An Investigation of AI-based Ensemble Methods for the Detection of Phishing Attacks

Yazan A. Alsariera
Department of Computer Science, College of Science, Northern Border University, Arar, Saudi Arabia | Department of Computer Science, College of Information and Communications Technology, Tafila Technical University, Jordan
yazan.sadeq@nbu.edu.sa (corresponding author)

Meshari H. Alanazi
Department of Computer Science, College of Science, Northern Border University, Arar, Saudi Arabia
meshari.alanazi@nbu.edu.sa

Yahia Said
Department of Electrical Engineering, College of Engineering, Northern Border University, Saudi Arabia
yahia.said@nbu.edu.sa

Firas Allan
Department of Computer Science, College of Science, Northern Border University, Saudi Arabia
firas.allan@nbu.edu.sa

Received: 20 March 2024 | Revised: 8 April 2024 | Accepted: 11 April 2024

 DOI: https://doi.org/10.48084/etasr.7267

ABSTRACT

Phishing attacks remain a significant cybersecurity threat in the digital landscape, leading to the development of defense mechanisms. This paper presents a thorough examination of Artificial Intelligence (AI)-based ensemble methods for detecting phishing attacks, including websites, emails, and SMS. Through the screening of research articles published between 2019 and 2023, 37 relevant studies were identified and analyzed. Key findings highlight the prevalence of ensemble methods such as AdaBoost, Bagging, and Gradient Boosting in phishing attack detection models. Adaboost emerged as the most used method for website phishing detection, while Stacking and Adaboost were prominent choices for email phishing detection. The majority-voting ensemble method was frequently employed in SMS phishing detection models. The performance evaluation of these ensemble methods involves metrics, such as accuracy, ROC-AUC, and F-score, underscoring their effectiveness in mitigating phishing threats. This study also underscores the availability of credible open-access datasets for the progressive development and benchmarking of phishing attack detection models. The findings of this study suggest the development of new and optimized ensemble methods for phishing attack detection.

Keywords-artificial intelligence; phishing attack detection; AdaBoost; bagging; gradient boosting

I. INTRODUCTION

The Internet and digital devices are crucial aspects of modern society, as they define individuals, companies, and government institutions' communication and interactions [1-2]. Active Internet users, approximately 4.66 billion people - about half of the world's population [3], depend on the digital world for communication, interaction, earning, and making purchases [4]. Following the 2019 pandemic [5], most organizations and government institutions transitioned or accelerated their digital platforms to maintain essential services and product offerings [6]. The mass adoption of digital platforms by various stakeholders brought tremendous benefits and vulnerabilities. With the expansion of the digital landscape in modern society, sophisticated threats and attacks are on the rise, with phishing attacks being a prevalent and expensive vulnerability attack [7].

Phishing attacks are a menace to the modern digital world and a major cybersecurity threat [8-11]. Phishing attacks account for approximately 90% of data breaches and have become increasingly rampant in recent years, posing quite a huge threat to digital platform stakeholders [6, 12]. Phishing attacks are executed adopting deceptive tactics to trick targets into divulging sensitive information (e.g., login credentials or credit card information) or performing actions (e.g., downloading malware or visiting virus-infected websites) that

compromise the security of digital systems [13-14]. From spam or deceptive emails that impersonate trusted entities to personalized spear-phishing attacks that exploit specific individuals or organizations, the dangers of phishing attacks cannot be overemphasized. The effect of phishing attacks ranges from substantial financial loss and compromised infrastructure to loss of trust in digital communication channels [8, 12, 13, 15]. Therefore, it becomes imperative to develop and implement defenses against such malicious cybersecurity attacks.

To defend the modern digital society against various phishing attacks, countermeasures are constantly being developed to address the latter's multifaceted nature. Traditional security methods and tools to detect phishing websites, Short Message Service (SMS), and emails, such as blacklist, antivirus and antimalware software, and filter rules, are often inadequate, as attackers continuously devise new tactics to evade detection [3, 7, 9, 16, 17]. This led to the need for more sophisticated and effective methods and tools for detecting dynamic and zero-day phishing attacks. The advent of AI-based approaches to phishing attack detection has shown promise in this area better than that of the traditional methods [7, 8, 13]. Specifically, AI-based solutions analyze a wide range of data sources, including network traffic, email or SMS content, and even user behavior, to identify phishing attack patterns and anomalies. Several studies have explored the use of various AI techniques, ranging from machine learning, ensemble learning, deep learning, and natural language processing, to spot phishing attacks. Some AI techniques showed better results than others, which may also be due to the data, leaving room for improvements. Therefore, this study aims to complete a comprehensive survey that identifies AI-based ensemble methods to detect three forms of phishing attacks, website, email, and SMS.

This study significantly contributes to the knowledge of the field through the identification of credible and accessible datasets to develop three types of phishing attack detection models (i.e., website, email, and SMS), a survey of AI-based ensemble algorithms used to detect various phishing attacks, and a summary of the performance of AI-ensemble-based phishing attack detection models.

II. METHODS AND MATERIALS

This study focused on investigating AI ensemble methods for detecting phishing attacks and the variables for characterizing them. To complete this survey, various steps proposed in [18-21] were followed. These steps include (a) research question formulation, (b) inclusion and exclusion eligibility criteria setting, (c) information source and search strategy formulation, and (d) selection of studies.

A. Research Questions

This study utilized the Population, Intervention, Context, and Outcome (PICO) framework to formulate the right research questions. Table I provides the criteria for formulating the research questions. The following research questions were formulated:

Q1: What open and credible datasets are available for phishing attack detection?

Q2: Which AI-based ensemble methods are used for the detection of various phishing attacks?

Q3: What are the results and accuracies of these phishing attack detection models?

TABLE I. CRITERIA FOR RESEARCH QUESTIONS

PICO criteria	Description
Population	Phishing website, email, and SMS attacks
Intervention	AI ensemble learning algorithms
Context	Phishing attack detection
Outcome	Available open and credible datasets, phishing detection AI ensemble-based models, and model performance (metrics)

B. Eligibility Criteria

This survey included studies that (a) were published within the period 2019-2023, (b) from both academic journals and conference proceedings, (c) written in English, and (d) utilized AI meta or ensemble methods for website, email, and SMS phishing attack detection. More so, studies that (a) were published as a survey, systematic, scoping, narrative, traditional, or conceptual review, (b) used other AI methods, such as traditional machine learning and deep learning, (c) did not complete an experimental or empirical analysis, and (d) were not written in English were excluded.

C. Information Source and Search Strategy

To carry out a comprehensive survey to provide answers to the research questions, a search string was formulated containing combinations of different terms of keywords that encapsulate the objective of this survey. The search string contained "phishing" OR "phishing attack" OR "phishing website" OR "email phishing" OR "SMS phishing" OR "smishing" AND "prediction" OR "detection" AND "meta-algorithm" OR "ensemble methods". The search string was used by different combinations on five online repositories, namely ScienceDirect, IEEE Xplore, Springer, ACM Digital Library, and Google Scholar. Existing works published in these repositories since 2019 were identified.

D. Study Selection

The screening and selection process consisted of two stages. Initially, studies were evaluated based on the relevance of their titles and abstracts according to the eligibility criteria. Subsequently, a full-text assessment was performed to select the studies. In cases of uncertainty, a full-text evaluation was applied. Any disagreements among co-authors were resolved through consensus. Additionally, the EndNote X20 software was used to eliminate duplicates and organize all citations. Initially, 947 papers were identified. After the removal of duplicates, 549 papers remained. Subsequently, four hundred of them were excluded based on title and abstract screening. The full text of the remaining 82 articles was thoroughly evaluated. Among these, 45 did not meet the inclusion and exclusion criteria. Consequently, 37 studies were selected for this review. Figure 1 illustrates the screening and selection procedures.

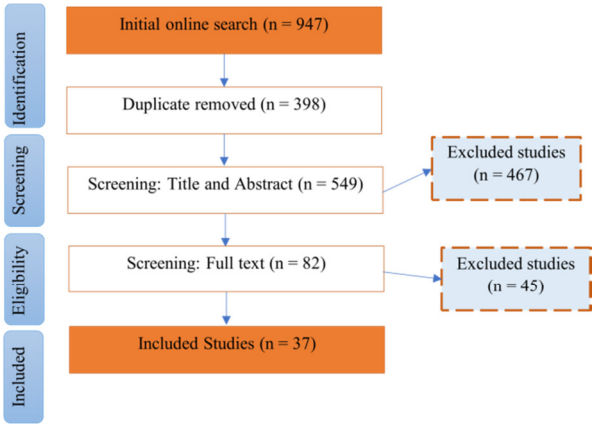

Fig. 1. Flowchart of article screening and selection.

III. RESULTS AND DISCUSSION

A. *Included Studies' Characteristics*

There has been a notable increase in the number of articles in recent years, suggesting a growing interest among scholars in using ensemble machine learning methods to detect various forms of phishing attacks. Figure 2 depicts this trend, with most of the articles included published between 2022 (n=13, 33.3%) and 2023 (n=11, 28%).

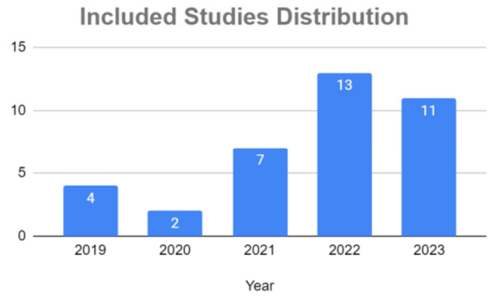

Fig. 2. Yearly distribution of published studies.

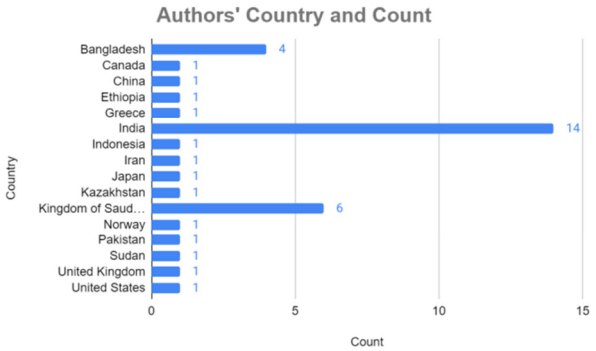

Fig. 3. Authors' country distribution.

Based on the first author's affiliation country, most of the studies originated in India (n = 14, 37.8%), the Kingdom of Saudi Arabia (n = 6, 16.2%), and Bangladesh (n = 4, 10.8%). In contrast, other countries contributed between 1 and 2 articles each (Figure 3). As observed in Figure 4, the included studies are mainly based on website phishing detection (n = 25, 67.6%) [22-46], while email (n = 6, 16.2%) [47-52], and SMS phishing (n = 6, 16.2%) [53-58] detection had the same distribution.

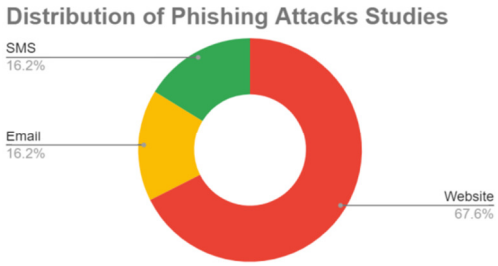

Fig. 4. Distribution of phishing attacks across published works.

B. *Datasets for Phishing Attack Detection*

For each type of phishing attack, various datasets were identified and employed in studies to complete the experimental analysis. The subsections below discuss the datasets used per attack.

1) *Phishing Website Detection*

The studies [22-46] utilized one or more phishing website datasets to develop ensemble-based models for website phishing detection. A website phishing dataset published in 2015 in the UC Irvine (UCI) machine learning repository (UCI 2015) [59] is the most deployed dataset, as 19 studies used it [22-27, 29-32, 34-36, 38-41, 45, 46]. The data consist of 11,055 instances, 6,157 legitimate and 4,898 phishing websites. 30 features characterize each instance along with one label feature. Another website phishing dataset, published in 2016 in the UCI machine learning repository [60] (UCI 2016), was applied in two studies [23, 25]. UCI 2016 contains 1,353 instances, 548 legitimate, 702 phishing, and 103 suspicious websites, characterized by 10 features.

Three other datasets [61-63] were used by other studies to develop ensemble-based phishing website detection models hosted at the Mendeley Data repository. One of the datasets, Mendeley 2018 [61], had 48 features and 10,000 (5,000 legitimate and phishing website) instances and was implemented in three studies [25, 29, 42]. The second dataset, Mendeley 2020 [54], has 111 features and is presented in two variants: "Mendeley 2020 Small" consisting of 58,645 (27,998 legitimate and 30,647 phishing) instances, and "Mendeley 2020 Full" consisting of 88,647 (58,000 legitimate and 30,647 phishing) instances. One study employed the Mendeley 2020 Small dataset [36], whereas four studies used the Mendeley 2020 Full dataset [28, 36, 43, 44]. There is another website phishing dataset, Mendeley 2021, utilized in [33], consisting of a total of 11,430 (50% legitimate and 50% phishing) website instances characterized by 87 features.

In [37] a dataset of 247,064 (149,991 legitimate and 97,073 phishing websites) was presented. Other examples include the study in [33] that curated data with 100,000 instances of legitimate and phishing websites characterized by 21 features, from PhishTank and OpenPhish for phishing websites and from Moz and the Canadian Institute of Cybersecurity for legitimate websites.

2) Phishing Email Detection

Six ensemble-based method studies [47-52] put into service one or more datasets to develop email phishing detection models. In [47], the Enrol1 dataset, accessed through the Kaggle online repository, was used. The dataset contained 5,975 emails, however, it is no longer accessible in the Kaggle URL cited in [47]. Studies [48, 49] implemented the UCI SpamBase dataset [64] that contained already extracted features from email content. The dataset consists of 4,601 emails characterized by 57 features already processed for phishing detection modeling. 2,788 instances are ham and 1,813 instances are spam. The HELPHED dataset employed in [49-50] consists of a hybrid feature set of 271 features (18 content-based and 253 text-based extracted by the Word2Vec method). The data comprise existing email spam datasets, such as the Enron corpus, SpamAssassin Public Corpus, the Nazario Phishing Corpus, and authors' mailboxes. The HELPHED dataset consists of 3,460 phishing and 32,051 benign emails from the period 2015-2021. In [51], the Enrol-spam dataset [65] was combined with a sample from the SpamAssassin dataset [66] into a single dataset. 6,047 SpamAssassin and 7,582 Enron-spam emails were merged into a total of 13,629 email messages. Content-based features were extracted from the email corpus through tokenization of emails, and all emails were transformed into vectors of tokens following the Term-Frequency-Inverse Document Frequency (TF-IDF) method.

3) Phishing SMS Detection

The studies [53-58] developed ensemble-based models to detect SMS phishing attacks using SMS-based messages. In [45], an SMS spam dataset containing 6000 instances, with 1000 spam and 5000 ham messages, was utilized. The methods for collection and preprocessing were not mentioned. In [57], Bengali text SMS messages were collected from phones and manually labeled spam or ham. The SMS messages were between 2020 and 2022, and about 250 of them were labeled ham, and about 300 were labeled spam. In [57], the TF-IDF count vectorizer (based on the bag-of-words approach) and N-gram methods were engaged for feature extraction and representation of SMS messages in a numerical structure, following the translation of Bengali text into English.

In [54-56], SMS phishing attack detection models were developed deploying the SMS spam collection dataset [67]. This dataset consists of 747 spam and 4,827 ham messages, a total of 5,574 SMS messages. In [54], features were extracted from the dataset adopting four techniques: GPT-3, TF-IDF, Word embedding, and BERT-based embedding. In [55], features from the SMS messages were extracted using the bag-of-words and TF-IDF techniques. In [44], the TF-IDF technique was utilized to extract features, adding the length of each SMS message as an additional feature. In [58], two SMS datasets were applied for model development. The first dataset was the SMS spam collection data, and the second dataset was the ExAIS_SMS [68]. The ExAIS_SMS data consists of 2,453 spam and 1,967 ham SMS messages, having a total of 4,420 SMS messages. The TF-IDF technique was implemented for feature extraction.

C. AI-based Ensemble Methods for Detecting Phishing Attacks

1) Phishing Website Detection

Fifteen ensemble methods, Voting, AdaBoost, Bagging, Gradient Boosting, Extra Trees, XGBoost, Multi-boost adaptive boost, Proposed Optimized Bagging Classifier, Light Gradient Boosting Machine (LightGBM), Random Forest, Stacking, Weighted Voting, Weighted soft voting, Histogram-Based Gradient Boosting, and Category Boosting, were used across the reviewed studies to develop phishing website detection models.

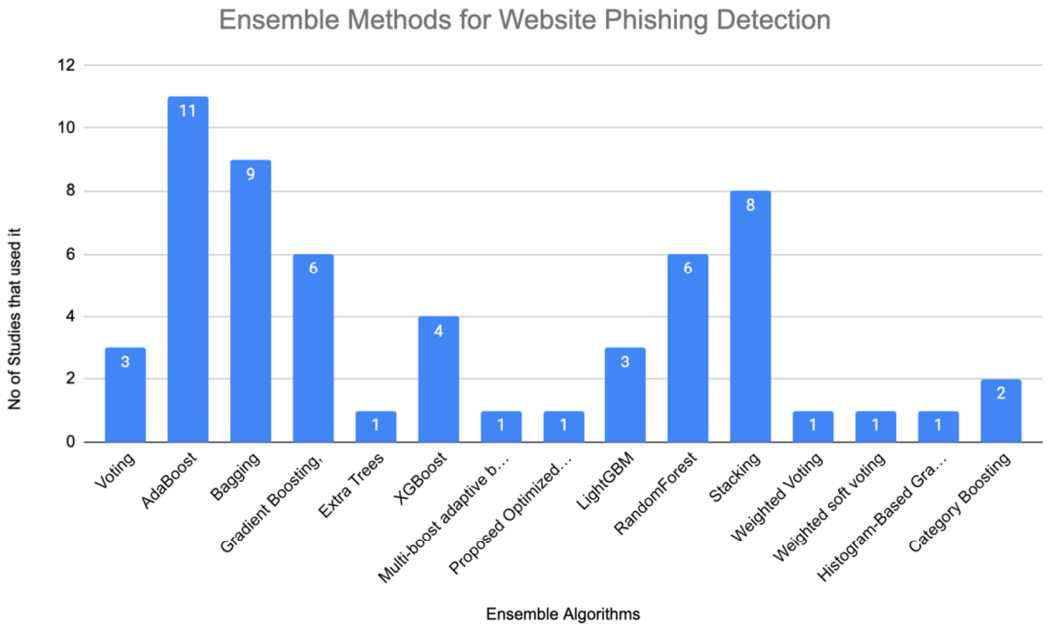

Fig. 5. Ensemble methods for phishing website detection.

The Adaboost ensemble algorithm was deployed in 11 studies [23-26, 29, 30, 34, 35, 38, 41, 43] for website phishing detection models and was the most popular ensemble method among others. Bagging-based ensemble phishing website detection models were employed in nine studies [23-26, 29, 32, 35, 37, 40], the Stacking ensemble method was addopted in eight studies [32, 35-37, 40, 42, 45], the Random Forest algorithm was applied in six studies [29, 30, 38, 39, 44, 46], and the Gradient Boosting algorithm was also used in six studies [24, 29, 35, 39, 43-44]. Other ensemble-based phishing website detection models were developed utilizing XGBoost [24, 27, 29, 31], LightGBM [29, 44], Voting [22, 24, 40], Category Boosting [43, 46], Extra Trees [24], Multi-boost Adaptive Boost [17], weighted Voting [33], weighted Soft-Voting [34], Histogram-based Gradient Boosting [44], and a customized optimized Bagging ensemble method [28].

2) Phishing Email Detection

From the reviewed studies, six ensemble methods, namely Soft voting, Stacking, Random Forest, AdaBoost, Gradient Boosting, Bagging, and Majority Voting, were utilized for email phishing detection models in various studies, as displayed in Figure 6. Both Adaboost and Stacking were the most employed ensemble methods. The AdaBoost ensemble method was used in three studies [47, 48, 52], and the Stacking ensemble method was deployed in three other studies [49-51]. The Majority-Voting ensemble was used in [47, 48], the Soft-Voting ensemble was implemented in [49, 50], the Bagging ensemble was put into service in [48], and both the Random Forest and Gradient Boosting ensemble methods were followed in [47].

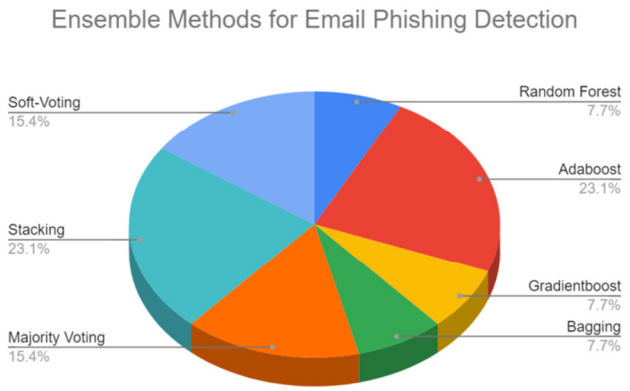

Fig. 6. Ensemble methods for phishing email detection.

3) Phishing SMS Detection

Nine existing ensemble methods were utilized for SMS phishing detection models by various studies. These ensemble methods are Weighted Voting [54], Majority Voting [53, 55], Random Forest [56], Gradient Boosting [56], Extra Trees [56], Bagging [57], Adaboost [57], XGBoost [57], and Stacking [57]. A newly proposed and developed ensemble method, which is based on a Custom Aggregation equation, was identified [50]. Among all ten ensemble methods, the Majority Vote ensemble is the most popular method for SMS phishing detection, as noticed in Figure 7.

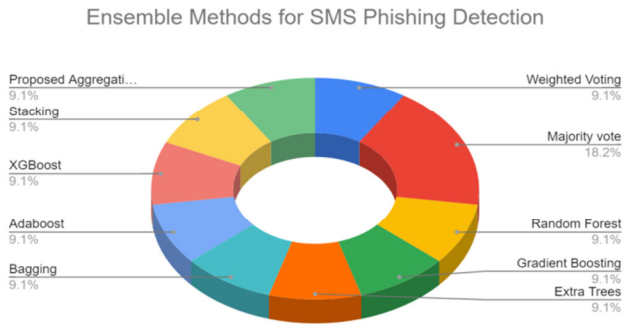

Fig. 7. Ensemble methods for phishing SMS detection.

D. Performance of Ensemble Methods for Phishing Attack Detection

For each type of phishing attack, the performance of the ensemble-based phishing attack models was identified. The following subsections discuss the methods used, per attack type. The performance of most ensemble-based models was evaluated utilizing Accuracy, Receiver Operator Characteristic - Area Under Curve (ROC-AUC), F-score, Recall, and Precision [69-70].

1) Phishing Website Detection

Based on the UCI 2016 dataset, Adaboost models achieved a maximum Accuracy of 89.73% (Logistic model tree base learner) [23] and a minimum Accuracy of 89.06% (Naive Bayes Tree base learner) [25]. Adaboost ensemble models, based on the UCI 2015 dataset, attained 97.42% (Logistic model tree base learner) [23] and 97.09% (Best-First Tree base learner) Accuracy [25, 70, 71]. The Decision Tree-based learner obtained 94.21% [24] and 94.3% [38] Accuracy among others. Bagging ensemble-based models for detecting phishing websites using the UCI 2015 dataset [23-26, 29, 32, 35, 37, 40], achieved a maximum Accuracy of 98.78%, 98% Precision, F-score and Recall of 99%, and 99.6% ROC-AUC.

Voting ensemble models for detecting phishing websites based on the UCI 2015 dataset attained a maximum Accuracy of 97.33% by assembling Random Forest and K-Nearest Neighbor algorithms [22]. However, the voting ensemble model of an Artificial Neural Network and Random Forest algorithms [22] had the highest ROC-AUC of 99.7%, maximum F-score of 97.6%, 98.3% Recall, and 97% Precision. Additionally, the Voting ensemble method obtained 97.15% Accuracy, 97% F-score, 98% Recall, and 97% Precision in [24], and 81.26% Accuracy, and 81.26% Recall in [40].

Other ensemble-based models for website phishing detection deploying the UCI 2015 dataset, such as Random Forest ensemble-based models [29, 30, 38, 39, 46], reached a maximum Accuracy of 97.47%. The Weighted Soft Voting ensemble-based model [34] attained a 95% F-score, Recall, and Precision. LightGBM ensemble-based models [29, 46] achieved a maximum Accuracy of 96.42%, and 95.30% F-score, Recall, and Precision. Stacking ensemble-based models [32, 35, 36, 40, 41, 45], accomplished a maximum Accuracy of 97.5% and an F-score of 96.54%. The Category Boosting ensemble model obtained 95.9% accuracy [46]. XGBoost ensemble-based models [24, 27, 29, 31] achieved a maximum

Accuracy of 99.18%, 99% ROC-AUC, 100% Recall, and 97% Precision.

The stacking ensemble model [36], using the Mendeley 2020 small dataset, achieved 96.5% Accuracy, 96.33% F-score, 96.25% Recall, and 96.42% Precision. Based on the Mendeley 2020 full dataset, the Adaboost ensemble method and the Gradient Boosting ensemble [43, 44] obtained a maximum accuracy of 99% and 98% ROC-AUC. The Category-Boosting ensemble-based model [43] on the same data attained 98% Accuracy and 96% ROC-AUC, while the Histogram-Based Gradient Boosting [44] model reached 96.54% Accuracy, 95.17% F-score, and 94.93% Recall and Precision. Based on the Mendeley 2021 dataset [42], Category boosting, Gradient boosting, Random Forest, Meta-category boosting and Hard Vote ensemble models achieved 97.36%, 97.27%, 96.37%, 97.18%, and 97.34% Accuracy, respectively.

2) Phishing Email Detection

Adaboost ensemble-based models for detecting phishing emails [48, 52] using the UCI Spambase dataset [64] accomplished a maximum accuracy of 94.41%. The majority voting ensemble model of Multilayered Perceptron, Naive Bayes, and Random Forest base learners [47] employing the Enrol1 dataset reached a maximum Accuracy of 97.25%, while the majority voting ensemble of Multinomial Naive Bayes, Support Vector Machine, and Random Forest base learners [48] deploying the UCI Spambase dataset achieved 98.5% Accuracy, 0.98 ROC-AUC, 98% F-score and Recall, and 97% Precision.

Stacking ensemble-based models fitted on the HELPHED dataset [65] obtained 99.09% Accuracy based on Decision Tree and Support Vector Machine base learners [41] and 99.07% Accuracy based on Decision Tree and k-Nearest Neighbor base learners [50]. Similarly, the Stacking ensemble model, fitted on the combined SpamAssassin [66] and EnronSpam [65] datasets and based on Logistic Regression as a meta-learner and Decision Tree, Gaussian Naive Bayes, k-Nearest Neighbor, and Adaboost base learners attained a 98.8% Accuracy. The Softing voting ensemble model reached 99.43% Accuracy, 97.14% ROC-AUC, 99.43% Precision and Recall, and 99.42% F-score [50]. The soft voting ensemble achieved 99.32% Accuracy, 97.02% ROC-AUC, and 99.32% F-score [49].

3) Phishing SMS Detection

The majority voting ensemble of MultiLayered Perceptron, Logistic regression, and Multinomial Naive Bayes base learners [40], via 10-fold cross-validation, resulted in a maximum Accuracy of 98.75%, Recall of 91.9%, and Precision of 100%. Meanwhile, the majority voting ensemble of Multinomial Naive Bayes, Logistic Regression, Support Vector Machine, Nearest Centroid, Extreme Gradient Boosting, K-Nearest Neighbor, and Perceptron base learners resulted in maximum accuracy of 98.11% and 98.91%. In [56], Random Forest, Gradient Boosting, and Extra Tree ensemble algorithms achieved 98.2%, 98.7%, and 98.2% accuracy and 93.2%, 95%, and 93.2% F-score, respectively. SMS phishing detection models based on Bagging, AdaBoost, XGBoost, and Stacking attained 75.8%, 79.76%, 83.87%, and 82.65% Accuracy, correspondingly [57]. The weighted-voting ensemble of Support Vector Machine, K-Nearest Neighbor, LightGBM, and Convolutional Neural Network reached 99.91% Accuracy [54], and a custom aggregation ensemble of Random Forest and Logistic Regression algorithms achieved 98.06% Accuracy [58].

TABLE II. TOP PERFORMING PHISHING ATTACK DETECTION MODELS DEVELOPED PER ATTACK PER DATASET

Attack Type	Study ref.	Dataset	Number of studies that used it	Ensemble Method	Base Learner	Accuracy	ROC	Year published
Website	[24]	UCI_2015	19	Bagging	Decision Tree	98.64%	Not reported	2022
	[25]	UCI_2016	2	Bagging	Naive Bayes Trees	90.61%	97.50%	2022
	[36]	Mendeley 2018	3	Stacking	XGBoost, Logistic Regression, Random Forest, Multi Layer Perceptron, and K-Nearest Neighbor	98.90%	Not reported	2022
	[43]	Mendeley 2020 (Full)	5	Adaboost	Decision Tree	99%	99	2022
Email	[49]	HELPED	2	Soft voting	Decision Tree and Support Vector Machine	99.43%	Not reported	2023
	[48]	UCI Spambase Dataset	2	Majority Voting	Multinomial Naive Bayes, Support Vector Machine, Random Forest	98.5	98	2019
SMS	[54]	SMS Spam Collection Dataset	3	Weighted Voting	Support Vector Machine, K-Nearest Neighbor, Light Gradient Boosting Machine, Convolutional Neural Network	99.91	Not reported	2023

IV. CONCLUSIONS

This study focused on the survey of AI-based ensemble methods used in the development of three phishing attack detection models (i.e., website, email, and SMS), within the 2019-2023 period. Through the review of relevant studies, seventeen (17) ensemble methods were identified: AdaBoost, Bagging, Gradient Boosting, Extra Trees, XGBoost, Multi-Boost Adaptive Boost, Proposed Optimized Bagging Classifier, LightGBM, Stacking, Weighted Voting, Weighted soft voting, Histogram-Based Gradient Boosting, Category Boosting, Random Forest, Majority Voting, Soft-Voting, and a custom Aggregation Method. Some studies developed new ensemble methods, while others optimized the existing ones. Table II presents a summary of the top-performing detection models per phishing attack per dataset.

Large, credible, and open-access datasets are available for developing and testing various phishing attack detection models. The data are available in Mendeley and UCI online repositories. Additionally, other repositories (e.g., PhishTank and the Canadian Institute of Cybersecurity) provide sources for obtaining both phishing and legitimate websites. Kaggle and GitHub also host some phishing data, specifically for SMS phishing. The Adaboost ensemble algorithm is the most popular among others for website phishing detection models. Both Stacking and Adaboost ensemble methods are popular for developing email phishing detection models. The Majority Voting ensemble methods are commonly used to develop SMS phishing detection models. The performance of AI-based ensemble methods for phishing attack detection is evaluated using Accuracy, ROC-AUC, F-score, Recall, and Precision. These performance metrics and open-access phishing datasets are available to benchmark various ensemble method phishing attack detection models. It is recommended that new AI-based ensemble methods should be developed to efficiently and effectively detect various forms of phishing attacks.

ACKNOWLEDGMENT

The authors gratefully acknowledge the approval and the support of this research study by the grant no. SCIA-2023-12-2176 from the Deanship of Scientific Research at Northern Border University, Arar, K.S.A.

CONFLICT OF INTEREST

The authors declare that they have no conflict of interest. The manuscript was written through the contributions of all authors. All authors have approved the final version of the manuscript.

REFERENCES

[1] S. Madakam, R. Ramaswamy, and S. Tripathi, "Internet of Things (IoT): A Literature Review," *Journal of Computer and Communications*, vol. 3, no. 5, pp. 164–173, May 2015, https://doi.org/10.4236/jcc.2015.35021.

[2] I. Mergel, N. Edelmann, and N. Haug, "Defining digital transformation: Results from expert interviews," *Government Information Quarterly*, vol. 36, no. 4, Oct. 2019, Art. no. 101385, https://doi.org/10.1016/j.giq.2019.06.002.

[3] P. Seuwou and V. F. Adegoke, "The Changing Global Landscape With Emerging Technologies and Their Implications for Smart Societies," in *Handbook of Research on 5G Networks and Advancements in Computing, Electronics, and Electrical Engineering*, IGI Global, 2021, pp. 402–423.

[4] S. Hussain, W. Guangju, R. M. S. Jafar, Z. Ilyas, G. Mustafa, and Y. Jianzhou, "Consumers' online information adoption behavior: Motives and antecedents of electronic word of mouth communications," *Computers in Human Behavior*, vol. 80, pp. 22–32, Mar. 2018, https://doi.org/10.1016/j.chb.2017.09.019.

[5] Y. K. Dwivedi *et al.*, "Impact of COVID-19 pandemic on information management research and practice: Transforming education, work and life," *International Journal of Information Management*, vol. 55, Dec. 2020, Art. no. 102211, https://doi.org/10.1016/j.ijinfomgt.2020.102211.

[6] R. Sujeetha, H. Das, T. Dhelawat, and M. Tanveer, "Cyber-Space and Its Menaces," in *2019 IEEE International Conference on System, Computation, Automation and Networking (ICSCAN)*, Pondicherry, India, Mar. 2019, https://doi.org/10.1109/ICSCAN.2019.8878848.

[7] A. Basit, M. Zafar, X. Liu, A. R. Javed, Z. Jalil, and K. Kifayat, "A comprehensive survey of AI-enabled phishing attacks detection techniques," *Telecommunication Systems*, vol. 76, no. 1, pp. 139–154, Jan. 2021, https://doi.org/10.1007/s11235-020-00733-2.

[8] A. Chakraborty, A. Biswas, and A. K. Khan, "Artificial Intelligence for Cybersecurity: Threats, Attacks and Mitigation," in *Artificial Intelligence for Societal Issues*, A. Biswas, V. B. Semwal, and D. Singh, Eds. Cham, Switzerland: Springer International Publishing, 2023, pp. 3–25.

[9] S. Garera, N. Provos, M. Chew, and A. D. Rubin, "A framework for detection and measurement of phishing attacks," in *Proceedings of the 2007 ACM workshop on Recurring malcode*, Alexandria, VA, USA, Aug. 2007, pp. 1–8, https://doi.org/10.1145/1314389.1314391.

[10] S. Nasiri, M. T. Sharabian, and M. Aajami, "Using Combined One-Time Password for Prevention of Phishing Attacks," *Engineering, Technology & Applied Science Research*, vol. 7, no. 6, pp. 2328–2333, Dec. 2017, https://doi.org/10.48084/etasr.1510.

[11] A. Darem, "Anti-Phishing Awareness Delivery Methods," *Engineering, Technology & Applied Science Research*, vol. 11, no. 6, pp. 7944–7949, Dec. 2021, https://doi.org/10.48084/etasr.4600.

[12] D. Aljeaid, A. Alzhrani, M. Alrougi, and O. Almalki, "Assessment of End-User Susceptibility to Cybersecurity Threats in Saudi Arabia by Simulating Phishing Attacks," *Information*, vol. 11, no. 12, Dec. 2020, Art. no. 547, https://doi.org/10.3390/info11120547.

[13] A. Sadiq *et al.*, "A review of phishing attacks and countermeasures for internet of things-based smart business applications in industry 4.0," *Human Behavior and Emerging Technologies*, vol. 3, no. 5, pp. 854–864, 2021, https://doi.org/10.1002/hbe2.301.

[14] K. Joshi *et al.*, "Machine-Learning Techniques for Predicting Phishing Attacks in Blockchain Networks: A Comparative Study," *Algorithms*, vol. 16, no. 8, Aug. 2023, Art. no. 366, https://doi.org/10.3390/a16080366.

[15] M. Z. Gashti, "Detection of Spam Email by Combining Harmony Search Algorithm and Decision Tree," *Engineering, Technology & Applied Science Research*, vol. 7, no. 3, pp. 1713–1718, Jun. 2017, https://doi.org/10.48084/etasr.1171.

[16] R. Yang, K. Zheng, B. Wu, C. Wu, and X. Wang, "Phishing Website Detection Based on Deep Convolutional Neural Network and Random Forest Ensemble Learning," *Sensors*, vol. 21, no. 24, Jan. 2021, Art. no. 8281, https://doi.org/10.3390/s21248281.

[17] Y. A. Alsariera, A. V. Elijah, and A. O. Balogun, "Phishing Website Detection: Forest by Penalizing Attributes Algorithm and Its Enhanced Variations," *Arabian Journal for Science and Engineering*, vol. 45, no. 12, pp. 10459–10470, Dec. 2020, https://doi.org/10.1007/s13369-020-04802-1.

[18] C. Romero and S. Ventura, "Educational Data Mining: A Review of the State of the Art," *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 40, no. 6, pp. 601–618, Aug. 2010, https://doi.org/10.1109/TSMCC.2010.2053532.

[19] Y. Baashar *et al.*, "Customer relationship management systems (CRMS) in the healthcare environment: A systematic literature review," *Computer Standards & Interfaces*, vol. 71, Aug. 2020, Art. no. 103442, https://doi.org/10.1016/j.csi.2020.103442.

[20] Y. A. Alsariera, Y. Baashar, G. Alkawsi, A. Mustafa, A. A. Alkahtani, and N. Ali, "Assessment and Evaluation of Different Machine Learning Algorithms for Predicting Student Performance," *Computational intelligence and neuroscience*, vol. 2022, Jan. 2022, Art. no. 4151487, https://doi.org/10.1155/2022/4151487.

[21] Y. Baashar *et al.*, "Toward Predicting Student's Academic Performance Using Artificial Neural Networks (ANNs)," *Applied Sciences*, vol. 12, no. 3, Jan. 2022, Art. no. 1289, https://doi.org/10.3390/app12031289.

[22] A. Basit, M. Zafar, A. R. Javed, and Z. Jalil, "A Novel Ensemble Machine Learning Method to Detect Phishing Attack," in *2020 IEEE 23rd International Multitopic Conference (INMIC)*, Bahawalpur, Pakistan, Nov. 2020, https://doi.org/10.1109/INMIC50486.2020.9318210.

[23] V. E. Adeyemo, A. O. Balogun, H. A. Mojeed, N. O. Akande, and K. S. Adewole, "Ensemble-Based Logistic Model Trees for Website Phishing Detection," in *Advances in Cyber Security*, Penang, Malaysia, 2021, pp. 627–641, https://doi.org/10.1007/978-981-33-6835-4_41.

[24] A. Awasthi and N. Goel, "Phishing website prediction using base and ensemble classifier techniques with cross-validation," *Cybersecurity*, vol. 5, no. 1, Nov. 2022, Art. no. 22, https://doi.org/10.1186/s42400-022-00126-9.

[25] Y. A. Alsariera, A. O. Balogun, V. E. Adeyemo, O. H. Tarawneh, and H. A. Mojeed, "Intelligent tree-based ensemble approaches for phishing website detection," *Journal of Engineering Science and Technology*, vol. 17, no. 1, pp. 563–582, 2022.

[26] H. Agrawal and R. R. Singh, "An Ensemble Approach for Detecting Phishing Attacks," *International Journal of Computer Sciences and Engineering*, vol. 9, no. 7, pp. 53–59, Jul. 2021, https://doi.org/10.26438/ijcse/v9i7.5359.

[27] J. Gu and H. Xu, "An Ensemble Method for Phishing Websites Detection Based on XGBoost," in *2022 14th International Conference on Computer Research and Development (ICCRD)*, Shenzhen, China, Jan. 2022, pp. 214–219, https://doi.org/10.1109/ICCRD54409.2022.9730579.

[28] P. Ponnusamy and P. Dhandayudam, "An Optimized Bagging Learning with Ensemble Feature Selection Method for URL Phishing Detection," *Journal of Electrical Engineering & Technology*, vol. 19, no. 3, pp. 1881–1889, Mar. 2024, https://doi.org/10.1007/s42835-023-01680-z.

[29] M. Al-Sarem *et al.*, "An Optimized Stacking Ensemble Model for Phishing Websites Detection," *Electronics*, vol. 10, no. 11, Jan. 2021, Art. no. 1285, https://doi.org/10.3390/electronics10111285.

[30] S. Menaka, J. Harshika, S. Philip, R. John, N. Bharathiraja, and S. Murugesan, "Analysing the Accuracy of Detecting Phishing Websites using Ensemble Methods in Machine Learning," in *2023 Third International Conference on Artificial Intelligence and Smart Energy (ICAIS)*, Coimbatore, India, Feb. 2023, pp. 1251–1256, https://doi.org/10.1109/ICAIS56108.2023.10073834.

[31] A. Das, F. I. Alam, S. Sharmin, and R. Uddin, "Boosting Guided Probabilistic Ensemble-based Approach For Phishing Website Detection," in *2022 International Conference on Innovations in Science, Engineering and Technology (ICISET)*, Chittagong, Bangladesh, Feb. 2022, pp. 402–407, https://doi.org/10.1109/ICISET54810.2022.9775819.

[32] Y. Chandra and A. Jana, "Improvement in Phishing Websites Detection Using Meta Classifiers," in *2019 6th International Conference on Computing for Sustainable Global Development (INDIACom)*, New Delhi, India, Mar. 2019, pp. 637–641, [Online]. Available: https://ieeexplore.ieee.org/abstract/document/8991353.

[33] A. Maini, N. Kakwani, R. B, S. M K, and B. R, "Improving the Performance of Semantic-Based Phishing Detection System Through Ensemble Learning Method," in *2021 IEEE Mysore Sub Section International Conference (MysuruCon)*, Hassan, India, Oct. 2021, pp. 463–469, https://doi.org/10.1109/MysuruCon52639.2021.9641614.

[34] A. Taha, "Intelligent Ensemble Learning Approach for Phishing Website Detection Based on Weighted Soft Voting," *Mathematics*, vol. 9, no. 21, Jan. 2021, Art. no. 2799, https://doi.org/10.3390/math9212799.

[35] A. F. Nugraha and L. Rahman, "Meta-Algorithms for Improving Classification Performance in the Web-phishing Detection Process," in *2019 4th International Conference on Information Technology, Information Systems and Electrical Engineering (ICITISEE)*, Yogyakarta, Indonesia, Nov. 2019, pp. 271–275, https://doi.org/10.1109/ICITISEE48480.2019.9003952.

[36] L. R. Kalabarige, R. S. Rao, A. Abraham, and L. A. Gabralla, "Multilayer Stacked Ensemble Learning Model to Detect Phishing Websites," *IEEE Access*, vol. 10, pp. 79543–79552, 2022, https://doi.org/10.1109/ACCESS.2022.3194672.

[37] D. M. Linh, H. D. Hung, H. M. Chau, Q. S. Vu, and T.-N. Tran, "Real-time phishing detection using deep learning methods by extensions," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 14, no. 3, pp. 3021–3035, Jun. 2024, https://doi.org/10.11591/ijece.v14i3.pp3021-3035.

[38] A. Soni and J. Tiwari, "Phishing Website Detection Using Ensemble Learning," *International Journal of Emerging Trends in Engineering Research*, vol. 11, no. 1, pp. 17–20, Jan. 2023, https://doi.org/10.30534/ijeter/2023/031112023.

[39] Z. Ghaleb Al-Mekhlafi *et al.*, "Phishing websites detection by using optimized stacking ensemble model," *Computer Systems Science and Engineering*, vol. 41, no. 1, pp. 109–125, 2022, https://doi.org/10.32604/csse.2022.020414.

[40] F. Hossain, L. Islam, and M. N. Uddin, "PhishRescue: A Stacked Ensemble Model to Identify Phishing Website Using Lexical Features," in *2022 5th International Conference of Computer and Informatics Engineering (IC2IE)*, Jakarta, Indonesia, Sep. 2022, pp. 342–347, https://doi.org/10.1109/IC2IE56416.2022.9970179.

[41] M. K. Pandey, M. K. Singh, S. Pal, and B. B. Tiwari, "Prediction of phishing websites using machine learning," *Spatial Information Research*, vol. 31, no. 2, pp. 157–166, Apr. 2023, https://doi.org/10.1007/s41324-022-00489-8.

[42] K. Adane, B. Beyene, and M. Abebe, "Single and Hybrid-Ensemble Learning-Based Phishing Website Detection: Examining Impacts of Varied Nature Datasets and Informative Feature Selection Technique," *Digital Threats: Research and Practice*, vol. 4, no. 3, Jul. 2023, Art. no. 46, https://doi.org/10.1145/3611392.

[43] D. Kaibassova, M. Nurtay, A. Tau, and M. Kissina, "Solving the Problem of Detecting Phishing Websites Using Ensemble Learning Models," *Scientific Journal of Astana IT University*, vol. 12, no. 12, pp. 55–64, Dec. 2022, https://doi.org/10.37943/12OYRS4391.

[44] Y. Wei and Y. Sekiya, "Sufficiency of Ensemble Machine Learning Methods for Phishing Websites Detection," *IEEE Access*, vol. 10, pp. 124103–124113, 2022, https://doi.org/10.1109/ACCESS.2022.3224781.

[45] Z. G. Al-Mekhlafi and B. A. Mohammed, "Using Genetic Algorithms to Optimized Stacking Ensemble Model for Phishing Websites Detection," in *Advances in Cyber Security*, Penang, Malaysia, 2021, pp. 447–456, https://doi.org/10.1007/978-981-16-8059-5_27.

[46] M. Khatun, M. A. I. Mozumder, Md. N. H. Polash, Md. R. Hasan, K. Ahammad, and Md. S. Shaiham, "An Approach to Detect Phishing Websites with Features Selection Method and Ensemble Learning," *International Journal of Advanced Computer Science and Applications*, vol. 13, no. 8, pp. 768–775, 2022, https://doi.org/10.14569/IJACSA.2022.0130888.

[47] A. K. Shrivas, A. K. Dewangan, S. M. Ghosh, and D. Singh, "Development of Proposed Ensemble Model for Spam e-mail Classification," *Information Technology and Control*, vol. 50, no. 3, Sep. 2021, https://doi.org/10.5755/j01.itc.50.3.27349.

[48] S. Suryawanshi, A. Goswami, and P. Patil, "Email Spam Detection : An Empirical Comparative Study of Different ML and Ensemble Classifiers," in *2019 IEEE 9th International Conference on Advanced Computing (IACC)*, Tiruchirappalli, India, Sep. 2019, pp. 69–74, https://doi.org/10.1109/IACC48062.2019.8971582.

[49] Q. Qi, Z. Wang, Y. Xu, Y. Fang, and C. Wang, "Enhancing Phishing Email Detection through Ensemble Learning and Undersampling," *Applied Sciences*, vol. 13, no. 15, Jan. 2023, Art. no. 8756, https://doi.org/10.3390/app13158756.

[50] P. Bountakas and C. Xenakis, "HELPHED: Hybrid Ensemble Learning PHishing Email Detection," *Journal of Network and Computer Applications*, vol. 210, Jan. 2023, Art. no. 103545, https://doi.org/10.1016/j.jnca.2022.103545.

[51] M. Adnan, M. O. Imam, M. F. Javed, and I. Murtza, "Improving spam email classification accuracy using ensemble techniques: a stacking approach," *International Journal of Information Security*, vol. 23, no. 1, pp. 505–517, Feb. 2024, https://doi.org/10.1007/s10207-023-00756-1.

[52] D. M. Ablel-Rheem, A. O. Ibrahim, S. Kasim, A. A. Almazroi, and M. A. Ismail, "Hybrid feature selection and ensemble learning method for spam email classification," *International Journal of Advanced Trends in Computer Science and Engineering*, vol. 9, no. 1.4 Special Issue, pp. 217–223, 2020, https://doi.org/10.30534/ijatcse/2020/3291.42020.

[53] A. Mahabub, M. I. Mahmud, and M. F. Hossain, "A Robust System for Message Filtering Using an Ensemble Machine Learning Supervised Approach," *ICIC Express Letters*, vol. 10, no. 9, pp. 805–811, 2019, https://doi.org/10.24507/icicelb.10.09.805.

[54] A. Ghourabi and M. Alohaly, "Enhancing Spam Message Classification and Detection Using Transformer-Based Embedding and Ensemble Learning," *Sensors*, vol. 23, no. 8, Jan. 2023, Art. no. 3861, https://doi.org/10.3390/s23083861.

[55] J. Fattahi and M. Mejri, "SpaML: a Bimodal Ensemble Learning Spam Detector based on NLP Techniques," in *2021 IEEE 5th International Conference on Cryptography, Security and Privacy (CSP)*, Zhuhai, China, Jan. 2021, pp. 107–112, https://doi.org/10.1109/CSP51677.2021.9357595.

[56] N. Sharma, "A Methodological Study of SMS Spam Classification Using Machine Learning Algorithms," in *2022 2nd International Conference on Intelligent Technologies (CONIT)*, Hubli, India, Jun. 2022, https://doi.org/10.1109/CONIT55038.2022.9848171.

[57] A. Al Maruf, A. Al Numan, Md. M. Haque, T. T. Jidney, and Z. Aung, "Ensemble Approach to Classify Spam SMS from Bengali Text," in *Advances in Computing and Data Sciences*, Kolkata, India, 2023, pp. 440–453, https://doi.org/10.1007/978-3-031-37940-6_36.

[58] S. Hosseinpour and H. Shakibian, "An Ensemble Learning Approach for SMS Spam Detection," in *2023 9th International Conference on Web Research (ICWR)*, Tehran, Iran, May 2023, pp. 125–128, https://doi.org/10.1109/ICWR57742.2023.10139070.

[59] R. Mohammad and L. McCluskey, "Phishing Websites.", UC Irvine Machine Learning Repository, 2012, https://doi.org/10.24432/C51W2X.

[60] N. Abdelhamid, "Website Phishing.", UC Irvine Machine Learning Repository, 2014, https://doi.org/10.24432/C5B301.

[61] C. L. Tan, "Phishing Dataset for Machine Learning: Feature Evaluation.", Mendeley Data, 2018, https://doi.org/10.17632/h3cgnj8hft.1.

[62] G. Vrbančič, "Phishing Websites Dataset." Mendeley Data, 2020, https://doi.org/10.17632/72ptz43s9v.1.

[63] A. Hannousse and S. Yahiouche, "Web page phishing detection." Mendeley Data, 2021, https://doi.org/10.17632/c2gw7fy2j4.3.

[64] M. Hopkins, E. Reeber, G. Forman, and J. Suermondt, "Spambase." UC Irvine Machine Learning Repository, 1999, https://doi.org/10.24432/C53G6X.

[65] P. Bountakas, "HELPED - Email Spam Dataset." 2021.

[66] W. W. Cohen, "Enron Email Dataset." 2015, [Online]. Available: https://www.cs.cmu.edu/~enron/.

[67] "Apache SpamAssassin." https://spamassassin.apache.org/.

[68] "SMS Spam Collection Dataset." [Online]. Available: https://www.kaggle.com/datasets/uciml/sms-spam-collection-dataset.

[69] AbayomiAlli, "SMS Spam Dataset." 2023, [Online]. Available: https://github.com/AbayomiAlli/SMS-Spam-Dataset.

[70] Y. A. Alsariera, V. E. Adeyemo, A. O. Balogun, and A. K. Alazzawi, "AI Meta-Learners and Extra-Trees Algorithm for the Detection of Phishing Websites," *IEEE Access*, vol. 8, pp. 142532–142542, 2020, https://doi.org/10.1109/ACCESS.2020.3013699.

[71] Y. A. Alsariera, "Detecting Generic Network Intrusion Attacks using Tree-based Machine Learning Methods," *International Journal of Advanced Computer Science and Applications*, vol. 12, no. 2, 2021, https://doi.org/10.14569/IJACSA.2021.0120275.