Enhancing Cybercrime Deterrence with Artificial Intelligence

Muyad Mahmmoud ALkharabsheh
Computer Science Department, The University of Jordan
Email : may8191448@ju.edu.jo

Mohammad Alshraideh
[1]Artificial Intelligence Department, The University of Jordan, Jordan,
Email : :mshridah@ju.edu.jo
[2]Information Technology College, Lusail University, Qatar
Email : mshridah@Ju.edu.jo
Imad Salah
[1]Computer Science Department, The University of Jordan
Email: isalah@ju.edu.jo
[2]Information Technology College, Lusail University, Qatar
Email : isalah@lu.edu.qa

ABSTRACT

Web phishing poses a significant security challenge for web users owing to three primary factors. First, it is easy to implement and does not require profound technical expertise in programming or networking. Second, it can be executed across various platforms, including the web, SMS, and social media platforms. Finally, this type of attack relies on social engineering, meaning that users' responses are influenced by the content presented to them. Over the past few decades, there has been a proliferation of methods and services designed for phishing detection. In this study, we introduced a novel approach to web phishing detection based on a hybrid weighted machine learning framework. Our method harnesses the capabilities of four distinct machine learning algorithms, including an unsupervised approach (K-means) and three supervised techniques. The outputs of these algorithms were strategically weighted to produce a final decision. To train and evaluate our proposed algorithm, we employed a vast dataset encompassing no content web features, totaling 111 distinct attributes. The correlations between these features and the classification outcomes were leveraged to streamline the feature set, and various correlation values were explored. Our findings from the training and validation phases underscore the significance of the correlation between the chosen features in determining the accuracy of the algorithm. In summary, our research introduces an innovative approach to combat web phishing, showcasing the potential of hybrid machine learning techniques and the critical role of feature selection through correlation analysis to enhance detection accuracy. The accuracy outcomes of the various algorithms exhibited a range of values, ranging from 0.6561 to 0.8833, across different correlation thresholds when considering all features.

Keywords - **Web Phishing, Cybersecurity, Machine Learning, Detection Methods, Cyber Threats, Security Challenges.**

Date of Submission: November 11, 2023 Date of Acceptance: December 19, 203

I. INTRODUCTION

The realm of technology is witnessing a significant surge in cybercrime, where malevolent actors exploit the personal data of Internet users to their advantage. This surge has reached unprecedented levels, resulting in annual losses amounting to billions of dollars for businesses and individuals. This trend is exacerbated by technological advancements and the growing prevalence of smart devices, which provide hackers with numerous entry points into user devices. Despite the efforts of law enforcement, cybercriminal numbers continue to rise, capitalizing on the inherent anonymity of the Internet [1].Cybercrime, in its broadest sense, encompasses criminal activities that target computers, networked devices, and networks. Cybercriminals employ these means to gain access to users' personal information, confidential business data, government records, and even the capability to deactivate devices. Various forms of cybercrime pose severe financial threats to users, including phishing, malicious software, and privacy breaches [1]. Consequently, this research focuses on the pervasive issue of phishing.

Phishing, as defined by the Anti-Phishing Working Group (APWG), is a type of cybercrime that combines social engineering and technical deception to procure user identifiers, passwords, bank account details, and other sensitive information. This is a grave threat, often resulting in data breaches, identity theft, and property damage [9]. Given the extensive scope of phishing, this study specifically focuses on phishing websites.

In line with Aslam and Rahul [4], a typical phishing attack unfolds in four stages. First, the perpetrator creates and deploys a deceptive website that is meticulously designed to mimic a legitimate counterpart. Second, as a reputable entity, the hacker transmits the forged website's URL link to their targeted victims. Third, the hacker endeavors to convince the victim to visit a fraudulent website. Finally, unsuspecting victims, lured by the deceit, click on the spurious website's link, and inadvertently provide the requested information. Subsequently, the phisher exploits the personal data to perpetrate fraud.

A critical research problem, namely, the urgent need to detect phishing websites at an early stage to safeguard web users, businesses, government entities, and banks. Phishing attacks are constantly evolving, posing a significant challenge for effective identification. Conventional methods that rely on static black and whitelisting databases struggle because new phishing websites can be created rapidly. The inability to dynamically assess new websites often results in legitimate websites being misclassified as phishing sites. To address these limitations, intelligent phishing detection approaches, particularly those employing machine learning techniques, have gained prominence. Machine learning, a subset of artificial intelligence, enables computers to learn from data and adapt autonomously, thus making it a promising tool for combating the ever-changing nature of phishing threats.

The significance of this research stems from the widespread menace of phishing, where cybercriminals exploit personal information for financial gain and identity theft, causing significant disruptions in businesses worldwide. Phishing attacks involve creating deceptive replicas of legitimate websites and emails, often impersonating financial institutions to deceive users into disclosing confidential data. The flexibility of HTML allows for the replication of graphical elements and entire websites, making fraudulent communications highly convincing [19].

While previous anti-phishing studies have explored domain characteristics, such as website URLs and content, this research underscores the need for effective tools to detect malicious URLs and protect users. Machine learning techniques offer a promising avenue for identifying malicious URLs on the Internet [5].

The primary objective of this study is to compare and evaluate various machine learning algorithms, with a particular emphasis on key features that can enhance the quality of detection. Detecting malicious websites is treated as a binary classification task that distinguishes between malicious and legitimate sites. Additionally, this study incorporates classification and clustering techniques to improve the detection accuracy and automate the detection process.

II. LITERATURE REVIEW AND BACKGROUND

A. Background

Phishing is a deceptive strategy that combines social engineering and technology to fraudulently obtain sensitive information, such as passwords and credit card numbers, by posing as a trustworthy entity in electronic communications. Phishers often employ fake emails that mimic real messages from reputable sources like financial institutions or e-commerce platforms. The goal is to persuade users to visit fraudulent websites through links provided in these phishing emails[4].

Phishing attacks typically follow four distinct phases:

1. The phisher creates a fake website that closely resembles the original.
2. Phishers send a large volume of deceptive emails, impersonating legitimate companies and organizations, in order to convince recipients to visit their fraudulent websites.
3. Victims access the counterfeit website by clicking on the provided link and inputting their personal information.
4. Phishers then use this acquired information to commit fraudulent activities, such as unauthorized fund transfers from victims' accounts[4].

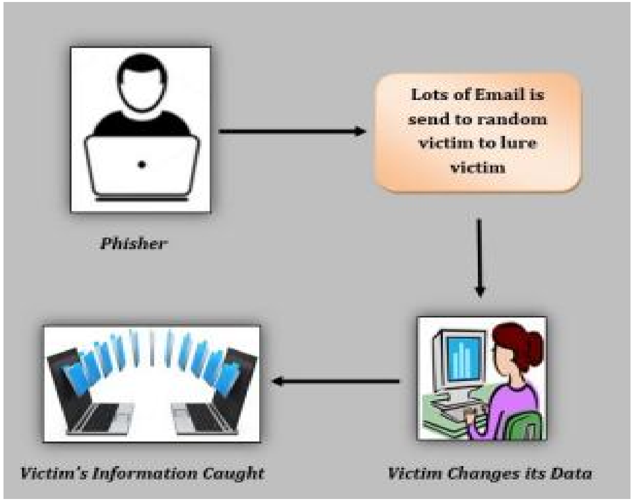

Lots of Email is send to random victim to lure victim
Phisher
Victim's Information Caught
Victim Changes its Data

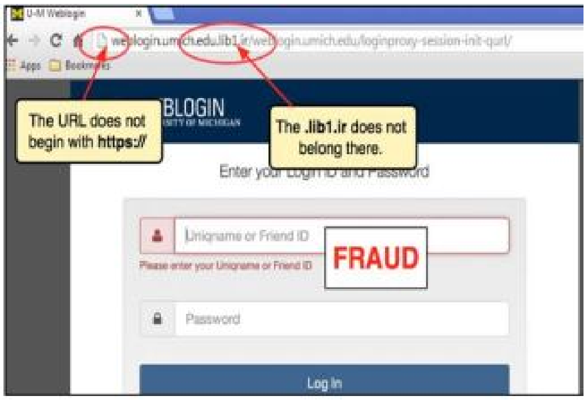
Figure 1: Process of Phishing

Fig.1, as presented in the work by Aslam and Rahul ([4], illustrates the phishing process. Detecting and avoiding phishing URLs has become increasingly challenging due to the proliferation of phishing campaigns, which enhance fraudsters' efficiency and their ability to conceal their activities using fraudulent tools. Phishing websites pose a significant threat to both consumers and institutions, jeopardizing electronic transactions. A phishing URL is designed to either download malware, initiate phishing attacks, or manipulate search engine results. Criminals continually enhance their technological skills, enabling them to build more robust infrastructures to sustain phishing attacks [3].

Every URL follows a common syntax (<protocol>://<hostname><path>). In a phishing attack, the target is typically enticed into clicking on a URL leading to a phishing site. Adversaries often obscure this URL using various obfuscation techniques. Some of the most common obfuscation techniques include:

1. Type 1: Host obfuscation using an IP address. In this type of attack, the URL's hostname is replaced with an IP address, and the path often includes the target organization. The IP address is sometimes expressed in hex or decimal format, in addition to the dotted quad form.
2. Type 2: Host obfuscation using another domain. Here, the URL's host contains a domain name that appears legitimate, while the path contains the phished organization. This technique aims to mimic URLs through redirects to make them appear genuine.
3. Type 3: Host obfuscation using long hostnames. In this type of attack, the host contains the phished organization, but a lengthy string of words and domains is appended after the hostname.
4. Type 4: Usage of unknown or misspelled domains. There is no apparent connection to the phishing organization, and the domain name is intentionally misspelt.

These obfuscation techniques are employed by adversaries to deceive users and make phishing URLs appear legitimate [3].


U-M Weblogin
weblogin.umich.edu.lib1.ir/weblogin.umich.edu/loginproxy-session-init-qurl/
Apps
Bookmarks
BLOGIN
SITY OF MICHIGAN
Enter your Login ID and Password
Uniqname or Friend ID
Please enter your Uniqname or Friend ID
Password
Log In

The URL does not begin with https://

The .lib1.ir does not belong there.

FRAUD

Figure 2: Example of a phishing website.

Fig.2, as presented in the work by Aslam and Rahul ([4], provides an example of a phishing website. Phishing website attacks have emerged as one of the most pervasive and damaging threats on the internet, inflicting immeasurable harm. This proliferation can be attributed to the rapid expansion of online financial services and e-commerce. In recent years, the discovery of phishing web pages has seen a significant uptick. Both online financial institutions and their customers must gain a comprehensive understanding of phishing and anti-phishing technology while implementing security measures to thwart phishing websites (Aslam & Rahul, 2018).

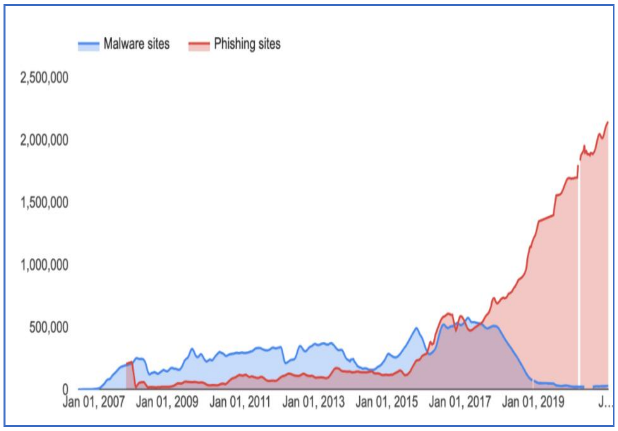


Figure 3: Number of websites considered phished between January 2016 and January 2021.

Fig.3, sourced from Google Safe Browsing, illustrates the growth in the number of websites classified as phishing websites between January 2016 and January 2021, as discussed in the study by [8]. This visual representation underscores the alarming increase in phishing websites over this period, emphasizing the growing threat posed by such malicious sites.

B. Related Works

Phishing website detection using machine learning.

This section addresses a critical research problem, namely, the pressing need for early detection of phishing websites to safeguard web users, businesses, government entities, and banks. Phishing attacks continuously evolve, presenting a significant challenge to effective identification. Traditional methods that rely on static black and whitelisting databases struggle because new phishing websites can be rapidly created. This inability to dynamically assess new websites often results in legitimate websites being erroneously classified as phishing sites. To overcome these limitations, intelligent phishing detection approaches, particularly those utilizing machine learning techniques, have gained prominence. Machine learning, a subset of artificial intelligence, empowers computers to learn from data and adapt autonomously, making it a promising tool for combatting the ever-changing nature of phishing threats.

The significance and motivation for this research stem from the widespread menace of phishing, where cybercriminals exploit personal information for financial gain and identity theft, causing significant disruptions to businesses worldwide. Phishing attacks involve crafting deceptive replicas of legitimate websites and emails, often impersonating financial institutions, to deceive users into disclosing confidential data. The flexibility of HTML enables the replication of graphical elements and entire websites, rendering these fraudulent communications highly convincing[19].

In [16] Wormhole attacks, once thought confined to Adhoc networks, now pose a threat to infrastructure-based wireless LANs. Our defense, reliant on shared information among Access Points, effectively mitigates the risk without needing location data or clock synchronization.

In [17] Wireless Sensor Networks (WSNs) find applications in diverse fields, yet their sensor nodes face vulnerability to security attacks, notably DoS jamming attacks. This work introduces a stepwise approach using Improved Exponentially Weighted Moving Average (IEWMA) with the Packet Break Advent Time (BAT) feature to efficiently detect jamming attacks in WSNs. Results from trace-driven simulations confirm the solution's effectiveness with minimal overhead.

In [18] Cyber-attacks demand comprehensive cybersecurity education. This research presents a card deck gaming platform, Cyber Awareness Learning Imitation Environment, enabling non-IT professionals to simulate and learn about various cyber threats. The system covers attacks like malware and phishing, with player feedback collected for understanding assessment.

In [20] Efficient cybersecurity requires robust malware detection. This study introduces a model using machine learning classifiers and a novel feature selection technique, genetic programming. Comparative analysis favors Random Forest, Random Forest (4), and Random Tree for superior performance. The proposed feature selection method, GPMP, outperforms Filter-based techniques with an accuracy of 0.881066 and an F1-score of 0.867546, utilizing fewer features for reduced computational complexity.

In [21] Malware, or harmful software, continually evolves, demanding effective classification methods. This paper introduces two Genetic Programming (GP)-based feature selection methods, GPM and GPMP, for malware prediction. Comparative analysis against popular techniques reveals their superior performance in accuracy and F-score, with reduced computation time. Evaluation on four datasets using Random Forest and Decision Tree classifiers demonstrates their efficiency, particularly with the Random Forest classifier.

While previous anti-phishing studies have explored domain characteristics such as website URLs and content, this research underscores the necessity for effective tools to detect malicious URLs and protect users. Machine learning techniques offer a promising avenue for identifying malicious URLs on the internet[6]).

The primary objective of this research is to compare and evaluate various machine learning algorithms, with a specific focus on key features that can enhance the quality of detection. Detecting malicious websites is approached as a binary classification task, distinguishing between malicious and legitimate sites. Additionally, the study incorporates classification and clustering techniques to enhance detection accuracy and automate the detection process.

Phishing Website Detection Using Deep Learning

In their 2021 research, Khana and Rana's [7] main focus was on developing robust anti-phishing models using innovative techniques. They incorporated ten distinct features and applied three key methods: Long Short-Term Memory (LSTM), Deep-Neural Network (DNN), and Convolution-Neural Network (CNN). Impressively, the LSTM-based model achieved a remarkable accuracy score of 98.67%, the DNN model reached 96.33%, and the CNN model scored 97.23%. These methods brought about significant improvements in the field of phishing detection.

The threat of phishing has become increasingly serious, resulting in substantial financial losses and data breaches. Traditional anti-phishing approaches often require experts to manually extract specific website elements and rely on third-party services, causing delays in detection. In [13] their 2021 research, Rundong et al. proposed an integrated approach that utilized convolutional neural networks (CNN) and random forest (RF) without depending on third-party services. Their model achieved an impressive accuracy rate of 99.35% by transforming URLs into matrices, extracting features through CNN, and employing RF classifiers.

In [10] study by Ping and colleagues, the primary focus was on leveraging deep learning to identify phishing websites. They introduced two categories of features, known as original and interaction features, and implemented a detection model based on the Deep Belief Network (DBN). This DBN model demonstrated an exceptional 90% true-positive rate and an impressively low false-positive rate of 0.6% in tests conducted with actual ISP IP traffic data.

Table 1: Summary of Related Works.

Reference	**Classification Method**	**Results**
[2]	Decision tree	90.39%
[22]	XGBoost	0.98%
[15]	Ensemble three classifiers Random Forest, SVM, and Decision Tree	98.52 %
[12]	Random forest	99.55%
[11]	Random forest	97.14%
[5]	Decision Tree	98.4%
[7]	Long Short-Term Memory	98.67%
[13]	Convolutional neural networks with random forest	99.35%
[10]	Deep Belief Networks	90%

C. *Machine Learning Methods*

1) *K-means Clustering:*

- K-means clustering is particularly useful when you want to group data points into clusters based on their similarity. It is widely used in applications like customer segmentation, image compression, and anomaly detection.
- The first step in K-means is determining the number of clusters (K) you want to create. This

can be done using various techniques, such as the elbow method or silhouette score.

- After determining K, the algorithm initializes cluster centroids. These centroids represent the centre of each cluster, although they are initially chosen randomly.
- Data points are then assigned to the nearest centroid based on a distance metric, often using the Euclidean distance formula.
- Once data points are assigned to clusters, the centroids are recalculated as the mean of all data points within each cluster.
- Steps 3 and 4 are repeated iteratively until the centroids no longer change significantly, indicating convergence.
- K-means is efficient for large datasets and provides interpretable results, but it assumes that clusters are spherical and equally sized, which may not always be the case.

2) *Logistic Regression:*

- Logistic regression is primarily used for binary classification tasks, where the target variable has two possible outcomes, such as spam detection (spam or not spam) or disease diagnosis (positive or negative).
- It models the relationship between the independent variables (features) and the probability of a data point belonging to the positive class.
- Logistic regression uses the sigmoid function to constrain the output probability between 0 and 1. The logistic curve smoothly transitions from 0 to 1, making it suitable for classification.
- The model learns coefficients for each feature to maximize the likelihood of observing the given data.
- Unlike linear regression, logistic regression deals with categorical outcomes and focuses on class separation rather than predicting a continuous value.
- Logistic regression can be extended to handle multi-class classification through techniques like one-vs-all (OvA) or SoftMax regression.

3) *Random Forest:*

- Random forest is an ensemble learning method that combines multiple decision trees to make predictions. Each decision tree is trained on a bootstrapped subset of the data (bagging), and randomness is introduced during feature selection.
- Random Forest's strength lies in its ability to handle high-dimensional data, noisy data, and datasets with complex relationships.
- The model produces a robust prediction by averaging the outputs of individual decision trees, reducing the risk of overfitting.
- It also provides a measure of feature importance, allowing you to identify which features contribute the most to predictions.
- Random forest can be used for both classification and regression tasks, making it versatile in various domains, including finance, healthcare, and natural language processing.

4) *Support Vector Machine (SVM):*

- SVM is a powerful classification algorithm known for its effectiveness in both linearly and non-linearly separable data.
- SVM aims to find a hyperplane that maximizes the margin between classes. The margin is the distance between the hyperplane and the nearest data points (support vectors).
- For non-linearly separable data, SVM can transform the data into a higher-dimensional space using a kernel trick (e.g., polynomial, or radial basis function kernels). This transformation makes the data linearly separable in the new space.
- SVM is effective in handling high-dimensional data, such as text classification and image recognition.
- It has fewer hyperparameters to tune compared to some other algorithms, making it relatively easy to use and adapt to different tasks.

These machine-learning algorithms are fundamental tools in data science and play vital roles in solving a wide range of real-world problems. The choice of algorithm depends on the nature of the data, the task at hand, and the desired model interpretability, among other factors. Researchers and practitioners continue to explore and develop these algorithms to enhance their capabilities and address emerging challenges in machine learning.

III. Research Methodology

This Section introduces a novel approach to developing an intelligent phishing detection model using data mining algorithms in an ensemble. The primary objective is to determine whether a website is engaged in phishing activities. The study explores various machine learning methods, focusing on four classification models, to identify phishing websites effectively. The aim is to create an ensemble model capable of predicting whether a website is legitimate or involved in phishing activities.

Phishing website detection is treated as a data mining classification problem, with the class attribute being "phishing." The classification process relies on feature analysis to differentiate between legitimate and phishing sites.

The anticipated outcomes of this research hold the potential to pave the way for new investigations in the domain of phishing website prediction and detection, particularly using ensemble and data mining approaches. To achieve this, a sophisticated two-layer ensemble learning model will be developed to forecast phishing websites.

This research aims to contribute to the advancement of intelligent systems capable of safeguarding users against online threats, further enhancing cybersecurity measures in the digital landscape. Fig.4 provides an overview of the proposed phishing prediction methodology's workflow.

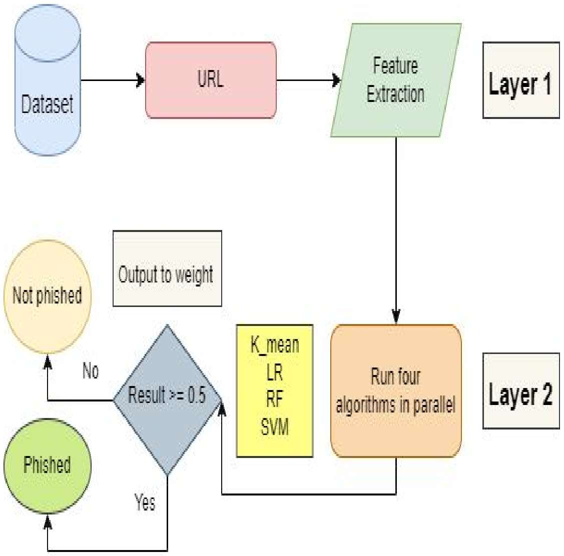


Figure 4: flowchart of the proposed phishing prediction methodology.

To begin, we initiate the process by taking the URL and implementing the Feature Extraction technique to select the most crucial features. These features are chosen based on their correlation, ensuring that they exhibit a correlation coefficient greater than 0.6 with other features.
Next, we execute four algorithms concurrently: K-means clustering, Logistic Regression (LR), Random Forest (RF), and Support Vector Machine (SVM). Suby, Ulate the value "**Result**".

$$Result\ =\ 0.4\ *\ K\_mean + 0.2 * LR + 0.2 * RF + 0.2 * SVM \ldots\ldots eq(1)$$

Classify the URL as either phishing or not phishing, we employ a threshold:

- *if* $Result \geq$ 0.5, it is designated as a phishing website.
- *if* $Result <$ 0.5, it is classified as a non-phishing website.

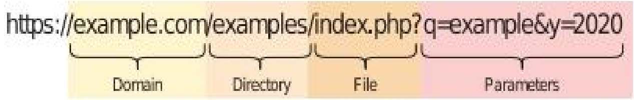
The constants used in eq.1 have been meticulously fine-tuned to attain the highest accuracy in our model. This optimization process ensures the effectiveness of our approach in distinguishing phishing websites from legitimate ones.

Dataset

In this research, we utilize a dataset consisting of 88,647 websites that have been categorized as either legitimate or phishing. This extensive dataset serves as the foundation for training classification models, creating phishing detection systems, and conducting association rule mining.

The data in this study was gathered and compiled by [14] to develop and analyze several categorization algorithms for identifying phishing websites using URL characteristics, URL resolving metrics and external services. Six groups of characteristics may be found in the prepared dataset:

- Attributes that are dependent on the whole of the URL's features (20 attributes).
- Attributes that are based on the features of the domain (21 attributes).
- Attributes depend on the features of the URL directory (18 attributes).
- Attributes depend on the features of the URL file (18 attributes).
- Attributes depend on the features of URL parameters (20 attributes).
- Attributes based on URL-resolving data and external metrics (16 attributes).



The values of the first four groups are based on the values of the attributes on the whole URL string, whilst the values of the next four groups are based on specific sub-strings. The final set of characteristics is based on URL resolve metrics as well as external services like the Google search index. The dataset has 111 variables in total, except the target phishing attribute, which indicates if a specific case is genuine (value 0) or phishing (value 1).

The dataset was offered in two different versions, and we opted for the second variant, which comprises a total of 88,647 instances. Among these instances, 30,647 are classified as *phishing* websites, while the remaining 58,000 are marked as *legitimate* sites. This choice was made to emulate a real-world scenario where the number of authentic websites typically outweighs phishing sites. In both versions of the dataset, the classes are distributed as follows: the data is represented by binary values, consisting of zeros and ones, while the outputs are categorized as either zero or one.

Experimental Results and Evaluation

In this Section, we delve into several key aspects of this research. We begin by discussing the dataset's size, providing insights into its scale and composition. Subsequently, we introduce the evaluation metrics employed to gauge the performance of our machine learning algorithms. Finally, we present and analyze the outcomes of various experiments involving feature selection and correlation, shedding light on the impact of these experiments on our results and findings.

4.1 Dataset Splitting

Before partitioning the dataset, we conducted a series of experiments aimed at feature selection based on correlations. These experiments encompassed different thresholds for feature inclusion, including all features, those with a correlation exceeding *0.20*, those surpassing *0.50*, those exceeding *0.60*, and those with a correlation value exceeding *0.75*. Table 2 provides an overview of the number of features included in each of these experiments, offering valuable insights into the scope of our feature selection process.

Table 2: Number of Features based on Correlation Value.

All Features	Features Correlation > 20%	Features Correlation > 50%	Features Correlation > 60%	Features Correlation > 75%
111	62	33	31	19

Subsequently, the dataset is divided into training and testing sets using the following criteria: *70%* of the entire dataset is allocated to the training set, while the remaining *30%* constitutes the testing set. This results in 62,053 instances within the training dataset and 26,594 instances within the testing dataset.

In the assessment of these algorithms, we employed a widely used evaluation metric known as Accuracy. Accuracy is calculated as the ratio of correctly predicted instances to the total number of instances, as illustrated in Equation (2) below:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad \dots\dots\dots \quad eq(2)$$

Here, the terms are defined as follows:

- *TP* (True Positive) represents the correctly predicted instances categorized as genuine with label 0.
- *TN* (True Negative) represents the correctly predicted instances categorized as phishing with label 1.
- *FN* (False Negative) corresponds to instances where the algorithm predicts label 1, but the actual label is 0.
- *FP* (False Positive) corresponds to instances where the algorithm predicts label 0, but the actual label is 1.

4.1 Results and Discussion

This section provides the Confusion Metrics and accuracy results for each of the previously mentioned algorithms in the correlation experiments conducted for the five machine learning algorithms.

All Features: Confusion Metrics and Accuracy Results Table 3 displays the confusion metrics for each algorithm in the initial experiment where all features were selected.

Table 3:Confusion Metrics for Machine Learning Algorithms in All Features

Algorithm	K-mean			RF			SVM			LG			Proposed		
		Prediction Label			Prediction Label		Prediction Label			Prediction Label			Prediction Label		
		0	1		0	1		0	1		0	1		0	1
Actual Label	0	56773	29255	0	46762	747	0	49854	13277	0	55003	7345	0	40306	258
	1	1227	1392	1	11238	29900	1	8146	17370	1	2997	23302	1	17694	30389

Fig.5 presents the accuracy results for five machine learning algorithms, namely K-means, Random Forest (RF), Support Vector Machine (SVM), Logistic Regression (LG), and the Proposed algorithm, across all feature experiments.

The respective accuracy scores for these algorithms were as follows: 0.6561, 0.8648, 0.6543, 0.8833, and 0.7975. In this Table 3:Confusion Metrics for Machine Learning Algorithms in All Features

experiment, the LG algorithm demonstrated the highest accuracy when compared to the other algorithms.

For instance, taking RF as an example, which is previously detailed in Table 3 the following terminology was used for the confusion metrics:

True Positives (TP): The count of instances correctly predicted as genuine (label 0), which amounted to 46,762 instances.

True Negatives (TN): The count of instances correctly predicted as phishing (label 1), totalling 29,900 instances.

False Negatives (FN): The count of instances predicted as phishing (label 1) but belonging to the genuine class (label 0), with a value of 747 instances.

False Positives (FP): The count of instances predicted as genuine (label 0) but being phishing instances (label 1), amounted to 11,238 instances.

The accuracy was calculated using the following formula for RF:

Accuracy = (TP + TN) / (TP + TN + FP + FN) = (46,762 + 29,900) / (46,762 + 29,900 + 11,238 + 747) = 76,662 / 88,647 ≈ 0.8648.

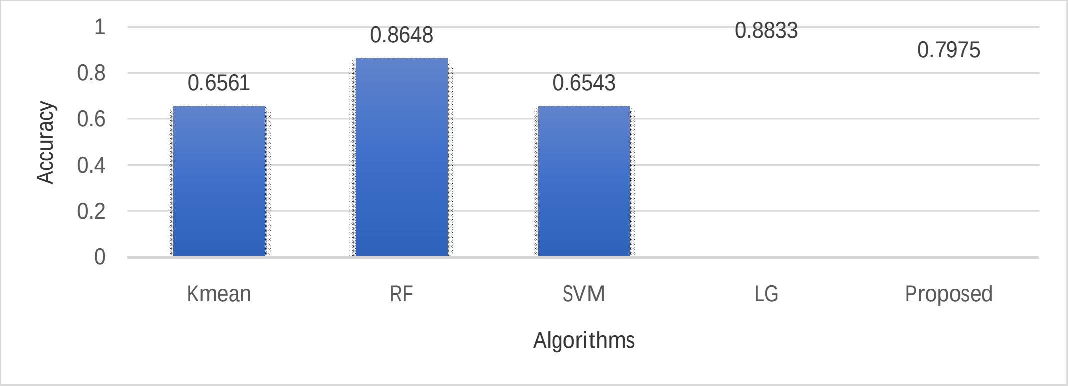


Figure 5: Machine Learning Algorithms Results with all Features

As explained, the study focuses on analyzing the impact of features with a correlation greater than 20% in a second experiment.

Table 4 presents the confusion metrics for each algorithm when 62 such features were selected.

Fig.6 then illustrates the accuracy results obtained from five machine learning algorithms (K-means, Random Forest - RF, Support Vector Machine - SVM, Logistic Regression - LG, and the Proposed algorithm) in this experiment. The accuracy rates achieved when these 62 features were selected are as follows: 0.6776, 0.869, 0.7645, 0.8861, and 0.8907, respectively. Notably, the Proposed algorithm demonstrated the highest accuracy among the algorithms in this experiment.

To provide a concrete example, consider the RF algorithm, as previously detailed in Table 4 the True Positives (TP) represent the instances correctly predicted as genuine (label 0), totalling 46,982 instances, while the True Negatives (TN) denote instances correctly predicted as phishing (label 1), amounting to 29,945 instances. On the other hand, the False Negatives (FN) represent instances predicted as phishing (label 1) but belonging to the genuine class (label 0), with a count of 702 instances. The False Positives (FP) signify instances predicted as genuine (label 0) but being phishing instances (label 1), totalling 11,018 instances.

The accuracy was calculated using the formula (RF as an example): *(TP + TN) / (TP + TN + FP + FN) = (46,982 + 29,945) / (46,982 + 29,945 + 11,018 + 702) = 77,727 / 88,647 ≈ 0.869.*

This demonstrates the method used to determine accuracy, with RF achieving an accuracy rate of approximately 0.869 in this experiment.

Table 4: Confusion Metrics for Machine Learning Algorithms in Features Correlation Greater than 20%

Algorithm	**K-mean**			**RF**			**SVM**			**LG**			**Proposed**		
	Prediction Label			**Prediction Label**			**Prediction Label**			**Prediction Label**			**Prediction Label**		
		0	1		0	1		0	1		0	1		0	1
Actual Label	0	57623	28235	0	46982	702	0	49954	12830	0	55205	7300	0	48566	262
	1	377	2412	1	11018	29945	1	8046	17817	1	2795	23347	1	9434	30386

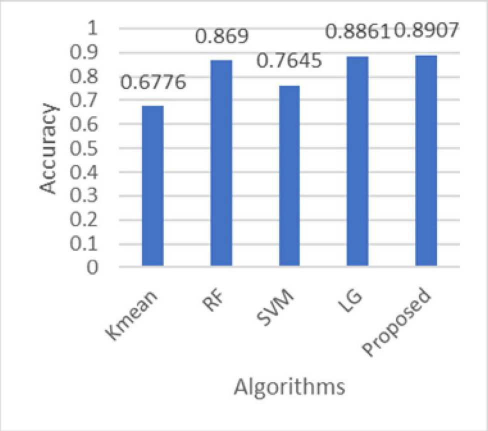


Figure 6: Machine Learning Algorithms Results with Features Correlation Greater than 20%.

Also, the analysis of features with a correlation exceeding 50% in a second experiment. Table 5 presents the confusion metrics for each algorithm when 33 highly correlated features were selected.

Fig.7 showcases the accuracy results attained from five machine learning algorithms (K-means, Random Forest - RF, Support Vector Machine - SVM, Logistic Regression - LG, and the Proposed algorithm) within this experiment. The accuracy rates achieved when these 33 features were selected are as follows: 0.7369, 0.9016, 0.8156, 0.9074, and 0.9688, respectively. Once again, the Proposed algorithm stands out with the highest accuracy among the algorithms in this experiment.

For instance, taking the RF algorithm as an example, as elaborated in Table 5, the True Positives (TP) represent the instances correctly predicted as genuine (label 0), totaling 49,709 instances. Meanwhile, the True Negatives (TN) denote instances correctly predicted as phishing (label 1), amounting to 30,136 instances. Conversely, the False Negatives (FN) signify instances predicted as phishing (label 1) but belonging to the genuine class (label 0), with a count of 511 instances. The False Positives (FP) indicate instances predicted as genuine (label 0) but being phishing instances (label 1), totalling 8,291 instances.

The accuracy was computed using the formula (RF as an example): (TP + TN) / (TP + TN + FP + FN) = (49,709 + 30,136) / (49,709 + 30,136 + 8,291 + 511) = 80,647 / 88,647 ≈ 0.9016. This illustrates the method employed for determining accuracy, with RF achieving an accuracy rate of approximately 0.9016 in this specific experiment.

Table 5: Confusion Metrics for Machine Learning Algorithms in Features Correlation Greater than 50%

Algorithm	**K-mean**			**RF**			**SVM**			**LG**			**Proposed**		
	Prediction Label			**Prediction Label**			**Prediction Label**			**Prediction Label**			**Prediction Label**		
		0	1		0	1		0	1		0	1		0	1
Actual Label	0	57683	23035	0	49709	511	0	52514	10860	0	56790	7000	0	55430	200
	1	317	7612	1	8291	30136	1	5486	19787	1	1210	23647	1	2570	30447

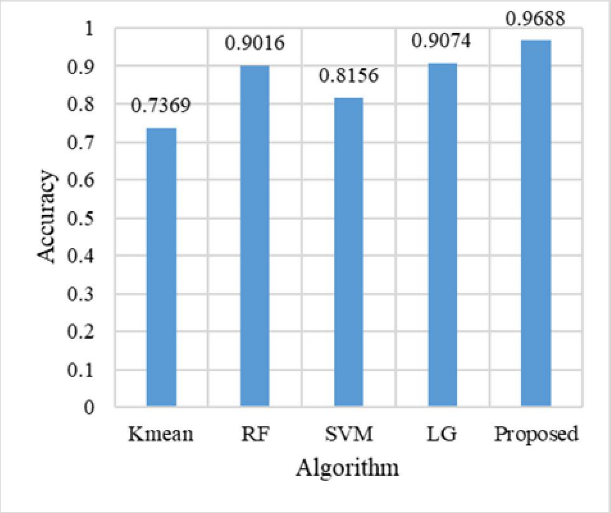


Figure 7: Machine Learning Algorithms Results with Features Correlation Greater than 50%.

Also, the study explores features with a correlation exceeding 60% in a second experiment. Table 6 displays the confusion metrics for each algorithm when 31 of these highly correlated features were selected.

Subsequently, Table 6 and Fig.8 depict the accuracy results obtained from five machine learning algorithms (K-means, Random Forest - RF, Support Vector Machine - SVM, Logistic Regression - LG, and the Proposed algorithm) within this experiment. The accuracy rates achieved when these 31 features were selected are as follows: 0.7369, 0.9306, 0.8276, 0.9277, and 0.9951, respectively. Once again, the Proposed algorithm stands out with the highest accuracy among the algorithms in this experiment.

For instance, considering the RF algorithm, as previously detailed in Table 6 the True Positives (TP) represent the instances correctly predicted as genuine (label 0), totalling 52,000 instances. Meanwhile, the True Negatives (TN) denote instances correctly predicted as phishing (label 1), amounting to 30,445 instances. On the other hand, the False Negatives (FN) signify instances predicted as phishing (label 1) but belonging to the genuine class (label 0), with a count of 202 instances. The False Positives (FP) indicate instances predicted as genuine (label 0) but being phishing instances (label 1), totalling 6,000 instances.

The accuracy was calculated using the formula (RF as an example): (TP + TN) / (TP + TN + FP + FN) = (52,000 + 30,445) / (52,000 + 30,445 + 6,000 + 202) = 83,245 / 88,647 ≈ 0.9306. This demonstrates the method employed to determine accuracy, with RF achieving an accuracy rate of approximately 0.9306 in this specific experiment as shown in Fig.6.

Table 6: Confusion Metrics for Machine Learning Algorithms in Features Correlation Greater than 60%

Algorithm	K-mean			RF			SVM			LG			Proposed		
	Prediction Label			Prediction Label			Prediction Label			Prediction Label			Prediction Label		
		0	1		0	1		0	1		0	1		0	1
Actual Label	0	57683	23035	0	52000	202	0	52514	9800	0	56790	5200	0	57660	91
	1	317	7612	1	6000	30445	1	5486	20847	1	1210	25447	1	340	30556

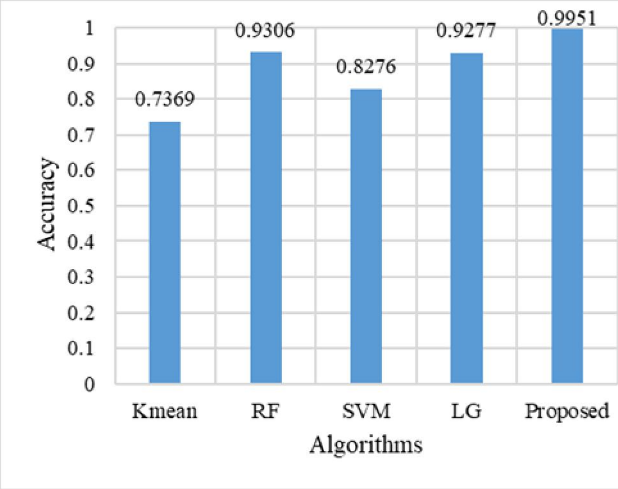


Figure 8: Machine Learning Algorithms Results with Features Correlation Greater than 60%.

Also, the study delves into the examination of features with an even higher correlation, surpassing 75%, in a second experiment. Table 7 provides an overview of the confusion metrics for each algorithm when 19 of these strongly correlated features were selected.

Following that, Fig.9 presents the accuracy results derived from five machine learning algorithms (K-means, Random Forest - RF, Support Vector Machine - SVM, Logistic Regression - LG, and the Proposed algorithm) within this experiment. The accuracy rates achieved when these 19 features were selected are as follows: 0.5375, 0.5799, 0.5606, 0.7356, and 0.6966, respectively. In this specific experiment, it is noteworthy that the LG algorithm demonstrated the highest accuracy among the algorithms.

For instance, focusing on the RF algorithm as described in Table 7, the True Positives (TP) represent the instances correctly predicted as genuine (label 0), totalling 33,560 instances. The True Negatives (TN) correspond to instances correctly predicted as phishing (label 1), amounting to 17,847 instances. Conversely, the False Negatives (FN) denote instances predicted as phishing (label 1) but belonging to the genuine class (label 0), with a count of 12,800 instances. The False Positives (FP) indicate instances predicted as genuine (label 0) but being phishing instances (label 1), totalling 24,440 instances.

The accuracy was computed using the formula (RF as an example): (TP + TN) / (TP + TN + FP + FN) = (33,560 + 17,847) / (33,560 + 17,847 + 24,440 + 12,800) = 51,407 / 88,647 ≈ 0.5799. This illustrates the method employed to calculate accuracy, with RF achieving an accuracy rate of approximately 0.5799 in this experiment.

Table 7: Confusion Metrics for Machine Learning Algorithms in Features Correlation Greater than 75%

Algorithm	K-mean			RF			SVM			LG			Proposed		
	Prediction Label			Prediction Label			Prediction Label			Prediction Label			Prediction Label		
		0	1		0	1		0	1		0	1		0	1
Actual Label	0	47682	30200	0	33560	12800	0	47700	28205	0	46650	12166	0	40006	8900
	1	10318	447	1	24440	17847	1	10300	2442	1	11350	18481	1	17994	21747

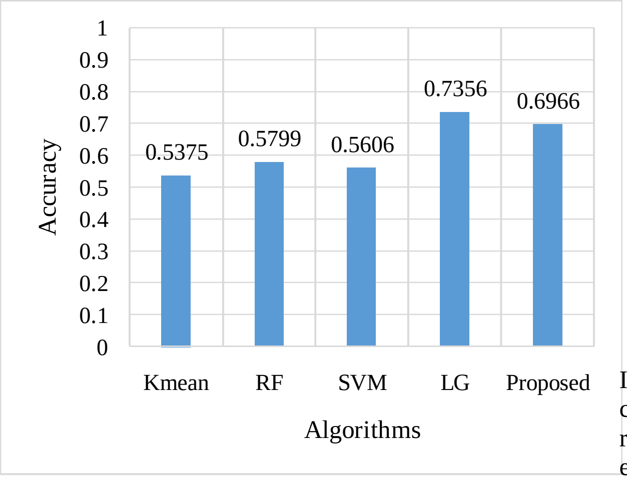


Figure 9: Machine Learning Algorithms Results with Features Correlation Greater than 75%.

As depicted in the preceding tables and figures, the highest accuracy outcomes in the All-Correlation experiment were achieved when 31 features with a correlation value exceeding 60% were selected. Furthermore, the proposed algorithm consistently delivered the most accurate predictions across three experiments, while LG demonstrated the best accuracy results in two experiments.

The following will discuss the research questions.

1. **Reasons for Website Phishing Attacks and Avoidance**: The discussion begins by outlining the reasons behind successful website phishing attacks. These reasons include a lack of user security awareness, criminals following financial incentives, and insufficient due diligence by organizations. To mitigate such attacks, security measures like recognizing phishing indicators, refraining from clicking on suspicious links, and regularly updating passwords are recommended.
2. **Role of Machine Learning**: The study explores how machine learning assists in achieving security goals. It emphasizes the application of state-of-the-art machine learning algorithms to create robust models for predicting and detecting phishing attacks. Machine learning is harnessed to protect data from various hacking attempts.
3. **High Performance with Low Error Rate**: The study confirms that high performance with a low error rate is attainable. By developing and selecting appropriate algorithms, particularly those tailored to detecting and classifying phishing attacks, the study achieves high accuracy and a low error rate. Features correlation experiments also contribute to accuracy improvement.
4. **Intelligent Addressing of Website Phishing**: The proposed solution intelligently addresses website phishing by achieving high-performance accuracy in the detection and classification of phishing attacks. The discussion highlights the effectiveness of the approach.
5. **Comparison with Related Works**: The proposed solution is shown to outperform related works in the field of phishing detection. It is highlighted that the proposed solution provides the highest level of performance accuracy when compared to other research papers. An example is given where the proposed solution achieved an accuracy value of 99.51%, surpassing the accuracy of a comparable study which used the Random Forest algorithm.

In essence, this section of the study presents a comprehensive discussion of the research findings in response to the research questions, highlighting the effectiveness of the proposed methods in addressing website phishing attacks and emphasizing their superior performance compared to related works in the field.

V Conclusions and Future Works

This research focused on the detection of email phishing attacks using five distinct machine learning algorithms, leveraging a well-established dataset related to phishing attacks. The algorithms employed were K-means, Random Forest, Support Vector Machine, Logistic Regression, and a Proposed algorithm. The dataset, containing 88,647 instances and 111 features, underwent a feature selection process based on correlation values. This experiment was conducted at various correlation thresholds: 1) using all features, 2) features with a correlation value greater than 20% (62 features), 3) features with a correlation value greater than 50% (33 features), 4) features with a correlation value greater than 60% (31 features), and 5) features with a correlation value greater than 75% (19 features). The dataset was divided into a 70% training set and a 30% testing set to build and evaluate the performance of the machine learning models.

The accuracy results for each algorithm at different correlation thresholds were as follows:

1. All features: Varied accuracy values ranging from 0.6561 to 0.8833.

2. Features with correlation > 20%: Accuracy ranged from 0.6776 to 0.8907.

3. Features with correlation > 50%: Accuracy varied from 0.7369 to 0.9688.

4. Features with correlation > 60%: Accuracy ranged from 0.7369 to 0.9951.

5. Features with correlation > 75%: Accuracy ranged from 0.5375 to 0.7356.

The best accuracy results were achieved when 31 features with a correlation value greater than 60% were used. Additionally, the Proposed algorithm demonstrated the best accuracy in three experiments, while Logistic Regression (LG) performed best in two experiments.

For future research, the study aims to explore other machine learning algorithms, including deep learning or pre-trained models, to further enhance the robustness of the detection models. The goal is to develop a practical system or framework based on these models, making them accessible as computer programs, web applications and smartphone applications. Furthermore, the study intends to apply the same methodology to other datasets to assess its generalizability and effectiveness in different contexts. This ongoing work will contribute to the continual advancement of phishing attack detection techniques and their practical implementation.

REFERENCES

[1] Adam M. Bossler & Tamar Berenblum. (2019) “Introduction: new directions in cybercrime research” Journal of Crime and Justice, 495-499, Doi: 10.1080/0735648X.2019.1692426.

[2] Arun Kulkarni, Leonard L. Brown, (2019), Phishing websites detection using machine learning, International Journal of Advanced Computer Science and Applications, 10(7).

[3] Ashit Kumar Dutta. (2021) “Detecting phishing websites using machine learning technique” PLoS ONE, doi: 10.1371/ journal. pone.0258361.

[4] Aslam Khan, Rahul Sharma. (2018)” A Survey Paper on Detection of Phishing Website by URL Technique” International Journal of Computer Science and Mobile Applications, 6 (4), 33-37.

[5] Ishant Tyagi, Jatin Shad; Shubham Sharma; Siddharth Gaur; Gagandeep Kaur. (2018) “A novel Machine Learning Approach to Detect Phishing Websites" 2018 5th International Conference on Signal Processing and Integrated Networks (SPIN), 425-430.

[6] Justin Ma, Lawrence K. Saul, Stefan Savage, Geoffrey M. Voelker. (2011) “Learning to detect malicious URLs” ACM Transactions on Intelligent Systems and Technology, 2 (3), Article No.: 30, 1–24.

[7] Khana Faisal Md, Rana B. L. (2021) “Detection of Phishing Websites Using Deep Learning Techniques” Turkish Journal of Computer and Mathematics Education (TURCOMAT), 12(10), 3880-3892.

[8] Maddie Rosenthal. (2022) “Must-Know Phishing Statistics” [Online]: available at: https://www.tessian.com/blog/phishing-statistics-2020/

[9] Peng yang,Guangzhen ahaus, and pungent .(2019) “Phishing Website Detection Based on Multidimensional Features Driven by Deep Learning” IEEE Access, 7, 15196-15209.

[10] Ping Yi, Yuxiang Guan, Futai Zou, Yao Yao, Wei Wang, and Ting Zhu. (2018) “Web phishing detection using a deep learning framework” Wireless Communications and Mobile Computing, Article ID 4678746, doi:10.1155/2018/4678746.

[11] Rishikesh Mahajan, Irfan Siddavatam. (2018) ‘Phishing Website Detection using Machine Learning Algorithms” International Journal of Computer Applications (0975 – 8887), 181(23).

[12] Routhu Srinivasa Rao & Alwyn Roshan Pais. (2019) “Detection of phishing websites using an efficient feature-based machine learning framework” Neural Comput & Applic 31, 3851–3873.

[13] Rundong Yang , Kangfeng Zheng , Bin Wu , Chunhua Wu , Xiujuan Wang. (2021) “Phishing Website Detection Based on Deep Convolutional Neural Network and Random Forest Ensemble Learning” Sensors 2021.

[14] Vrbančič, G., Fister Jr, I., & Podgorelec, V. (2020) “Datasets for phishing websites detection” Data in Brief, 33, 106438.

[15] Yussra M. AL-Shareef, Hesham Abusaimeh. (2020) “How to Detect Phishing Website Using Three Model Ensemble Classification” Computer Science Department, Faculty of Information Technology, Middle East University.

[16] A.K. Patel, Jaypalsinh A. Gohil, D.T. Meva, (2010), Proposed Methodology for Securing Wireless LANs from Wormhole Attack, Int. J. of Advanced Networking and Applications 342 Volume: 01, Issue: 06, Pages:342-346.

[17] Sg, Hymlin & R, Thilothi & Banu.N, Rubiya & A, Sandhiya & Sumithra,. (2021). Detection Of Jamming Attack Using IEWMA In Clustered Wireless Sensor Network. International Journal of Advanced Networking and Applications. 12. 4691-4696.

[18] P.Mohana Priya, Abhijit Ranganathan, , (2022), Cyber Awareness Learning Imitation Environment (CALIE): A Card Game to provide Cyber Security Awareness for Various Group of Practitioners, Int. J. Advanced Networking and Applications Volume: 14 Issue: 02 Pages: 5318-5328.

[19] Hung Le, Quang Pham, Doyen Sahoo, Steven C.H. Hoi. (2018) “URLNet: Learning a URL Representation with Deep Learning for Malicious URL Detection” arXiv, Computer Science, Cryptography and Security, arXiv:1802.03162.

[20] Heba Harahsheh, Mohammad Shraideh, Saleh Sharaeh, (2021), PERFORMANCE OF Malware Detection Classifier Using Genetic Programming in Feature Selection, Informatica, 45(4).

[21] Heba Harahsheh, Mohammad Alshraideh, Saleh Al-Sharaeh & Rizik Al-Sayyed (2023) Improving Classification Performance for Malware Detection Using Genetic Programming Feature Selection Techniques, Journal of Applied Security Research, 18:3, 627-647.

[22] Methaq Kadhum, Enas Rawashdeh, Mohammad Alshraideh, (2021), An Efficient Bug Reports Assignment for IOT Application with Auto-Tuning Structure of ELM Using Dragonfly Optimizer, Journal of Hunan University Natural Sciences, 48(7).

Authors Biography

Moiad Alkharabsheh, Lt. Col. Moiad Alkharabsheh holds a Master's degree in Computer Science from the King Abdullah II Bin Al Hussein College of Information Technology at the University of Jordan, under the supervision of Professor Dr. Mohammad Alshraideh. He earned his bachelor's (BSc) degree in computer science from Mut'ah University and his master's degree (MSc) in computer science from Jordan University. Lt. Col. Alkharabsheh currently serves as a lieutenant colonel in the Jordanian Public Security Directorate (PSD).

Mohammad Aref Alshraideh currently holds the position of Professor specializing in Software Testing and Computational Intelligence at Lusail University's Information Technology College. Before joining Lusail University, he was affiliated with the Artificial Intelligence department at the University of Jordan in Jordan. Dr Alshraideh earned his PhD in Software Engineering with a focus on Testing from the University of Hull in the United Kingdom in 2007. He boasts an extensive 15-year career in the IT industry, encompassing roles such as software system design and development, as well as project management. Throughout his professional journey, Dr. Alshraideh has served in various administrative capacities, including Head Director Assistant for Computer Technology at the University Hospital, Human Resource Director, Registrar General at the University of Jordan, and Dean of the Graduate School. Moreover, he has held the distinguished position of Vice President for Administrative and Financial Affairs at the University of Jordan. Dr. Alshraideh has been the recipient of numerous academic project grants and has made significant contributions to his field, with over fifty published papers in areas related to software testing, artificial intelligence, data mining, and digital humanity. He actively engages in workshops, seminars, and conferences within the domains of software engineering and computational intelligence. His research interests encompass Software Testing, Artificial Intelligence, Data Mining, and Digital Humanity. he can be contacted at email: mshridah@ju.edu.jo or malshraideh@lu.edu.qa

Imad Khaled Salah currently holds the position of Professor of Computer Engineering at the Information Technology College of Lusail University. Before joining Lusail University, he was affiliated with the computer science department at the University of Jordan in Jordan. Dr. Imad earned his PhD from the National Technical University of Ukraine 'Kyiv Polytechnic Institute.' Throughout his professional journey, he has served in various administrative capacities, including Registrar General at the University of Jordan and Dean of Student Affairs. Moreover, he has held the distinguished position of Vice President for Administrative and Financial Affairs at the University of Jordan. Dr. Imad possesses a wealth of experience in the fields of Information Technology, Strategic Planning, Lecturing, Research, Higher Education, and Leadership. He is an experienced Professor with a demonstrated history of working in the research industry and is skilled in Computer Engineering. Dr. Imad is a strong educational professional with a Doctor of Philosophy (PhD) focused on Computer Engineering. he can be contacted at email: isalah@lu.edu.qa or isalah@ju.edu.jo