Spectrum of Engineering Sciences
ISSN (e) 3007-3138 (p) 3007-312X Volume 3, Issue 12, 2025

MALICIOUS QR CODE DETECTION: A COMPREHENSIVE REVIEW OF METHODS AND CHALLENGES

Faheem Ahmed[*1], Isha Fatima[2], Mehmood Ul Hassan[3], Dr. Iftikhar Rasheed[4], Dr. Umar Fayyaz[5]

[*1,2,3,4,5]*Department of Information and Communication Engineering, The Islamia University of Bahawalpur, Bahawalpur, Pakistan*

[*1]faheemahmedkhan32@gmail.com, [2]ishafatima171@gmail.com, [3]sbmehmood85@gmail.com, [4]iftikhar.rasheed@iub.edu.pk, [5]umar.fayyaz@iub.edu.pk

DOI: https://doi.org/10.5281/zenodo.18196524

Keywords
Malicious QR Codes, Quishing, QR Code Secu-rity, Phishing Detection, Machine Learning, Deep Learning, QR Detection

Article History
Received: 01 October 2025
Accepted: 15 December 2025
Published: 30 December 2025

Corresponding Author: *
Faheem Ahmed

Abstract
QR codes have become popular to permit the speedy and touch-free access to online services in the spheres of payments, health care, transport, and smart surroundings. Nevertheless, their opaque and machine-readable properties have made them a very appealing target of attacks by phishing, the distribution of malware, financial fraud, and unauthorized redi-rection, also known as quishing attacks. Conventional blacklist- based and rule-based security systems are mostly inefficient against such threats, especially in dynamic and physical worlds. The current paper is a review of the research published in the field of malicious QR code detection between 2023 and 2025. The systematically reviewed machine learning, deep learning, hybrid detection frameworks, and infrastructure-level security mechanisms are analyzed, such as URL-based analysis, image- level QR inspection, multimodal learning, and preventive mech-anisms, such as blockchain-based verification and digital water- marking. The important areas like feature extraction techniques, data properties, assessment procedures, and reported trends of performance are scrutinized. It also addresses that if a model works well in theory, it must also handle real data, real users, limited resources, and possible attacks in the real world. Judging by the reviewed literature, this paper establishes an open research gap and presents the possible future directions in the creation of powerful, scalable, and user-friendly QR code security solutions acceptable in the real-world implementation.

INTRODUCTION

There is a popular interface between the physical and digital environment that is brought about by Quick Response QR Codes, as they are simple to create and generate, and they are compatible with mobile devices. They have been extensively integrated in various applications such as digital payments, au-thentication systems, transport services, marketing campaigns, healthcare systems, and supply chain traceability. Further embracement of QR codes in public and commercial places, especially due to the pandemic, has further brought the QR technology to the daily interaction of users [1], [2].

Although QR codes are widely used, they were not initially developed with security in mind. QR payloads are normally plaintext-encoded, do not inherently authenticate, and hence cannot be read by the users before scanning. The design features provide an asymmetric trust model by making users

incapable of assessing the safety of a QR code before engaging with it. Consequently, the QR codes have become a promising target of cybercriminals to carry out a type of attack often known as quishing [3], [4]. When turning off attacks, attackers use QR codes to navigate to phishing sites, deliver malware, corrupt transactions, or steal confidential information [5], [6]. The initial research on the topic of the security of QR codes mainly positioned the QR-based attacks as a continuation of classical phishing. The fact that QR codes are a unique phishing medium since they are opaque and require scanning devices was as one of the first works to formalize this point of view. [3]. The malicious URLs embedded in QR payloads were found to be identified using blacklist-based and heuristic methods, which were later found to be ineffective in identifying attack vectors generated recently, short-lived, or
obfuscated [7], [8].

To overcome these drawbacks, researchers resorted to using machine learning methods more and more in detecting mali- cious URLs, using lexical, structural and host-based features to enhance the generalization [9], [10]. They were used in con- trolled environments to enhance detection accuracy, whereas they were only viable in component situations where non-URL payloads, URL shorteners, or visually manipulated QR codes were not at risk.

More recent studies have turned to deep learning based methods that do not need to decode a payload, but instead analyze the images of the QR code and structural features of the same [11], [12]. The visual-only version of image-based detection techniques that use convolutional neural networks proved that it is possible to detect malicious or manipulated QR codes by visual attention alone. The further improvement in ability against adversarial manipulation and obfuscation came with advanced architectures that incorporated both con- volutional neural networks and recurrent neural networks [13], [14].

In parallel, system-level QR security frameworks have been proposed to integrate detection mechanisms directly into scanning applications and infrastructure environments. These systems combine classification models, anomaly detection, and risk signaling to provide end-to-end protection during QR code interaction [15]–[17]. Preventive approaches, including tamper-resistant QR generation and digital watermarking, aim to ensure QR integrity and authenticity rather than relying solely on post-scan detection [18], [19].

New studies have also examined blockchain-enhanced QR security to offer traceability and assurances of trust in sensi-tive application areas like in supply chain management and finance [20]. Although these solutions prove to be excellent in providing high levels of integrity, they are limited in terms of scalability and complexity of deployment.

In addition to the technical detection mechanisms, recent studies that are people-centered have shown that the behavior of the user is a determining factor in the success of QR- based attacks. Real-world and naturalistic empirical studies indicate that users often, without examining destination URLs and security indicators, scan QR codes, which is often because of implied trust in physical location [21], [22]. The results here point to the fact that high algorithmic detection accuracy may not be directly related to working real-world protection.

Even though an increasing body of literature exists on the topic of malicious QR code detection from a variety of perspectives, the current research is still disjointed in terms of the detection paradigm, datasets, evaluation protocols, and application contexts. The existing reviews tend to concentrate on specific features, including URL-based detecting or scanner performance, which lack a complete analysis that incorporates technical, system-level, and human-oriented orientations [23], [24].

In order to overcome these drawbacks, the current review paper includes an in-depth evaluation of malicious QR code detection methods that were reported from 2014 to 2025. This review has threefold contributions: (i) to systematically identify the existing detection strategies into the following categories: URL-based, image-based, hybrid, preventative and user-centered paradigms; (ii) to examine the performance attributes, assessment practices and consideration, and a (iii) to suggest the research gaps and future research directions to realize effective, scalable and user-obligatory QR code security solutions.

II. LITERATURE REVIEW

Studies concerning malicious detection of QR codes have increased swiftly in line with the increase in abuse of QR codes as a delivery method of phishing, malware, and fraud. QR codes are non-visual because they do not allow the user to visually examine the content codes embedded in them, which makes them highly vulnerable to misinformation [1], [2]. As a result, there has been a broad spectrum of detection approaches investigated by researchers, including URL-based machine- learned systems, image-based deep-learned systems, structure- based deep-learned systems, hybrid system-level systems, and systems-level anthropocentric studies of security.

A. Early QR Code Security and URL-Based Detection

Initial research efforts treated malicious QR codes primarily as a variant of phishing attacks, focusing on the analysis of URLs embedded within QR payloads. One of the earliest and most influential works, QRphish, demonstrated that QR codes represent a distinct phishing vector and proposed decoding QR payloads followed by machine learning–based URL classifica- tion [3]. This work established the foundation for subsequent QR-focused phishing detection studies.

Building on this direction, numerous studies investigated malicious URL detection using classical machine learning models such as Random Forest, Support Vector Machines, and gradient boosting techniques [7]–[9]. These methods are usually based on lexical, structural, and host-based criteria, and they encompass the length of the URL, entropy, frequency of the token, domain age, and the hosting data. These perform very well in structured but shortened URLs, obfuscations, or Quishing comes tricky to detect where attackers change tactics quickly.

Some extensions of the URL-based detection to specific QR code contexts and adding QR decoding steps and scanner- level analysis were made [6], [10]. These studies highlight the fact that URL-based detection alone is not sufficient for QR- related attacks, especially when the QR Code contains non-URL-based data.

Besides the main methods discussed above, some other studies also expand on the topic of ensemble learning, feature engineering, and other machine learning classifiers that are used to detect malicious URLs in the QR code payloads and show better resilience to the changes in phishing tactics [8], [9], [25].

B. Image-Based and Structural Deep Learning Approaches

In order to break the drawbacks of URL-based detection, more recent studies have shifted attention to the analysis of QR code images and structural patterns per se. Conventional neural networks were developed to be lightweight to predict QR code images by treating them based on spatial and texture patterns of activities such as tampering and malicious motive [11], [12]. Such techniques proved that malice can be detected even prior to payload execution by QR image attributes.

Further complicated deep learning models integrate the con- volutional feature extractors with recurring models to represent the spatial and temporal characteristics in the QR structures. CNN-RNN hybrid systems with gated recurrent unit or long short-term memory network have been reported to achieve high detection rates and, at the same time, be feasible in real- time [13], [14]. These models are notably useful to counter the obfuscated, adversarial, and multi-layer QR code attacks. An alternative to pixel-level learning, structural analysis has also investigated. Analyses of the architecture of QR codes, their alignment, and irregularities during error corrections show that malicious manipulation frequently causes some structural variations [2], [24]. Nonetheless, the methods are still sensitive to the quality of printing, environmental noise,
and the variety of datasets.

Other papers examine lightweight and resource-efficient deep learning models in the context of QR code attack detection, with a focus on the trade-off between the detection error and athe bility to run on a mobile and edge device [12], [26], [27].

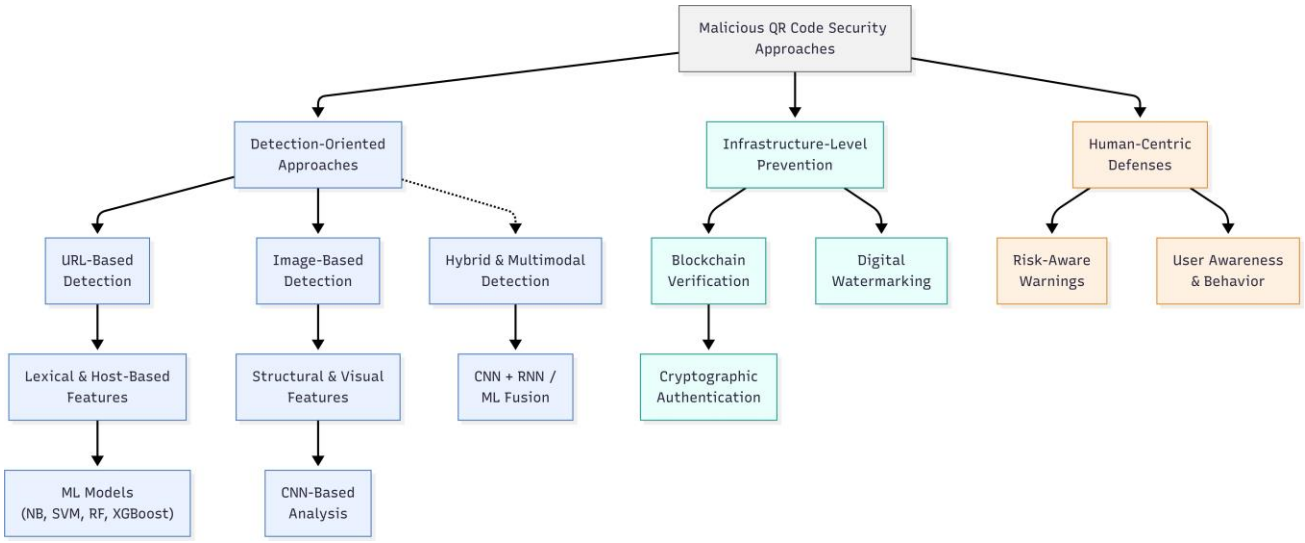

Fig. 1. Taxonomy of malicious QR code security approaches, highlighting detection-oriented, infrastructure-level, and human-centric strategies

C. Hybrid, Multimodal, and System-Level Frameworks

It has been realized that no single modality can be used to identify all types of attacks, and thus, a number of studies have been conducted on hybrid and multimodal detection models. QR Shield proposed to use a dual-machine-learning architecture, which uses multiple classifiers to enhance ro- bustness and minimize the instances of false positives when scanning a QR [15]. On the other hand, QsecR is a secure QR scanner developed here, which does feature extraction from the URLs, its classification, and privacy-conscious risk assessment metrics [16].

Other system-level solutions focus on end-to-end security that involves direct integration of the detection mechanisms within the QR scanner applications. QR security frameworks that are enhanced with AI show a higher level of real-world applicability since they incorporate an anomaly detection element, attack simulation, and a risk signaling element [5], [17]. These systems are deployment-oriented as opposed to independent classification precision.

Instead of detecting attacks from the QR Codes, these preventive methods aim to make sure the QR code hasn't been altered. The digital watermarking of the tamper-resistant QR code generation and neural network-based verification of authenticated allow the validation of authenticity without in- volving an external database. [18], [19]. These approaches can prevent code replacement attacks, but they typically involve the alteration of how QR is generated.

A number of other studies suggest framework-based and scanner-based solutions that balance the accuracy of detection with the implementability of these solutions, such as response latency, usability, and compatibility with the existing QR ecosystem [28], [29].

D. Blockchain-Based and Domain-Specific QR Security

QR security facilitated by blockchain has become a prospec- tive field of application in the sphere of activities that need traceability and trust guarantee. A number of research works use blockchain to document the QR transactions and make it immutable, and cannot be changed [1], [20]. The systems that need trust and traceability, like supply chains and payment systems, or authentication, still face issues like scalability and latency.

Domain-specific inquiries also testify to the fact that QR- based attacks may appear in diverse ways within the context of applications. The research and works on supply chain traceability,

online payments, and smart infrastructure identify that the security requirements of QRs can differ significantly, depending on the deployment context.

Table I
SUMMARY OF TECHNICAL APPROACHES FOR MALICIOUS QR CODE DETECTION

Author(s)	Year	Approach	Key Contribution
Kharraz et al.	2016	URL-based ML	Introduced QR codes as a phishing vector and applied ML-based URL analysis for QR phishing detection.
Nigam	2025	Survey / Review	Provided a performance-oriented review of QR phishing detection methods and high- lighted scanner-level limitations.
Khan et al.	2024	URL-based ML	Proposed a dual-model framework using lexical and host-based URL features for real-time quishing detection.
Tayachi and Ouni	2025	ML/DL (URL)	Developed a real-time quishing detection system using Random Forest and LSTM models for malicious URL identification.
Alsulami et al.	2025	Image-based DL	Proposed a CNN–RNN framework enabling pre-scan detection of tampered and mali- cious QR code images.
QR Shield	2024	Hybrid ML	Introduced a dual-classifier framework to improve robustness and reduce false posi- tives during QR scanning.
Alsuhibany	2025	Preventive / Watermarking	Proposed tamper-proof QR code genera- tion using digital watermarking and neural network-based verification.
Ahamed et al.	2025	Blockchain-based	Utilized blockchain to ensure QR code au- thenticity, integrity, and traceability in sup- ply chain environments.
MultiPhishNet	2025	Multimodal DL	Demonstrated the effectiveness of multi- modal phishing detection by combining URL, visual, and contextual features.

E. Human-Centered and Usable Security Studies

The Intelligent Human Behavior towards QR Code-based at- tacks is a real success, besides technical detection techniques. Empirical research carried out in real-life settings indicates that the users do scan QR codes often, without checking destination URLs or security indicators [21]. Such results prove that the

effectiveness of QR phishing is frequently higher than that of standard phishing because of implicit confidence in the physical location of QRs.

Mass user surveys also indicate that QR-based phishing is far more difficult to identify in users than email-based phishing, even with a security-conscious user base. [22]. Sup- plementary surveys and usability-related research are found to focus on these underlying areas of scanner interface design, warning visibility, and user attention, meaning a great deal to the attack success rate [1], [23].

Overall, the literature reveals a clear evolution from URL- based machine learning toward deep learning, hybrid, and system-level solutions for malicious QR code detection. While image-based and multimodal approaches demonstrate superior robustness, challenges related to generalization, explainability, deployment feasibility, and user interaction remain unresolved. These observations motivate the need for integrated, user- aware, and practically deployable QR security solutions.

TABLE II
HUMAN-CENTRIC STUDIES ON QR CODE ATTACKS

Author(s)	Year	Key Finding
Sharevski et al.	2024	Users frequently scan QR codes without inspecting destination URLs, increasing susceptibility to quishing attacks.
Kowalewski et al.	2025	QR-based phishing is significantly harder for users to detect than traditional email phishing attacks.
USEC Study	2024	Persuasive pretexts and lack of security indicators in QR scanners lead to high success rates of real-world QR phishing.

III. PERFORMANCE ANALYSIS AND DISCUSSION

The key feature of the malicious QR code detection study is performance evaluation, as it not only defines the classification accuracy, but also evaluates its potential in the real world and its strength. In the literature reviewed, the metrics that are most widely used to evaluate the performance include accuracy, precision, recall, F1-score, and inference latency. Nevertheless, it is difficult to subjectively compare studies directly since datasets, representations of features, attack models, and eval- uation protocols differ.

Machine learning methods that are URL-based tend to show high accuracy on benchmark datasets, and the accuracy is normally reported in the range of 90% to 97%. These have the advantage of low computational cost, interpretability, and deployability, which is why they can be used in real-time QR scanning systems. However, according to several studies, performance is always negatively affected in the face of truncated URLs, obscured payloads, or newly created domains. Deep learning based systems that work directly on QR code images have greater resistance to obfuscation and tampering. Convolutional neural networks based on images and CNN- RNN hybrid networks are likely to achieve detection rates of over 95% in controlled settings. One of the most important benefits of such methods is the possibility of conducting a pre-scan detection, which allows detecting malicious QR codes before code execution. Nevertheless, their performance is strongly related to the diversity of the dataset, the quality of images, and the environmental factors, including light, camera resolution, and printer artifacts.

Hybrid and multimodal frameworks are always reported to have the highest performance in the literature. Hybrid systems are able to detect with more accuracy instead of single model systems, which have fewer false positives with the integration of URL and visual characteristics and contextual cues. These multi-model systems are tough enough that single ones are not easy to fool. With the benefits, hybrid systems have high computational power, and

their cost of integration is also high, which can be a significant weakness in deploying on actual systems.

QR security solutions at the system level focus on end-to- end performance as opposed to individual classification scores. Real-time detection, anomaly analysis, and risk signaling frameworks based on secure scanners exhibit good results in real-world scanning. The systems show that the accuracy of detection is important, as well as latency, usability, and privacy. Preventive systems such as tamper-resistant QR code generation and digital watermarking are believed to be largely integrity-oriented but not classification-oriented. Even though these techniques can be utilized to prevent QR replacement and manipulation attacks, they would be applied only with regard to existing QR ecosystems.

In the supply chain and finance sector, QR verification systems are very crucial because blockchains provide a good guarantee of authenticity and traceability. In such studies, performance analysis is on transaction latency, scalability, and system overhead instead of traditional detection metrics. Blockchain-based solutions might not be practicable for high- frequency or consumer-grade QR scanning because of infras- tructure requirements, although practical in controlled settings. Outside the algorithmic performance, the human-centred research indicates the existence of a significant gap between detection and actual security performance. There are some real-life experiments that prove that users do not think before scanning the QR Codes, as well as in cases where detection systems are in place. These results indicate that high accuracy of classifiers is not sufficient to make attacks unsuccessful. As a result, a number of studies point to the integration of usable security, known as explainable warnings, clear risk indicators, and delayed execution strategies to enhance decision-making
among users.

When it comes to comparing different systems, there is no standard datasets which makes the comparison a little bit difficult. Most research works use proprietary or synthetic data, which restricts reproduction and raises the issue of extrapolation to the real-world attack scenario. Moreover, not many of them consider long-term resilience in case of adaptive adversaries or implement large-scale field deployments.

Altogether, the analysis of the performance of existing studies suggests that deep learning and hybrid solutions are superior in detecting QR codes, but implemented in the real world, a combination of factors, such as computational effectiveness, deployment, explainability, and user behavior, must be chosen to ensure the high-quality security of QR codes used in practice. Future performance measurement should shift its focus from individual accuracy to the standardized, real- life, user-conscious evaluation techniques to guarantee that it is adequately ready to defend against the emerging QR-based attacks.

Additional comparative discussions indicate that the report- ing measures of performance are highly subject to experimen- tal assumptions and scanner-related limitations, which contin- ues to highlight the significance of holistic measurements of performance as opposed to isolated measurements of accuracy.

IV. CROSS-PARADIGM ANALYSIS AND RESEARCH INSIGHTS

Although the research on malicious QR code detection is rapidly growing, the current literature is still disjointed in terms of detection paradigms, threat assumption, dataset, and evaluation practice. Although the individual research findings indicate that the results are positive in their respective areas of concern, holistic view on how the two methods correlate with each other and in what major aspects they diverge is yet to be achieved. It is in this section where a cross-paradigm synthesis of the literature reviewed is provided showing some of the important conceptual insights that are not apparent when the studies are studied separately.

A. Detection Timing and Trust Boundary

Among the most important, yet least research studied distinctions, in malicious QR code research is the question of when the detection is performed with reference to user interaction. The methods that are currently in use can broadly be divided into post-scan detection, pre-scan detection, and preventive-by-design mechanisms. Machine learning systems based on URLs normally work following the decoding of QR codes, i.e. the trust boundary is already breached

when the user scans the code. Contrarily, image-based and structural deep learning methods can be used to perform pre-scan detection where visual or structural QR characteristics are analyzed prior to payload execution. Even more preventive procedures, like digital watermarking, blockchain-based verification, impose even earlier security by including integrity and authenticity at the stage of generating or verifying the QR.

This difference is essential since the presence of detection accuracy does not ensure the security of users. After scanning has been performed, the user can already be exposed to phish- ing sites, the delivery of malware or unauthorized redirections. This means that strategies with lower lifetime in the interaction lifecycle have a higher security guarantee even though the classification accuracy as reported is the same as that of post scan strategies.

TABLE III
Performance Summary of Representative Malicious QR Code Detection Approaches

Study	Approach Type	Dataset Used	Evaluation	Accuracy	Recall	F1-score
Khan et al. (2024)	URL-based ML (RF)	Custom	10-fold CV	94.2%	0.93	0.94
Tayachi et al. (2024)	ML/DL (RF + LSTM)	Phishing URL dataset	Train/Test split	97.8%	0.98	0.98
Alsulami et al. (2025)	Image-based DL (CNN)	Synthetic + real QR im- ages	Hold-out	98.6%	0.97	0.98
Multiple Authors (2024)	Hybrid (CNN + RNN)	Mixed QR + URL data	Cross-validation	99.1%	0.99	0.99
Alsuhibany (2025)	Preventive (Watermarking)	Integrity-protected QR codes	Verification accu- racy	Not classification- based	Not classification- based	Not classification- based
Blockchain-based (2025)	Infrastructure-level	Supply-chain QR sys- tem	Latency and in- tegrity analysis	Integrity-oriented	Integrity-oriented	Integrity-oriented

B. Threat Model Misalignment Across Studies

The implicit application of divergent threat models is also another significant source of unregulated results in the litera- ture. Other studies suppose opportunistic attackers using fixed phishing QR codes and other ones assume adaptive adversaries using URL shortening, high-speed domain rotation, or adver- sarial obfuscation. Attacks on the physical world, including QR code replacement, overlays, or even partial damage in public spaces, can be considered as a specific type of threats, which URL-centric models do not fully cover. Another dimen- sion of adversarial layout that is introduced by infrastructure- level attacks in the supply chain and payment systems is that integrity and provenance are of primary importance compared to payload classification.

Performance measurements are often viewed out of context as many studies are silent on which threat models they assumed. A model that works very well against one type of threat could be totally broken against another adversarial assumption. Such incongruity explains why threat-conscious assessment

is necessary and discourages the direct comparison of performance in the heterogeneous studies.

C. Dataset Realism and Evaluation Blind Spots

Absence of standardized and publicly available datasets of malicious QR codes has been one of the most important obstacles in the field. The majority of them are based on the synthetic generation of the QR code or URL-based dataset, which is subsequently converted to QR symbols. Though these datasets can be experimented and scaled, they are not always successful in representing the actual environmental factors like changing light, camera distortion, printing artifacts, occlusion, and physical degeneration.

In addition, existing data sets mostly overlook the human aspect of the QR code interaction. Most of the evaluation pipelines do not provide user behavior, cues of contextual trust, or interface design, which have been demonstrated to have a significant effect on attack success. Consequently, high reported accuracy with curated datasets does not in any way give significant protection in the real world.

D. Accuracy Versus Real-World Protection Gap

The classical machine learning metrics of accuracy, pre- cision, recall, and F1-score are greatly considered by the literature. Nonetheless, such measures are not enough to measure the QR code security in the real world. False negative consequences can be much more serious than false positives in real-world applications, the latency to inference can be a critical factor to usability and inference explainability is a key element in user acceptance and trust.

Preventive and infrastructure-level solutions also pose an additional challenge to traditional evaluation procedures, since their usefulness is more effectively evaluated in terms of in- tegrity assurance, latency, scalability, and operational stability than in terms of classification accuracy. This discrepancy be- tween measure of evaluation and the real-life goals constitutes a subject matter of a gap in the modern research practice.

E. Deployment Readiness and Operational Maturity

One more cross-cutting insight is related to the operational maturity of suggested solutions. The majority of the work is at the prototype and simulation level, with little thought given to the long-term implementation issues of model maintenance, adversarial drift, privacy protection, and compatibility with existing QR ecosystems. A very limited portion of the works incorporates detection into the scanning applications or sub- systems on an infrastructure level, where usability, resource limits, and system interoperability are particularly important. This finding implies that future studies need to go be- yond stand-alone model demonstrations and to end-to-end, deployment-aware security solutions. These issues should be tackled in order to transfer academic gains into viable and credible systems of QR code protection.

V. LIMITATIONS AND FUTURE WORK

Although there is a considerable advancement in detecting malicious QR codes, there are still a number of weaknesses in

the existing bodies of knowledge. Lack of standardized data and publicly available data is one of the major challenges. Most of the research works are based on proprietary, synthetic, or small-scale data, which ensures that the reproducibility is reduced, and it is not possible to perform a fair performance assessment between the various detection methods. Conse- quently, the reported measures of performance might not be properly reflected in real-life performance.

The other significant disadvantage is that most detection models are not widely generalizable. Image-based deep learn- ing models can be vulnerable to changes in lighting conditions, noise, printing quality, and camera resolution, whereas URL- based approaches will have difficulty with shortened and obfuscated links as well as dynamically generated links. These aspects may have a huge effect on the accuracy of the detection in practical applications.

The issue of explainability of the model is also a burn- ing question, especially in the case of deep learning-based methods. On the one hand, the complex models have high detection accuracy, but their decision-making processes are typically black-

box, and it is hard to use and implement system designers to comprehend why a QR code is a malicious object. Such untransparency may decrease user confidence and adoption in security-sensitive applications.

Computational overhead and resource constraints are further challenges as far as deployment is concerned. Hybrid and multimodal systems are also very accurate, but they might not be suitable in low-power and real-time settings, which require greater inference latency and memory. Moreover, blockchain- based solutions at the infrastructure level, e.g., by utilizing verification, present scalability and integration issues that restrict their use in general-purpose QR scanning.

Another limitation that is under-examined in the research is human factors. The detection systems most of them assume that the user is rational and that they do not properly take into consideration the inattention of users, excessive trust in the position of physical QR codes, or lack of meaningful security indicators on the scanner interfaces. Consequently, quality technical accuracy is not necessarily a good fit in the field of practical protection.

The next step in research development should focus on the creation of uniform benchmarking data and evaluation proce- dures to conduct reproducible and similar performance studies. More attention should also be given to the explainable and interpretable detection models capable of giving transparent risk indicators to end users. To positively impact the real-life level of security, it is critical to introduce user-focused design features to the QR scanning software, including the provision of visual alerts, delayed execution, and better display of the destination.

Further investigations should also focus in the future on the lightweight and adaptive detectors frameworks that would be able to offer accuracy with low-resource and comput- ing requirements, allowing their implementation on resource- constrained devices. Multimodal and cross-domain detection approaches incorporating visual, lexical, contextual, and be havioral cues are very promising in terms of enhancing re- silience to adaptive attack methods. Lastly, longitudinal trials testing the detection capabilities on real world conditions and an adaptive adversarial environment are required to further promote the effectiveness of QR code ecosystems in the real world.

VI. CONCLUSION

In this review, a critical and structured analysis was done on the malicious QR code detection research, encompassing the detection and prevention methods that have been devised as a reaction to the increased security risks posed by the use of QR codes. The synthesis of the studies published in 2014 to 2025 enabled the paper to describe the shift of tradi- tional rule-based defenses to intelligent machine learning, deep learning, and multimodal detection systems and infrastructure- level mechanisms designed to guarantee QR code integrity and authenticity.

Although learning-based methods have shown high detec- tion rates when using controlled conditions, this review un- derlined the fact that practical use is limited by the drawbacks of datasets, a lack of standardized, benchmarked results, com- putational bottlenecks, adversarial resilience, and human be-haviour. The results highlight the idea that technical accuracy cannot be used on its own but it needs realistic assessment, comprehensible decision-making, and user-conscious design.

The areas of focus in future research include standardized datasets, sound multimodal architecture, lightweight deploy- ment with privacy and risk communication. The analysis of these issues is critical to creating robust and reliable QR code security systems with the potential to work in dynamic and hostile environments.

REFERENCES

P. F. Katharina Krombholz, “Qr code security: Attacks, challenges, and countermeasures,” Lecture Notes in Computer Science, 2023.

Y. Y. Ye Tiana, “From past to present: Evolution of qr code security,”
Journal of Information Assurance, 2025.

A. Kharraz, E. Kirda, and W. Robertson, “Qrphish: An automated approach for detecting qr code phishing attacks,” IEEE Security & Privacy, 2016.

J. N. David Njuguna, “Qr code security: Attacks, challenges, and countermeasures,” Journal Of Cyber Security, 2025.

L. Rivas, V. K. Singh, et al., "Securing qr codes infrastructure using ai to detect malicious activity," IEEE CCWC, 2025.

A. Tayachi and B. Ouni, "Quishing attack detection and mitigation using machine learning and deep learning for malicious url identification," International Wireless Communications and Mobile Computing, 2025.

C. R. Nuria Reyes-Dorta, Pino Caballero-Gil, "Detection of malicious urls using machine learning," International Journal of Computer Appli- cations, 2024.

M. K. Fuat Tu¨rk, "Malicious url detection using advanced machine learning techniques," Advances in Engineering Software, 2025.

A. S. M. K. Suresh Sankaranarayanan, "Ensemble learning-based mali- cious url detection," PLOS ONE, 2024.

V. R. M V H Sai Sriraj, "Malicious urls and qr code classification using machine learning and deep learning techniques," Journal of Cybersecurity Research, 2023.

S. M. Mousa Sarkhi, "Detection of qr code-based cyber attacks using a lightweight deep learning model," Engineering, Technology & Applied Science Research, 2024.

Y. Alaca, "Lightweight deep learning-based qr code attack detection," Engineering, Technology & Applied Science Research, 2023.

A. Alsulami, Q. Abu Al-Haija, and B. Alturki, "Efficient malicious qr code detection system using an advanced deep learning approach," Computer Modeling in Engineering & Sciences, 2025.

N. A. Yahya Tashtoush, Mahar Khashim, "Comparing machine learn- ing and deep learning models for qr code phishing detection," IEEE Conference Proceedings, 2025.

H. Y. A. Suliman Alsuhibany, "Qr shield: A dual machine learning approach towards malicious qr codes," Electronics, 2024.

A. S. Rafsanjani and N. Kamaruddin, "Qsecr: Secure qr code scanner according to a novel malicious url detection framework," IEEE Access, 2023.

Z. Mukhammedali, "Ai-enhanced security framework for qr codes," IEEE Conference Proceedings, 2025.

H. Z. Tianyu Wang, "A texture-hidden anti-counterfeiting qr code and authentication method," Sensors, 2023.

S. A. Alsuhibany, "Innovative qr code system for tamper-proof genera- tion and fraud-resistant verification," Sensors, 2025.

N. Ahamed, T. Murugan, and B. Mohanta, "Qr code phishing (quish- ing): Food supply chain management traceability risks combined with blockchain technology," IEEE ICBATS, 2025.

F. Sharevski, "Exploring phishing threats through qr codes in naturalistic settings," in Workshop on Usable Security (USEC), 2024.

M. Kowalewski, L. Lassak, M. Du¨rmuth, and T. Schnitzler, "Scanned and scammed: Insecurity by obsqrity? measuring user susceptibility and awareness of qr code-based attacks," in USENIX Security Symposium, 2025.

N. Nigam, "Performance analysis of qr phishing detection approaches," Journal of Information Security and Emerging Technologies, 2025.

L. S. Khedekar, "A comparative review of qr code scanners according to a malicious url detection framework," International Journal of Infor- mation Security, 2025.

S. Abad, H. Gholamy, and M. Aslani, "Classification of malicious urls using machine learning," Sensors, 2023.

H. Y. Dongwan Shin, "Detection of qr code based cyber attacks," IJRASET, 2024.

S. S. H. R. Kaushik Goswami, Subhagata Sardar, "Malicious qr code detection using machine learning," IJCRT, 2024.

A. S. RAFSANJANI, "Qr-based security frameworks for malicious url detection," Electronics, 2024.

R. Shevchuk, "Qr code scanner with anti-qrishing real-time technique," International Journal of Computer Applications, 2024.