], a random mix-up of both legitimate and phishing websites datasets greatly contributed to the optimized performance of the machine learning model.

As shown in Table 1 of the finding section, 10 of the reviewed studies did not collect Phish-legitimate website datasets in equal ratios, nor did they use any dataset balancing methods. This indicates that the models in the aforementioned studies were biased because they exclusively favored the majority class. Only 7 of the reviewed studies had balanced dataset rations, and the other 13 studies did not specify how many phishing and legitimate websites were used in their research.

Table 2. Dataset train-test split ratio.

Dataset train-test split ratios	Lists of Authors	Dataset size	Total No. of Studies
75%:25%	Sabahno and Safara [18], 2021.	11.055	3
	Alswailem, et al. [36], 2019	16.000	
	Rao and Pais [12], 2019.	3.526	
80%: 20%	Gupta, et al. [19], 2021.	19.964	5
	Mourtaji, et al. [20], 2021.	40.000	
	Abedin, et al. [25], 2020.	11.504	
	Harinahalli Lokesh and BoreGowda [34], 2021.	N/A.	
	Pratiwi, et al. [39], 2018.	2.455	
67% :33%	Lakshmi, et al. [16], 2021.	11.000	1
90%:10%	Jain and Gupta [38], 2019.	2.544	2
	Jain and Gupta [6], 2018.	4.059	
70%:30%	Odeh, et al. [8], 2020.	2,456	5
	Zhu, et al. [21], 2020.	25,637	
	Suryan, et al. [28] 2020.	11.055	
	Kumar, et al. [33], 2020.	100.000	
	Chiew, et al. [35], 2019.	10.000	

4.6. Issues with the Types of Website Features Used

To detect phishing websites, a variety of features are available. This study found four different categories of website features in a recent review of several studies: URL-based, domain-based, web-content/source-code-based and page-based features, as shown in Figure 5.

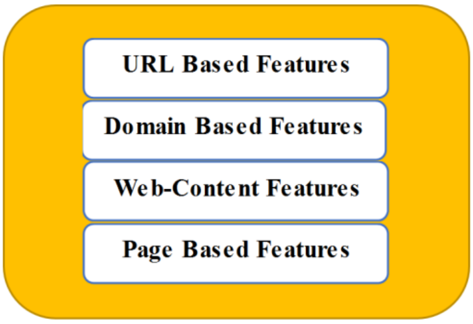

Figure 4. Types of website features.

The study findings reveal that there is still a lack of common consensus among the scientific community on the choice of features for phishing websites detection. For example, there were studies that excluded domain-based and page-based features due to not being suitable for run-time analysis [27, 35, 38]. The content-based features were excluded from the study due to the non-availability of web-content-based features as a result of the short life duration of phishing websites and the lack of suitability for run-time analysis [19, 33, 44]. However, the study by Hannousse and Yahiouche [3]; Subasi and Kremic [24] refuted the claims of those studies stating domain-based and page-based features were not suitable for run-time analysis by saying that extracting DNS and page-based features from third-party services was computationally faster than extracting web-content-based features.

The scientific community has yet to come to an agreement on what defines a "phishing website short life span." For example, according to the study [12], a website is phishing if the domain age record in the WHOIS database is less than a year, whereas a website is phishing if the domain age record in the WHOIS database is less than 6 months [13, 14, 16, 21]. Using high-speed Internet access and alternative methods, the network delay or non-suitability for run-time analysis during phishing website detection can be handled [38]. To address the short life span of phishing websites, the study by Hannousse and Yahiouche [3] proposed to generate a Document Object Model (DOM) tree of webpages using the available tool 'HTML DOM Parser for Python', and stored them in a separate dataset along with URLs index, assisting them to extract more web-content-based features regardless of the dead links.

4.7. Issues with Relevant Website Feature Selection Technique

As all website features are not equally important to detect phishing websites, making use of relevant feature selection techniques is crucial to improve the machine learning model accuracy to speed up the time taken for training and testing as well as to address overfitting issues [7]. As shown in Table 1 of the finding section, 9 of the reviewed studies did not employ any feature selection technique. Principal Component Analysis, Recursive Feature Elimination, Pearson Correlation Coefficient, Info-Gain, Chi-squire, Relief-ranking, Gain Ratio, and Gini coefficient were among the most commonly used feature selection strategies in the reviewed papers, and were applied on an individual basis in the majority of the assessed research. To combat the challenge of phishing website identification, just one study [35] used the hybrid ensemble feature selection technique and achieved a better run-time analysis in contrast to a single-based feature selection technique.

4.8. Issues with Run-Time Analysis

Before internet visitors hand over their personal information to fraudulent websites, machine learning and deep learning algorithms must provide a fast prediction time along with the highest level of accuracy. However, 20 of the 30 studies reviewed did not conduct a run-time analysis of the model as shown in Table 1 of the finding section. To tackle the problem of phishing website detection, the study by Subasi and Kremic [24] used boosting-type ensemble learning learners that combined Support Vector Machine (SVM) and Adaboost, and the study by Abedin, et al. [25] used bagging-type ensemble learning learners that combined Neural Network and Random Forest. Despite the highest model accuracy scores, both experiments found that predicting phishing websites requires a lot of computational time.

5. CONCLUSIONS & FUTURE RESEARCH DIRECTIONS

In this study, an extensive effort has been made to rigorously review recent studies focusing on Machine Learning and Deep Learning Based Phishing Websites Detection to dig out the main gaps and offer suitable solutions. As a result, significant research gaps were identified. These gaps are mainly related to imbalanced dataset use, selection of dataset source, dataset size adequacy, dataset train-test split ratios, website feature inclusion and exclusion, the issue with relevant feature selection techniques, and run-time analysis. This study clearly presented a summary of the comparative analysis performed on each reviewed study so that future researchers could use it as a structured guideline to develop a novel anti-phishing website attack solution.

The findings reveal that Random Forest has the best overall accuracy in the majority of peer-reviewed research articles. In the remaining 13 different studies, algorithms such as SVM, MLP, Logistic Regression, Extreme Learning Machine, Gradient Boost, ANN, CNN, and DNN performed the highest overall accuracy. High computational time requirement was reported by some studies that utilized bagging- and boosting-type ensemble learning learners, despite the highest model accuracy scored. Fast computational time is shown in the study that utilized the Hybrid Ensemble Feature Selection technique. There is still a lack of common consensus reached on what is defining small dataset size and the exact threshold of phishing websites' short lifespan; there are no clearly established rules for how much dataset train-split ratio to use for how much dataset size. Future research will require the construction of benchmark datasets that will represent both machine learning and deep learning algorithms. The details of each machine learning and deep learning algorithm, as well as the details of each relevant feature selection technique, were not included, and the study did not conduct an experiment to address the identified research gaps.

Funding: This study received no specific financial support.
Competing Interests: The authors declare that they have no competing interests.
Authors' Contributions: Both authors contributed equally to the conception and design of the study.

REFERENCES

[1] S. Patil and S. Dhage, "A methodical overview on phishing detection along with an organized way to construct an anti-phishing framework.," in *2019 5th Int Conf Adv Comput Commun Syst*, 2019, pp. 588–93.

[2] N. Abdelhamid, A. Ayesh, and F. Thabtah, "Phishing detection based associative classification data mining," *Expert Systems with Applications*, vol. 41, pp. 5948-5959, 2014.Available at: https://doi.org/10.1016/j.eswa.2014.03.019.

[3] A. Hannousse and S. Yahiouche, "Towards benchmark datasets for machine learning based website phishing detection: An experimental study," *Engineering Applications of Artificial Intelligence*, vol. 104, p. 104347, 2021.Available at: https://doi.org/10.1016/j.engappai.2021.104347.

[4] APWG, "Phishing activity trends report 1st Quarter 2021,Anti-Phishing Working Group(APWG), quarter1(June,8), pp.1-13. Retrieved from: https://docs.apwg.org/reports/apwg_trends_report_q1_2021.pdf," 2021.

[5] P. Labs, "2019 phishing trends and intelligence report: The growing social engineering threat, PhishLabs, pp.1- 30. Retrieved from: https://info.phishlabs.com/hubfs/2019%20PTI%20Report/2019%20Phishing%20Trends%20and%20Intelligence%20Report.pdf," 2019.

[6] A. K. Jain and B. B. Gupta, "Towards detection of phishing websites on client-side using machine learning based approach," *Telecommunication Systems*, vol. 68, pp. 687-700, 2018.Available at: https://doi.org/10.1007/s11235-017-0414-0.

[7] L. Tang and Q. H. Mahmoud, "A survey of machine learning-based solutions for phishing website detection," *Machine Learning and Knowledge Extraction*, vol. 3, pp. 672-694, 2021.Available at: https://doi.org/10.3390/make3030034.

[8] A. Odeh, A. Alarbi, I. Keshta, and E. Abdelfattah, "Efficient prediction of phishing websites using multilayer perceptron (mlp)," *J Theor Appl Inf Technol*, vol. 98, pp. 3353–63, 2020.

[9] APWG, "Phishing activity trends report 2nd Quarter 2021,Anti-Phishing Working Group(APWG), quarter2(September, 22), pp.1-12. Retrieved from: https://docs.apwg.org/reports/apwg_trends_report_q2_2021.pdf," 2021.

[10] APWG, "Phishing activity trends report 3rd Quarter 2021,Anti-Phishing Working Group(APWG), quarter3(November,22),pp.1-12. Retrieved from: https://docs.apwg.org/reports/apwg_trends_report_q3_2021.pdf," 2021.

[11] APWG.. Phishing Activity Trends Report 3rd Quarter. (July-September)2021.

[12] 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.Available at: https://doi.org/10.1007/s00521-017-3305-0.

[13] Y. Sönmez, T. Tuncer, H. Gökal, and E. Avci, "Phishing web sites features classification based on extreme learning machine," in *6th Int Symp Digit Forensic Secur ISDFS 2018 - Proceeding*, 2018, pp. 1–5.

[14] A. Suryan, C. Kumar, M. Mehta, R. Juneja, and A. Sinha, "Learning model for phishing website detection," *ICST Trans Scalable Inf Syst*, 2018.

[15] M. S. Rahman and M. Sultana, "Performance of Firth-and logF-type penalized methods in risk prediction for small or sparse binary data," *BMC Medical Research Methodology*, vol. 17, pp. 1-15, 2017.Available at: https://doi.org/10.1186/s12874-017-0313-9.

[16] L. Lakshmi, M. P. Reddy, C. Santhaiah, and U. J. Reddy, "Smart phishing detection in web pages using supervised deep learning classification and optimization technique adam," *Wireless Personal Communications*, vol. 118, pp. 3549-3564, 2021.Available at: https://doi.org/10.1007/s11277-021-08196-7.

[17] K. P. Pavan, T. Jaya, and V. Rajendran, "SI-BBA – A novel phishing website detection based on Swarm intelligence with deep learning," *Elsevier, Materials Today: Proceedings*, vol. 30, pp. 1-11, 2021.Available at: 10.1016/j.matpr.2021.07.178.

[18] M. Sabahno and F. Safara, "ISHO: Improved spotted hyena optimization algorithm for phishing website detection," *Multimedia Tools and Applications*, pp. 1-20, 2021.Available at: https://doi.org/10.1007/s11042-021-10678-6.

[19] B. B. Gupta, K. Yadav, I. Razzak, K. Psannis, A. Castiglione, and X. Chang, "A novel approach for phishing URLs detection using lexical based machine learning in a real-time environment," *Computer Communications*, vol. 175, pp. 47-57, 2021.Available at: https://doi.org/10.1016/j.comcom.2021.04.023.

[20] Y. Mourtaji, M. Bouhorma, D. Alghazzawi, G. Aldabbagh, and A. Alghamdi, "Hybrid rule-based solution for phishing URL detection using convolutional neural network," *Wireless Communications and Mobile Computing*, vol. 2021, pp. 1-24, 2021.Available at: https://doi.org/10.1155/2021/8241104.

[21] E. Zhu, Y. Ju, Z. Chen, F. Liu, and X. Fang, "DTOF-ANN: An artificial neural network phishing detection model based on decision tree and optimal features," *Applied Soft Computing*, vol. 95, p. 106505, 2020.Available at: https://doi.org/10.1016/j.asoc.2020.106505.

[22] M. N. Alam, I. Saha, D. Sarma, R. Ulfath, F. Lima, and S. Hossain, "Phishing Attacks detection using Machine learning approach," in *Proceedings 3rd International Conference Smart System Inven Technol ICSSIT 2020. 2020;(Icssit):1173-9*, 2020.

[23] I. Saha, D. Sarma, R. Chakma, M. Alam, A. Sultana, and S. Hossain, "Phishing attacks detection using deep learning approach," in *Proc 3rd Int Conf Smart Syst Inven Technol ICSSIT 2020*, 2020, pp. 1180–5.

[24] A. Subasi and E. Kremic, "Comparison of adaboost with multiboosting for phishing website detection," *Procedia Computer Science*, vol. 168, pp. 272-278, 2020.Available at: https://doi.org/10.1016/j.procs.2020.02.251.

[25] N. Abedin, R. Bawm, T. Sarwar, M. Saifuddin, M. Rahman, and S. Hossain, "Phishing attack detection using machine learning classification techniques," in *Proceedings 3rd International Conference Intell Sustain System ICISS 2020*, 2020, pp. 1125–30.

[26] A. Zamir, H. U. Khan, T. Iqbal, N. Yousaf, F. Aslam, A. Anjum, and M. Hamdani, "Phishing web site detection using diverse machine learning algorithms," *The Electronic Library*, vol. 38, pp. 65-80, 2020.Available at: https://doi.org/10.1108/el-05-2019-0118.

[27] S. Hossain, D. Sarma, and R. Chakma, "Machine learning-based phishing attack detection," *Int J Adv Comput Sci Appl*, vol. 11, pp. 378–88, 2020.

[28] A. Suryan, C. Kumar, M. Mehta, R. Juneja, and A. Sinha, "Learning model for phishing website detection," *ICST Trans Scalable Inf Syst*, vol. 7, pp. 1-9, 2020.Available at: https://doi.org/10.4108/eai.13-7-2018.163804.

[29] E. Gandotra and D. Gupta, "Improving spoofed website detection using machine learning," *Cybernetics and Systems*, vol. 52, pp. 169-190, 2021.Available at: https://doi.org/10.1080/01969722.2020.1826659.

[30] S. Singhal, U. Chawla, and R. Shorey, "Machine learning concept drift based approach for malicious website detection," in *2020 Int Conf Commun Syst NETworkS, COMSNETS 2020*, 2020, pp. 582–5.

[31] N. Zaini, D. Stiawan, M. Razak, A. Firdaus, W. Din, and S. Kasim, et al., "Phishing detection system using machine learning classifiers," *Indones J Electr Eng Comput Sci*, vol. 17, pp. 1165–71, 2019.

[32] S. Shabudin, N. S. Sani, K. A. Z. Ariffin, and M. Aliff, "Feature selection for phishing website classification," *Int. J. Adv. Comput. Sci. Appl*, vol. 11, pp. 587-595, 2020.Available at: https://doi.org/10.14569/ijacsa.2020.0110477.

[33] J. Kumar, A. Santhanavijayan, B. Janet, B. Rajendran, and B. Bindhumadhava, "Phishing website classification and detection using machine learning," in *Int Conf Comput Commun Informatics, ICCCI 2020*, 2020, pp. 20-5.

[34] G. Harinahalli Lokesh and G. BoreGowda, "Phishing website detection based on effective machine learning approach," *Journal of Cyber Security Technology*, vol. 5, pp. 1-14, 2021.Available at: https://doi.org/10.1080/23742917.2020.1813396.

[35] K. L. Chiew, C. L. Tan, K. Wong, K. S. Yong, and W. K. Tiong, "A new hybrid ensemble feature selection framework for machine learning-based phishing detection system," *Information Sciences*, vol. 484, pp. 153-166, 2019.Available at: https://doi.org/10.1016/j.ins.2019.01.064.

[36] A. Alswailem, B. Alabdullah, N. Alrumayh, and A. Alsedrani, "Detecting phishing websites using machine learning," in *2nd International Conference on Computer Applications & Information Security (ICCAIS)*, 2019, pp. 1-6.

[37] P. Tumuluru and R. M. Jonnalagadda, "Extreme learning model based phishing vlassifier," *International Journal of Recent Technology and Engineering (IJRTE) ISSN*, vol. 8, pp. 2277-3878, 2019.

[38] A. K. Jain and B. B. Gupta, "A machine learning based approach for phishing detection using hyperlinks information," *Journal of Ambient Intelligence and Humanized Computing*, vol. 10, pp. 2015-2028, 2019.Available at: https://doi.org/10.1007/s12652-018-0798-z.

[39] M. Pratiwi, T. Lorosae, and F. Wibowo, "Phishing site detection analysis using artificial neural network," in *J Phys Conf Ser*, 2018.

[40] H. Shirazi, K. Haefner, and I. Ray, "Improving auto-detection of phishing websites using fresh-phish framework," *International Journal of Multimedia Data Engineering and Management*, vol. 9, pp. 51-64, 2017.

[41] A. Althnian, D. AlSaeed, H. Al-Baity, A. Samha, A. B. Dris, N. Alzakari, A. Abou Elwafa, and H. Kurdi, "Impact of dataset size on classification performance: an empirical evaluation in the medical domain," *Applied Sciences*, vol. 11, p. 796, 2021.Available at: https://doi.org/10.3390/app11020796.

[42] J. Prusa, T. Khoshgoftaar, and N. Seliya, "The effect of dataset size on training tweet sentiment classifiers," in *Proc - 2015 IEEE 14th Int Conf Mach Learn Appl ICMLA 2015*, 2016, pp. 96–102.

[43] T. Goswami and U. Roy, "Classification accuracy comparison for imbalanced datasets with its balanced counterparts obtained by different sampling techniques," in *International Conference on Communications and Cyber Physical Engineering, Lecture Notes in Electrical Engineering*, 2020, pp. 45-53.

[44] 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.Available at: https://doi.org/10.3837/tiis.2020.07.001.