Detecting and Preventing Cyber Threats Using Machine Learning Techniques (June 2022)

Raed A. Alshehri[1]

[1]Intelligent Secure Systems Research Center, King Fahd University of Petroleum and Minerals, 31261, Dhahran, Saudi Arabia

Corresponding author: Raed A. Alshehri (e-mail: Alshehri.RaedA@gmail.com).

This work was supported by the Office of Undergraduate Research at KFUPM under Grant "Uxplore212".

ABSTRACT This report explores the usage of machine learning and heuristic optimization in the field of cybersecurity. Precisely, this report assesses the application of machine learning algorithms in predicting and preventing cyber threats. It evaluates and analyzes a dataset containing spam messages. Then, it uses Particle Swarm Optimization and Genetic Algorithm for selecting the related features. After that, this study builds predictive models using multiple machine learning algorithms and illustrates the key findings. It concludes with useful recommendations that can be applied by private organizations and governmental agencies to ensure the security of systems against the prevailing cyber threats.

INDEX TERMS Cybersecurity, Genetic Algorithms, Heuristic Optimization, Machine Learning, Particle Swarm Optimization, Phishing emails, Spam messages.

I. INTRODUCTION

The recent growth and development of information technology has been accompanied by an increasing number of cyberattacks. Consequently, conventional cybersecurity systems are unable to keep pace with the advancements of cyber threats. This limitation in the traditional security systems propelled the use of machine learning in detecting cyberattacks. As machine learning can learn from patterns, and due to the frequent occurrence as well as the severity of cyber threats, the use of machine learning is ideal for detecting and preventing these threats.

Machine learning is an extremely effective tool that is used nowadays more than ever in many areas of information security. Precisely, and as noted by many researchers in the literature, cyber threats, such as spam messages and phishing emails are more relevant to the current times, and these threats are continuously being examined in greater detail. Furthermore, the abundance of machine learning algorithms and their functionalities paved the way for an extensive set of mechanisms, which are required to assess the viability of detecting these cyber threats. The purpose of this study is to examine the application of different machine learning techniques to enhance the accuracy in detecting and preventing cyber threats, and the shortcomings that limit its efficient utilization.

As the field of information security is continuously evolving to address the ever-changing threat of cybercrimes, machine learning represent a relatively new field of investigation. Moreover, there are many models and techniques that have been developed in recent years to enhance the current problems in security. Many organizations and government agencies could benefit from the capabilities of machine learning in securing their information. Private companies can secure their systems and prevent malicious attacks by implementing machine learning models. Similarly, other organizations can use these models to protect their information against the prevailing cyber threats.

This study is limited to an examination of machine learning algorithms in detecting spam messages and phishing emails. Mainly, this study will go through different datasets and use machine learning classification algorithms to evaluate the accuracy of predicting these cyber threats. This study is restricted to the use of shallow learning algorithms in the procedure of classifying the malicious emails and messages. Other cyber threats, and the use of deep learning algorithms are beyond the scope of this study. The goal of this study is to enhance the accuracy of classifying cyber threats using different machine learning algorithms. Specifically, this study will analyze different datasets and select features using heuristic optimization. After that, a variety of machine learning classifiers will be tested until high accuracy in predicting spam and phishing emails is reached.

II. LITERATURE REVIEW

The Recent research analyzed a variety of different machine learning algorithms in detecting cyber threats and used accuracy and precision to accurately evaluate the results. In [2], the authors reviewed the application of machine learning algorithms and the high accuracy they achieved when dealing with several cybercrimes, such as detecting malware, classifying phishing attacks, and identifying intruders. The proposed machine learning algorithms that achieved high accuracy and accurate result when dealing with these cyber threats included: Naïve Bayes, Logistic Regression, Support Vector Machines, and Random Forest [14]. Nevertheless, even with the prevailing advantages, some limitations arise.

An in-depth review of the literature revealed that machine learning classifiers yield high accuracy when used in detecting spam and phishing emails. In [1], the authors proposed a framework in which K-Nearest Neighbor and Naïve Bayes classifiers are used to classify spam SMS messages. Additionally, K-Means clustering algorithm was used to group the SMS messages from each group and assess their risk, and then classify them accordingly. In [3], the author used multiple machine learning algorithms and concluded that Random Forest algorithm achieved the highest accuracy of 98.2% in detecting spam SMS messages. Likewise, the authors in [14] extracted the features from a dataset and proposed the use of Random Forest algorithm to detect phishing emails, which achieved a high accuracy of 98.87%. In [13], the authors delivered astonishing results in detecting phishing emails, attaining around 99.8% accuracy when using Random Forest and Support Vector Machine classifiers.

Many anomalous behaviors have emerged as a result of the vast number of devices that rely on the internet, particularly with the recent spike in the use of Internet of Things technologies (IoT), which has increased Internet consumption. These developments paved the way for Intrusion Detection Systems (IDS), which are defined as tools that analyze and monitor network traffic to alert system networks of suspicious activities [11]. Machine Learning Algorithms are used by the majority of high-performance IDSs since they are better at detecting relationships between data than predefined algorithms. According to [5], in a review of different machine learning algorithms used to detect intrusions, the authors discovered that decision trees outperformed other ML techniques with an overall accuracy of 98.45 % in Denial-of-Service (DoS), probe, User to Root (U2R), and Remote to Local (R2L) attacks, followed by Neural Networks which excelled only in detecting DoS attacks with 99.87 % and had 80% performance in other attacks.

Malware is one of the most used methods to access users' information without their knowledge. In fact, the authors in [4] found that one of the top methods to acquire user information is Trojan horse malware. In order to prevent these Malicious programs, machine learningalgorithms have been tested and showed significant achievement in detecting these threats. In [10], the authors useddifferent algorithms to detect malware including Random Forest, Gradient Boosting, and Decision Tree, and got an accuracy result of 98.9%, 97.5%, and 97.5 respectively, andit was done using opcode image, n-gram, and combined classifications. Other tests done by [12], used Random Forest classifier as well as Principal Component Analysis and got a 96.05% accuracy result. In [7], the authors proposed an interesting model to detect malware by classifying malicious behavior by machine learning using virtual memory access patterns which outperformed other methods by achieving a 99% accuracy result.

A major limitation in the current literature is the lack of appropriate datasets to conduct experiments. This might be due to the fact that many organizations feel reluctant to share confidential information. Moreover, it has been emphasized that machine learning algorithms can yield different results when dealing with different datasets [4]. In addition, several studies have discussed that the classifiers and machine learning algorithms themselves are subject to cyberattacks. It is evident that the current research found encouraging results when dealing with spam and phishing emails. Yet, some datasets used in the literature might not replicate real life scenarios [13]. Likewise, although machine learning algorithms are detecting intrusions accurately, low-frequency attacks such as U2R and R2L cannot be detected easily even with high examination because it is close to the normal connection. Similarly, all these findings show the high ability of machine learning to detect Malware threats. Nevertheless, in real-world domains, the capability of cybersecurity models is limited since it is usually built in a priori known risks and the nature of real- world threats is unpredictable [5].

The use of machine learning algorithms in detecting cyber threats is being analyzed extensively because of the valuable capabilities they provide. Nonetheless, the absence of proper datasets and the vulnerability of machine learning classifiers against cyberattacks represent stumbling limitations to the researchers. The current literature highlighted encouraging results and high accuracy in predicting the most relevant cyber threats. Yet, many shortcomings and limitations still persist. It is apparent that significant improvements are expected in this field but only when the limiting factors are eliminated, or at least properly dealt with.

III. RESEARCH METHODOLOGY

The methodology used to conduct this study consists of four phases. It starts by finding an appropriate dataset for phishing emails and spam messages. As noted in the literature, the availability of datasets poses a challenge to many researchers. For this reason, a publicly available dataset will be used to conduct the research. Next, the data will undergo a feature selection technique in which heuristic optimization algorithms will be utilized. After that, the dataset will be split into training data to train the prediction model, and testing data to ensure that the fitting is proper. Finally, the testing data will be used to build a classification model and the accuracy of the model will be calculated to measure its effectiveness. The detailed research methodology is exhibited in Fig. 1 below.

The limitations of this study may come mainly through finding a common dataset to apply the experiments and testing required. As organizations are unwilling to share confidential datasets that contain sensitive information related to cyberattacks, a publicly available database will be used. Furthermore, the dataset used was encoded so that only testing could be performed and hence, drawing a clear conclusion related to a specific cyber threat was somewhat difficult. Moreover, achieving higher accuracy is a challenge that needs a deep understanding of the dataset to appropriately extract the most useful information that allow for accurate predictions. Finally, the application of different machine learning algorithms could lead to subjectivity in the interpretation and analysis of the acquired data. All necessary steps have been made to minimize the effects of these limitations on the study.

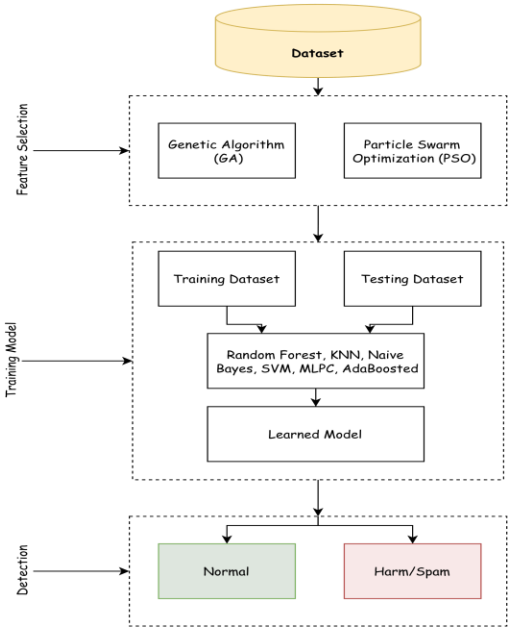

FIGURE 1. Detailed Research Methodology

IV. FINDINGS

This study was designed to enhance the current status of using machine learning in detecting and preventing cyber threats. A publicly available spam messages dataset was used in testing and building the machine learning models, which can be found in [6]. The findings of this study will be presented in two main sections: Feature Selection and Machine Learning Classifiers.

A. FEATURE SELECTION

In the investigated database, there are 57 different features that affect the output (whether the message is harmful/spam or safe). To build a classification model that can predict the status of the message with high accuracy, the related features must be selected properly. Rather than using all features, which might not all be related and could result in misclassification, the utilized feature selection method increases the effect of the selected features while simultaneously ignoring the unrelated attributes. Even though many machine learning classifiers achieved relatively high accuracy in predicting spam and phishing messages, this study improved the resulting accuracy through the use of heuristic optimization for selecting the related features.

In this study, feature selection is simply treated as an optimization problem to improve the accuracy of predicting spam and phishing emails. Mainly, two heuristic optimization methods are used: Genetic Algorithm and Particle Swarm Optimization. Many machine learning classifiers were used with the heuristics and the following results were found and can be observed in Fig. 2 and Fig. 3 The figures above exhibit how many times the algorithms run in the x-axis, against the measure of selection error for a specific feature (fitness) in the y-axis. As can be seen from the figures, the selection error converges a minimum value with each iteration. Therefore, by considering the feature selection process as an optimization problem, a minimum error was found, and the optimal features were selected accordingly.

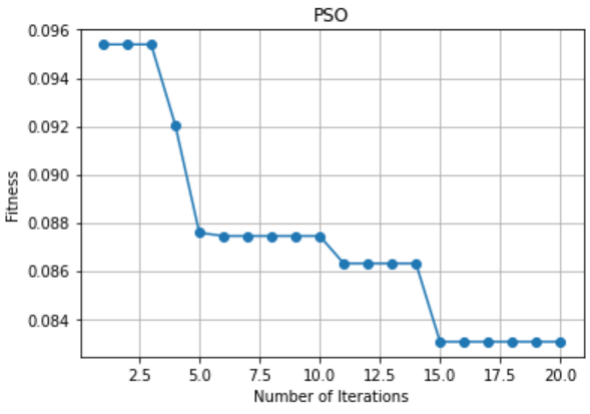

FIGURE 2. Implementing Feature Selection Using Particle Swarm Optimization

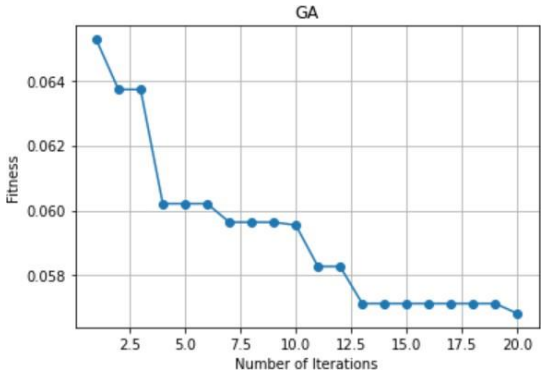

FIGURE 3. **Implementing Feature Selection Using Genetic Algorithm**

B. MACHINE LEARNING CLASSIFIERS

After the use of heuristic optimization for selecting the related features, many machine algorithms were used to build classification models. Namely, Decision Tree, Random Forest, Naïve Bayes, Support Vector Machine, and K-Nearest Neighbor were used. The purpose of using a variety of machine learning algorithms was to measure the enhanced accuracy due to the use of heuristics in selecting the inputs. In Table 1, the measured accuracy of the different machine learning classifiers are displayed, along with the accuracy after applying heuristic optimization for feature selection.

As can be seen from Table 1, the accuracy of machine learning classifiers changed due to application of heuristic optimization in selecting the related features. It is evident that the overall accuracy of some machine learning algorithms improved due to the use of heuristics. Yet, some algorithms such as Support Vector Machine (SVM) were not affected significantly by the use of heuristic optimization for selecting the related inputs. Thus, the enhanced accuracy of these models demonstrates an improved method of predicting these spam and phishing messages.

TABLE I
COMPARISON OF THE ACCURACY AFTER APPLYING HEURISTIC OPTIMIZATION FOR DIFFERENT MACHINE LEARNING ALGORITHMS

Classifier	Accuracy	Accuracy After Applying Heuristics	
		PSO	GA
Decision Tree	91.75%	93.12%	92.83%
Random Forest	95.73%	95.15%	93.77%
Naïve Bayes	82.11%	83.06%	85.01%
Support Vector Machine	93.56%	92.69%	92.47%
K-Nearest Neighbor	90.51%	90.88%	91.02%

V. CONCLUSION

Based on the results from the extensive testing, many conclusions can be drawn regarding the application of heuristic optimization and machine learning for detecting and preventing cyber threats. The results of this study indicated that machine learning are capable of enhancing and securing security systems against cyber threats, such as spam and phishing messages. The prevailing cyber threat of spam messages and phishing emails can be mitigated by the utilization of machine learning techniques and heuristic optimization algorithms for feature selection. From this study, it can be concluded that Random Forest Algorithm achieves the highest accuracy when predicting spam and phishing messages. Genetic Algorithms and Particle Swarm Optimization exhibit better performance than traditional feature selection methods, and especially with datasets containing many features. The machine learning models built after selecting the features scores higher accuracy. Some

machine learning algorithms perform better by using different optimization technique but there is no clear rationale for these results. Specifically, Decision Tree algorithm, appears to be performing better by using Particle Swarm Optimization for feature selection, while Naïve Bayes and K-Nearest Neighbor perform better when Genetic Algorithm is used.

The accuracy scores of some machine learning classifiers were not affected by implementing heuristic optimization for feature selection, and hence this process of feature selection is not appropriate for all machine learning algorithms.

Based on the findings and conclusions in this study, the following recommendations are made: Organizations and governmental agencies should use machine learning rather than traditional security systems to deal with spam and phishing messages, as it will achieve high accuracy in predicting and preventing these cyber threats.

Private organizations and companies should release more data regarding the cyber threats they face, as the availability of real-life datasets will allow researchers to build machine learning model with enhanced capabilities. Organizations utilizing machine learning when detecting spam and phishing emails ought to use Random Forest algorithm, as it was found to achieve the highest accuracy. Heuristic optimization should also be integrated when building the model as it will improve the resulting accuracy even more. Organizations utilizing machine learning should use Particle Swarm Optimization and Genetic Algorithm when selecting the related features rather than traditional input-selection methods

Additional research should be done to investigate the use of machine learning in the field of cybersecurity. Moreover, the effect of integrating advanced feature selection techniques can improve the overall accuracy of the model, but some algorithms appear to perform poorly when heuristics are used. Other areas negatively affected by the use of machine learning should be studied to determine the security of the models themselves against cyber threats.

ACKNOWLEDGMENT

I am truly grateful and thankful to Allah, and to all of those whom I have had the pleasure to work with. This work would not have been possible without the support of the Office of Undergraduate Research. I am indebted to Dr. Mujahid who has provided me with extensive personal and professional guidance and taught me a great deal about scientific research.

REFERENCES

[1] Adewole, K. S., Anuar, N. B., & Kamsin, A. (2016). Ensemble based streaming framework for spam detection and risk assessment in microblogging social networks. International Conference on Computer Science and Computational Mathematics (ICCSCM). https://www.researchgate.net/publication/301975526_Ensemble_based_streaming_framework_for_spam_detection_and_risk_assessment_in_microblogging_social_networks.

[2] Al-Khater, W. A., Al-Maadeed, S., Ahmed, A. A., Sadiq, A. S., & Khan, M. K. (2020). Comprehensive Review of Cybercrime Detection Techniques. IEEE Access, 8, 137293–137311. https://doi.org/10.1109/access.2020.3011259

[3] Alshahrani, A. (2021). Intelligent Security Schema for SMS Spam Message Based on Machine Learning Algorithms. *International Journal of Interactive Mobile Technologies (iJIM)*, *15*(16), 52. https://doi.org/10.3991/ijim.v15i16.24197

[4] Apruzzese, G., Colajanni, M., Ferretti, L., Guido, A., & Marchetti, M. (2018). On the effectiveness of machine and deep learning for cyber security. 2018 10th International Conference on Cyber Conflict (CyCon). https://doi.org/10.23919/cycon.2018.8405026

[5] Dasgupta, D., Akhtar, Z., & Sen, S. (2020). Machine learning in cybersecurity: a comprehensive survey. The Journal of Defense Modeling and Simulation: Applications, Methodology, Technology, 19(1), 57–106. https://doi.org/10.1177/1548512920951275

[6] Dua, D., & Graff, C. (2019). UCI Machine Learning Repository. UCI Machine Learning Repository. http://archive.ics.uci.edu/ml

[7] Geetha, R., & Thilagam, T. (2020). A Review on the Effectiveness of Machine Learning and Deep Learning Algorithms for Cyber Security. Archives of Computational Methods in Engineering, 28(4), 2861–2879. https://doi.org/10.1007/s11831-020-09478-2

[8] Gomez, F. (2022). Genetic algorithms for feature selection | Neural Designer. Neural Designer Is a Registered Trademark of Artificial Intelligence Techniques, S.L. https://www.neuraldesigner.com/blog/genetic_algorithms_for_feature_selection

[9] Kok, A., Ilic Mestric, I., Valiyev, G., & Street, M. (2020). Cyber Threat Prediction with Machine Learning. Information & Security: An International Journal, 47(2), 203–220. https://doi.org/10.11610/isij.4714

[10] Liu, L., Wang, B. S., Yu, B., & Zhong, Q. X. (2017). Automatic malware classification and new malware detection using machine learning. Frontiers of Information Technology & Electronic Engineering, 18(9), 1336–1347. https://doi.org/10.1631/fitee.1601325

[11] Mishra, P., Varadharajan, V., Tupakula, U., & Pilli, E. S. (2019). A Detailed Investigation and Analysis of Using Machine Learning Techniques for Intrusion Detection. IEEE Communications Surveys & Tutorials, 21(1), 686–728. https://doi.org/10.1109/comst.2018.2847722

[12] Moon, J., Kim, S., Song, J., & Kim, K. (2021). Study on Machine Learning Techniques for Malware Classification and Detection. KSII Transactions on Internet and Information Systems, 15(12). https://doi.org/10.3837/tiis.2021.12.003

[13] Rawal, S., Rawal, B., Shaheen, A., & Malik, S. (2017). Phishing Detection in E-mails using Machine Learning. International Journal of Applied Information Systems, 12(7), 21–24. https://doi.org/10.5120/ijais2017451713

[14] Smadi, S., Aslam, N., Zhang, L., & Hossain, M. A. (2015). Detection of phishing emails using data mining algorithms. 2015 9th International Conference (SKIMA). https://doi.org/10.1109/skima.2015.7399985