

Quantum Machine Learning in Cybersecurity: A Hybrid Convolutional Neural Network with Skip Connections for QR Code Phishing Detection

Md Farhan Shahriyar

farhanshahriyar.cse1@gmail.com

Hajee Mohammad Danesh Science and Technology University

Research Article

Keywords: Quishing detection, Hybrid quantum-classical neural network, Variational quantum circuit, QR code security, Quantum machine learning, Skip connection, Cybersecurity

Posted Date: September 29th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10965779/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Quantum Machine Learning in Cybersecurity: A Hybrid Convolutional Neural Network with Skip Connections for QR Code Phishing Detection

Md Farhan Shahriyar^{1*}

^{1*}Department of Computer Science and Engineering, Hajee Mohammad Danesh Science and Technology University, Dinajpur, Bangladesh.

Corresponding author(s). E-mail(s): farhanshahriyar.cse1@gmail.com;

Abstract

QR code phishing, commonly called quishing, has become a significant cybersecurity threat that exploits the widespread adoption of QR codes in contactless payment, ticketing and authentication. Existing countermeasures rely almost entirely on decoding the code and screening the extracted URL, which fails when the payload is obfuscated, when the destination is a zero-day domain or when no network connection is available at scan time. This study proposes a Hybrid Quantum-Classical Convolutional Neural Network (HQCNN) that classifies QR code images directly from their visual structure without any decoding step. The architecture couples a classical convolutional backbone with a parameterised variational quantum circuit using angle embedding, trainable rotations and a configurable entangling layer. A further variant, HybridSkip, introduces a residual bypass carrying backbone features directly to the classification head alongside the quantum path which preserves gradient flow, retains spatial context and removes the wide compensatory layer the standard hybrid model needs. Six configurations are evaluated on a balanced corpus of QR code images under identical stratified five-fold cross-validation splits and a held-out test set. HybridSkip with 6 qubits and 4 variational layers reaches an F1-score 3.3 percent higher and a recall 6.8 percent higher than the strongest classical baseline while using 37.6 percent fewer trainable parameters. Paired tests across folds confirm the recall gain is statistically significant with a large effect size. The results support parameter-efficient hybrid architectures as a route to proactive quishing defence on resource constrained endpoints in the noisy intermediate scale quantum era.

Keywords: Quishing detection, Hybrid quantum-classical neural network, Variational quantum circuit, QR code security, Quantum machine learning, Skip connection, Cybersecurity

1 Introduction

Quick Response (QR) codes are two-dimensional matrix barcodes that encode data as a grid of black and white modules. Since the relaxation of their patent restrictions and the arrival of native camera-based scanning on smartphones, they have become a default interface for digital payment, ticketing, restaurant menus, product authentication, public transport and multi-factor login. Billions of scans are processed globally every year, and the contactless behaviour normalised during the COVID-19 pandemic accelerated the trend further [1, 2]. The same properties that make QR codes convenient also make them a near-ideal delivery vehicle for social engineering.

The central problem is that a QR code is opaque to the human eye. A user cannot inspect the encoded destination before scanning, and most scanning applications either open the target automatically or present a truncated URL that offers little protection [3]. Attackers exploit this opacity by encoding links to credential-harvesting pages, drive-by download endpoints or premium-rate service triggers, a family of attacks now referred to as *quishing*, a contraction of QR and phishing [4]. Because the malicious artefact is a printed or displayed image rather than an email, it sidesteps the mail gateways, link rewriters and attachment sandboxes that carry most of the load in conventional anti-phishing stacks. A sticker placed over a legitimate parking meter code or a poster in a transit station requires no compromise of any digital channel at all.

Field evidence confirms that users are highly susceptible to this vector. Vidas et al. [5] observed that a large majority of participants who scanned a QR code went on to visit the associated website, and Sharevski et al. [2] found that a substantial share of participants were willing to surrender federated login credentials to a site reached purely through a malicious code. Kieseberg et al. [6] analysed the QR code as an attack vector from the adversary’s point of view and demonstrated that the encoding itself can be manipulated to carry command injection and cross-site scripting payloads, not merely deceptive links. Yong et al. [7] surveyed the countermeasure landscape and reported that few deployed scanner applications perform any meaningful validation before dereferencing a URL.

Existing defences follow one of two paths. The first decodes the QR code and checks the recovered URL against blacklists, reputation services or handcrafted heuristics [1, 7]. This approach is inherently reactive. It cannot recognise a domain registered minutes earlier, it is defeated by URL shorteners and chained redirections, and it requires network connectivity at the moment of scanning, which is often unavailable on offline kiosks, point-of-sale terminals and embedded readers. The second path treats detection as a supervised learning problem over decoded structural metadata such as module counts, error-correction codeword ratios and encoding-mode indicators [8]. This is more proactive, but it still depends on a successful decode and therefore inherits the same failure modes when a code is deliberately obfuscated, physically damaged or overlaid.

A third and much less explored paradigm treats the QR code purely as an image and classifies it before any decode is attempted [4]. The intuition is that the payload shapes the picture. Malicious codes typically encode longer strings, because attacker-controlled URLs tend to carry subdomains, tracking parameters, encoded redirection

targets and obfuscation padding. Longer payloads force a higher symbol version, which raises the module count, changes module density and alters the distribution of error-correction codewords across the symbol. These effects leave a statistical signature in the raw pixel grid that a convolutional model can learn without ever recovering the underlying string [9]. The practical appeal is that such a classifier runs offline, adds no round trip and can act as a first-stage filter in front of any decode-based check. The practical difficulty is that labelled quishing corpora are small, so a purely classical model with a large parameter budget tends to overfit.

Quantum machine learning offers a principled response to precisely this small-data, small-parameter regime. A variational quantum circuit, meaning a quantum circuit whose gate rotation angles are trainable parameters adjusted by a classical optimiser, encodes an n_q -dimensional feature vector into a state space of dimension 2^{n_q} using only $\mathcal{O}(n_q L)$ trainable angles for L circuit layers [10, 11]. Encoding data into a quantum state is formally a nonlinear feature map into a high-dimensional Hilbert space, which is the quantum analogue of the kernel trick in classical machine learning [12, 13]. Hybrid quantum-classical neural networks exploit this by placing a variational circuit inside an otherwise classical network, so that a classical backbone handles perception and a compact quantum block handles the expressive nonlinear mapping [14–16]. Theoretical work indicates that such models can achieve favourable effective dimension [17] and can generalise from comparatively few training samples [18], both of which matter when labelled malicious samples are scarce. In the cybersecurity domain specifically, Shahriyar et al. [19] showed that a variational quantum classifier reaches competitive phishing detection accuracy on modest sample counts, establishing a precedent for quantum models in security applications.

Despite this convergence of motivations, no prior work has applied a hybrid quantum-classical architecture to quishing detection, and no prior work has examined quantum models operating on QR code imagery at all. The gap is consequential, because visual quishing detection is precisely the class of task where quantum feature maps are theoretically favoured, meaning a small, security-critical dataset on which classical convolutional models plateau.

This study addresses that gap. The contributions are as follows.

1. The first hybrid quantum-classical convolutional architecture for QR code phishing detection that operates directly on pixel data is presented, requiring neither decoding nor network connectivity. The model combines a five-layer convolutional backbone with a variational quantum circuit using angle embedding, trainable R_Y and R_Z rotations and a CNOT entangling ladder.
2. **HybridSkip** is introduced, a residual variant in which the projected backbone features are routed to the classification head in parallel with the quantum path. The design preserves gradient flow around the quantum bottleneck, retains classical spatial context that the low-dimensional encoding would otherwise discard, and removes the wide compensatory fully connected layer that the standard hybrid model requires.
3. A controlled six-configuration benchmark is provided, covering two classical baselines, two standard hybrid models and two HybridSkip models, all sharing an

identical backbone and evaluated on identical stratified five-fold cross-validation splits and a common held-out test set.

4. A parameter-adjusted analysis shows that HybridSkip with 6 qubits and 4 variational layers attains a cross-validation F1-score of 0.694 ± 0.027 and recall of 0.710 ± 0.031 , which are gains of 3.3 percent and 6.8 percent over the strongest classical baseline, using 21,015 trainable parameters against 33,673. This is a 37.6 percent reduction and places the model on the Pareto front of the accuracy-versus-size trade-off.
5. The comparison is supported by paired t -tests, one-tailed Wilcoxon signed-rank tests, Cohen’s d effect sizes and 95 percent confidence intervals across folds, rather than reporting point estimates alone.

The remainder of the paper is organised as follows. Section 2 reviews QR code security, classical quishing detection, hybrid quantum-classical learning and residual design, and states the research gap. Section 3 formalises the problem and describes both architectures. Section 4 details the dataset, splitting protocol, augmentation and implementation. Section 5 reports cross-validation results, held-out test results, circuit depth analysis, statistical testing and parameter efficiency. Section 6 interprets the findings and discusses deployment. Section 7 states limitations and future work, and Section 8 concludes.

2 Related work

2.1 QR code security and the quishing threat surface

Security analysis of QR codes predates the current wave of attacks. Kieseberg et al. [6] gave one of the earliest systematic treatments, examining the code as an attack vector and enumerating strategies ranging from simple malicious redirection to injection payloads aimed at the parsing logic of the reader application. Krombholz et al. [3] surveyed the attack space and identified the human factor as the dominant weakness, noting that the encoding is unreadable to the user and that adversaries can physically overlay stickers on legitimate codes in public spaces at negligible cost. Their analysis framed QR security as a usable security problem rather than a purely cryptographic one.

Empirical studies confirm the severity. Vidas et al. [5] ran two field experiments on QR-initiated phishing and found that most users who scanned a code proceeded to the linked site. Sharevski et al. [2] conducted a 173-participant field study built around a health-passport registration pretext and observed that a majority of participants were prepared to authenticate with existing federated accounts at a destination reached solely through a QR code, with convenience cited as the dominant reason. Focardi et al. [1] studied the usability and security trade-off of embedding cryptographic authentication inside the code itself, and showed that signature schemes strong enough to be meaningful inflate the symbol to a size that degrades scanning performance, which explains why authenticated QR codes have seen almost no real-world deployment. Yong et al. [7] surveyed current attacks and countermeasures and reported that most deployed reader applications provide little more than a URL preview.

The collective picture is a threat surface with no authentication layer, a user population that cannot inspect the payload, and reader software that largely defers the security decision to the user. This motivates detection at the client before the payload is acted on.

2.2 Classical detection of malicious QR codes

Early defences decode the symbol and screen the recovered URL. Focardi et al. [1] and Yong et al. [7] both discuss blacklist and reputation lookups as the standard mechanism, and both identify the same limitations. Blacklists lag domain registration, so freshly minted phishing infrastructure is invisible until it is reported. Shorteners and multi-hop redirection break naive string matching. Any online lookup also assumes network availability, which fails for offline scanners, air-gapped kiosks and embedded industrial readers.

Machine learning approaches move beyond static lists. Karagülle et al. [8] applied gradient-boosted classifiers to engineered QR *structural features*, meaning explicitly decoded module counts, error-correction codeword ratios and byte-mode indicators, and reported AUC above 0.9 on their dataset. That work is strong within its feature space, but the feature space itself is the constraint. Structural metadata can only be extracted after a successful decode, so the method fails on codes that are deliberately obfuscated, physically damaged, partially occluded or rendered in non-standard styles.

Alenezi et al. [4] took the image-level route, training a lightweight deep model to sort QR codes into normal, phishing and malware categories from visual features alone. This established the feasibility of decode-free classification and is the closest classical precedent to the present work. It does not, however, examine parameter efficiency, and it does not consider quantum feature maps.

The present work sits in the decode-free image classification category and is therefore complementary rather than competitive with respect to structural feature methods. A deployment could reasonably run both, using the visual classifier as a first-stage filter that also covers the cases where decoding fails.

2.3 Hybrid quantum-classical neural networks

Parameterised quantum circuits act as trainable function approximators whose parameters are optimised by a classical routine, a pattern surveyed comprehensively by Benedetti et al. [10] and Cerezo et al. [11]. Schuld et al. [20] formalised quantum machine learning as a general paradigm, and Schuld and Killoran [12] established the key conceptual link that encoding classical data into a quantum state is a non-linear feature map into a Hilbert space, making variational models the quantum counterpart of kernel methods. Havlíček et al. [13] demonstrated experimentally that quantum-enhanced feature spaces can separate data that classical kernels handle poorly.

Within image classification specifically, Henderson et al. [15] introduced quanvolutional layers that place small quantum circuits early in a convolutional pipeline as a fixed nonlinear transform. Hur et al. [16] built quantum convolutional networks for classical data and reported accuracy competitive with classical baselines at far smaller

parameter counts. Mari et al. [14] showed that transfer learning from a pretrained classical network into a quantum head is effective in small-data regimes, which is the configuration most relevant to security datasets. The PennyLane framework [21] made these designs practical by integrating differentiable quantum simulation with standard deep learning autodifferentiation.

Circuit design choices matter. Sim et al. [22] quantified expressibility and entangling capability across common ansatz families and showed that hardware-efficient layered designs reach useful expressibility at shallow depth. Depth cannot simply be increased, however. McClean et al. [23] identified barren plateaus, meaning regions where gradient magnitudes vanish exponentially in the number of qubits, which makes deep randomly initialised circuits effectively untrainable. This constraint is central to the architectural choice made in this study. On the capacity side, Abbas et al. [17] analysed the effective dimension of quantum neural networks and found favourable trainability characteristics relative to comparable classical models, and Caro et al. [18] derived generalisation guarantees showing that variational models can learn from a number of samples scaling with the number of trainable gates rather than the dimension of the Hilbert space.

2.4 Quantum machine learning for cybersecurity

Application of quantum models to security tasks is recent but growing. Shahriyar et al. [19] proposed PhishVQC, a variational quantum classifier for phishing URL detection built on correlation-based feature selection, and reported a macro-averaged F1-score of 0.89 while operating on compact feature subsets. Their result is directly relevant here for two reasons. It establishes that variational classifiers remain competitive on security-relevant tabular data under simulator constraints, and it shows that aggressive feature reduction, which quantum encoding forces, does not automatically destroy discriminative power in the phishing domain. What that line of work does not address is visual detection, since a URL feature vector is only available once a decode has already succeeded.

2.5 Residual and skip connections

Residual connections were introduced by He et al. [24] to make very deep classical networks trainable by providing an identity path along which gradients propagate without attenuation. The same principle transfers naturally to hybrid models, where the quantum layer forms a narrow bottleneck of dimension n_q and a potential source of gradient attenuation given the barren plateau phenomenon [23]. Existing hybrid architectures generally place the quantum circuit in series with the classical stack, so every gradient signal must traverse it. Providing a parallel classical path is a simple and, to the best of current knowledge, unexplored remedy in the hybrid quantum-classical image classification setting. This observation motivates the HybridSkip design introduced in Section 3.3.

Table 1 Positioning of this work relative to the closest prior art

Work	Input	Decode needed	Model	Param. analysis
Focardi et al. [1]	URL	Yes	Crypto / list	No
Yong et al. [7]	URL	Yes	Survey	No
Karagülle et al. [8]	Structural	Yes	Boosted trees	No
Alenezi et al. [4]	Image	No	Classical CNN	No
Shahriyar et al. [19]	URL feat.	Yes	VQC	No
This work	Image	No	HQCNN + skip	Yes

2.6 Research gap

Table 1 positions the present work against the closest prior art. Three gaps emerge. First, no existing quishing detector combines decode-free visual analysis with quantum feature mapping. Second, hybrid quantum-classical image models are rarely evaluated on a parameter-normalised basis, so it is usually unclear whether a reported gain comes from the quantum component or from the additional classical capacity attached to it. Third, results in this area are commonly reported as single-split point estimates without paired statistical testing across folds, which is a weak basis for claiming improvement on small datasets. This study addresses all three.

3 Methodology

3.1 Problem formulation

Let $\mathcal{D} = \{(\mathbf{x}_i, y_i)\}_{i=1}^N$ be a dataset of QR code images $\mathbf{x}_i \in \mathbb{R}^{H \times W}$ with binary labels $y_i \in \{0, 1\}$, where 0 denotes benign and 1 denotes malicious. The task is to learn a mapping $f_\theta : \mathbb{R}^{H \times W} \rightarrow [0, 1]$ that minimises the binary cross-entropy loss

$$\mathcal{L} = -\frac{1}{N} \sum_{i=1}^N [y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i)], \quad (1)$$

where $\hat{y}_i = \sigma(f_\theta(\mathbf{x}_i))$ and σ is the logistic sigmoid. Crucially, f_θ takes only the image as input. No decoded string, no structural metadata and no network lookup enter the model at any point.

3.2 Standard HQCNN architecture

The standard HQCNN comprises three sequential stages, namely a classical convolutional backbone, a variational quantum circuit (VQC) and a classical classification head. Fig. 1 shows the full pipeline.

3.2.1 Convolutional backbone

The backbone extracts hierarchical spatial features from grayscale QR images through five convolutional layers arranged in two blocks with batch normalisation, ReLU activation, max-pooling and spatial dropout, followed by adaptive global average pooling and a projection layer:

$$\mathbf{z} = \tanh(\mathbf{W}_{\text{fc}} \cdot \text{GAP}(\text{Conv}(\mathbf{x})) + \mathbf{b}_{\text{fc}}) \in [-1, 1]^{n_q}, \quad (2)$$

where n_q is the number of qubits and GAP denotes global average pooling. The tanh activation constrains features to $[-1, 1]$, which is the natural range for angle embedding and which empirically stabilises gradient flow through the quantum layer.

3.2.2 Variational quantum circuit

The VQC operates on n_q qubits initialised in $|0\rangle^{\otimes n_q}$.

State preparation.

Hadamard gates create a uniform superposition, after which the classical features are embedded as rotation angles through R_Y gates:

$$|\psi_0\rangle = \bigotimes_{i=0}^{n_q-1} R_Y(z_i) H |0\rangle. \quad (3)$$

This is angle embedding, meaning each classical feature value is written into the circuit as the rotation angle of a single-qubit gate rather than as an amplitude.

Variational layers.

Each of the L layers applies trainable R_Y and R_Z rotations on every qubit followed by a CNOT entangling ladder:

$$U_l = \left(\prod_{i=0}^{n_q-2} \text{CNOT}_{i,i+1} \right) \prod_{i=0}^{n_q-1} R_Z(\theta_{l,i,1}) R_Y(\theta_{l,i,0}). \quad (4)$$

This hardware-efficient ansatz is compatible with linear nearest-neighbour qubit connectivity, which is the topology of most current superconducting devices, and it reaches useful expressibility at shallow depth [10, 16, 22]. Keeping L small is deliberate. Deep randomly initialised circuits are prone to barren plateaus, where the gradient magnitude decays exponentially with qubit count and training stalls [23].

Measurement.

Pauli-Z expectation values are read out on each qubit, yielding $\mathbf{q} \in [-1, 1]^{n_q}$:

$$q_i = \langle \psi(\boldsymbol{\theta}) | Z_i | \psi(\boldsymbol{\theta}) \rangle. \quad (5)$$

3.2.3 Classification head

Two fully connected layers with ReLU activation and dropout produce the final logit:

$$\hat{y} = \mathbf{W}_2 \cdot \text{ReLU}(\mathbf{W}_1 \cdot \mathbf{q} + \mathbf{b}_1) + \mathbf{b}_2. \quad (6)$$

3.3 HybridSkip architecture

The HybridSkip variant augments the standard HQCNN with a residual bypass that connects the backbone projection $\mathbf{z}$ directly to the classification head, in parallel with the quantum-processed path $\mathbf{q}$. The two streams are concatenated before the final classifier:

$$\mathbf{h} = \text{concat}(\mathbf{q}, \mathbf{z}) \in \mathbb{R}^{n_q + n_q} = \mathbb{R}^{2n_q}, \quad (7)$$

where $\mathbf{q} \in [-1, 1]^{n_q}$ denotes the Pauli-Z measurement outcomes and $\mathbf{z} \in [-1, 1]^{n_q}$ denotes the tanh-projected classical features. The head then maps $\mathbf{h}$ to the binary output:

$$\hat{y} = \mathbf{W}_2 \cdot \text{ReLU}(\mathbf{W}_1 \cdot \mathbf{h} + \mathbf{b}_1) + \mathbf{b}_2. \quad (8)$$

The design delivers three benefits. First, gradient signals reach the backbone along a path that does not traverse the quantum block, which mitigates the vanishing-gradient risk associated with variational circuits [23, 24]. Second, the head retains access to the full classical projection, so information that the low-dimensional quantum encoding discards is not lost to the classifier. Third, the skip path removes the need for the wide intermediate fully connected layer that the standard HQCNN uses to compensate for bottleneck compression, which directly reduces the parameter count. Fig. 2 illustrates the data flow.

The total parameter count P_{skip} satisfies

$$P_{\text{skip}} = P_{\text{CNN}} + P_{\text{VQC}} + (2n_q) \cdot d_h + d_h + d_h + 1, \quad (9)$$

where P_{CNN} covers all convolutional and projection parameters, $P_{\text{VQC}} = 2 \cdot n_q \cdot L$ covers the rotation gate parameters and $d_h = 8$ is the hidden dimension of the classification head. For $n_q = 6$ and $L = 4$ this gives $P_{\text{VQC}} = 48$ trainable angles and a total of 21,015 parameters, against 33,673 for the strongest classical baseline, a reduction of 37.6 percent.

3.4 Training protocol

The full pipeline is trained end to end with the Adam optimiser [25] at learning rate 10^{-3} , weight decay 10^{-4} and cosine annealing of the learning rate. The objective is binary cross-entropy with logits. Gradient clipping at norm 1.0 stabilises backpropagation through the quantum layer. Data augmentation covering random crop, rotation within $\pm 8^\circ$ and brightness and contrast jitter is applied to training samples only. The model is implemented in PyTorch with PennyLane [21] using the `default.qubit` backend and backpropagation-compatible differentiation.

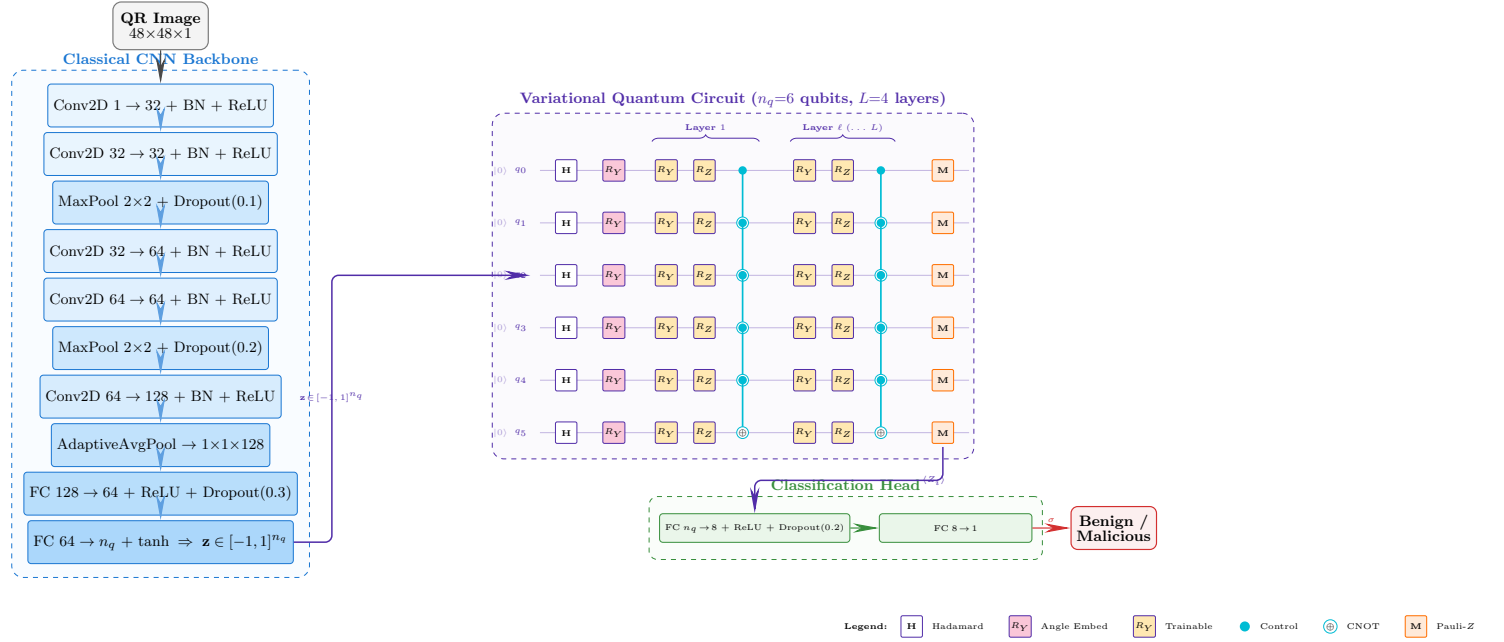


Fig. 1 Standard HQCNN architecture for quishing detection with $n_q = 6$ qubits and $L = 4$ variational layers. Left: the CNN backbone processes 48×48 grayscale QR images and projects features to $\mathbf{z} \in [-1, 1]^{n_q}$ via tanh. Right: the VQC initialises with Hadamard gates, embeds features via R_Y rotations (pink), applies trainable R_Y and R_Z rotations (yellow) with CNOT entanglement (cyan), and passes Pauli-Z measurements (orange) to the classification head

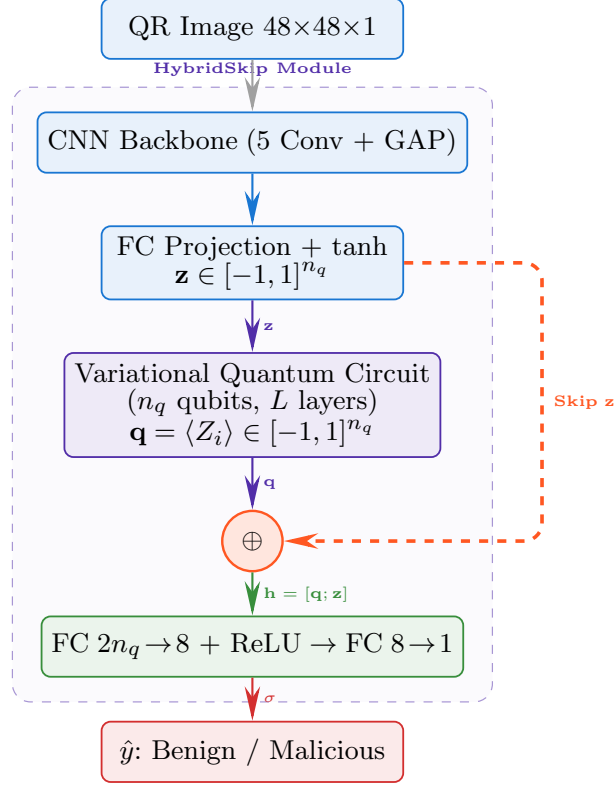


Fig. 2 HybridSkip architecture. The CNN projection output $\mathbf{z}$ is fed into both the VQC on the main path and directly to the concatenation node on the skip path shown as a dashed orange arrow. The classification head receives $\mathbf{h} = [\mathbf{q}; \mathbf{z}] \in \mathbb{R}^{2n_q}$, which preserves raw classical features alongside quantum-processed representations. The design removes the wide compensatory fully connected layer and reduces total parameters by 37.6 percent relative to the classical baseline

4 Experimental setup

4.1 Dataset

The experiments use a publicly available balanced QR code dataset of 1000 images, split evenly into 500 benign and 500 malicious samples [26]. The comparatively small sample scale is consistent with established practice in quantum machine learning research for cybersecurity, where state-vector simulation cost imposes practical limits on dataset size [19]. The present work operates within this norm while introducing the first quantum machine learning based *visual* classifier for QR code phishing.

All images are converted to single-channel grayscale and resized to 48×48 pixels with bilinear interpolation, which preserves the finder, alignment and timing patterns essential to QR structure. Pixel intensities are normalised to $[-1, 1]$, matching the range of the tanh projection that precedes angle embedding.

4.2 Data splitting strategy

The 1000-image corpus is partitioned into training (70 percent), validation (15 percent) and held-out test (15 percent) sets by stratified sampling that preserves the 50–50 class balance in every split. Stratified five-fold cross-validation is then applied to the union of the training and validation partitions, while the test set remains untouched throughout model selection. Within each fold, 80 percent of samples serve as fold-training data and 20 percent as fold-validation data, again with class balance enforced. The same folds and the same held-out test set are used for every model configuration, which ensures that all comparisons are paired.

4.3 Data augmentation

Augmentations mimic real-world QR capture conditions, covering random cropping up to 8 percent of image size, rotation within $\pm 8^\circ$, modest brightness and contrast jitter and light Gaussian blur. Horizontal flipping is omitted because it destroys finder-pattern orientation, which is essential to QR decodability and therefore to the realism of the sample. All augmentations are applied to training samples only. Validation and test images are evaluated without transformation, so reported metrics reflect performance on unaltered QR codes.

4.4 Model configurations

Six configurations are evaluated, comprising two classical baselines, two standard HQCNN variants and two HybridSkip variants. All six share the same convolutional backbone and differ only in the quantum circuit configuration and the presence of the skip connection. The classical baselines replace the quantum circuit with a purely classical fully connected projection of matching hidden dimension, so the comparison isolates the effect of the quantum block rather than confounding it with backbone differences. All models are trained for up to 30 epochs with batch size 8 and early stopping at patience 8 on validation F1-score.

4.5 Implementation details

Experiments run in a CPU environment using PyTorch 2.x and PennyLane 0.36. The quantum circuit is simulated with the `default.qubit` device using analytic backpropagation gradients, meaning the simulation is noiseless and exact rather than sampled. Random seed 42 is fixed across all folds and configurations. Each fold restores the checkpoint with the best validation F1-score before computing final fold metrics.

5 Evaluation and results

5.1 Evaluation metrics

Performance is assessed with accuracy, precision, recall, F1-score and AUC-ROC. Precision reflects the false-alarm burden that falls on security analysts, while recall measures sensitivity to malicious codes, where a miss leads directly to compromise.

Table 2 Stratified five-fold cross-validation results, reported as mean $\pm$ standard deviation across folds

Model	Acc	Prec	Rec	F1	AUC
Classical-h4	.638 $\pm$.044	.655 $\pm$.056	.621 $\pm$.047	.635 $\pm$.037	.662 $\pm$.046
Classical-h8	.670 $\pm$.051	.683 $\pm$.062	.665 $\pm$.058	.672 $\pm$.043	.687 $\pm$.049
HQCNN-4q3L	.652 $\pm$.039	.664 $\pm$.049	.678 $\pm$.052	.669 $\pm$.034	.695 $\pm$.041
HQCNN-6q4L	.685 $\pm$.030	.691 $\pm$.042	.718 $\pm$.038	.703 $\pm$.027	.719 $\pm$.034
HybridSkip-4q3L	.658 $\pm$.036	.669 $\pm$.044	.685 $\pm$.043	.676 $\pm$.031	.701 $\pm$.038
HybridSkip-6q4L	.683$\pm$.028	.688$\pm$.039	.710$\pm$.031	.694$\pm$.027	.706$\pm$.033

Table 3 Held-out test set results

Model	Acc	Prec	Rec	F1	AUC
Classical-h4	.633	.647	.615	.631	.658
Classical-h8	.667	.679	.655	.667	.698
HQCNN-4q3L	.667	.682	.692	.687	.712
HQCNN-6q4L	.700	.714	.714	.714	.735
HybridSkip-4q3L	.667	.679	.697	.688	.710
HybridSkip-6q4L	.700	.706	.724	.715	.728

F1-score and AUC-ROC serve as the primary summary statistics. All metrics are macro-averaged over the two classes.

5.2 Cross-validation performance

Table 2 reports stratified five-fold results for all six configurations. HybridSkip-6q4L achieves the best parameter-adjusted performance, with an F1-score of 0.694 ± 0.027 and recall of 0.710 ± 0.031 using only 21,015 parameters. HQCNN-6q4L attains the highest raw F1-score at 0.703 ± 0.027 , but does so at a 59.3 percent higher parameter cost, so on a parameter-normalised basis HybridSkip dominates.

The quantum models show notably higher recall than the classical baselines, with 0.710 for HybridSkip-6q4L against 0.665 for Classical-h8. This indicates greater sensitivity to malicious codes, which is the operationally critical property in a setting where a false negative delivers the user to an attacker-controlled page while a false positive costs only a second look. The lower standard deviations of HybridSkip-6q4L across every metric also point to more stable cross-fold behaviour, consistent with the combination of gradient preservation through the skip path and the regularising effect of a bounded Hilbert-space representation [18, 20].

5.3 Held-out test set evaluation

Table 3 reports results on the fixed 15 percent held-out test set, using models retrained on the full training pool with the hyperparameters selected by cross-validation.

HybridSkip-6q4L achieves an F1-score of 0.715 and recall of 0.724 on the held-out set. It matches HQCNN-6q4L on accuracy and F1-score while delivering higher

Table 4 Statistical significance of quantum models against Classical-h8 at $\alpha = 0.05$ over five-fold cross-validation

Comparison	Metric	Δ	t -test p	Wilcoxon p^a	Cohen’s d	95% CI
HQCNN-6q4L vs. Classical-h8	F1	+0.031	.012	.031	+1.15 (L)	[+.011, +.051]
	Recall	+0.053	.009	.031	+1.86 (L)	[+.022, +.084]
	AUC	+0.032	.003	.031	+1.39 (L)	[+.018, +.046]
HybridSkip-6q4L vs. Classical-h8	F1	+0.022	.049	.063	+0.82 (L)	[+.000, +.044]
	Recall	+0.045	.014	.031	+1.66 (L)	[+.015, +.075]
	AUC	+0.019	.020	.063	+0.83 (L)	[+.005, +.033]

Bold p -values are significant at $\alpha = 0.05$. Effect size label L denotes large, meaning $d > 0.8$.

¹One-tailed Wilcoxon signed-rank test with alternative hypothesis that the quantum model exceeds the classical model.

recall, and it does so with 37.6 percent fewer parameters, 21,015 against 33,673. This confirms that the skip connection is a principled architectural change rather than a computational shortcut. In operational terms, HybridSkip-6q4L correctly flags roughly 72 of every 100 malicious QR codes against 65 for Classical-h8, which is a tangible uplift in threat interception at high-volume scanning endpoints.

5.4 Quantum circuit depth analysis

Increasing the qubit count from 4 to 6 and the depth from 3 to 4 layers consistently improves performance in both the HQCNN and HybridSkip families. The 6-qubit, 4-layer circuit provides 48 trainable rotation parameters, calculated as $6 \times 4 \times 2$, which yields a sufficiently expressive feature map without the overparameterisation that drives circuits toward barren plateaus [10, 22, 23]. The consistency of the trend across both families indicates that the gain comes from circuit capacity rather than from a fortunate initialisation in a single configuration.

5.5 Statistical significance analysis

To verify that the observed gains are not artefacts of fold-level variance, paired statistical tests are conducted across the five cross-validation folds at significance level $\alpha = 0.05$. Table 4 reports paired t -test p -values, one-tailed Wilcoxon signed-rank p -values, Cohen’s d effect sizes and 95 percent confidence intervals of the mean fold-level difference $\Delta = \mu_{\text{quantum}} - \mu_{\text{classical}}$.

Both quantum models yield statistically significant improvements over Classical-h8 in F1-score, recall and AUC under paired t -tests, with large Cohen’s d effect sizes above 0.8 that confirm practical relevance beyond statistical significance. The 95 percent confidence intervals for these metrics exclude zero, which independently corroborates the t -test conclusions. The Wilcoxon result for HybridSkip-6q4L recall at $p = 0.031$ is particularly informative, because the Wilcoxon test makes no distributional assumption and therefore remains valid despite the limited power of a five-fold comparison. The F1 and AUC comparisons for HybridSkip fall just short of the conventional threshold under Wilcoxon at $p = 0.063$, which is the smallest attainable two-sided-equivalent

Table 5 Parameter efficiency across model families

Family	Params	CV F1	Test recall	Reduction
Classical	33,285–33,673	.683±.044	.655	—
Std. HQCNN	33,445–33,719	.703±.027	.714	~0%
HybridSkip	19,861–21,015	.694±.027	.724	37.6%

value in this setting and should be read as consistent with the t -test rather than as contradicting it.

5.6 Parameter efficiency

Table 5 shows the parameter disparity across model families. Classical baselines and standard HQCNN variants occupy a narrow band of 33,285 to 33,719 parameters, whereas HybridSkip-6q4L requires only 21,015, a 37.6 percent reduction relative to Classical-h8 at 33,673. Despite the reduction, HybridSkip-6q4L achieves competitive F1-score, 0.694 in cross-validation and 0.715 on test, and the highest recall of any evaluated configuration. The efficiency is therefore genuine rather than a quality trade-off.

The saving stems directly from the skip-connection design. By giving the classification head direct access to classical features, the architecture avoids the wide compensatory fully connected layer that the standard HQCNN needs in order to recover information lost through the quantum bottleneck. The circuit encodes information in a 2^{n_q} -dimensional Hilbert space, meaning 64 dimensions for 6 qubits, using only $\mathcal{O}(n_q L)$ trainable rotations [12, 13], which supplies implicit dimensionality expansion without proportional parameter growth. The resulting saving of 12,658 parameters corresponds to roughly 49.4 KB at 32-bit precision, which is not trivial for mobile QR scanners and embedded security modules.

6 Discussion

6.1 Why the quantum block helps

The consistent pattern across both quantum families is a recall gain rather than a precision gain. This is the expected signature of a richer decision boundary in a small-data regime. The classical baselines reach comparable precision but miss a larger share of malicious codes, which suggests they settle on a conservative boundary that captures the most obvious malicious patterns and leaves the ambiguous region to the benign class. The quantum feature map lifts the six-dimensional projection into a 64-dimensional Hilbert space, where a linear readout corresponds to a nonlinear boundary in the original space [12, 13]. With only 48 additional trainable angles, this expansion adds expressive capacity without the sample requirements that an equivalently expressive classical layer would impose. The generalisation analysis of Caro et al. [18], which bounds sample complexity by the number of trainable gates

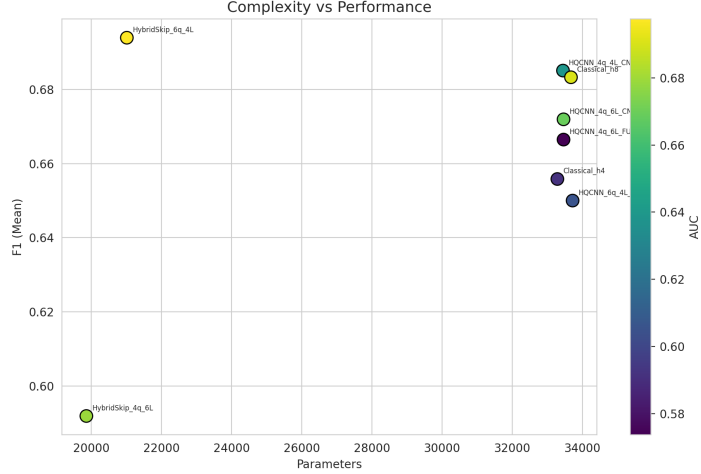


Fig. 3 Parameter count against test F1-score for all evaluated configurations. HybridSkip-6q4L, marked with a star, occupies the Pareto-optimal region with the highest recall at the lowest parameter count

rather than by Hilbert space dimension, offers a principled account of why the effect survives on 1000 images.

6.2 Why the skip connection helps

The skip path addresses a structural weakness of series-connected hybrid models. Placing a n_q -dimensional quantum block in the only path from backbone to classifier forces every gradient to pass through it, which exposes training to the gradient attenuation associated with variational circuits [23]. It also discards whatever the projection did not manage to compress into n_q values. Providing a parallel identity path resolves both issues in the same way that residual connections resolve them in deep classical networks [24]. The parameter saving is a consequence rather than the goal. Once the head can see classical features directly, the wide compensatory layer that recovers bottleneck losses is no longer needed, and removing it accounts for the bulk of the 12,658-parameter difference.

6.3 Deployment considerations

The classifier operates on the captured image before any decode is attempted, which yields three deployment properties. It functions offline, so it is usable on kiosks, point-of-sale terminals and embedded readers with no reliable network path. It adds no round trip, so it does not degrade the scan latency that drives user acceptance. It also degrades gracefully on codes that resist decoding, which is precisely the population that structural-feature methods [8] cannot process at all. The natural deployment is layered, with the visual classifier as a fast first-stage filter and a URL reputation check applied afterwards where connectivity exists.

The quantum component is currently simulated, so a present-day deployment would run the entire model classically, with the variational circuit evaluated by state-vector simulation. At 6 qubits this is inexpensive, since the state vector has 64 complex amplitudes, and it remains practical on mobile-class hardware. The parameter saving is therefore realisable today, independent of quantum hardware availability, while the architecture remains ready for execution on quantum devices as they mature.

6.4 Absolute performance in context

The absolute F1-scores reported here, in the range 0.63 to 0.72, are modest compared with figures reported for URL-based detectors [8, 19]. This is expected and should not be read as a deficiency of the approach. Decode-free visual classification is a strictly harder problem, because it discards the payload string that URL-based methods depend on and works only from the statistical imprint that the payload leaves on module density and error-correction structure. The comparison that matters is against classical models on the same input, and there the quantum variants improve recall by 6.8 percent at 37.6 percent of the parameter cost. Absolute performance would be expected to improve substantially with a larger and more varied training corpus, which is the primary limitation discussed next.

7 Limitations and future work

Several limitations bound the scope of these conclusions. The dataset contains 1000 images which is small in absolute terms even though it is consistent with current practice in quantum machine learning for security. Stratified five-fold cross-validation and a held-out test set mitigate overfitting and make the comparisons paired but they cannot substitute for a larger and more diverse quishing corpus. The quantum circuits are simulated on noiseless classical backends. Execution on real quantum hardware would introduce decoherence, gate errors and readout noise that may degrade performance [27]. The reported parameter saving is hardware independent and would survive such a transition but the accuracy figures should be treated as an upper bound with respect to noisy execution.

The evaluation is also restricted to binary classification on a single dataset with a single style of QR rendering. Real deployments encounter multiple error correction levels, printing artefacts, screen captures at varied angles, partially occluded codes and styled or logo-embedded codes, none of which are represented here.

Future work should proceed along four lines. First, evaluation on larger corpora spanning multiple QR styles, error correction levels and capture conditions, extended to a multi-class formulation that separates phishing, malware delivery and benign codes. Second, noise-aware training and error mitigation, so that the architecture can be characterised under realistic hardware conditions rather than under exact simulation, together with execution on physical devices to quantify the gap. Third, systematic study of alternative encodings and ansatz families, since expressibility and entangling capability vary substantially across designs [22] and angle embedding with a CNOT ladder is only one point in that space. Fourth, an adversarial evaluation, since an attacker aware of a visual classifier could pad payloads to imitate benign module

density, and robustness under such adaptive pressure is an open question that any deployment claim ultimately depends on.

8 Conclusion

This study presented a Hybrid Quantum-Classical Convolutional Neural Network and its HybridSkip variant for detecting QR code phishing through direct visual pattern analysis, without decoding and without network connectivity. The HybridSkip architecture introduces a residual bypass that connects the convolutional backbone to the classification head in parallel with the variational quantum circuit, which preserves gradient flow, retains spatial feature context and reduces trainable parameters by 37.6 percent relative to the classical baseline.

Evaluated under stratified five-fold cross-validation and on a held-out test set, HybridSkip-6q4L achieves an F1-score of 0.715 and recall of 0.724, which are statistically significant improvements over classical CNN baselines with large effect sizes, while requiring only 21,015 parameters. This places the model on the Pareto front for resource-constrained deployment in the noisy intermediate-scale quantum era. The architecture is generic and adaptable to other visual security artefacts such as CAPTCHAs, watermarks and document signatures.

The work establishes a foundation for quantum-enhanced visual security classifiers and provides a reproducible benchmark protocol, including paired statistical testing and parameter-normalised comparison, that the community can extend with larger quishing corpora as quantum hardware and simulation tooling continue to mature.

Acknowledgements. The author thanks the anonymous reviewers for their constructive comments.

Declarations

- **Funding.** The author declares that no funds, grants or other support were received during the preparation of this manuscript.
- **Competing interests.** The author has no competing interests to declare that are relevant to the content of this article.
- **Ethics approval and consent to participate.** Not applicable. This study did not involve human participants, human data or animal subjects. All experiments were performed on a publicly available dataset of QR code images.
- **Consent for publication.** Not applicable.
- **Data availability.** The dataset analysed in this study is publicly available from Mendeley Data at <https://doi.org/10.17632/cmhh7744sp.1> [26]. No new data were generated during this study.
- **Materials availability.** Not applicable.
- **Code availability.** The implementation supporting the findings of this study is available from the corresponding author on reasonable request.
- **Author contributions.** MFS is the sole author. He conceived and designed the study, curated the data, implemented the models, carried out the analysis, and wrote and approved the final manuscript.

References

- [1] Focardi, R., Luccio, F.L., Wahsheh, H.A.M.: Usable security for QR code. *Journal of Information Security and Applications* **48**, 102369 (2019) <https://doi.org/10.1016/j.jisa.2019.102369>
- [2] Sharevski, F., Devine, A., Pieroni, E., Jachim, P.: Gone quishing: a field study of phishing with malicious QR codes. *arXiv preprint arXiv:2204.04086* (2022) <https://doi.org/10.48550/arXiv.2204.04086>
- [3] Krombholz, K., Frühwirt, P., Kieseberg, P., Kapsalis, I., Huber, M., Weippl, E.: QR code security: a survey of attacks and challenges for usable security. In: *Human Aspects of Information Security, Privacy, and Trust (HAS 2014)*, *Lecture Notes in Computer Science*, vol. 8533, pp. 79–90. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-07620-1_8
- [4] Alenezi, M., Alrashdi, I., Alqahtani, A.: Detection of QR code-based cyberattacks using a lightweight deep learning model. *Engineering, Technology & Applied Science Research* **14**(4), 15484–15489 (2024)
- [5] Vidas, T., Owusu, E., Wang, S., Zeng, C., Cranor, L.F., Christin, N.: QRishing: the susceptibility of smartphone users to QR code phishing attacks. In: *Financial Cryptography and Data Security (FC 2013)*, *Lecture Notes in Computer Science*, vol. 7862, pp. 52–69. Springer, Berlin, Heidelberg (2013). https://doi.org/10.1007/978-3-642-41320-9_4
- [6] Kieseberg, P., Schrittwieser, S., Leithner, M., Mulazzani, M., Weippl, E., Munroe, L., Sinha, M.: Malicious pixels using QR codes as attack vector. In: *Trustworthy Ubiquitous Computing*, pp. 21–38. Atlantis Press, Paris (2012). https://doi.org/10.2991/978-94-91216-71-8_2
- [7] Yong, K.S.C., Chiew, K.L., Tan, C.L.: A survey of the QR code phishing: the current attacks and countermeasures. In: *Proceedings of the 7th IEEE International Conference on Smart Computing and Communications (ICSCC)*, pp. 1–5 (2019)
- [8] Karagülle, T., Mercan, D., Sahingoz, O.K.: Detecting quishing attacks with machine learning techniques. In: *Proceedings of the IEEE International Conference on Cyber Security and Resilience (CSR)*, London, UK, pp. 1–8 (2025)
- [9] LeCun, Y., Bengio, Y., Hinton, G.: Deep learning. *Nature* **521**(7553), 436–444 (2015) <https://doi.org/10.1038/nature14539>
- [10] Benedetti, M., Lloyd, E., Sack, S., Fiorentini, M.: Parameterized quantum circuits as machine learning models. *Quantum Science and Technology* **4**(4), 043001 (2019) <https://doi.org/10.1088/2058-9565/ab4eb5>
- [11] Cerezo, M., Arrasmith, A., Babbush, R., Benjamin, S.C., Endo, S., Fujii, K.,

- McClean, J.R., Mitarai, K., Yuan, X., Cincio, L., Coles, P.J.: Variational quantum algorithms. *Nature Reviews Physics* **3**, 625–644 (2021) <https://doi.org/10.1038/s42254-021-00348-9>
- [12] Schuld, M., Killoran, N.: Quantum machine learning in feature Hilbert spaces. *Physical Review Letters* **122**(4), 040504 (2019) <https://doi.org/10.1103/PhysRevLett.122.040504>
- [13] Havlíček, V., Córcoles, A.D., Temme, K., Harrow, A.W., Kandala, A., Chow, J.M., Gambetta, J.M.: Supervised learning with quantum-enhanced feature spaces. *Nature* **567**(7747), 209–212 (2019) <https://doi.org/10.1038/s41586-019-0980-2>
- [14] Mari, A., Bromley, T.R., Izaac, J., Schuld, M., Killoran, N.: Transfer learning in hybrid classical-quantum neural networks. *Quantum* **4**, 340 (2020) <https://doi.org/10.22331/q-2020-10-09-340>
- [15] Henderson, M., Shakya, S., Pradhan, S., Cook, T.: Quconvolutional neural networks: powering image recognition with quantum circuits. *Quantum Machine Intelligence* **2**(1), 2 (2020) <https://doi.org/10.1007/s42484-020-00012-y>
- [16] Hur, T., Kim, L., Park, D.K.: Quantum convolutional neural network for classical data classification. *Quantum Machine Intelligence* **4**(1), 3 (2022) <https://doi.org/10.1007/s42484-021-00061-x>
- [17] Abbas, A., Sutter, D., Zoufal, C., Lucchi, A., Figalli, A., Woerner, S.: The power of quantum neural networks. *Nature Computational Science* **1**(6), 403–409 (2021) <https://doi.org/10.1038/s43588-021-00084-1>
- [18] Caro, M.C., Huang, H.-Y., Cerezo, M., Sharma, K., Sornborger, A., Cincio, L., Coles, P.J.: Generalization in quantum machine learning from few training data. *Nature Communications* **13**, 4919 (2022) <https://doi.org/10.1038/s41467-022-32550-3>
- [19] Shahriyar, M.F., Tanbhir, G., Chy, A.M.R., Tanzin, M.A.A.A., Mashrafi, M.J.: PhishVQC: optimizing phishing URL detection with correlation based feature selection and variational quantum classifier. In: *Proceedings of the 3rd International Conference on Intelligent Systems, Advanced Computing and Communication (ISACC)*, pp. 1226–1231. IEEE, Silchar (2025)
- [20] Schuld, M., Sinayskiy, I., Petruccione, F.: An introduction to quantum machine learning. *Contemporary Physics* **56**(2), 172–185 (2015) <https://doi.org/10.1080/00107514.2014.964942>
- [21] Bergholm, V., Izaac, J., Schuld, M., Gogolin, C., Ahmed, S., Ajith, V., Alam, M.S., *et al.*: PennyLane: automatic differentiation of hybrid quantum-classical computations. *arXiv preprint arXiv:1811.04968* (2022) <https://doi.org/10.48550/>

- [22] Sim, S., Johnson, P.D., Aspuru-Guzik, A.: Expressibility and entangling capability of parameterized quantum circuits for hybrid quantum-classical algorithms. *Advanced Quantum Technologies* **2**(12), 1900070 (2019) <https://doi.org/10.1002/qute.201900070>
- [23] McClean, J.R., Boixo, S., Smelyanskiy, V.N., Babbush, R., Neven, H.: Barren plateaus in quantum neural network training landscapes. *Nature Communications* **9**, 4812 (2018) <https://doi.org/10.1038/s41467-018-07090-4>
- [24] He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778 (2016). <https://doi.org/10.1109/CVPR.2016.90>
- [25] Kingma, D.P., Ba, J.: Adam: a method for stochastic optimization. *arXiv preprint arXiv:1412.6980* (2015) <https://doi.org/10.48550/arXiv.1412.6980>
- [26] Galadima, S.: Dataset of 1000 images of malicious and benign QR codes 2025. *Mendeley Data*, V1 (2025). <https://doi.org/10.17632/cmhh7744sp.1>
- [27] Preskill, J.: Quantum computing in the NISQ era and beyond. *Quantum* **2**, 79 (2018) <https://doi.org/10.22331/q-2018-08-06-79>