Phishing Detection Techniques: Exploring Machine Learning and Deep Learning Models

Hadi S. Hadi and Ahmed J. Obaid*
Faculty of Computer Science and Mathematics, University of Kufa, Iraq
Email: hadis.alhasan@student.uokufa.edu.iq (H.S.H.), ahmedj.aljanaby@uokufa.edu.iq (A.J.O.)

Manuscript received April 5, 2025; revised May 25, 2025; accepted June 21, 2025
*Corresponding author

Abstract—**The number of Internet users has increased significantly in recent years, driven by the growing popularity of online education, e-commerce, and other digital services. E-commerce, in particular, has seen significant growth due to increasing consumer demand for a convenient and secure online shopping experience. The COVID-19 pandemic has significantly sped up the uptake of e-commerce, changing consumer habits and propelling online transactions at an extraordinary rate. Nevertheless, this swift digital shift has heightened vulnerability to cyber threats, resulting in a significant rise in phishing attacks targeting the theft of confidential user data. In this article, we explore the use of Machine Learning (ML) and Deep Learning (DL) methods for detecting phishing, emphasizing conventional models, ensemble techniques, and hybrid systems. We analyze the key obstacles in this domain, such as data imbalance, significant computational expenses, and the challenges of real-time applications. Regarding research, this research emphasizes the potential of hybrid models and advanced methods to enhance the accuracy, efficiency, and scalability of Phishing systems. This result emphasizes the urgent need for a reliable, adaptive, and flexible detection system in order to express the growing risk of personal and organizational security in the digital development environment and to fight the increase in phishing attack.**

Index Terms—**phishing detection, machine learning, deep learning, ensemble models, hybrid approaches**

I. Introduction

In recent years, the field of e-commerce has witnessed significant development, due to the increase in customer demand for online shopping, which provides security, speed, and convenience. During the COVID-19 pandemic, the growth of e-commerce has significantly accelerated, as many users have relied on conducting their services online [1]. The expansion of digital markets has concurrently increased vulnerability to cyber threats, particularly through the proliferation of phishing attacks that exploit weaknesses associated with the rise in online shopping, causing numerous data breaches and credential fraud, with damages and financial losses estimated in the millions of dollars [2–5]. Phishing operations have caused financial losses for many large projects [6].

Phishing is viewed as a major cybercrime that threatens security in technology, with the attacker seeking to obtain sensitive user data or information using various techniques, such as dispatching fraudulent emails or URLs that mimic authentic websites. In recent years, these attacks have escalated markedly, posing an important risk to internet users. The attacker replicates a genuine website of a business or organization and distributes it through email or social media platforms, causing numerous users to click on these links and fall victim [7]. About 50% to 80% of unlawful sites were restricted after facing a monetary setback [8]. During the second quarter of 2024, a total of 877,536 phishing attempts were documented. While the count of documented attacks stayed consistent, the methods employed in phishing have greatly varied. Attackers have progressively begun using new techniques like phone phishing (vishing) and SMS phishing (smishing) to reach customers in financial services and payment sectors, showcasing the advancement of phishing strategies over time [9]

Users must understand attackers' methods and become familiar with anti-phishing techniques to protect themselves. However, many users still lack sufficient awareness of these types of attacks [10]. The ability to recognize phishing websites within a reasonable timeframe is of high importance for these websites [11], [12].

Traditional methods for detecting phishing websites involve updating antivirus databases with suspicious IP addresses and URLs, commonly referred to as the “blacklist” method. However, the attackers are able to evade this by using sophisticated techniques such as URL obfuscation, which disguises malicious links as legitimate. They also rely on quick camouflage techniques, deploying automated systems that create fake websites, algorithms that churn out new URLs endlessly. The major drawback of the blacklist method is its incapacity to identify the phishing attack in real-time [13]. Inspired by that, the majority of these techniques fail at detecting whether a new fraudulent URL is legitimate.

The features for URL and Email are generally analyzed through Machine Learning (ML) and Deep Learning (DL) techniques in the identification of attempted phishing attacks. These techniques have demonstrated high detection rates and adaptability to the evolving landscape of cyber threats. Ensemble-based algorithms and conventional ML models such as Random Forest (RF) have been successfully utilized in phishing detection due to their robustness and potential for handling large datasets [14]. Similarly, DL models such as Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) networks have proven successful in learning complex

doi: 10.18178/ijeetc.14.5.274-281

patterns from sequential data, making them essential tools for building scalable phishing detection systems [15]. These techniques provide powerful tools for detecting phishing attacks in real-time environments.

This paper provides a comprehensive survey of recent phishing detection techniques using Machine Learning (ML) and Deep Learning (DL) models. It summarizes previous works, discusses important techniques, and identifies gaps in the literature. We explore both classical/modern ML/DL techniques for real-time phishing detection. Lastly, we suggest possible research paths from the issues raised in the study and provide recommendations on how to further improve the progress of robust, adaptive, and scalable models that are able to fight against increasingly sophisticated phishing attacks.

II. Related Work

Many studies have investigated different methods for phishing detection in recent years, including ML and DL techniques. The ability of ensemble models, such as stacking and boosting, to combine the results of several base models has made them especially popular. This improves prediction accuracy and robustness. To ensure the best performance of these models, hyperparameters have also been fine-tuned using parameter optimization techniques including particle swarm optimization (PSO) and Genetic Algorithms (GA) [16–19].

RF had a maximum accuracy of 97.52% when Tamal *et al*. (2024) used the optimal feature vectorization algorithm (OFVA) in conjunction with 15 ML classifiers [20]. With an impressive accuracy of 98.69%, Othman and Hassan (2022) presented an ensemble stacking model that integrated classifiers such as RF, decision tree, and k-Nearest Neighbors (KNN) [10]. A stacking ensemble model combining Recursive Feature Elimination (RFE) and Multilayer Perceptron (MLP) was presented by Newaz and Haq (2023). It showed 97.48% accuracy but had issues with computational overhead [21]. Similar to this, Jaber *et al*. (2022) used a hybrid strategy that combined a Variational Autoencoder (VAE) with the Multi-Objective Grey Wolf Optimizer (MOGWO) for feature selection. This approach achieved a competitive accuracy of 97.49%, however, scalability was still an issue [22]. With the use of XGBoost, RF, and KNN, Kalabarige *et al*. (2022) created a multi-layer stacked ensemble learning model that achieved 98.43% accuracy; nevertheless, issues with dataset imbalance were encountered [23]. In a comparison analysis, Tubyte and Paulauskaite-Taraseviciene (2021) discovered that RF was 95.00% accurate, particularly when handling URL parameters for phishing detection [24]. Together with DL models like CNN and LSTM, Wei and Sekiya (2022) investigated ensemble techniques like AdaBoost, Gradient Boosting (GB), and LightGBM (LGBM). They were able to achieve an accuracy of 96.94% but faced difficulties with dataset imbalance [25]. Akour *et al*. (2021) addressed lexical and host-based features and concentrated on conventional ML techniques like Support Vector Machine (SVM) and Naive Bayes (NB). They found that SVM was the most successful, with an accuracy of 96.30% [26]. Puli Raju *et al*. (2024) introduced an optimized feature selection approach using extra trees classifier (ET) and classical methods like Chi-square, Information Gain, and correlation coefficient, achieving an accuracy of 96.95%. However, they noted challenges with high false positives, computational demands, and scalability [27]. Al-Sarem *et al*. (2021) introduced an optimized stacking ensemble model that achieved 97.39% accuracy. The model integrated three primary classifiers Genetic Algorithm (GA)–GB, GA–eXtreme Gradient Boosting (XGB), and GA–Bagging with SVM as the meta-learner. While the approach demonstrated strong phishing detection capabilities, it faced challenges with high computational costs and scalability, limiting its real-world application [17].

III. Methodology

The methodology section explains the general methodology used in phishing detection studies, with a focus on ML and DL techniques. The methodology includes several main stages, from data collection and preprocessing to model development and evaluation using evaluation metrics. Fig. 1 illustrates the flowchart used in phishing detection research.

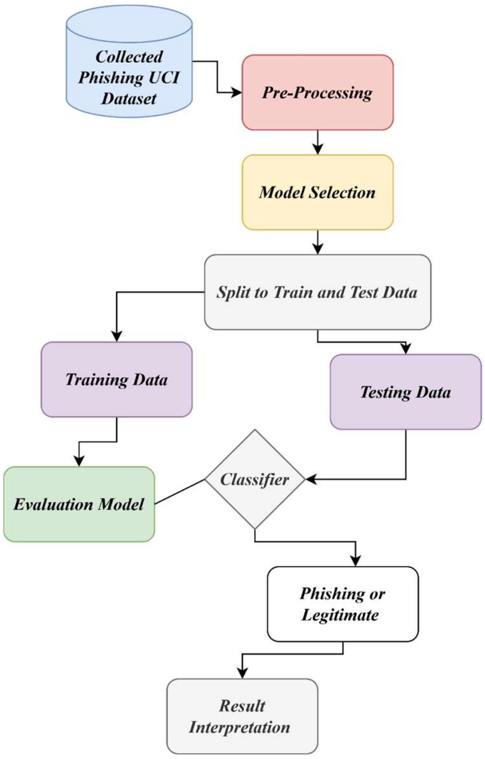

Fig. 1. The flowchart of the basic steps of the general methodology for developing a phishing detection model.

A. Data Collection

Data collection is one of the most important basic steps in building a model because the quality and diversity of data greatly affects the accuracy of the model during the training process. Table I presents the two most important datasets used in phishing detection research and outlines their main characteristics.

TABLE I: OVERVIEW OF DATASETS USED FOR PHISHING DETECTION

Suggested name for the dataset	Source	Total Instances	Phishing Instances	Legitimate Instances	Features
Dataset-1	[28]	88,647	30,647	58,000	111
Dataset-2	[29]	247,950	119,409	128,541	41

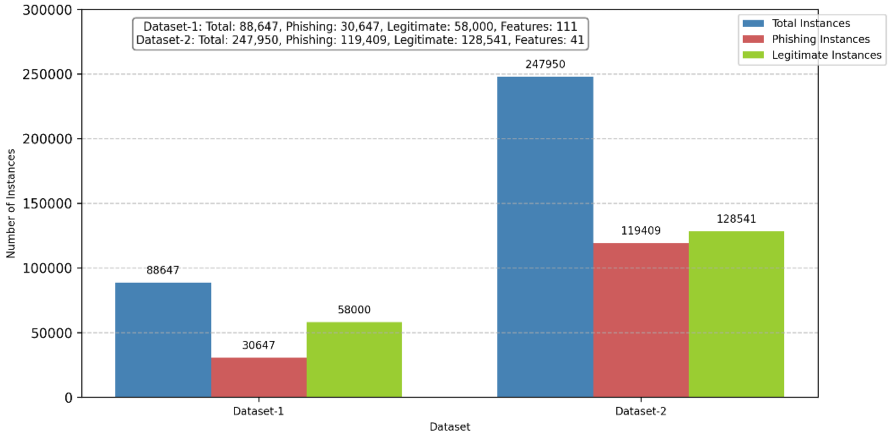

Fig. 2. Instance comparison: Dataset-1 vs Dataset-2.

Dataset-1 is characterized by a large number of features, including 111, which gives high accuracy to the model. Dataset-1 is the most widely used dataset in previous research. Dataset-2 is the newest and largest dataset, consisting of 247,950 datapoints and have 41 features. This large size makes it suitable for DL models. These datasets were also selected for their comprehensive features and accurate preprocessing. Fig. 2 shows the comparison between Dataset-1 and Dataset-2.

One of the most difficult challenges in data collection is the continuous updating of datasets and the development of phishing techniques. Also, the imbalance of the dataset is a challenge because it affects the quality of the system.

Each dataset was compiled differently. In Dataset-1, features were extracted based on the URL syntax. In Dataset-2, features were extracted from the URL syntax and domain metadata. These features are considered important because they powerfully help in detecting fake URLs.

B. Data Preprocessing

Preprocessing is considered one of the most important steps as it aims to clean and transform data into a suitable format to ensure optimal performance when fed into ML and DL algorithms. It plays a crucial role in reducing noise and handling missing values, thereby improving the model's overall quality. Since URL data comes in raw, unstructured form and cannot be directly entered into ML models, it is necessary to convert it to a structured, tabular format that includes features and data points.

The steps for converting this unstructured data are as follows:

- Feature Extraction: Extracting numerical features from the URL (such as URL length, number of slashes, count of special characters).
- Metadata Extraction: Extracting features from the domain infrastructure (e.g., domain age via WHOIS lookup, number of nameservers, SSL certificate validity).
- Encoding: Converting categorical features to numeric values. For example, we extract a numeric column with binary values (0, 1) representing the URL protocol type, whether https or http.

Fig. 3 shows the URL structure and the parameters used to extract the features. Fig. 4 shows the parts of the domain metadata and some of the features extracted in Dataset-1.

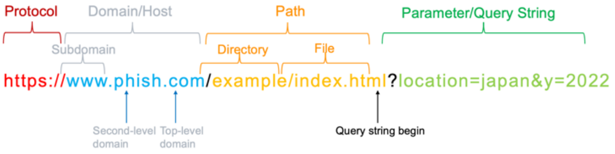

Fig. 3. URL structure.

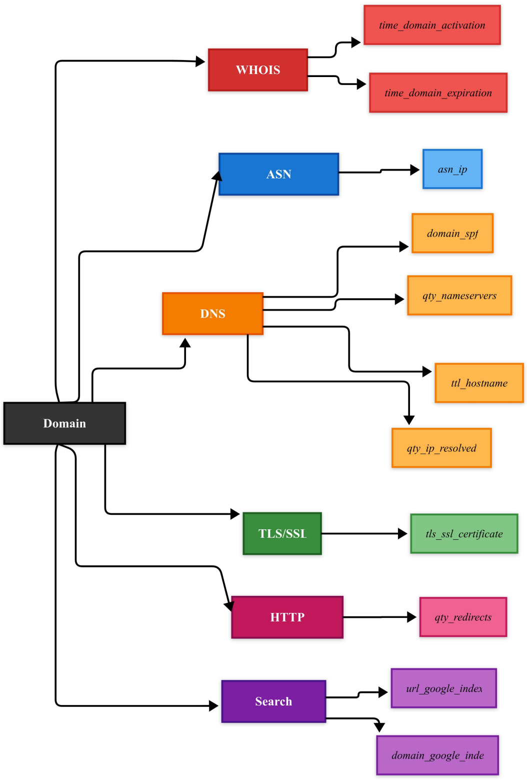

Fig. 4. Parts of the domain metadata.

The basic steps of preprocessing are as follows.

After converting the data into structured data, it enters the data cleaning phase, which includes removing duplicate entries, fixing corrupted records, and handling missing values using techniques such as mean imputation or interpolation.

Data can now be fed into ML, but sometimes issues with data dimensionality (variability in the range of values) can arise, which can impair the system's accuracy. For example, URL length, which has a wide range, is often due to phishing sites being long compared to legitimate sites, which are short.

To address these issues, specific techniques exist, including normalization and standardization.

Normalization is one of the essential steps in the preprocessing stage as it ensures that the model is trained without bias because it sets the data range from a specific range. There are two methods for normalization, one of which is the Min-Max method, which sets all values between 0-1. This process reduces the variance between values, which leads to an increase in the quality and accuracy of the model.

$$X' = \frac{X - X_{\min}}{X_{\max} - X_{\min}} \tag{1}$$

where X' refer to the normalized value and X is the original value, $X_{\min}$ is the minimum value, and $X_{\max}$ is the maximum value of the dataset.

Standardization of this process focuses on distributing data to a mean and a standard deviation of one, where the mean is zero and the standard deviation is one. This is very important in some situations because it reduces bias and reduces computational efficiency, such as the PCA algorithm.

$$Z = \frac{X - \mu}{\sigma} \tag{2}$$

where Z is the standardized value, X is the original value, μ is the mean of the dataset, and σ is the standard deviation of the dataset.

Standardization is typically applied when the dataset has significant variation, as some algorithms are sensitive to feature scaling differences.

C. Feature Selection

The technique of selecting a subset of the most relevant features in a dataset for a problem is called feature selection. Feature selection helps ML and DL algorithms to learn more efficiently and effectively by reducing memory usage and time complexity while keeping features with the greatest impact on the class label.

In feature selection, features that do not influence the class label are removed either automatically or based on a defined threshold. The methods used include:

- Filter methods: Use statistical measures (e.g., Chi-square, Information Gain, ANOVA) to evaluate each feature's impact.
- Wrapper methods: Evaluate various subsets of features using models (e.g., RFE).
- Embedded methods: Select features during the training process (e.g., RF, which can rank feature importance).

D. Model Selection

ML and DL techniques play a very important role in solving many problems in most scientific disciplines. Through these techniques, a model can be built to detect phishing, and its advantages include its ability to learn patterns, predict, and classify. There is a difference between ML and DL in extracting features. In ML, features are extracted manually or using techniques and then entered into machine learning models, unlike what is found in DL, which has the advantage of extracting features automatically, which makes it suitable for large and complex data sets.

Many ML algorithms have been used in phishing detection studies, such as RF, logistic regression (LR), GB, etc. These models rely on URL structural features, such as URL length and the number of characters in a subdomain, etc. [16]. Ensemble models such as RF, XGB, and ET have shown better performance because they reduce overfitting by combining multiple classifiers and selecting the best [17]. DL models are also excellent but require big data for their pattern-learning ability.

The application of DL models is also important in phishing detection because they excel at learning complex patterns from data, despite challenges such as the need for large datasets and high computational costs. Nonetheless, ML and DL remain some of the most powerful tools for phishing detection. There are many methods, including hybrid approaches, that improve accuracy and reduce false positives. Integrating these methods with ML or DL yields robust and scalable solutions that keep pace with evolving technological demands.

Several ML models were evaluated based on key performance metrics: accuracy, precision, recall, and F1 score to select the most suitable classifier for phishing detection. These metrics will be explained in more detail later. However, in summary, these metrics provide a comprehensive assessment of the model's performance. Fig. 5 shows the comparative performance of the best classifiers (from the related work analysis) applied to both Dataset-1 and Dataset-2 (see next page).

Among the evaluated models, ensemble-based methods such as stacking and RF showed excellent performance. In the discussion section, we will analyze the results with challenges and solutions.

E. Evaluation Metrics

Evaluation metrics are important steps to building any ML or DL model and these metrics help determine the system's quality and accuracy in prediction and classification and enabling researchers to identify the best model for use. Table II presents the evaluation metrics along with their descriptions and formulas.

Some challenges include the impact of imbalanced data on model accuracy and the difficulty of achieving an optimal balance between Precision and Recall.

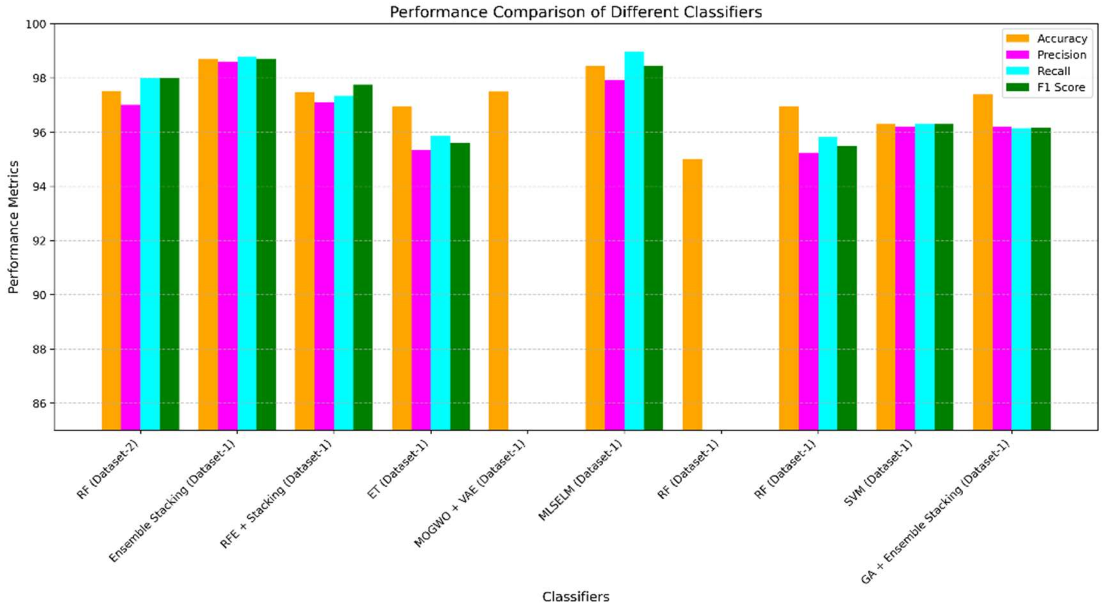

Fig. 5. Performance of best classifiers for Datasets 1 and 2.

TABLE II: SUMMARY OF EVALUATION METRICS

Metric	Description	Formula
Accuracy	Correct predictions percentage over all samples.	$(\text{TP} + \text{TN})/(\text{TP} + \text{TN} + \text{FP} + \text{FN})$
Precision	True positives over predicted positives.	$(\text{TP})/(\text{TP} + \text{FP})$
Recall	True positives over actual positives.	$(\text{TP})/(\text{TP} + \text{FN})$
F1-Score	Harmonic mean of Precision and Recall.	$(\text{Precision}\times\text{Recall}\times 2)/(\text{Precision} + \text{Recall})$
ROC	Trade-off between True positive rate and false positive rate.	TPR (True Positive Rate): $\text{TPR} = (\text{TP})/(\text{TP} + \text{FN})$ FPR (False Positive Rate): $\text{FPR} = (\text{FP})/(\text{FP} + \text{TN})$
AUC	Area under the ROC curve.	$\text{AUC} = \int_0^1 \text{TPR}(\text{FPR})d(\text{FPR})$

This methodology review presents a structured approach to phishing detection research, starting with data collection and preprocessing, selecting the best ML or DL model for the specific problem, and finally evaluating the model by testing it on new data. This helps us identify weaknesses and challenges we may face. Table II shows the most important evaluation metrics used in ML fields.

IV. DISCUSSION

In this section, we will discuss previous research studies and the findings of researchers on phishing detection, which were reviewed in the Related Work section. Phishing detection has emerged as a critical and challenging topic for researchers because it threatens the security of users and organizations alike. Many researchers have used ML or DL models to detect phishing attacks by analyzing URL structures and classifying them as fraudulent or legitimate. These studies have demonstrated the effectiveness of ensemble models such as RF and XGB in detection accuracy and speed compared to individual models such as KNN and LR. This is because the ensemble model provides more than one classification and selects the best one.

Additionally, DL models like CNN and FNN have the ability to analyze sequential data, making them ideal for URLs analysis. Since Dataset-2 contains features extracted from the URL structure, features can be easily extracted from it when new links become available and merged with the original dataset, making it suitable for DL algorithms. Because these models require large datasets and significant computational resources, they can be applied to real-time phishing detection systems. Table III provides a summary of related work, focusing on the methodologies, datasets, challenges, and results from previous studies on phishing detection.

TABLE III: SUMMARY OF RELATED WORK

Ref.	Year	Dataset	Methodology	Main Findings	Challenges
[20]	2024	Dataset-2	OFVA + RF + Hyperparameter Tuning	RF achieved an accuracy of 97.52%	High computational demands, preprocessing requirements, limited use of DL
[21]	2023	Dataset-1	RFE + Stacking (MLP Meta)	The stacking ensemble model achieved an accuracy of 97.48%	Computational overhead, heuristic dependency, real-time adaptability challenges
[27]	2024	Dataset-1	ET, RF, LGBM, GB, DT, AdaBoost (ADA), Ridge Classifier, linear discriminant analysis (LDA), LR, Quadratic Discriminant Analysis (QDA), KNN, NB, SVM, Dummy Classifier	ET achieved an accuracy of 96.95%	High false positives, computational demands, scalability issues.

[22]	2022	Dataset-1	MOGWO + VAE Hybrid	MOGWO-VAE achieved an accuracy of 97.49%	Feature dependence, MOGWO cost, limited real-time application.
[23]	2022	Dataset-1	Stacked Ensemble (XGB + RF + MLP + KNN + LR + XGB Meta)	MLSELM achieved an accuracy of 98.43%	Stacking overhead, imbalanced datasets without resampling
[24]	2021	Dataset-1	LR, LDA, DT, SVM, RF	RF achieved an accuracy of 95.00%	Feature extraction demands, dataset balancing for real-world conditions
[25]	2022	Dataset-1	Ensemble Methods + DL (FCNN, LSTM, CNN + ML: SVM, NB, KNN, LR)	RF achieved an accuracy of 96.94%	Dataset imbalance, DL cost, adapting features for zero-day attacks
[26]	2021	Dataset-1	Lexical + Host-based Features (SVM, KNN, NB, LR)	SVM achieved an accuracy of 96.30%	Dataset imbalance, phishing technique diversity, public dataset reliance.
[17]	2021	Dataset-1	GA + Stacking (GB, XGB, Bagging + SVM Meta)	SVM achieved an accuracy of 97.39%.	GA cost, public dataset reliance, limited DL exploration

TABLE IV: BEST CLASSIFIER PERFORMANCE FOR DATASET 2

Paper/year	Classifier	Accuracy	Precision	Recall	F1 Score
[20]/2024	(RF)	97.50%	97.00%	98.00%	98.00%

TABLE V: BEST CLASSIFIER PERFORMANCE FOR DATASET 1

Paper/year	Classifier	Accuracy	Precision	Recall	F1 Score
[10]/2022	Ensemble Stacking with LR Meta	98.69%	98.59%	98.80%	98.69%
[21]/2023	RFE + Stacking (MLP Meta)	97.48%	97.11%	97.34%	97.75%
[27]/2024	ET	96.95%	95.35%	95.87%	95.60%
[22]/2022	MOGWO + VAE	97.49%	-	-	-
[23]/2022	MLSELM	98.43%	97.93%	98.96%	98.44%
[24]/2021	RF	95.00%	-	-	-
[25]/2022	RF	96.94%	95.24%	95.83%	95.49%
[26]/2021	SVM	96.30%	96.20%	96.30%	96.30%
[17]/2021	GA + Ensemble Stacking with SVM Meta	97.39%.	96.20%	96.14%	96.17%

Table IV highlights the best-performing algorithms along with their evaluation metrics for Dataset-2, showcasing the top models that achieved the highest performance. Similarly, Table V presents the best-performing algorithms and their evaluation metrics for Dataset-1. Together, these summaries facilitate a clear comparison of methodologies and their impact across different datasets.

A. Challenges and Research Gaps

Most researchers also lack techniques for merging datasets from different sources, for several reasons. These include the fact that a column may not be merged with another column despite similar feature functions. For example, a link length feature might be called "URL_length" in one group and "length_of_URL" in another, even though they both refer to the same concept. There are also problems with value variation, which can cause discrepancies in the number of samples. However, several proposed solutions exist, including the Feature Unification Table, which helps unify the names of features that have the same function but with a different name. This makes it easier to merge different datasets. As for value variation, there are algorithms that address these problems, including LGBM, which handles missing values excellently without the need to fill in the missing values using traditional techniques or imputation, which can reduce the accuracy of predictions. LGBM handles missing values when building a tree. It tries to send missing values to the right branch and then to the left, choosing the direction that yields the best accuracy, without you having to fill in or delete values.

Current research gaps have focused on the importance of extracting features to enhance accuracy, as well as the importance of selecting influential features through some statistical methods that were not mentioned, such as ANOVA, which analyzes the extent of the influence of each feature on the class label.

The limited focus on parameter tuning and cross-validation methods hinders the generalization of the model, as many studies have ignored their importance. It is also necessary to address the challenges associated with false positives and negatives. In addition, the combination of ML and DL models remains underexplored despite its power. As the imbalance of data is recognized, there are some solutions that have not been used, such as oversampling and undersampling to balance datasets and data augmentation to increase data size.

The reviewed studies show significant progress in phishing detection using ML and DL models. However, some challenges such as imbalanced datasets, computational costs, and feature extraction techniques highlight the need for more novel approaches such as hybrid and data processing techniques.

Addressing these gaps through unified preprocessing and the use of modern ML algorithms with parameter optimization methods and hybrid frameworks, whether ML or neural networks, will significantly contribute to real-time phishing detection.

V. CONCLUSION

Internet security is compromised by criminal activities through malicious websites. Hence, a suspicious website detection framework is needed to prevent users from visiting these malicious URLs, but this is a challenging task nowadays as malicious content on web pages changes from time to time. There are many studies on phishing attacks, but they are not able to achieve completely accurate results. As technology penetrates every aspect of our lives, attackers try to find new ways to steal data. The world needs to improve security methods and predict and prevent financial losses and data theft. In this paper, we discuss previous works on developing phishing detection models and their results. Through analysis, ML and DL models have made great progress in phishing detection. RF and GB models have proven their effectiveness and good performance. DL models such as CNN and LSTM have shown high efficiency in analyzing complex patterns. However, this field faces significant challenges such as lack of large data, data imbalance, high computational costs, and difficulty in implementing models in real time. This paper highlights the importance of developing innovative approaches, such as hybrid models and feature selection techniques, to overcome these challenges. In the future, with future studies, we plan to focus on feature selection methods, ensemble model, and hybrid models as they have achieved high performance. Moreover, focus on updating data to implement deep learning to detect phishing sites as it needs large data for its ability to learn patterns from complex data. Moreover, we need further study to detect phishing attacks via mobile devices. Nowadays, “smartphones” are a very popular technological offspring. These smartphones are also a common point where attackers converge, where phishing attacks occur. Mobile users prefer to read emails on their phones right away. Therefore, it is imperative to find new solutions that will detect phishing attacks that occur on mobile devices.

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

Hadi Salah conducted the literature review, collected and curated the data, and drafted the manuscript under Prof. Dr. Ahmed J. Obaid, who proposed the research topic, provided guidance and academic oversight throughout the study, and reviewed and edited the manuscript. Prof. Obaid also checked all the results and verified the implementation of the proposed system. Prof. Obaid also maintained the editing and checking of the final version of this work.

References

[1] L. S. Alaimo, M. Fiore, and A. Galati, “How the COVID-19 pandemic is changing online food shopping: The model of TAM and new variables in the equation,” *International Journal of Information Management*, vol. 60, 102594, 2021. https://doi.org/10.1016/j.ijinfomgt.2021.102594

[2] D. Hillman, Y. Harel, and E. Toch, “Evaluating Organizational Phishing Awareness Training on an Enterprise Scale,” *Computers & Security*, vol. 132, 103364, 2023. doi: 10.1016/j.cose.2023.103364

[3] X. Liu, S. F. Ahmad, M. K. Anser *et al.*, “Cyber security threats: A never-ending challenge for E-commerce,” *Frontiers in Psychology*, vol. 13, 927398, 2022. doi: 10.3389/fpsyg.2022.927398

[4] Y. Wang, Z. Li, T. Wu *et al.*, “An empirical study: Automated subdomain takeover threat detection,” in *Proc. 2021 IEEE International Conference on Cyber Situational Awareness*, 2021. doi: 10.1109/CyberSA52016.2021.9478220

[5] S. S. Roy, U. Karanjit, and S. Nilizadeh, “A large-scale analysis of phishing websites hosted on free web hosting domains,” arXiv, arXiv:2212.02563, 2024. doi: 10.48550/arXiv.2212.02563

[6] A. O. Balogun, K. S. Adewole, M. O. Raheem *et al.*, “Improving the phishing website detection using empirical analysis of Function Tree and its variants,” *Heliyon*, vol. 7, no. 7, e07437, Jun. 29, 2021. doi: 10.1016/j.heliyon.2021.e07437

[7] B. B. Gupta, N. A. G. Arachchilage, and K. E. Psannis, “Defending against phishing attacks: Taxonomy of methods, current issues and future directions,” *Telecommun. Syst.*, vol. 67, pp. 247–267, 2018. doi: 10.1007/s11235-017-0334-z

[8] A. K. Jain and B. B. Gupta, “A machine learning-based approach for phishing detection using hyperlinks information,” *J. Ambient Intell. Humaniz. Comput.*, vol. 10, pp. 2015–2028, 2019. doi: 10.1007/s12652-018-0798-z

[9] Anti-Phishing Working Group. (2024). Phishing Activity Trends Report. Washington, DC, USA. [Online]. Available: https://apwg.org/trendsreports/

[10] M. Othman and H. Hassan, “An empirical study towards an automatic phishing attack detection using ensemble stacking model,” *Future Comput. Inform. J.*, vol. 7, no. 1, 2022. doi: 10.54623/fue.fcij.7.1.1

[11] R. M. Mohammad, F. Thabtah, and L. McCluskey, “Tutorial and critical analysis of phishing websites methods,” *Comput. Sci. Rev.*, vol. 17, pp. 1-24, 2015. doi: 10.1016/j.cosrev.2015.04.001

[12] R. M. Mohammad, F. Thabtah, and L. McCluskey, “Predicting phishing websites based on self-structuring neural network,” *Neural Comput. Appl.*, vol. 25, pp. 443–458, 2013. doi: 10.1007/s00521-013-1490-z

[13] A. Basit, M. Zafar, and X. Liu, “A comprehensive survey of AI-enabled phishing attacks detection techniques,” *Telecommun. Syst.*, vol. 76, pp. 139–154, 2021. doi: 10.1007/s11235-020-00733-2

[14] C. Catal, G. Giray, B. Tekinerdogan, *et al.*, “Applications of deep learning for phishing detection: A systematic literature review,” *Knowl. Inf. Syst.*, vol. 64, pp. 1457–1500, 2022. doi: 10.1007/s10115-022-01672-x

[15] L. Tang and Q. H. Mahmoud, “A survey of machine learning-based solutions for phishing website detection,” *Mach. Learn. Knowl. Extr.*, vol. 3, pp. 672–694, 2021. doi: 10.3390/make3030034

[16] W. Ali and A. A. Abdullah, “Hybrid intelligent phishing website prediction using deep neural networks with genetic algorithm-based feature selection and weighting,” *IET Inf. Secur.*, vol. 13, pp. 659–669, 2019. doi: 10.1049/iet-ifs.2019.0006

[17] M. Al-Sarem, F. Saeed, Z. G. Al-Mekhlafi *et al.*, “An optimized stacking ensemble model for phishing websites detection,” *Electronics*, vol. 10, 2021. doi: 10.3390/electronics10111285

[18] W. Ali and S. Malebary, “Particle swarm optimization-based feature weighting for improving intelligent phishing website detection,” *IEEE Access*, vol. 8, pp. 1–1, 2020. doi: 10.1109/ACCESS.2020.3003569

[19] T. Nagunwa, “Comparative analysis of nature-inspired metaheuristic techniques for optimizing phishing website detection,” *Analytics*, vol. 3, pp. 344–367, 2024. doi: 10.3390/analytics3030019

[20] M. A. Tamal, M. K. Islam, T. Bhuiyan, A. Sattar, and N. U. Prince, “Unveiling suspicious phishing attacks: Enhancing detection with an optimal feature vectorization algorithm and supervised machine learning,” *Front. Comput. Sci.*, vol. 6, 1428013, 2024. doi: 10.3389/fcomp.2024.1428013

[21] A. Newaz and F. S. Haq, “A sophisticated framework for the accurate detection of phishing websites,” *arXiv preprint* arXiv:2403.09735, 2024. doi: 10.48550/arXiv.2403.09735

[22] A. Jaber, L. Fritsch, and H. Haugerud, “Improving phishing detection with the grey wolf optimizer,” in *Proc. Int. Conf. Electr. Eng. Inform. (ICEE)*, Oslo, Norway, 2022. doi: 10.1109/ICEIC54506.2022.9748592

[23] L. Kalabarige, S. Routhu, and A. Abraham, “MLSELM: Multi-layer stacked ensemble learning model to detect phishing websites,” *IEEE Access*, vol. 10, 2022. doi: 10.1109/ACCESS.2022.3194672

[24] M. Tubyte and A. Paulauskaite-Taraseviciene, “Research on phishing email detection based on URL parameters using machine learning algorithms,” M.S. thesis, Kaunas Univ. Technol., Kaunas, Lithuania, 2021.

[25] Y. Wei and Y. Sekiya, “Sufficiency of ensemble machine learning methods for phishing websites detection,” *IEEE Access*, vol. 10, 2022. doi: 10.1109/ACCESS.2022.3224781

[26] I. Akour, N. Alnazzawi, A. Aburayya, R. Alfaisal, and S. Salloum, “Using classical machine learning for phishing websites detection from URLs,” *Int. J. Inf. Decis. Sci.*, vol. 24, pp. 1–9, 2021

[27] P. Raju, K. V. N. Aditya, and P. Karthik, “Optimized and relevant features for enhanced phishing detection accuracy,” *Int. J. Innov. Res. Technol.*, vol. 11, no. 3, pp. 1749-1755, Aug. 2024.

[28] G. Vrbančič, I. Fister Jr., and V. Podgorelec, “Datasets for phishing websites detection,” *Data in Brief*, vol. 33, 2020. doi: 10.1016/j.dib.2020.106438

[29] M. Tamal, “Phishing detection dataset,” *Mendeley Data*, vol. 1, 2023. doi: 10.17632/6tm2d6sz7p.1

Hadi S. Hadi received a B.Sc. degree in computer science from the University of Kufa, Najaf, Iraq. He is currently pursuing his M.Sc. degree in computer science at the Faculty of Computer Science and Mathematics, University of Kufa, Najaf, Iraq. His research interests include artificial intelligence and data science.

Ahmed J. Obaid is an assistant professor in the Department of Computer Science at the University of Kufa, Iraq, with over 14 years of experience. He holds a bachelor’s in computer science and information systems, a master’s in computer science engineering, and a Ph.D. degree in web mining. Dr. Obaid specializes in web mining techniques, image processing, and medical health applications, serving as an associate editor for the Brazilian Journal of Operations & Production Management and editorial board member for several international journals. He has authored over 180 research articles, edited 18 books, and actively participates in international conferences as a keynote speaker and session chair.