Enhance the accuracy of malicious uniform resource locator detection based on effective machine learning approach

Haifa Alqahtani[1], Ahmed Abu-Khadrah[2]
[1]College of Computing and Informatics, Saudi Electronic University, Riyadh, Saudi Arabia
[2]Department of Electrical Engineering, College of Engineering Technology, Al-Balqa Applied University, Amman, Jordan

Article Info

Article history:

Received Oct 28, 2023
Revised Apr 2, 2024
Accepted May 17, 2024

Keywords:

Bagging
Malicious
Phishing
Random forest
Support vector machine

ABSTRACT

Phishing attacks are increasing with the rise in web users. Addressing them requires understanding the techniques and employing effective response strategies. Phishing websites mimic authentic ones to deceive users into divulging personal information like bank account details, national insurance numbers, and passwords. Therefore, victims face financial loss from breached information security, constituting high-level internet fraud. Detecting phishing websites necessitates an intelligent model capable of recognizing suspicious features. To that purpose, this paper examines three classification methods for detecting phishing website attacks. This analysis allows to reconsider our awareness of phishing attacks and prevent the damage caused by phishing attempts in advance. Phishing website detection algorithm using three classification algorithms is proposed in this paper. It achieves high phishing website detecting accuracy, because three classification algorithms random forest (RF), support vector machine (SVM), and Bagging are combined in one system. The result of this research is found accuracy on validation set is 92.33%, the precision on validation set is 92.13%, the recall is 92.09% and F1 score is 92.10%. That prove that the result obtained in this research is more accurate than all the results of all the algorithms were applied in the same dataset that was train the proposed algorithm on it.

Corresponding Author:

Ahmed Abu-Khadrah
Department of Electrical Engineering, College of Engineering Technology, Al-Balqa Applied University
Amman, Jordan
Email: abosuliman2@yahoo.com

1. INTRODUCTION

Due to the increase in the number of uniform resource locators (URL), the number of cyber-attacks is increasing not only in the United Kingdom but across the world. It is reported that since January 2015 the number of ransom ware attacks has increased by 300% till 2022 [1]. This indicates that, increased number of malicious URLs is increasing issues of cyber-attack and phishing to the websites of educational institutes and business organizations. It has been identified that ransom ware attacks on websites mainly targeting home users, networks of educational institutes and business websites to steal personal and private information. As a result, civilians, business owners, and students are facing the issues related to the privacy violation. Temporary and permanent loss of personal data of users is finally resulting in loss of sensitive and proprietary information. Hence, the entire operational process is facing disruptions and regular functionality getting accessed. Malicious URLs is attacking home users or businesses by tricking users through sending passive and virus laden email attachments [2].

Finally, the disruptions in the operational process are causing revenue loss in business organizations while home users are losing personal information which is one of the key issues in the present scenario. Hence, it can be commented that the increased number of malicious users is a present issue as with the digital transformation the number of authentic websites is increasing rapidly. In this regard, this study is focusing to understand how usefulness of machine learning can help in secure website development. Along with that, this study has mentioned how combination techniques can help to develop a framework with the collaboration of support vector machine (SVM) algorithms, random forest (RF) approaches, and Bagging. It is being expected that, after developing a framework with a combined algorithm, users could be able to face limited issues caused by cyber-attack. Proactive prevention with a combined algorithm is imperative to establish effective defense against the malicious URL. Hence, it is indicating that, after completing this study, home users could get the opportunity to develop awareness among them [3]. On the other hand, business organizations could get the opportunity to establish a strong framework that can prevent incidents of cyber-attack on their websites. Therefore, this study is said to focus on the development of a framework with a combined algorithm.

Digital transformation in the contemporary business world is increasing the number of malicious URLs as it is directing users to other fraudulent web pages. It is reported that URL generally uses resources available on the internet which is leading to failure of detecting malicious URLs [4]. On the other hand, the similar study also highlighted that inclusion of machine learning techniques in the malevolent URL detection helps web developers to implement behaviour analysis techniques. As a result of it, web developers get the opportunity to construct a deep learning algorithm to classify characteristics of URLs by utilising SVM [5]. This is indicating that use of machine learning in the URL characteristics detection holds the capability to detect doubtful URLs. Along with that, it helps to configure and control access of files, networks, and shares permission information to end users. As a result of it, network restriction policies and other control systems successfully prevents incidents of theft and phishing to websites. Only the SVM algorithm is not able to categorise data based on the nature and characteristics of URLs. As a result, end users are facing difficulty in securing the privacy of users while losing important information [6].

On the contrary, several studies have highlighted that SVM algorithm is not appropriate to analyse large datasets, as a result of it, users are leading to fraudulent web pages [7]. This is indicating that the SVM algorithm is not performing accurately if the dataset is large and consists of more noise. Therefore, it is indicating that, whenever web developers are using SVM algorithm for large datasets witnessing overlapping for target classes. SVM algorithms underperformed if the number of features exceeds each data point. This is clearly demonstrating that for large data sets linear SVM algorithm is not appropriate for supporting vector classifiers as there are no behaviour based probabilistic explanations for the classification. The education sector of the United Kingdom witnessed 4000 ransom ware attacks from the year 2020 to 2022 which is 300% higher than 2015 [4]. This is indicating that URL characteristics detection model news to be improved by combining different algorithms as only SVM algorithm is not appropriate to protect against cyber-attack.

Phishing websites are fake websites that can be built by attaching to imitate and represent authentic websites to defraud others by stealing their personal important data such as bank account information, national insurance number, and passwords. As a result, the victim incurs a financial loss because of a breach of information security caused by the theft of confidential data. In an essence, it is high-level internet fraud or delinquency. As a result, assessing or identifying phishing websites necessitates the use of an intelligent model capable of recognizing and detecting suspicious features associated with phishing websites. The objective of this study is to develop a new model for predicting the legitimacy of websites, distinguishing between genuine ones and potential phishing sites. This will be accomplished by employing three machine learning algorithms: SVM, Bagging, and RF.

2. MACHINE LEARNING ALGORITHMS

Malicious websites are created for stealing data and different kinds of resources from any official websites. URL promotes cyber-attacks as clicking on infected network results in commencement of unauthorised data transfer process. Unethical hackers often upload malware in the websites by creating malicious URLs. The malicious URLs are clickable. It often occurs that people out of curiosity click on the malware and it causes a big loss for them. Sometimes, the people may face huge financial loss along with some crucial data loss. Additionally, malicious URL attraction is also now becoming a part of political war [8].

Machine learning is used globally in order to prevent attacks from malicious websites. Protocol identifier is a basic characteristic of machine learning that can track the IP address or a particular domain which plays an important role in causing malware. Cyber attackers often change certain components of the URL. It deceives users and causes a massive loss of data and financial resources. It has been reported that the global cyber security scam might reach about $265 billion by 2031 [9]. Therefore, machine learning is used widely to catch suspicious URL activities. Machine learning comprises a diverse set of algorithms, such as SVM, RF, K-nearest neighbor (KNN), deep neural networks (DNNs), and voting. The algorithms such as SVM use dynamic and static analysis for detecting malwares. SVM traces runtime of executables for analysing dynamic and static analysis. Moreover, these certain algorithms are utilised for automated object characterisation after detecting the object. In order to determine the malicious URL a machine learning algorithm testing procedure is taken place. This algorithm testing procedure is done based on the Synthetic image possessing noises. Hence, in this way SVM machine learning is used to detect malicious websites [10].

Machine learning has improved information management by supporting uses by providing the facilities with smart algorithms, data-driven decision making, and others. It has been reflected that the implementation of machine learning can develop the evaluation quality by improving their data analysis with algorithms and logistics that improved the decision making in this present situation. It has been reflected that the application of the algorithm within the machine learning necrosis the memorising method along with learning that improved the future output with low mistakes. Hence, appropriate training based on the previous record also develops the understanding of factors execution for further implementation.

The evidence reflected that the implication of machine learning helps to detect the antivirus that developed the security factors to some enough extent. Machine learning methods are also used for improving the detection procedure of the phishing sites to improve the relevance and confidentiality of the research. In that case, the application of the recurrent neural network method is the key element of detecting phishing sites. It has been reflected that sending reports regarding the suspicious emails to the security team in a proper way. As an impact, its developers do antivirus executions in a proper way in this digitised era. More to the icon text, the resorts have reflected that in this present situation most of the users' faced problems related to insecure browsing that has read the rate of cyber hacking up to 15%. In that case, the implication of machine learning can help the users to classify the authentic URLs and the suspicious URLs that mitigate the rate of cyber hacking properly [11]. On the other hand, malicious software significantly emitted the URLs care impacted cyber security in this competitive age.

Ubing *et al.* [12] concentrated on improving the accuracy of phishing website detection. As a result, a feature elicitation technique was chosen and integrated with an ensemble learning approach, which is based on plurality voting and is comparable to a range of classification models such as RF, logistic regression, prediction model, and so on. According to the study, existing phishing detection technologies have an accuracy rate ranging from 70% to 92.52%.

Nagaraj *et al.* [13] conducted a comprehensive evaluation of different machine learning classifiers. Their results revealed that the RF algorithm initially demonstrated a prediction accuracy of 93%, surpassing all other machine learning algorithms considered. Notably, the RF algorithm excelled in identifying phishing websites, leading to its integration into a dual ensemble model alongside feedforward neural networks, Bagging, and boosting neural networks. This combined approach aims to create a robust and accurate predictive model for effectively classifying unknown data instances.

Kiruthiga and Akila [14] four classifiers are used; decision trees, Nave Bayes' classifiers, SVM, and neural network. Phishing URLs were detected using the classifiers. There are two steps to identifying phishing URLs. The first stage is to extract features from the URLs, and the second step is to categorize the URLs using the model that was constructed using the training set data. The decision tree that had been pruned had the maximum classification accuracy of 90%.

Altaher [15] proposes a two-stage technique that intelligently combines the KNN and the SVM algorithm. To beginning, the KNN was used as a robust and effective classifier for noisy data. Second, the SVM is used as a strong classifier. The suggested method combines the ease of use of KNN with the efficacy of SVM. When compared to other approaches, the testing findings demonstrate that the proposed hybrid strategy attained the greatest accuracy of 90.04%.

Assegie [16], a KNN based model is suggested for detecting phishing attacks by categorizing URLs as malicious or phishing and legal. The experimental results suggest that the proposed model is effective at detecting phishing attacks when accuracy metrics are used as performance measures. Overall, the proposed model has an accuracy of 85.08%. The proposed approach is useful for combating the growing number of phishing attacks in the commercial market.

The most of previous papers used a single data mining classifier with training and validation to detect phishing websites. Furthermore, several of them used numerous classifiers to detect the phishing website. Furthermore, the previous studies have proven a little slow and inaccurate in identifying the phishing website. As a result, improved approaches based on ensemble classifiers are necessary to detect phishing websites quickly and reliably [17]-[21].

In fact, when implemented in real-world circumstances such as phishing detection, adopting a single classifier in the field of machine learning may lack resilience in terms of training and validation performance. As a result, in order to improve performance prediction, a new intelligent technique that integrates many classifiers is required. Combining numerous classifiers improves classification accuracy while also making it easier for decision makers to distinguish between legitimate and illegal websites [22]-[25].

3. PROPOSED MODEL

This research aims to investigate several machine learning approaches in examine the possible uses of three ensemble classification models in detecting phishing websites. The objective is to construct an ensemble model that will be used to predict if a website is phishing or legitimate, and if so, to what degree. At this point, determining the phishing website can be considered as a data mining classification problem. The classification process is based on attributes and characteristics used to differentiate phishing sites, such as spelling errors, long URLs, personalization, prefixes, and suffixes. Various technologies are used to retrieve these attributes from input websites.

The proposed model uses multiple machine learning models to encourage all classification to be able to construct a hybrid ensemble learning model. The models used include the RF, Bagging classifier, and the SVM. Therefore, an intelligent three ensemble learning model to predict phishing websites will be designed and developed. Figure 1 depicts the general framework of the proposed phishing prediction methodology.

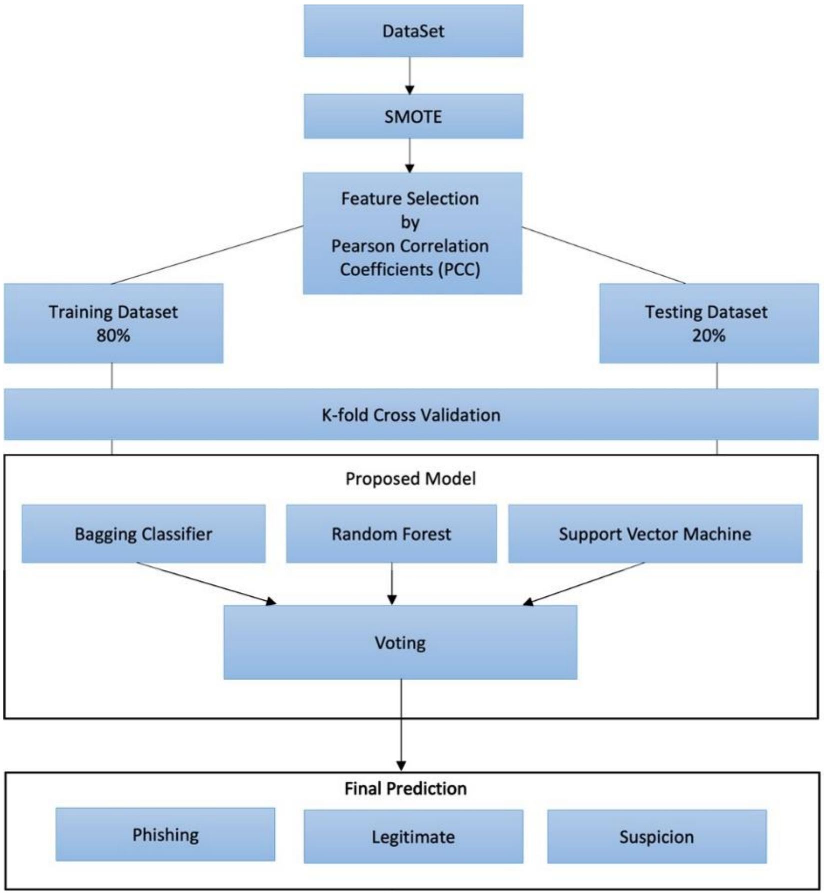

Figure 1. Proposed model method

This methodology starts with dealing with imbalanced data by using synthetic minority oversampling technique (SMOTE) balances class distributions through its reproduction of the minority classes. This technique is the simplest approach and involves duplicating examples in the minority class, although these examples do not add any new information to the model, these new data can be synthesized from the existing data. In the modelling phase, a three-classification ensemble algorithm will be developed to handle the selected most significant features.

The RF employs a stochastic strategy in constructing decision trees, where each tree is trained on randomly chosen instances and features through the random subspace method. Subsequently, predictions are generated based on the outcomes of each tree, and various aggregation methods, such as a rapid majority vote, can be employed. This randomized approach proves effective in minimizing model errors and reducing the variance in predictions.

While Bagging and RF share similarities, Bagging entails training N independent decision trees on bootstrapped subsamples without replacement, utilizing the entire feature data space concurrently. Each tree in Bagging learns independently without knowledge of the results from other trees. Similar to the RF, the final prediction for each instance is derived from the results of individual trees. In contrast, SVMs focus on maximizing the margin from the dividing plane, relying on dot products of instances rather than the instances themselves. This allows the incorporation of kernels for enhanced flexibility. These SVM approaches enable faster model training compared to Bagging and RF methods. All classifications work in parallel each one will result prediction all of three predictions will be voting to get the result. In evaluation phase which aims to assess the overall performance of the suggested classification framework, therefore, the most widely applied evaluation metrics for phishing detection problems such as classification accuracy, precision, recall, and F1 score.

4. RESEARCH RESULTS AND DISCUSSION

In this section, the evaluation results of the proposed algorithm with RF, SVM, and Bagging algorithm is explained and analyse results using a thorough attribute-relation file format (ARFF) dataset. The data set used in this paper was downloaded from the University of California, Irvine Machine Learning Repository, Centre for Machine Learning, and Intelligent Systems. It contains features from 1353 URLs. Out of these, 548 are legitimate, 702 are phishing, and 103 are suspicious.

Choosing the suitable features that will be used in the training of our model will strongly affect the performance and results. There is a need to extract the most important features among the entire features in our dataset. It is preferable to use fewer features to reduce the complexity of the model while keeping an eye on accuracy. The feature selection can be done using many methods. In our model, the filter method and specifically the correlation method are used. The idea here is to build the correlation matrix that describes the linear relationship between the features and target to gain information about the most important features of our model. The approach begins by addressing imbalanced data through the utilization of the SMOTE. This technique equalizes class distributions by replicating instances of the minority classes. This technique is the simplest approach and involves duplicating examples in the minority class, although these examples do not add any new information to the model, these new data can be synthesized from the existing data.

The purpose of this study is to explore multiple machine learning methods to assess the potential applications of three ensemble classification models for identifying phishing websites. The goal is to develop an ensemble model capable of predicting the legitimacy of a website and, if it is a phishing site, to what extent. Identifying a phishing website is treated as a classification problem in data mining. The classification relies on attributes and features that distinguish phishing sites, including spelling errors, lengthy URLs, personalization, prefixes, and suffixes. Various technologies are employed to extract these attributes from input websites. To assess the model performance, some evaluation metrics are used based on the problem domain. Accuracy, precision, recall, F1 score, and confusion matrix are calculated. To evaluate the model's results, accuracy will be calculated as the assessment metric. Accuracy is one metric for evaluating classification models. Accuracy is the fraction of predictions our model got right.
Accuracy formula:

$$Accuracy = \frac{Number\ of\ correct\ prediction}{Total\ number\ of\ prediction} \tag{1}$$

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{2}$$

after applying the proposed algorithm, the accuracy is 92.33%. Precision represents the portion of positive identifications that was correct.

$$Precision = \frac{TP}{TP+FP} \tag{3}$$

By using precision formula, it will count the number of positive identifications that was correct actually, in this research the result it predicts 91% of phishing website, 96% of suspicious websites and 89% of legitimate websites. Recall represents the portion of actual positive examples that was identified correctly.

$$Recall = \frac{TP}{TP+FN} \tag{4}$$

By using recall formula, it will count the portion of actual positive predicts that was identified correctly. In this research the result it predicts 90% of phishing website, 99% of suspicious websites and 88% of legitimate websites. The F1 score is defined as the harmonic mean of precision and recall.

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{5}$$

By using F1 score formula, it will count the harmonic mean of precision and recall. In this research the result it predicts 90% of phishing website, 97% of suspicious websites and 88% of legitimate websites. confusion matrix summarizes the results of predictions showing the number of correct and incorrect predictions with count values for each class. Confusion matrices are used to visualize important predictive analytics like recall, specificity, accuracy, and precision. Confusion matrices are useful because it gives direct comparisons of values like true positives, false positives, true negatives, and false negatives. Figure 2 shows the confusion matrix. Table 1 summarize the result of evaluating the proposed algorithm. Table 2 shows the comparison of proposed algorithm with existing algorithms.

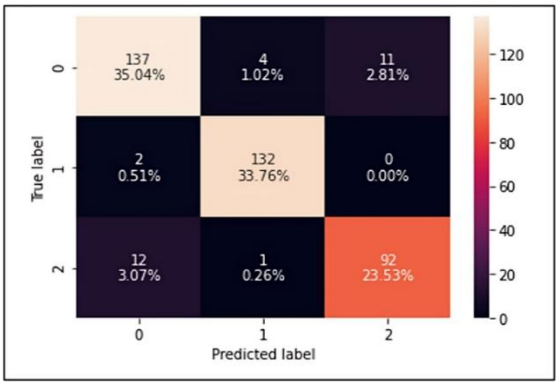

Figure 2. Confusion matrix

Table 1. Evaluation of proposed algorithm

Accuracy (%)	Precision (%)	Recall (%)	F1_Score (%)
92.33	92.13	92.09	92.10

Table 2. Comparison of the recall achieved by the proposed algorithm and other algorithm

Used approach	Suspicious (%)	Legitimate (%)	Phishing (%)	AVG (%)
Naive Bayes [13]	17.65	85.29	90.30	64.41
Neural network [13]	66.67	83.96	91.33	80.65
SVM [13]	0.00	83.02	92.67	58.56
DT [13]	13.33	88.68	92.67	64.89
KNN [16]	80.00	83.02	91.33	84.78
The hybrid KNN-SVM [15]	86.67	90.57	90.00	89.08
Proposed algorithm SVM, Bagging, and RF	99.00	88.00	90.00	92.33

5. CONCLUSION

Malicious websites are created for stealing data and different kinds of resources from any official websites. URL promotes cyber-attacks as clicking on infected network results in commencement of unauthorized data transfer process. Unethical hackers often upload malware in the websites by creating malicious URLs. This research states different detection methods and tools of phishing websites. There are two types of websites such as legitimate and phishing websites that are active in online databases. Machine learning has a huge impact in determining the websites and preventing malicious websites for preventing cyber-attack. Different aspects of machine learning algorithm framework and model are discussed which are associated with detecting phishing and legitimate websites. The research suggests new model that can be used to predict if the website is legitimate or phish by using three machine learning algorithms SVM, Bagging, and RF. The suggested approach attained the greatest accuracy of 92.33% and outperformed the other machine learning classifiers, indicating the advantage of the proposed technique.

ACKNOWLEDGEMENTS

Thanks to ALLAH who helped us in this work and thanks to the Saudi Electronic University for sponsoring this work.

REFERENCES

[1] M. A. Mos and M. M. Chowdhury, "The Growing Influence of Ransomware," in *2020 IEEE International Conference on Electro Information Technology (EIT)*, Jul. 2020, pp. 643–647, doi: 10.1109/EIT48999.2020.9208254.

[2] A. Vazhayil, R. Vinayakumar, and K. P. Soman, "comparative study of the detection of malicious URLs using shallow and deep networks," in *2018 9th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*, Jul. 2018, pp. 1–6, doi: 10.1109/ICCCNT.2018.8494159.

[3] C. Johnson, B. Khadka, R. B. Basnet, and T. Doleck, "Towards detecting and classifying malicious urls using deep learning," *Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications*, vol. 11, no. 4, pp. 31–48, 2020, doi: 10.22667/JOWUA.2020.12.31.031.

[4] A. Zervopoulos, A. G. Alvanou, K. Bezas, A. Papamichail, M. Maragoudakis, and K. Kermanidis, "Hong Kong protests: using natural language processing for fake news detection on Twitter," in *Artificial Intelligence Applications and Innovations: 16th IFIP WG 12.5 International Conference,* Cham: Springer International Publishing, 2020, pp. 408–419, doi: 10.1007/978-3-030-49186-4_34.

[5] A. Bondielli and F. Marcelloni, "A survey on fake news and rumour detection techniques," *Information Sciences*, vol. 497, pp. 38–55, Sep. 2019, doi: 10.1016/j.ins.2019.05.035.

[6] K. G. Al-Hashedi and P. Magalingam, "Financial fraud detection applying data mining techniques: a comprehensive review from 2009 to 2019," *Computer Science Review*, vol. 40, p. 100402, May 2021, doi: 10.1016/j.cosrev.2021.100402.

[7] K. Demertzis and L. Iliadis, "Cognitive web application firewall to critical infrastructures protection from phishing attacks," *Journal of Computations & Modelling*, vol. 9, no. 2, pp. 1792–8850, 2019.

[8] I. A. Sawaneh, F. K. Kamara, and A. Kamara, "Cybersecurity: a key challenge to the information age in Sierra Leone," *Asian Journal of Interdisciplinary Research*, pp. 35–46, Feb. 2021, doi: 10.34256/ajir2114.

[9] D. Vekshin, K. Hynek, and T. Cejka, "DoH insight: detecting DNS over HTTPS by machine learning," in *ACM International Conference Proceeding Series*, Aug. 2020, pp. 1–8, doi: 10.1145/3407023.3409192.

[10] X. Gao, C. Shan, C. Hu, Z. Niu, and Z. Liu, "An adaptive ensemble machine learning model for intrusion detection," *IEEE Access*, vol. 7, pp. 82512–82521, 2019, doi: 10.1109/ACCESS.2019.2923640.

[11] V. E. Adeyemo, A. O. Balogun, H. A. Mojeed, N. O. Akande, and K. S. Adewole, "Ensemble-based logistic model trees for website phishing detection," in *Advances in Cyber Security: Second International Conference*, Singapore: Springer, 2021, pp. 627–641, doi: 10.1007/978-981-33-6835-4_41.

[12] A. A. Ubing, S. Kamilia, A. Abdullah, N. Jhanjhi, and M. Supramaniam, "Phishing website detection: an improved accuracy through feature selection and ensemble learning," *International Journal of Advanced Computer Science and Applications*, vol. 10, no. 1, 2019, doi: 10.14569/IJACSA.2019.0100133.

[13] K. Nagaraj, B. Bhattacharjee, A. Sridhar, and S. GS, "Detection of phishing websites using a novel twofold ensemble model," *Journal of Systems and Information Technology*, vol. 20, no. 3, pp. 321–357, Nov. 2018, doi: 10.1108/JSIT-09-2017-0074.

[14] R. Kiruthiga and D. Akila, "Phishing websites detection using machine learning," *International Journal of Recent Technology and Engineering*, vol. 8, no. 2S11, pp. 111–114, Nov. 2019, doi: 10.35940/ijrte.B1018.0982S1119.

[15] A. Altaher, "Phishing websites classification using Hybrid SVM and KNN Approach," *International Journal of Advanced Computer Science and Applications*, vol. 8, no. 6, 2017, doi: 10.14569/IJACSA.2017.080611.

[16] T. A. Assegie, "K-nearest neighbor based url identification model for phishing attack detection," *Indian Journal of Artificial Intelligence and Neural Networking*, vol. 1, no. 2, pp. 18–21, Apr. 2021, doi: 10.54105/ijainn.B1019.041221.

[17] R. M. Qibtiah, Z. M. Zin, and M. F. A. Hassan, "Artificial intelligence system for driver distraction by stacked deep learning classification," *Bulletin of Electrical Engineering and Informatics (BEEI)*, vol. 12, no. 1, pp. 365–372, Feb. 2023, doi: 10.11591/eei.v12i1.3595.

[18] B. Chanakot and C. Sanrach, "Classifying thai news headlines using an artificial neural network," *Bulletin of Electrical Engineering and Informatics*, vol. 12, no. 1, pp. 395–402, Feb. 2023, doi: 10.11591/eei.v12i1.4228.

[19] R. Sharma, H. Pandey, and A. K. Agarwal, "Exploiting artificial intelligence for combating COVID-19: a review and appraisal," *Bulletin of Electrical Engineering and Informatics*, vol. 12, no. 1, pp. 514–520, Feb. 2023, doi: 10.11591/eei.v12i1.4366.

[20] S. A. M. Al-Juboori, F. Hazzaa, Z. S. Jabbar, S. Salih, and H. M. Gheni, "Man-in-the-middle and denial of service attacks detection using machine learning algorithms," *Bulletin of Electrical Engineering and Informatics*, vol. 12, no. 1, pp. 418–426, Feb. 2023, doi: 10.11591/eei.v12i1.4555.

[21] C. Sawangwong, K. Puangsuwan, N. Boonnam, S. Kajornkasirat, and W. Srisang, "Classification technique for real-time emotion detection using machine learning models," *IAES International Journal of Artificial Intelligence (IJ-AI)*, vol. 11, no. 4, pp. 1478–1486, Dec. 2022, doi: 10.11591/ijai.v11.i4.pp1478-1486.

[22] T. H. Prasanna, M. Shantha, A. Pradeep, and P. Mohanan, "Identification of polar liquids using support vector machine based classification model," *IAES International Journal of Artificial Intelligence (IJ-AI)*, vol. 11, no. 4, pp. 1507–1516, Dec. 2022, doi: 10.11591/ijai.v11.i4.pp1507-1516.

[23] M. Sari, M. A. Berawi, T. Y. Zagloel, and R. W. Triadji, "Machine learning model for green building design prediction," *IAES International Journal of Artificial Intelligence (IJ-AI)*, vol. 11, no. 4, pp. 1525–1534, Dec. 2022, doi: 10.11591/ijai.v11.i4.pp1525-1534.

[24] R. A. Rahma, R. A. Nugroho, D. Kartini, M. R. Faisal, and F. Abadi, "Combination of texture feature extraction and forward selection for one-class support vector machine improvement in self-portrait classification," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 13, no. 1, pp. 425–434, Feb. 2023, doi: 10.11591/ijece.v13i1.pp425-434.

[25] M. Wafi, F. A. Bachtiar, and F. Utaminingrum, "Feature extraction comparison for facial expression recognition using adaptive extreme learning machine," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 13, no. 1, pp. 1113–1122, Feb. 2023, doi: 10.11591/ijece.v13i1.pp1113-1122.

BIOGRAPHIES OF AUTHORS

Haifa Alqahtani received the master's degree from the college of information computing, majoring in cyber security from Saudi Electronic University. Her research interests in cyber security, machine learning, and malicious URL detection. She can be contacted at email: alhatlan.hayfaa@hotmail.com.

Ahmed Abu-Khadrah was born in United Arab Emirates in 1981. He received bachelor of engineering in computer engineering from Alblqa Applied University in 2003. He received the master's degree in electronic engineering (computer engineering) from Universiti Teknikal Malaysia Melaka (UTeM) in 2013. He received a Ph.D. in computer engineering and communications from Universiti Teknikal Malaysia Melaka (UTeM) in 2017. He is currently Faculty member at Department of Electrical Engineering, College of Engineering Technology, Al-Balqa Applied University, Amman, Jordan. His research interests in wireless network protocols, networking, communications, wireless mathematical model, also multimedia service over the networks. He can be contacted at email: abosuliman2@yahoo.com.