Quantum Convolutional Neural Networks with Interaction Layers for Classification of Classical Data

Jishnu Mahmud[1], Raisa Mashtura[2], Shaikh Anowarul Fattah[3*], Mohammad Saquib[4]

[1,2,3]Department of Electrical & Electronic Engineering, Bangladesh University of Engineering & Technology, Dhaka, 1000, Bangladesh.
[4]Department of Electrical Engineering, The University of Texas at Dallas, Texas, 75083-0688, USA.

*Corresponding author(s). E-mail(s): fattah@eee.buet.ac.bd;
Contributing authors: jishnu.mahmud@gmail.com;
raisamashtura@gmail.com; saquib@utdallas.edu;

Abstract

Quantum Machine Learning (QML) has come into the limelight due to the exceptional computational abilities of quantum computers. With the promises of near error-free quantum computers in the not-so-distant future, it is important that the effect of multi-qubit interactions on quantum neural networks is studied extensively. This paper introduces a Quantum Convolutional Network with novel Interaction layers exploiting three-qubit interactions, while studying the network's expressibility and entangling capability, for classifying both image and one-dimensional data. The proposed approach is tested on three publicly available datasets namely *MNIST*, *Fashion MNIST*, and *Iris* datasets, flexible in performing binary and multiclass classifications, and is found to supersede the performance of existing state-of-the-art methods.

Keywords: Quantum Machine Learning, classification, entanglement, quantum gates, qubits.

1 Introduction

In this era of artificial intelligence a constant improvement in computation speed, accuracy, and precision is a necessity. This widespread success in the world of computing over the last decade can be attributed to both the development of efficient software algorithms and the advancements in computational hardware. However, the physical limits of semiconductor fabrication in the post-Moore's Law era raise concerns about the extrapolation of its effectiveness in the future. On the other hand, significant advancements have been made in the field of quantum computing, which has shown promise as a potential solution for modern computing problems. Quantum computing exploits the laws of quantum mechanics to store and process information in quantum devices, using qubits instead of classical bits, which enables them to solve problems intractable for classical computers (Arute et al (2019)). Qubits have several properties that make them superior to the classical bit (Yanofsky and Mannucci (2008)). Firstly, they are the superposition of the two fundamental bit states. This property enables a collective number of qubits to have exponentially higher processing power than the same number of classical bits. The second noteworthy property is entanglement, which creates interdependencies among the qubits, suggesting that the change of state of one will affect the state of the qubit it is entangled with. Superposition, Entanglement along with Interference, which is a method of controlling the qubit wavefunctions to reach their desired states, are the three properties that give qubits their formidable potential.

The era of quantum computing, currently referred to as the Noisy Intermediate Scale Quantum (NISQ) era, is characterized by the lack of absolute control over the qubits due to errors arising from quantum decoherence, crosstalk, and imperfect calibration, thereby limiting the number of qubits used on quantum computers. However, the revelation in January 2022 that quantum computing in silicon hit 99% fidelity (Madzik et al (2022)) indicates a greater similarity between the desired and actual quantum states. This result promises near-error-free quantum computing and indicates that they are close to being utilized in large-scale applications, which further motivates the development of various machine learning algorithms to be implemented on quantum devices.

2 Related Works

Quantum machine learning (QML) involves constructing a sequential circuit of different types of parameterized quantum gates that act on a specific number of qubits. The main tasks of these QML algorithms are to logically place these parameterized gates and train their parameters to minimize the objective cost function. QML has already been implemented to address one of the most fundamental machine learning problems, i.e., classification. It is shown in Rebentrost et al (2014) and Mengoni and Di Pierro (2019) that kernel-based algorithms, such as Quantum-enhanced Support Vector Machine (QSVM) can classify data efficiently and accurately. The search for devising various convolutional networks in the quantum domain is introduced in Cong et al (2019) in which the concept of quantum convolutional neural networks

(QCNNs) is proposed. Their architecture also claims to be able to tackle the exponential complexity of many-body problems, making them suitable for use in quantum physics problems. Advancing the field of QCNNs, a parameterized quantum neural network (QNN) with an enhanced feature mapping process has been designed in Liu et al (2021). Their proposed network is called a quantum-classical CNN (QCCNN) which is suitable for NISQ devices. It appears that several quantum counterparts of a large variety of classical machine learning models have been proposed over the recent years, all of which claim superior performances in various categories, such as accuracy and speed. Therefore, it is of no surprise that quantum networks have been shown to have a wide range of applications in medicine (Jain et al (2020)), weather forecasting (Enos et al (2021)), quantum chemistry (Von Lilienfeld (2018)), and many more.

Ever since the proposal of QCNNs and the availability of quantum simulators and quantum computers, much attention has been drawn to devising various methods to improve the performance of classification problems. This is driven by the fact that QCNN models are immune to barren plateau problems (Pesah et al (2021)) contrary to other structures. The architecture proposed in Hur et al (2022) has been benchmarked for binary image classification on the *MNIST* and *Fashion MNIST* datasets. A multiclass classification method using a quantum network is also reported in Chalumuri et al (2021), which has been proven to perform well on 1D data such as the *Iris dataset.* All these prior studies confirm that a QNN aids speed with a significantly lower number of parameters with better accuracy than their existing classical counterparts using a comparable number of training parameters. However, a crucial aspect of designing parameterized quantum circuits is to maintain sufficient expressibility and entangling capability, while keeping it cost-effective Sim et al (2019). Expressibility is the ability of a quantum circuit, comprising quantum gates to explore the Hilbert space. The cost of a quantum circuit is judged by the number of layers and hence its parameters, as well as the number of two-qubit operations. This paper explores the relative changes in the performance of a QCNN network upon the incorporation of three-qubit interactions while maintaining a relatively low depth and a small number of parameters for comparability. Although three-qubit gates are practically difficult to implement on NISQ devices and their synthesis using 1 and 2-qubit interactions exponentially increases depth, it is important to explore the comparative changes in the performance of a network resulting from their addition which is expected to be a reality considering the rapid growth in the performance of quantum hardware in recent times. These three-qubit interactions are brought forward using novel Interaction Layers in our network which use a minimal number of trainable parameters. Furthermore, to explore the performance of the proposed network we have used an ancilla-based classifier as the final layer of the circuit to carry out binary and multiclass classification tasks.

The major contributions of this work are summarized as follows:

1. A new QCNN architecture is proposed which uses *Amplitude* and *Angle Encoding* schemes separately, considering two different data reduction and encoding techniques.

2. Novel Interaction Layers are introduced, which exhibit sufficient expressibility and exploit the entanglement property of qubits further to help the quantum network learn more nuanced information from the data.

3. A classifier layer involving ancilla qubits and $CNOT$ gates are cascaded with the quantum convolutional structure to accommodate both binary and multiclass classifications.

The use of ancilla qubits in the classifier layer cascaded with the Interaction Layers in a QCNN structure is a first to the best of our knowledge.

4. The proposed network is tested on three publicly available datasets for binary and multiclass classification, and it is seen that the performance supersedes that of the existing state-of-the-art models using a similar number of parameters. The versatility of the network is further demonstrated in its ability to perform equally well in both image and 1-dimensional data.

3 Proposed Architecture

A simplified block diagram representation of the proposed architecture is depicted in Fig.1. A quantum circuit, designed in the spirit of QCNN structure and possessing minimal trainable parameters, has been proposed. The robustness of QCNNs against the "barren plateau" issue is expected to be exhibited by the proposed network (Pesah et al (2021)).

Interaction Layers are introduced in various stages of the proposed quantum architecture and are designed to leverage three-qubit interactions through the use of *Toffoli* and parameterized rotational gates. The implementation of these Layers in various stages of the network can be observed in Fig.1. It differs from the earlier quantum convolutional methods, which relied on the reduction of qubits through sequences of convolutional and pooling layers alone. The rapid advancement of quantum hardware indicates that considerably more intricate operations on qubits will soon be possible. Although it is practically challenging to achieve three-qubit interactions on current technology, this paper investigates the comparative advantage in a network's performance with the addition of such layers. It must be noted that the number of trainable parameters and the total number of qubits have been kept minimum such that they can be implemented on NISQ devices for the purpose of comparison with other methods. The paradox of deploying three-qubit interactions on the network that was, at its core, intended to operate on NISQ devices is acknowledged by the authors. When moving from NISQ to more potent quantum computers, more qubits would be used, the circuit depth would expand, and there would be more qubit interactions with multi-qubit gates. By keeping a small number of qubits and trainable parameters, this study aims to evaluate how expanded multi-qubit interaction improves QCNN network performance in comparison to networks limited to two-qubit operations.

It is expected that the incorporation of these novel Interaction Layers will enable the network to have the ability to extensively span the Hilbert space as well as exploit the entanglement property further for improved classification performance. The use

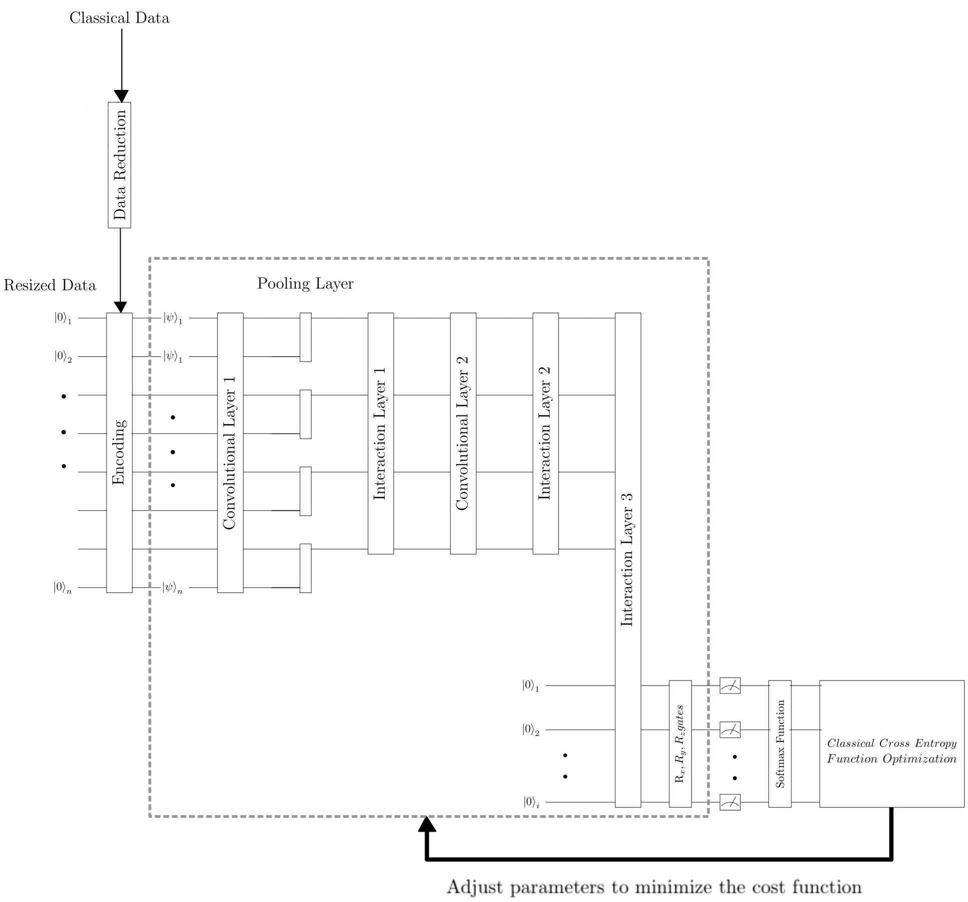

Fig. 1 Simplified Block diagram of the proposed architecture

of Toffoli gates, enabling three-qubit interactions in QCNN networks is a first to the best of our knowledge.

3.1 Data Preprocessing

The number of qubits and, therefore, the size of a QNN is bound by the current limitations of NISQ computing technology, in contrast to classical models, which often possess many trainable parameters due to their substantial size and depth. In the next stage where quantum feature encoding is performed, the features of the data to be classified are inserted as parameters of quantum gates, which perform various operations on these qubits. Therefore, a limited number of qubits also sets a bar on the total number of gate parameters; thus, the dimensionality reduction of classical data prior to its utilization within a quantum network is deemed imperative.

Standard classical techniques, such as the autoencoder and simple resizing, are chosen as they allow for efficient compression of high-dimensional input data, which is important for reducing the computational complexity of quantum machine learning models. The autoencoder is particularly useful in this regard, as it can learn to represent the input data of dimensions $p \times p$, to a lower q-dimensional space, $q < p$, extracting a reduced set of features of size $q \times 1$, while still preserving important features and minimizing information loss. As an alternative, the simple resizing operation can also be effective in reducing the dimensionality of input data. It converts input data of dimensions $p \times p$ to a desired dimension of $q \times q$, $q < p$.

3.2 Quantum Feature Encoding

The projection of the reduced classical values, received as output from the previous layer, into quantum states is referred to as quantum feature encoding. Mathematically, the mapping of the classical input data, X, into higher dimensional quantum states, represented in the Hilbert space and denoted by H, is represented as

$$\phi : X \mapsto H$$

where ϕ is the feature map. In this stage of the network, n qubits initialized to the state of $|0\rangle$, are fed. The qubits are then subjected to state operations via quantum gates, parameterized by the classical data X which is the output of the block performing classical data reduction. This results in the mapping of the classical data to the Hilbert space and the resulting state is represented by $|\psi\rangle$. This process of quantum state preparation encodes the classical values into the input qubits which can then exploit the unique properties of superposition, entanglement, and interference to achieve superior performance. Two widely employed state preparation techniques, known as amplitude and *Angle Encoding*, have aided in achieving significant results and are discussed in subsequent sections.

3.2.1 Amplitude Encoding

In this encoding scheme, the normalized classical vectors from the Data Preprocessing Layer are represented as amplitudes of the n input qubits in the Quantum Feature Encoding Layer. This displays a particular quantum advantage as normalized feature vectors of size 2^n can be encoded into only n-qubits (Schuld and Petruccione (2018)). The following equation shows the states prepared after performing *Amplitude Encoding* on the input qubits.

$$|\psi_x\rangle = \sum_{i=1}^{N} x_i |i\rangle \tag{1}$$

Here $|\psi_x\rangle$ is the quantum state corresponding to the N-dimensional classical datapoint X after reduction, where $N = 2^n$, x_i is the i-th element of the datapoint X and $|i\rangle$ is the i-th computational basis state.

In a classical neural network, each binary value necessitates a distinct trainable weight or bias, resulting in a considerable number of parameters. In contrast, *Amplitude Encoding* permits the representation of data through the amplitudes of a limited number of quantum states, thereby enabling a more compact representation. This has been demonstrated to result in a significant decrease in the number of trainable parameters, contributing to the simplification of the model and enhancement of its performance. While this method provides this benefit, it also increases the depth of the quantum circuit as *O(poly(n))* or as *O(n)* if the number of qubits fed in this layer is increased (Araujo et al (2021)).

3.2.2 Angle Encoding

Angle Encoding is another technique employed in quantum machine learning for the representation of data, which utilizes the rotation of quantum gates (R_x, R_y and R_z) to encode classical information. This method involves encoding the N features of classical data as the angles of n input qubits between quantum states (Schuld (2021)). In this method, N has been kept equal to n to allow us to use the maximum size of classical features possible. The advantage of this approach lies in its ability to represent continuous data more naturally and efficiently compared to *Amplitude Encoding* (Schuld (2021)). The states resulting from performing *Angle Encoding* on the input qubits are:

$$\left|\psi_x\right\rangle = \otimes_{i=1}^{n} R(x_i)\left|0^n\right\rangle \tag{2}$$

Here $R(.)$ can be either of the rotation gates R_x, R_y, or R_z. In *Angle Encoding*, the angles between the quantum states can be varied continuously to capture the intricacies of the data. This leads to a more precise and nuanced representation of the data and can result in improved performance for certain types of quantum machine learning models. Although, unlike *Amplitude Encoding*, it can only encode one qubit with one feature value, resulting in the reduction of noise, which makes it particularly advantageous in NISQ computing.

The selection of encoding techniques for this design is contingent upon the classical dimensionality reduction technique employed in the first layer. It can be recalled from the previous section that the *Amplitude Encoding* method, which uses n input qubits, can accommodate a maximum of 2^n data points. This requires the use of simple resizing to $2^{n/2} \times 2^{n/2}$ dimension followed by flattening, which is essential according to this state preparation method. Conversely, the *Angle Encoding* technique encodes the flattened N data points into n qubits and thus relies on the use of an autoencoder to reduce the dimensions accordingly.

3.3 Proposed Layers

3.3.1 The Quantum Convolutional Layer

The proposed model for the classification problem is comprised of two Convolutional Layers with one Pooling Layer in between. As shown in Fig.2, each of these layers is constructed of blocks of quantum gates called ansatzes, which are parameterized quantum circuits. In this study, an ansatz is composed of different configurations of single and multi-gate operations as illustrated in Fig.3 and 4.

The first ansatz in Fig.3 consists of a relatively large number of parameters, i.e., 15, which helps increase flexibility, and the controlled R gates help increase expressibility. The ansatz in Fig.4 has five fewer parameters, which is a parametrized form of a reduced version of the circuit that recorded the best expressibility in a study carried out by Sim et al (2019). The ansatzes consist of the R_x, R_y, R_z gates, which cause qubit rotations about the x, y, and z axes, respectively. Ansatz 1 additionally has the $U3$ gate, which can be decomposed to rotation and phase shift gates and is represented by the matrix as follows:

$$U3(\theta_1, \theta_2, \theta_3) = \begin{bmatrix} \cos(\frac{\theta_1}{2}) & -\mathrm{e}^{i\theta_3}\sin(\frac{\theta_1}{2}) \\ \mathrm{e}^{i\theta_2}\sin(\frac{\theta_1}{2}) & \mathrm{e}^{i(\theta_2+\theta_3)}\cos(\frac{\theta_1}{2}) \end{bmatrix}$$

Here, θ_1, θ_2, and θ_3 are the parameters of the $U3$ gate, and the matrix represents a unitary operation on a qubit. The reason for the selection of two different ansatzes is to inspect the flexibility of the proposed network performance on slight changes in the structure of the ansatz and the number of trainable parameters.

3.3.2 The Quantum Pooling Layer

The main purpose of the Pooling Layer in any convolutional neural network is to reduce the spatial size of the data representation and to maintain the most important information. In the process, the layer helps reduce the computational cost of the network and improve its generalization capabilities by decreasing overfitting, making it robust to translations, rotations, and other minute changes.

The quantum Pooling Layer in Fig.5 traces out one qubit from the two qubits it is fed and thus reduces the two-qubit states to a one-qubit state. The Pooling Layer uses two controlled rotation gates and a $Pauli - X$ gate. The Pooling Layers, along with the Convolutional Layers, extend qubit interactions beyond nearest neighbors and hence establishes further dependencies. The output qubits are further processed by an Interaction Layer to aid the learning process.

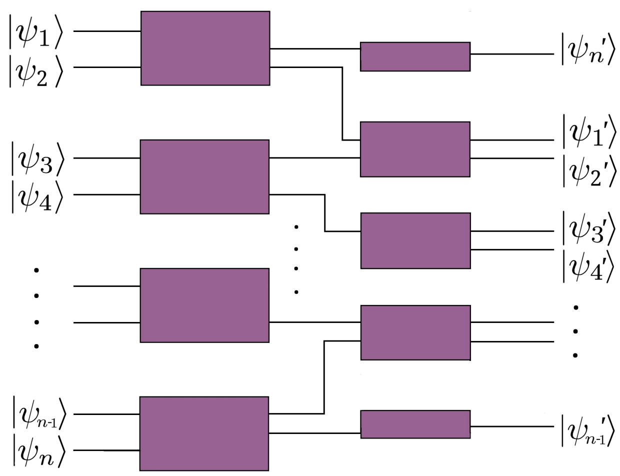

Fig. 2 The Convolutional Layer comprising of ansatzes, all of which use two-qubit interactions. The input qubits have their states changed by parameterized gate operations in the layer

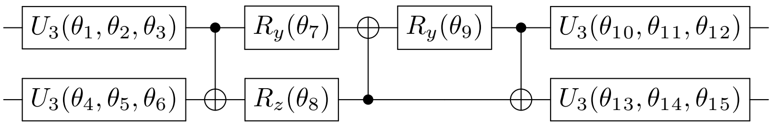

Fig. 3 Ansatz 1 containing 15 trainable parameters

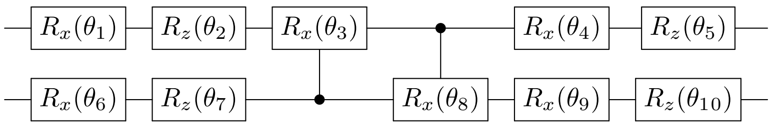

Fig. 4 Ansatz 2 containing 10 trainable parameters

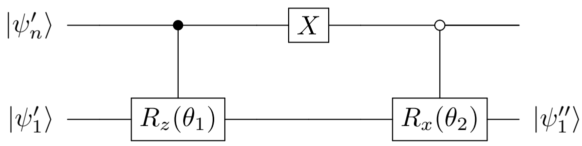

Fig. 5 The ansatz used for building the Pooling Layer

3.3.3 The Interaction Layers

The novelty brought forward in the proposed quantum architecture involves making variational quantum layers designed to introduce extensive entanglement and expressibility in the overall quantum network. In order to bring forward this special quantum phenomenon, in this proposed *Interaction layer*, *Toffoli* gates are cascaded with the convolutional and rotational gates and are expected to establish three-qubit interactions, as shown in Fig.6 and 7. In order to bring forward this special quantum phenomenon in this proposed layer, *Toffoli* and *CNOT* gates are cascaded with the convolutional and rotational gates and are expected to establish three and two-qubit interactions, as shown in Fig.6 and 7.

In this era of NISQ computing, the implementation of two-qubit gates is more difficult than single-qubit ones. However, with the recent developments in quantum hardware and the promise of near error-free quantum computers in the not-so-distant future, the use of multi-qubit gates in quantum computers is expected to be more common. It must be noted that the use of two-qubit gates, such as the *CNOT* gate, in quantum machine learning models is motivated by the increase in entanglement and expressibility of the network, which helps it to learn more complex features of classical data Sim et al (2019). It is therefore imperative that studies are conducted exploring the effectiveness of three-qubit gates in various quantum networks.

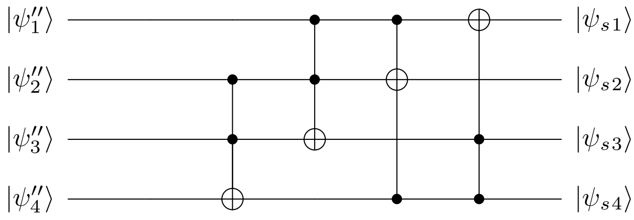

Fig. 6 The first proposed Interaction Layer which comes before the second Convolutional Layer

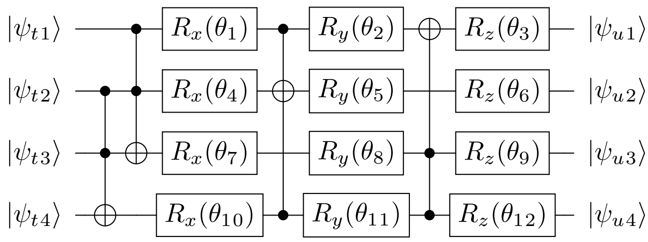

Fig. 7 The second novel Interaction Layer which comes after the second Convolutional Layer

This paper explores the comparative advantage of three-qubit *Toffoli* gates with the introduction of Interaction layers in the conventional QCNN structure. The difference in performance upon this addition is expected to indicate the extent of effectiveness which may result from the successful implementation of such gates in quantum hardware.

In the first Interaction Layer, the four *Toffoli* gates entangle the four qubits in a circuit-block interaction configuration, as illustrated in Fig.6, which means that interdependency has been established between these quantum states, so the measured value of one state will depend on the others. This particular configuration is chosen over a nearest-neighbor or all-to-all configuration, as it tends to display favorable expressibility and less qubit connectivity requirements. The placement of these *Toffoli* gates after the previous layers enable the network to span all the basis states more strongly (i.e., with a higher probability for the basis states that previously had near-zero probabilities) than without them.

After the second Convolutional Layer, the qubits are passed through Interaction Layer 2. This Interaction Layer differs from the first in the inclusion of R_x, R_y, and R_z rotation gates with trainable parameters between its *Toffoli* gates as shown in Fig.7. The parameterized gates in between the *Toffoli* gates increase the degree of freedom of the quantum states, increasing the flexibility of the learning process and therefore has the potential to learn more nuanced features of the training data.

The last Interaction Layer in Fig.8 acts as a classifier utilizing *CNOT* gates to entangle the remaining qubits with the ancilla qubits, which are used to store the

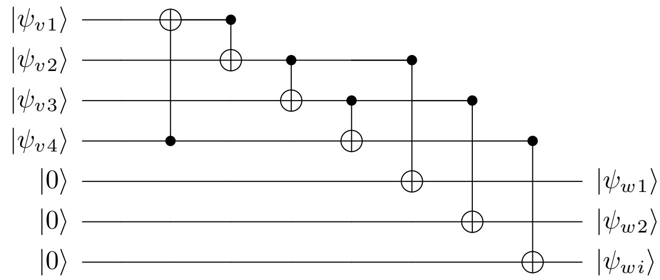

Fig. 8 The final Interaction Layer acts as the classifier for the classification of three classes. Note that the three qubits from the bottom are ancilla qubits for the three classes and are initialized to $|0\rangle$

entangled states. It must be noted that the number of ancilla qubits is equal to the number of classes that are to be classified using the network and have been set to $|0\rangle$ initially. The ancilla qubits interact with the remaining qubits of the network through the *CNOT* gates as shown in Fig.8 and are passed through the three rotational gates at the terminal of the quantum network. Finally, the measurement operation is performed on the ancilla qubits which causes the wavefunction to collapse into deterministic values. The expectation values of the ancilla qubits are measured and fed into the *Softmax* function.

The introduction of these Interaction Layers in the middle of the conventional QCNN along with an ancilla-based classifier (Interaction Layer 3) is expected to provide promising results and is inspected in detail later in the next section 4. The proposed Interaction Layers consisting of a combination of *Toffoli* and *CNOT* gates along with trainable parameters in between the Convolutional Layers, and the use of ancilla qubit-*CNOT* classifier in QCNNs is a first to the best of our knowledge.

3.3.4 Cost Function and Softmax

Following the measurement of the ancilla qubits, the classical values are sent to the *softmax* function to calculate the probability vectors for each class. Then the losses are calculated using the classical categorical cross-entropy loss function which can be expressed as:

$$loss = \sum_{i=1}^{output\ size} y_i.log(\bar{y}_i) \tag{3}$$

where y_i is the true-label and $\bar{y}_i$ is the predicted probability of the corresponding class. The parameters of the quantum gates are optimized through gradient descent using classical computational techniques, after which the parameters are updated accordingly through back-propagation.

4 Simulation and Results

4.1 Dataset

The widely utilized standard datasets, namely *MNIST* Deng (2012) and *Fashion MNIST* Xiao et al (2017) are employed to benchmark the proposed QCNN model. Binary classification involving classes (0, 1) and three-class classification involving classes (0, 1, 2) are performed. In the *MNIST dataset*, the number of training (test) images, for class 0 is 5,923 (980); for class 1 it is 6,742 (1,135); for class 2 it is 5,958 (1,032). The *Fashion MNIST dataset* consists of 6,000 training and 1,000 test images per class. The original size of the images from either dataset is 28×28, and a reduction in dimension is accomplished through a classical autoencoder or simple resize to the desired shape. A third dataset known as the *Iris dataset* De Marsico et al (2015) is used solely for the purpose of multiclass classification. It consists of feature data of three classes of iris species with 50 samples per class. The features include 4 attributes per sample namely sepal length, sepal width, petal length & petal width. The dataset is such that one flower class is linearly separable from others but the other two classes are not linearly separable from each other.

4.2 Simulation

The simulation of the proposed QCNN model is conducted using Pennylane (Bergholm et al (2018)). The variational circuit is trained through the use of the Nesterov Moment Optimization algorithm (Nesterov (1983)). A loop is executed through the training process where a batch of randomly selected images is fed into the network in each iteration, reducing run time and preventing the gradient from becoming trapped in a

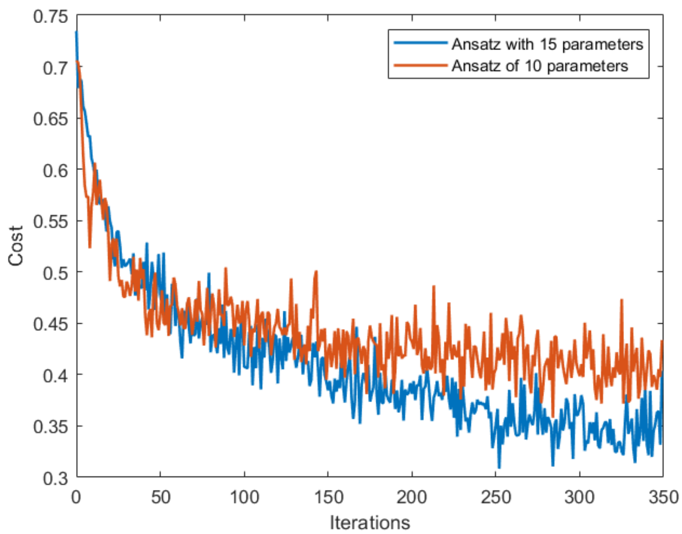

Fig. 9 Cost vs. Iteration plot using *Angle Encoding*

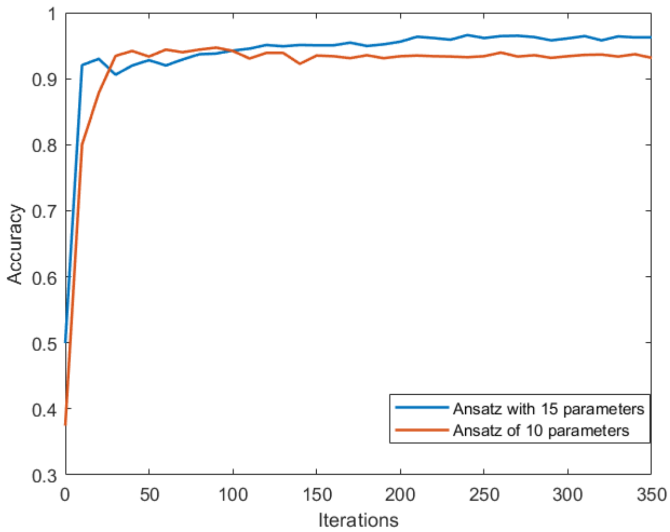

Fig. 10 Accuracy vs. iteration plot using *Angle Encoding*

local minimum. The optimization of the learning process is further facilitated through the use of an adaptive learning rate, where the learning rate is decreased as the rate of change of the output of the cost function is decreased.

4.3 Performance evaluation

4.3.1 Binary classification

For the binary classification problem, classes 0 and 1 are chosen for both datasets in order to be able to compare with previous works. A total of eight input qubits along with two ancilla qubits for the two classes are used in the proposed model. The Convolutional and the Pooling Layers are arranged as illustrated in Fig.1. During training, the batch size is kept at 50 images, which are randomly selected in each iteration. The learning rate in the Nesterov Optimizer is tuned to be 0.05 at the beginning of the learning process and in the later stages, it is reduced in accordance with the decrease in the cost and the improvement of test accuracy. The trainable parameters are initialized randomly using the normal distribution and the average classification accuracy is calculated over five random initializations.

Table 1 Results of the proposed binary classification model

Dataset	Ansatz	Encoding Method	Trainable parameters	Accuracy (%)
Fashion MNIST	1	Angle	50	96.60
		Amplitude		92.50
	2	Angle	40	94.70
		Amplitude		91.30
MNIST	1	Angle	50	98.22
		Amplitude		99.24
	2	Angle	40	93.52
		Amplitude		94.28

Table 2 Table showing a comparison of the results of our proposed model to that of existing models for different datasets

Dataset	Model used	Accuracy (%)
Fashion MNIST	Proposed	96.60
	QCNNFCDC (Hur et al (2022))	94.30
	Proposed without E	92.00
MNIST	Proposed	99.24
	QCNNFCDC (Hur et al (2022))	98.70
Iris	Proposed	94.74
	HCQAMC (Chalumuri et al (2021))	92.10

Note: 'E' refers to Interaction Layers 1 and 2.

The effect of the following different approaches on overall performance is investigated:

1. Quantum encoding by either *Amplitude Encoding* or *Angle Encoding*.
2. The two parameterized ansatzes given in Fig.3 and 4 used to construct the Convolutional Layers in Fig.2.

The classification of the *Fashion MNIST* is benchmarked to an accuracy of 96.60% using the combination of autoencoder with *Angle Encoding* and ansatz 1 as the Convolutional Layer filters, for which the total number of trainable parameters in the quantum network is 50. In order to demonstrate the convergence of the cost function of the training stage, Fig.9 and Fig.11 are shown. It can be observed from Fig.10 that over 90% accuracy is reached within only 50 iterations for both the ansatzes which suggests that the network converges rapidly with respect to the number of iterations. The peak accuracy attained for the *MNIST* dataset is 99.24%. The number of trainable parameters, in this case, is also 50, and simple resizing with ansatz 1 is used. The accuracies for the different combinations are summarized in Table 1.

The superiority in performance of ansatz 1 over ansatz 2, due to its additional parameters is demonstrated in Fig.11 and 12, where the performances are compared with respect to the same Quantum Feature Encoding methods.

It can be observed from table 1, that the type of Quantum Encoding method used to bear the best accuracies is dependent on the dataset that is used. The peak accuracy for the *Fashion MNIST* dataset results from reducing the classical data by an autoencoder followed by *Angle Encoding* whereas, for *MNIST*, it is simple resizing followed by *Amplitude Encoding*. Comparison of the results with other existing quantum machine learning models for binary classification, such as that proposed in Hur et al (2022),

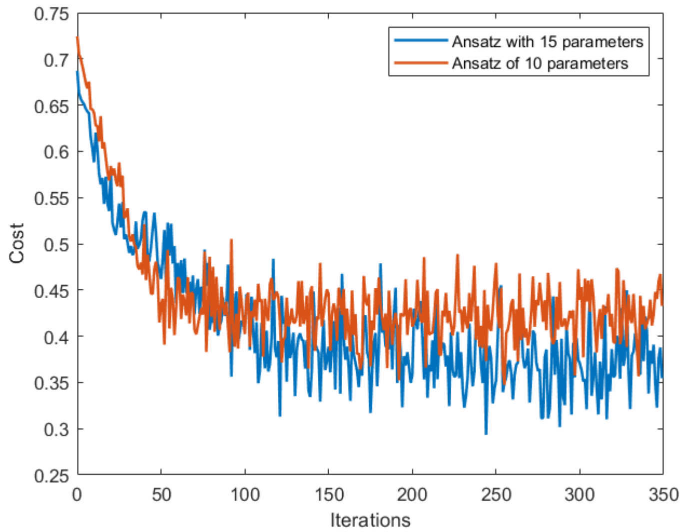

Fig. 11 Cost vs. iteration plot using *Amplitude Encoding*

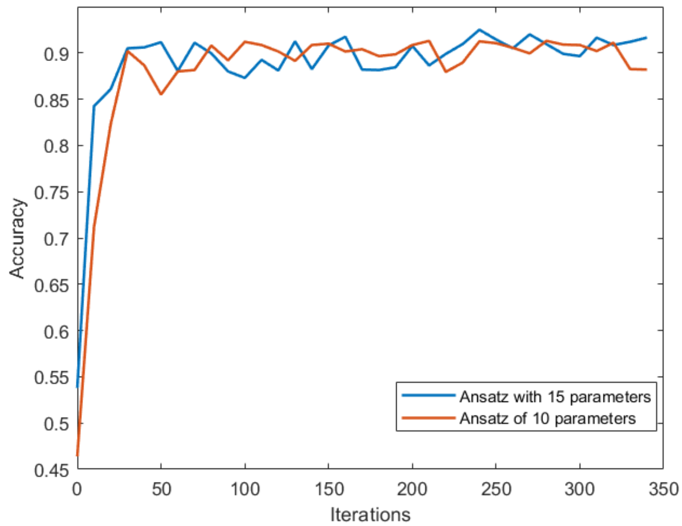

Fig. 12 Accuracy vs. iteration plot using *Amplitude Encoding*

shows that our model, surpasses their accuracy, as shown in the first two rows of table 2. The accuracy for the binary classification of classes 0 and 1 for *Fashion MNIST* is 2.6% and for *MNIST* it is 0.5% more than that found in Hur et al (2022). This increase in accuracy can be attributed to the incorporation of the Interaction Layers and the use of the ancilla-based final classifier (Interaction Layer 3). Hur et al (2022)

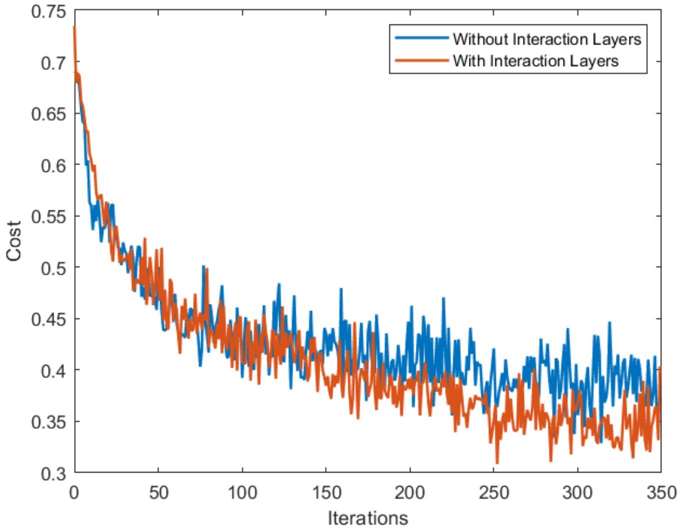

Fig. 13 Cost vs Iteration plots comparison between models with and without Interaction Layers 1 and 2

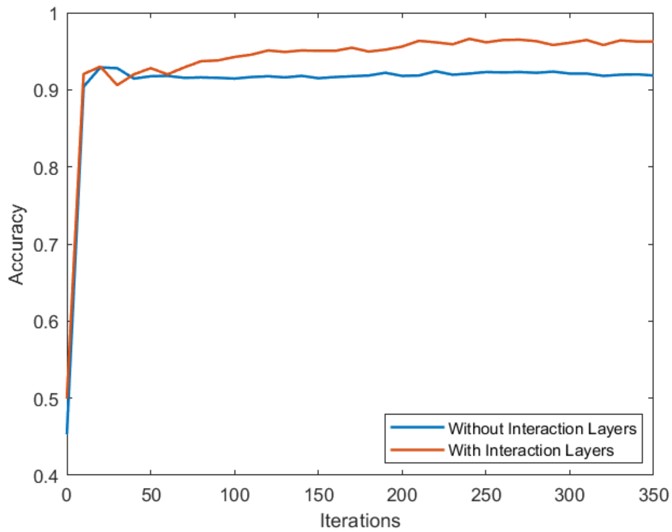

Fig. 14 Accuracy vs Iteration plots comparison between models with and without Interaction Layers 1 and 2

have further shown that their quantum network outperforms classical counterparts using a similar number of trainable parameters for the binary classification problem. It can therefore be concluded that the results of the study in our paper exhibit a clear

Table 3 Results of the Multiclass Classification problems.

Dataset	Encoding Method	Trainable Parameters	Accuracy
Fashion MNIST	Angle	53	91.76%
MNIST	Amplitude	53	85.11%
Iris	Angle	31	94.74%

Note: In all of the cases, ansatz 1 is used to construct the Convolutional Layers.

superiority in performance compared to the classical networks with a similar number of trainable parameters.

The peak accuracy for ansatz 1 used to construct the Convolutional Layers is 96.60% compared to 94.70% when ansatz 2 is used. This increase can be related to the number of trainable parameters available for each ansatz which is more in the case of ansatz 1. Additionally, the effect of the proposed Interaction Layers 1 and 2 on the performance of the network is demonstrated by comparing the performances with and without their presence. As evident in Fig.13 and Fig.14, it can be concluded that these layers help reduce cost and increase accuracy by creating further dependencies between quantum states and making them more capable of spanning the *Hilbert Space* adding only 12 more trainable parameters. In both cases, the data reduction and quantum encoding technique used is autoencoder and *Angle Encoding* respectively on the *Fashion MNIST* dataset with ansatz 1 in Fig.3 used as the convolutional filter.

4.3.2 Multiclass classification

Multiclass classification is performed on the *MNIST* and *Fashion MNIST* datasets with the network slightly modified to include two Convolutional Layers cascaded together in each convolutional stage. It must be noted that these cascaded Convolutional Layers share the same weight and therefore the number of trainable parameters in the circuit does not significantly increase. The number and placement of the Interaction Layers remain unchanged from the network for Binary classification. Classes 0, 1, and 2 are selected for both datasets and the batch size is kept at 100 with the learning rate set at 0.05 in the beginning and adapted to 0.01 after 50 iterations. The peak classification accuracy obtained is 91.76% for the *Fashion MNIST* dataset and 85.11% for the *MNIST* dataset using ansatz 1. The total number of trainable parameters in the network is only 53.

It is noticed that the combination of ansatz 1 and *Angle Encoding* as the Quantum Encoding Method provides the highest accuracy for the *Fashion MNIST* dataset but the combination of ansatz 1 with the *Amplitude Encoding* dataset performs better for *MNIST* dataset.

To demonstrate the flexibility of the proposed circuit, performance on the *Iris* dataset is also tested. In order to accommodate data of smaller dimensions, a cut-down version of the proposed circuit, with only 4 qubits was sufficient. The test accuracy with a batch size of 50 and a learning rate of 0.005 is found to be 94.74%. A three-class classifier with a variational quantum circuit has been proposed in Chalumuri et al (2021), where classification was performed on classical one-dimensional feature data. The accuracy of 94.74% supersedes the accuracy of the network proposed in Chalumuri et al (2021) (92.10%) as shown in the third row of table 2. It must also be noted that the network used for benchmarking the *Iris* dataset has only 31 parameters and is much shallower than the one in Chalumuri et al (2021). It is therefore understood that the network is not only limited to image classification but performs equally well in one-dimensional feature data. The results of Multiclass classification problems are summarized in 3.

The high accuracy achieved with only 53 parameters (for *Fashion MNIST* and *MNIST*) and 47 parameters (for *Iris*) can be directly attributed to the incorporation of the Interaction Layers. When expanded qubit interactions are used, it is possible to achieve such accuracy while using a few parameters. This implies that these interactions will speed up the training of QCNNs while producing better outcomes with shallower circuits, enabling the development of networks that are more resistant to the barren plateau issues that result from greater depth.

5 Conclusion

In this work, a shallow entangled QCNN with a minimal number of trainable parameters is proposed, which provides very satisfactory performance in binary and multiclass classification problems. The incorporation of weighted Interaction Layers, consisting of trainable parameters and utilizing three-qubit interactions between the quantum Convolutional Layers, has played a major role in enhancing the performance of the network. In doing so, it also studies the effect of the addition of such parameterized 3-qubit layers in a QCNN structure, which is a first of its kind. This result indicates the significance of increased qubit interaction on the substantial increase in the ability of a quantum network to learn more complex information from the training data whilst only using a few parameters.

This approach constitutes a novel way towards the development of a generalized parameterized QNN that performs equally well for binary and multiclass classification on both image data and one-dimensional feature data. It further explores the possibilities of performance enhancement of quantum networks upon the use of increased qubit interaction which is expected to be a reality in the not-so-distant future.

The simulation results indicate a quantum advantage of such networks, showing a clear superiority in performance compared to their classical counterparts using a similar number of parameters. Further research could be conducted to gain a more comprehensive understanding of the quantum advantage of these networks. An extensive investigation of the underlying causes of data dependencies on the proposed feature encoding methods can be done. Other future milestones may also include an extension of the work for big data analysis and the solution of more complex problems utilizing more resources and power on real quantum computers.

Statements and Declaration

The authors declare no competing interest in any other work or publication.

Data Availability

The simulation code used in this paper can be found at the following link: Simulation and Code.

The datasets used in this paper are publicly available and can be found in the works of Deng (2012), Xiao et al (2017), and De Marsico et al (2015).

References

Araujo IF, Park DK, Petruccione F, et al (2021) A divide-and-conquer algorithm for quantum state preparation. Scientific Reports 11(1):1–12. https://doi.org/https://doi.org/10.1038/s41598-021-85474-1

Arute F, Arya K, Babbush R, et al (2019) Quantum supremacy using a programmable superconducting processor. Nature 574(7779):505–510. https://doi.org/https://doi.org/10.1038/s41586-019-1666-5

Bergholm V, Izaac J, Schuld M, et al (2018) Pennylane: Automatic differentiation of hybrid quantum-classical computations. arXiv preprint arXiv:181104968 https://doi.org/https://doi.org/10.48550/arXiv.1811.04968

Chalumuri A, Kune R, Manoj B (2021) A hybrid classical-quantum approach for multi-class classification. Quantum Information Processing 20(3):1–19. https://doi.org/https://doi.org/10.1007/s11128-021-03029-9

Cong I, Choi S, Lukin MD (2019) Quantum convolutional neural networks. Nature Physics 15(12):1273–1278. https://doi.org/https://doi.org/10.1038/s41567-019-0648-8

De Marsico M, Nappi M, Riccio D, et al (2015) Mobile iris challenge evaluation (miche)-i, biometric iris dataset and protocols. Pattern Recognition Letters 57:17–23

Deng L (2012) The mnist database of handwritten digit images for machine learning research [best of the web]. IEEE signal processing magazine 29(6):141–142

Enos GR, Reagor MJ, Henderson MP, et al (2021) Synthetic weather radar using hybrid quantum-classical machine learning. arXiv preprint arXiv:211115605 https://doi.org/https://doi.org/10.48550/arXiv.2111.15605

Hur T, Kim L, Park DK (2022) Quantum convolutional neural network for classical data classification. Quantum Machine Intelligence 4(1):1–18. https://doi.org/https://doi.org/10.1007/s42484-021-00061-x

Jain S, Ziauddin J, Leonchyk P, et al (2020) Quantum and classical machine learning for the classification of non-small-cell lung cancer patients. Springer Nature Applied Sciences 2(6):1–10. https://doi.org/https://doi.org/10.1007/s42452-020-2847-4

Liu J, Lim KH, Wood KL, et al (2021) Hybrid quantum-classical convolutional neural networks. Science China Physics, Mechanics and Astronomy 64(9):1–8. https://doi.org/https://doi.org/10.1007/s11433-021-1734-3

Madzik MT, Asaad S, Youssry A, et al (2022) Precision tomography of a three-qubit donor quantum processor in silicon. Nature 601(7893):348–353. https://doi.org/https://doi.org/10.1038/s41586-021-04292-7

Mengoni R, Di Pierro A (2019) Kernel methods in quantum machine learning. Quantum Machine Intelligence 1(3):65–71. https://doi.org/https://doi.org/10.1007/s42484-019-00007-4

Nesterov YE (1983) A method for solving the convex programming problem with convergence rate. In: Dokl. Akad. Nauk SSSR,, pp 543–547

Pesah A, Cerezo M, Wang S, et al (2021) Absence of barren plateaus in quantum convolutional neural networks. Physical Review X 11(4):041011. https://doi.org/https://doi.org/10.1103/PhysRevX.11.041011

Rebentrost P, Mohseni M, Lloyd S (2014) Quantum support vector machine for big data classification. Physical Review Letters 113(13):130503. https://doi.org/10.1103/PhysRevLett.113.130503

Schuld M (2021) Supervised quantum machine learning models are kernel methods. arXiv preprint arXiv:210111020 https://doi.org/https://doi.org/10.48550/arXiv.2101.11020

Schuld M, Petruccione F (2018) Supervised learning with quantum computers, vol 17. Springer, https://doi.org/https://doi.org/10.1007/978-3-319-96424-9

Sim S, Johnson PD, Aspuru-Guzik A (2019) Expressibility and entangling capability of parameterized quantum circuits for hybrid quantum-classical algorithms. Advanced Quantum Technologies 2(12):1900070

Von Lilienfeld OA (2018) Quantum machine learning in chemical compound space. Angewandte Chemie International Edition 57(16):4164–4169. https://doi.org/https://doi.org/10.1002/anie.201709686

Xiao H, Rasul K, Vollgraf R (2017) Fashion-mnist: a novel image dataset for benchmarking machine learning algorithms. arXiv preprint arXiv:170807747

Yanofsky N, Mannucci M (2008) Quantum Computing for Computer Scientists. Cambridge University Press