

# Towards Efficient Quantum Anomaly Detection: One-Class SVMs Using Variable Subsampling and Randomized Measurements

Michael Kölle<sup>1</sup>, Afrae Ahouzi<sup>1,2</sup>, Pascal Debus<sup>2</sup>, Robert Müller<sup>1</sup>, Daniëlle Schuman<sup>1</sup>  
and Claudia Linnhoff-Popien<sup>1</sup>

<sup>1</sup>*Institute of Informatics, LMU Munich, Munich, Germany*

<sup>2</sup>*Fraunhofer AISEC, Garching, Germany*

**Keywords:** Quantum Machine Learning, Anomaly Detection, OC-SVM.

**Abstract:** Quantum computing, with its potential to enhance various machine learning tasks, allows significant advancements in kernel calculation and model precision. Utilizing the one-class Support Vector Machine alongside a quantum kernel, known for its classically challenging representational capacity, notable improvements in average precision compared to classical counterparts were observed in previous studies. Conventional calculations of these kernels, however, present a quadratic time complexity concerning data size, posing challenges in practical applications. To mitigate this, we explore two distinct approaches: utilizing randomized measurements to evaluate the quantum kernel and implementing the variable subsampling ensemble method, both targeting linear time complexity. Experimental results demonstrate a substantial reduction in training and inference times by up to 95% and 25% respectively, employing these methods. Although unstable, the average precision of randomized measurements discernibly surpasses that of the classical Radial Basis Function kernel, suggesting a promising direction for further research in scalable, efficient quantum computing applications in machine learning.

## 1 INTRODUCTION

In the prevailing digital landscape, anomaly detection algorithms safeguard numerous systems by identifying irregularities, such as unauthorized intrusion into a network, unexpected machinery behavior, or abnormal medical readings (Fernando et al., 2021; Boloka et al., 2021). In e-commerce, these algorithms are pivotal in protecting both customers and companies from fraudulent transactions, which could precipitate substantial financial losses (ECB, 2023). Anomaly detection faces challenges, including the diverse and rarely similar nature of anomalies, and often encounters high-dimensional, correlated, and sometimes unlabeled data sets.

Quantum Machine Learning (QML), a synthesis of machine learning and quantum computing, promises potential solutions to some of these challenges by leveraging the capability of quantum algorithms to compute classically challenging kernels (Havlíček et al., 2019). Some initiatives have aimed at ameliorating existing anomaly detection methods with quantum methods. A notable attempt (Kyriienko and Magnusson, 2022) achieved a 20% elevation in

average precision using one-class support vector machine models with a quantum kernel, though confronted quadratic scaling challenges with data size, impacting both training and testing times.

Addressing the aforementioned time complexity challenge, this work replicates the results from (Kyriienko and Magnusson, 2022) and employs them as benchmarks to explore the efficacy of two linear time complexity methods based on data size: *randomized measurements* for quantum kernel measurement and an ensemble method termed *variable subsampling*. The focus rests on two pivotal questions:

1. Can the quantum kernel extract superior information from the data compared to the classical Radial Basis Function kernel, and thereby offer overall enhanced results?
2. Do randomized measurements and variable subsampling models maintain the performance of the quantum kernel derived through the inversion test, while diminishing the time complexity related to data size?

The evaluation engages synthetic data and the Credit Card Fraud data set, inspecting performance and time

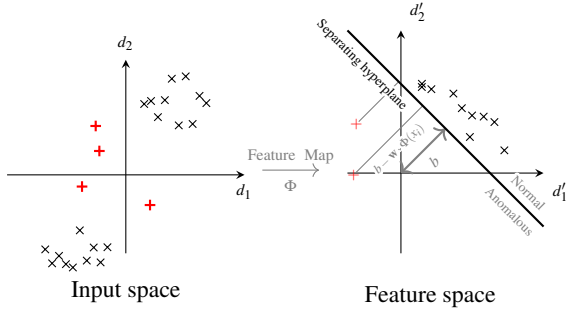


Figure 1: The linearly inseparable points on the input space are mapped using a quantum feature  $\Phi$  into a feature space where they are linearly separable (Aggarwal, 2017).

complexities contingent on data size and qubit number. Our findings, diverging from (Kyriienko and Magnusson, 2022), reveal that while attainable improvements in average precision and F1 score over the classical kernel are discernible, they are minimally significant. Models utilizing variable subsampling with the inversion test exhibited stability, whereas those employing the randomized measurement method presented high variance. Nevertheless, variable subsampling did manifest considerable enhancements in training and testing times, indicating potential performance elevation opportunities through alternate hyperparameters. All Code and Experiments can be found here<sup>1</sup>.

## 2 PRELIMINARIES

### 2.1 One-Class Support Vector Machines

The One-Class Support Vector Machine (OC-SVM), originally presented by (Schölkopf et al., 1999), offers an unsupervised variant of the SVM for anomaly and novelty detection. Differing from the conventional SVM, which finds a maximum margin hyperplane to distinguish normal from anomalous data, the OC-SVM presupposes that the origin represents the anomalous class when labels are absent. It seeks to maximize the margin  $b$  between origin and the input data, penalizing points below the hyperplane (Fig. 1). These are governed by a hyperparameter  $\nu \in (0, 1]$  which dictates the portion of points categorized as anomalous.

$$\min_{\mathbf{w}, b} \frac{1}{2} \mathbf{w}^2 + \frac{1}{\nu N} \sum_{i=1}^N \max\{b - \mathbf{w} \cdot \Phi(x_i), 0\} - b \quad (1)$$

<sup>1</sup><https://github.com/AfraeA/q-anomaly>

To address non-linearly separable data, feature maps  $\Phi : \mathcal{X} \rightarrow \mathcal{F}$  elevate data from input space  $\mathcal{X}$  to a higher dimensional feature space  $\mathcal{F}$ . Since direct calculations of feature maps are computationally intensive, a kernel function  $k : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$  is used to ascertain data point similarity in the embedded feature space.

$$k(x_i, x_j) = \langle \Phi(x_i), \Phi(x_j) \rangle \quad (2)$$

The use of the kernel function, which can be summed up in matrix form by the kernel matrix  $G = [k(x_i, x_j)]$ , facilitates a dual problem formulation. Solving this dual problem leads to an implicit parameterization of the hyperplane based on the support vectors  $\alpha_i$ .

Relying on this parameterization, scoring new data depends on its relation to the separating hyperplane:

$$\text{Score}(x_{\text{new}}) = \sum_{i=1}^N \alpha_i \cdot k(x_{\text{new}}, x_i) \quad (3)$$

The label is determined by the score's sign: negative connoting an anomaly and positive signifying normalcy.

### 2.2 Quantum Kernel Embedding

Quantum feature maps allow for the embedding of classical data into quantum states within Hilbert space  $\mathcal{H}$  through the use of data-dependent quantum unitary gates  $U_\Phi(x)$ , mathematically expressed as  $|\Phi(x)\rangle = U_\Phi(x)|0\rangle$ . The IQP-like (Instantaneous Quantum Polynomial) feature map, recognized for being hard to simulate classically (Havlíček et al., 2019), encodes a  $d$ -dimensional input  $x_i$  into  $d$  qubits as follows:

$$|\Phi(x_i)\rangle = U_Z(x_i) H^{\otimes d} U_Z(x_i) H^{\otimes d} |0^d\rangle, \quad (4)$$

$$U_Z(x_i) = \exp\left(\sum_{j=1}^d \lambda x_{ij} Z_j + \sum_{j=1}^d \sum_{j'=1}^d \lambda^2 x_{ij} x_{ij'} Z_j Z_{j'}\right),$$

where  $\lambda$ , influenced by data reuploading counts, impacts the kernel bandwidth analogously to  $\gamma$  (Shaydulin and Wild, 2022). We refer to FigureFig. 2a for a visual representation. Upon data encoding, fidelity, distilled to the overlap of states for pure quantum states as  $F(x, x') = |\langle \Phi(x') | \Phi(x) \rangle|^2$ , quantifies data similarity.

Two prominent methods for measuring fidelity are the inversion and swap tests. The inversion test, detailed in FigureFig. 2b, calculates the overlap of two data points' pure quantum states with  $O(n^2)$  kernel evaluations, but necessitates unitary feature maps and

yields deeper circuits. Conversely, the swap test, depicted in Figure Fig. 2c, is applicable to both pure and mixed states but necessitates wider circuits. It is based on the swap trick (Hubregtsen et al., 2022), which deduces the inner product from the tensor product of density matrices  $\rho_i$  and  $\rho_j$  utilizing a swap gate  $\mathbb{S}$ , expressed in Eq. 5:

$$\text{Tr}(\rho_i \rho_j) = \text{Tr}(\mathbb{S} \rho_i \otimes \rho_j). \quad (5)$$

### 2.3 Time Complexity for Quantum Kernels

Quantum one-class Support Vector Machines provide a nuanced method for detection but are hindered by formidable time complexity, particularly given the existing operational frequencies of quantum hardware. The intrinsic quadratic time complexity of one-class SVMs is significantly amplified in a quantum computing context, requiring a substantial number of repetitions (shots) to accurately measure fidelity between points due to its probabilistic nature; typically, at least 1000 shots per circuit measurement are necessary to obtain replicable results. Training a quantum one-class SVM on a large dataset, exemplified by the 284,000 instances from the Credit Card dataset in our experiments, would theoretically necessitate approximately  $4 \times 10^{10}$  kernel function value calculations to construct the symmetrical kernel matrix. This implies a staggering requirement of  $4 \times 10^{13}$  shots, which, with a measurement rate of 5kHz (Haug et al., 2021), equates to a minimum training time of 255 years using a swap or inversion test kernel.

Reducing the dataset may abbreviate training time but risks degrading algorithmic performance and stability due to decreased representativity of training samples. This diminution can result in less dependable support vectors and decision functions, jeopardizing the reliability and consistency of the one-class SVM, especially when training data substantially deviates from the overall distribution. The challenge also permeates inference times, as predicting a new point demands evaluating the kernel function against all training points, elongating detection times and hindering real-time applications like patient monitoring and fraud prevention.

To surmount these obstacles, optimization of kernel calculation algorithms and investigation into innovative quantum measurement techniques, which could minimize the requisite shots, are pivotal. Viable approaches may encompass adapting classical methods to minimize data needed for kernel matrix computations while preserving performance, employing clustering, and applying matrix decomposition and

approximation methods to avoid evaluating the kernel across the entire training set.

## 3 RELATED WORK

### 3.1 Quantum Anomaly Detection

This work augments the methodology propounded in (Kyriienko and Magnusson, 2022), amalgamating the one-class SVM with a computationally intricate kernel based on the IQP feature map (Figure Fig. 2a) to secure a 20% enhancement in average precision vis-à-vis their classical benchmark. Subsequent experiments herein adhere to this protocol as a quantum yardstick, exploring two strategies to diminish the time complexity relative to data size.

Hybrid quantum-classical models manifest as a prominent archetype in anomaly detection research. For instance, (Sakhnenko et al., 2022) innovatively refines the auto-encoder (AE) hidden representation by interfacing a parameterized quantum circuit (PQC) with its bottleneck, segueing into an unsupervised model post-training by substituting the decoder with an isolation forest, and vetting performance across multifarious data sets and PQC architectures. Concurrently, (Herr et al., 2021) pioneers an adaptation of the classical AnoGAN (Schlegl et al., 2017) by deploying a Wasserstein GAN, wherein the generator is substituted with a hybrid quantum-classical neural network, and subsequently trained via a variational algorithm.

Contrastingly, quantum annealing approaches, such as the QUBO-SVM presented in (Wang et al., 2022), reformulate the conventional SVM optimization predicament into a quadratic unconstrained binary optimization problem (QUBO) which is amenable to resolution via quantum annealing solvers. Although retaining the conventional SVM optimization problem, this methodology expedites accurate predictions through adept kernel function identification, thereby facilitating plausible real-time anomaly detection.

(Ray et al., 2022) explores hybrid ensembles, constructing an amalgamation of bagging and stacking ensembles from assorted quantum and classical components, each playing a pivotal role in the anomaly detection framework. The amalgamated quantum components encapsulate disparate variable quantum circuit architectures, kernel-based, and quantum annealing-based SVMs, while the classical constituents encompass logistic regression, graph convolutional neural networks, and light gradient boosting models. Despite superficial similarity to the variable subsampling utilized herein, it's noteworthy that the

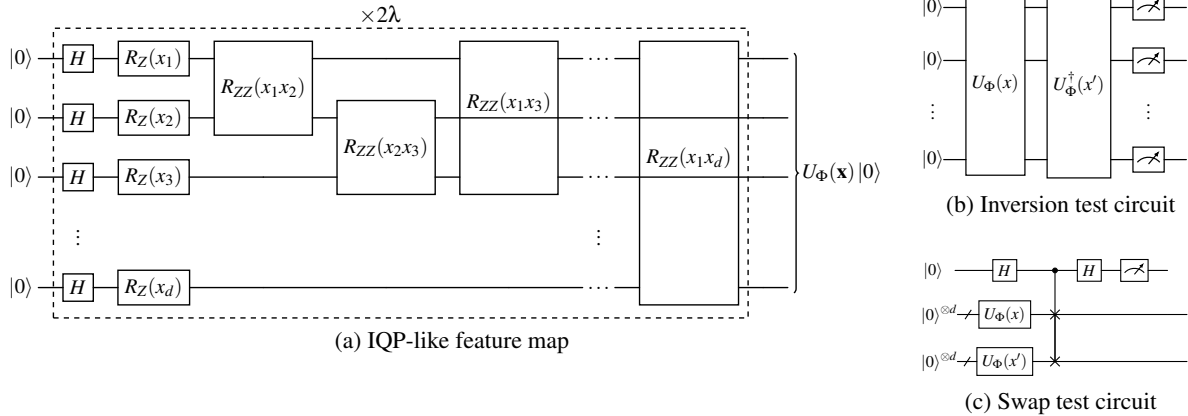


Figure 2: Quantum circuits for IQP-like feature map, inversion test and swap test.

latter method’s adoption of varying sub-sample sizes uniquely addresses the OC-SVM’s parameterization dependence.

### 3.2 Efficient Gram Matrix Evaluation

Quantum kernel methods, pivotal for various quantum machine learning applications, grapple with notable computational demands in matrix evaluations. Approaches to mitigate this complexity are: (i) *Quantum-Friendly Classical Methods*, reducing kernel matrix elements to evaluate, and (ii) *Quantum Native Methods*, minimizing the shot requirements yet necessitating classical post-processing, albeit feasibly parallelizable or vectorizable.

Randomized measurement kernels, pioneered by (Haug et al., 2021) and utilized with hardware-efficient feature maps, achieved an expedited kernel measurement while approximating Radial Basis Function (RBF) kernels, demonstrated via both synthetic and MNIST data. Conversely, the classical shadow method, proposed by (Huang et al., 2020), employs a similar quantum protocol but diverges in classical post-processing to provide classical state snapshots via the inversion of a quantum channel, often attaining reduced error in predicting second Rényi Entropy.

Variable subsampling, introduced by (Aggarwal and Sathe, 2015), and its sophisticated counterpart, variable subsampling with rotated bagging, offer an efficient ensemble training approach, leveraging varied sample sizes and rotational orthogonal axis system projections respectively. These methods not only confer computational efficiency but also harness an adaptive ensemble model training strategy, tested efficaciously with algorithms like local outlier factor (LOF) models and the k-Nearest Neighbors algorithm.

The Distance-based Importance Sampling and

Clustering (DISC) approach, by (Hajibabae et al., 2021), and Block Basis Factorization, by (Wang et al., 2019), represent variants of matrix decomposition based methods for kernel approximation. While DISC employs cluster centroids as landmarks to formulate approximation matrices and assumes kernel matrix symmetry, block basis factorization avails randomized spectral value decomposition on cluster samples and creates a smaller, computationally tractable inner similarity matrix, demonstrating superior performance relative to the k-means Nyström method.

## 4 APPROACHES

In the preceding review of related work, several methodologies of efficient gram matrix evaluation were highlighted. However, moving forward, we will focus on two specific approaches which can be applied to both symmetric training kernel matrices and asymmetric ones used during prediction. Although the classical shadows and block basis factorization methods satisfy this criterion, we have decided to explore randomized measurements and variable subsampling methods because of their intuitive conceptual frameworks.

### 4.1 Randomized Measurements Kernel

Expanding upon the method of randomized measurements, suggested for future exploration in (Kyriienko and Magnusson, 2022), and practically employed in kernel calculation for classification by (Haug et al., 2021), this method endeavors to adeptly meld linear and quadratic complexities concerning data size. This is achieved respectively through acquiring measurements of feature maps and subsequent classical post-



processing. The method notably diminishes the requisite quantum shots, thereby alleviating the overall computational burden.

The inception of the fidelity calculation was motivated by the possibility of conceiving the swap operator  $\mathbb{S}$  as a quantum twirling channel  $\Phi_N^{(2)}$  (Elben et al., 2019). Quantum twirling channels are operations commonly used in error correction. A 2-fold local quantum twirling channel applied to an arbitrary operator  $O$  is articulated by

$$\Phi_N^{(2)}(O) = \overline{(U^{\otimes 2})^\dagger O U^{\otimes 2}}, \quad (6)$$

with  $\overline{\dots}$  denoting the average over the unitaries  $U = \bigotimes_k^N U_k$  and  $U_k$  sampled from a unitary 2-design. A unitary t-design approximates property averages over all possible unitaries using a finite set, while the Haar measure provides a uniform sampling mechanism across unitary matrices. (Elben et al., 2019) demonstrate that the expectation value of applying such a twirling channel on an operator  $O$  acting on two copies of a quantum state  $\rho$  is obtainable from the probability products  $P_U$  resulting from the measurements:

$$\sum_{s,s'} O_{s,s'} \overline{P_U(s)P_U(s')} = \text{Tr}(\Phi_N^{(2)}(O) \rho \otimes \rho) \quad (7)$$

The coefficients  $O_{s,s'}$  specific to the swap operator are derived by employing Weingarten calculus of Haar random unitaries and Schur-Weyl duality (Roberts and Yoshida, 2017) to calculate the purity of the state  $\rho$ , yielding:

$$O_{s,s'} = d^N (-d)^{D(s,s')}. \quad (8)$$

A formula for quantum fidelities in terms of randomized measurement probabilities is obtained by utilizing the swap trick, then employing Eq. 6 with the coefficients from Eq. 7.

Local Haar random unitaries  $U_{\text{Haar}}$  are constructed by tensoring sampled unitary  $U_k \in SU(2)$  for each qubit. As illustrated in Fig. 3, each quantum circuit leverages a unitary  $U_\Phi$ , corresponding to the quantum feature map, and one of the  $r$  different local Haar random unitaries  $U_{\text{Haar}}$ . Every circuit requires  $s$  different shots. This yields  $r$  sets of strings  $s_A$  and their measurement probabilities  $P_U^{(i)}(s_A)$  for the different random basis rotations  $U_{\text{Haar}}$ . Post-processing harnesses the formula:

$$\begin{aligned} K(x_i, x_j) &= \text{Tr}(\rho_i \rho_j) \\ &= 2^d \sum_{s_A, s'_A} (-2)^{-D(s_A, s'_A)} \overline{P_U^{(i)}(s_A) P_U^{(j)}(s'_A)} \quad (9) \end{aligned}$$

The statistical error in fidelity measurement, quantified as  $\Delta G \approx \frac{1}{s\sqrt{r}}$ , necessitates error mitigation, par-

ticularly pertinent in noisy hardware scenarios. A facilitative aspect of the randomized measurements approach is its provision for straightforward error mitigation, involving purities calculated and recorded in the diagonal of the training kernel matrix and exploiting a minor computational overhead in the testing phase for the asymmetrical kernel matrix.

The quantum kernel calculation segment, employing randomized measurements, presents a time complexity of  $nrs$  ( $n$ : data size,  $r$ : basis rotation unitaries,  $s$ : shots per rotation), while classical post-processing demands  $n^2$  complexity, albeit with an unfavorable exponential time complexity concerning the number of qubits (features). Implementation aligning with (Haug et al., 2021) is accessible via the Large Scale QML<sup>2</sup> GitHub repository, whereas randomized measurement processing and combination utilize functionalities from the qc\_optim<sup>3</sup> repository. Accommodating the IQP-like feature map and enabling interim kernel copy retention and calculation time-keeping mandated the development of a novel implementation.

## 4.2 Variable Subsampling

Introduced by (Aggarwal, 2017), variable subsampling addresses sensitivities in the one-class SVM to kernel choice and hyperparameter  $\nu$  values by exploiting ensemble methods. Unlike bagging ensembles, variable sampling ensembles, while utilizing random subset selection for model training, employ varying sub-sample sizes, permitting sampling over parameter spaces, particularly those related to data size, like the expected anomaly ratio  $\nu$  in the one-class SVM.

Supposing a variable subsampling ensemble comprises 3 OC-SVMs trained with  $\nu = 0.1$  and sample sizes of 53, 104, and 230, differing support vector lower bounds among components yield varied decision boundaries. These can subsequently be combined in a manner that diminishes the bias or variance in predictions.

Ensemble construction commences with uniformly sampling  $c$  different subsample sizes between 50 and 1000. Data subsets, corresponding to sampled sizes, are randomly selected from the dataset and employed to train base model versions, here, the quantum one-class SVM with inversion test. Although subsamples may contain identical elements, each avoids reusing data points to better reduce variance. Predictions are calculated by combining normalized (to zero mean and unit standard deviation) decision functions of all components, considering

<sup>2</sup><https://github.com/chris-n-self/large-scale-qml>

<sup>3</sup>[https://github.com/chris-n-self/qc\\_optim](https://github.com/chris-n-self/qc_optim)

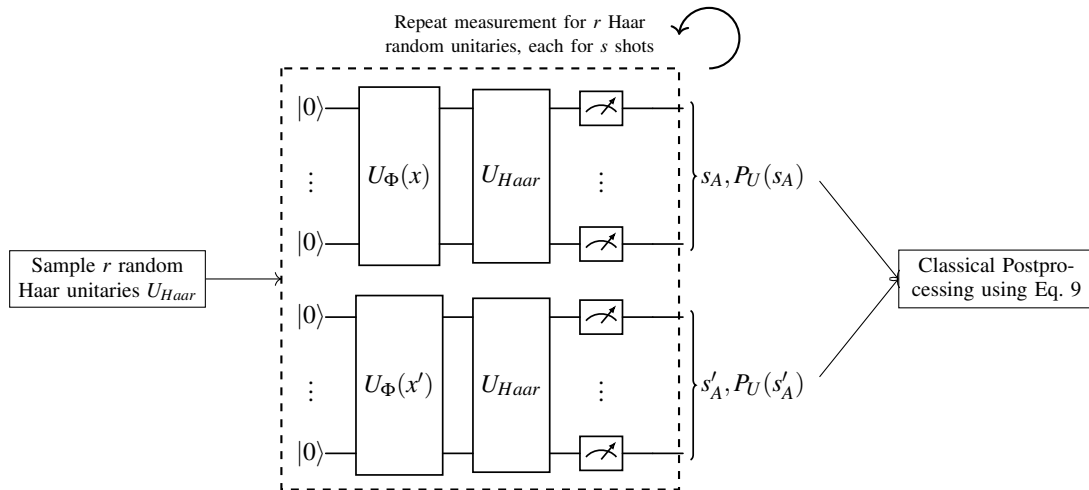


Figure 3: The protocol and the circuit architecture for calculating quantum kernel functions using randomized measurement.

each is trained with distinct data sizes, thus possessing varied decision function value ranges (Aggarwal and Sathe, 2015). Simple averaging of outlier scores is advantageous for reducing variance and superior performance with smaller datasets. Applying the maximum score, conversely, curtails bias but elevates variance. Post decision function value extraction, the threshold function  $\text{sgn}(\cdot)$  derives the class label.

More components and a higher maximum subsample size facilitate optimal variance reduction, at the expense of increased computational resource and time demands. However, this trade-off can be managed through wise hyperparameter selection. For example, we choose a maximum subsample size of 100, instead of the recommended 1000 points from the original research, and use  $\lfloor \frac{n}{100} \rfloor$  components instead of 100. While this may negatively affect performance, it provides insight into the ensemble’s behavior when scalability takes precedence.

The training phase exhibits approximately  $c \cdot (\frac{n_{\min} + n_{\max}}{2})^2$  time complexity, where  $c$  is the number of ensemble components, and  $n_{\min} = 50$  and  $n_{\max} = 100$  are the minimum and maximum subsample sizes respectively. Incorporating scalability modifications, this becomes  $\lfloor \frac{n}{100} \rfloor \cdot (\frac{50+100}{2})^2$ , indicating linear complexity. Testing similarly maintains linear time complexity, formulated as  $c \cdot \frac{n_{\min} + n_{\max}}{2} \cdot n_{\text{test}}$ , with  $n_{\text{test}}$  representing test samples.

## 5 EXPERIMENTAL SETUP

Ensuring the reproducibility of our experiments, this section meticulously delineates implementation aspects, encompassing pre-processing, various kernel calculation methodologies, and hyperparameter se-

lection. Two distinct experiment sets facilitate the comparative analysis of our approaches:

1. The first set aspires to examine performance, training, and testing durations with respect to data size. Primarily, it endeavors to contrast the computational efficiencies afforded by our methodologies, delineated in Section 4, and to comprehend model responses to elevated data volumes.
2. Focusing on the relationship between performance, computational time, and the number of features (or qubits), the second experiment set seeks not only to discern whether the methods impose detrimental effects on performance but also to elucidate the time feature/qubit relationship.

Experiments are conducted using 15 distinct seeds, ranging from 0 to 14.

### 5.1 Datasets

Our experiments utilize two datasets, distinguished by their synthetic or real-world origin, to investigate various methodologies. The synthetic dataset is employed exclusively in the first experiment set, given its two-dimensional nature and the associated limitations in exploring numerous features.

#### 5.1.1 Synthetic Data

Our synthetic dataset is a bi-dimensional, non-linearly separable dataset, derived by modifying an SKlearn OC-SVM demonstration<sup>4</sup> to yield training samples of diverse sizes. Testing samples consistently

<sup>4</sup>[https://scikit-learn.org/stable/auto\\_examples/svm/plot\\_oneclass.html](https://scikit-learn.org/stable/auto_examples/svm/plot_oneclass.html)

comprise 125 points, incorporating a 0.3 anomaly ratio.

### 5.1.2 Credit Card Fraud Data

The credit card fraud data<sup>5</sup> encompasses approximately 284,000 datapoints, with 492 classified as anomalous (class 1), across 31 features. Omitting 'time' and 'amount', 28 PCA-applied, anonymized numerical features are retained. Each seed corresponds to a unique data split. The size varies for the set of experiments exploring the effects of data size, while it is kept at a constant 500 data samples for the set of experiments exploring the effects of the qubit/feature number. Training data only contains non-anomaly samples while the test set always contains 125 points, which include 6 anomalies, achieving a 0.05 anomaly ratio.

## 5.2 Data Pre-Processing

Distinctive pre-processing methodologies were necessitated based on the quantum kernel measurement technique applied and the data type (synthetic or real).

**Radial Basis Function (RBF) Kernel.** For implementations utilizing the RBF kernel, standard scaling was executed post data partitioning into training and test subsets, ensuring zero mean and unit standard deviation across all features. Subsequently, Principal Component Analysis (PCA) was employed to consolidate the data into the requisite number of features.

**Inversion Test Kernels.** The application of inversion test kernels warranted an additional step following the pre-processing used for the RBF kernel. Given that data was utilized as rotation angles within the quantum circuit, a scaling by a factor of 0.1 was an imperative post-PCA application.

**Randomized Measurements Kernels.** Adhering to guidelines by (Haug et al., 2021), the randomized measurements kernels necessitated a unique rescaling approach. Post-PCA, a second round of standard scaling was administered, succeeded by an additional rescaling using factor  $\frac{1}{\sqrt{M}}$ , with  $M$  representing the post-PCA data dimensionality.

Note that while real data pre-processing was adapted in accordance with the specific quantum kernel measurement technique deployed, synthetic data

was subjected to pre-processing solely when implementing the randomized measurement kernel, only requiring a step of standard scaling with an additional rescaling by factor  $\frac{1}{\sqrt{M}}$ .

## 5.3 Baselines

For the real dataset, we aim to replicate the one-class SVM results from (Kyriienko and Magnusson, 2022), adopting them as quantum and classical benchmarks. Lacking detailed insight into their sampling method and explicit test set sizes, we employ random sampling to generate data sets comprising 500 training and 125 test data points, training the OC-SVM solely on non-anomalous data, like the original authors. An anomaly ratio of 0.05 is maintained in the test set.

The experiments concerning model responses to various data sizes retain these architectures, holding the number of PCA features steady while manipulating data sizes. Specifically, we utilize 2 features for synthetic and 6 for real data sets, exploring data sizes  $n \in \{250, 500, 750, 1000, 1250, 1500\}$ .

## 5.4 Models and Parameter Selection

Employing both classical and quantum versions of the OC-SVM, we utilize `OneClassSVM` from the `SKlearn`<sup>6</sup> library. All classical models adopt the RBF kernel with  $\gamma = \frac{1}{N \cdot \text{Var}(M)}$  and a consistent  $\nu = 0.1$ .

Quantum circuits, pivotal for kernel calculations, are crafted with the `qiskit`<sup>7</sup> library, simulated via `qiskit_aer.QasmSimulator`. Employing  $\lambda = 3$  data reuploadings for all quantum methods' feature maps, inversion test kernel elements are determined using 1000 shots each.

Randomized measurement circuits, realized with  $r = 30$  measurement settings and subjected to  $s = 9000$  shots each, leverage crucial classical post-processing, which emphasizes minimal embedded loops and prioritizes efficient matrix operations. The variable subsampling method, utilizing  $c = \lfloor \frac{n}{100} \rfloor$  components ( $n$  representing the training set size) and a subsample size  $n_s \in [50, 100]$  to ensure scalable model performance, invokes the inversion test for kernel calculations with 1000 shots per element. Employing a consistent  $\nu$  value, the desired kernel matrix calculation function is passed as a callable kernel parameter, designed to accept two data sets and method-specific hyper-parameters, subsequently returning the computed kernel matrix.

<sup>5</sup><https://www.kaggle.com/datasets/mlg-ulb/creditcardfraud>

<sup>6</sup><https://scikit-learn.org/stable/modules/generated/sklearn.svm.OneClassSVM.html>

<sup>7</sup><https://qiskit.org/>

## 5.5 Performance Metrics for Imbalanced Data Sets

Analyzing highly imbalanced data sets, prevalent in anomaly detection, necessitates alternative metrics to accuracy due to its incapability to reflect model performance accurately across classes (Aggarwal, 2017). Hence, derivatives of precision and recall like the F1 score and average precision gain prominence.

**F1 Score:** represents the harmonic mean of precision and recall, providing a balanced perspective on model performance regarding false positives and negatives. It is defined as

$$\text{F1 score} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}. \quad (10)$$

**Average Precision:** quantifies the model's capability to discern anomalies, irrespective of threshold, by measuring the area under the precision-recall curve. Specifically,

$$\text{AP} = \sum_k [\text{Recall}(k) - \text{Recall}(k+1)] \cdot \text{Precision}(k). \quad (11)$$

An average precision equal to the data's anomaly ratio signifies a model with no anomaly detection capability, while a score of 1 indicates a flawless detector. Consequently, our analysis prioritizes average precision, complemented by precision, recall, and F1 score insights.

## 6 RESULTS

### 6.1 Performance Analysis Synthetic Dataset

The performance results utilizing synthetic data are presented in the first column of Fig. 4a. Quantum methods demonstrate superior average precision compared to classical Radial Basis Function (RBF) models with the synthetic data. Despite this, given the anomaly ratio of 0.3, most methods underperform relative to a random detector, with inversion test and variable subsampling (using the maximum score) as exceptions. Variable subsampling not only surpasses randomized measurements in average precision but also exhibits comparable variance in ensemble results across different combination functions. Notably, it presents augmented results with the maximum over the average function for data sizes exceeding 750 points.

Upon applying the sign as a threshold, classical models outperform quantum models with the RBF kernel models securing an approximate F1 score of 0.88 and demonstrating lower variance. The randomized measurements and inversion test models follow in performance, with versions of the variable subsampling ensemble trailing. Despite the mitigated version of randomized measurements models achieving superior precision, F1 scores, and recall values were alike between mitigated and unmitigated versions. Contrarily, variable subsampling employing the maximum score delivers superior precision but inferior recall compared to its average score counterpart, suggesting its unsuitability for anomaly detection using the current threshold.

Although results might appear conflicting, disparities may stem from suboptimal threshold selection for quantum methods. Given that average precision operates independently of a threshold, findings might imply an optimal threshold where quantum methods supersede classical RBF.

### 6.2 Training and Testing Durations Synthetic Dataset

Fig. 5a depicts the notable efficiency of the variable subsampling method in reducing training durations, achieving the lowest among all employed quantum methods for both VS Max and VS Average. Interestingly, varied score functions do not impact the time consumption. Variable subsampling not only realizes linear time but also accomplishes a  $\sim 95\%$  reduction in training time with 1500 data points, suggesting potential similar results in significantly less time given their average precision closely mirroring that of an individual one-class SVM utilizing the inversion test (Section 2). Furthermore, a  $\sim 25\%$  enhancement in testing time is observed at 1500 training data points.

Mitigated and unmitigated randomized measurements methods yielded analogous times, corroborating that error mitigation imposes minimal overhead. While randomized measurements facilitate substantial training time reduction, complexity is non-linear relative to data size, attributed to the inclusion of classical post-processing in training time, which demands quadratic time complexity. Albeit the training time reduction from the randomized measurement method falls below that of variable subsampling, it yet provides  $\sim 35\%$  and  $\sim 12\%$  reduced testing times compared to the inversion test and variable subsampling respectively at 1500 training data points.



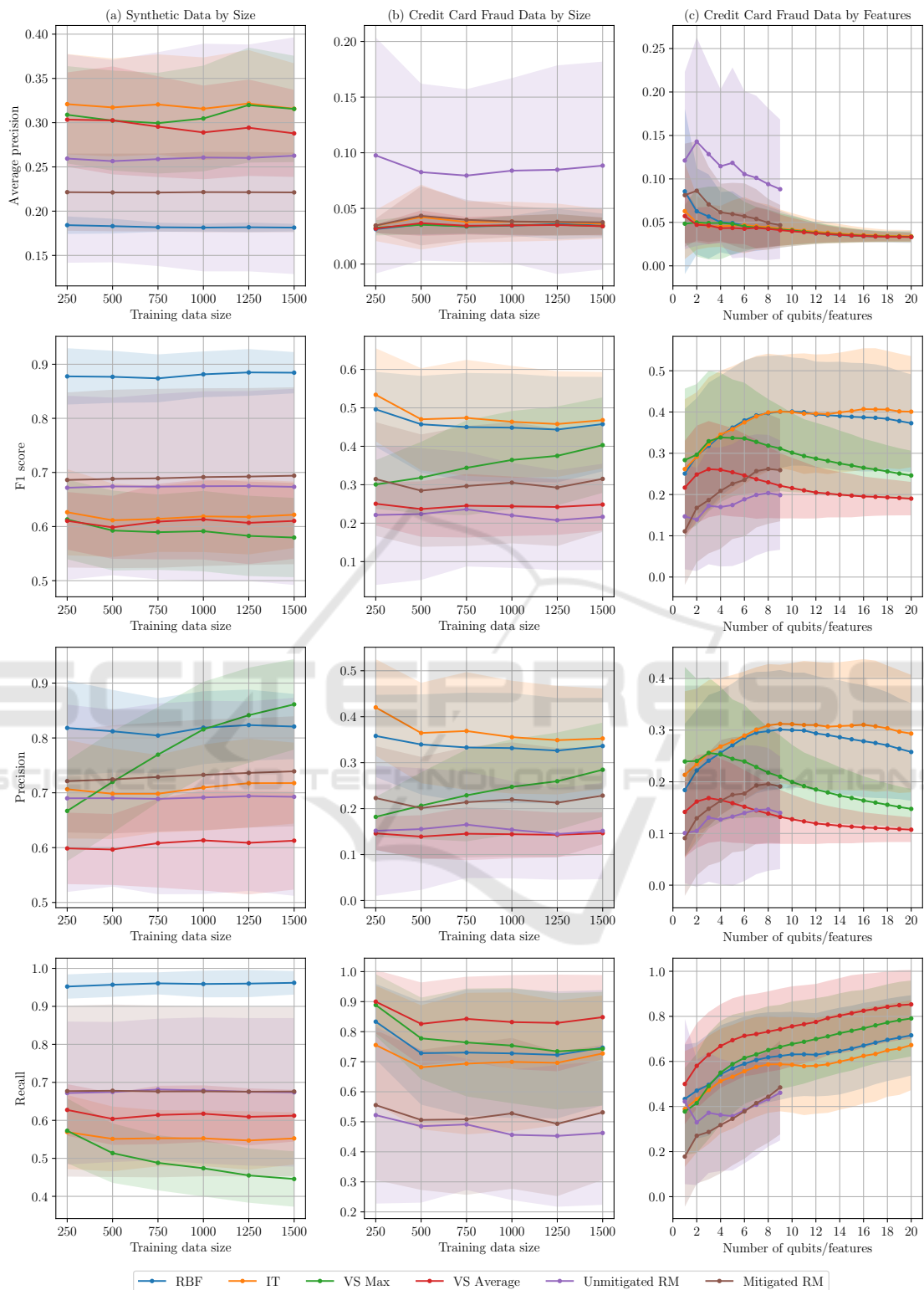


Figure 4: The performance of the models trained using the different methods. The first column of the figure corresponds to the synthetic dataset, used with 2 features and various data sizes. The second and third columns correspond to the Credit Card dataset, with 6 as the number of features using various data sizes, then used with various numbers of features and a constant data size of 500.

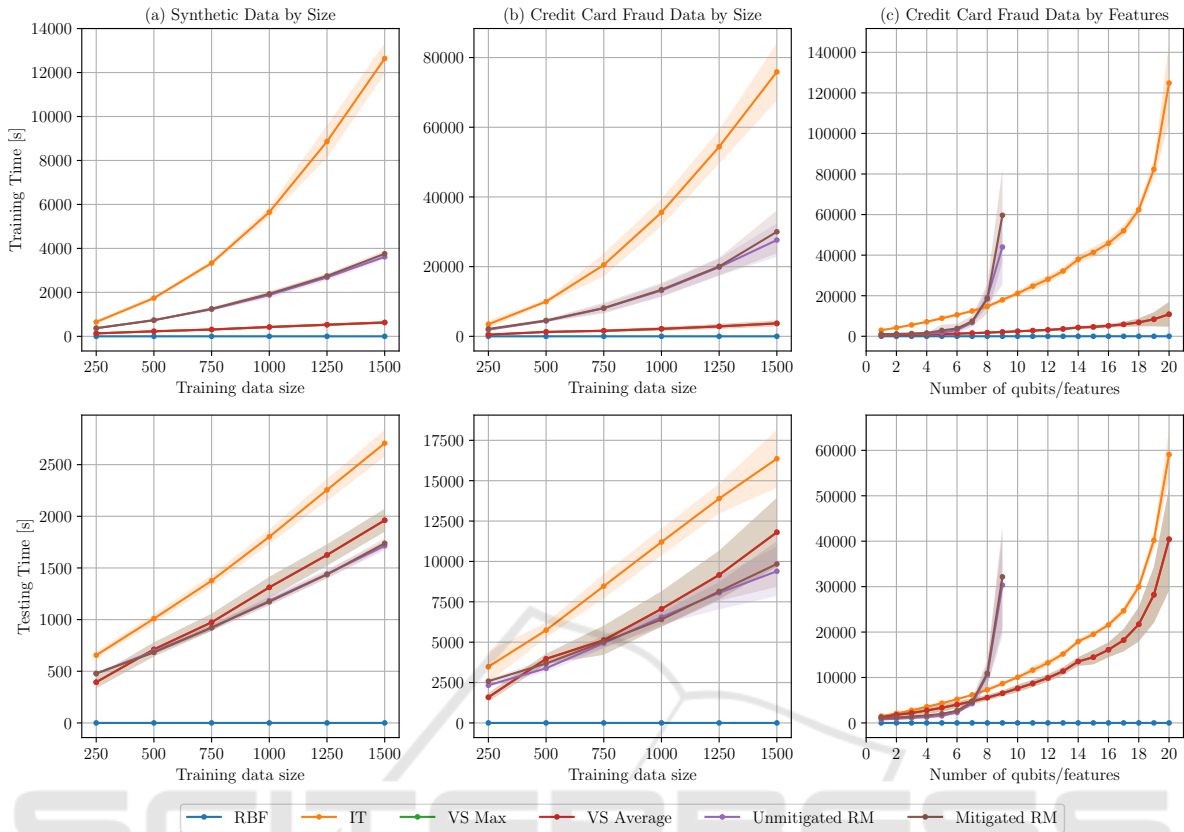


Figure 5: Training and testing durations in seconds, based on the data size.

### 6.3 Performance Analysis Relative to Data Size Credit Card Data

Fig. 4b reveals a predominant equivalence to a random detector across most models, with the unmitigated randomized measurements kernel being a notable exception. Despite this hinting at a potential performance threshold, the model’s notable variance signifies an inherent instability.

At the utilized threshold, the inversion test and classical RBF exhibit comparable F1 scores in Fig. 4b, succeeded by variable subsampling models employing maximum scoring, which seem to exhibit an enhancement with increased data usage. Despite affording the highest recall, both variable subsampling variations grapple with suboptimal precision.

Meanwhile, both randomized measurement kernels, under the current threshold, struggle with notably low recall, undermining their utility for anomaly detection.

### 6.4 Time Complexity Relative to Data Size Credit Card Data

In Fig. 5b, we demonstrate that, analogous to synthetic data outcomes, employing variable subsampling assures linear time complexity and substantially curtails training time. Furthermore, randomized measurements realize up to a 50% reduction in training time, even while tripling data utilization. Both variable subsampling and randomized measurements maintain comparable, and notably reduced, inference times relative to the inversion test method.

### 6.5 Performance Analysis Concerning Qubit Number - Credit Card Fraud Dataset

In our endeavor to replicate the results from (Kyriienko and Magnusson, 2022), we also explore additional metrics including precision, recall, and F1 score, employing thresholds in model evaluation. The original study lacks explicit data selection methodology for training and testing, along with absent details on the number of runs, prompting us to

opt for a uniformly random data selection.

Our findings, notably divergent from (Kyriienko and Magnusson, 2022), are presented in Fig. 4c.

Average Precision reveals considerably less favorable outcomes for classical RBF and quantum inversion test than the aforementioned study, particularly over 15 different runs, underperforming even a random detector. Increasing feature/qubit usage evens the average precision of the inversion test kernel, utilizing the IQP-like kernel, to that of the classical RBF, suggesting the original paper’s results might hinge more on beneficial sample selection for the quantum IQP-like kernel rather than demonstrating a clear quantum advantage. Notably, models exhibit enhanced stability—diminishing average precision variance—with augmented feature/qubit use.

Thresholded metrics indicate marginal improvement with inversion tests over classical methods, yet a similar F1 score is attainable via the classical kernel with fewer qubits. Variable subsampling, while offering the highest recall and consistent results, underachieves in F1 score due to suboptimal precision—potentially attributable to conservative hyperparameter selections. An exploration into its performance with increased components and a larger maximum subsample could be interesting.

## 6.6 Time Complexity Concerning Qubit Number - Credit Card Fraud Dataset

Fig. 5c illustrates a quadratic relationship between the training time and the number of qubits for the inversion test. Despite a similar relationship, variable subsampling effectively diminishes training time, potentially attributed to its ensemble components utilizing the inversion test independently of qubit number, constrained by limited and non-correlated component and data sizes. Conversely, the randomized measurements method sees escalated training times with 8 or more qubits, hindering its utility for high-dimensional data sets. A rise in testing times is also observed with increasing qubit numbers.

## 7 CONCLUSION

This work explores two methodologies aimed at enhancing the scalability of the quantum one-class SVM in a semi-supervised framework: (1) the randomized measurements and (2) variable subsampling ensemble methods. The former, inherently quantum, realizes linear quantum time complexity but demands

a quadratically complex classical post-processing based on data size. It generates kernel matrices by combining measurements from quantum feature maps of the data, executed through a randomized scheme. The latter adopts an ensemble approach, training multiple base model instances on varied-sized data subsets.

Two experimental datasets, synthetic and Credit Card, were deployed, revealing a marginal average precision improvement over the classical RBF for all methods in the synthetic data. Discrepancies emerged between outcomes from the Credit Card Fraud dataset and those in (Kyriienko and Magnusson, 2022). While models utilizing the inversion test and variable subsampling approximated the RBF and were relatively stable, those yielding higher average precision, like unmitigated randomized measurements, exhibited instability.

Future research trajectories include exploring the integration of variable subsampling with randomized measurements kernels and employing randomized measurements or variable subsampling alongside alternative kernels, such as learnable ones. Incorporating importance sampling in selecting the Haar random unitaries (Rath et al., 2021), as suggested by (Haug et al., 2021), could refine the randomized measurements method. The classical shadow method, due to its lower average error compared to randomized measurements, and the examination of Nyström and block basis factorization approximations of a quantum kernel, warrant further exploration. Finally, enhancing the quantum SVM’s performance through the proposed variable sampling method, by utilizing more components and enlarging maximum subsample sizes, remains a viable avenue, given the resultant training and testing time reductions.

## ACKNOWLEDGEMENTS

This research is part of the Munich Quantum Valley, which is supported by the Bavarian state government with funds from the Hightech Agenda Bayern Plus.

## REFERENCES

- Aggarwal, C. C. (2017). *An introduction to outlier analysis*. Springer.
- Aggarwal, C. C. and Sathe, S. (2015). Theoretical foundations and algorithms for outlier ensembles. *ACM SIGKDD explorations newsletter*, 17(1):24–47.
- Boloka, T., Crafford, G., Mokuwe, W., and Eden, B. V. (2021). Anomaly detection monitoring system for healthcare. In *2021 Southern African*

- Universities Power Engineering Conference/Robotics and Mechatronics/Pattern Recognition Association of South Africa (SAUPEC/RobMech/PRASA)*, pages 1–6.
- ECB (2023). Report on card fraud in 2020 and 2021.
- Elben, A., Vermersch, B., Roos, C. F., and Zoller, P. (2019). Statistical correlations between locally randomized measurements: A toolbox for probing entanglement in many-body quantum states. *Physical Review A*, 99(5):052323.
- Fernando, T., Gammulle, H., Denman, S., Sridharan, S., and Fookes, C. (2021). Deep learning for medical anomaly detection – a survey.
- Hajibabae, P., Pourkamali-Anaraki, F., and Hariri-Ardebili, M. A. (2021). Kernel matrix approximation on class-imbalanced data with an application to scientific simulation. *IEEE Access*, 9:83579–83591.
- Haug, T., Self, C. N., and Kim, M. (2021). Large-scale quantum machine learning. *arXiv e-prints*, pages arXiv–2108.
- Havlíček, V., Córcoles, A. D., Temme, K., Harrow, A. W., Kandala, A., Chow, J. M., and Gambetta, J. M. (2019). Supervised learning with quantum-enhanced feature spaces. *Nature*, 567(7747):209–212.
- Herr, D., Obert, B., and Rosenkranz, M. (2021). Anomaly detection with variational quantum generative adversarial networks. *Quantum Science and Technology*, 6(4):045004.
- Huang, H.-Y., Kueng, R., and Preskill, J. (2020). Predicting many properties of a quantum system from very few measurements. *Nature Physics*, 16(10):1050–1057.
- Hubregtsen, T., Wierichs, D., Gil-Fuster, E., Derks, P.-J. H., Faehrmann, P. K., and Meyer, J. J. (2022). Training quantum embedding kernels on near-term quantum computers. *Physical Review A*, 106(4):042431.
- Kyriienko, O. and Magnusson, E. B. (2022). Unsupervised quantum machine learning for fraud detection. *arXiv preprint arXiv:2208.01203*.
- Rath, A., van Bijnen, R., Elben, A., Zoller, P., and Vermersch, B. (2021). Importance sampling of randomized measurements for probing entanglement. *Physical Review Letters*, 127(20).
- Ray, A., Guddanti, S. S., Ajith, V., and Vinayagamurthy, D. (2022). Classical ensemble of quantum-classical ml algorithms for phishing detection in ethereum transaction networks. *arXiv preprint arXiv:2211.00004*.
- Roberts, D. A. and Yoshida, B. (2017). Chaos and complexity by design. *Journal of High Energy Physics*, 2017(4).
- Sakhnenko, A., O’Meara, C., Ghosh, K. J., Mendl, C. B., Cortiana, G., and Bernabé-Moreno, J. (2022). Hybrid classical-quantum autoencoder for anomaly detection. *Quantum Machine Intelligence*, 4(2):1–17.
- Schlegl, T., Seeböck, P., Waldstein, S. M., Schmidt-Erfurth, U., and Langs, G. (2017). Unsupervised anomaly detection with generative adversarial networks to guide marker discovery.
- Schölkopf, B., Williamson, R. C., Smola, A., Shawe-Taylor, J., and Platt, J. (1999). Support vector method for novelty detection. *Advances in neural information processing systems*, 12.
- Shaydulin, R. and Wild, S. M. (2022). Importance of kernel bandwidth in quantum machine learning. *Physical Review A*, 106(4).
- Wang, H., Wang, W., Liu, Y., and Alidaee, B. (2022). Integrating machine learning algorithms with quantum annealing solvers for online fraud detection. *IEEE Access*, 10:75908–75917.
- Wang, R., Li, Y., Mahoney, M. W., and Darve, E. (2019). Block basis factorization for scalable kernel evaluation. *SIAM Journal on Matrix Analysis and Applications*, 40(4):1497–1526.