1. Introduction

A Quick Response code (QR code) is a two-dimensional matrix composed of black and white pixels, offering a compact and easily scannable method for storing information. QR codes have surged in popularity due to their higher data capacity and improved readability compared to traditional barcodes. As smartphone usage continues to soar, businesses are increasingly adopting QR codes as a convenient means to direct users to their websites and products. QR codes are engineered to be readable from any orientation and can still be scanned even if partially damaged or obscured, making them versatile for consumer applications [1]. Typically, QR codes are "scanned" by capturing an image of the code with a smartphone camera and then interpreted by a QR code reader app installed on the device. The reader decodes the message and performs an action based on its content, such as launching a marketplace app for downloading mobile applications if the encoded data contains a link. Often, QR codes represent hyperlinks, prompting the device's web browser to visit the specified website. Despite their convenience, the ease of creating and distributing QR codes has also attracted malicious actors looking to engage in phishing attacks, known as QRishing. This poses a security threat to the widespread adoption of QR codes, as attackers may overlay legitimate QR codes with malicious ones or create entirely fake QR code advertisements to deceive users [2].

The QR Code Phishing Attacks

QR Code Phishing Attacks can be divided into three categories:

i. QR Code Replacement: Legitimate QR codes can be manipulated into malicious versions by replacing the entire code or by defacing the existing code, adding black modules to the white ones. Manipulation can occur through various elements of the QR code, such as its mask, character encoding mode, character count indicator, mixing modes, data part, and error correction section. Simple tools like a pen can be used to change white modules to black, resulting in phishing attacks. Malicious QR codes typically redirect users to phishing or exploit sites rather than directly downloading malware [3]. An example is illustrated in Fig. 1, where both white and black modules are altered to change a link originally pointing to "ebay.com" to "gbay.com".

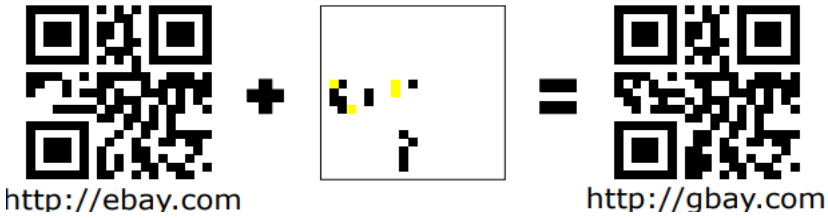
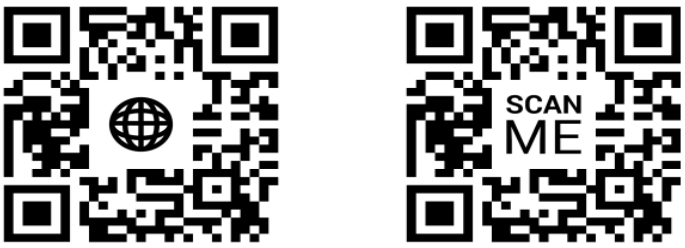


Fig. 1. Example of a manipulated QR code

ii. QR Code Manipulation: The risk posed by malicious QR codes is heightened by applications that automatically launch the embedded link without allowing users to verify it beforehand. Additionally, the use of URL shorteners complicates visual verification for users. Furthermore, mobile operating systems often hide the full URL of the loaded website for ease of use on small screens, making it challenging for users to assess the nature of the link. This limitation, coupled with the inability of most QR code scanners to detect modifications, exacerbates the risks associated with QR code usage [4]. Although many mobile applications display the decoded URI, they may not thoroughly check URIs and web content for malicious intent.

iii. Barcode-in-barcode attack: Another type of QR code manipulation reported is the barcode-in-barcode attack. QR codes have error correction features that allow QR code decoders to disregard segments with unknown coding. This area can be exploited to display an icon or text, as depicted in Fig. 2a and 2b.



(a) A QR code with an icon. (b) A QR code with text.

Fig. 2 Example of QR codes.

This feature can be exploited to embed another barcode within the QR code, while still enabling the original QR code to be decoded [5].

Proposed Countermeasures

- Increase the level of strict data correction.

- Prioritize barcodes based on their importance.
- Display the border of the detected code to indicate if an inner barcode is being scanned.
- Notify the user of the detection of multiple barcodes.
- Enhance barcode reader functionality by showing the decoded URL before launching and implementing URL check methods.

2. Literature Review

S. Ismail, et.al (2021) suggested a novel framework to authenticate QR Code and detect phishing [6]. This framework was effective for detecting the phishing and malicious URLs while validating the QR Code. It was employed for detecting abnormal and malicious links, embedded in the QR codes and utilized as a vector attack. A testing was done prior to suggest this framework. To ensure that the code was scanned perfectly, the suggested framework aimed to detect the malicious and phishing page. Moreover, it was proved useful for prevent users from harmful QR Codes.

A. S. Rafsanjani, et.al (2023) presented a secure and privacy-friendly QR code scanner (QsecR) relied on a model of detecting malicious URL [7]. This model was designed on the basis of classifying predefined static features. The URLhaus and PhishTank were considered to gather a dataset of 4000 real-time random URLs. Diverse QR code scanners were used to compute its security and privacy. The experiments exhibited the supremacy of presented model over others and its accuracy was counted 93.50% and precision was 93.80% to detect phishing URL in QR.

B. Herlina, et.al (2023) projected a model to detect phishing URLs in QR codes based on Machine Learning (ML) methods [8]. Such URLs were similar to real websites, which the cybercriminals had created for attaining user information. This work focused on gathering a dataset from several sources, having features of address bar, domain and HTML. Six ML methods: DT, RF, SVM, MLP, AENN and XGBoost were adopted. The simulation indicated that the last method was performed well and offered an accuracy of 92% to detect phishing URLs in QR codes.

M. Sahay, et.al (2024) aimed at deploying machine learning (ML) techniques for detecting SQL injections in applications [9]. The malicious actions were recognized in a cloud-based SaaS scenario. An analysis was conducted to verify whether the QR codes were practical and secure in the anti-phishing recommendations for the suggested method (secure QR code). These

recommendations were deployed after exploring the extensive cases and attack vectors. The social engineering called phishing was found a frequent scam in QR codes as an attack vector.

G. A. Amoah, et.al (2022) aimed to discover the privacy issues related to QR codes [10]. The introduced method employed a count vectorizer to extract words of every URL, and considered the URLs comprised in QR code for extracting features. The genuine URLs were differentiated from the phishing ones after tokenizing the traits and words. The Naive Bayesian machine learning (NB-ML) methods were employed in a recursive loop with logistic regression (LR). This resulted in generated an effective mechanism to mitigate the quishing (QR-phishing) behavior.

M. Thakare, et.al (2023) suggested to deploy machine learning (ML) and convolutional neural networks (CNNs) into the QR code scanning procedure for detecting and blocking malicious codes. The secure QR codes were assisting in making the online transactions more secure and preventing web phishing attacks. The developed method was adopted to detect malware URLs with Support Vector Machine (SVM) method. This method was designed to make the communication more scalable, flexible, and secure.

2.1 Research Gaps

Following are the various research gaps: -

1. The schemes which are already proposed for the phishing detection are unable to establish relation between attribute set and target set due to which optimal level of accuracy is not achieved.

2. The models for the phishing detection are based on the machine learning techniques. The machine learning models are supervised model. In the previous research no, one proposed unsupervised model for the phishing detection.

3. The models which are already proposed are unable to work on the QR codes for the phishing detection. The model needs to propose which can detect phishing from QR codes.

3. Problem Formulation

Phishing leverages email deception and fraudulent websites to extract sensitive information from unsuspecting individuals. Despite various efforts to address this issue, there remains a lack of comprehensive solutions to combat phishing effectively. Therefore, leveraging machine learning techniques is crucial in mitigating cybercrimes, particularly those involving phishing attacks. The proposed study utilizes a phishing URL-based dataset sourced from a renowned repository, comprising attributes of both phishing and legitimate URLs gathered from over 11,000 websites. Following preprocessing, numerous machine learning algorithms are employed and tailored to thwart phishing URLs and safeguard users. These algorithms include decision tree (DT), linear regression (LR), random forest (RF), naive Bayes (NB), gradient boosting classifier (GBM), K-neighbors classifier (KNN), support vector classifier (SVC), and a novel hybrid LSD model. The hybrid LSD model, which combines logistic regression, support vector machine, and decision tree (LR+SVC+DT) with both soft and hard voting mechanisms, is particularly promising in defending against phishing attacks with remarkable accuracy and efficiency. It is analyzed that hybrid method give good performance of phishing URL's about with the change in the trend it donot perform of phishing QR codes. The model needs to propose which can perform well on phishing QR codes.

4. Objectives

Study and Analysis of Phishing QR Detection Techniques

Phishing QR detection techniques can be broadly categorized into three main approaches: heuristic-based methods, feature-based methods, and machine learning-based methods. Heuristic-based methods rely on predefined rules and patterns to identify malicious QR codes. These methods are straightforward and fast but often lack the flexibility to detect new and evolving phishing techniques. Feature-based methods, on the other hand, analyze specific characteristics of QR codes, such as the embedded URL's structure, domain reputation, and the presence of obfuscation techniques. While these methods provide better detection rates than heuristic-based methods, they still fall short in identifying sophisticated phishing attacks that continually adapt to bypass detection mechanisms. Machine learning-based methods represent the most advanced and promising approach to phishing QR detection. These methods leverage various algorithms to learn from vast datasets, identifying subtle patterns and anomalies that might indicate phishing attempts. Supervised learning techniques, such as decision trees, support vector machines, and neural networks, have shown significant success in this domain. Additionally, unsupervised learning methods, like clustering and anomaly detection, can

identify new phishing techniques by recognizing deviations from normal QR code behavior. Despite their effectiveness, machine learning-based methods require substantial computational resources and well-curated training datasets to achieve high detection accuracy.

Implementation of Machine Learning Methods for Phishing QR Detection

To enhance phishing QR detection, various machine learning methods can be implemented and evaluated. Supervised learning algorithms such as Random Forest, Gradient Boosting Machines, and Convolutional Neural Networks (CNNs) can be trained on labeled datasets containing both benign and phishing QR codes. These algorithms can learn complex patterns and relationships between the QR code features and their corresponding labels, enabling accurate classification of new, unseen QR codes. In addition to supervised learning, semi-supervised and unsupervised learning methods can also be employed. Semi-supervised learning can leverage a small amount of labeled data combined with a larger pool of unlabeled data, effectively reducing the dependency on extensive labeled datasets. Techniques such as self-training and co-training can iteratively improve the model's performance. Unsupervised learning methods, like K-means clustering and autoencoders, can detect anomalies in QR codes that may indicate phishing attempts. These methods are particularly useful in identifying zero-day phishing attacks that have not been seen before.

Designing a Novel Approach for Phishing QR Detection

Building on the strengths and limitations of existing methods, a novel approach for phishing QR detection can be designed. This approach could integrate multiple machine learning techniques to leverage their complementary strengths. For instance, a hybrid model combining supervised learning for initial classification and unsupervised learning for anomaly detection can provide robust and adaptive phishing QR detection. The proposed approach could involve several stages. Initially, a CNN could extract features from the QR code images, capturing both visual and contextual information. These features could then be fed into a Random Forest classifier for preliminary classification. Concurrently, an autoencoder could analyze the QR code data to detect any anomalies, flagging potential phishing codes that deviate from the learned patterns. By combining the outputs of these models, the hybrid approach can achieve higher detection accuracy and adaptability.

Implementation and Comparison with Existing Methods

The proposed approach will be implemented and compared with existing methods using standard evaluation metrics such as accuracy, precision, and recall. Accuracy measures the overall correctness of the model, while precision and recall provide insights into the model's ability to identify true positives and minimize false positives. A comprehensive dataset containing a diverse range of QR codes, including benign and phishing examples, will be used for training and testing. Initial experiments will focus on optimizing the hyperparameters of each component model to ensure the best performance. Cross-validation techniques will be employed to validate the model's robustness and prevent overfitting. The hybrid model's performance will be benchmarked against existing heuristic-based, feature-based, and standalone machine learning models. Expected outcomes include improved detection rates, lower false positive rates, and enhanced capability to identify new phishing techniques.

Following are the various objectives: -

1. To study and analyse various phishing QR detection techniques

2. To implement various machine learning methods for the phishing QR detection techniques

3. To design novel approach for the phishing QR detection techniques

4. Implement proposed approach and compare with existing in terms of accuracy, precision and recall

5. Research Methodology

The QR code phishing can be detected in diverse stages such as to pre-process the data, extract the attributes and classification. The research methodology is defined as: -

1. **Data set input and Pre-processing**: - The data will be taken as input and it will have processed to remove and missing values from the dataset.

2. **Feature Extraction**: - This stage is executed to establish the association of each attribute with the target set. The FP (false positive) is a scenario in which a sample seems normal but determined as intrusion. The situation in which a sample is actually an intrusion but determined as normal is known as FN (false negative). In case, the earlier case unable to detect the intrusion that is considered as poor FN.

3. **Classification**: - This stage focuses on splitting the entire data into training and testing. The voting classification technique is implemented to classify the network traffic. Multiple classifiers are integrated in this technique to classify the QR codes. The predicted set is allocated to the test set in this stage.

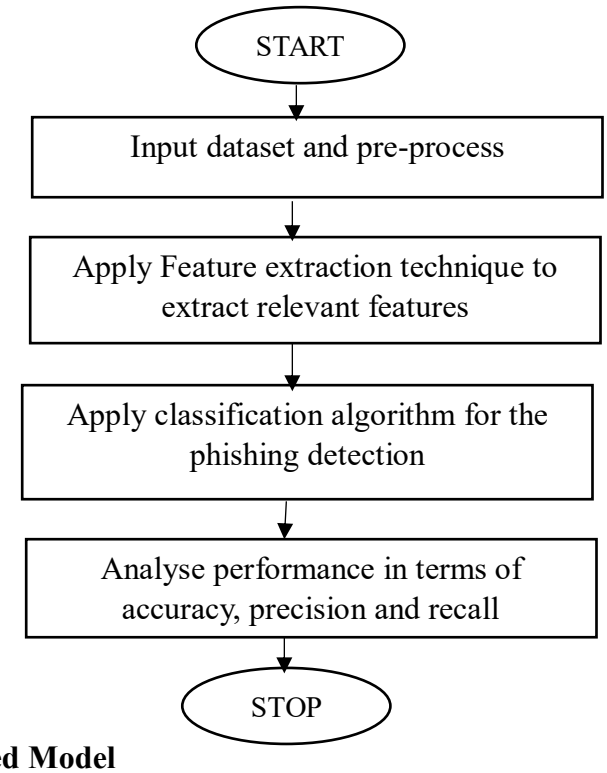


Fig 3: Proposed Model

References

[1] H. A. M. Wahsheh and F. L. Luccio, “Security and privacy of QR code applications: A comprehensive study, general guidelines and solutions,” Information, vol. 11, no. 4, p. 217, Apr. 2020, doi : 10.1007/978-3-319-65127-9.

[2] V. Mavroeidis and M. Nicho, “Quick response code secure: A cryptographically secure anti-phishing tool for QR code attacks,” in Computer Network Security (Lecture Notes in Computer Science), vol. 10446. Cham, Switzerland: Springer, 2017 doi: 10.1007/978-3-319-65127-9.

[3] A. Y. Alnajjar, M. Anbar, S. Manickam, O. Elejla, and H. El-Taj, “QRphish: An automated QR code phishing detection approach,” J. Eng. Appl. Sci., vol. 11, no. 3, pp. 553–560, 2016

[4] X. Xiao, D. Zhang, G. Hu, Y. Jiang, and S. Xia, “CNN–MHSA: A convolutional neural network and multi-head self-attention combined approach for detecting phishing websites,” Neural Netw., vol. 125, pp. 303–312, May 2020

[5] Y. Mourtaji, M. Bouhorma, D. Alghazzawi, G. Aldabbagh, and A. Alghamdi, “Hybrid rule-based solution for phishing URL detection using convolutional neural network,” Wireless Commun. Mobile Comput., vol. 2021, Sep. 2021, Art. no. 8241104.

[6] S. Ismail, M. H. Alkawaz and A. E. Kumar, "Quick Response Code Validation and Phishing Detection Tool," 2021 IEEE 11th IEEE Symposium on Computer Applications & Industrial Electronics (ISCAIE), Penang, Malaysia, 2021, pp. 261-266

[7] A. S. Rafsanjani, N. B. Kamaruddin, H. M. Rusli and M. Dabbagh, "QsecR: Secure QR Code Scanner According to a Novel Malicious URL Detection Framework," in IEEE Access, vol. 11, pp. 92523-92539, 2023

[8] B. Herlina and H. Soeparno, “Machine Learning Model to Improve Classification Performance in The Process of Detecting Phishing URLs in QR Codes”, Journal of Theoretical and Applied Information Technology, vol. 10, no. 18, pp. 13-20, 2023

[9] M. Sahay, S. Vanjale and M. Mane, “Software as Service Attack Detection and Prevention for Deceitful QR code”, International Journal of Intelligent Systems and Applications in Engineering, vol. 12, no. 4, pp. 454-462, 2024

[10] G. A. Amoah and H.-A. J.B., “QR Code Security: Mitigating the Issue of Quishing (QR Code Phishing)”, International Journal of Computer Applications, vol. 184, no. 33, 2022

[11] M. Thakare, S. Patil, H. Pawar, A. Sawant and K. Vatekar, “Qrshield: Qr Code-Based Attack Detection and Prevention for Software-As-A-Service (Saas) Applications”, International Research Journal of Modernization in Engineering Technology and Science, vol. 5, no. 4, pp. 12-20, 2023s