Quantformers: Bits to Qubits*

Submission Id 274

ABSTRACT

This research presents a novel method utilizing quantum circuits for image classification, a task traditionally handled by classical patch-wise vision transformers. These classical methods, using binary bits, have been highly effective on standard computers. However, the potential of quantum computing in AI applications is rapidly growing. Our study explores the early application of quantum circuit algorithms, termed "Quantformers," in image classification. Quantum circuits have three main components: universal gates, measurement trajectory, and qubits entanglement. We focus on all, already proven effective in quantum circuits challenges. Our work involves designing and analyzing Quantum Transformers. These are advanced versions of classical transformer neural networks, which are known for their effectiveness in image analysis. Quantum Transformers, built with shallow quantum circuits, offer distinct classification models compared to classical methods. We conducted thorough simulations on standard medical image datasets. The results demonstrate that Quantum Transformers are not only competitive but, in some cases, surpass the best classical transformers and other benchmarks. An advantage of our quantum architectures is their reduced parameter count compared to classical methods with millions of parameters, leading to lower computational demands. Our research introduces a quantum circuit-based algorithm as a foundational approach in the evolving field of quantum-enhanced artificial intelligence.

CCS CONCEPTS

• **Theory of computation → Quantum computation theory**; • **Computing methodologies → Computer vision**; *Image representation*; *Object recognition.*

KEYWORDS

Quantum computing, Quantum gates, computer vision, Image classification.

ACM Reference Format:

Submission Id 274. 2025. Quantformers: Bits to Qubits. In *Proceedings of 16th Indian Conference on Computer Vision, Graphics and Image Processing (ICVGIP'25).* ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

Quantum learning theory, emerging as a subfield within quantum physics, introduces groundbreaking techniques that have the potential to revolutionize traditional machine learning methodologies [8, 47]. Leveraging quantum techniques for linear algebra tasks in both traditional learning algorithms and deep learning [32] has proven to be an effective strategy, as highlighted in [25]. These methods encompass various applications, including NLP [30], computer vision [44, 46], dynamical systems [3, 23], climate change [42, 49], and energy [48].

This study focuses on transformers, a modern neural network architecture introduced for various applications, including computer vision tasks [24]. Transformers have exhibited exceptional performance across a wide range of tasks and datasets [36]. At its core, transformers are neural networks that employ an attention mechanism to comprehensively consider the entire context while processing input data element by element. In the context of visual analysis, images are divided into smaller patches [18]. Instead of using fixed-size kernels for patch-wise operations, a transformer model is employed. This transformer model learns attention coefficients for each patch, determining the extent to which each patch interacts with other parts of the image. In both visual recognition and text interpretation, understanding the context of each element plays a crucial role [26].

In recent times, several quantum-based methods have emerged in various subfields of machine learning. For instance, classical transformer architectures with attention mechanisms have found applications in quantum tomography [43]. Additionally, a recent investigation by introduced a quantum-enhanced transformer for sentiment analysis [16]. Moreover, [35] utilized a self-attention mechanism for text classification. In all these cases, standard variational quantum circuits [11, 14] were utilized for the neural networks, with attention coefficients computed using classical methods. In another development proposed in [20], there is a suggestion to employ a quantum attention mechanism in reinforcement learning. Additionally, it's worth noting the proposal put forth by [13], which discusses variational circuits sharing similarities with convolutional neural networks, especially in the context of general-purpose image classification.

Our methodology diverges significantly from those previously discussed in three main areas. First, we employ a Quantum Circuit Vision Transformer, or Quantformer, for tasks such as classification. This contrasts with traditional image classification tasks based on quantum theory, which rely on adiabatic quantum computing for parameter updates, as detailed in [2]. The Quantformer is distinctive in its use of complex algebraic operations, typically challenging to perform using classical theory. As outlined in Section 3, we introduce two quantum circuits: QLinear, which converts image patches into quantum states, and the core circuit, which executes quantum operations. Secondly, our Quantformer is better suited for Noisy Intermediate-Scale Quantum (NISQ) environments due to its innovative handling of bit-flip probability. Unlike methods that calculate the flipping-phase directly, our approach involves pre and post-measurement state trajectories, as elaborated in Section 4. Specifically, it allows for direct control of each circuit state by monitoring and analyzing measurement trajectories. This enables the implementation of feedback control (akin to back-propagation

*Produces the permission block, and copyright information

ICVGIP'25, December 2025, IIT Mandi, India

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM.
https://doi.org/10.1145/nnnnnnn.nnnnnnn

in classical systems) within quantum systems [37]. By adjusting parameters in response to measurement outcomes, we can guide the system toward a desired state or result, paralleling techniques in classical control theory. Finally, using the Neumann entropy method explained in Section 6, we've concentrated on and successfully created a strong connection between qubits. This is crucial because it enhances our quantum circuits, making them more effective and efficient than standard ones. Thus, the overall workflow of the proposed Quantformer method is described in the Figure 1.

2 FOUNDATIONS OF QUANTUM COMPUTING

Quantum Bit or Qubit is the foundational element of quantum computing. This differs markedly from the classic bit by having the capacity to exist in a state that represents both $0 = [1\ 0]^T$ and $1 = [0\ 1]^T$ simultaneously, expressed mathematically as a state vector $\psi = \alpha 0 + \beta 1$, where α and β are complex numbers that comply with the normalization condition $|\alpha|^2 + |\beta|^2 = 1$. This property of superposition enables a system with n qubits to simultaneously embody all 2^n possible states, a stark contrast to a traditional system of n bits that can only represent a single state at any given moment [21].

Quantum circuit process information using these qubits, quantum computers deploy quantum circuits, which consist of a series of quantum gates. These gates are essentially unitary transformations that evolve the state of qubits from one quantum state to another, denoted as $\psi \rightarrow U\psi$ with U being a unitary operator. The outcomes from such quantum circuits are determined by measuring the qubits, which inherently collapses the qubit's superposed state to a definite state of either 0 or 1, influenced by the probability amplitude of α and β [38].

Parametric quantum circuits utilize adjustable parameters, often symbolized as angles in rotational gates, to shape a quantum state $\psi(x, \theta)$. The process begins with constructing a circuit specific to the desired task and then optimizing the parameters θ via quantum-classical techniques. Effective training of Parametric quantum circuits is crucial but can be impeded by issues like "barren plateaus" in the optimization landscape. Parameters are typically optimized using gradient-based methods such as backpropagation, adjusting them according to the gradient of the loss function $\mathcal{L}$ with respect to each parameter θ_i and a chosen learning rate α [6].

Quantum gates are the building blocks of quantum circuits. They perform operations on qubits, changing their states in a manner that can be used to perform calculations [1, 17]. Unlike classical logic gates, which can only be in one state at a time (0 or 1), quantum gates operate on the principles of superposition and entanglement, and their action is reversible. Below are some of the fundamental quantum gates:

- **Hadamard (H) gate**: The Hadamard gate acts on a single qubit. It creates a superposition state from a definite input state. When applied to a qubit in the state 0 or 1, it results in an equal superposition of states, transforming the basis states as follows:

$$H0 = \frac{1}{\sqrt{2}}(0 + 1), \quad H1 = \frac{1}{\sqrt{2}}(0 - 1). \tag{1}$$

The matrix representation of the Hadamard gate is:

$$H = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}. \tag{2}$$

- **Rotation Z (Rz) gate**: This gate rotates a qubit around the Z-axis of the Bloch sphere by an angle ϕ. The action of an Rz gate on a qubit is described by the matrix:

$$Rz(\theta_i) = \begin{bmatrix} e^{-i\theta_i/2} & 0 \\ 0 & e^{i\theta_i/2} \end{bmatrix} \tag{3}$$

where ϕ is the rotation angle.

- **Controlled Rotation Z (CRz) gate**: The CRz gate is a two-qubit gate that performs a rotation around the Z-axis on the second qubit (target qubit) only when the first qubit (control qubit) is in state 1. The matrix for CRz gate is:

$$CRz(\theta_i) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & e^{-i\theta_i/2} & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & e^{i\theta_i/2} \end{bmatrix} \tag{4}$$

- **Controlled NOT (CNOT) gate**: Also known as the controlled-X gate, it flips the state of a target qubit if the control qubit is in the state 1. It is fundamental for creating entanglement between qubits. The matrix representation of a CNOT gate is:

$$CNOT = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix} \tag{5}$$

3 QUANTFORMER

3.1 Pre-Insertion of Quantformers

Data Prepossessing The dataset $\{\mathcal{X}_i, y_i\}^M \in \mathbb{R}^{H \times W \times C}$ consists of M samples. Each sample represents an image structured as a three-dimensional matrix with dimensions (H, W, C), where $C = 1$ signifies that the images are in grayscale. Hence, for analytical processing purposes, we obtain smaller segments known as "patches" from every image, with each patch having dimensions (p, p). The total number of patches, denoted as t, is determined for each individual image during the patch extraction process. Once the patches have been extracted, the transformed dataset representing the i-th image $\mathcal{X}$ is described by a four-dimensional matrix with dimensions (M, t, p, p).

3.1.1 QLinear. The QLinear transformation aims to utilize quantum parallelism and entanglement to perform operations similar to classical linear transformations. A Qlinear block is typically an arrangement of quantum gates that apply successive operations to the qubits in a linear, sequential manner. This contrasts with classical computation [21, 40], where processing is often non-linear and may involve complex branching and multiple paths through a circuit. The QLinear block, with its straightforward sequence of operations, is critical because it maintains the delicate quantum states necessary for quantum computation. These operations include quantum gates that generate superposition, enable entanglement, and carry out parameterized rotations—each essential for quantum algorithms to exploit quantum mechanical properties that are not present in classical systems [41]. By manipulating qubits

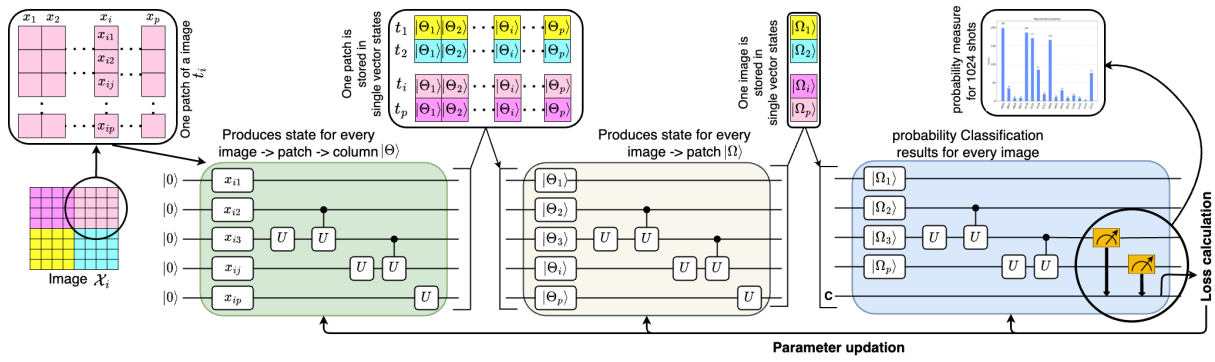


Figure 1: The Quantformer architecture pipeline is divided into three main stages: QLinear, Core, and Classification. In the first stage, QLinear, each column of a single image patch is taken as input, with the focus on extracting and collecting the states from every column of the patch. Next, the Core block plays a crucial role. It takes these patch column states and combines all the information from each patch into a single, cohesive state. Finally, in the Classification stage, these consolidated states are analyzed and categorized. This architecture effectively streamlines the journey of an image from initial feature detection to reducing computational complexity, ensuring efficient and precise image processing.

in a controlled, linear sequence, these blocks can prepare states, perform computations, and eventually interpret outcomes in ways that harness the full power of quantum mechanics [22].

Flow of QLinear The algorithm iteratively applies the quantum circuit across the entirety of the tensor, effectively treating each column of every patch as an independent input vector for the quantum processing unit. Each Patch t_i of training sample is utilizing a p-bit(where p is patch size) precision configuration i.e, one patch at a time. A single patch, considered column-wise, can be represented as a set $t_i = \{x_1, x_2, \ldots, x_p\}$. Hence, a quantum circuit, initialized in the state ψ and comprising P qubits, all set to 0, can be represented in the following manner:

$$\psi = 0^{\otimes P}. \tag{6}$$

Therefore, we encode a each column x_k of t_i into the quantum circuit using custom gates p. Each gate $\mathbb{P}(x_{kj})$ encodes a value x_{kj} corresponding to the j-th qubit. This process transforms the initial state as follows:

$$\Psi_{\mathrm{k}} = \bigotimes_{j=1}^{p} \mathbb{P}(x_{kj})0_i j. \tag{7}$$

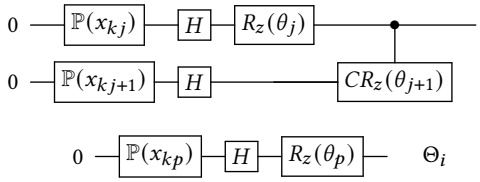
Subsequently, we apply the Hadamard gate(H) to each qubit, resulting in a superposition as follows:

$$\Psi_{\mathrm{k}} = H^{\otimes p}\Psi_{\mathrm{k}} \tag{8}$$

Finally, we introduce a sequence of rotation and controlled rotation gates to the qubits. For each qubit q_j, a rotation $(R_z(\theta_j))$ is applied, and a controlled rotation $(CR_z(\theta_j))$ is applied between qubits q_j and q_{j+1}. This process leads to the formation of an entangled state:

$$\Theta_i = \bigotimes_{j=1}^{p-1} \left(R_z(\theta_j)\Psi_j\right) \,\&\, \bigotimes_{j=1}^{p-1} \left(CR_z(\Psi_j)\Psi_j\Psi_{j+1}\right) \tag{9}$$

To enhance comprehension, consider the following circuit as an illustrative demonstration of the entanglement between qubits j and $j+1$, enabling the seamless transmission of information from the upper to the lower part of the circuit.



This is the process by which Θ_i is gathered for each column of patch t_i. Thus, we have $t_i = \Theta_1, \Theta_2, \ldots, \Theta_p$.

3.2 Core circuit

This core circuit helps the patches to learn internally and learn information from external patches. So, this circuit has two sub circuits (i) Inward patch learning (ii) Patch Entanglement.

3.2.1 Inward patch learning. Image patches are small blocks of an image that contain specific features such as edges, textures, and color intensities [18]. In classical computing, handling this data requires substantial memory and computational resources, especially when dealing with high-resolution images [15] or performing complex operations like convolutional processing [34] for machine learning applications. However, a quantum circuit that can effectively learn and compress the data from these patches into a single qubit could vastly reduce the computational overhead [38]. This is due to the quantum property of superposition, which allows a qubit to hold a combination of states simultaneously, providing a multidimensional space to represent complex information compactly [45].

In Section 3.1.1, we transformed our image into (t, p, Θ) i.e, encoded (p) dimension information into p states. Now, if a state can store important information of image patch in a single state, then the computational power can be reduced. This sub-circuit takes p states of a image patch and encoded all important information regarding this patch into single state Ω.

Let us consider a single patch, denoted as t_i, which carries a quantum representation of our parameter set (p, Θ). Subsequently, this patch undergoes a transformation reminiscent of passing a point of light through a prism, unveiling its inherent spectrum [5]. This transformation is achieved through the application of a Hadamard gate(H), which shifts our perception of the state from a singular certainty to a harmonious amalgamation of potential outcomes [29]. Hence, we can express this transformation in the following manner:

$$|\Theta' i\rangle = (H \otimes I^{\otimes(n-1)})|\Theta_i\rangle \tag{10}$$

$$|\Theta'' i\rangle = \left(\bigotimes j = 1^{n-1} CRZ(\theta(P+i))_j\right)|\Theta' i\rangle \tag{11}$$

$$|\Theta''' i\rangle = \left(\bigotimes j = n^2 CNOT\, j, j-1\right)|\Theta_i''\rangle \tag{12}$$

$$\Omega = |RZ(\Phi)|\Theta_i\rangle = RZ(\Phi)^{\otimes n}|\Theta_i'''\rangle \tag{13}$$

The final segment of the quantum circuit invokes a reconvening of initial operations in a mirrored fashion. The Hadamard(H) and controlled rotation gates($CRZ(.)$) are applied in a symmetrically inversed sequence. This set of operations can be seen as a form of quantum inversion, wherein the evolved states are sought to be reverted to a near-original form, under an altered parametric evolution [28]. The Inward Patch Learning phase is defined to interconnect distinct quantum patches, utilizing entanglement, phase rotation, and controlled operations to create a larger, correlated quantum system. By manipulating parameters within the quantum circuit, various quantum states can be explored, potentially unveiling new quantum behaviors and correlations within the given image.

3.2.2 Patch Entanglement. Quantum circuits use special quantum properties like superposition, entanglement, and interference to process data more efficiently than regular computers. In image processing, they can check many parts of an image at once, which is faster. Entanglement has the potential to capture the intrinsic relationships and patterns required for image recognition tasks by correlating features across these regions. The utilisation of quantum interference has the potential to significantly enhance the learning process by increasing the likelihood of accurate results and decreasing the likelihood of erroneous ones. In addition, the utilisation of high-dimensional Hilbert spaces for quantum information processing may facilitate a more accurate and condensed depiction of the extensive feature space occupied by image regions. This, in turn, may contribute to the development of more robust learning algorithms. Thus, we have obtained a set of quantum states $\{|\Omega_i\rangle\}_{i=1}^{t}$, where each state $|\Omega_i\rangle$ resides in the Hilbert space ($\mathcal{H}$). To be more precise, we can define the process of entanglement among patches and the classification as follows:

$$|\Omega_i'\rangle = (U_{backward} \cdot U_{forward})|\Omega_i\rangle \tag{14}$$

where, $U_{forward}$ and $U_{backward}$ are tensor products of gate operations entangling each patch with its neighbors. Additionally, the unitary transformation operation for optimizing the state is represented as:

$$|\Omega_i''\rangle = U(\Theta)|\Omega_i'\rangle \tag{15}$$

where $U(\Theta)$ is a parameterized unitary matrix acting on the quantum state to enhance distinguishable for classification. Hence, the observable used for classification is defined as follows;

$$O = \sum_{k=1}^{\Gamma} \lambda_k |o_k\rangle\langle o_k| \tag{16}$$

The classification is based on the measurement probabilities:

$$p_{ik} = |\langle o_k|\Omega_i''\rangle|^2 \tag{17}$$

with the assigned class is given by:

$$C(i) = \underset{k}{\text{argmax}}\, p_{ik} \tag{18}$$

The overall process uses entanglement, parametric transformations, and measurements within a quantum computational framework to categorize quantum states into distinct classes.

3.3 Post-Insertion of Quantformers

For the purpose of training, we introduce a loss function $\mathcal{L}(\theta, \phi)$ that gauges the discrepancy between the predicted probabilities and the true labels. A common choice for this function is the cross-entropy loss:

$$\mathcal{L}(\theta, \phi) = -\sum_{i=1}^{M}\sum_{k=1}^{\Gamma} y_{ik} \log(p_k) \tag{19}$$

Here, y_{ik} functions as a binary indicator, denoting whether sample i belongs to class k or not.

Optimizing parameters θ, ϕ in quantum gradient methods, like the parameter-shift rule [50], minimizes the loss function ($\mathcal{L}$). The gradient of the expectation value with respect to a parameter (θ_i, ϕ) is computed as follows:

$$\frac{\partial O_{\theta,\phi}}{\partial \theta_i, \phi} = \frac{1}{2}\left(O_{(\theta_i,\phi)+\pi/2} - O_{(\theta_i,\phi)-\pi/2}\right) \tag{20}$$

where $(O_{(\theta_i,\phi)+\pi/2})$ and $(O_{(\theta_i,\phi)-\pi/2})$ are the expectation values when the parameter (θ_i) is shifted by $\pi/2$ in the positive and negative directions, respectively.

Thus, the gradient of the loss function with respect to θ_i, ϕ is:

$$\frac{\partial \mathcal{L}(\Theta, \phi)}{\partial \theta_i, \phi} = \frac{1}{2}\left(\mathcal{L}((\theta_i, \phi) + \pi/2) - \mathcal{L}((\theta_i, \phi) - \pi/2)\right) \tag{21}$$

Using this gradient, optimization algorithms (like gradient descent [4]) can be applied to adjust the parameters (θ) to minimize the quantum loss $(\mathcal{L}(\theta))$ and bring the output of the quantum circuit closer to the desired target.

4 MEASUREMENT TRAJECTORY OF STATES FOR CONTROLLING NISQ

In the context of the circuits designed for controlling Noisy Intermediate-Scale Quantum (NISQ) devices, a measurement trajectory entails the process of measuring quantum states at various stages to steer the computation [7, 10, 39]. This practice is particularly crucial in NISQ systems due to their inherent noise and limitations in coherence time. Here's an explanation of how this measurement trajectory is implemented for the described circuits:

Measurement in QLinear Circuit In the QLinear circuit, the goal is to apply a series of quantum gates to transform the state of

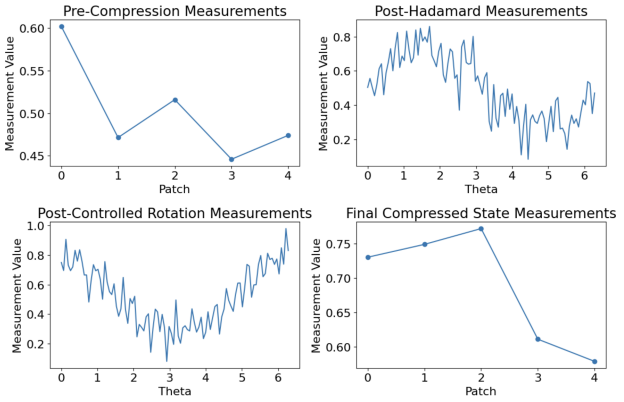


Figure 2: The figure contains four graphs of TissueMNIST, each illustrating a key phase in quantum image patch processing under NISQ conditions. These phases are Pre-Compression, Post-Hadamard, Post-Controlled Rotation, and Final Compressed State Measurements, all with added noise to mimic real-world quantum environments. The series effectively demonstrates the journey from initial quantum states to compressed outcomes amidst quantum noise

qubits to represent the data encoded from the image patches. In this context, Figure 2 provides a visual representation of the specific details of the intermediate state measurement trajectory.

Initial State Measurement: Prior to applying any gates, it is customary to measure the initial state of all qubits, typically set to 0, as a baseline to verify proper system initialization.

Post-Encoding Measurement: After applying custom gates to encode each column x_k from a training sample into the quantum circuit, we carry out a measurement to confirm the effectiveness of the encoding process.

Intermediate Gate Measurement: After applying the H and rotation gates (CRZ, RZ), we take intermediate measurements to evaluate the evolution of the quantum state and verify that the desired superposition and entanglement have been achieved.

Final State Measurement: Once the entire circuit has been executed, the final measurement of the qubits yields the output of the QLinear transformation. This pivotal measurement serves as the basis for subsequent processing steps.

Measurement in Core Circuit and Inward Patch Learning The core circuit, particularly the inward patch learning part, focuses on compressing the data from image patches. The measurement trajectory here involves:

Pre-Compression Measurement: This involves measuring the states representing the patch information before the compression process begins. This gives an understanding of the quantum representation of the patch data.

$$M_{pre}(|\Theta_i\rangle) = \langle\Theta_i|\Theta_i\rangle \tag{22}$$

Post-Hadamard Measurement: After applying the Hadamard gate, a measurement is taken to confirm the creation of a superposition of states.

$$M_H(|\Theta_i'\rangle) = \langle\Theta_i'|H|\Theta_i\rangle \tag{23}$$

Post-Controlled Rotation Measurement: Subsequent to the application of rotation and controlled rotation gates, measuring the quantum states is provide insight into the entanglement and correlations between different parts of the patch.

$$M_{CR}(|\Theta_i''\rangle) = \langle\Theta_i''|CR_zR_z|\Theta_i\rangle \tag{24}$$

Final Compressed State Measurement: The most crucial measurement is after the full compression circuit has been applied, where the state Ω represents the compressed information of the patch. This measurement provides the output for the patch learning.

$$M_\Omega(\Omega) = \langle\Omega|\Omega\rangle \tag{25}$$

5 EXPERIMENTS

5.1 MEDICAL IMAGE CLASSIFICATION

5.1.1 Datasets. For the purpose of evaluating our algorithms, we employed MedMNIST, an assortment of 12 uniformly processed, 2D medical image public datasets [51]. This compilation has been normalized for categorization tasks across 12 varied imaging modalities, each incorporating medical images with dimensions of 28 × 28 pixels. Our transformers were utilized to execute simulations across all 12 MedMNIST datasets. We used our transformers to run tests on all 12 MedMNIST datasets. We chose MedMNIST because it's not too big, making it good for simulating quantum circuits and hardware experiments. It's also important for medical image categorization. First, we tested our models on the 2D MedMNIST datasets. The results showed that our new state control quantum attention method performs just as well as, and sometimes even better than, regular classical models in terms of accuracy.

5.1.2 Simulation settings. To evaluate the effectiveness of the Quantformer mechanisms discussed in this paper, we conducted an extensive simulation of the training processes for the different quantum circuits that compose each network. Additionally, we included two classical methods in our comparison. First, we used the Vision Transformer [19] as our initial benchmark, representing classical neural networks commonly used in various image classification tasks. We also introduced the Orthogonal Fully-Connected Neural Network (OrthoFNN) [27], which was trained using the MedMNIST dataset, as our reference point for quantum comparison. Furthermore, we assessed the performance of a recent quantum vision transformer based on orthogonal constraints [12] in comparison to our proposed Quantformer.

5.1.3 Classification Results. Our research encompassed various datasets, classification challenges, and architectural frameworks, which we thoroughly describe below. The outcomes, presented in Figures 3, as well as in Table 1, highlight the performance metrics such as the Area Under the Curve (AUC) and Accuracy (ACC). These metrics were recorded for both training and test phases. As general remarks from these results, one can see that from some datasets, both the AUC and ACC are quite close or higher than other classical algorithms, showing that the quantum bades experiments reach the performance of the classical algorithms.

Table 1 present a comparative analysis of various neural network models, including VisionTransformer [18], OrthoFNN [31],

Network	PathMNIST		ChestMNIST		DermaMNIST		OCTMNIST		PneumoniaMNIST		RetinaMNIST	
	AUC	ACC	AUC	ACC	AUC	ACC	AUC	ACC	AUC	ACC	AUC	ACC
VisionTransformer	0.957	0.755	0.718	0.948	0.895	0.727	0.879	0.608	0.957	0.902	0.736	0.548
OrthoFNN	0.939	0.643	0.701	0.947	0.883	0.719	0.819	0.516	0.950	0.864	0.731	0.548
OrthoPatchWise	0.953	0.713	0.692	0.947	0.898	0.730	0.861	0.554	0.945	0.867	0.739	0.560
OrthoTransformer	0.964	0.774	0.703	0.947	0.891	0.719	0.875	0.606	0.947	0.885	0.745	0.542
CompoundTransformer	0.957	0.735	0.698	0.947	0.901	0.734	0.867	0.545	0.947	0.885	0.740	0.565
Quantformer	0.961	0.755	0.737	0.951	0.893	0.729	0.876	0.607	0.948	0.886	0.811	0.601

Network	BreastMNIST		BloodMNIST		TissueMNIST		OrganAMNIST		OrganCMNIST		OrganSMNIST	
	AUC	ACC	AUC	ACC	AUC	ACC	AUC	ACC	AUC	ACC	AUC	ACC
VisionTransformer	0.824	0.833	0.985	0.888	0.880	0.596	0.968	0.770	0.970	0.787	0.934	0.620
OrthoFNN	0.815	0.821	0.972	0.820	0.819	0.513	0.916	0.636	0.923	0.672	0.875	0.481
OrthoPatchWise	0.830	0.827	0.984	0.866	0.845	0.549	0.973	0.786	0.976	0.805	0.941	0.640
OrthoTransformer	0.770	0.744	0.982	0.860	0.856	0.557	0.968	0.763	0.973	0.785	0.946	0.635
CompoundTransformer	0.859	0.846	0.985	0.870	0.841	0.544	0.975	0.789	0.978	0.819	0.943	0.647
Quantformer	0.821	0.792	0.985	0.871	0.916	0.608	0.966	0.762	0.974	0.798	0.954	0.651

Table 1: By conducting performance analyses on each test dataset of MedMNIST using AUC and ACC, we are able to compare our proposed quantum architecture to both state-of-the-art quantum methods [12] and the vision transformer [18]

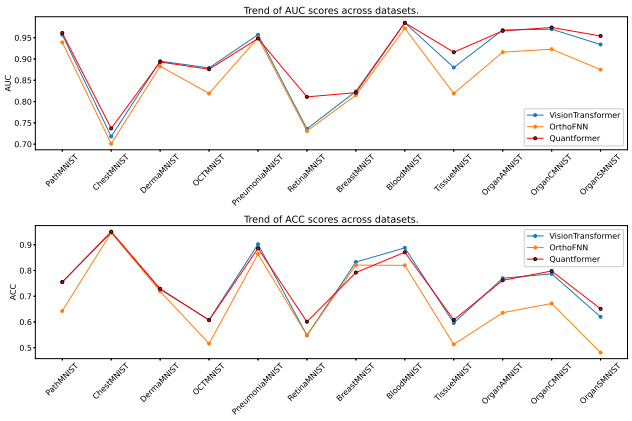


Figure 3: The top chart displays the Area Under the Curve (AUC) scores, a measure of the model's ability to distinguish between classes. The bottom chart shows the Accuracy (ACC) scores, indicating the percentage of correct predictions. The results suggest varying degrees of effectiveness of each model, with distinct peaks and troughs, indicating their relative strengths and weaknesses across different types of medical data.

OrthoPatchWise [12], OrthoTransformer [12], CompoundTransformer [12] with our Quantformer, across a range of medical imaging datasets like PathMNIST, ChestMNIST, DermaMNIST, OCTMNIST, PneumoniaMNIST, RetinaMNIST, BreastMNIST, BloodMNIST, TissueMNIST, OrganAMNIST, OrganCMNIST, and OrganSMNIST. Quantformer demonstrates a consistently strong performance across most datasets. Particularly noteworthy is its superior performance on the BloodMNIST, OrganSMNIST, and TissueMNIST datasets, where it achieves the highest AUC scores compared to the other models. This indicates a robust ability of Quantformer in handling these specific types of medical images. On the other datasets, Quantformer also performs competitively, often ranking among the top models in terms of AUC and ACC. Its performance is particularly strong when compared to traditional models like VisionTransformer and OrthoFNN, suggesting that the Quantformer's architecture may be more suited for processing complex features images.

In a comparative analysis across various medical imaging datasets, Quantformer notably excels in specific cases, demonstrating its effectiveness in medical image analysis. On BloodMNIST, Quantformer achieves an AUC score of 0.985, identical to the best-performing model CompoundTransformer, resulting in a negligible difference of 0%. However, it slightly edges out with a 0.11% higher accuracy. In OrganSMNIST, Quantformer's AUC of 0.954 surpasses the second-best OrthoTransformer (0.946) by approximately 0.85%, and its accuracy of 0.651 is marginally better than CompoundTransformer's 0.647, showing a difference of 0.62%. The most significant performance disparity is observed in TissueMNIST, where Quantformer's AUC of 0.916 substantially outperforms OrthoPatchWise's 0.845, marking an impressive difference of approximately 8.4%. These results highlight Quantformer's superior analytical capabilities in certain datasets, particularly in discerning complex patterns in medical images. **Details on computational complexity, additional results, hardware simulations (IBM), and code can be found in the supplementary materials.**

6 ENTANGLEMENT OF QUBITS

Entanglement, a non-classical correlation between quantum systems, is highly sensitive to the sequence of quantum operations applied. In a typical quantum circuit, entanglement is generated and manipulated through the application of two-qubit gates, such as the controlled rotation gates (CRZ). These gates apply a rotation to a target qubit conditioned on the state of a control qubit, intertwining their states and generating entanglement.

6.1 Entanglement calculation

The state of a quantum system is fully described by its density matrix ρ, which in the case of a pure state ψ is given by $\rho = |\Psi\rangle\langle\Psi|$. To understand the fluctuations in entanglement over time as depicted,

we must consider the measure of entanglement in the system's evolving state. One such measure is the von Neumann entropy of a reduced density matrix, defined for a bipartite system AB as $S(\rho_A) = -\text{Tr}(\rho_A \log \rho_A)$, where $\rho_A = Tr(\rho_B)$. For a system evolving under time-dependent Hamiltonians, which is typically the case in parameterized quantum circuits with time-varying parameters $\theta(t)$, the unitary operator would also be time-dependent. The evolution of the system's state vector would be governed by the Schrödinger equation:

$$\frac{d}{dt}|\Psi(t)\rangle = -iH(t)|\Psi(t)\rangle \tag{26}$$

The entanglement measures for the state at any given time t, such as von Neumann entropy or entanglement of formation [9, 33] can be calculated from the time-evolved density matrix $\rho(t)$. These measures will exhibit peaks and troughs corresponding to the variations in entanglement throughout the evolution of the quantum state as dictated by the applied quantum gates and the temporal modulation of their parameters.

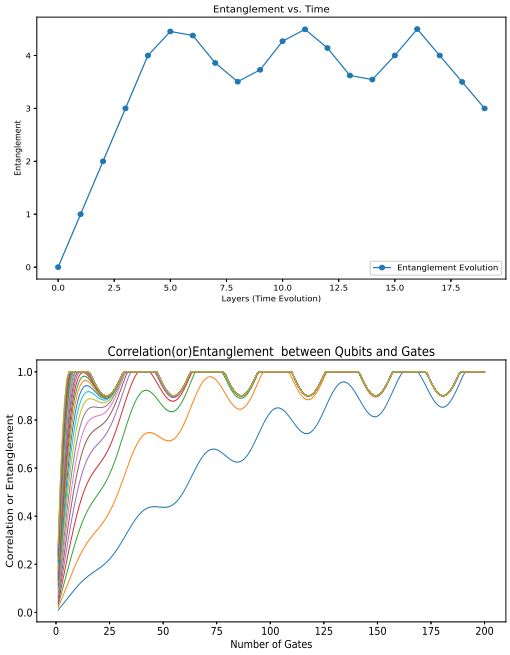


Figure 4: Using mathematical modeling, calculated how entanglement might change as the circuit progresses, providing insights into the circuit's behavior and potential efficiency for the TissueMNIST dataset. Hadamard gates initially create superpositions, and CRZ and CNOT gates then build entanglement between qubit pairs, crucial for quantum correlations. This leads to an oscillation phase, where controlled operations fluctuate entanglement levels, redistributing it in a way beneficial for quantum computation. Finally, a decrease in entanglement occurs, a necessary step in many quantum algorithms for extracting classical information by disentangling qubits.

6.2 Analysis of entanglement

The initial steep ascent in the figure 4 reflects the rapid establishment of entanglement as Hadamard gates (H) create superpositions and CRZ gates induce phase relationships between qubits. The subsequent undulations in the graph can be interpreted as the manifestation of constructive and destructive interference patterns emerging from the cumulative effect of subsequent gate applications. Each peak in the graph corresponds to a state where the quantum gates have synergistically interacted to enhance entanglement. Conversely, troughs represent states where the specific configurations of applied gates lead to a partial or complete disentanglement. This occur due to the intrinsic reversibility of quantum operations, where certain gate sequences can undo entanglement, a phenomenon that does not have a classical analog. The evolution of such a system would lead to a state vector that oscillates between entangled and separable states as a function of the angles θ used in the rotation gates. If these angles are varied over time—perhaps as a parameter of the circuit's operation—the graph would naturally exhibit the observed fluctuations. In a multi-qubit system like the one described, the total state is a complex, high-dimensional vector, formed by combining the states of individual qubits and the effects of the gates. We can track the state's changes using entanglement measures like concurrence or entanglement entropy. The graph's fluctuations reflect the gate sequence and parameters in the quantum circuit, highlighting the fine control needed for quantum computation, where entanglement is key to exceeding classical computing power.

7 CONCLUSION

We present a new quantum vision transformer, named Quantformer, which is based on three core principles: managing complex Clifford (Stabilizer) gates, adapting to NISQ environments through innovative bit-flip probability management via state trajectory, and estimating qubit entanglement using the Von-Neumann entropy measure. To assess the effectiveness of our proposed method, we tested it on twelve medical MNIST datasets. Our experiments demonstrate that our method not only matches but sometimes exceeds the performance of classical vision transformers. Additionally, to understand the qubit flow process, we conducted further analysis using NISQ state trajectory and measured the entanglement to explore the physical relationship between patch information. As quantum computing research progresses, quantum circuits are expected to not only outperform classical methods consistently but also provide greater accuracy with less computational power. This work highlights the future potential of quantum computing as a transformative technology.

REFERENCES

[1] Nabila Abdessaied, Mathias Soeken, and Rolf Drechsler. 2014. Quantum circuit optimization by Hadamard gate reduction. In *Reversible Computation: 6th International Conference, RC 2014, Kyoto, Japan, July 10-11, 2014. Proceedings 6*. Springer, 149–162.
[2] Tameem Albash and Daniel A Lidar. 2018. Adiabatic quantum computation. *Reviews of Modern Physics* 90, 1 (2018), 015002.
[3] Robert Alicki and Mark Fannes. 2001. Quantum dynamical systems. (2001).
[4] Pierre Baldi. 1995. Gradient descent learning algorithm overview: A general dynamical systems perspective. *IEEE Transactions on neural networks* 6, 1 (1995), 182–195.

[5] M Bayer, Pawel Hawrylak, K Hinzer, S Fafard, Marek Korkusinski, ZR Wasilewski, O Stern, and A Forchel. 2001. Coupling and entangling of quantum states in quantum dot molecules. *Science* 291, 5503 (2001), 451–453.

[6] Marcello Benedetti, Erika Lloyd, Stefan Sack, and Mattia Fiorentini. 2019. Parameterized quantum circuits as machine learning models. *Quantum Science and Technology* 4, 4 (2019), 043001.

[7] Kishor Bharti, Alba Cervera-Lierta, Thi Ha Kyaw, Tobias Haug, Sumner Alperin-Lea, Abhinav Anand, Matthias Degroote, Hermanni Heimonen, Jakob S Kottmann, Tim Menke, et al. 2022. Noisy intermediate-scale quantum algorithms. *Reviews of Modern Physics* 94, 1 (2022), 015004.

[8] Jacob Biamonte, Peter Wittek, Nicola Pancotti, Patrick Rebentrost, Nathan Wiebe, and Seth Lloyd. 2017. Quantum machine learning. *Nature* 549, 7671 (2017), 195–202.

[9] Paul Boes, Jens Eisert, Rodrigo Gallego, Markus P Müller, and Henrik Wilming. 2019. Von Neumann entropy from unitarity. *Physical review letters* 122, 21 (2019), 210402.

[10] Adam Callison and Nicholas Chancellor. 2022. Hybrid quantum-classical algorithms in the noisy intermediate-scale quantum era and beyond. *Physical Review A* 106, 1 (2022), 010101.

[11] Marco Cerezo, Andrew Arrasmith, Ryan Babbush, Simon C Benjamin, Suguru Endo, Keisuke Fujii, Jarrod R McClean, Kosuke Mitarai, Xiao Yuan, Lukasz Cincio, et al. 2021. Variational quantum algorithms. *Nature Reviews Physics* 3, 9 (2021), 625–644.

[12] El Amine Cherrat, Iordanis Kerenidis, Natansh Mathur, Jonas Landman, Martin Strahm, and Yun Yvonna Li. 2022. Quantum vision transformers. *arXiv preprint arXiv:2209.08167* (2022).

[13] Iris Cong, Soonwon Choi, and Mikhail D Lukin. 2019. Quantum convolutional neural networks. *Nature Physics* 15, 12 (2019), 1273–1278.

[14] Andrew J Daley, Immanuel Bloch, Christian Kokail, Stuart Flannigan, Natalie Pearson, Matthias Troyer, and Peter Zoller. 2022. Practical quantum advantage in quantum simulation. *Nature* 607, 7920 (2022), 667–676.

[15] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. 2009. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*. Ieee, 248–255.

[16] Riccardo Di Sipio, Jia-Hong Huang, Samuel Yen-Chi Chen, Stefano Mangini, and Marcel Worring. 2022. The dawn of quantum natural language processing. In *ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 8612–8616.

[17] David P DiVincenzo. 1998. Quantum gates and circuits. *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences* 454, 1969 (1998), 261–276.

[18] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, et al. 2020. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929* (2020).

[19] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, and Neil Houlsby. 2021. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. arXiv:2010.11929 [cs.CV]

[20] Vedran Dunjko, Jacob M Taylor, and Hans J Briegel. 2017. Advances in quantum reinforcement learning. In *2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*. IEEE, 282–287.

[21] Matthew PA Fisher, Vedika Khemani, Adam Nahum, and Sagar Vijay. 2023. Random quantum circuits. *Annual Review of Condensed Matter Physics* 14 (2023), 335–379.

[22] Samuel J Garratt, Zack Weinstein, and Ehud Altman. 2023. Measurements conspire nonlocally to restructure critical quantum states. *Physical Review X* 13, 2 (2023), 021026.

[23] Valentin Gebhart, Raffaele Santagati, Antonio Andrea Gentile, Erik M Gauger, David Craig, Natalia Ares, Leonardo Banchi, Florian Marquardt, Luca Pezzè, and Cristian Bonato. 2023. Learning quantum systems. *Nature Reviews Physics* 5, 3 (2023), 141–156.

[24] Kai Han, Yunhe Wang, Hanting Chen, Xinghao Chen, Jianyuan Guo, Zhenhua Liu, Yehui Tang, An Xiao, Chunjing Xu, Yixing Xu, et al. 2022. A survey on vision transformer. *IEEE transactions on pattern analysis and machine intelligence* 45, 1 (2022), 87–110.

[25] Tak Hur, Leeseok Kim, and Daniel K Park. 2022. Quantum convolutional neural network for classical data classification. *Quantum Machine Intelligence* 4, 1 (2022), 3.

[26] Khawar Islam. 2022. Recent advances in vision transformer: A survey and outlook of recent work. *arXiv preprint arXiv:2203.01536* (2022).

[27] Kui Jia, Shuai Li, Yuxin Wen, Tongliang Liu, and Dacheng Tao. 2019. Orthogonal Deep Neural Networks. arXiv:1905.05929 [cs.LG]

[28] Bjarni Jónsson, Bela Bauer, and Giuseppe Carleo. 2018. Neural-network states for the classical simulation of quantum computing. *arXiv preprint arXiv:1808.05232* (2018).

[29] Cherie R Kagan and Christopher B Murray. 2015. Charge transport in strongly coupled quantum dot solids. *Nature nanotechnology* 10, 12 (2015), 1013–1026.

[30] Dimitri Kartsaklis, Ian Fan, Richie Yeung, Anna Pearson, Robin Lorenz, Alexis Toumi, Giovanni de Felice, Konstantinos Meichanetzidis, Stephen Clark, and Bob Coecke. 2021. lambeq: An efficient high-level python library for quantum NLP. *arXiv preprint arXiv:2110.04236* (2021).

[31] Iordanis Kerenidis, Jonas Landman, and Natansh Mathur. 2021. Classical and quantum algorithms for orthogonal neural networks. *arXiv preprint arXiv:2106.07198* (2021).

[32] Iordanis Kerenidis, Jonas Landman, and Anupam Prakash. 2019. Quantum algorithms for deep convolutional neural networks. *arXiv preprint arXiv:1911.01117* (2019).

[33] Alexei Kitaev and John Preskill. 2006. Topological entanglement entropy. *Physical review letters* 96, 11 (2006), 110404.

[34] Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. 1998. Gradient-based learning applied to document recognition. *Proc. IEEE* 86, 11 (1998), 2278–2324.

[35] Guangxi Li, Xuanqiang Zhao, and Xin Wang. 2022. Quantum self-attention neural networks for text classification. *arXiv preprint arXiv:2205.05625* (2022).

[36] Tianyang Lin, Yuxin Wang, Xiangyang Liu, and Xipeng Qiu. 2022. A survey of transformers. *AI Open* (2022).

[37] Seth Lloyd. 2000. Coherent quantum feedback. *Physical Review A* 62, 2 (2000), 022108.

[38] Kosuke Mitarai, Makoto Negoro, Masahiro Kitagawa, and Keisuke Fujii. 2018. Quantum circuit learning. *Physical Review A* 98, 3 (2018), 032309.

[39] Prakash Murali, Jonathan M Baker, Ali Javadi-Abhari, Frederic T Chong, and Margaret Martonosi. 2019. Noise-adaptive compiler mappings for noisy intermediate-scale quantum computers. In *Proceedings of the twenty-fourth international conference on architectural support for programming languages and operating systems*. 1015–1029.

[40] Adam Nahum, Jonathan Ruhman, and David A Huse. 2018. Dynamics of entanglement and transport in one-dimensional systems with quenched randomness. *Physical Review B* 98, 3 (2018), 035118.

[41] Crystal Noel, Pradeep Niroula, Daiwei Zhu, Andrew Risinger, Laird Egan, Debopriyo Biswas, Marko Cetina, Alexey V Gorshkov, Michael J Gullans, David A Huse, et al. 2022. Measurement-induced quantum phases realized in a trapped-ion quantum computer. *Nature Physics* 18, 7 (2022), 760–764.

[42] Karen L O'Brien. 2016. Climate change and social transformations: is it time for a quantum leap? *Wiley Interdisciplinary Reviews: Climate Change* 7, 5 (2016), 618–626.

[43] Chengwei Pan and Jiaoyang Zhang. 2022. Deep Learning-Based Quantum State Tomography With Imperfect Measurement. *International Journal of Theoretical Physics* 61, 9 (2022), 227.

[44] Daniel K Park, Carsten Blank, and Francesco Petruccione. 2020. The theory of the quantum kernel-based binary classifier. *Physics Letters A* 384, 21 (2020), 126422.

[45] Dénes Petz and Csaba Sudár. 1996. Geometries of quantum states. *J. Math. Phys.* 37, 6 (1996), 2662–2673.

[46] Varadi Rajesh, Umesh Parameshwar Naik, et al. 2021. Quantum Convolutional Neural Networks (QCNN) using deep learning for computer vision applications. In *2021 International Conference on Recent Trends on Electronics, Information, Communication & Technology (RTEICT)*. IEEE, 728–734.

[47] Maria Schuld, Ilya Sinayskiy, and Francesco Petruccione. 2015. An introduction to quantum machine learning. *Contemporary Physics* 56, 2 (2015), 172–185.

[48] Makhamisa Senekane, Benedict Molibeli Taele, et al. 2016. Prediction of solar irradiation using quantum support vector machine learning algorithm. *Smart Grid and Renewable Energy* 7, 12 (2016), 293.

[49] Joanna Slawinska, Abbas Ourmazd, and Dimitrios Giannakis. 2019. A quantum mechanical approach for data assimilation in climate dynamics. In *International Conference on Machine Learning Workshop on*.

[50] David Wierichs, Josh Izaac, Cody Wang, and Cedric Yen-Yu Lin. 2022. General parameter-shift rules for quantum gradients. *Quantum* 6 (2022), 677.

[51] Jiancheng Yang, Rui Shi, Donglai Wei, Zequan Liu, Lin Zhao, Bilian Ke, Hanspeter Pfister, and Bingbing Ni. 2023. MedMNIST v2-A large-scale lightweight benchmark for 2D and 3D biomedical image classification. *Scientific Data* 10, 1 (2023), 41.