Research Article

Shubham Vashisth*, Ishika Dhall, and Garima Aggarwal

Design and analysis of quantum powered support vector machines for malignant breast cancer diagnosis

Abstract: The rapid pace of development over the last few decades in the domain of machine learning mirrors the advances made in the field of quantum computing. It is natural to ask whether the conventional machine learning algorithms could be optimized using the present-day noisy intermediate-scale quantum technology. There are certain computational limitations while training a machine learning model on a classical computer. Using quantum computation, it is possible to surpass these limitations and carry out such calculations in an optimized manner. This study illustrates the working of the quantum support vector machine classification model which guarantees an exponential speed-up over its typical alternatives. This research uses the quantum SVM model to solve the classification task of a malignant breast cancer diagnosis. This study also demonstrates a comparative analysis of distinct forms of SVM algorithms concerning their time complexity and performances on standard evaluation metrics, namely accuracy, precision, recall, and F1-score, to exemplify the supremacy of quantum SVM over its conventional variants.

Keywords: quantum machine learning, supervised machine learning, support vector machines, quantum support vector machines, breast cancer classification

1 Introduction

Breast cancer disease remains the second most common form of cancer diagnosed in women in the United States of America. This cancer matures in breast cells, typically in the ducts or lobules of the breast. The early signs in an infected patient may include swelling, skin irritation, a lump in the breast, nipple retraction, redness, changes in the lymph node, and so on. In the past few decades, there have been consistent developments in designing reliable methods [1] that can diagnose and deal with this form of cancer with high efficiency. Within the domain of computer science, after an exponential increase in the amount of data available and upgraded computational hardware coming into the picture, researchers are now able to implement various data mining [2] and machine learning [3] techniques to perform accurate classification on real-life data.

The task of diagnosing breast cancer primarily remains as a binary classification problem that consists of the malignant tumor that is cancerous in nature and the other being the non-cancerous benign tumor. Many of the existing machine learning classifiers namely support vector machine (SVM), k nearest neighbor, decision trees, random forest, artificial neural networks, and so on can model most of the available breast cancer datasets. However, with an increase in the number of deciding factors, i.e., features

---

* Corresponding author: Shubham Vashisth, Department of CSE, Amity University, Noida Sector-125, Uttar Pradesh, India,
e-mail: shubham.vashisth.delhi@gmail.com
Ishika Dhall, Garima Aggarwal: Department of CSE, Amity University, Noida Sector-125, Uttar Pradesh, India

of the problem, and an increase in the size of the problem, most of the above-mentioned classifiers demand high-end computation power for the training of the classifier.

With advances made in other forms of computing, it is understandable to explore such fields and implement machine learning algorithms to perform the required calculations which require less computation power. Using the principles of quantum computing and the present-day noisy intermediate-scale quantum technology, it is possible to speed up conventional machine learning classifiers without compromising classification accuracy. The concepts of quantum speedup [4] has the potential for algorithms used in quantum machine learning to overshadow the classical algorithms. Also, it has ability to re-transform machine learning and take it one step closer in developing the most optimal and efficient classification systems. One of the most worked conventional classifiers, SVM classifiers, is the best for challenges like breast cancer classification where there are a large number of features that are taken into consideration to perform the task of classification. Especially the nonlinear SVM which can transform data into higher dimensional space by using its kernel trick [5] is expected to perform well on such complex classification problems. The nonlinear SVM classification algorithm is computationally expensive as it requires the calculation of the kernel matrix and solving the least-squares formulation or quadratic dual problem [6], making it challenging to be computed using a classical computer system. Thus, a significant speedup can be obtained in the computation capability of the conventional SVM classifier by using the principles of quantum machine learning [7] and quantum SVMs [8]. The computational speedup is achieved by quantum SVM using two sources, the first being faster calculation of the kernel matrix and the second source being an effective solution of the linear equations on quantum hardware as SVMs transform the optimization problem into a set of linear equations.

The main contribution of this research is the implementation and simulation of quantum SVM classification algorithms on quantum systems using the Qiskit library as mentioned in ref. [9] to generate an optimized breast cancer machine learning classifier. This model further includes the application of the principal component analysis (PCA) algorithm to reduce the total number of features in consideration based on the number of qubits made accessible. The proposed quantum-powered classification model would be able to solve the binary task of breast cancer diagnosis in an optimized manner in logarithmic order as opposed to the polynomic way applied conventionally, which results in lower computation power and accelerated classification.

Another contribution of this study is the comparative analysis of distinct SVM classification algorithms, typically linear SVMs, nonlinear SVMs, and quantum SVMs keeping into consideration their performances on several evaluation metrics (accuracy, precision, recall, and F1-score) and the computation cost in solving the problem of a breast cancer diagnosis. This analysis clears the most efficient SVM classifier out of the above-mentioned models for the specified problem that has been taken into consideration.

The rest of the paper has been structured as follows: Section 2 introduces the preliminaries; Section 3 offers the literature review; Section 4 discusses the methodology; Section 5 presents the results; and Section 6 presents the discussion. Finally, the conclusion and future implications are discussed in Section 7.

2 Preliminaries

An SVM classifier [10] is a type of supervised machine learning model. SVM learns a given independent and identically distributed training set $\{(x_1, y_1), \ldots, (x_M, y_M)\}$, where $x_i \in R^N$ represents the data points and $y \in \{-1, 1\}$ denotes the binary classes which in our breast cancer classification problem represents the benign and malignant classes. Here N denotes the number of parameters that represent a single instance, i.e., features and M denotes the number of training instances. The goal of SVM is to maximize the closest distance between margin and support vectors. In the case of linear SVM, the output is given using the equation $y_i := \mathrm{sgn}(w^T x_i + b)$, where x_i is the ith training instance, y_i is the respective output from the linear SVM, w is the normal vector to the hyperplane, and b denotes the offset or the bias. For complex and

real-life nonlinear information, linear SVM is not able to model data efficiently. Therefore, SVM utilizes nonlinear kernel mapping [5] that performs mapping of training data from an input space to a higher dimensional feature space. This mapping is crucial for problems that are not linearly distinguishable.

Using the ϕ embedding operation, SVM needs the solution to the following optimization problem:

$$\text{Minimize } \frac{1}{2}w^Tw + C\sum_{i=1}^{M}\varepsilon_i, \tag{2.1}$$

subject to $y_i(w^T\phi(x_i) + b) \geq 1 - \varepsilon_i, \ \varepsilon_i \geq 0, \ i = 1,..., N$, where ε_i denotes the classification error, and C refers to the regularization or cost parameter. In cases where the misclassification rate increases, an SVM model with a higher number of support vectors is required to increase the complexity of the classifier. The decision function is given by

$$f(x) = \text{sgn}(w^T\phi(x) + b).$$

As equation (2.1) is a constrained optimization problem, it can be converted to an unconstrained optimization problem. Lagrange duality is used to obtain the optimization problem's dual formulation. By application of the Lagrangian multipliers, it is possible to accommodate constraints in the optimization problem. The saddle point of the Lagrangian is defined using the partial derivatives in $w, \ b$, and ε that transform the dual form in a quadratic programming problem:

$$\text{Maximize } \sum_{i=1}^{M}\alpha_i - \frac{1}{2}\sum_{i=1}^{M}\sum_{j=1}^{M}\alpha_iy_i\alpha_jy_jK(x_i, x_j),$$

subject to $\sum_{i=1}^{M}\alpha_iy_i = 0, \ \alpha_i \in [0, \ C], \ i = 1,...,M$. Here, $K(x_i, x_j)$ represents the kernel function, the dot product of embedding space. Kernel function evades evaluation of embedding by itself, and it is applied directly to classify new data instances. Therefore, the decision function becomes:

$$D(z) = \sum_{i=1}^{M}\alpha_iy_iK(x_i, z) + b.$$

The quadratic programming problem is resolved using sequential minimum optimization on classical computer systems. By using the l_2 norm in the regularization term of the primal problem, conventional SVMs modify the goal function:

$$\text{Minimize } \frac{1}{2}w^Tw + \frac{\gamma}{2}\sum_{i=1}^{M}e_i^2,$$

subject to $y_i(w^T\phi(x_i) + b) = 1 - e_i, \ i = 1,..., N$. Here parameter γ acts as the cost parameter. The following least-squares problem is obtained seeking the saddle point of the consequent Lagrangian:

$$\begin{pmatrix} 0 & 1 \\ 1 & K + \gamma^{-1}I \end{pmatrix}\begin{pmatrix} b \\ \alpha \end{pmatrix} = \begin{pmatrix} 0 \\ y \end{pmatrix}, \tag{2.2}$$

where $K_{ij} = K(x_i, x_j)$. The complexity of the classifier increases because the SVM trades off zero α_is for non-zero error terms e_is.

For a linear or polynomial kernel, the calculation of entry in the kernel matrix has $O(N)$ time complexity. Therefore, evaluating the entire kernel matrix requires $O(M^2N)$ time. To solve the quadratic dual problem, $O(M^3)$ complexity is exploited. Linking both these steps, the conventional SVM has a minimum of $O(M^2(M + N))$ complexity. However, by the application of the quantum formulation, significant speedup can be achieved in the above-mentioned two steps resulting in an exponentially reduced complexity.

The quantum SVM algorithm utilizes quantum states to represent the training instances. This representation of quantum states is incorporated by querying the QRAM $O(\log N)$ times. For performing the dot product between any two training vectors, the following steps are included:

1. Generation of two quantum states ($|\Psi\rangle$ and $|\Phi\rangle$) with an ancilla variable.
2. Estimation of the sum of the squared norms of the two respective training instances given by $Z = |x_i|^2 + |x_j|^2$.
3. Comparing the quantum two states by performing a projective measurement on the ancilla variable.

$|\Psi\rangle$ is constructed by querying QRAM, whereas the estimation of the other state $|\Phi\rangle$ and the sum of the squared norms are done together. This takes the complexity of $O(\epsilon^{-1})$, where ϵ represents the estimation of the desired accuracy [11]. This results in the complexity of $O(\epsilon^{-1} \log N)$ for the estimation of a single dot product. Therefore, this estimation of kernel matrix by quantum means yields a complexity of $O(M^2(M + \epsilon^{-1} \log N))$. To reduce the complexity, even more, the speedup is obtained by calculating the optimum of the least-squares dual formation.

Q-SVM solves equation (2.2), by performing the inversion of matrix F where

$$F = \begin{pmatrix} 0 & 1 \\ 1 & K + \gamma^{-1}I \end{pmatrix}.$$

The quantum inversion algorithm [12] simulates the matrix exponential of F. This requires splitting of matrix F in the form of $F = J + K$, where $J = \begin{pmatrix} 0 & 1^T \\ 1 & 0 \end{pmatrix}$ and $K = \begin{pmatrix} 0 & 0 \\ 0 & K + \gamma^{-1}I \end{pmatrix}$.

Using the Baker–Campbell–Hausdorff formula, F is normalized with its trace and the exponential is given by

$$e^{i\hat{F}\Delta t} = e^{\frac{-iJ\Delta t}{trK_\gamma}} e^{\frac{-i\gamma^{-1}I\Delta t}{trK_\gamma}} e^{\frac{-iK\Delta t}{trK_\gamma}} + O(\Delta t^2). \tag{2.3}$$

In the case of the sparse matrix J, the simulation of constant identity matrix multiplication is easy to simulate [13]. On the other hand, matrix K is non-sparse. Therefore, in this case, quantum self-analysis [11] eases the measurement by exponentiating it as a Hamiltonian. The quantum estimation of the dot product and access to the QRAM result in the normalized kernel matrix which is given by

$$\hat{K} = \frac{K}{trK} = \frac{1}{trK} \sum_{i,j=1}^{M} \langle x_i | x_j \rangle |x_i||x_j||i\rangle\langle j|,$$

where $\hat{K}$ is a normalized Hermitian matrix. The exponentiation is performed in $O(\log N)$ steps. $e^{i\hat{K}\Delta t}$ is obtained by

$$e^{-i\mathcal{L}_{\hat{K}}\Delta t} \approx I - i\Delta t[\hat{K}, .] + O(\Delta t^2),$$

where $\mathcal{L}_{\hat{K}} = [K, .]$. To obtain the eigenvalues and eigenvectors, equation (2.3) is used. This is followed by expressing y in (2.2) in eigenbasis and inverting the eigenvalues to acquire the solution $|b, \alpha\rangle$. Hence, with the application of the inversion algorithm, training the support vectors leads to the overall time complexity of $O(\log NM)$.

3 Literature review

In the past few decades, there has been a significant amount of research to perform a reliable diagnosis of breast cancer. Various methods and techniques have been proposed by researchers for successfully predicting the breast cancer type, especially approaches that fall within the domain of data mining, soft computing, and artificial intelligence. One such work [14] includes the construction of three machine learning classifiers (SVM, decision tree, and artificial neural network) to perform breast cancer diagnosis. These models were trained on the data that were gathered from the Iranian Center for Breast Cancer (ICBC) program, 1997–2008. The results were evaluated using the cross-validation technique to prevent biased

prediction scores. The finalized SVM, decision tree, and ANN models were able to accomplish a testing score of 0.957, 0.936, and 0.947, respectively.

The application of machine learning to solve classification, regression, pattern recognition, dimensionality reduction, and other problems is increasing day after day. Because of the algorithmic advancement and access to high computation power these days, it is relatively less challenging for engineers and researchers to build deep and complex machine learning models. Quantum systems can generate counter-intuitive patterns that are resisted by classical computer systems. The field of quantum machine learning that explores quantum-powered algorithms that can offer such advantages has been proposed in ref. [7]. The proposed work opens several paths for developing quantum-powered machine learning systems contemplating software and hardware challenges.

Diagnosing breast cancer primarily is a binary classification problem. There are several machine learning algorithms like the k-nearest neighbor, SVM, decision tree, Naïve Bayes, etc. which serve as a reliable option for creating an efficient classifier. Various approaches have been presented to take the performance of the existing classifiers one step further. According to ref. [15], one such approach involves the principles of quantum mechanics. Quantum mechanics considers a combination of unique theoretical frameworks that serve as the key in designing the state-of-the-art classification model. The research done in ref. [15] indicates that advances in classification are possible to obtain, keeping partial exploitation of quantum detection into consideration.

For the task of binary classification, there are several efficient supervised machine learning classifiers, one of them being the SVM classifier. The study by Rebentrost et al. [8] proposed an optimized variant of SVM algorithm which is possible to be implemented on a quantum computer. The proposed algorithm has logarithmic complexity that depends upon the number of training instances and the size of the vector. This algorithm guarantees an exponential speed-up over the conventional SVM algorithm that requires polynomial time. The quantum SVM algorithm operates using a non-sparse matrix exponentiation approach method that enables it to efficiently calculate matrix inversion of the training instances.

Quantum computing has a lot to offer to the domain of machine learning (both supervised and unsupervised). Studies conducted in ref. [16] emphasize the quantum machine learning algorithms, specifically quantum SVM algorithms that have a significant decrease in time complexity over its classical variant. In this study, researchers illustrated the task of recognizing hand gestures on a four-qubit MNR test bench. The paper illustrates that the quantum variant of SVM can be implemented with the complexity of $O(\log(M \times N))$, which is a considerable speedup over the classical variant of SVM with the training complexity of $O(M^2(M + N))$.

Apart from the SVM classifier, the nearest neighbor classifier and its various modifications are some of the most applied classification models in machine learning and data mining for accomplishment of classification tasks. This conventional algorithm can be powered to generate more high-level models that can perform complex classification using fast and comprehensible quantum methods. The study shown in ref. [17] was able to verify higher bounds on the number of queries to the required input to compute inner distance and Euclidean distance matrices. In the worst case, they were able to reduce the complexity of polynomial nature exponentially in comparison to the classical algorithm.

Reinforcement machine learning is one of the major types of learning fields under the domain of machine learning. To speed up the learning, a combination of the concept of quantum theory with the standard reinforcement learning to produce quantum reinforcement learning which is highly influenced by the concepts behind quantum superposition and quantum parallelism has been performed [18]. The proposed model was able to achieve impressive optimality and convergence, overall illustrating high efficiency while modeling complex problems.

There has been a significant amount of research in predicting breast density and breast cancer development by using computer-aided diagnosis (CAD) systems. Work done in ref. [19] illustrates that an interconnection does exist between the early prediction of breast tissue density and cancer development. The research was executed on the MIAS dataset. Entropy, kurtosis, mean, standard deviation, and skewness were estimated followed by the reduction in space dimensionality of these five texture feature vectors using the PCA algorithm. These estimations were carried out using Laws' texture energy images with mask

lengths 5, 7, and 9. PNN and SVM models were used to perform classification. The findings demonstrate that the SVM classifier (mask length 5) reported a higher accuracy score of 94.4% as compared to the PNN (mask length 7) classifier's accuracy score of 92.5% using the first four principal components. These results imply that the proposed CAD system design could be effective in breast density classification.

Breast cancer is responsible for the highest number of cancer deaths among women. CAD systems are often considered to be the default prototype for performing mammography which allows efficient and early detection of breast abnormalities. Studies in ref. [20] suggest an ensemble machine learning approach for accurate classification of malignant and benign breast tumors. By application of ensemble learning, diversity among classifiers is being promoted. This process involves the selection of various feature subsets and unique classification components from the initial feature pool. Multiple ensemble methods, namely AdaBoost, Bagging, and RSS, were compared to the proposed RSS-SCS approach. The experimentations were carried out using sensitivity, specificity, and accuracy as the evaluation metrics, the results of which give the RSS-SCS classifier a higher edge over the conventional ensemble learning approaches.

Gradient descent serves as the fundamental learning algorithm for training an artificial neural network. By building a quantum-based version of gradient descent optimizer, the learning can be accelerated exponentially for higher dimension input problems. Factorization machines which use factorized parameterization unlike other common models such as SVMs that use dense parameterization have been used in ref. [21]. The paper discusses the implementation of the same on quantum computers, this learning algorithm takes the complexity of $O(K \text{ polylog } N)$ where K and N represent parameter dimensions.

The rapid increase in the use of CAD systems for the detection of breast abnormalities has opened room for applying distinctive classification approaches for the efficient diagnosis of breast cancer. Traditional approaches focused upon supervised machine learning models which require a large volume of gathered data to be labeled before training the model. Ref. [22] offers a semi-supervised machine learning approach for solving the classification problem of mammography abnormalities. This approach uses a TSVM-CAD system, the performance of which is examined on the DDSM dataset. The performance of the transductive SVM model was compared with the Gaussian, Triangle, and Linear kernel functions. Results suggest that the TSVM-CAD system with the Gaussian kernel function outperformed its other variants with an accuracy score of 0.929 and a sensitivity of 0.89.

With the considerable increase in using CAD systems for early and secure breast cancer diagnosis, efficient classification of mammary masses persists to be a hard challenge for researchers. Benzebouchi et al. [23] proposed a deep learning approach to solve the same by designing a convolutional neural network classifier. Automated selection of optimal features rather than hand tuning the features as done in the case of conventional machine learning models motivated the selection of CNN. The classification model was trained on segmented data gathered from the Digital Database for Screening and Mammography. The proposed CNN model offered an improved accuracy score of 97.89%, making it a more secure and reliable classifier for breast cancer diagnosis.

Besides supervised machine learning problems, quantum speedup could be brought to other types of machine learning problems like clustering which falls under the category of unsupervised machine learning. In ref. [24], Grover's algorithm and its variations, which are at the core of the clustering algorithm quantization, are well explained. It presents three quantum subroutines to fast-track conventional clustering algorithms. Also, it proposed four algorithms for various clustering techniques, for instance, k-means, c-means, and divisive clustering. It speeds up the selection of classical clustering algorithms by a new quantized method, but the approach used is not exactly realistic but requires the accessibility of a black box used to inquire the distance between pairs of points in principles of superposition.

Most of the machine learning models function by carrying out certain matrix operations on vectors in high-dimensional space, which also serves as the basis of quantum mechanics. The reviewed papers presented in this section illustrate how various machine learning classifiers are trained on classical computer systems using supervised and semi-supervised learning approaches. One of the major drawbacks of the conventional classifiers is their considerably high training time complexity. Also, a considerable amount of research argues regarding the capability of a hybrid quantum system to generate patterns more efficiently than a classical computer in certain machine learning challenges. Understandably, quantum

computing can be applied to current machine learning algorithms to reduce their computation cost exponentially. This study, therefore, uses quantum SVM classifier for performing breast cancer classification in an optimized approach compared to the conventional SVM classifiers (linear and nonlinear SVMs) which are typically trained on classical computer systems.

4 Methodology

Demonstrably there are computational drawbacks in the conventional SVM classifier. By using the concepts of quantum physics and machine learning, it is possible to outclass the classical model. Quantum systems can enable computation advantages that further allow classification in higher dimensions. The problem of a breast cancer diagnosis is typically viewed as a supervised machine learning [25] classification task which involves a series of steps as illustrated in Figure 1.
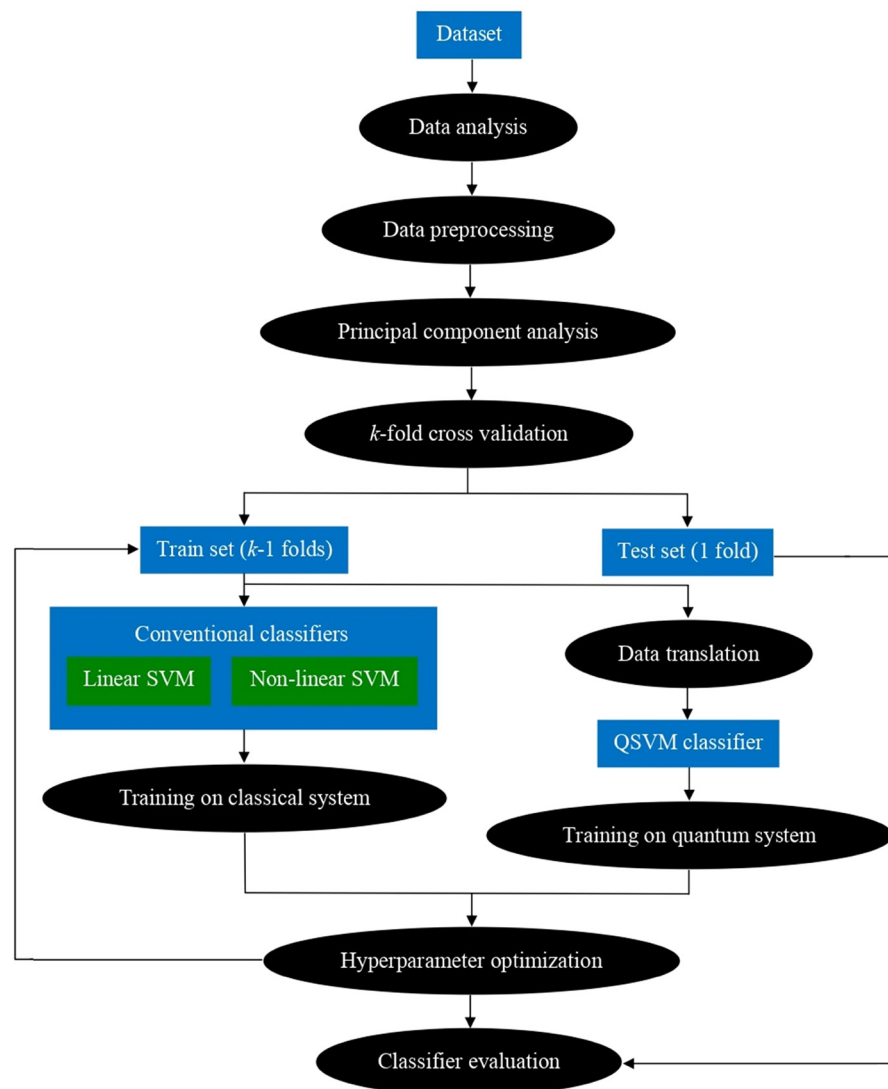

Figure 1: Object process diagram of constructing, training, and testing the SVM classifiers.

To demonstrate the most efficient model between the linear, nonlinear, and quantum SVM classifiers, three different simulations were performed. Linear and nonlinear SVM algorithms were simulated on classical computer systems, whereas the quantum SVM algorithm was simulated on the quantum systems.

4.1 Data analysis, preprocessing, and translation

Data serve as the fuel that enables learning of the classifiers that must be constructed. Machine learning algorithms are data-driven and thus the initial phase in building an efficient breast cancer classifier requires an adequate amount of data. For the problem taken into consideration, the Breast Cancer Wisconsin Database has been selected and collected from ref. [26]. The dataset consists of two classes (benign and malignant). It has a length of 569 reports, i.e., the total number of samples, out of which 357 reports belong to the class benign and the remaining 212 reports belong to the malignant class. Also, the dataset comprises 30 features, i.e., the features for generating the classifying model. These features include deciding considerations for the breast cancer problem like mean radius, texture, area, etc. Since for this research, we are simulating the quantum SVM algorithm using a quantum system with a restricted number of qubits (2-qubit quantum simulator) it raised the motivation to reduce the dimensionality of our original dataset. Thus, to reduce the data dimensions from 30 to 2 dimensions, PCA is used after scaling and normalizing the data using the MinMaxScaler method. The application of PCA not only solves the problem of tackling the problem of the limited number of qubits, but it also makes it easier to explore and visualize data as displayed in Figure 2.

In both cases, i.e., classical as well as quantum simulation involving the three SVM classifiers, transformed data with reduced dimensions (two principal components as features for 569 total samples) served as the final dataset for training and testing the classifiers. These two principal components cumulatively were able to explain up to 70% of the variance in the dataset as shown in Figure 3.
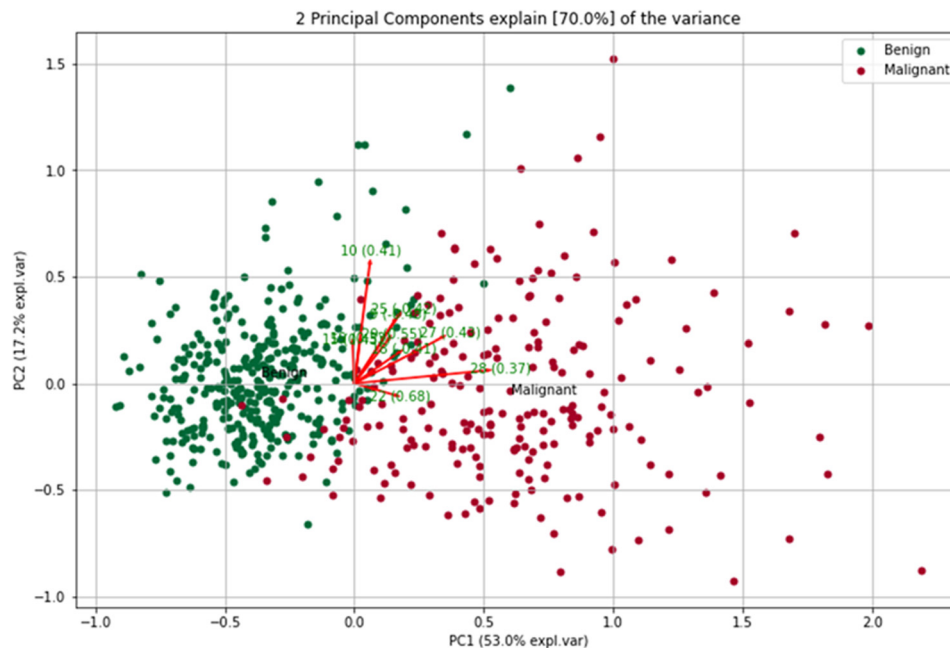

Figure 2: Visualization of the breast cancer dataset with reduced dimensions using PCA algorithm.
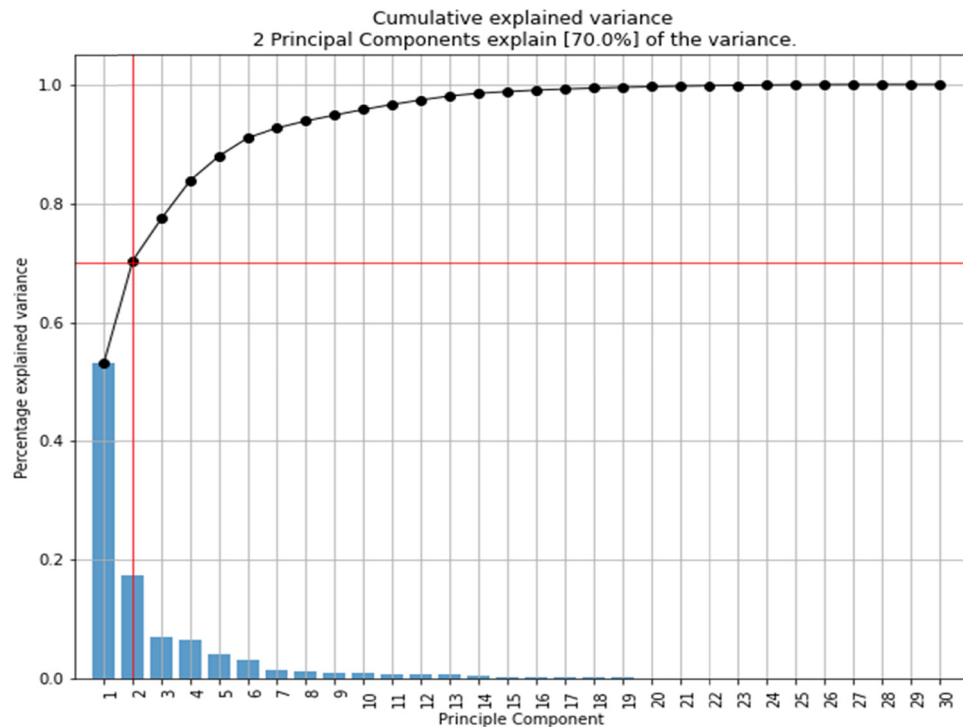
Figure 3: Cumulative variance is explained using the two principal components.

For simulating the quantum SVM algorithm, a classical datapoint says x must be transformed into quantum datapoint $|\Phi(\vec{x})\rangle$. This transformation is made possible using a circuit $V(\Phi(\vec{x}))$, where $\Phi(\ )$ could be any classification function applied to classical data $\vec{x}$. A parameterized quantum circuit $W(\theta)$ with θ parameters is used to process these data. This is followed by obtaining the respective labels of the transformed data by applying measurements that output the classical value of $-1,\ 1$ for every classical input $\vec{x}$. We can describe an ansatz for such a problem that can be represented using $W(\theta)V(\Phi(\vec{x}))|0\rangle$ which defines the quantum variational circuit.

4.2 Constructing and training of the classification model

To construct and train our three SVM classification models, three unique simulations were performed using the principles of supervised machine learning on the dimensionally reduced dataset. For training each of the three SVM classifiers, the k-fold cross-validation technique (with the value of $k = 10$) was deployed to produce fair and well-rounded metric scores [27]. The two of the SVM classifiers (linear and nonlinear SVM) are trained using a classical computer system, whereas the quantum SVM classification model is trained using a quantum system. Hence, the training phase has been further divided into two sections based on their computation requirements.

4.2.1 Classical simulations

The conventional SVM models were trained on classical computer systems. This was done by initially importing the necessary dependencies including the required packages and libraries. In the case of constructing the linear SVM model, the LinearSVC class provided by the scikit-learn library was used. The LinearSVC class at its core uses the liblinear library which implements an optimized algorithm [28].

LinearSVC model does not perform the kernel trick, which makes it computationally less expensive but also takes away the ability of the model to perform efficient classification in case of complex datasets which further increases the misclassification loss.

In the case of training a nonlinear SVM model, the libsvm library is used which allows the nonlinear SVM model to perform the kernel trick according to ref. [29]. The kernel trick is computationally expensive but allows the classifier to perform well on complex problems. This transformation increases the current dimension of the data to a higher dimension space which allows the model to fit an optimized hyperplane for performing classification efficiently. The libsvm library provides a variety of kernel options that satisfies Mercer's condition and enable the selection of the best-suited kernel function for solving the nonlinear separation problem. The kernel options comprise the following:

- Linear kernel: $K(x_i, x_j) = x_i . x_j$,
- Polynomial kernel: $K(x_i, x_j) = (\gamma \langle x_i, x_j \rangle + c)^p$,
- Gaussian kernel (radian-basis function): $K(x_i, x_j) = e^{-\frac{||x_i - x_j||^2}{2\sigma^2}}$,
- Sigmoid kernel: $K(x_i, x_j) = \tanh \beta_o x_i . x_j + \beta_1$,

where x_i, x_j are the two feature vectors, p is the degree of the polynomial, and β_o, β_1 are two constants. Except for the kernel functions, both SVM classifiers require other hyperparameters as well which include the regularization parameter, tolerance, and gamma which specifies the influence of plausible line of separation on the model's performance. To train the two SVM classifiers, a batch learning or offline learning mechanism is applied where samples are randomly picked and fed to the model to minimize the loss after each iteration.

4.2.2 Quantum simulation

The quantum SVM classification model is trained similarly as done in the case of the classical one once the quantum kernel is obtained. Therefore, the process of training a quantum SVM model can be explained in the following steps.

4.2.2.1 Obtaining the quantum kernel

The acquired quantum states that were translated by the quantum feature maps given by $V(\Phi(\vec{x}))$ enable the construction of the quantum kernel. Once the kernel matrix is successfully obtained on the quantum system, training of the model is performed similarly. The quantum kernel is computed by taking the inner product of the quantum feature maps. As these quantum feature maps are hard to simulate on classical computer systems, we aim to obtain the quantum advantage from the quantum simulation. Employing the following ansatz [30]:

$$V(\Phi(\vec{x})) = U(\Phi(\vec{x})) \otimes H^n,$$

lets two qubits interact at a time which leads to interaction ZZ and non-interaction term Z. Here, H^n is the Hadamard gate that is applied to each qubit, and n denotes the number of qubits. As $n = 2$ in our case, we repeat the ansatz twice for the depth of $S = 2$. This leads the feature map to take the following form:

$$V(\Phi(\vec{x})) = U(\Phi(\vec{x})) \otimes H^n \otimes U(\Phi(\vec{x})) \otimes H^n.$$

Then, information about the quantum kernel is needed to be extracted from the quantum circuit to feed the same to the training algorithm which is a non-trivial task. This is so because, the kernel $K(x_i, x_j) = |\langle \Phi(x_i)|\Phi(x_j)\rangle|^2$ consists of two overlapped states which are required to be measured. According to a study [30], the circuit shown in Figure 4 allows the estimation of this overlap.

This is based on the fact that if $V(\Phi(\vec{x})) = V(\Phi(\vec{z}))$ the entire unitary would unravel to the unitary matrix and the info ought to be equivalent to the yield.
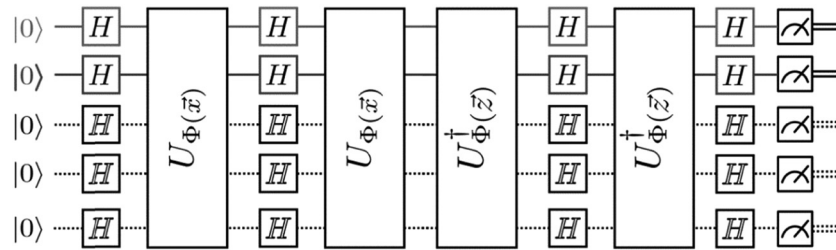
Figure 4: Estimating the $|\langle \Phi(x_i)|\Phi(x_j)\rangle|^2$ overlap for a circuit of depth $S = 2$, where $V(\Phi(\vec{x})) = U(\Phi(\vec{x})) \otimes H^n \otimes U(\Phi(\vec{x})) \otimes H^n$.

4.2.2.2 Building the feature map and simulating quantum SVM using the Qiskit library

To generate the quantum circuit for the feature map $V(\Phi(\vec{x}))$, we have used the SecondOrderExpansion function, which is provided by the Qiskit Aqua library. This function takes feature dimensions (which is the same as the number of qubits) and depth which is defined as the number of recurrences of the feature map.

Also, to train the quantum SVM classifying model, Qiskit provides a pre-defined training function that takes the feature map along with the training and testing set for each fold. Applying the quantum kernel and obtaining the dual formulation allow the quantum SVM to perform the classification with less computational cost. The quantum SVM model minimizes the loss by optimizing the parameter W as follows:

$$L(W) = \sum_i w_j - \frac{1}{2} \sum_{ij} y_i y_j \alpha_i \alpha_j K(x_i, x_j), \tag{4}$$

Finally, the constructed model is sent to IBM Quantum Experience where the model is simulated, and its performance is evaluated by a real quantum computer and the results are fetched back for tuning the hyperparameters and minimizing the loss.

4.3 Hyperparameter tuning and testing

All three of the generated SVM classifiers comprise several crucial parameters that are required to be tuned according to the problem taken into consideration, which in this case is a binary task of malignant breast cancer tumor diagnosis. In our case, the linear SVM classifier requires the regularization parameter, tolerance parameter, and penalty type as a parameter. Nonlinear SVM classifier requires the regularization parameter, kernel type, degree of the kernel, tolerance parameter, and gamma as a parameter. And the quantum SVM requires depth and feature dimensions as a necessary set of parameters to function. Therefore, to obtain the optimal hyperparameters for the models, we have used the GridSearchCV technique provided by the Scikit-learn library that takes all the potential optimal values of the considered hyperparameters and brute forces the classifier's performance over every pair of these hyperparameters possible to estimate the best configuration, which leads the classifying model to the minimalized loss. The value of the tuned parameters for each of the three classifiers is discussed in Section 5.

5 Results

The three simulations taken into consideration gave the expected outcome for the very same dataset. Two of these classical simulations, i.e., linear and nonlinear SVM, were carried on a classical computer system using the Python Scikit-learn library, whereas the third simulation was carried out using a quantum computer simulator IBM cloud and the Qiskit library. The trained quantum SVM model was successfully able to perform the binary classification task and thereby diagnosing the malignant breast cancer tumor. Out of the three tuned SVM classifiers (linear, nonlinear, and quantum classifying models), the

performance of the nonlinear and quantum SVM classifiers was significantly improved as compared to the linear SVM classifier. The evaluation metric scores for each of the classifiers were evaluated by estimating the mean scores of the model on each of the folds of the k-fold cross-validation technique. Figure 5 shows the testing accuracy scores of the three classifiers on the breast cancer classification dataset.
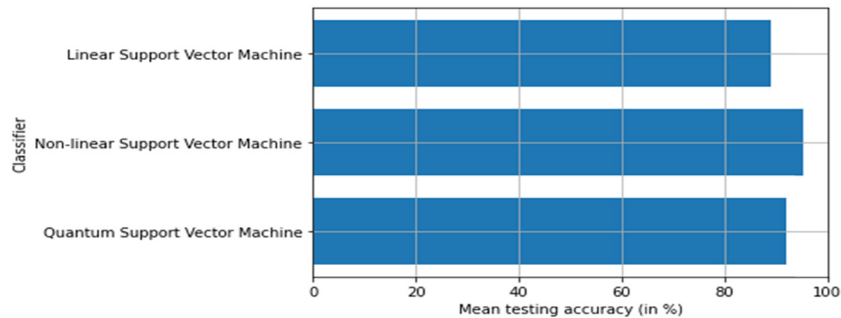

Figure 5: Mean testing accuracy of the three SVM classification model on the Breast Cancer Dataset.

In both the classical and quantum simulations, the corresponding linear, nonlinear, and quantum SVM classifiers were trained using the two principal components and the respective label values for every data sample. The configured set of optimized hyperparameters for which the three classification models produced the minimized classification loss are shown in Tables 1–3, respectively.

In Tables 1 and 2, C (regularization parameter or penalty parameter) represents the misclassification rate and controls the strength of regularization which acts inversely proportional to C, Tol (tolerance parameters) defines the tolerance for the termination criterion and gamma describes the influence of data points in the calculation of the decision boundary. While comparing the performance of the quantum and conventional models, it was observed that in the case of the quantum SVM model, FP = 29 and FN = 16, whereas for the conventional SVM models, Type II error in the case of the linear model was the highest, coming out to be 32 and the same was the least in the case of the nonlinear model with a value of 16. A similar trend can be observed in terms of the Type I error, where it was the most in the case of the linear classifier, the minimum in the case of the nonlinear classifier, and balanced for the quantum SVM classifier (Figure 6). This makes the nonlinear SVM and quantum SVM classifier much more reliable models for the classification task as compared to the linear model.

Table 1: Tuned hyperparameters for linear support vector machine classifier

Linear-support vector machine classifier	Configuration
C (regularization parameter)	1.0
Penalty type	l2 regularization
Tol (tolerance parameter)	0.0001

Table 2: Tuned parameters for nonlinear support vector machine classifier

Nonlinear support vector machine classifier	Configuration
Kernel type	Polynomial
Degree of polynomial	3
C (regularization parameter)	100
Gamma	1
Tol (tolerance parameter)	0.001

Table 3: Tuned for quantum support vector machine classifier

Parameters	Configuration
Feature dimensions	2
Number of qubits	2
Depth	2

a)

b)

c)
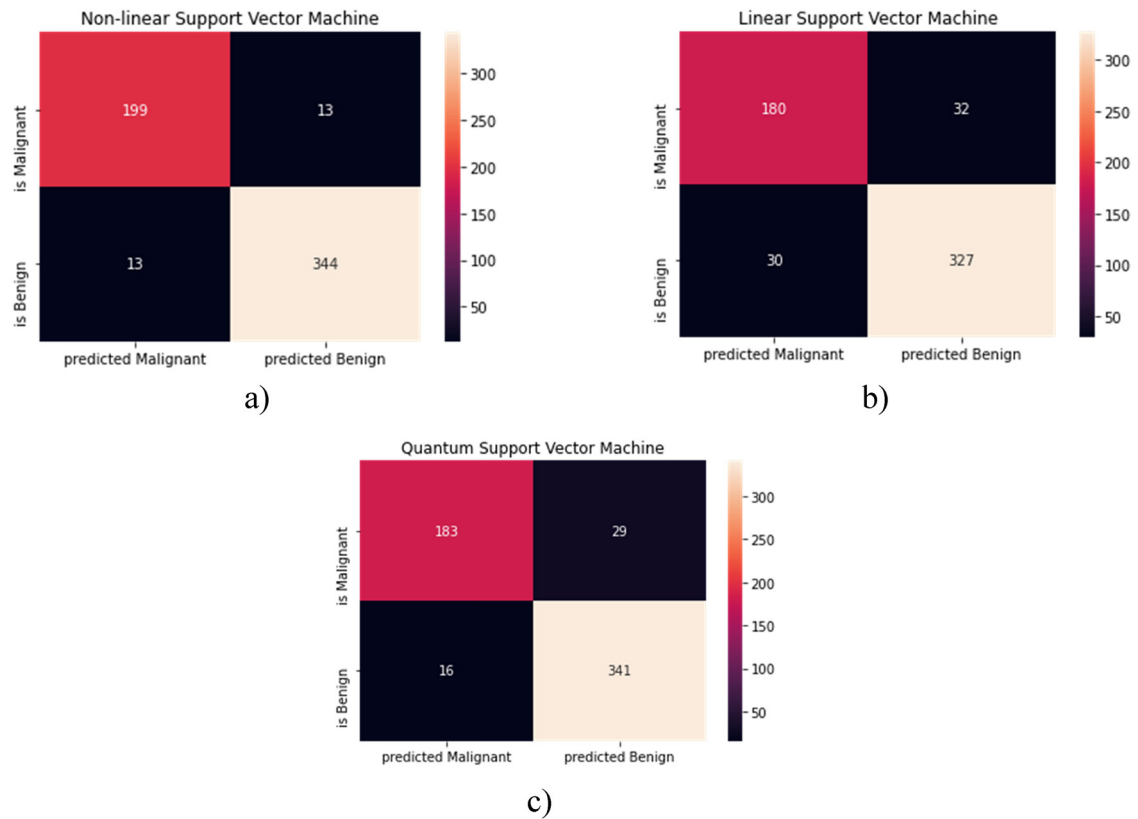
Figure 6: Confusion matrix for (a) linear SVM model, (b) nonlinear SVM model, and (c) quantum SVM model.

Upon comparing the three models (trained on two constrained principal components) concerning various evaluation metrics, the linear SVM in the case of classifying the malignant samples had the lowest scores (precision, recall, and F1-score) of approximately 0.85. This can be explained because of the absence of the kernel trick that is used in the case of the nonlinear SVM model. This absenteeism of the kernel trick makes it much more challenging for the linear model to perform classification on complex data in lower dimension feature space. On the other hand, the precision and F1-score for quantum SVM come out to be 0.92 and 0.89, respectively (as shown in Table 4), which is a significant improvement over the linear SVM. However, the most promising results were achieved in the case of the nonlinear SVM classifier with a maximum score of 0.94 on every evaluation metric taken into consideration.

While training the quantum SVM classifier, the evaluation of quantum kernel K and solving (2) to obtain $|b, \alpha\rangle$ were carried out inside the quantum systems, whereas the classification task using the data samples x, and the results of previous training $|b, \alpha\rangle$ are performed using the classical system giving it the structure of a variational algorithm. Figure 7a illustrates the classification achieved by the quantum SVM classifier for the first fold out of the total number of folds ($k = 10$) in the cross-validation technique during its training.

The proposed quantum SVM model gave satisfactory results on most of the evaluation metrics for the problem of performing malignant breast cancer diagnosis. Despite the restricted access to a large qubit

quantum system, the quantum SVM model overshadowed the linear SVM classifier and was able to achieve a very similar precision score for classifying the malignant samples. Table 5 shows the overall performance report on all the three simulations performed on a classical and quantum computing environment concerning their theoretical worst-case time complexities.
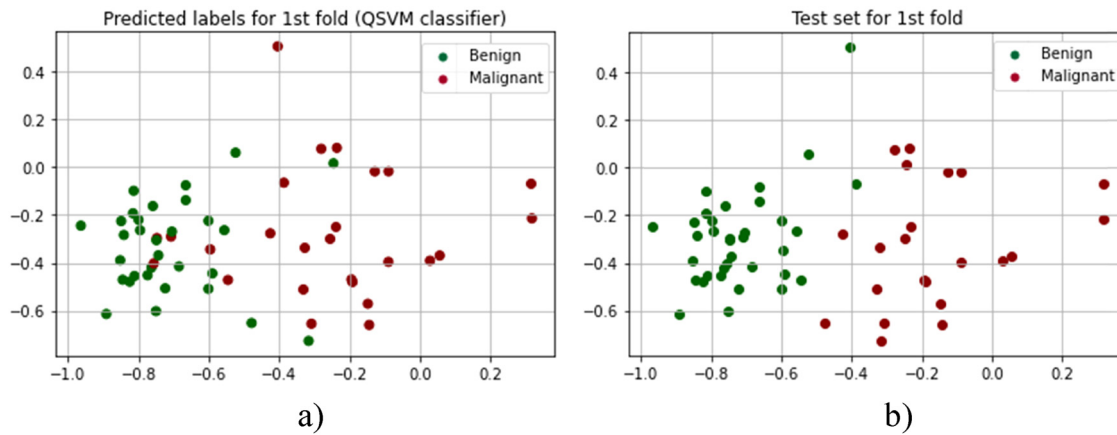

a)   b)

Figure 7: Visualizing classification achieved by quantum SVM: (a) predicted labels on (b) test set for the first fold during cross-validation.

Table 4: Precision, recall, and F_1-score for the three SVM variants

Classifier	Class	Precision	Recall	F_1-score
Linear support vector machine	Malignant	0.86	0.85	0.85
	Benign	0.91	0.92	0.91
Nonlinear support vector machine	Malignant	0.94	0.94	0.94
	Benign	0.96	0.96	0.96
Quantum support vector machine	Malignant	0.92	0.86	0.89
	Benign	0.92	0.96	0.94

Table 5: Performance tradeoff for the three SVM variants

Classifier	Accuracy	Theoretical worst-case time complexity
Linear SVM [28]	0.89	$O(M \times N)$
Nonlinear SVM [29]	0.95	$O(M^2(M + N))$
Quantum SVM	0.92	$O(\log(M \times N))$

Even though the nonlinear SVM model produced the most prominent results by attaining the testing accuracy score of 0.95 utilizing the kernel trick, the theoretical worst time complexity for the same comes down to $O(M^2(M + N))$, which is significantly high as compared to that of the quantum SVM model. Moreover, as the size or the number of features in the dataset increases, the performance of the nonlinear SVM model will degrade exponentially. Comparing the time complexities and testing accuracy of the other two models, the linear SVM model, although it has a considerably low time complexity of $O(M \times N)$, was not able to fit on malignant samples that well (testing accuracy of 0.89) due to the absence of the kernel trick operation. Therefore, it is safe to say that the quantum SVM model is successful in providing an exponential speedup in training complexity over what is being offered by its conventional variants. With

the training time complexity of $O(\log(M \times N))$ and an accuracy score of 0.92, quantum SVM is effective in maintaining the tradeoff between the time complexity and the classification accuracy of the models.

6 Discussion

Many proficient techniques have been deployed to solve the classification task of breast cancer diagnosis. Many of these conventional approaches often rely on a machine learning formulation to crack the problem. In most instances, a learning-based approach aids in the production of a highly accurate and precise classification model. While following such an approach, one of the drawbacks that is often left unresolved is the computational efficiency of the generated machine learning model. This study, therefore, intends to explore one such approach for minimizing the computational cost of traditional machine learning classifiers by constructing and analyzing a quantum SVM classifier that resolves the breast cancer classification problem by following a much more computation friendly manner.

7 Conclusion and future implications

The proposed quantum SVM model in this study was successful in solving the task of binary classification of a malignant breast cancer diagnosis. The presented model utilizes the concepts of quantum computing to attain exponential speed-up over conventional machine learning classifiers. This study illustrated the complete working of the quantum SVM classification model which guaranteed an exponential speed-up over its typical variant. This research also effectively demonstrated a comparative analysis of distinct forms of SVM algorithms, namely linear, nonlinear, and quantum SVM, concerning their time complexities and various evaluation metrics which exemplified the supremacy of quantum SVM over the conventional SVM algorithms.

The quantum SVM classifier works sufficiently well with dense training vectors. However, the major limitation of this research revolves around the number of features taken into consideration while training the classifiers as the two principal components cumulatively can only explain up to 70% of the variance in data. The simulation and analysis of the QSVM model on a complex and large-scale dataset utilizing a large qubit available quantum simulator is still one of the major challenges at this time of research. This research justifies how the domain of quantum computing could benefit conventional machine learning models. With increasing investment and research in quantum technology, studies in resolving such limitations will increase substantially. Moreover, access to quantum systems with a greater number of qubits is expected to increase considerably which will further promote the generation of quantum machine learning models on much more sizeable datasets for solving complex machine learning problems. This will enable conventional machine learning algorithms to be redesigned in a manner that can enable their simulation in the quantum environment allowing a significant reduction in computation cost without compromising on their prediction performance.

Conflict of interest: Authors state no conflict of interest.

References

[1]   Waks AG, Winer EP. Breast cancer treatment: a review. Jama. 2019;321(3):288–300.
[2]   Gupta S, Kumar D, Sharma A. Data mining classification techniques applied for breast cancer diagnosis and prognosis. Indian J Comput Sci Eng (IJCSE). 2011;2(2):188–95.

[3]   Salama GI, Abdelhalim M, Zeid MA-E. Breast cancer diagnosis on three different datasets using multi-classifiers. Breast Cancer (WDBC). 2012;32(569):2.

[4]   Paparo GD, Dunjko V, Makmal A, Martin-Delgado MA, Briegel HJ. Quantum speedup for active learning agents. Phys Rev X. 2014;4(3):031002.

[5]   Murty MN, Raghava R. "Kernel-based SVM." Support vector machines and perceptrons. Cham: Springer; 2016. p. 57–67

[6]   Mangasarian OL, Musicant DR. Lagrangian support vector machines. J Mach Learn Res. 2001;1(Mar):161–77.

[7]   Biamonte J, Wittek P, Pancotti N, Rebentrost P, Wiebe N, Lloyd S. Quantum machine learning. Nature. 2017;549(7671):195–202.

[8]   Rebentrost P, Mohseni M, Lloyd S. Quantum support vector machine for big data classification. Phys Rev Lett. 2014;113(13):130503.

[9]   Aleksandrowicz G, Alexander T, Barkoutsos P, Bello L, Ben-Haim Y, Bucher D, et al. Qiskit: an open-source framework for quantum computing. Zenodo; 2019. Accessed on: March, 16.

[10] Polat K, Güneş S. Breast cancer diagnosis using least square support vector machine. Digital Signal Process. 2007;17(4):694–701.

[11] Lloyd S, Mohseni M, Rebentrost P. Quantum self analysis. Arxiv Prepr 1307. 2013.

[12] Harrow AW, Hassidim A, Lloyd S. Quantum algorithm for linear systems of equations. Phys Rev Lett. 2009;103(15):150502.

[13] Berry DW, Ahokas G, Cleve R, Sanders BC. Efficient quantum algorithms for simulating sparse Hamiltonians. Commun Math Phys. 2007;270(2):359–71.

[14] Ahmad LG, Eshlaghy AT, Poorebrahimi A, Ebrahimi M, Razavi AR. Using three machine learning techniques for predicting breast cancer recurrence. J Health Med Inf. 2013;4(124):3.

[15] Di Buccio E, Li Q, Melucci M, Tiwari P. Binary classification model inspired from quantum detection theory. Proceedings of the 2018 ACM SIGIR International Conference on Theory of Information Retrieval. New York, NY, United States: Association for Computing Machinery; 2018.

[16] Li Z, Liu X, Xu N, Du J. Experimental realization of a quantum support vector machine. Phys Rev Lett. 2015;114(14):140504.

[17] Wiebe N, Kapoor A, Svore KM. Quantum nearest-neighbor algorithms for machine learning. Quant Inf Comput. 2018;15(3–9):318–58.

[18] Dong D, Chen C, Li H, Tarn TJ. Quantum reinforcement learning. IEEE Trans Systems, Man, Cybern, Part B (Cybernetics). 2008;38(5):1207–20.

[19] Virmani J, Dey N, Kumar V. PCA-PNN and PCA-SVM based CAD systems for breast density classification. Applications of intelligent optimization in biology and medicine. Cham: Springer; 2016. p. 159–80

[20] Cheriguene S, Azizi N, Zemmal N, Dey N, Djellali H, Farah N. Optimized tumor breast cancer classification using combining random subspace and static classifiers selection paradigms. Applications of intelligent optimization in biology and medicine. Cham: Springer; 2016. p. 289–307

[21] Samani NA, Aghababa H. Application of quantum gradient descent as a learning algorithm for factorization machines: quantum artificial intelligence. Proceedings of the 10th Hellenic Conference on Artificial Intelligence. New York, NY, United States: Association for Computing Machinery; 2018.

[22] Zemmal N, Azizi N, Dey N, Sellami M. Adaptive semi supervised support vector machine semi supervised learning with features cooperation for breast cancer classification. J Med Imaging Health Inform. 2016;6(1):53–62.

[23] Benzebouchi NE, Azizi N, Ayadi K. A computer-aided diagnosis system for breast cancer using deep convolutional neural networks. Computational intelligence in data mining. Singapore: Springer; 2019. p. 583–93

[24] Aïmeur E, Brassard G, Gambs S. Quantum clustering algorithms. Proceedings of the 24th International Conference on Machine Learning. New York, NY, United States: Association for Computing Machinery; 2007.

[25] Kotsiantis SB, Zaharakis I, Pintelas P. Supervised machine learning: a review of classification techniques. Emerg Artif Intell Appl Comput Eng. 2007;160:3–24.

[26] Dua D, Graff C. UCI machine learning repository. Irvine, CA: University of California, School of Information and Computer Science; 2019. http://archive.ics.uci.edu/ml

[27] Fushiki T. Estimation of prediction error by using K-fold cross-validation. Stat Comput. 2011;21(2):137–46.

[28] Hsieh CJ, Chang KW, Lin CJ, Keerthi SS, Sundararajan S. A dual coordinate descent method for large-scale linear SVM. Proceedings of the 25th International Conference on Machine Learning. New York, NY, United States: Association for Computing Machinery; 2008.

[29] Platt J. Sequential minimal optimization: a fast algorithm train support vector machines. Johns Hopkins University Baltimore, Maryland: Microsoft; 1998.

[30] Havlíček V, Córcoles AD, Temme K, Harrow AW, Kandala A, Chow JM, et al. Supervised learning with quantum-enhanced feature spaces. Nature. 2019;567(7747):209–12.