www.ijprems.com
editor@ijprems.com

INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)
(Int Peer Reviewed Journal)
Vol. 04, Issue 11, November 2024, pp : 321-325

e-ISSN : 2583-1062
Impact Factor : 7.001

# ENHANCING SECURITY IN QR CODE (SHORT URL) USING NLP & CNN BASED FRAMEWORK

**Vishaka Kotian[1], Dr. Rakhi Gupta[2], Miss Nashrah Gowalkar[3]**

[1]Masters of Science in I.T. K.C. College, HSNC University Mumbai-4000020, India
vishakakotian162000@gmail.com

[2]Head of I.T. Department K.C. College, HSNC University Mumbai-4000020, India
rakhi.gupta@kccollege.edu.in

[3]Asst. Professor, I.T. Department K.C. College, HSNC University Mumbai-4000020, India
nasrah.gowalkar@kccollege.edu.in

DOI: https://www.doi.org/10.58257/IJPREMS36542

## ABSTRACT

This research paper focuses on the identification of malicious QR codes, a growing concern in cybersecurity as their usage expands across various sectors. Malicious QR codes can lead to phishing attacks, data breaches, and malware installations, posing significant risks to users and organizations alike. The study presents a novel framework for detecting and analyzing malicious QR codes, integrating machine learning algorithms with traditional security measures. Through a comprehensive dataset of QR codes both benign and malicious, the research evaluates various features that distinguish harmful codes from legitimate ones. Results indicate that certain patterns, such as URL characteristics and code structure, can effectively predict malicious intent with high accuracy. Additionally, user awareness and education are assessed as vital components in mitigating risks. The paper concludes with practical recommendations for developing effective scanning tools and implementing security protocols, thereby enhancing protection against the threats posed by malicious QR codes in an increasingly digital environment.

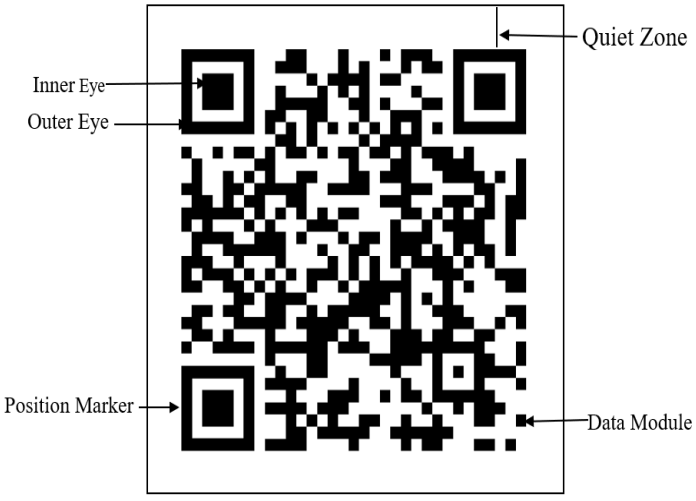
**Keywords-** Cyber Security, Artificial Intelligence, QR code, NLP, CNN, Short URL, Machine learning, Detection System.

## 1. INTRODUCTION

QR (Quick Response) codes have gained widespread popularity as a convenient means of sharing information, directing users to websites, and facilitating transactions. However, their increasing use has also made them a target for malicious actors seeking to exploit unsuspecting users. Malicious QR codes can redirect users to phishing sites, distribute malware, or engage in fraudulent activities, posing significant security risks.

The rise of mobile technology has facilitated the adoption of QR codes, where contactless interactions became essential. While QR codes provide a seamless way to access information, the lack of inherent security features makes it difficult for users to identify potentially harmful links.

Detection tools for malicious QR codes are crucial for mitigating these risks. Such tools employ various techniques, including machine learning, natural language processing (NLP), and image recognition, to analyze both the visual structure of the QR code and the content it directs to. By integrating multiple approaches, these detection systems can effectively discern safe codes from those that pose a threat.



www.ijprems.com
editor@ijprems.com

INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)
(Int Peer Reviewed Journal)
Vol. 04, Issue 11, November 2024, pp : 321-325

e-ISSN : 2583-1062
Impact Factor : 7.001

**Fig. 1.** Elements of QR Code.

A. Basic Structure of a QR Code

The most important parts of QR code are:

1) Data module: This is the standard unit of the QR code. It's typically a black square set against a white background. Though the colors and contrast can be different, black-on-white is the most optimal when creating a custom QR code. The arrangement of these black squares, or data modules, is what makes up the majority of a QR code.
2) Position marker: There are three position markers on every QR code. Consisting of an inner and outer eye, they allow scanners and cameras to quickly and accurately locate the data modules and the scanning direction.
3) Quiet zone: This is the blank area on all sides of the data module matrix that contains all the data modules and position markers. It allows scanners and readers to optically place where the QR code begins and ends.
4) Outer Eye: The largest square located in the corner, which helps the QR code reader detect the code and determine its orientation.
5) Inner Eye: The smaller square inside the outer eye. It helps the reader accurately interpret the data encoded in the QR code.

B. Process to retrieve information

QR code consists of several key components:

1) Scanning: A user scans the QR code using a smartphone camera or a dedicated QR code reader.
2) Decoding: The scanner detects the finder patterns and uses them to align the code. It then interprets the data area, translating the pattern of black and white squares into binary data.
3) Error Correction: If the QR code is partially damaged or obscured, the error correction algorithms reconstruct the original data.
4) Data Retrieval: The decoded binary data is then converted into a readable format, such as a URL, text, or other types of information.
5) Action: Depending on the content, the device may open a web page, display text, or perform another action, allowing users to access the information linked to the QR code.

**QR CODE INDUSTRY INSIGHTS [2]**

With the increasing number of smartphone users and the trend toward mobile-first marketing, QR code scan rates have grown significantly, reaching up to 57 percent across 50 countries.

According to QR Tiger's latest statistical report, global scans have quadrupled in 2024, reaching 26.95 million scans.

8 customized QR codes are created in one minute-clear evidence if the upsurge in QR code usage.

47.68% of QR code users use URL QR codes.

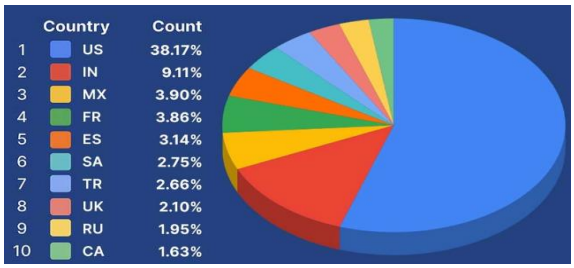



**Fig. 2.**Usage of QR code globally.

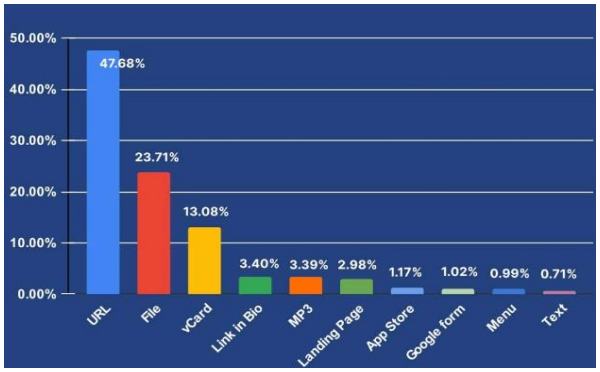



**Fig. 3.**Most popular QR code solutions.

This number tells us that almost half of the global population knows that they can use QR codes to store URLs or website links, leading scanners to different pages online. No wonder it's the most popular QR solution.

## 2. PROBLEM STATEMENT

Consumers mindset toward scanning QR codes, as well as their intentions to do so, are greatly influenced by their perceptions of the codes ease of use and usefulness. Cybercriminals exploit QR codes to spread malware or direct users to phishing sites that can compromise their credentials, personal data, and other sensitive information.

Types of cyber attack using QR Codes:

a) Phishing Attacks: Malicious QR codes can direct users to fraudulent websites designed to steal personal information, such as login credentials or payment details. When scanned, the QR code might lead to a site that mimics a legitimate service.[3]

b) Malware Distribution: Scanning a malicious QR code can initiate the download of malware onto the user's device. This malware could range from ransomware to spyware, compromising the device's security.[4]

c) Wi-Fi Credential Hijacking: QR codes can be used to trick users into connecting to a rogue Wi-Fi network. When scanned, the code might configure the device to connect to a malicious network, allowing attackers to intercept data.[5]

d) SMS or Call Manipulation: QR codes can be used to generate automated calls or text messages to premium-rate numbers, resulting in unauthorized charges to the victim.

e) Data Exfiltration: Attackers can use QR codes to facilitate the unauthorized transfer of sensitive data from compromised devices.

f) Rogue Application Installation: Scanning a QR code can prompt users to download malicious apps that mimic legitimate software, compromising device security.

g) Social Engineering: QR codes can be used in social engineering schemes, where users are led to believe they are engaging with a trusted entity. For example, a QR code on a poster might promise discounts but lead to a phishing site.[6]

### DISTINCTIVE FEATURE OF NLP OVER OTHER ALGORITHMS

Natural Language Processing (NLP) can offer distinct advantages over algorithms like Support Vector Machines (SVM), K-Nearest Neighbors (KNN), and Random Forests in malicious QR code detection, particularly when analyzing the textual elements associated with QR codes. Here are some reasons why NLP may be more effective in this context:

h) Contextual Understanding: NLP techniques are adept at understanding the semantics and context of text encoded in QR codes, such as URLs or commands, which helps identify malicious intent. In contrast, traditional methods may struggle with the nuances of language.[7]

i) Advanced Feature Extraction: NLP utilizes advanced techniques like word embeddings (e.g., Word2Vec, BERT) that automatically capture meaningful features from text, reducing reliance on manual feature engineering. Traditional models like SVM and KNN often require explicit numerical features.[8]

j) Adaptability and Fine-Tuning: NLP models can be fine-tuned on various datasets, allowing them to adapt to emerging threats more effectively than traditional models, which may require extensive retraining.[9]

k) Intent and Sentiment Analysis: NLP techniques facilitate the analysis of intent and sentiment, which can be crucial in identifying phishing or malicious URLs in QR codes. This level of analysis is often beyond the capabilities of SVM, KNN, and Random Forests.[10]

### FRAMEWORK FOR ENHANCING SECURITY OF QR CODE USING CNN & NLP FRAMEWORK

Combining CNN and NLP techniques for malicious QR code detection leverages the strengths of both approaches, providing a robust framework for identifying threats effectively. This integrated method enhances accuracy, adaptability, and overall security, making it well-suited for addressing the challenges posed by malicious QR codes.

Short URLs mask the actual destination, making it difficult for users and automated systems to assess the safety of the link without resolving it first [1]

**CNN(QR Image Analysis)**

Pattern Recognition: CNN is used to analyze the visual aspects of QR codes to recognize visual patterns, making them suitable for analyzing the graphical structure of QR codes. They can detect anomalies or suspicious features in the QR code's design that may indicate malicious intent. Dataset Creation: To train a CNN for this purpose, a dataset containing both benign and malicious QR codes would be required. The dataset should include various styles and designs of QR codes, along with examples of known malicious codes.

INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)
(Int Peer Reviewed Journal)
Vol. 04, Issue 11, November 2024, pp : 321-325
www.ijprems.com
editor@ijprems.com
e-ISSN : 2583-1062
Impact Factor : 7.001

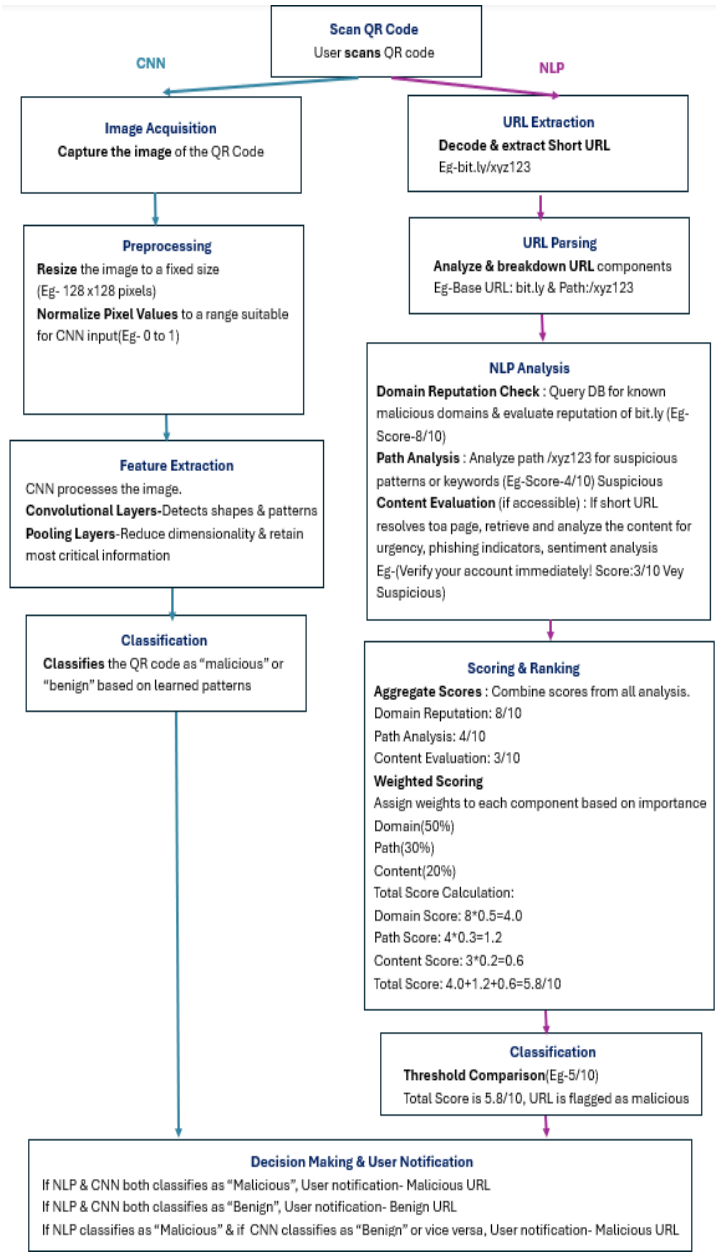



**Fig. 2.** Detailed process flow of Malicious QR code detection using CNN & NLP Framework.

Hierarchical Learning**:** CNN automatically extracts hierarchical features from the QR code images, learning to differentiate between safe and harmful codes based on their visual characteristics.

Binary Classification: Once trained, CNN can classify new QR codes as either malicious or benign based on the learned features.

Multimodal Approaches: Combining CNNs with other techniques (like NLP for analyzing extracted URLs) can enhance the overall detection system, providing a more robust analysis of potential threats.

**NLP(Shorl URL Analysis)**

**URL Analysis-**

Domain Recognition: NLP techniques can analyze the URLs extracted from QR codes, identifying suspicious or known malicious domains (e.g., misspelled domains or newly registered ones).

Keyword Detection: NLP can look for red flags in the URL, such as certain keywords commonly associated with phishing or scams (e.g., "login," "update," "secure").

**Sentiment and Context Analysis-**

Content Evaluation: If the QR code leads to text or webpages, NLP can assess the content for signs of phishing or malicious intent, such as aggressive language or urgent calls to action.

Contextual Understanding: NLP can help understand the context in which the QR code is used (e.g., marketing materials vs. unsolicited flyers), aiding in risk assessment.

**Anomaly Detection-**

Pattern Recognition: By analyzing large datasets of legitimate URLs and text, NLP can identify deviations or unusual patterns that may indicate malicious content.

URL Analysis via domain recognition: NLP techniques can analyse the URLs extracted from QR codes, identifying suspicious or known malicious domains (e.g., misspelled domains or newly registered ones).

## 3. CONCLUSION

While NLP focuses on language-related tasks, CNN is a powerful tool within machine learning that can be applied to a variety of domains, including image and text data. When used together, they can enhance the capabilities of models for tasks like malicious QR code detection by analyzing both the visual aspects of the QR code (using CNNs) and the textual information (using NLP techniques).

Overall, the primary goal of a malicious QR code detection tool is to enhance cybersecurity by preventing users from falling victim to threats associated with malicious QR codes. These tools contribute to a safer digital environment by leveraging advanced detection techniques to analyze both the visual and textual components of QR codes.

## 4. LIMITATION

Both NLP and CNN models typically require large amounts of labeled data for effective training. In the case of malicious QR code detection, gathering comprehensive datasets that accurately represent both benign and malicious samples can be difficult.

Many URL shortening services permit changes to the destination URL even after the short link has been generated. As a result, a short URL that initially directed to a safe site could later redirect to a malicious one.

Malicious actors continuously adapt their tactics, including how they use QR codes. As new threats emerge, models trained on previous data may become outdated quickly.

## 5. FUTURE SCOPE

The establishment of industry standards for QR code safety can guide developers and organizations in implementing effective detection mechanisms. Eg: URL preview before the scanning the code, Expiration Dates of QR Code, ensuring the destination URL is secure with HTTPS.

Future detection systems can be integrated into broader cybersecurity frameworks, allowing for seamless communication and data sharing between various security tools (e.g., antivirus software, web filters)

## 6. REFERENCES

[1] M. N. Islam et al. (2021). "QR Code Security: A Survey of Threats and Solutions." International Journal of Information Security, 20(5), 561-578.

[2] https://www.qrcode-tiger.com/qr-code-statistics-2022-q1

[3] Enck, W., & Octeau, D. (2011). "A Study of Android Application Security." ACM SIGSAC Conference on Computer and Communications Security.

[4] Zarefsky, R., & Beekman, C. (2021). "QR Code Malware: A Growing Threat." Journal of Cybersecurity Research.

[5] Kordzadeh, N., & Ghasemi, M. (2018). "QR Code Scanning Vulnerabilities." International Journal of Information Security.

[6] Agarwal, R., & Mital, M. (2019). "Social Engineering and QR Codes: An Emerging Threat." Cybersecurity and Digital Privacy Journal.

[7] Jelodar, H., et al. (2019). "Text classification algorithms: A survey." Computer Science Review, 27, 1-13. DOI: 10.1016/j.cosrev.2018.10.001.

[8] Mikolov, T., et al. (2013). "Distributed representations of words and phrases and their compositionality." Advances in Neural Information Processing Systems, 26.

[9] Howard, J., & Ruder, S. (2018). "Universal language model fine-tuning for text classification." Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics. DOI: 10.18653/v1/P18-1002

[10] Bhatia, R., et al. (2016). "Analyzing text data for sentiment and intent: A survey." Artificial Intelligence Review, 46(4), 765-798. DOI: 10.1007/s10462-016-9507-4.