Enhanced machine learning algorithm for detection and classification of phishing attacks

Mutaz Abdel Wahed

***Abstract*—Machine learning employs artificial neural networks to acquire representations. Phishing is duplicitous behavior or threat where attackers aim to obtain credential information from websites. Phishing websites are fraudulent attempts by cybercriminals to impersonate reputable sites with mission of deceiving victims in divulging personnel information such as credentials, and personal data. Detecting and categorizing these malicious sites has been a topic of interest, with various methods focusing on URL-based techniques proving effective. Among these, machine learning and artificial intelligence approaches leveraging URL features have demonstrated superior results, contingent on the specific features utilized. This research proposes a novel model using a decision tree and a specific feature set to enhance the accuracy of phishing website detection, especially IoT devices. The research explores the impact of selecting a the most common attributes from the well-trained datasets to optimize performance and speed in classification and categorization phishing attacks on IoT devices. Experimental findings and comparative analysis present that the implemented algorithms achieve exceptional performance, with the proposed model achieving an impressive accuracy in identifying phishing URLs.**

Keywords— phishing attack, Phishing detection, Machine learning, decision tree

I. INTRODUCTION

Due to rise of the online technology, it became apparent that a new hi-tech revolution was underway, marked by advancements in cybercrimes. This shift has seen many business sectors and information technologies transition from traditional to online platforms. Exploiting the widespread use of online activities, numerous criminal activities have also shifted to the digital realm, known as cybercrime. One prevalent form of cybersecurity threats is phishing attacks. In 2020, worldwide losses due to real-time banking fraud reached $1,950 million, with $320 million traced back to phishing attacks. This type of fraud has emerged as one of the most impactful and prevalent forms of internet fraud. [1–2].

In phishing attacks, regular internet users are tricked into entering their personal information, often through a deceptive URL. Typically, the URL used in these attacks is disguised in web browsers by utilizing lengthy sequences of numbers and letters or by using similar characters to mimic legitimate URLs (example, http://www.hcbsbank.com instead of http://www.hsbcbank.com). When malicious URLs are received on hand held devices, like tablets or smart phones, the effectiveness of phishing attacks increases, enabling cybercriminals to conduct fraud more efficiently. Furthermore, in internet browsers, the URL address bar for inputting URL or IP addresses is often minimized or even concealed from users. These tactics are particularly effective on Internet of Things (IoT) devices, which encompass various gadgets like mobile phones, smart home devices, and more, commonly used for tasks such as messaging, entertainment, online shopping, and communication. Consequently, cyber attackers are increasingly targeting IoT devices and their users. Among different cyber threats, phishing attacks are anticipated to grow rapidly, appealing to cybercriminals due to the vulnerabilities and limited security measures present in IoT devices.

The study primarily concentrates on leveraging features in used training data to enhance the effectiveness of ML and AI algorithms in detecting and categorizing phishing attacks. However, many research efforts in this area tend to focus on evaluating the performance of different classifiers based on predefined features from third-party services and public training dataset sources. These studies commonly employ intricate data structures, data representations, and computationally intensive processes, rendering them impractical for implementation on Internet of Things (IoT) devices. Furthermore, a few investigations talk encountering suspicious net pages, suggesting that the users have potentially been subjected to phishing assaults. Given that IoT hardware have limited computational resources and power efficiency, using such class methods and algorithms on those structures is deemed impractical [5].

In such situations, classification algorithms designed for IoT gadgets must prioritize being lightweight and energy-green. It is beneficial to influence clear of complex records structures and opt for sincere assets, training datasets, and functions. To meet those standards, this research indicates employing a tree based totally algorithm, outlined in [6], for figuring out phishing IoT devices URLs in settings. This technique pursuits to optimize the detection charge and accuracy in classifying phishing assaults. The careful choice of capabilities is crucial in formulating a sensible phishing detection strategy. Moreover, the proposed technique is capable of figuring out attacks directly, which include zero-day threats, without reliance on external offerings.

The key contributions of this text include:

1. Choosing applicable features from the education dataset to correctly become aware of and categorize

phishing URLs, tailored specifically for IoT structures.

2. Serving as a foundational aid for researchers and practitioners seeking to deal with the venture of classifying phishing assaults in structures with constrained capabilities like IoT.

1.1 Phishing attacks: Often exploit URLs accessed via web browsers. These assaults contain embedding special phrases or characters in URLs. Some common strategies include:

A. Generating comparable phrases with minor errors.
B. Using unique characters or letters to redirect customers.
C. Employing excessively lengthy or overly short URLs.
D. Incorporating appealing keywords.
E. Adding malicious documents to hyperlinks that mechanically download to victims' IoT gadgets.

The most common algorithm used to detect and classify phishing domains and addresses relies on phishing lists, which maintain a data of already categorized URLS like "Phish Tank". While this technique is speedy and powerful, it has barriers. For example, it could now not trap URLs that don't exist or are not but at the blacklist. Traditional machine gaining knowledge of algorithms are often used to deal with those challenges in detecting phishing URLs. These strategies are specifically relevant for IoT devices, that have exceedingly limited processing strength and sources.

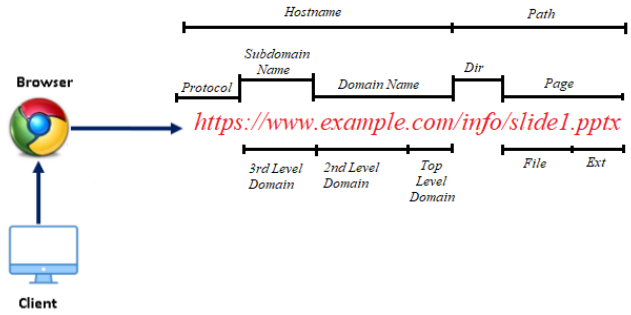

Figure 1. URL directing to a file

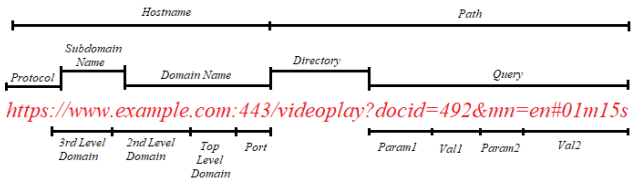

Figure 2. URL with query

1.2 Types of phishing Attacks

Phishing assaults are deceptive techniques utilized by cybercriminals to trick individuals into revealing touchy information or acting moves that compromise their safety. Here are some common forms of phishing attacks [8-10]:

1. **Spear Phishing:** In spear phishing, attackers specifically target an individual within an organization. They gather information about the person (such as their name, position, and contact details) and then craft personalized messages to steal login credentials.
2. **Vishing (Voice Phishing):** Vishing involves the use of phone calls to steal records. Attackers might also pose as trusted individuals or representatives to lie to victims.
3. **Email Phishing:** In email phishing, attackers send reputedly legitimate emails designed to trick recipients into presenting sensitive records.
4. **HTTPS Phishing:** Attackers ship sufferers emails containing links to faux websites. These web sites aim to misinform sufferers into entering non-public data.
5. **Pharming:** In pharming assaults, malicious code is mounted on a victim's laptop. This code redirects the sufferer to a fake internet site designed to collect login credentials.
6. **Pop-up Phishing:** Pop-up phishing exploits pop-up messages approximately pc security issues to trick users into clicking on them.

II. RELATED WORKS

The identification of phishing websites has been a focal point of good sized hobby within the studies community.
The performance of phishing detection fashions varies across special varieties of phishing assaults based on the strategies and datasets used. Various research have shown that device gaining knowledge of models, consisting of Random Forest (RF), Gradient Boosting (XGB), and Extreme Gradient Boosting (XGB), gain high accuracy fees ranging from ninety seven.44% to 98.27% whilst detecting phishing URLs [1] [2]. Additionally, the use of Natural Language Processing (NLP) and Deep Learning (DL) algorithms, like Long Short-Term Memory (LSTM) and Bidirectional GRU, has shown promising outcomes with mean accuracies exceeding ninety six.7% for detecting phishing attacks based totally on internet web page content material [3]. Furthermore, the assessment of different machine gaining knowledge of algorithms, which includes Naive Bayes, K-Nearest Neighbor, Random Forest, Decision Tree, Support Vector Machine, and Logistic Regression, has highlighted the effectiveness of Random Forest in reaching the highest accuracy of 98 % in identifying phishing URLs [4] [5].
In the table 1, Some of the carried out research might be summarized.

TABLE 1. RELATED WORKS

No.	*Authors*	*Threat*	*Methodology*	*Findings*
[14]	Ubing et al., (2019)	legitimate or phishing	analyze URLs	high accuracy of 92.52%
[15]	Mande and Thosar, (2018)	phishing attack	algorithm of (ELM) extreme learning machine	classification accuracy respectively
[16]	Nagaraj et. al., (2018)	IDS detection that nullifies phishing threats	Classification phishing URLs using ensemble twofold.	Random Forest produced a high accuracy of 93.4%

III. ETHODOLOGY

1.3 Techniques for categorization classifying phishing URLs:

1. **Variation approach:** Phishing attackers regularly make a mess of URLs with specific variations. This approach objectives to divert the eye of ordinary net customers, appreciably growing the chances of a hit phishing assault. For instance:
   - Introducing multiple forward slashes in the URL path to mimic legitimate directories.
   - Adding extra dots and alphanumeric characters to the domain name, creating a sense of validity.
2. **Character Substitution:** Malicious URLs frequently replace alphanumeric characters with other symbols, such as Unicode characters or hexadecimal representations. The predictability of English text allows for effective detection by analyzing changes in entropy when different symbols are introduced.
3. **Lightweight URL Representation and Improved Algorithms:** The article proposes a lightweight representation for URLs and an enhanced algorithm to detect and classify phishing URLs specifically in Internet of Things (IoT) systems. These systems often have limited processing power and resources [8].
4. **Delivery via Regular Applications:** During phishing attacks, cybercriminals commonly deliver malicious URLs through everyday applications like email, Telegram, Twitter, and Facebook. If an inexperienced internet user accesses such a phishing URL, they unwittingly become vulnerable to the cybercriminal's malicious activity.

1.4 Recommended features for detecting phishing URLs, considering the specific challenges posed by IoT devices:

Selecting features for detecting phishing attacks [12-17], the authors explored some methods for detecting phishing URLs. These methods use many features from used URLs. This work proposed the lasts' set of features that implemented by phishing attacks and seen in the URLs, for example, attackers try to confuse users of IoT devices by making URLs unreadable and unfamiliar. The composed phishing URLs become longer and use different symbols and string or digits than the legal URL. Moreover, this research recommends using all indicators that detect the length of any segment of the URL (string, special characters, digits and signs related to HTTP/S. The following indicators kit is recommended for this study [1,4-8,10-12]:

1. **having_ip_address:** Indicates whether the URL has an IP address instead of a domain name.
2. **url_length:** Length of the URL.
3. **shortening_service:** a URL shortening service.
4. **having_at_symbol:** Presence of "@" symbol in the URL.
5. **Double_slash_redirecting:** Whether the URL makes use of double slash redirecting.
6. **Prefix_suffix:** Presence of a prefix or suffix in the URL.
7. **Having_sub_domain:** Presence of a subdomain in the URL.
8. **Sslfinal_state:** SSL very last kingdom of the URL.
9. **Domain_registration_length:** Length of time the domain has been registered for.
10. **Favicon:** Presence of a fave icon on the web site.
11. **Port:** Port wide variety in the URL.
12. **Https_token:** Presence of an HTTPS token inside the URL.
13. **Request_url:** Whether there's a request URL in the web site.
14. **Url_of_anchor:** URL of the anchor textual content.
15. **Links_in_tags:** Number of hyperlinks gift in the tags.
16. **Sfh:** Server form handler (SFH) attribute within the shape tag.
17. **Submitting_to_email:** Whether the web site is submitting records to an email.
18. **Abnormal_url:** Indicates whether or not the URL is abnormal.
19. **Redirect:** Presence of redirection within the URL.
20. **On_mouseover:** On mouseover event within the web site.
21. **Rightclick:** Right-click on is enabled on the website.
22. **Popupwindow:** Presence of popup windows.
23. **Iframe:** Presence of inline frame in the website.
24. **Age_of_domain:** Age of the area.
25. **Dnsrecord:** DNS record of the area.
26. **web_traffic:** Web traffic statistics.
27. **page_rank:** Page rank of the webpage.
28. **google_index:** Webpage is indexed by Google.
29. **links_pointing_to_page:** Number of external links pointing to the webpage.
30. **statistical_report:** Statistical report of the webpage.

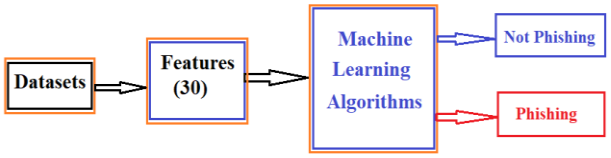

Figure 3: Methodological scheme for classifying phishing URLs

IV. EXPERIMENTAL TESTING

This paper employed eight distinct categorization and classification techniques (Neural Networks, Deep Learning, Decision Tree, etc.) as the core machine learning framework for the proposed system, followed by a comparative evaluation. One significant challenge in testing the proposed system is the utilization of a widely accepted datasets from various sources, including:

Kaggle: Is a famous platform for finding and sharing datasets. You can look for datasets related to phishing URLs or cybersecurity. Visit Kaggle's internet site and look for applicable key phrases to locate datasets contributed with the aid of customers.

UCI Machine Learning Repository: The UCI Machine Learning Repository hosts an extensive range of datasets for machine learning research. They might have datasets associated with phishing URLs or cybersecurity. Visit their website and explore the to be had datasets.

PhishTank: Is a community-driven clearing residence for phishing facts. They provide a dataset of recognized phishing URLs that you may use for studies functions. Visit their internet site to get entry to the dataset and analyze extra about their facts.

Microsoft Malware Classification Challenge: Microsoft has released a dataset containing URLs categorized as phishing or valid, along with various capabilities extracted from the ones URLs. This dataset is part of the Microsoft Malware Classification Challenge. You can look for it on-line or discover it on platforms like Kaggle.

GitHub: Many researchers and companies proportion datasets related to phishing detection on GitHub. You can discover repositories and look for "phishing" to discover applicable datasets.

In the context of the dataset being saved in a CSV file, this setup means that the data is organized in a spreadsheet-like format, where each row contains information about a single URL, and each column represents a specific attribute or characteristic of those URLs.

CSV File Format: CSV stands for Comma-Separated Values. It's a commonplace record layout used to save tabular information, wherein every row represents a document or commentary, and columns represent exceptional attributes or capabilities of that record. In this paper, the dataset is stored in a CSV document, because of this it may be effortlessly examine and manipulated using programming languages like Python.

Each Row Represents a URL: In the dataset, every row corresponds to a single URL. This URL could be an internet site address, an e-mail hyperlink, or any other type of Uniform Resource Locator (URL). Each URL serves as a unique statistics point in the dataset.

A test was conducted on a dataset comprising 75,655 URLs, consisting of 36,500 valid URLs and 38,20 phishing URLs. The tests were handled by HP device deployed with a 3.6 GHz intel Core i7 1165G7 quad CPU, 16GB DDR5 RAM Memory and Nvidia GeForce M450 2GB Graphic Card. Throughout the tests, 10-fold cross-validation and default parameter values were applied across all algorithms. Each test set was executed using the aforementioned eight machine learning algorithms. The confusion matrix for the evaluated learning algorithms is depicted in Table II.

TABLE 1I. CONFUSION MATRIX

Confusion Matrix		***Predicted***	
		P	***N***
Decision Tree	P	36728	447
	N	1252	31052
Random Forest	P	36806	35280
	N	1120	9512
KNN (K=3)	P	36214	961
	N	2082	34318
Logistic Regression	P	35652	1394
	N	3804	32721
Neural Network	P	36628	32052
	N	1352	547
Naïve Bayes	P	27663	9512
	N	1247	35153
Deep Learning	P	36500	547
	N	1120	30012
Adaboost	P	35813	1362
	N	3609	32791

True Positive (TP): The model efficiently predicts the effective class.

True Negative (TN): The version correctly predicts the terrible elegance.

False Positive (FP): Also called Type I errors, the version incorrectly predicts the fine class while it is in reality terrible.

False Negative (FN): Also called Type II mistakes, the model incorrectly predicts the terrible class whilst it's simply fine.

From the confusion matrix, numerous metrics may be derived to evaluate the overall performance of the model, consisting of accuracy, precision, consider, and F1 score. These metrics assist in knowledge how nicely the model is performing and whether or not it's biased toward any particular magnificence. Utilizing the statistics from the confusion matrix, four wonderful metrics— Precision, Sensitivity, F-measure, and Accuracy—are employed to assess the effectiveness and efficiency of the algorithms. These metrics, delineated in equations (1-4), are essential for benchmarking hooked up gadget learning methodologies:

1. **Precision (Positive Predictive Value):**

$$\textbf{Precision} = \frac{TP}{TP+TF} \quad (1)$$

Precision measures the accuracy of fantastic predictions. It is the ratio of efficiently predicted effective observations to the full expected positives. High precision indicates a low false wonderful fee.

2. **Sensitivity (Recall or True Positive Rate):**

$$\textbf{Sensitivity} = \frac{TP}{TP+FN} \quad (2)$$

Sensitivity measures the proportion of actual positive instances that were efficaciously recognized through the classifier. This gives the average of efficiently expected superb monitoring to the whole real positives. Very sensitivity detects the low false negative index.

3. **F-measure (F1-score):**

$$\textbf{F} - \textbf{Measure} = 2x\frac{\text{Precision x Sensitivity}}{\text{Precision x Sensitivity}} \quad (3)$$

F-measure is the harmonic mean of precision and sensitivity. It provides a balance between precision and sensitivity and is particularly useful when the class distribution is imbalanced.

4. **Accuracy**

$$\textbf{Accuracy} = \frac{TP+TN}{TP+TN+FN+FP} \quad (4)$$

Accuracy measures the overall correctness of the classifier. It is the ratio of correctly predicted observations (both positive and negative) to the total number of observations.

In these equations, TP denotes true positive, TN represents true negative, FP indicates false positive, and FN signifies false negative rates of classification algorithms. Following these formulas, the computational outcomes of the implemented machine learning algorithms are detailed in Table 2 in a comparative manner.

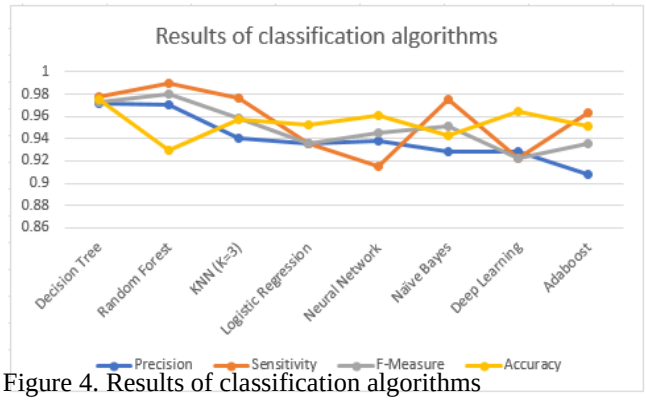

Figure 4. Results of classification algorithms

From the figure, it's evident that the decision tree exhibits the best classification performance, achieving an accuracy of 97.50%. Furthermore, the significance of features can be observed from the results.

This level of accuracy is considered satisfactory and commendable for phishing detection. Achieving 100% accuracy is unattainable due to the constant evolution of both defensive techniques employed by system security managers and the strategies employed by attackers to circumvent existing or emerging anti-phishing systems. Moreover, the proposed technique is based on reading the URL of the phishing webpage. Upon investigating undetected phishing webpages, it was noted that some of these pages feature short domain names and subdomains without any paths. Consequently, if a URL comprises solely a single domain name (e.g., "www.testname.org"), it may not be detected by the proposed solution, which is based on natural language processing (NLP).

In a typical phishing attack, perpetrators design webpages to resemble legitimate ones, often employing lengthy URLs and special terms to deceive users. Shorter URLs, however, may be more easily identified by users with basic knowledge of phishing attacks.

For an overall performance comparison, Table 3 is compiled. Across nearly the most ML algorithms, the CL-DR algorithm consistently delivers more enhancements for URL classification, with an average improvement of 10.9%. Additionally, incorporating hybrid functions further enhances system performance, yielding an improvement of 2.25% with NLP functions and 13.1% with feature words.

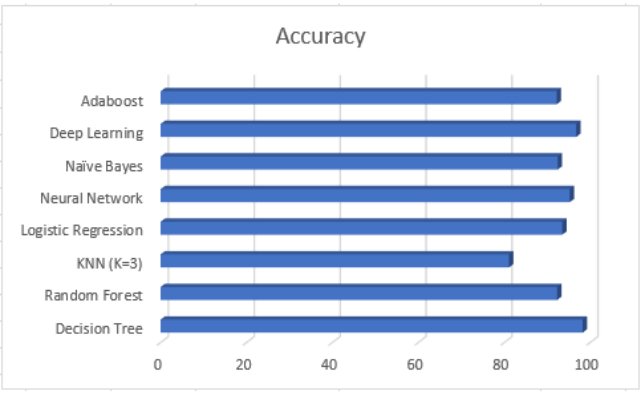

Figure 5. comparison of different algorithms

V. Conclusion

This research introduces an intelligent system for detecting phishing threats in IoT hardware. This approach focuses on simplicity, making it suitable for devices with limited capabilities. By employing decision trees in various training algorithms, the proposed method demonstrates effectiveness in IoT systems. Furthermore, it shows potential for enhanced performance in spite of extra concise version structures as compared to existing strategies. To enhance detection accuracy, creating an effective feature list remains a crucial task.

References

[1] Ghazi-Tehrani, Adam Kavon, and Henry N. Pontell. "Phishing evolves: Analyzing the enduring cybercrime." In The New Technology of Financial Crime, pp. 35-61. Routledge, 2022.

[2] Nadeem, Muhammad, Syeda Wajiha Zahra, Muhammad Nouman Abbasi, Ali Arshad, Saman Riaz, and Waqas Ahmed. "Phishing attack, its detections and prevention techniques." Int. J. Wirel. Secur. Netw 1 (2023): 13-25.

[3] Carroll, Fiona, John Ayooluwa Adejobi, and Reza Montasari. "How good are we at detecting a phishing attack? Investigating the evolving phishing attack email and why it continues to successfully deceive society." SN Computer science 3, no. 2 (2022): 170.

[4] Chanti, S., and T. Chithralekha. "A literature review on classification of phishing attacks." International Journal of Advanced Technology and Engineering Exploration 9, no. 89 (2022): 446-476..

[5] Desolda, Giuseppe, Lauren S. Ferro, Andrea Marrella, Tiziana Catarci, and Maria Francesca Costabile. "Human factors in phishing attacks: a systematic literature review." ACM Computing Surveys (CSUR) 54, no. 8 (2021): 1-35.

[6] Goenka, Richa, Meenu Chawla, and Namita Tiwari. "A comprehensive survey of phishing: Mediums, intended targets, attack and defence techniques and a novel taxonomy." International Journal of Information Security 23, no. 2 (2024): 819-848.

[7] Wang, Mengli, and Lipeng Song. "Efficient defense strategy against spam and phishing email: An evolutionary game model." Journal of Information Security and Applications 61 (2021): 102947.

[8] Bhardwaj, Akashdeep, Fadi Al-Turjman, Varun Sapra, Manoj Kumar, and Thompson Stephan. "Privacy-aware detection framework to mitigate new-age phishing attacks." Computers & Electrical Engineering 96 (2021): 107546.

[9] Aslan, Ömer, Semih Serkant Aktuğ, Merve Ozkan-Okay, Abdullah Asim Yilmaz, and Erdal Akin. "A comprehensive review of cyber security vulnerabilities, threats, attacks, and solutions." Electronics 12, no. 6 (2023): 1333.

[10] Gomes, Vanessa, Joaquim Reis, and Bráulio Alturas. "Social engineering and the dangers of phishing." In 2020 15th Iberian Conference on Information Systems and Technologies (CISTI), pp. 1-7. IEEE, 2020.

[11] Andriu, Adrian-Viorel. "Adaptive phishing detection: Harnessing the power of Artificial Intelligence for enhanced email security." Romanian Cyber Secur. J 5, no. 1 (2023): 3-9.

[12] M. A. Wahed, "Real-Time Intrusion Detection and Traffic Analysis Using AI Techniques in IoT Infrastructure," 2024 1st International Conference on Emerging Technologies for Dependable Internet of Things (ICETI), Sana'a, Yemen, 2024, pp. 1-6, doi: 10.1109/ICETI63946.2024.10777213.

[13] Jimmy, F. N. U. "Cyber security Vulnerabilities and Remediation Through Cloud Security Tools." Journal of Artificial Intelligence General science (JAIGS) ISSN: 3006-4023 2, no. 1 (2024): 129-171.

[14] Rains, Tim. Cybersecurity Threats, Malware Trends, and Strategies: Learn to mitigate exploits, malware, phishing, and other social engineering attacks. Packt Publishing Ltd, 2020.

[15] Lim, Wei Heng, W. Foong Liew, Chun Yew Lum, and Seah Fang Lee. "Phishing security: Attack, detection, and prevention mechanisms." In Proceedings of the International Conference on Digital Transformation and Applications (ICDXA). 2020.

[16] Basit, Abdul, Maham Zafar, Xuan Liu, Abdul Rehman Javed, Zunera Jalil, and Kashif Kifayat. "A comprehensive survey of AI-enabled phishing attacks detection techniques." *Telecommunication Systems* 76 (2021): 139-154.

Mutaz Abdel Wahed - Computer Networks and Cybersecurity Department, Jadara University, Irbid, Jordan (email: mutaz@jadara.edu.jo)