



QShield-NS: A Variational Quantum Machine Learning Model for Zero-Day Cyber Threat Detection in National Security Systems

Md Sajedul Karim Chy¹, Salman Mohammad Abdullah², Mahbub Ahmed Nabil³, Abidul Alam⁴,

Md Himeluzzaman⁵, Tofayel Ahmed Onik⁶, Shaown Mahamud Shakil⁷, Hafiz Aziz Khan⁸

M.S. in Project Management, Washington University of Science and Technology, VA, USA¹

M.S. in Information Technology (IT), Washington University of Science and Technology, VA, USA²

Master of Science in Information Technology, Washington University of Science and Technology, VA, USA³

Bachelor of Business Administration (BBA), International University of Business, Agriculture and Technology, Bangladesh⁴

M.Sc. in Information Technology Systems and Management, Washington University of Science and Technology, VA, USA⁵

Bachelor of Science in Electrical & Electronics Engineering, Metropolitan University, Sylhet, Bangladesh⁶

Bachelor in Human Resource Management, Independent University, Bangladesh⁷

Master of Science in Information Technology, Washington University of Science and Technology, USA⁸

Sajedulkarimchy661@gmail.com; farsisalman310@gmail.com; ahmednabilnsucse@gmail.com;

Abidulalam633@gmail.com; Mzzaman1995@gmail.com; tofayelahmedonik@gmail.com;

shawon95mahmud@gmail.com; hakhan.student@wust.edu

ABSTRACT: Machine-learning intrusion detection systems are regularly benchmarked against attack classes that also exist in their training data, a practice that inflates preparedness against the new types of threats most pertinent to national-security networks. We introduce QShield-NS, a variance-reduced quantum machine learning classifier capable of zero-day threat detection, and demonstrate its ability via strict Leave-One-Attack-Out (LOAO) evaluation on the UNSW-NB15 dataset, where each of nine different attack families are omitted from training in turn. Eight principal components are angle-encoded on eight qubits, processed by strongly entangling variational layers and read through Pauli-Z expectation values into a compact classical head that is trained end-to-end. QShield-NS is compared against a multilayer perceptron, a random forest and XGBoost on the same folds, with the same features and decision threshold. A rank inversion is what the results reveal. If however the training is known-attack control split QShield-NS ranks last (macro-F1 0.8793 v s 0.9158 random forest), and under the zero-day protocol first (0.7303 against 0.7044). Its generalisation gap, at 0.1489 to 0.2084–0.2302, is the smallest for the four models and it lowered the mean false-positive rate from 20.24 percent to 10.72 percent, scoring precision on eight of nine folds. It is a calibration, not discrimination advantage: the classical baselines get better ROC-AUC and PR-AUC in every fold but achieve distribution shift recall at the lower price of over-predicting the attack class. A depth ablation and a study of gradient-variance establish robust trainability up to 8 qubits. We do not claim any quantum advantage; all results were obtained by classical simulation.

KEYWORDS: variational quantum circuits; quantum machine learning; zero-day attack detection; intrusion detection systems; UNSW-NB15; leave-one-attack-out evaluation; barren plateaus; national security; NISQ



I. INTRODUCTION

Most evaluations that have been published lack the adversarial asymmetry which prevails when network intrusion detection systems (NIDSs) are deployed to protect government, defense and critical-infrastructure networks. A defender must generalize, and an attacker needs only to find a single behavior not catalogued by the defender. This is definitionally blind to novel exploits (signature-based detection), and even the machine-learning approaches have been introduced specifically in order to close this gap, but through a more dominant evaluation methodology, the authors reintroduce it. Since models were trained and tested on random partitions of the same corpus, every attack category at test time was also present during training. Such an experiment measures interpolation in a static taxonomy, not the ability to identify something truly novel.

Sommer and Paxson expressed this concern nearly a decade ago, noting that the aspect of anomaly detection that is primarily atypical among machine-learning applications is that its operational cost in terms of a false positive can be significantly severe, whilst also observing that the feature to be detected has an absence from the training distribution [4]. Later surveys have repeatedly observed that accuracy scores reported on standard intrusion benchmarks group unreasonably close to 1 while deployments in the field remain overwhelmed by false alerts [5,6,48]. The difference is not so much that we model poorly; it is to evaluate poorly. Kasongo and Sun use the same benchmark (which is described here) and clearly note that machine-learning intrusion detectors experience high false-positive rates, and decreased accuracy when the training distribution becomes imbalanced or shifted [9].

However, the national security context strengthens the stakes in both directions. First, the adversaries of interest are highly capable and intentionally using methods not present in publicly available threat corpora, so again the zero-day paradigm is actually the operating regime as opposed to an outlier. Second, an alert queue that cannot be triaged by a human analyst functions as if no detection were taking place at all: flag one out of five benign flows and our detector will be tuned out or turned off. Therefore, detection rate is a poor figure of merit by itself, and an acceptable evaluation should provide the burden in false positives incurred to achieve it.

At the same time, quantum computing shifted from a concern for SEC to a new source of modelling primitives. The long-term threat of fault-tolerant quantum machines for deployed public-key cryptography was established by Shor's and Grover's algorithms [18,19], motivating a decade of post-quantum standardisation activity [22,23], along with a wider realization that security architectures need to be designed across a quantum transition lifetime [21]. Compared to the more negative side, it should be mentioned: The most constructive way is less represented in the security literature. In this NISQ era [20], we have created a huge family of variational quantum circuits (VQC), which behave as parametrized differentiable machine-learning models [24,25,28,29,30,31,32]. These circuits embed classical inputs into a high-dimensional Hilbert space using a parameterized feature map [26,27] and optimise a shallow entangling ansatz in the spirit of gradient descent, where the quantum device is treated as a structured hard-to-simulate functional class rather than an accelerator.

For the security practitioner, the relevant question is not whether such a circuit runs faster (on current hardware and in simulation it still does not) but whether its inductive bias differs usefully from that of a decision tree ensemble, or of a fully connected layer network. Theoretical work suggests it may. For example, analyses of the expressivity and effective dimension of quantum neural networks suggest that variational models induce classes of functions with significantly different geometries than their classical counterparts [37], while studies of data encoding show that the feature map largely governs what family of functions the circuit can represent [39]. Both results suggest that a VQC generalizes differently perhaps not better in-distribution, but differently under distribution shift. Distribution shift is the only object of interest under a zero-day protocol.

This hypothesis has not yet been tested in existing applications of quantum machine learning to intrusion detection. Gouveia and Correia proposed an early quantum-assisted NIDS presenting simulation results on two datasets, NSL-KDD and UNSW-NB15, achieving near SVM [43] accuracy. Payares and Martinez-Santos also compared quantum support vector machines, hybrid quantum-classical networks and a two-circuit ensemble on distributed denial-of-service detection, claiming accuracies near unity [44]. The clustering-based intrusion detection has also been attended to by quantum-inspired optimisation [45]. These studies are mainly, however, valuable as feasibility demonstrations (where all evaluate on random splits in which attack categories overlap between training and test partitions). As a result, they are measuring the same interpolation quantity as the classical literature from which they draw comparisons, and their high reported accuracies suffer from the same caveat.



This paper addresses that gap. We treat zero-day detection as a Leave-One-Attack-Out (LOAO) generalization setting in which UNSW-NB15 data [1, 2] is used to derive our empirical results across three classical baselines under both the same splits, the same preprocessing, and with the same decision thresholds per class between QShield-NS, a hybrid variational quantum classifier. Our contributions are as follows.

We first design and develop a leakage-controlled LOAO protocol for zero-day evaluation on UNSW-NB15, along with an adjacent known-attack control split, so that novelty-induced degradation may be decoupled from baseline model performance. To the best of our knowledge, this is the first application of any such protocol to quantum machine learning (QML) for intrusion detection.

Second, we find a rank inversion between the two regimes. Among the four models, QShield-NS is currently the weakest on known attacks but also the strongest on unseen attacks (in terms of macro-F1, accuracy, and F1), while degrading by 30 to ~40% less than classical baselines when shifting from control to zero-day.

Finally, we describe the mechanism that leads to this inversion. While the variational model is not a better ranker in a threshold-agnostic sense: it lags behind on ROC-AUC and PR-AUC over all nine folds; but it does produce a much more conservative and better-calibrated decision boundary under distribution shift. The false-positive rate has been halved relative to MLP, reduced by about one-third relative to random forest.

Fourth, we perform a trainability analysis. Across the considered depths at eight qubits, a depth ablation over $L \in \{1, 2, 4, 6\}$ and direct measurement of gradient variance indicate stable non-vanishing gradients with no evidence of barren-plateau onset $\lambda(L)$ on this regime [33,34,35].

The rest of the paper is organised as follows. In Section 2, we provide a survey of intrusion detection using machine learning, the UNSW-NB15 benchmark, and variational quantum machine learning. Section 3 presents the dataset and the preprocessing procedure, the LOAO (Local-Attention Object-based Approach) Protocol, the QShield-NS architecture, and model configuration. We present control, zero-day, per-family, calibration, depth, and gradient-variance in Section 4. Section 5 presents rank inversion; then we show its operating background and limitations of the study. Section 6 concludes.

II. RELATED WORK

2.1 Machine learning for network intrusion detection

Machine learning-based approach in studying anomaly-based intrusion detection for over two decades and accompanies Chandola et al. Chan et al., Buczak and Guven [5], Khraisat et al. Mitchell and Chen [47] p. 1344; Ahmad et al. [48] describes a continuous scale across this progression from statistical outlier tests through classical supervised learners to deep architectures. Vinayakumar et al. Deep neural networks were evaluated on a number of intrusion corpora and have demonstrated solid in-distribution performance [8]; ensemble methods using statistical flow features achieved similar gains for constrained network environments over the same kind of flows [10].

A recurrent methodological critique of this literature revolves around evaluation, rather than modelling. The semantic from the statistical outlier to an actionable security event differs for intrusions detection than typical classification tasks, the high cost of false positives defines training [4] and so does the define of a target class as novel. Tavallae et al. previously showed redundancy in the KDD Cup 99 corpus exaggerated reported accuracies and biased learners toward majority records [3]. Newer datasets only hope to fully address each critique one at time: as long as the evaluation partitions a fixed taxonomy randomly, the model is not forced to recognise a class it has never n before.

Previous works tackle the novelty indirectly via (one-class or reconstruction-based) formulations, training exclusively on benign traffic and consequent deviations are reported. These approaches naturally cope with novelty but discard the discriminative information leveraged in labelled attacks, and are mostly tested on identical mixed test sets. Here, the LOAO formulation used is between these two extremes: The model still gets supervised from eight known families while being tested on a family that is completely unseen during training time, which realistically simulates a situation in which a defender with robust historical threat intelligence is facing an attacker with an entirely new technique.



2.2 The UNSW-NB15 benchmark

To further overcome limitations of the KDD Cup 99 and NSL-KDD, such as their age, synthetic traffic properties and unrealistic attack distributions [1], Moustafa and Slay have proposed a new benchmark dataset: UNSW-NB15. This dataset contains the forty-nine flow, content and time-based features along with a nine-category attack taxonomy, which was generated at the Cyber Range Lab of the Australian Centre for Cyber Security using IXIA PerfectStorm traffic generator whose attack behaviours were based on contemporary sources of vulnerabilities. Furthermore, in a parallel study it was shown that the statistical properties of UNSW-NB15 were significantly more complex than KDD99 and classifiers with near perfect classification accuracy on the former can deteriorate substantially on the latter [2].

The benchmark has since become a common reference point. Kasongo and Sun also used filter-based feature selection with tree ensembles and provided the accuracy values along with a caveat about false-positive inflation under class imbalance [9]. Support vector machines were evaluated by Jing and Chen on the corpus [49], Kumar et al. gave a statistical characterization of its feature distributions [50]. Binary-classification accuracies reported in the rest of this literature are generally between the mid-eighties and mid-nineties, which is consistent with the control results in Section 4 and provides a sanity check on our classical baselines. Importantly, none of these studies break out data by attack family, meaning their figures reflect the known-attack regime only.

UNSW-NB15 being nine-family in its structure makes it applicable to the current research. Excuse us while we scream! In terms of size Families have dramatically different population sizes, from Generic with tens of thousands of records to Worms with under two hundred, or in behavioural signature between high-volume denial-of-service floods to low-footprint shellcode delivery. Such heterogeneity allows for the generation of nine qualitatively different zero-day scenarios from a single corpus, one of which is an rare low-volume family that operationally matters.

2.3 Variational quantum machine learning

Consider, for example, early quantum machine learning proposals that required fault-tolerant hardware and coherent access to quantum memory, achieving speedups asymptotic with the problem size on primitives such as principal component analysis [41] and support vector classification [42]; these results have been reviewed in depth by Biamonte et al. [24] and Schuld et al. [25]. However, we can't execute these algorithms on today's devices. Instead, the NISQ era is defined by several tens to hundreds of noisy qubits without error correction; this gives rise to a model dominated by so-called variational algorithms [20,32]. A vqc consists of a data-encoding stage that maps classical inputs to quantum states, a parameterised unitary ansatz, and a measurement stage whose expectation values make up the model output. The circuit can thus be optimised via classical gradient methods and embedded inside a neural network layer as the underlying circuit is differentiable w.r.t its parameters [29,31]

The interest in these models for classification has two lines of theory. The feature-map perspective (Havlíček et al. As previously explored by both Guan et al. Circuit-centric classifiers makes this ansatz formal by treating it as a trainable linear map in that space [28]. Schuld et al. take the complementary point of view, analyzing the resulting function class directly. found that the Fourier spectrum a circuit can easily represent is determined by the nature of encoding [39], and Abbas et al. showed that quantum neural networks can have larger effective dimension and more advantageous optimisation geometry than similarly sized classical networks [37]. Huang et al. have cooled these expectations, showing that access to data can enable classical models to approximate quantum ones on a wide variety of tasks and that true separation may require an intricate construction of the problem [38].

Para: The barren plateau is the dominating challenge in scaling variational models McClean et al. showed that for sufficiently expressive random circuits the variance of the gradient decays exponentially in n such that the optimisation landscape is actually flat [33] Cerezo et al. Wang et al. [38] recently provided evidence that the effect is dependent on locality of the cost function, showing that local observables maintain trainability at shallow depth where global observables do not [34]. proves hardware noise produces plateaus, regardless of circuit expressivity [35]. Mitigation strategies are based on approaches such as structured parameter initialisation [36] or intentional ansatz depth truncation. These findings inspire the empirical gradient-variance experiment in Section 4.6; we operate within a preserve-trainability regime (eight qubits, local Pauli-Z observables, depth at most six), and confirm the same through measurement.



2.4 Quantum machine learning for intrusion detection

There are relatively few applications of quantum learning to network security. First, Gouveia and Correia proposed a quantum-assisted NIDS which merged an autoencoder and a quantum kernel estimator that was assessed in simulation on NSL-KDD [42] and UNSW-NB15, achieving accuracies similar to those observable with classical SVMs [43]. However, Payares and Martinez-Santos compared three quantum models for denial-of-service detection on their partition without any partition [44], obtaining accuracies near to the unity. Chen et al. used hybrid k-means clustering for intrusion detection with the help of quantum-inspired optimisation [45]. In addition to security, variational circuits have seen applications in reinforcement learning [46] and multiple supervised tasks [28, 30, 31].

Shared across the security-focused part of these works are three limitations. As evaluation is performed on different random splits using the same attack categories, novelty is never evaluated. Most comparison works against the classical baseline rather than a benchmark suite Trainability is typically not investigated, meaning it is unknown if reported results stem from a well conditioned optimisation or merely an accident of initialisation. This study explores all three: A family-disjoint protocol, three classical baselines (both neural and tree-ensemble) evaluated in the same conditions from which we explicitly calculate gradient variance.

III. METHODOLOGY

This section specifies the complete QShield-NS framework. Figure 1 gives the end-to-end pipeline, from corpus construction through preprocessing, the Leave-One-Attack-Out partitioning scheme, the four competing models, and the evaluation battery. Each stage is formalised in the subsections that follow.

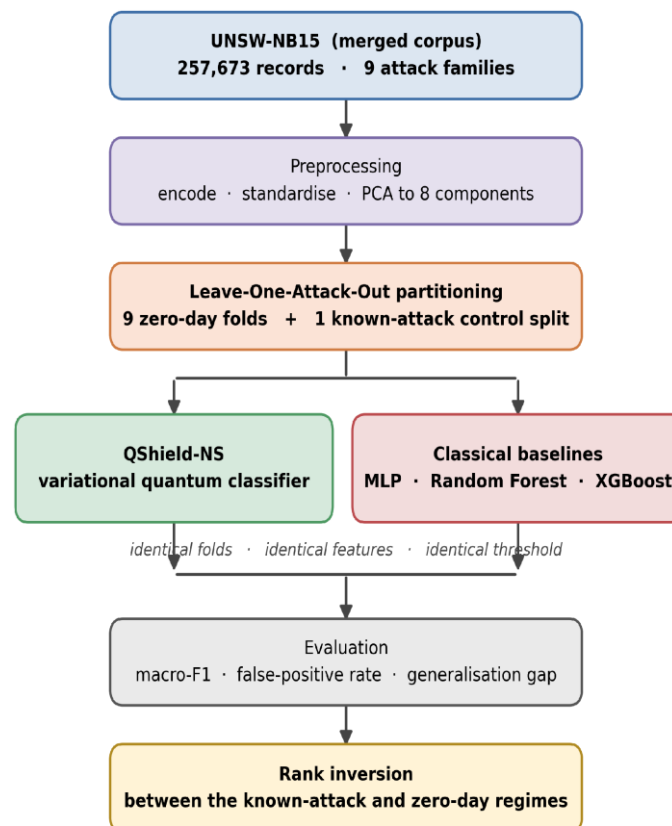


Figure.1. Overview of the QShield-NS methodology



3.1 Problem formulation and the zero-day protocol

Denote with in Eq. {1}, where each record consists of a feature vector, a binary label taking the value 1 to indicate an attack and fine-grained category taken from the set of nine attack families as well as benign class.

$$D = \{(x_i, y_i, c_i)\}_{i=1}^N, \quad x_i \in \mathbb{R}^d, \quad y_i \in \{0, 1\}, \quad c_i \in \mathcal{C} \cup \{\text{Normal}\} \quad (1)$$

The detection task is a binary mapping from a feature vector to the label. The zero-day requirement is enforced not via the loss function but by constructing the training and test partitions, contrasting with standard evaluation schemes. The training partition for a specified unseen family is defined by Eq. Eq. (2) and the test partition (3).

$$D_{\text{train}}^{(f)} = \{(x_i, y_i) : c_i = \text{Normal}, i \in \mathcal{I}_{\text{tr}}\} \cup \{(x_i, y_i) : c_i \in \mathcal{C} \setminus \{f\}\} \quad (2)$$

$$D_{\text{test}}^{(f)} = \{(x_i, y_i) : c_i = \text{Normal}, i \in \mathcal{I}_{\text{te}}\} \cup \{(x_i, y_i) : c_i = f\} \quad (3)$$

The withheld family does not participate in training, validation, feature scaling or principal-component basis. The key quantity we care about is the recall on that family at test time, which we call the zero-day detection rate—are instances of a new target class (of noise) detected at all by the decision boundary learned from eight families in which the model was trained on, and how well. For each family, we do this relative construction nine times.

In a paired control experiment, we use an entire stratified random split, such that all nine families are present in both partitions and use the same pre-processing and same models. We isolate the cost of novelty from the intrinsic quality of each model by comparing performance on control and LOAO, which is not possible when calculating either quantity alone; instead, this difference - between such a comparison - constitutes our main experimental signal.

3.2 Dataset and preprocessing

We use the distribution partitioning of UNSW-NB15 and merge published training file and the testing files to a single corpus which is equal to 257,673 records. We need merging because the official partition is stratified by proportion rather than category, so a family-disjoint split from it is impossible. The merged corpus consists of 93,000 normal records and 164,673 attack records that are split into nine families as illustrated in Table 1. The skew is extreme: Generic contributes nearly 36 percent of total attack traffic while Worms only about one-tenth of one percent, a four-order-of-magnitude gap that makes the rare families a true generalization challenge.

Table.1.Composition of the merged UNSW-NB15 corpus used in this study

Category	Records	Share of corpus (%)	Share of attack traffic (%)
Normal	93,000	36.09	—
Generic	58,871	22.85	35.75
Exploits	44,525	17.28	27.04
Fuzzers	24,246	9.41	14.72
DoS	16,353	6.35	9.93
Reconnaissance	13,987	5.43	8.49
Analysis	2,677	1.04	1.63
Backdoor	2,329	0.90	1.41
Shellcode	1,511	0.59	0.92
Worms	174	0.07	0.11
Total	257,673	100.00	100.00



The preprocessing consists of four stages as shown in Stage 2 in Fig. 1. Columns having non-predictive identifiers are eliminated and the column with fine-grained category is only kept to build the fold assignment, then removed from the feature matrix as even this would leak information we want the protocol designed to sequester. Numerical encoding of three categorical attributes, transport protocol, application service and connection state. The data is pre-preprocessed with standardised numeric attributes then the representation is projected onto the leading components using principal component analysis [17] as shown in Eq. (4).

$$z = W_n^T \left(\frac{x - \mu_{tr}}{\sigma_{tr}} \right) \in \mathbb{R}^n, \quad n = 8 \quad (4)$$

In this case, the standardisation statistics and the projection matrix are estimated only on the training partition of this fold. Question C: Dimensionality Reduction is not something to be fine-tuned this is an architectural decision QShield - NS uses an angle-encoding scheme which depends on committing one qubit per input dimension; simulating a circuit of n -qubits involves recording a state vector of 2^n complex amplitudes. Hence, eight qubits are a well-chosen trade-off between expressiveness and simulatability that aligns with achievable qubit numbers on current NISQ devices [20]. Importantly, all four models get the same eight-component representation provided to them, so no baseline is either helped or hurt by having the compression.

Leakage control is enforced strictly. And since the unseen family is absent from the training partition by design, it has no contribution to either the scaling statistics or component basis. This means there is no synthetic oversampling (to oversample the classes in the training set) used to address class imbalance; generating synthetic minority samples would be contra indicated in the LOAO setting since the family must not have been seen by the model at test time. The cumulative explained variance of the retained components is given in figure 2.

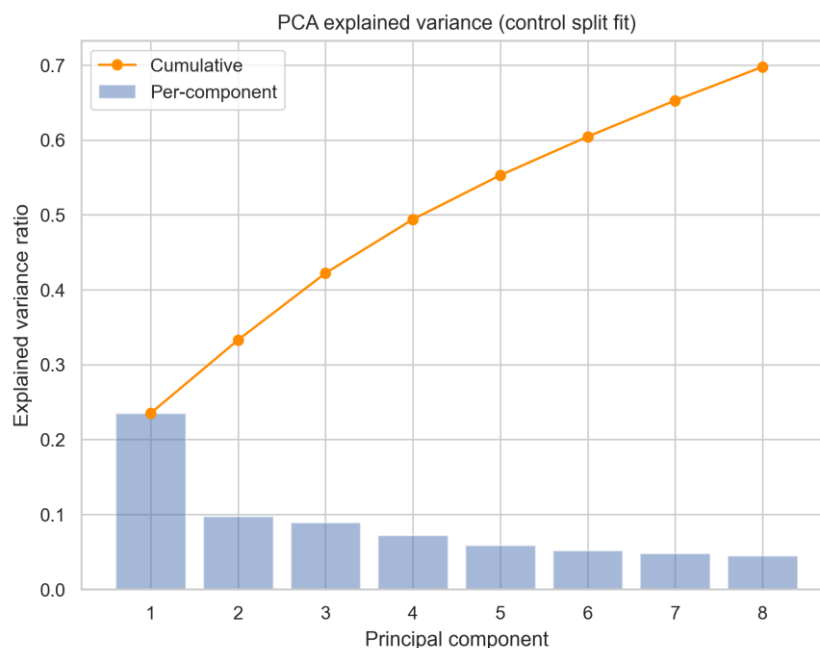


Figure.2. Cumulative explained variance of the eight retained principal components.

3.3 The QShield-NS architecture

QShield-NS is a hybrid quantum-classical classifier in which a variational quantum circuit acts as a differentiable feature transformer and a compact classical head produces the decision. Figure 3 shows the complete architecture, comprising an encoding stage, a variational ansatz, a local measurement stage and a classical readout.

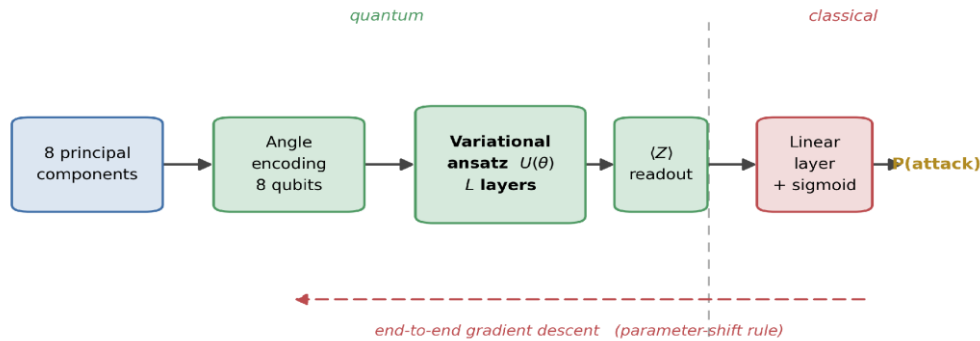


Figure.3. The QShield-NS hybrid architecture

Encoding. Angle embedding maps the 8-dimensional principal-component vector to an 8-qubit register, where its components parameterise a single-qubit rotation on our own wire (as given in Eq. (5)).

$$|\psi_0(z)\rangle = \bigotimes_{j=1}^n R_Y(z_j) |0\rangle^{\otimes n}, \quad R_Y(\phi) = \exp(-i\phi Y/2) \quad (5)$$

We use angle embedding instead of amplitude embedding because (1) it requires circuit depth linear rather than exponential in the input dimension and is therefore realisable on near-term hardware, and (2) the resulting model has a well-characterized Fourier spectrum [39].

Variational ansatz. L strongly entangling layers act on the encoded state. The layers consist of three parameterised rotations applied to each qubit followed by a ring of controlled-NOT gates connecting adjacent wires Eq. (6); the complete ansatz is an ordered product of these layers acting on a limited number of copies of the state, as seen in Eq.

$$U_\ell(\theta_\ell) = \left[\prod_{j=1}^n \text{CNOT}_{j, (j \bmod n) + 1} \right] \bigotimes_{j=1}^n R(\theta_{\ell,j,1}, \theta_{\ell,j,2}, \theta_{\ell,j,3}) \quad (4) \quad (6)$$

$$U(\theta) = \prod_{\ell=1}^L U_\ell(\theta_\ell), \quad |\psi(z, \theta)\rangle = U(\theta) |\psi_0(z)\rangle, \quad \theta \in \mathbb{R}^{3nL} \quad (7)$$

Thus, $3n$ trainable parameters per layer and consequently fully $3nL$ for all the ansatz. Although this ansatz is standard for the variational classification literature [28,31], it has enough expressivity to generate entanglement across the register, and sufficiently shallow that it avoids the expressivity-induced barren plateaus outlined by McClean et al. [33]. In the default configuration, we use $L=4$, for a total of 96 quantum parameters; in Section 4.5 (depth ablation), we probe options with $L \in \{1, 2, 4, 6\}$.

Measurement and classical head. Independently for each wire, the expectation value of the Pauli-Z observable is measured as in Eq. (8), which produces an eight-dimensional real vector that is fed into a linear layer with a sigmoid function output to get the attack probability, Eq. (9).

$$m_j(z, \theta) = \langle \psi(z, \theta) | Z_j | \psi(z, \theta) \rangle \in [-1, 1], \quad j = 1, \dots, n \quad (8)$$

$$\hat{y}(z) = \sigma(w^\top m(z, \theta) + b), \quad \sigma(u) = (1 + e^{-u})^{-1} \quad (9)$$



This is actually a design choice that leads to using local single-qubit observables instead of a global multi-qubit observable: Cerezo et al. demonstrated that in regimes where global cost functions onset non-vanishing gradients at shallow depth, local cost functions do not [34]. Training. The entire model, with the quantum parameters combined along the overhead and classical head is optimised end-to-end by minimising the class-weighted binary cross-entropy of Eq. (10), where weights are used to balance the attack records with benign records in this training fold.

$$\mathcal{L}(\theta, w, b) = -\frac{1}{N} \sum_{i=1}^N [w_+ y_i \log \hat{y}_i + w_- (1 - y_i) \log (1 - \hat{y}_i)] \quad (10)$$

Using PennyLane [40] interfaced to PyTorch [15], the circuit was implemented as a differentiable layer in the classical automatic-differentiation graph. For the parameter-shift rule, Eq. (D06) gives gradients of the measured observables with respect to the circuit parameters analytically. an alternative (11), where the same circuit at two shifted values of α is evaluated, requiring no finite-difference approximation.

$$\frac{\partial m_j}{\partial \theta_k} = \frac{1}{2} [m_j(\theta_k + \frac{\pi}{2}) - m_j(\theta_k - \frac{\pi}{2})] \quad (11)$$

Optimisation uses Adam [16]. Since classical simulation of the circuit dominates wall-clock cost, when training our quantum model its trained on a balanced subsample consisting of at most 25,000 records per fold; however evaluation is always run on the complete test partition so that no fold is scored against a reduced or filtered test set. The parameter count is low: 96 quantum parameters for the default depth and nine classical parameters for the readout head. This is two to three orders of magnitudes less than the classical neural baseline (which we will discuss in Section 5 via an analysis of its results)—the robustness under distribution shift we observe is obtained from a substantially lower-capacity model.

3.4 Classical baselines

Three classical baselines are reviewed under the same conditions as those for QShield-NS: The identical folds, the use of eight principal components in total, a decision threshold after refitting (25%) and metrics that are computed on predictions. The feedforward network is a multilayer perceptron with full connectivity, trained using Adam and binary cross-entropy as the closest classical equivalent to the hybrid model as a parametric classifier for the same input operating in gradient-descent. The random forest, which is an ensemble of bootstrapped decision trees [12], is the class of methods that currently yield state-of-the-art published results on tabular intrusion corpora. The strongest reported classical performer on UNSW-NB15 in comparable studies, XGBoost is a regularised gradient-boosted tree ensemble [13]. For classical models we use scikit-learn [14] and the XGBoost library; for the neural baseline we use PyTorch [15].

Moreover, reporting all three issues for the interpretation of the outcomes, since classical baselines do not behave consistently below the zero-day protocol. Separating a true property of the variational model from an artefact of one weak comparison requires a suite that encompasses both families, neural and tree-ensemble.

IV. RESULTS

We present the results for: known-attack control experiment (4.1), an examination of the zero-day LOAO regime and rank inversion between the two regimes (4.2), a per-family breakdown (4.3), the false-positive analysis that motivates some of those inverting findings (4.4) and, lastly, circuit-depth ablation (4.5) and gradient-variance study (4.6).

4.1 Known-attack control

Table 2 presents performances against the stratified control split, where each of the nine attack families are seen in both train and test partitions. This is formal evaluation setting, and twofold: First it checks that the pipeline is working and matches up to what others in the field have published; second it provide a baseline for comparing cover any degradation of zero-day against.



Table.2. Control split (all attack families seen during training). Models ordered by macro-F1.

Model	Accuracy	Precision	Recall	F1	Macro-F1	ROC-AUC	PR-AUC
Random Forest	0.9231	0.9292	0.9523	0.9406	0.9158	0.9840	0.9909
XGBoost	0.9168	0.9231	0.9488	0.9358	0.9088	0.9817	0.9898
MLP	0.9093	0.9288	0.9293	0.9291	0.9017	0.9781	0.9877
QShield-NS (VQC)	0.8889	0.9107	0.9160	0.9133	0.8793	0.9660	0.9804

The classical baselines operate as predicted by the literature. Out of these, random forest performs best with macro-F1 0.9158 and accuracy 0.9231, followed by XGBoost and the multilayer perceptron. These numbers fit well into the spectrum provided by similar works using UNSW-NB15 binary classification [9,49] and confirm that while having aggressively reduced to an eight-component representation they still contain enough information for competitive classification, thus also proving there are no major defects within the pipeline.

In this regime, QShield-NS is last on every metric, with macro-F1 0.8793 - 0.0365 behind random forest. This is what we would expect and conforms to the wider result that classical models with data are hard to beat on tabular tasks [38]. A low-capacity model of 96-parameter variational circuit using eight compressed features competing with ensembles of hundreds of trees. If reported alone, then Table 2 would indicate that the quantum model is not better. That conclusion would be premature, as Section 4.2 shows.

4.2 Zero-day detection and the rank inversion

The macro-average across the nine LOAO folds, in which each model is given a score on an attack family it has never seen, is shown in Table 3. On the threshold-dependent metrics, the ordering of the four models is exactly opposite that of Table 2.

Table.3. Zero-day LOAO results, macro-averaged over nine folds. Models ordered by macro-F1.

Model	Accuracy	Precision	Recall	F1	Macro-F1	ROC-AUC	PR-AUC
QShield-NS (VQC)	0.8348	0.6118	0.7673	0.5986	0.7303	0.9281	0.7318
Random Forest	0.8042	0.5596	0.8270	0.5643	0.7044	0.9488	0.8413
XGBoost	0.7989	0.5409	0.8515	0.5628	0.7004	0.9514	0.8456
MLP	0.7643	0.5186	0.8214	0.5306	0.6715	0.9370	0.8050

QShield-NS achieves both the highest macro-F1 (0.7303), accuracy (0.8348), precision (0.6118) and attack-class F1 (0.5986). The leader in the control regime, random forest, drops to second on macro-F1 (0.7044); the multilayer perceptron which was third in the control drops to last (0.6715). The 0.0259macro-F1 margin of QShield-NS over the best classical model is on a level comparable in size with the 0.0365 qualitative deficit that the quantum model bore in control mode, making this reversal not an orbital fluctuation but one of similar size to the original gap.

The uniform pattern is not sustained across the board, and the exceptions are telling. Out of the 4 models, QShield-NS has the lowest recall (0.7673 vs 0.8515 for XGBoost) and most surprisingly also the lowest ROC-AUC (0.9281) and PR-AUC (0.7318). A study of these results without any thresholding method shows that on both metrics the quantum model is lower ranked than some comparison method in all nine folds. This means that the variational classifier separates the two classes at a ranking level less cleanly than any classical baseline but does so with more balanced performance at the operating threshold. We resolve this apparent contradiction in Section 4.4.

In Table 4 we quantify the reduction in performance each model experienced while transitioning from control regime to the zero-day regime, which is known as the generalisation gap (Eq. (15), which is the main quantity of this work.



Table.4. Generalisation gap: degradation from the known-attack control to the zero-day LOAO regime.

Model	Macro-F1 (control)	Macro-F1 (LOAO)	Macro-F1 gap	Accuracy (control)	Accuracy (LOAO)	Accuracy gap
QShield-NS (VQC)	0.8793	0.7303	0.1489	0.8889	0.8348	0.0540
XGBoost	0.9088	0.7004	0.2084	0.9168	0.7989	0.1178
Random Forest	0.9158	0.7044	0.2114	0.9231	0.8042	0.1189
MLP	0.9017	0.6715	0.2302	0.9093	0.7643	0.1450

The degradation for the quantum model is of 0.1489 macro-F1, compared to 0.2084 to 0.2302 for the classical baselines a reduction in degradation of between 29 and 35 percent! On accuracy the contrast is even starker: QShield-NS loses 0.0540 while all classical models lose between 0.1178 and 0.1450, a reduction in performance of between 54% and 63%. The QShield-NS is the weakest model that starts out strong and ends up even stronger, not because it gets an upgrade but rather decays significantly less. In particular, the multilayer perceptron—the classical model that is architecturally closest to the hybrid and shares its gradient-based training and input representation—exhibits by far the greatest degradation of all, suggesting an effect due to variational structure rather than to optimisation procedure or feature space.

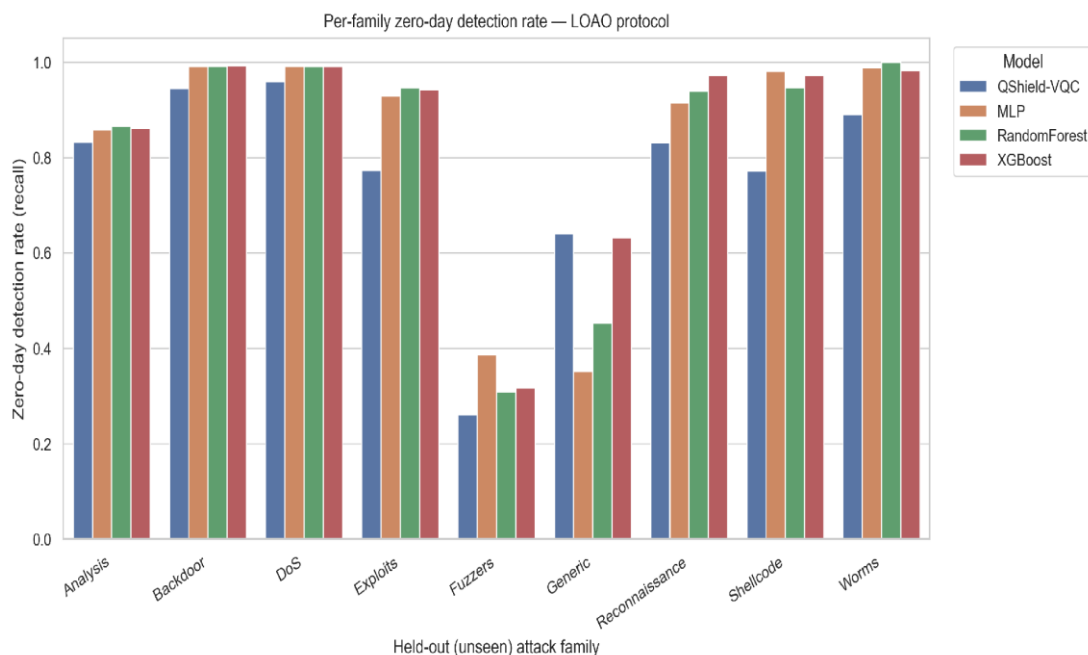


Figure.4. Zero-day detection rate (recall on the withheld family) for each of the nine LOAO folds

4.3 Per-family analysis

Table 5 shows the detection rate at zero-day across each family and model. This is due to the intrinsic separability of the held-out family dwarfing any classifier choice since the variation across families has a much larger range than the variation across models.



Table.5.Zero-day detection rate (recall on the unseen family) for each LOAO fold, with family size for reference

Unseen family	Records	QShield-NS	MLP	Random Forest	XGBoost
Analysis	2,677	0.8323	0.8581	0.8663	0.8614
Backdoor	2,329	0.9450	0.9910	0.9914	0.9923
DoS	16,353	0.9585	0.9911	0.9906	0.9910
Exploits	44,525	0.7737	0.9299	0.9457	0.9428
Fuzzers	24,246	0.2608	0.3872	0.3084	0.3169
Generic	58,871	0.6408	0.3511	0.4534	0.6323
Reconnaissance	13,987	0.8312	0.9153	0.9401	0.9722
Shellcode	1,511	0.7723	0.9808	0.9471	0.9722
Worms	174	0.8908	0.9885	1.0000	0.9828
Macro-average	164,673	0.7673	0.8214	0.8270	0.8515

Every model reliably observes three families. Detection rates for Backdoor, DoS and Worms are above 0.89 for all four classifiers, but reach over 0.98 for classical baselines. The families generate flow signatures volumetric anomalies, stray connection states, outlier duration and byte-count profiles that are all shared with the eight remaining families held in the training data so transfer learning experienced by the learner from known attacks is painless. The Worms result is the strongest example: The family is simply too small with 174 records to really have any impact on a boundary, and yet it is perfectly discovered because its pattern looks similar to attack traffic in general.

By our experiments, the fuzzers are the most difficult family for all models with detection rates from 0.2608 to 0.3872. Fuzzing is the act of generating malformed or randomised inputs against a target service, and at the flow level rich traffic for this is little different to benign protocol use—normal connection states, unremarkable volumes, ordinary durations. The key is that the discriminative signal lies in payload semantics neglected by the flow-level feature set. This single place in the study stands out as being operationally important negative result, indicating that feature representation is limited rather than any classifier, for all architecturally diverse models a variational circuit, a neural network and two tree ensembles.

The biggest attack family on the corpus is Generic fold, for which the pattern inverts so it deserves attention. Here QShield-NS scores 0.6408, ahead of XGBoost (0.6323), random forest (0.4534) and the multilayer perceptron (0.3511) the quantum model beats the weakest classical baseline by a factor of 1.8. Generic consists of attacks against block ciphers that are agnostic to the specific cryptographic primitive, and its flow properties differ dramatically from those belonging to the much larger exploit-driven and reconnaissance families in the rest of the training data. This means that this is the fold in which the held-out family is furthest from distribution in terms of training data; and it is also precisely where the variational model shows its best relative performance. This is consistent with the analysis of generalisation gaps from Section 4.2 and also the theoretical argument that variational feature maps induce a different shaped decision boundary [37, 39].

For the other fold analyses, Analysis, Exploits, and Reconnaissance, Shellcode QShield-NS achieves lower detection rates than the classical baselines in some cases, much worse (Exploits 0.7737 against 0.9457 for random forest; Shellcode 0.7723 against 0.9808 for ML perceptron). On their own, these numbers would imply that the quantum model is only less sensitive. However, QShield-NS achieves the higher macro-F1 in six out of nine folds and higher precision in eight out of nine. The next subsection addresses the resolution of this tension.



4.4 False-positive analysis: the mechanism of the inversion

This apparent contradiction between smaller recall and lower AUC on the one hand, and higher macro-F1 and higher accuracy on the other, is resolved by looking at false-positive burden. Since each LOAO fold is evaluated on the same 27,900 held-out normal records as the baseline comparison for Eq. The third metric, AUC (13), is directly comparable across folds and models. The corresponding rates are shown in table 6.

Table.6. False-positive rate (%) on the fixed pool of 27,900 held-out normal records, by fold and model.

Unseen family	QShield-NS	Random Forest	XGBoost	MLP
Analysis	9.67	17.07	19.91	23.33
Backdoor	8.00	17.00	20.29	20.40
DoS	14.38	16.98	20.16	20.86
Exploits	7.14	14.98	17.99	21.64
Fuzzers	4.19	5.41	5.85	8.28
Generic	19.75	16.46	18.87	23.92
Reconnaissance	10.82	16.19	19.66	16.33
Shellcode	12.23	16.87	20.20	22.73
Worms	10.32	17.16	20.18	24.71
Mean	10.72	15.35	18.12	20.24

QShield-NS achieves lowest false-positive rate across all nine folds, with the exception of Generic where it performed second best. The approach averages across folds, marking a mere 10.72 percent of benign traffic as malicious compared to 15.35 percent for random forest, 18.12 percent for XGBoost and 20.24 percent for the multilayer perceptron. As a result, the quantum model believably cuts the false-alarm rate in half compared to both the neural baseline and nearly 30 percent relative to even the best classical performer.

This discrepancy cannot be overemphasized in its absolute magnitude. For the Worms fold, multilayer perceptron raises about 6,894 alarms to capture 172 true positives and QShield-NS raises approximately 2,880 false alarms to capture 155. The classical model is able to detect seventeen more attack records but at the cost of approximately four thousand extra benign flows that must be read by a human analyst. Such a trade is bad enough in an operational security centre: at scale, it is impossible.

This also the reason why QShield-NS is outperformed on ROC-AUC and PR-AUC in each fold, whilst leading on threshold-dependent metrics. Metrics in the two families measure different properties. This invariant to the threshold, as AUC measures the goodness of equality of ranking over instances and so classical models rank better, which means that mean- per-fold threshold re-calibration could basically be tuned to a more advantageous operating point. Macro-F1 at a threshold evaluates where exactly the boundary is placed. In deployment, per-family threshold recalibration is impossible by definition—the family is unknown so there exists no labelled data on which to calibrate. It is therefore this type of fixed-threshold comparison that is operationally meaningful, and it is the strong-fixed-threshold comparison QShield-NS wins.

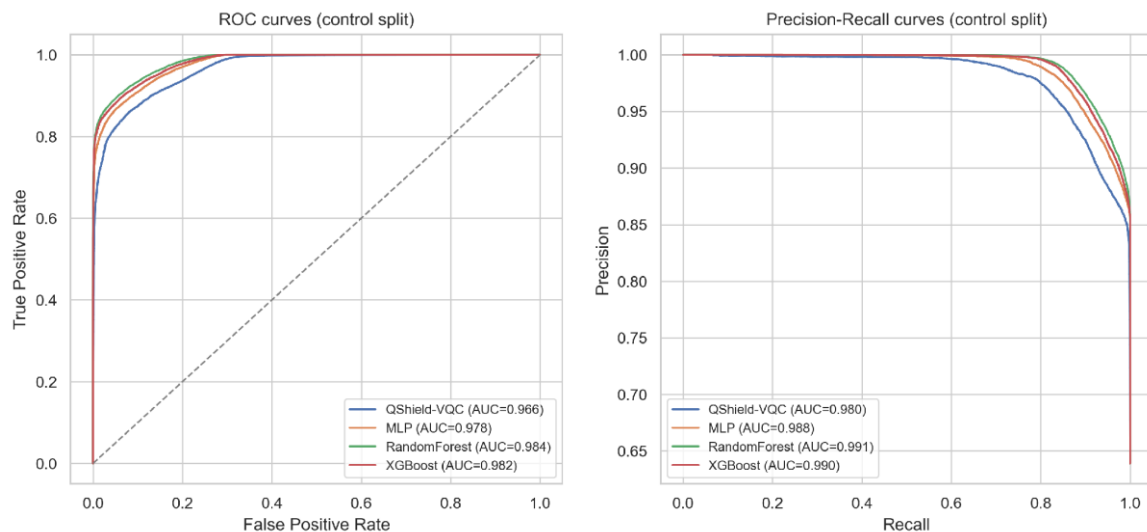


Figure.5. ROC and precision–recall curves

4.5 Circuit depth ablation

In Table 7, we report the control-split performance of QShield-NS as a function of L , while the other settings remain constant.

Table.7. QShield-NS performance versus variational circuit depth L (control split).

Depth L	Quantum parameters	Accuracy	Macro-F1	Macro-F1 over $L-1$	gain
1	24	0.8549	0.8424	—	
2	48	0.8813	0.8735	+0.0311	
4	96	0.8889	0.8793	+0.0058	
6	144	0.8916	0.8828	+0.0035	

Depth is monotonic but shows very rapid diminishing returns in performance. Going from one to two layers gives a macro-F1 gain of 0.0311; doubling this again to four layers gives 0.0058, and another doubling to six layers yields 0.0035 more. Obviously a single layer is not enough it only allows a single round of entangling across the register while from two layers onward, the model is expressive enough that extra parameters have diminishing returns. The standard configuration of $L = 4$ lies on a sensible point of this curve: it captures 95 percent of the overall gain available up to $L = 6$ but using two thirds fewer parameters, and thus significantly less simulation time as the parameter-shift gradient evaluation scales linearly with parameter count.

Remarkably, the lack of performance degradation at a greater depth is worthy in itself. This is consistent with the absence of optimization degradation from barren plateaus or overfitting in overparameterised variational circuits, as neither was observed here (see gradient measurements next).

4.6 Gradient variance and trainability

For variational models, the dominant scaling resource is barren plateaus, where the gradient variance decays exponentially in qubit count or circuit depth [33,34,35]. Table 8 summarizes the empirical variance obtained by Eq. We define the circuit depth as (16) along with the absolute average of the gradient magnitudes at randomly initialised data over 200 repetitions, each with a different random parameter initialisation at (eight qubits for each depth).



Table.8. Empirical gradient statistics for the QShield-NS ansatz at eight qubits (200 random initialisations per depth)

Depth L	Gradient variance	Mean absolute gradient	Initialisations	Qubits
1	3.97×10^{-34}	1.39×10^{-17}	200	8
2	1.82×10^{-3}	1.93×10^{-2}	200	8
4	2.13×10^{-3}	3.16×10^{-2}	200	8
6	1.55×10^{-3}	3.12×10^{-2}	200	8

$L = 1 \rightarrow$ Needs to be read with caution and should not be considered an extreme desert plateau. For instance, a variance of the order 10^{-34} and a mean absolute gradient of the order 10^{-17} are numerically indistinguishable from exact zero in double precision. It is a structural degeneracy of the single-layer configuration rather than a statistical concentration effect: with one entangling layer and local Pauli-Z readout, the gradient of the measured observable with respect to certain rotation parameters vanishes identically regardless of the parameter values. In the literature, the barren plateau is described as a probabilistic concentration of gradients around zero as the circuit grows, whereas the $L = 1$ result has a deterministic origin due to the ansatz geometry. This correlates with the poor performance seen at $L = 1$ in Table 7 and shows why depth one is an impossible configuration to learn from.

The scientifically meaningful trend lies in the range $L \in \{2, 4, 6\}$, where gradient variance remains of order 10^{-3} and does not decay with depth: it rises from 1.82×10^{-3} at $L = 2$ to 2.13×10^{-3} at $L = 4$ before falling to 1.55×10^{-3} at $L = 6$. The mean absolute gradient increases from 1.93×10^{-2} to 3.16×10^{-2} and then stabilises. No exponential decay is present. At eight qubits with local observables and depth at most six, the QShield-NS ansatz is comfortably trainable, and the mild reduction between $L = 4$ and $L = 6$ is consistent with the onset of the concentration effect but far from a plateau.

Such a result can not be extended to larger registers. Barren plateaus asymptote in qubit count, and eight is well below the point at which they become severe. This measurement shows that the observed results are not an artefact of a poorly conditioned optimisation landscape and also suggests that simply scaling QShield-NS to larger registers will not suffice: intrinsic near-term error mitigation strategies will be needed (such as structured initialisation [36] or explicit locality constraints on the cost function [34]).

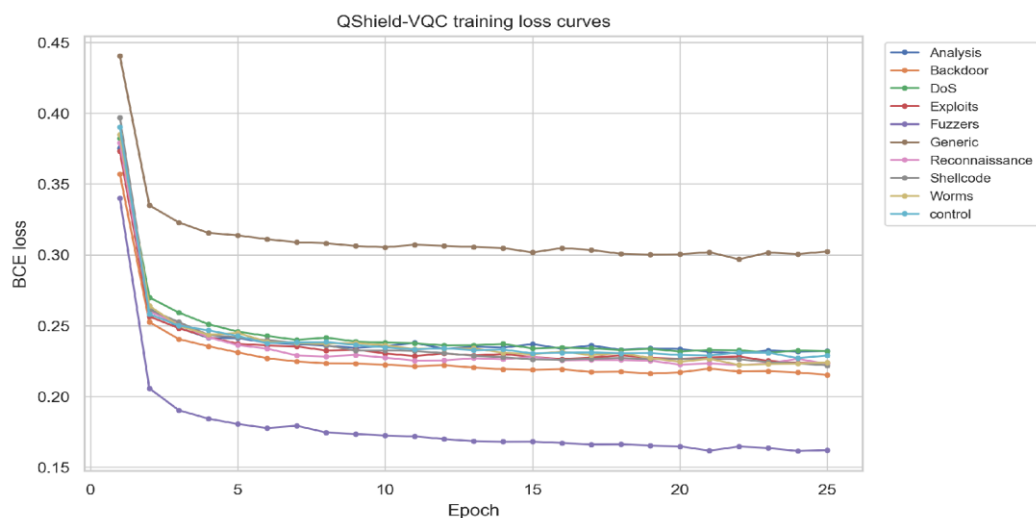


Figure.9. QShield-NS training loss curves. Convergence is smooth and stable across folds, corroborating the gradient-variance measurements.



VI. CONCLUSION AND FUTURE WORK

We introduced QShield-NS, a nimble variational quantum classifier to detect zero-day cyber threats and tested it against nine popular attack families on UNSW-NB15 under the leave-one-attack-out protocol where in turn each of the attack families is withheld from training. The results provide four established findings against a multilayer perceptron, a random forest and XGBoost under the same splits, features and thresholds.

Under novelty, we see that the variational model is the least affected. Its macro-F1 decreases by 0.1489 across regimes, as opposed to decrease of 0.2084 to 0.2302 for the classical baselines and its accuracy drops by 0.0540 against drop of 0.1178 to 0.1450. Worst of all, the neural baseline—closest to the hybrid in terms of optimiser and input space—degrades most, locating the effect in the variational structure.

Several directions follow. Disentangling quantum structure from capacity is most critical: classical generic models would be the ultimate test with data-driven capacity matched consumers on size and limited to comparable function classes determining whether that model matching could only come if training occurred largely in a low-capacity one. To establish whether the rank inversion originates from the model class or merely from UNSW-NB15, one must replicate on CIC-IDS2017 and additional corpora. Whether the advantage survives realistic noise is established by execution on physical NISQ hardware. Registers with structured initialisation [36] could be another option for testing scaling in excess of 8 qubits. Lastly, the fact that Fuzzers fail uniformly motivates using hybrid architectures leveraging flow-level and payload-level features, as in this case it is a limit of representation instead of any classifier.

In fact, we recommend that family-disjoint evaluation with explicit false-positive reporting at a fixed threshold become commonplace for the evaluation of any intrusion detection system—a quantum or classical classifier—operated in an adversarial environment. The best-performing models when randomly partitioned are not necessarily the same as the best-performing systems when they are faced with something new.

REFERENCES

- [1] Moustafa, N.; Slay, J. UNSW-NB15: A comprehensive data set for network intrusion detection systems (UNSW-NB15 network data set). In Proceedings of the 2015 Military Communications and Information Systems Conference (MilCIS), Canberra, Australia, 2015; pp. 1–6.
- [2] Moustafa, N.; Slay, J. The evaluation of Network Anomaly Detection Systems: Statistical analysis of the UNSW-NB15 data set and the comparison with the KDD99 data set. *Information Security Journal: A Global Perspective* 2016, 25, 18–31.
- [3] Tavallaee, M.; Bagheri, E.; Lu, W.; Ghorbani, A.A. A detailed analysis of the KDD CUP 99 data set. In Proceedings of the 2009 IEEE Symposium on Computational Intelligence for Security and Defense Applications (CISDA), Ottawa, Canada, 2009; pp. 1–6.
- [4] Sommer, R.; Paxson, V. Outside the Closed World: On Using Machine Learning for Network Intrusion Detection. In Proceedings of the 2010 IEEE Symposium on Security and Privacy, Oakland, CA, USA, 2010; pp. 305–316.
- [5] Buczak, A.L.; Guven, E. A Survey of Data Mining and Machine Learning Methods for Cyber Security Intrusion Detection. *IEEE Communications Surveys & Tutorials* 2016, 18, 1153–1176.
- [6] Khraisat, A.; Gondal, I.; Vamplew, P.; Kamruzzaman, J. Survey of intrusion detection systems: techniques, datasets and challenges. *Cybersecurity* 2019, 2, 20.
- [7] Chandola, V.; Banerjee, A.; Kumar, V. Anomaly detection: A survey. *ACM Computing Surveys* 2009, 41, 1–58.
- [8] Vinayakumar, R.; Alazab, M.; Soman, K.P.; Poornachandran, P.; Al-Nemrat, A.; Venkatraman, S. Deep Learning Approach for Intelligent Intrusion Detection System. *IEEE Access* 2019, 7, 41525–41550.
- [9] Kasongo, S.M.; Sun, Y. Performance analysis of intrusion detection systems using a feature selection method on the UNSW-NB15 dataset. *Journal of Big Data* 2020, 7, 105.
- [10] Moustafa, N.; Turnbull, B.; Choo, K.-K.R. An Ensemble Intrusion Detection Technique Based on Proposed Statistical Flow Features for Protecting Network Traffic of Internet of Things. *IEEE Internet of Things Journal* 2019, 6, 4815–4830.
- [11] Sharafaldin, I.; Lashkari, A.H.; Ghorbani, A.A. Toward Generating a New Intrusion Detection Dataset and Intrusion Traffic Characterization. In Proceedings of the 4th International Conference on Information Systems Security and Privacy (ICISSP), Funchal, Portugal, 2018; pp. 108–116.
- [12] Breiman, L. Random Forests. *Machine Learning* 2001, 45, 5–32.



- [13] Chen, T.; Guestrin, C. XGBoost: A Scalable Tree Boosting System. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 2016; pp. 785–794.
- [14] Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; et al. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* 2011, 12, 2825–2830.
- [15] Paszke, A.; Gross, S.; Massa, F.; Lerer, A.; Bradbury, J.; Chanan, G.; et al. PyTorch: An Imperative Style, High-Performance Deep Learning Library. In *Advances in Neural Information Processing Systems 32 (NeurIPS)*, Vancouver, Canada, 2019; pp. 8024–8035.
- [16] Kingma, D.P.; Ba, J. Adam: A Method for Stochastic Optimization. In Proceedings of the 3rd International Conference on Learning Representations (ICLR), San Diego, CA, USA, 2015.
- [17] Jolliffe, I.T. *Principal Component Analysis*, 2nd ed.; Springer: New York, NY, USA, 2002.
- [18] Shor, P.W. Polynomial-Time Algorithms for Prime Factorization and Discrete Logarithms on a Quantum Computer. *SIAM Journal on Computing* 1997, 26, 1484–1509.
- [19] Grover, L.K. A fast quantum mechanical algorithm for database search. In Proceedings of the 28th Annual ACM Symposium on Theory of Computing (STOC), Philadelphia, PA, USA, 1996; pp. 212–219.
- [20] Preskill, J. Quantum Computing in the NISQ era and beyond. *Quantum* 2018, 2, 79.
- [21] Mosca, M. Cybersecurity in an Era with Quantum Computers: Will We Be Ready? *IEEE Security & Privacy* 2018, 16, 38–41.
- [22] Chen, L.; Jordan, S.; Liu, Y.-K.; Moody, D.; Peralta, R.; Perlner, R.; Smith-Tone, D. Report on Post-Quantum Cryptography; NISTIR 8105; National Institute of Standards and Technology: Gaithersburg, MD, USA, 2016.
- [23] Alagic, G.; Alperin-Sheriff, J.; Apon, D.; Cooper, D.; Dang, Q.; Kelsey, J.; et al. Status Report on the Second Round of the NIST Post-Quantum Cryptography Standardization Process; NISTIR 8309; National Institute of Standards and Technology: Gaithersburg, MD, USA, 2020.
- [24] Biamonte, J.; Wittek, P.; Pancotti, N.; Rebentrost, P.; Wiebe, N.; Lloyd, S. Quantum machine learning. *Nature* 2017, 549, 195–202.
- [25] Schuld, M.; Sinayskiy, I.; Petruccione, F. An introduction to quantum machine learning. *Contemporary Physics* 2015, 56, 172–185.
- [26] 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* 2019, 567, 209–212.
- [27] Schuld, M.; Killoran, N. Quantum Machine Learning in Feature Hilbert Spaces. *Physical Review Letters* 2019, 122, 040504.
- [28] Schuld, M.; Bocharov, A.; Svore, K.M.; Wiebe, N. Circuit-centric quantum classifiers. *Physical Review A* 2020, 101, 032308.
- [29] Mitarai, K.; Negoro, M.; Kitagawa, M.; Fujii, K. Quantum circuit learning. *Physical Review A* 2018, 98, 032309.
- [30] Farhi, E.; Neven, H. Classification with Quantum Neural Networks on Near Term Processors. *arXiv* 2018, arXiv:1802.06002.
- [31] Benedetti, M.; Lloyd, E.; Sack, S.; Fiorentini, M. Parameterized quantum circuits as machine learning models. *Quantum Science and Technology* 2019, 4, 043001.
- [32] Cerezo, M.; Arrasmith, A.; Babbush, R.; Benjamin, S.C.; Endo, S.; Fujii, K.; et al. Variational quantum algorithms. *Nature Reviews Physics* 2021, 3, 625–644.
- [33] McClean, J.R.; Boixo, S.; Smelyanskiy, V.N.; Babbush, R.; Neven, H. Barren plateaus in quantum neural network training landscapes. *Nature Communications* 2018, 9, 4812.
- [34] Cerezo, M.; Sone, A.; Volkoff, T.; Cincio, L.; Coles, P.J. Cost function dependent barren plateaus in shallow parametrized quantum circuits. *Nature Communications* 2021, 12, 1791.
- [35] Wang, S.; Fontana, E.; Cerezo, M.; Sharma, K.; Sone, A.; Cincio, L.; Coles, P.J. Noise-induced barren plateaus in variational quantum algorithms. *Nature Communications* 2021, 12, 6961.
- [36] Grant, E.; Wossnig, L.; Ostaszewski, M.; Benedetti, M. An initialization strategy for addressing barren plateaus in parametrized quantum circuits. *Quantum* 2019, 3, 214.
- [37] Abbas, A.; Sutter, D.; Zoufal, C.; Lucchi, A.; Figalli, A.; Woerner, S. The power of quantum neural networks. *Nature Computational Science* 2021, 1, 403–409.
- [38] Huang, H.-Y.; Broughton, M.; Mohseni, M.; Babbush, R.; Boixo, S.; Neven, H.; McClean, J.R. Power of data in quantum machine learning. *Nature Communications* 2021, 12, 2631.
- [39] Schuld, M.; Sweke, R.; Meyer, J.J. Effect of data encoding on the expressive power of variational quantum-machine-learning models. *Physical Review A* 2021, 103, 032430.
- [40] Bergholm, V.; Izaac, J.; Schuld, M.; Gogolin, C.; Ahmed, S.; Ajith, V.; et al. PennyLane: Automatic differentiation of hybrid quantum-classical computations. *arXiv* 2018, arXiv:1811.04968.
- [41] Lloyd, S.; Mohseni, M.; Rebentrost, P. Quantum principal component analysis. *Nature Physics* 2014, 10, 631–633.



- [42] Rebentrost, P.; Mohseni, M.; Lloyd, S. Quantum Support Vector Machine for Big Data Classification. *Physical Review Letters* 2014, 113, 130503.
- [43] Gouveia, A.; Correia, M. Towards Quantum-Enhanced Machine Learning for Network Intrusion Detection. In *Proceedings of the 2020 IEEE 19th International Symposium on Network Computing and Applications (NCA)*, Cambridge, MA, USA, 2020; pp. 1–8.
- [44] Payares, E.D.; Martinez-Santos, J.C. Quantum machine learning for intrusion detection of distributed denial of service attacks: a comparative overview. *Proceedings of SPIE 11699, Quantum Computing, Communication, and Simulation*, 116990B, 2021.
- [45] Chen, J.; Qi, X.; Chen, L.; Chen, F.; Cheng, G. Quantum-inspired ant lion optimized hybrid k-means for cluster analysis and intrusion detection. *Knowledge-Based Systems* 2020, 203, 106167.
- [46] Chen, S.Y.-C.; Yang, C.-H.H.; Qi, J.; Chen, P.-Y.; Ma, X.; Goan, H.-S. Variational Quantum Circuits for Deep Reinforcement Learning. *IEEE Access* 2020, 8, 141007–141024.
- [47] Mitchell, R.; Chen, I.-R. A survey of intrusion detection techniques for cyber-physical systems. *ACM Computing Surveys* 2014, 46, 1–29.
- [48] Ahmad, Z.; Shahid Khan, A.; Wai Shiang, C.; Abdullah, J.; Ahmad, F. Network intrusion detection system: A systematic study of machine learning and deep learning approaches. *Transactions on Emerging Telecommunications Technologies* 2021, 32, e4150.
- [49] Jing, D.; Chen, H.-B. SVM Based Network Intrusion Detection for the UNSW-NB15 Dataset. In *Proceedings of the 2019 IEEE 13th International Conference on ASIC (ASICON)*, Chongqing, China, 2019; pp. 1–4.
- [50] Kumar, V.; Das, A.K.; Sinha, D. Statistical analysis of the UNSW-NB15 dataset for intrusion detection. In *Computational Intelligence in Pattern Recognition*; Springer: Singapore, 2020; pp. 279–294.