Optimizing Phishing Threat Detection: A Comprehensive Study of Advanced Bagging Techniques and Optimization Algorithms in Machine Learning

Samer Kadhim Jawad*, Satea H. Alnajjar**

* Department of Computer Engineering, College of Engineering, Al-Iraqia University, Iraq
Email: samer.k.jawad@aliraqia.edu.iq
https://orcid.org/0009-0002-4311-2216

** Department of Network Engineering, College of Engineering, Al-Iraqia University, Iraq
Email: sateaahn@gmail.com
https://orcid.org/0000-0002-2828-3167

Abstract

Bagging constitutes a prominent ensemble learning technique in contemporary machine learning. With this process, various instances of the base model are trained using various subsets of the training data that are extracted by bootstrapping. The resulting models are then aggregated, often through voting in a classification problem, to enhance performance and predictive power. Recent advances in bagging techniques include variants such as Random Forests, which introduce additional randomness by selecting a random subset of features in each partition and boosting algorithms that iteratively optimize the model's focus on misclassified instances. The efficacy of these strategies in enhancing the generality and adaptability of machine learning models has been impressive. There are many studies that confirm the ability of ensemble learning models to detect phishing attacks. However, the techniques used by these models that have enhanced their detection capabilities have not been highlighted. The study reached important results in terms of accuracy of up to 97% through the random forest model and the Particle swarm optimization algorithm. This study seeks to contribute to advancing the field of cybersecurity by providing a robust and interpretable machine learning-based classifier that can be integrated into a framework to detect phishing attacks by distinguishing between legitimate URLs and phishing URLs.

Keywords- Bagging Techniques, Ensemble learning, Particle swarm optimization algorithm, Phishing, Random Forests.

I. Introduction

The constant development of cyberthreats in the digital age presents a serious threat to the protection of private and sensitive data. Phishing attacks are particularly noteworthy among these risks because they are sneaky strategies used by attackers to trick people and obtain unauthorized access to sensitive data [1]. The complexity of phishing techniques has increased despite ongoing improvements in cybersecurity safeguards, necessitating a proactive and flexible approach to detection.

Phishing attacks have evolved beyond simplistic methods, incorporating elements of social engineering and multiple forms of deception to gain access to victims' information [2]. Nevertheless, a phishing assault consists of four main steps [3]. The first step involves the attacker building a website that appears authentic and targets the victim. The second is to create a phishing URL that directs the victim to that fake website. The third is to obtain sensitive information from the victim such as username, password, email, and the like. Finally, the attacker collects this information after the victim provides it to use it according to the attacker's goals and intentions.

In contrast, a lot of work is done to identify these assaults and try to lessen their impact by organizations, businesses, and researchers who specialize in cybersecurity. This is accomplished by offering tools and systems that recognize phishing attempts [4]. But because these systems can't keep up with how often these attacks evolve, attackers have been reported to be able to go around and escape these measures.

Accurate detection of phishing attacks lies in proposing a framework that detects these client-side attacks from the first stage. Phishing attempts generally begin when the attacker sends the victim a URL via email or other communication platforms and are detected by

http://doi.org/10.58564/IJSER.3.1.2024.146

analyzing and summarizing the sequential stages of the attack as shown in Figure 1. From this analysis, we can consider it the basis of the presented study, which is to detect phishing based on the distinction between phishing URLs and legitimate URLs. Thus, this work is a framework that proactively detects before users become victims of this type of attack.

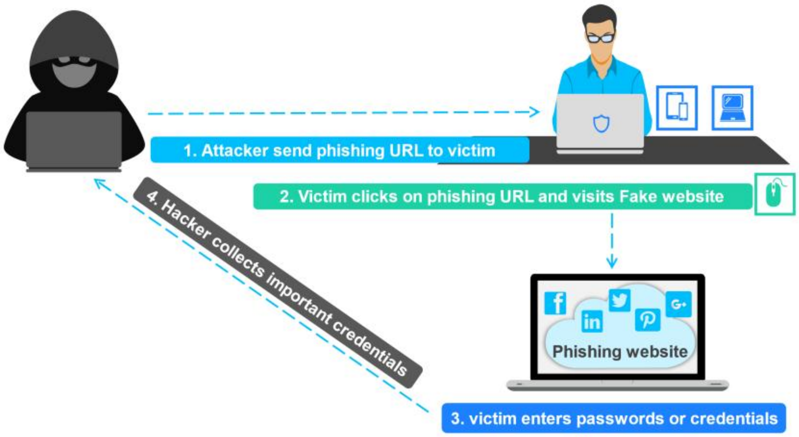
1. Attacker send phishing URL to victim
2. Victim clicks on phishing URL and visits Fake website
Phishing website
3. victim enters passwords or credentials
4. Hacker collects important credentials

Figure 1 , Sequence of stages of a phishing attack

Machine learning stands out among other methods with its high ability to distinguish between phishing URLs and legitimate URLs through the binary classification task built into its models [5]. Machine learning can be considered a powerful modeling and prediction tool if it is fed with a good dataset that is closely related to phishing attacks.

Therefore, the field of machine learning is broad and rapidly growing, and aims to address modern problems and experiences such as phishing attacks [6]. Therefore, utilizing state-of-the-art machine learning frameworks and methods to lessen overfitting and prediction bias and increase result accuracy is one of the study's objectives.

One of the newest machine learning methods for ensemble learning is bootstrap aggregating (bagging). Bagging is a potent method that reduces overfitting, boosts forecast accuracy, and stabilizes the model. It enhances performance overall by merging forecasts from several models. The bagging method has become very popular and is frequently utilized in contemporary machine learning applications. Random Forest is a dedicated decision tree-based bagging application [7]. Because of its versatility, capacity to handle complex data, robust prediction, and insight into feature relevance, Random Forest has become a standard in many domains.

In addition to contemporary machine learning models and methods for phishing attack detection, optimization algorithms are required. Particularly, more comprehensive algorithms with several optimization tasks are needed in order to handle unbalanced data sets, enhance the model's features, and modify the model's parameters. In order to increase the diversity among the trees in the random forest model and produce a more reliable ensemble model, the particle swarm optimization (PSO) algorithm is crucial in this situation [8].

This research has two objectives. Firstly, it aims to offer significant perspectives on how cybersecurity is changing and to give practitioners practical methods for fortifying their defenses. The second goal is to contribute to providing a roadmap to not only mitigate the risks associated with phishing attacks but also to advance the broader discourse around the intersection of machine learning and cybersecurity.

The main contribution of this study is to present the random forest model that relies on ensemble learning, but in an interpretable way and not a mysterious black box with configurations that are difficult to understand, by highlighting the techniques it uses that distinguish it and make it a model with high predictive accuracy, in addition to providing a comprehensive algorithm for improving Hyperparameters of the model which reduces the time and effort that researchers spend searching for the ideal algorithm that matches this model and the nature of the binary classification problem of legitimate URLs and fraudulent URLs. This, in turn, contributes to strengthening the field of cybersecurity in developing phishing attack detection system.

This paper is organized according to the structure that includes the following: Section II includes the relevant literature in the field of phishing detection. The third section is then devoted to the background of cyberattacks and the developments that led to phishing attacks Machine learning was covered in the fourth chapter, the study's methodology was covered in the fifth, and the study's findings and their commentary were covered in the sixth. The final section presents the study's conclusion and future work.

II. LITERATURE REVIEW

The study in this section will present current methods for phishing detection. And revealing studies that used bagging techniques and optimization algorithms in cybersecurity, as shown in the TABLE 1.

TABLE 1. LITERATURE REVIEW

Authors	Paper	publication	Description
M.Alanezi [9]	Phishing Detection Methods: A Review	Technium 2021	The writer of this research paper break down the most effective techniques for identifying phishing into two categories in order to present a review of these techniques. The first is detection through traditional methods, which includes systems that rely on awareness and education, as well as systems that rely on the white and black list approach. The second section is detection based on unconventional methods, which include systems that rely on machine learning and other modern systems.
R.Zieni et al. [10]	Phishing or Not Phishing? A Survey on the Detection of Phishing Websites	IEEE 2023	The researchers in this study conducted a recent survey that highlights the importance of the role of machine learning in detecting phishing attacks by dividing detection methods using machine learning into two main parts, the first is detection based on URLs and the second is detection based on page content by fetching HTML.
Z.Zhou and C.Zhang [11]	Phishing website identification based on double weight random forest	IEEE 2022	In this study, the authors highlighted the importance of the bagging technique in increasing the accuracy of results in detecting phishing attacks by using the random forest model, and reached significant results of more than 96%.
M.H.Alkawaz et al.[12]	Identification and Analysis of Phishing Website based on Machine Learning Methods	IEEE 2022	In this research, the decision tree model—an individual learning model—and the random forest model—an ensemble learning model that employs the parallel bagging technique—were simply compared by the authors. The investigation showed that the random forest model outperforms the decision tree model in terms of effectiveness.
K.Subashini and V. Narmatha [13]	Phishing Website Detection using Hyper-parameter Optimization and Comparison of Cross-validation in Machine Learning Based Solution	IEEE 2023	This study looks at the value of optimization strategies in raising the accuracy of machine learning model outputs by adjusting hyperparameters. The outcomes demonstrated the contribution of the optimization procedure in raising the accuracy of the random forest model's results to above 97%.

This research paper aims to provide a proactive approach to detect phishing attacks from the beginning on the client side by distinguishing between URLs that reach users from different social media platforms, using the latest and most accurate machine learning models and techniques, and developing and improving the performance of these models and techniques by integrating Optimization algorithms.

III. BACKGROUND

The term "cyber-attacks" is used in association with the digital domain; it refers to harmful operations that are conducted in the digital domain with the aim of targeting computer systems, networks, and data. With the continuous development, the transformation of the lifestyle into digital, and the emergence of online services, the spread of these attacks and their significant effects have been observed at various levels. These attacks exploit vulnerabilities in systems and networks [14].

However, there was a noticeable shift in the strategies and tactics used in cyberattacks, as they started to go in a different path. Attack strategies changed from emphasizing network and system flaws to concentrating on "humans," the weakest link in the security chain. exploiting the lack of awareness of these attacks among the public. We refer to these kinds of cyberattacks as "social engineering attacks"[15].

Phishing attacks are among the most prominent attacks that use social engineering methods. The fact that it is so simple to use and can get past security and detection measures makes it one of the most popular attacks employed by attackers. The ease of carrying out this attack lies in the fact that it does not require programming experience to implement it. The attacker only needs to create a URL and design a web page very similar to pages on reputable websites [16].

The primary cause of the proliferation of phishing assaults is the insufficiency of existing systems in identifying and detecting these types of attacks. Although many detection systems are available, it is not a radical solution to this problem. Anti-phishing systems that use a blacklist and whitelist strategy are among these systems. This approach cannot be relied upon, especially with the continuous development of these attacks and the attackers' use of URLs that are not included in the lists, and this is known as zero-day phishing attacks [17].

In addition, there are content-based phishing detection systems that attackers can avoid, especially on sites written in languages other than English [18]. In addition, there are sites for which it is not possible to obtain the HTML and compare their content with the content of phishing sites because this may be part of the site's internal security and part of privacy. As for systems that rely on behavior analysis, they are one of the methods that are difficult to implement, especially in organizations that contain a large number of employees, in addition to that attackers can imitate user behavior, and this leads to false expectations.

Making the distinction between legitimate and fraudulent URLs is one of the best ways to recognize and detect phishing efforts. For many reasons, the most important of which is that it targets the “spoofed link” entry point. Compared to approaches that rely on content and user behavior analysis, this approach promises to be a proactive response to these attacks, as well as a faster response and reduced resource usage. Machine learning has proven its high ability to distinguish between these two titles through its powerful models that excel in the binary classification task [19].

IV. MACHINE LEARNING IN PHISHING DETECTION

Machine learning is important for phishing detection because it provides an automated, scalable, dynamic solution that can adapt to the changing nature of phishing attempts and thus enhance the detection and identification of these assaults. The basic types of machine learning can be divided into three primary categories [20] based on the problem-solving approach, the kind of data set utilized, and the learning style. As shown in Figure 2.

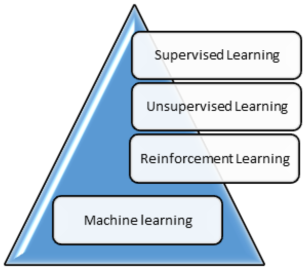

Figure 2 , Types of machine learning

A. *Supervised Learning*

Using a labeled dataset, supervised learning entails training a model with a target output for every input. Using the examples given, the algorithm picks up the mapping from inputs to outputs.

B. *Unsupervised Learning*

Unsupervised learning eliminates the requirement for explicit instruction in the form of labeled outputs by using unlabeled data and an algorithm to search the data for structures, relationships, or patterns.

C. *Reinforcement Learning*

A decision-making agent gains decision-making skills through interactions with its surroundings in reinforcement learning. Through rewards and penalties, the agent is given input, which helps it gradually figure out the best course of action.

In addition, there are three basic tasks in machine learning [21], each serving different purposes as shown in Figure 3.

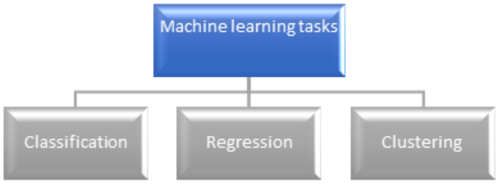

Figure 3 , Machine learning tasks

A. Classification

assigning labels or classes to the supplied data is the aim of this supervised learning activity.

B. Regression

Regression is a type of supervised learning job in which continuous numerical values are predicted as opposed to discrete classes. The algorithm learns the relationship between input variables and outputs a continuous value as the prediction.

C. Clustering

an unsupervised learning task where the goal is to group similar instances together based on certain properties, without knowing the actual class labels.

This study aims to detect phishing attacks based on analyzing URL addresses and working to distinguish and identify phishing and legitimate URLs, where it can be concluded that the solution to this problem is through binary classification. Numerous algorithms in machine learning have demonstrated efficacy in classification tasks; however, these algorithms can be broadly classified into two categories, as seen in Figure 4.

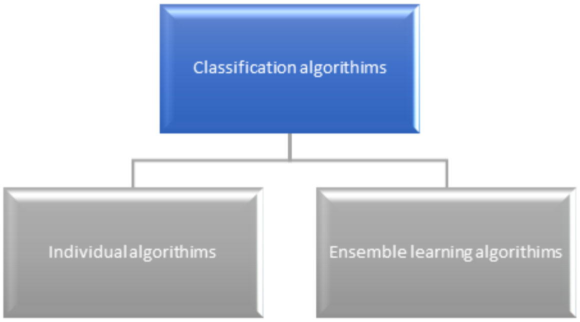

Figure 4 , Classification algorithms

A. Individual Algorithms

These autonomous algorithms use the available input features to anticipate things on their own. K-Nearest Neighbors (KNN), Native Bayes, logistic regression decision trees, and support vector machines (SVM) are a few of them.

B. Ensemble Learning Algorithms

With the help of these algorithms, a more reliable and accurate final forecast is produced by merging the predictions of several different distinct models. These consist of gradient boosting methods (such as XGBoost, LightGBM), AdaBoost, and random forest (an ensemble of decision trees).

Ensemble learning algorithms use three important techniques to combine predictions from several models [22], as shown in Figure 5.

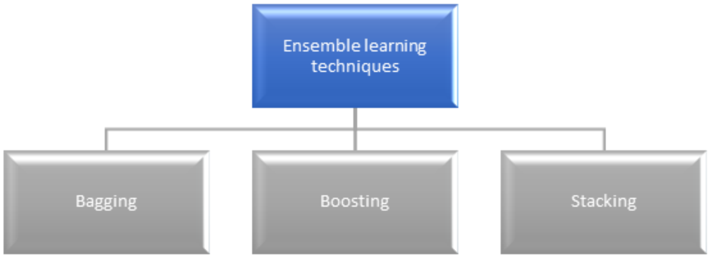

Figure 5, Ensemble learning techniques

A. Bagging (Bootstrap Aggregating)

In bagging, different subsets of the training data are used to train several concurrent instances of the same learning algorithm. Usually, replacement samples are taken (run sampling). Every model undergoes independent training, and in the case of regression or classification, the results are aggregated by voting or averaging.

B. Boosting

Boosting is the process of training a series of weak learners (models that perform marginally better than chance) and assigning greater weight to examples that the prior models incorrectly identified. A weighted total of the poor learners makes up the final prediction.

C. Stacking (Stacked Generalization)

Stacking involves training multiple diverse models and combining their predictions using another model (meta-model) The meta-model takes the predictions of the base models as input and learns to make the final prediction.

V. METHODOLOGY

This section of the paper presents a solution based on an ensemble learning algorithm and bagging technique that shows how machine learning may be used correctly in phishing detection. It also suggests a thorough optimization algorithm to adjust the model's hyperparameters. as shown in Figure 6.

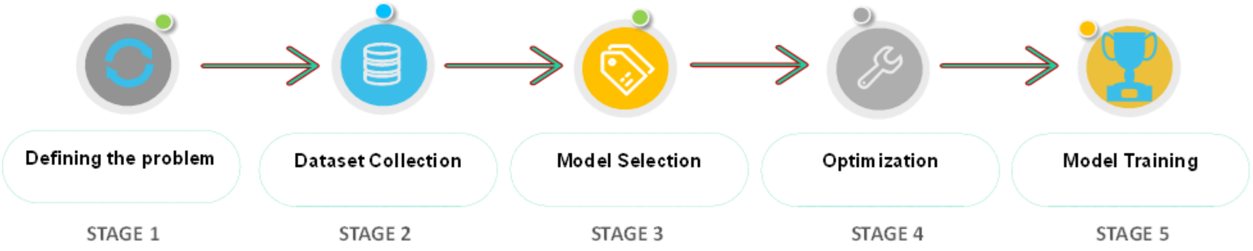

Figure 6 , Methodology

A. Defining the problem

The first step in launching any project or framework is to accurately characterize the problem to make sure that machine learning models and techniques can handle it. Since phishing URLs are a vital tool used by attackers to conduct phishing attacks, identifying URLs is an important goal of study. Machine learning has proven its high ability to discover these addresses through the binary classification tasks that characterize its models.

B. Data collection

The process of gathering data with the most crucial elements pertinent to the issue is then carried out. The model will be trained using these features as input. The prediction findings will be more accurate and valuable the better and more pertinent the features are to the issue. The dataset used in the study was taken from Kaggle, which includes 30 features and targets in addition to over 11,000 real and phony URLs.

The fundamental component of machine learning is problem-related data, which must be described and visualized for pre-processing prior to providing training data to the model. Finding and correcting any missing numbers or outliers was one of the most important phases in processing the research data. Creating two groups from the data set—one for the model's training and the other for testing—was another critical step (80–20).

In addition, feature extraction and cross-correlation map are adopted to visualize the most relevant features of the target and a feature selection threshold to exclude less relevant features in order to reduce feeding the model with less relevant features to the detection process and to focus the model for training and testing on the most relevant features.

The process of trying several models, training on them, and comparing the results is not a correct step and is not really a professional process. In this process, having a robust dataset that contains features that are strongly related to the problem can reduce and summarize many models and guide you to models that help solve the problem faster and more accurately by analyzing, describing, and visualizing that dataset.

C. Model selection.

Supervised machine learning methods can be used by describing the issue and dataset. Ensemble learning algorithms are noteworthy because they involve multiple techniques, and the combination of several predictions helps identify phishing attempts based on phishing URLs very efficiently.

Bagging techniques play a key role in dealing with variance-bias trade-offs and reducing variance in a prediction model. Bagging avoids overfitting of data and is used in both regression and classification models.

One of the most famous algorithms that inherently uses Bootstrap Aggregating as a core element of its methodology is the Random Forest algorithm. By training many instances of the model on different subsets of the training data, bagging is an effective ensemble learning strategy that seeks to improve the stability and accuracy of the prediction model. Figure 7 illustrates the bagging technique.

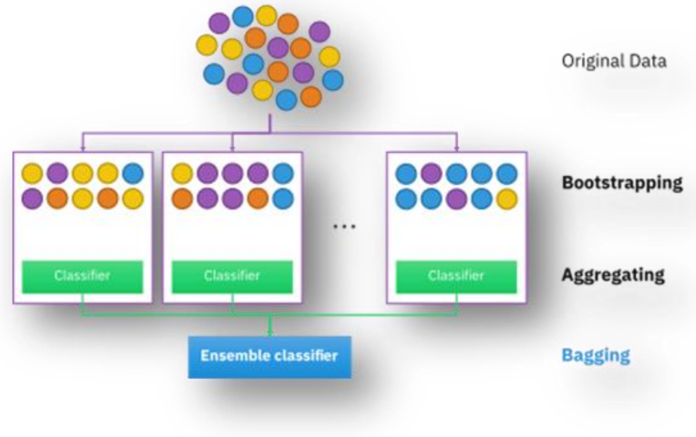

Figure 7, Bagging technique

D. Optimization

Algorithms for optimization are techniques that help choose the best answer from a range of workable options for a particular problem. When maximizing or minimizing an objective function—also referred to as a cost or loss function—these techniques are applied in machine learning and other domains. To attain the best outcome, modify the factors or parameters to reach the ideal outcome.

The PSO algorithm was adopted to improve and adjust the parameters of the random forest model. Based on the social behavior of fish and birds, Particle Swarm Optimization is a method for optimization influenced by nature. PSO is a heuristic algorithm first presented by Eberhart and Kennedy in 1995. It is designed to emulate the cooperative and adaptive behavior seen in natural swarms.

E. Model Training

The model is supplied training data during this phase, and it uses the data to identify patterns and relationships.

F. Model Evaluation and testing.

The model's performance is assessed after testing on an unseen test dataset. F1 score, recall, precision, and Accuracy are examples of common evaluation criteria.

Accuracy: This is the proportion of cases that were accurately predicted to all instances. It offers a general indicator of the model's performance.

$$\text{Accuracy}=\frac{Number\ of\ correct\ predictions}{Total\ number\ of\ predictions} \qquad (1)$$

Precision: The ratio of accurately predicted positive observations to all expected positives is known as precision. The accuracy of the optimistic predictions is the main emphasis.

$$\text{Precision}=\frac{True\ Positives}{True\ Positives+\ False\ Positives} \qquad (2)$$

Recall (Sensitivity): The ratio of accurately predicted positive observations to all of the observations made during the actual class is known as recall. Other names for it are True Positive Rate and Sensitivity.

$$\text{Recall}=\frac{True\ Positives}{True\ Positives+False\ Negatives} \qquad (3)$$

F1-score: The F1-score is calculated as the precision and recall harmonic means. It is especially helpful when there is an unequal class distribution since it strikes a compromise between recall and precision.

$$\text{F1-score}=\frac{2\times Precision\times Recall}{Precision\ +\ Recall} \qquad (4)$$

VI. Results and Discussion

The Random Forest model optimized using Particle Swarm Optimization (PSO), which is based on the parallel bagging technique, showed commendable performance in detecting phishing attempts. The evaluation metrics are summarized in the TABLE 2.

TABLE 2. Evaluation metrics

ML	Accuracy	Precision	Recall	F1-score
Random forest	97.1	97.1	97.7	97.4

The model's high accuracy level suggests that it is effective at accurately identifying situations. Precision signifies the proportion of predicted positives that are true positives, highlighting the model's capability to minimize false positives. The high recall rate indicates the model's effectiveness in capturing a significant portion of actual phishing instances. The F1 Score, which balances precision and recall, reflects the model's ability to provide a harmonized measure of performance. These evaluation metrics were obtained based on the results extracted from the confusion matrix. As shown in the Figure 8. In addition to the improvement in accuracy achieved by using the PSO optimization algorithm in several iterations to reach the best accuracy results, as shown in Figure 9, which shows the accuracy curve.

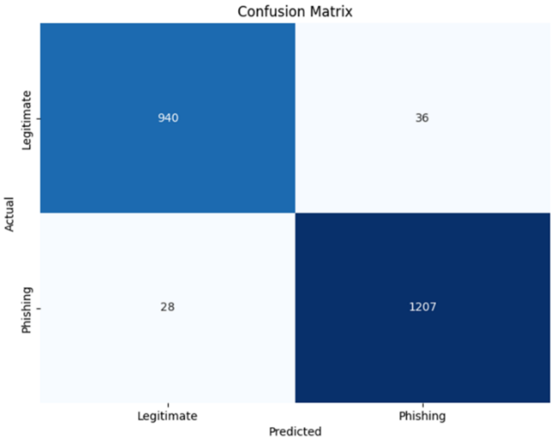

Figure 8, Confusion matrix

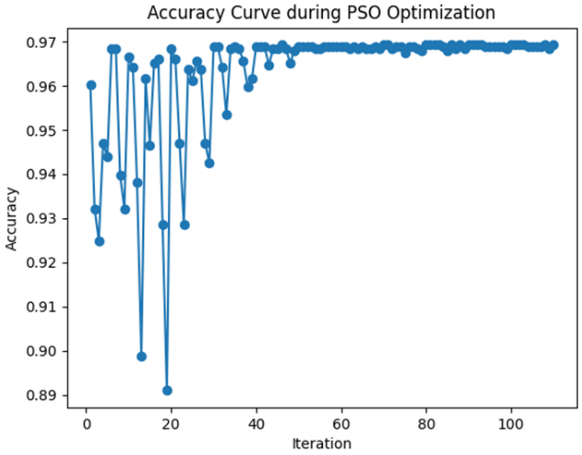

Figure 9, Accuracy curve

Compared to previous work conducted by a group of researchers in the same field of research on phishing detection and relying on the same model and techniques, the approach proposed in the study has highly accurate results in detecting and distinguishing between legitimate URLs and phishing URLs, by relying on parallel training. For the model. Through bagging technology and global optimization algorithm for PSO optimization purposes.

VII. Conclusion and future work

Phishing detection has been extensively researched using machine learning approaches. This research provided insights into two key areas: the optimization (PSO) algorithm, which played a crucial role in modifying the model's hyperparameters, and the bagging approach utilized in ensemble learning through the Random Forest model during the model training stage on the training data set. Resulting in an improved model with outstanding performance. The resultant Random Forest model, which was developed by combining optimization and machine learning, demonstrated an impressive 97% accuracy. The results of the F1, recall, and precision metrics also confirmed the efficiency of the model in distinguishing between phishing and legitimate cases. The study effectively demonstrated a robust phishing detection model that uses Random Forest and Particle Swarm Optimization techniques. The high accuracy achieved, coupled with balanced precision and recall, positions the model as a promising solution in current efforts to combat phishing threats. Future research will include integrating the model proposed in this study into an integrated framework for detecting phishing attacks by providing a web application with an easy graphical interface for users, in addition to mixing its predictions with predictions of other models that use techniques other than bagging. To achieve a hybrid model that provides results with higher accuracy and fills the gaps. Which is not covered by the model and bagging technique.

REFERENCES

[1] S. Aslam and A. B. Nassif, “Phish-identifier: Machine Learning based classification of Phishing attacks,” in 2023 Advances in Science and Engineering Technology International Conferences (ASET), 2023, pp. 1–6. doi: 10.1109/ASET56582.2023.10180869.

[2] V. Gomes, J. Reis, and B. Alturas, “Social Engineering and the Dangers of Phishing,” Iberian Conference on Information Systems and Technologies, CISTI, vol. 2020-June. 2020. doi: 10.23919/CISTI49556.2020.9140445.

[3] A. Kaur and S. M. Mian, “A Review on Phishing Technique: Classification, Lifecycle and Detection Approaches,” in 2023 3rd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE), 2023, pp. 336–339. doi: 10.1109/ICACITE57410.2023.10183292.

[4] K. Patil and S. R. Arra, “Detection of Phishing and User Awareness Training in Information Security: A Systematic Literature Review,” in 2022 2nd International Conference on Innovative Practices in Technology and Management (ICIPTM), 2022, pp. 780–786. doi: 10.1109/ICIPTM54933.2022.9753912.

[5] K. Subashini and V. Narmatha, “Website Phishing Detection of Machine Learning Approach using SMOTE method,” in 2023 Fifth International Conference on Electrical, Computer and Communication Technologies (ICECCT), 2023, pp. 1–5. doi: 10.1109/ICECCT56650.2023.10179745.

[6] G. Apruzzese et al., “The Role of Machine Learning in Cybersecurity,” Digit. Threat., vol. 4, no. 1, Mar. 2023, doi: 10.1145/3545574.

[7] I. D. Mienye and Y. Sun, “A survey of ensemble learning: Concepts, algorithms, applications, and prospects,” IEEE Access, vol. 10, pp. 99129–99149, 2022.

[8] M. Ajdani and H. Ghaffary, “Introduced a new method for enhancement of intrusion detection with random forest and PSO algorithm,” Secur. Priv., vol. 4, no. 2, p. e147, 2021, doi: https://doi.org/10.1002/spy2.147.

[9] M. Alanezi, “Phishing Detection Methods: A Review,” Tech. Rom. J. Appl. Sci. Technol., vol. 3, pp. 19–35, 2021, doi: 10.47577/technium.v3i9.4973.

[10] R. Zieni, L. Massari, and M. C. Calzarossa, “Phishing or not phishing? A survey on the detection of phishing websites,” IEEE Access, vol. 11, pp. 18499–18519, 2023.

[11] Z. Zhou and C. Zhang, “Phishing website identification based on double weight random forest,” in 2022 3rd International Conference on Computer Vision, Image and Deep Learning & International Conference on Computer Engineering and Applications (CVIDL & ICCEA), 2022, pp. 263–266. doi: 10.1109/CVIDLICCEA56201.2022.9824544.

[12] M. H. Alkawaz, S. Joanne Steven, O. F. Mohammad, and M. Gapar Md Johar, “Identification and Analysis of Phishing Website based on Machine Learning Methods,” in 2022 IEEE 12th Symposium on Computer Applications & Industrial Electronics (ISCAIE), 2022, pp. 246–251. doi: 10.1109/ISCAIE54458.2022.9794467.

[13] K. Subashini and V. Narmatha, “Phishing Website Detection using Hyper-parameter Optimization and Comparison of Cross-validation in Machine Learning Based Solution,” in 2023 Third International Conference on Advances in Electrical, Computing, Communication and Sustainable Technologies (ICAECT), 2023, pp. 1–6. doi: 10.1109/ICAECT57570.2023.10117851.

[14] I. Emmanuel O., E. V. C., O. E. I., and N. P. C., “Overview of Recent Cyberattacks: A Systematic Review,” in 2023 International Conference on Science, Engineering and Business for Sustainable Development Goals (SEB-SDG), 2023, pp. 1–8. doi: 10.1109/SEB-SDG57117.2023.10124473.

[15] W. Syafitri, Z. Shukur, U. A. Mokhtar, R. Sulaiman, and M. A. Ibrahim, “Social Engineering Attacks Prevention: A Systematic Literature Review,” IEEE Access, vol. 10, pp. 39325–39343, 2022, doi: 10.1109/ACCESS.2022.3162594.

[16] A. Basit, M. Zafar, X. Liu, A. R. Javed, Z. Jalil, and K. Kifayat, “A comprehensive survey of AI-enabled phishing attacks detection techniques,” Telecommun. Syst., vol. 76, no. 1, pp. 139–154, 2021, doi: 10.1007/s11235-020-00733-2.

[17] S. Merugula, K. S. Kumar, S. Muppidi, and C. Vidyadhari, “Stop Phishing : Master Anti-Phishing Techniques,” in 2022 IEEE North Karnataka Subsection Flagship International Conference (NKCon), 2022, pp. 1–5. doi: 10.1109/NKCon56289.2022.10126569.

[18] I. Dunđer, S. Seljan, and M. Odak, “Data Acquisition and Corpus Creation for Phishing Detection,” in 2023 46th MIPRO ICT and Electronics Convention (MIPRO), 2023, pp. 533–538. doi: 10.23919/MIPRO57284.2023.10159904.

[19] D. Savchuk and A. Doroshenko, “Investigation of machine learning classification methods effectiveness,” in 2021 IEEE 16th International Conference on Computer Sciences and Information Technologies (CSIT), 2021, pp. 33–37. doi: 10.1109/CSIT52700.2021.9648582.

[20] S. V Mahadevkar et al., “A Review on Machine Learning Styles in Computer Vision—Techniques and Future Directions,” IEEE Access, vol. 10, pp. 107293–107329, 2022, doi: 10.1109/ACCESS.2022.3209825.

[21] G. Gupta, M. Sharma, S. Choudhary, and K. Pandey, “Performance Analysis of Machine Learning Classification Algorithms for Breast Cancer Diagnosis,” in 2021 9th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO), 2021, pp. 1–6. doi: 10.1109/ICRITO51393.2021.9596230.

[22] T. T. Khoei, G. Aissou, W. C. Hu, and N. Kaabouch, “Ensemble Learning Methods for Anomaly Intrusion Detection System in Smart Grid,” in 2021 IEEE International Conference on Electro Information Technology (EIT), 2021, pp. 129–135. doi: 10.1109/EIT51626.2021.9491891.