The Workings of Quantum Anomaly Detection(One-class Classifiers)

Isabel Maria Adrover Cabot, Christos Koromilas,
Ismael Gómez Garrido, Gokul Jayakumar, Ivana Pallister

March 2023

Abstract

Quantum computing has sparked a global frenzy, captivating minds with its immense potential to transform the world of computation. Among its many promises, anomaly detection stands tall as a domain ripe for disruption. Why, some might ask? The answer lies in this technology's potential speed and precision, enabling us to identify elusive and extraordinary data points or behaviors. In this project, we explore the current landscape of anomaly detection methods, pitting the quantum realm against its classical counterpart. Our quest involves comparing tried-and-true methodologies, including kernel PCA, support vector machines, and fully connected autoencoders, against their quantum counterparts — the variational quantum classifier as one class classifier and the quantum support vector machine. As our investigation unfolds, a revelation emerges, a noticeable implementation gap separating these methodological categories. The scales tip overwhelmingly in favor of classical methods, their established prowess shining through. However, as we delve deeper into the intricacies of complex and inseparable data, a lingering question emerges: Could quantum anomaly detection techniques hold a hidden advantage for specific data types intertwined with quantum elements?

1 Introduction

Anomaly detection plays a vital role in various fields, such as finance, cybersecurity, and healthcare, where identifying unusual data points or behavior is crucial. particularly, One-Class Classification (OCC) algorithms have proven to be a successful approach to detecting anomalies in large datasets where the ratio of abnormalities is low, and the nature of this is unknown.

In recent years, quantum computing has garnered significant attention and generated a buzz across scientific and technological communities. This emerging discipline, rooted in the principles of quantum mechanics, holds immense promise for revolutionizing computation as we know it. By harnessing the peculiar behavior of quantum bits, or qubits, quantum computers possess the potential to solve complex problems at an unprecedented speed. Quantum Machine Learning (QML) has also emerged as a promising technique due to its ability to significantly reduce computational complexity, enabling the efficient solution of complex problems that are currently intractable using classical computing methods. QML has the potential to significantly improve the performance of anomaly detection algorithms by enabling faster and more accurate predictions, identifying patterns and relationships that classical methods may overlook, and reducing the energy and resources required for machine learning tasks.

This project aims to explore the current state of one-class classification algorithms in a simulated quantum computing environment. Studying how anomaly detection with one-class classifiers compares using quantum and classical algorithms is vital to gain insights into the capabilities and limitations of this technology.

2 State of the Art

OCC, as commented above, can be a good approach for Anomaly Detection. That is why trying to implement a Quantum One-Class Classifier (QOCC) has caught the attention of many researchers due to the possible benefits and possible favorable results that can be achieved with it. In this section we will comment some studies of the advances and alternatives presented for OCC for anomaly detection.

In [1] the authors created the quantum version of two usual classical machine learning methods, Principal Component Analysis (PCA) and Support Vector Machine (SVM). A detailed description of the algorithms is presented for pure and mixed states. The anomaly detection can be achieved using resources only logarithmic in the dimension of the quantum states. In addition, for pure states, the resources can also be logarithmic in the number of training quantum states used.

In this paper [2] a new quantum classifier is developed, the authors call it Quantum One-Class Classifier (QOCC). This method is based on the Hadamard Classifier (HC) [3] and the probabilistic quantum memory concept. They conclude that QOCC is more accurate and efficient than the HC. Additionally, for binary classification this new algorithm shows competitive results compared with classical classifiers. This approach can also be used to classify quantum data.

In this study [4] the authors present a new algorithm which consist in a Quantum Autoencoder (QAE) connected to a Parametrized Quantum Circuit (PQC). They tested the algorithm with a semi-supervised dataset that does not require decoding. Hence, the optimization of the circuit was more efficient in terms of computational resources. They obtained similar results for OCSVM and PCA methods and, in most cases, the outperformed DCAE.

Neural networks and, more specifically, Generative Adversarial Neural Networks (GANs) are very used in classical machine learning for image generation and anomaly detection. GANs consist of two different networks that compete one with each other, one generating and one discriminating. In [5] the authors take this idea into the quantum field building a variational quantum-classical Wasserstein GANs (WGANs). The resulting model is hybrid quantum-classical and was tested on anomaly detection showing alike results with classical methods regarding the F1 score.

In conclusion, many papers have proved the efficiency and high performance of the QOCC algorithms. However, further research is needed to deal with the problems that high dimensionality and big amounts of data present. Another field of research is to know with what kind of data these algorithms can offer good and reliable results.

3 Objectives and Motivation

This research project aims to achieve several objectives in the context of anomaly detection and the practical comparison of classical and quantum methods using three diverse datasets. Firstly, it compares the performance of one-class classification algorithms using classical and quantum computing techniques. Through this comparison, the project aims to assess the efficiency and accuracy of quantum machine learning (QML) in solving complex anomaly detection problems. Furthermore, the research aims to evaluate the potential improvements offered by QML at its current state compared to classical machine learning approaches. By conducting a practical comparison using the selected datasets, the project aims to provide insights into the effectiveness and limitations of quantum computing in anomaly detection. The ultimate goal is to contribute to understanding the practical implications of utilizing quantum computing techniques, offering recommendations, and highlighting the advantages of quantum computing while recognizing areas where classical methods may still prevail.

4 Research Questions

In this section, We aim to present the research questions that We want to be able to answer after the project:

- How effective are quantum one-class classification methods in comparison to classical one-class classification methods?
- How computationally complex are quantum one-class classification methods in comparison to classical one-class classification methods?
- How do different datasets affect the performance of a quantum one class classification method?
- How does the quantum approach perform in comparison to classical methods depending on the number of parameters?
- What benefits are apparent by using quantum computing with one classification?

5 Implementation and Approach

5.1 Selection of Datasets

The primary objective of this project is to address the anomaly detection problem, we wanted to select datasets that demonstrate precise performance in One-class classification scenarios. We will employ three popular datasets for anomaly detection: the Iris dataset, the Breast Cancer dataset, and the Vowel dataset. These datasets have been widely used in anomaly detection research and provide valuable insights into the performance of quantum machine learning on anomaly detection.

The Iris dataset offers a comprehensive collection of iris flower measurements, enabling us to identify anomalies in the floral characteristics. This dataset is separable and does not possess a large number of features. With these characteristics, the performance with quantum methods is straightforward and has helped us better understand how the quantum machine learning model is performing.

Following the research article [6], where They conducted an in-depth analysis to identify datasets that produce superior results for this specific problem, the authors assert that the effective sample size, denoting the ratio between the number of objects and the dimensionality, along with the average performance measured by the Area Under the Curve (AUC), serve as key variables to explain the variance in real-world one-class datasets. They analyzed 101 datasets, one of which is the already mentioned "Iris dataset", claiming that this dataset shows a high sample size and that the data is separable, as we commented before.

Motivated by these findings, we chose other datasets mentioned in the paper that possess similar characteristics to the "Iris dataset" but have a higher dimensionality. Consequently, we opted for the "Vowel Dataset" and performed anomaly detection on Vowel 1. The vowels in this dataset are interconnected, and it may be more difficult to seperate certain vowels. After researching the dataset we have decided to use vowel 1.

For our third dataset, we choose a dataset with a higher sample size and reduced separability compared to the others, the "Breast Cancer dataset". This dataset provides crucial information about breast cancer tumors, allowing us to detect any irregularities in tumor patterns.

By using these datasets, we aim to facilitate a comprehensive evaluation and draw meaningful conclusions regarding the different methods under consideration.

Table 1: Dataset Information

Dataset	Number of samples	Number of features
Iris	150	4
Vowel1	90	12
Vowel2	90	12
Breast Cancer	569	30

5.2 Classical methods

5.2.1 Support Vector Machine

In our study, we have incorporated the One-Class SVM as one of the baseline methods to compare the quantum version of this approach. This method creates a "decision boundary," which is essentially the optimal line that maximally separates the classes. To achieve this, we have utilized the kernel trick, which allows us to transform the data into higher dimensions where linear separation becomes feasible.

To ensure the best performance for each dataset, we conducted various experiments with different parameters. These experiments were carried out using a GridSearchCV with a cross-validation value of 5. The parameters that we check were the kernel, nu, and gamma.

Firstly, we explored different kernels such as linear, polynomial, rbf, and sigmoid, and selected the kernel that produced the most favorable output. Secondly, we conducted tests with varying values for the parameter 'nu' in the 'OneClassSVM()' function. This parameter represents an upper bound on the fraction of training errors and a lower bound on the fraction of support vectors. Determining the optimal 'nu' requires careful consideration of the trade-off between overfitting and underfitting. Setting a high value tends to result in underfitting, while setting a low value tends to lead to overfitting. Finally, we evaluated different values for the gamma parameter, which represents the kernel coefficient for 'rbf', 'poly', and 'sigmoid' kernels.

By performing these experiments and parameter tuning, we aimed to optimize the performance of the One-Class SVM method and obtain the most accurate results for our analysis.

5.2.2 Kernel PCA

Another classical method is the Kernel-PCA. Kernel PCA applies a nonlinear kernel function to the input data and then applies standard PCA in this transformed space. This makes Kernel PCA capable of capturing complex, nonlinear relationships in the data. We implemented Kernel PCA based on Hoffman's Novelty Detection[7]. Novelty detection is a One-Class classification and works by initially training the model with only inliers, which are considered normal data points. This process involves first fitting the data with Kernel PCA, transforming it into a feature space, and then reconstructing it back to the original space. The effectiveness of Kernel PCA lies in the calculation of reconstruction errors both in the training and test datasets, which represent the novelty scores. The rationale behind using reconstruction error in feature space as a measure of novelty is that inliers will typically have smaller reconstruction errors compared to anomalies. This is because the kernel PCA model is trained only on inliers, hence it is optimized to accurately reconstruct these instances. The detection of anomalies is accomplished by establishing a threshold based on this function:

$$t = serr[n - o - 1] \tag{1}$$

where,

- t represents the threshold for distinguishing normal data from anomalies.
- $serr$ is the sorted array of reconstruction errors in ascending order.
- n denotes the total count of training instances.
- o stands for the number of known outliers in the training set.

Given the training data, actual labels, reconstruction errors, and a contamination ratio (almost 0), determines the threshold of reconstruction error beyond which data points are considered anomalies. Finally, the test data is classified as either normal or anomalous. This classification is performed by comparing the reconstruction error of each data point with the previously determined threshold. Data points with a reconstruction error greater than the threshold are classified as anomalies, while others are classified as normal.
The final stage involves the implementation of a grid search to ascertain the optimal parameters that yield the most effective results. We employ the grid search technique specifically for Radial Basis Function (rbf) and polynomial (poly) kernels. Furthermore, we experiment with varying gamma values to enhance the performance of the model. This strategic approach is key to fine-tuning the model and driving its overall performance to the peak potential.

5.2.3 Fully Connected Autoencoder

In anomaly detection, fully connected autoencoders have gained considerable attention and have proven to be a valuable technique. A fully connected autoencoder offers a powerful approach to tackling this challenge. Fully connected autoencoders can effectively capture complex data representations by leveraging an unsupervised learning paradigm, where anomalies are detected based on deviations from standard learned patterns. This introductory section delves into the concept of fully connected autoencoders for anomaly detection, highlighting their utility and discussing why they have emerged as a valuable tool in identifying anomalies within large datasets.
This research project employed a fully connected autoencoder architecture with six linear layers and twelve hidden neurons. The autoencoder is an artificial neural network designed to reconstruct its input data by passing it through an encoder-decoder framework. Each linear layer in the autoencoder represents a set of interconnected nodes or neurons responsible for capturing and encoding relevant features from the input data.
Between every layer, the Rectified Linear Unit (ReLU) activation function introduces non-linearity into the network, enabling the autoencoder to learn complex representations of the input data. By applying ReLU activation, the autoencoder can handle both positive and negative input values, enhancing its ability to model intricate relationships and patterns.

To optimize the autoencoder's objective function, the Adam optimizer was employed. The Adam optimizer is a popular gradient-based optimization algorithm that adapts the learning rate during the training process. By dynamically adjusting the learning rate for each parameter, Adam provides faster convergence and improved training performance compared to traditional optimization algorithms.

The rationale behind selecting this particular autoencoder architecture with ReLU activation and the Adam optimizer is to establish a relatively straightforward model. By opting for a simplified structure, the focus of this research project is to gain a comprehensive understanding of the fundamental functioning of the autoencoder. This choice also provides a baseline for evaluating and contrasting the performance of more complex autoencoder architectures in subsequent analyses.
By implementing this autoencoder configuration, the research project seeks to explore the behavior and capabilities of the autoencoder model, enabling insights into its ability to reconstruct input data and identify anomalies effectively.

All data fed into the model is normalized prior to training in order to eliminate the potential bias introduced by the varying ranges of features. Moreover, to evaluate the performance of the implemented autoencoder, a series of experiments were conducted on each of the datasets. The training phase focused on the "normal" class, aiming to train the model on representative instances of standard data patterns. Subsequently, during the testing phase, a new label, not encountered during the training process, was introduced to the model. The reconstruction error was then recorded for both the training and testing sets, serving as a metric to assess the autoencoder's ability to identify anomalies in the data. In order to condense the errors, defined as ϵ, per prediction into a single numerical value we use the following formula,

$$\epsilon = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2,$$

where y is the original data and $\hat{y}$ is the reconstruction $\forall \in \{1, ..., n\}$.
Several evaluation measures were employed to determine the reconstruction error threshold for labeling a reconstruction as either an anomaly or a regular entry. Initially, we plot the distribution of reconstruction errors for labels 1 and 0, allowing for an assessment of the spread and density of the reconstruction errors.
Next, we extract a precision and recall curve from the evaluation process. The precision and recall curve depicts the trade-off between precision (the ability to identify anomalies accurately) and recall (the ability to capture all anomalies correctly) at various threshold levels. It provides valuable insights into the performance of the anomaly detection model across different levels of strictness or leniency.
Subsequently, we plot the Receiver Operating Characteristic (ROC) curve. The ROC curve illustrates the relationship between the true positive rate (sensitivity) and the false positive rate (1 - specificity) at varying classification thresholds. It is a graphical representation of the model's performance in distinguishing anomalies from standard instances, providing a comprehensive view of its discriminatory power.
Finally, we compute a Youden statistic using the information derived from the precision and recall curve and the ROC curve. The Youden statistic, also known as the Youden's J statistic, is a single-value metric that captures the optimal cut-off point for distinguishing anomalies from normal entries. It maximizes the difference between the true and false positive rates, thereby identifying the threshold that best balances sensitivity and specificity.

By employing these evaluation measures, including the distribution plot, precision and recall curve, ROC curve, and the computation of the Youden statistic, an appropriate reconstruction error threshold can be established to effectively label reconstructions as anomalies or normal entries, enabling accurate and reliable anomaly detection.

5.3 Quantum methods

5.3.1 Quantum-SVM

The first selected quantum method is the Quantum Support Vector Machine (QSVM). This algorithm is based on the classical SVM explained above but employing a quantum circuit to calculate the kernel. The idea behind this is that a quantum kernel can project the data into a new high dimensional Hilbert space which is hard to obtain by using only classical computation. The quantum kernel objective is to separate the classes in the new Hilbert space such that the linear classification task results simpler.
The pipeline of this method consists of three main elements. The first one is the quantum encoding; using a feature map the preprocessed classical data is encoded in the qubits of the circuit. Once the classical data is loaded in the quantum circuit, the quantum kernel is calculated by executing a quantum circuit using a quantum simulator. The last part is to use the previously calculated kernel to perform the classification using SVM on a classical computer.
During the process of optimization, we tried different options for the following parameters

1. Feature map: The selection of the feature map can influence the final results since it determines the way data is represented in the quantum circuit. In the experiments we tried ZFeatureMap, and ZZFeatureMap.

2. Parameters of the feature map: The feature maps have some parameters which we

also tried to optimize:

(a) Number of repetitions: Number of repeated circuits, we tried 2, 3 and 4.

(b) Type of entanglement: Entanglement structure of the circuit in the ZZFeaturemap, can be linear or full.

3. Parameter 'nu' of the classical SVM: The influence of this parameter has been discussed in the subsection of the classical SVM.

5.3.2 VQC

The variational quantum classifier (VQC) algorithm based on the Quantum Autoencoder (QAE) which is designed to work with classical data. The algorithm consists of three main components: encoding layer which converts classical data into quantum states, Parametrized Quantum Circuit (PQC), and finally quantum measurement of errors for one-class classification. We have used feature maps for data encoding which helps to transform the classical data into quantum states. The encoded data is then processed through PQC, which a quantum circuit with adjustable parameters that can be further optimized to during training to classify the data effectively. We train the PQC to minimize the cost function of the training class and the cost function for the outlier class is expected to give values far from zero. The final classification is based on a cross entropy loss function which takes the output probabilities and measure the distance from the truth values. In the model different parameters are optimized to minimize the loss function.

1. Data Encoding: The data encoding is achieved through feature maps. This feature map prepares the input data to be processed by the quantum circuit by encoding it into a quantum state representation suitable for quantum processing. We have run experiments using ZFeatureMap, ZZFeatureMap and PauliFeatureMap. In order to process large datasets, we perform dimensionality reduction using PCA.

2. Parameterized Quantum Circuit: We define a parametrized quantum circuit having a sequence of quantum gates with adjustable parameters(θ) that will be optimized during the training process.

3. Cost function: VQC uses a cross entropy loss function which is a most important cost function. It is used to optimize classification model. It loss is used when adjusting model weights during training. The cross entropy takes the output probability and measure the distance from the actual values. The aim is to make the model output be as close as possible to the desired output or to minimize the loss. A perfect model has a cross-entropy loss of 0. Cross-entropy is defined as

$$L_{CE} = -\sum_{i=1}^{n} t_i \log(p_i) \tag{2}$$

where t_i is the truth label and p_i is the Softmax probability for i^{t^h} class.

6 Results and Discussion

In this section, we will present all the results we obtained and conduct a comparison between the classical methods and the quantum methods. All the results were obtained using the same classes as inliers and outliers. Specifically, for the "Iris dataset," we designated class 1 as the inlier and [0, 2] as outliers. In the case of the "Vowel dataset," we considered 1 as the inlier and [2, 5, 6, 7, 9] as outliers. Lastly, for the "Breast cancer dataset," we used the benign class as the inlier and malign as outliers.

6.1 Comparison and results kernel based methods

In this section, we first analyze the results of Kernel PCA, followed by the OneClassSVM, and then the Quantum OneClassSVM.

For Kernel PCA, we determined the optimal parameters for each dataset, specifically the kernel type and gamma value, to achieve the highest model performance.

The Iris dataset produced perfect scores across all metrics, with an AUC, accuracy, F1 score all at 1 and the AUCROC curve was also perfect fig.22. These exceptional results were obtained using a polynomial kernel and a gamma value of 0.1256. In the case of the Vowel dataset, we achieved an AUC score of 0.9421, an accuracy of 0.9306, and F1 score of 0.9296 the AUCROC curve was less perfect than Iris dataset as we can see in the fig.23. We used a polynomial kernel with a gamma value of 0.3367. As for the Breast Cancer dataset, with the same polynomial kernel but a gamma value of 0.0671, we obtained an AUC score of 0.9251, an accuracy of 0.7867, and F1 score of 0.7448 the AUCROC curve in fig.24. A summary of these results can be found in Table 2.

Table 2: Results for different datasets using KernelPCA

Dataset	**Iris**	**Vowel**	**Breast**
Kernel	poly	poly	poly
Gamma	0.1256	0.3367	0.0671
Accuracy	1	0.9306	0.7867
F1 score	1	0.9296	0.7448

Regarding the classical OneClassSVM method, we conducted several tests to determine the best parameters, similar to Kernel PCA. The selected parameters, along with their corresponding results, are provided in Table 3. These parameters are the ones we used to evaluate the Quantum version of the method. The rationale behind using the same parameters is to assess whether the same configuration that yields good results in classical methods performs equally well in quantum.

Table 3: Results for different datasets using classical methods

Dataset	**Iris**	**Vowel**	**Breast**
Kernel	sigmoid	Sigmoid	Sigmoid
Gamma	0.5	0.5	1
Nu	0.2	0.2	0.1
Accuracy	0.78	0.74	0.89
F1 score	0.875	0.897	0.897

After conducting numerous tests with Quantum OneClassSVM, the best results were achieved using the configuration shown in Table 4. However, it is notable that the performance is significantly worse compared to using classical methods. Along these lines, we attempted to understand how the quantum method operates in order to gain insights into why we are obtaining these results.

The first observation we make is that the simplest dataset, the Iris Dataset, is the only one where we obtain an acceptable result when using the ZFeature map with 3 repetitions. This leads us to consider that increasing the number of required qubits negatively impacts the performance. Another insight we can derive from these results is that we consistently achieve better results with the ZFeature map compared to the ZZFeature map. The ZFeature map essentially encodes classical data into the quantum state space using quantum circuits. It is a first-order Pauli Z-evolution circuit, meaning that the resulting circuit lacks interactions between the encoded data features and, therefore, does not exhibit entanglement. On the other hand, the ZZFeature map is a second-order Pauli-Z evolution circuit, allowing for interactions in the data to be encoded based on the connectivity graph and the classical data map.

Table 4: Results for different datasets using quantum methods

Dataset	**Iris**	**Vowel**	**Breast**
Kernel	ZFeature	ZFeature	ZFeature
Reps	3	1	2
Gamma	0.5	0.5	1
Nu	0.2	0.2	0.1
Accuracy	0.77	0.3	0.29
F1 score	0.84	0.36	0.34

6.2 Fully Connected Autoencoder

As mentioned in the previous section, the performance results of a fully connected autoencoder on the three datasets demonstrate its effectiveness in anomaly detection across different domains. The autoencoder model has been applied to the Iris, Breast Cancer, and Vowel datasets to identify and isolate anomalies.

First, the autoencoder was applied to the iris dataset. During testing and training, we recorded the distribution of the reconstruction error in figure 1 and figure 2. In the distribution of the reconstruction error, we see that there is a fairly clear separation of both errors. Figure 4 shows the precision-recall curve declining as recall increases. To find the optimal threshold for separation, the j-statistics is represented in figure 5, the results shown using this threshold can be found in the confusion matrix reported in figure 3. There is a clear separation between the two classes. The accuracy of this method is 0.871, precision: 0.918, and recall: 0.9. This leaves us with a final F1 score of 0.909.

Next, we move on to the vowel dataset. When we look at the distribution of the reconstruction errors in figures 6 and 7, we see that there is far more overlap in the distribution. Therefore, the model will likely struggle with the separation. The precision-recall curve in figure 9 is steeper in its decline compared to the previous curve. We can also see this in the ROC curve in figure 10. In the roc curve, there is a clear point where the optimal separation threshold lies. When we use that threshold, we find an accuracy of 0.921, precision of 0.955, and recall of 0.933, leading to a combined F1 score of 0.944. The autoencoder seems to perform well on the dataset. Finally, the confusion matrix in figure 8, shows the stellar performance of the model when separating different vowels.

Finally, we applied the autoencoder to the breast cancer dataset. This was the most complex dataset of the three. That is clear when we look at the performance of the autoencoder. In figure 12 we can see that the reconstruction error of the anomalies and the standard data points are extremely similar. There is very little separation between the two classes. When looking at the ROC and precision-recall curves, we observe a similar problem. There is no clearly defined point of decline. Therefore the J-statistic does not seem to be a significant separation threshold. As such, the results of this dataset are lacking compared to the other two. This time, we can only find an accuracy of 0.615 but with a precision of 0.947. Therefore, the model outweighs precision over recall which is only 0.602, which leads us to an F1 score of 0.736.

Table 5: Results for different datasets using Autoencoder

Dataset	**Iris**	**Vowel**	**Breast**
Accuracy	0.871	0.921	0.615
F1 score	0.909	0.944	0.736

6.3 Variational Quantum Classifier

After running the iterations of the VQC as OCC model by changing different parameters, we have found the best F1 score of 0.94 for the Iris data set with 1 layer of circuit using ZFeatureMap and the SPSA optimizer. Model predicted the labels with the best F1 scores for breast cancer and vowels data set with 3 layers and SPSA optimizers.

1. Experiments with different number of layers: In Fig:28, we see the effect of the number of layers in the circuit using the ZFeatureMap on the final accuracy our algorithm produces. In this experiment, we see that except for the Iris data set, there has been a steady decline in the accuracy values upon increasing the number of layers. The model showed fewer variations while predicting the Iris data set with different layers; it had an accuracy of 0.89 for both 1 layer and 15 layers.

2. Experiments with different feature maps: The Fig:29 represents the performance of the model with different feature maps

across different layers for the Iris data set. It is very evident that the ZFeatureMap gives better results than both the ZZFeatureMap and the PauliFeatureMap. The performance of ZZFeatureMap and PauliFeatureMap was found to be different for different numbers of layers. We have observed similar results for both vowel and breast cancer data sets.

3. Experiments with different optimizers: The Fig:30 represents the performance of the model with different optimizers using ZFeatureMap for the Iris dataset. In this experiment, we see that the SPSA optimizers yield the best results when compared with both COBYLA and SLSQP optimizers. The trials revealed that SLSQP takes a very long time to execute compared to the other two optimizers, whereas the COBYLA optimizer produces results more quickly.

Table 6: Results for different datasets using VQC as OCC

Dataset	**Iris**	**Vowel**	**Breast**
Accuracy	0.89	0.820	0.8
F1 score	0.94	0.900	0.89

7 Research Question Answers

- How effective are quantum one-class classification methods in comparison to classical one-class classification methods?

  Based on our experiments, it is inconclusive whether quantum one-class classification methods are more effective than classical ones. Limited resources for executing quantum methods pose a challenge for anomaly detection. For example, the Breast Cancer dataset with 30 features requires more qubits than currently available in IBM Lab, making complex tasks difficult for quantum approaches. As quantum computing advances, these limitations may be overcome. Further research is needed to accurately compare the effectiveness of quantum and classical methods.

- How computationally complex are quantum one-class classification methods in comparison to classical one-class classification methods?

  We are in NISQ (Noise Intermidiate Scale Quantum) era and we were expecting that some quantum machine learning algorithms were going to perform competitive run times as classical machine learning algorithms. But because of the limitations we faced with available resources in IBM's quantum computers and simulators we are not able to give a clear answer on that.

- How do different datasets affect the performance of a quantum one class classification method?

  This answer is strongly related to the first one, as the performance of quantum methods is adversely affected by the dimensionality of the datasets. Therefore, it can be asserted that as the number of features increases, the possibility of running the quantum methods with the available resources decrease.

- How does the quantum approach perform in comparison to classical methods depending on the parameters?

  To accurately address this question, further research is necessary. However, based on our current understanding, it is evident that the parameters yielding the best results in classical methods may not necessarily yield superior performance in quantum methods.

  The distinctive nature of quantum computing needs a fresh approach to parameter specification. Quantum algorithms have unique characteristics that require careful consideration. While classical methods may offer insights and initial parameter choices, they cannot be directly translated into optimal settings for quantum algorithms.

To bridge this gap and determine the parameters that yield favorable results in quantum methods, additional research efforts must be undertaken. This entails exploring the quantum algorithms, analyzing the interplay between classical and quantum features, and conducting empirical studies to identify the parameter configurations that exhibit superior performance specifically in quantum contexts.

- What benefits are apparent by using quantum computing with one classification?

  This question cannot be answered based on the work conducted in this project. However, we have realized that in order to address this question adequately, a different approach is required. Specifically, we propose selecting a dataset that demonstrates poor performance when analyzed using classical methods. By identifying such a dataset, we can then explore how incorporating a quantum component can potentially yield benefits and improve the overall results.

8 Conclusions and Future Work

In this research report, we conducted an analysis to compare the performance of classical and quantum methods for one-class classification across different datasets. Our evaluation revealed that classical methods consistently outperformed quantum methods in most scenarios. Specifically, the classical methods exhibited remarkable accuracy on the Iris and Vowel datasets, achieving F1 scores well above 90%. These datasets, characterized by relatively simpler structures and lower dimensionalities, were well-suited for classical algorithms to extract meaningful patterns and make accurate predictions.

However, when we turned our attention to the breast cancer dataset, a more complex and high-dimensional dataset, the classical methods faced significant challenges in understanding the underlying patterns and achieving satisfactory results. This observation suggests that classical methods may struggle when confronted with intricate and nuanced datasets, potentially leading to suboptimal performance.
While the limitations of classical methods become apparent in handling complex datasets, there arises an opportunity for quantum computing to address such challenges. Quantum computing possesses the potential to leverage the principles of superposition and entanglement to process and analyze intricate datasets with greater efficiency and accuracy. However, it is important to acknowledge that the implementation and documentation of quantum methods are still in their infancy. If there would have been more access to computational power and documentation, the implementation of the VQOOC would have been improved. It was challenging to do the exact implementation of VQOCC as in [4]. Finishing the full implementation was beyond this project's scope but would have yielded better results. Consequently, making a direct and fair comparison between classical and quantum methods at this stage is impractical. Particularly within the constraints of this project.
For future work regarding kernel-based methods, we propose conducting more in-depth research to determine the parameters that yield good performance in the quantum methods. Our assumption was that the parameters that yield good results in classical data would also yield good results in the quantum context. However, as we have observed, this assumption does not hold true. Therefore, it is crucial to investigate and identify the specific parameters contributing to achieving favorable outcomes in quantum settings.

It is unrealistic to expect quantum methods to surpass classical methods in terms of performance at present for general use. Furthermore, the runtime comparison between quantum and classical methods strongly favors

classical computing, as quantum computing technologies are still evolving and face significant technical hurdles.
Nonetheless, as we confront the limitations of classical methods in handling complex datasets, the emergence of quantum computing provides promising opportunities for specific data types and problem domains. The ability of quantum algorithms to exploit quantum phenomena such as superposition and entanglement holds the potential to revolutionize future data analysis and computational modeling.

In conclusion, while classical methods currently outperform quantum methods and enjoy a significant advantage in terms of runtime, the limitations of classical approaches on complex datasets present an opportunity for quantum computing to make significant contributions. As quantum implementation and documentation continue to advance, quantum methods may offer novel solutions and enhanced performance for specific data types, addressing the challenges faced by classical algorithms.

References

[1] Nana Liu and Patrick Rebentrost. "Quantum machine learning for quantum anomaly detection". In: (2017). URL: https://journals.aps.org/pra/abstract/10.1103/PhysRevA.97.042315.

[2] Nicolas M. de Oliveira et al. "Quantum One-class Classification With a Distance-based Classifier". In: (2020). URL: https://arxiv.org/abs/2007.16200.

[3] M. Fingerhuth M. Schuld and F. Petruccione. "Implementing a distance-based classifier with a quantum interference circuit". In: *IOP Publishing* (2017). URL: https://iopscience.iop.org/article/10.1209/0295-5075/119/60002.

[4] Gunhee Park, Joonsuk Huh, and Daniel K Park. "Variational quantum one-class classifier". In: *IOP Publishing* (2023). URL: https://doi.org/10.1088/2632-2153/acafd5.

[5] Benjamin Obert Daniel Herr and Matthias Rosenkranz. "Anomaly detection with variational quantum generative adversarial networks". In: *IOP Publishing* (2021). URL: https://iopscience.iop.org/article/10.1088/2058-9565/ac0d4d.

[6] David M. J. Tax and Robert P. W. Duin. "Characterizing one-class datasets". In: *Information and Communication Theory Group Delft University of Technology* (2006). URL: http://rduin.nl/papers/prasa_05_occ.pdf.

[7] Heiko Hoffmann. "Kernel PCA for novelty detection". In: *Hoffmann, Heiko* (2007). URL: https://www.sciencedirect.com/science/article/pii/S0031320306003414.

9 Appendix

9.1 Results Autoencoder

9.1.1 Iris Dataset

The following figures show the performance of the fully connected autoencoder on the Iris dataset.

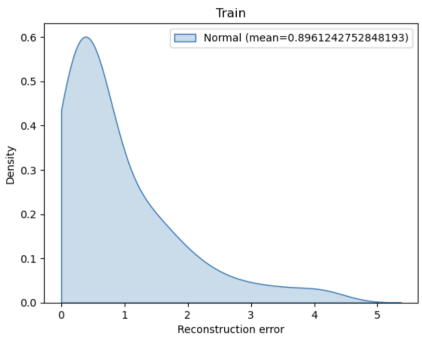

Figure 1: Training Reconstruction Error Distribution on Iris Dataset

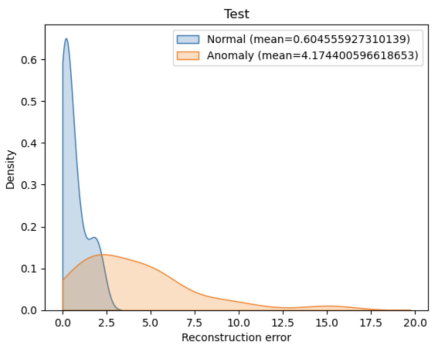

Figure 2: Test Reconstruction Error Distribution on Iris Dataset

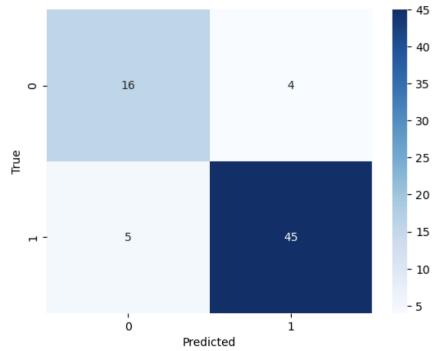

Figure 3: Confusion Matrix on Iris Dataset

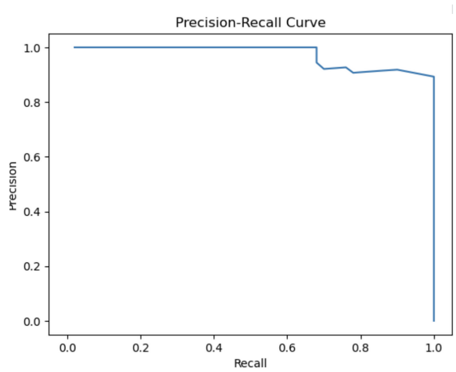

Figure 4: Precision-Recall Curve on Iris Dataset

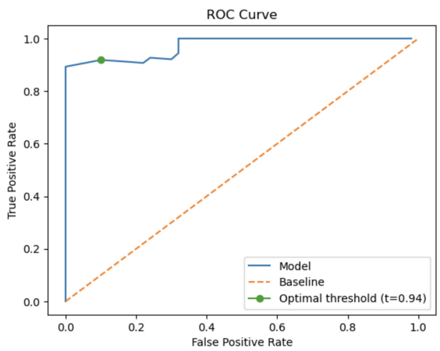

Figure 5: ROC Curve on Iris Dataset

9.1.2 Vowel Dataset

The following figures show the performance of the fully connected autoencoder on the Vowel dataset.

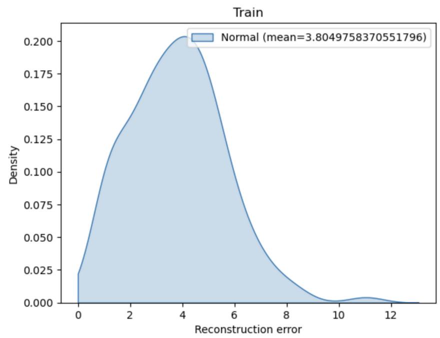

Figure 6: Training Reconstruction Error Distribution on Vowel Dataset

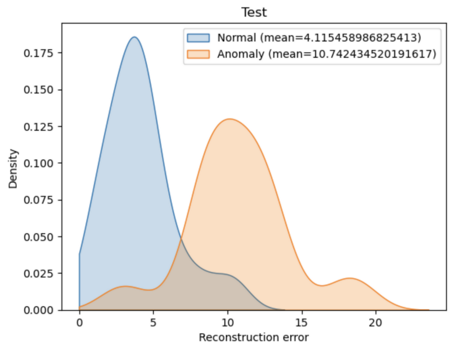

Figure 7: Test Reconstruction Error Distribution on Vowel Dataset

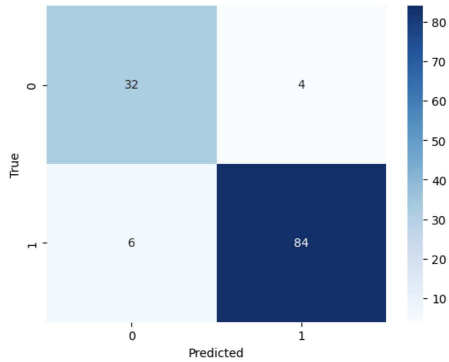

Figure 8: Confusion Matrix on vowel Dataset

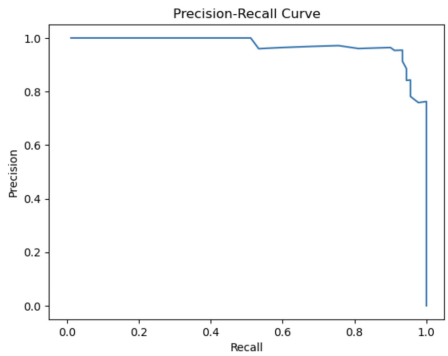

Figure 9: Precision-Recall Curve on Vowel Dataset

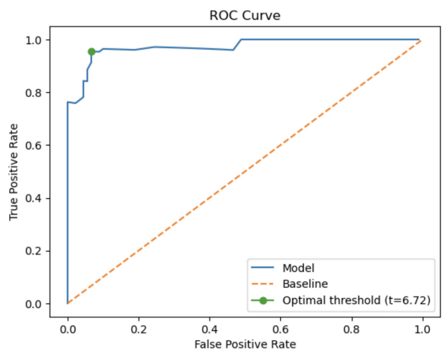

Figure 10: ROC Curve on vowel Dataset

9.1.3 Breastcancer Dataset

The following figures show the performance of the fully connected autoencoder on the Breast Cancer dataset.

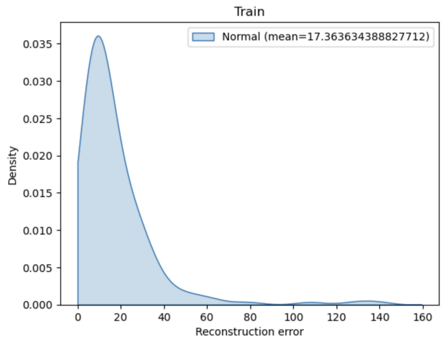

Figure 11: Training Reconstruction Error Distribution on Breast Cancer Dataset

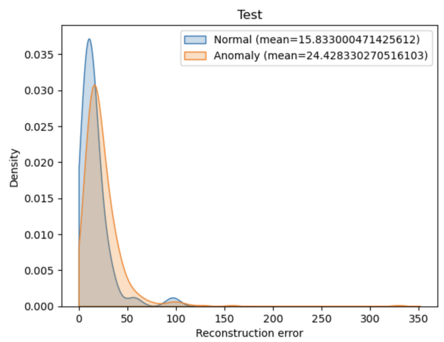

Figure 12: Test Reconstruction Error Distribution on Breast Cancer Dataset

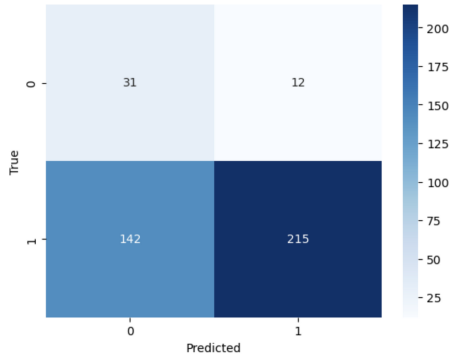

Figure 13: Confusion Matrix on Breast Cancer Dataset

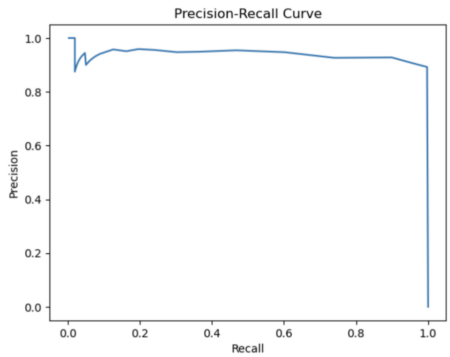

Figure 14: Precision-Recall Curve on Breast Cancer Dataset

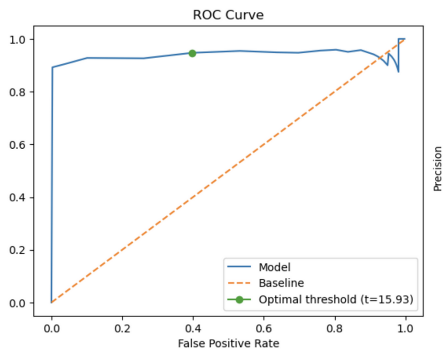

Figure 15: ROC Curve on Breast Cancer Dataset

9.2 Results SVM

The following figures show the performance of the Support Vector Machine on all three datasets.

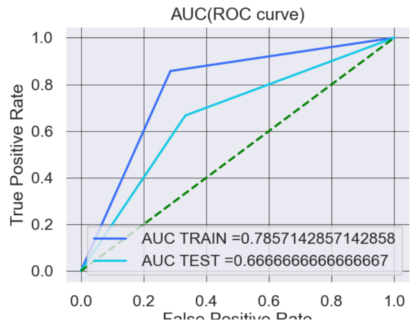

Figure 16: Roc curve for the Iris dataset

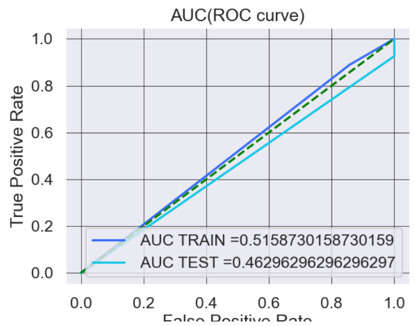

Figure 17: Roc curve for the Vowel dataset

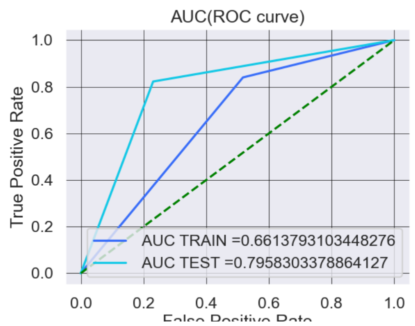

Figure 18: Roc curve for the Breast cancer dataset

9.3 Results QSVM

The following figures show the performance of the Quantum Support Vector Machine on all three datasets.

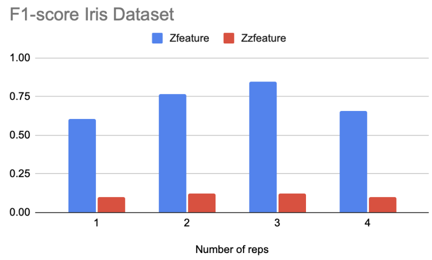

Figure 19: Here we can notice the difference between using Zfeature map and ZZfeature map

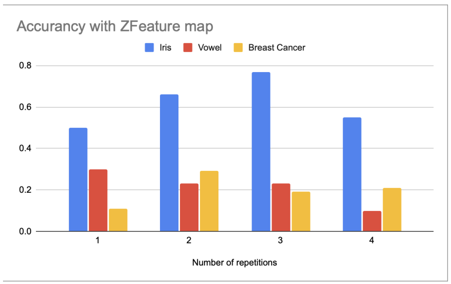

Figure 20: Here we can see the accurancies for each dataset when using Zfeature map

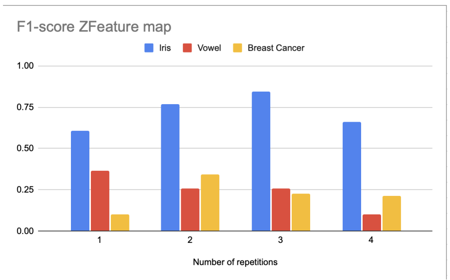

Figure 21: Here we can see the f1-scores for each dataset when using Zfeature map

9.4 Results Kernel PCA

9.4.1 Iris Dataset

The following figures show the performance of kernel PCA on the Iris dataset.

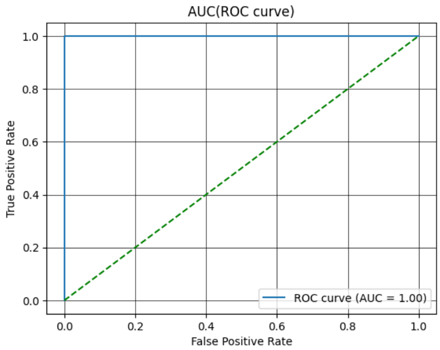

Figure 22: Roc curve for the Iris dataset

9.4.2 Vowel Dataset

The following figures show the performance of kernel PCA on the Vowel dataset.

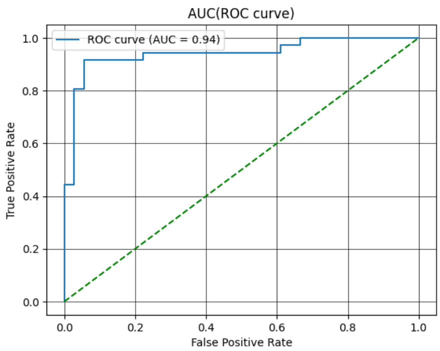

Figure 23: Roc curve for the Vowel dataset

9.4.3 Breast Cancer Dataset

The following figures show the performance of kernel PCA on the Breast Cancer dataset.

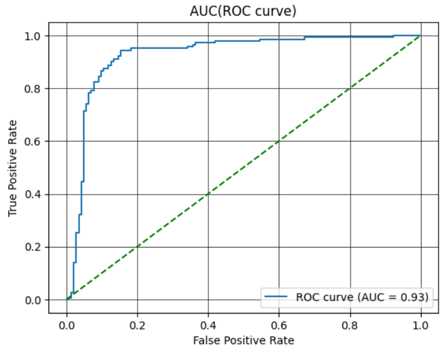

Figure 24: Roc curve for the Breast Cancer dataset

9.5 Results of VQC as OCC

9.5.1 Iris Dataset

The following figures show the performance of the Variational Quantum One-Class-Classifier on the Iris dataset.

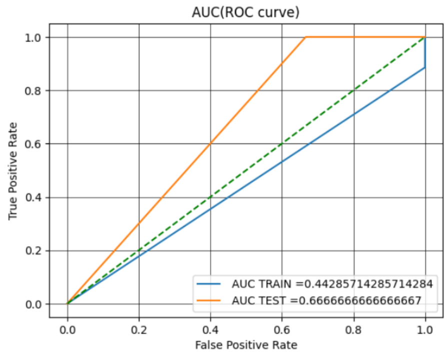

Figure 25: Roc curve for the Iris dataset

9.5.2 Vowel Dataset

The following figures show the performance of the Variational Quantum One-Class-Classifier on the Vowel dataset.

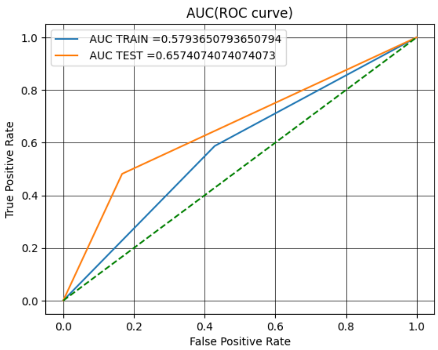

Figure 26: Roc curve for the Vowel dataset

9.5.3 Breast Cancer Dataset

The following figures show the performance of the Variational Quantum One-Class-Classifier on the Breast Cancer dataset.

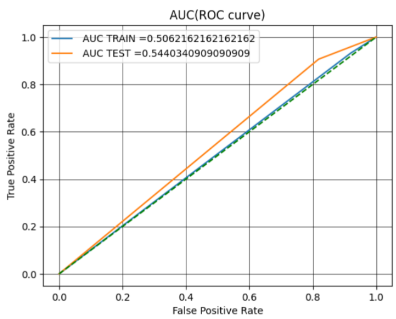

Figure 27: Roc curve for the Breast cancer dataset

9.5.4 Models Performance

The following figures show the performance of the Variational Quantum One-Class-Classifier.

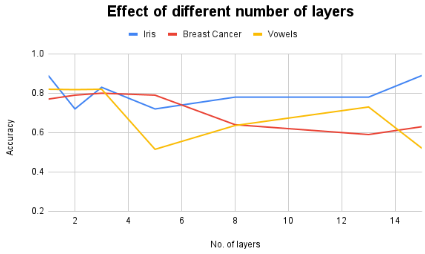

Figure 28: The plot represents the effect of different number of layers for each of the data set

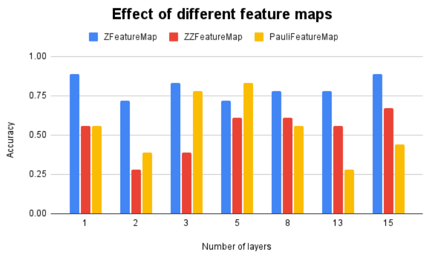

Figure 29: The plot represents the effect of different feature maps on Iris dataset

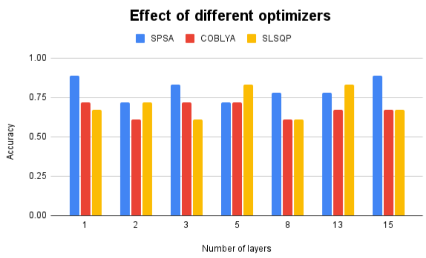

Figure 30: The plot represents the effect of different optimizers using ZFeature Map for Iris dataset