QViSTA: A Novel Quantum Vision Transformer for Early Multi-Stage Alzheimer's Diagnosis Using Optimized Variational Quantum Circuits

Anonymous Author(s)
Affiliation
Address
email

Abstract

Magnetic resonance imaging (MRI) is widely used by neurologists to detect brain abnormalities such as strokes, tumors, and various forms of dementia, including Alzheimer's disease. However, accurately diagnosing the different stages of Alzheimer's disease remains a challenge, with nearly one in five patients misdiagnosed due to symptom overlap with other conditions. This paper introduces QViSTA, a novel hybrid quantum vision transformer (QViT) model that exploits quantum parallelism to improve early diagnosis and differentiation of Alzheimer's disease stages. By integrating quantum variational circuits (VQCs) with vision transformers (ViTs), QViSTA addresses the data scalability and computational efficiency limitations of classical machine learning models. Using a balanced, multi-class dataset of 40,000 MRI images, QViSTA achieved a validation area under the receiver operating characteristic (AUC) of 87.86% and a test AUC of 86.67%, closely matching the performance of a benchmarked classical ViT while reducing feature space by 3.18%. Early and accurate detection of Alzheimer's disease is critical, as it allows for timely interventions that can significantly improve the quality of life for patients and their caregivers. As more hospitals adopt AI for biomedical imaging, QViSTA's innovative approach could dramatically reduce misdiagnosis rates, improve patient outcomes, and reduce costs.

1 Introduction

Alzheimer's disease (AD) is the leading progressive neurodegenerative disorder globally, accounting for nearly 70% of all dementia cases. Alzheimer's leads to cognitive decline and severe memory loss. The prevalence of dementia is projected to nearly double every 20 years, reaching 78 million by 2030 and 139 million by 2050, posing substantial challenges to global healthcare systems and society [1, 2]. Despite these statistics, the cause and validated disease-modifying treatments for AD remain unknown. Consequently, there is a 20-25% misdiagnosis rate due to overlap with other conditions like Lewy body dementia and mild cognitive impairment (MCI) [3, 4]. Past studies have leveraged artificial intelligence (AI) to address the challenges of early diagnosis and differentiation of AD. For instance, Bi et al. [5] developed a deep learning model combining transfer learning and multi-task learning to improve the accuracy of Alzheimer's diagnosis, achieving improvements over traditional methods. For a comprehensive review, Zhao et al. [6] provides an overview of AI advancements in diagnosing Alzheimer's. However, these studies primarily focus on classical machine learning and deep learning models, which suffer from data scalability and computational efficiency limitations. Hence, we introduce QViSTA, a novel hybrid quantum vision transformer (QViT) model, to address these challenges. Kim [7] introduced the first quantum machine learning (QML) approach by leveraging a hybrid quantum convolutional neural network (QCNN) for Alzheimer's classification. However, the

Submitted to 38th Conference on Neural Information Processing Systems (NeurIPS 2024). Do not distribute.

approach was limited to a binary classification task (non-demented and demented images), utilized a small dataset, and used CNNs. In contrast, QViSTA handles a multiclassification task to better reflect real-world usage in a clinical setting. Additionally, QViSTA employs a larger and balanced dataset to leverage the superior performance of hybrid QML models compared to classical models when dealing with larger datasets, due to their inherent parallelism and ability to explore vast solution spaces [8]. Maurício et al. [9] compares CNNs with ViTs, demonstrating that ViTs' self-attention mechanism allows overall image information to be accessible from the surface to the deepest layers and that their parameter efficiency provides higher accuracy while using fewer computational resources and reduced training time. As QViSTA leverages a hybrid version of a ViT, it can capitalize on the strengths of ViTs, making it better suited for image classification tasks.

2 Methodology

2.1 Dataset and Preprocessing

To conduct our multi-class classification experiments, we use the dataset published by uraninjo [10] on Kaggle. This dataset contains 40,384 skull-stripped, pre-augmented MRI images. The dataset is categorized into four stages of Alzheimer's disease: Non-Demented, Very Mildly Demented, Mildly Demented, and Moderately Demented. However, we find a significant class imbalance among the labels, which could lead to a biased model. To address this, we apply additional augmentations (random flips and 5°rotations) to upsample underrepresented classes to 10,000 images and downsample classes over 10,000 images, ultimately achieving a balanced dataset of 40,000 images. To prepare the dataset for model development, we convert the images to grayscale to reduce dimensionality and better replicate MRI scans. We further reduce the dimensionality of the images to 128 by 128 pixels and normalize them using mean and standard deviation normalization. Finally, we perform an 80-10-10 training-validation-test split to run our experiments. Sample images from the final dataset(`https://www.kaggle.com/datasets/aryansinghal10/alzheimers-multiclass-dataset-equal-and-augmented`) are depicted in Figure 1. The codebase for QViSTA can be found in the following GitHub repository: `https://github.com/3x-dev/QViSTA`.

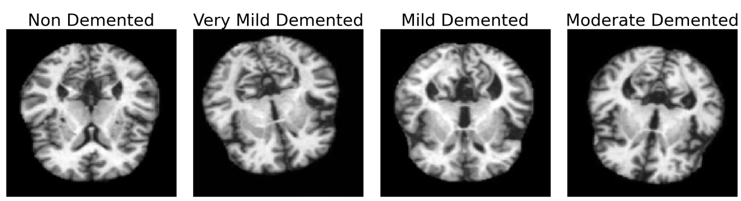

Figure 1: Sample images for each stage of Alzheimer's from the final dataset.

2.2 QViSTA Development

To develop QViSTA we first leverage a multi-layer perceptron (MLP), described as a composition of elementwise non-linearities (activation function) with affine transformations of the data [11].

The affine transformation is defined as:

$$a(x) = Wx + b,$$

and the activation function is applied to each component of the output vector a:

$$f(x) = \sigma(a(x)),$$

where σ denotes the activation function. For our activation function, we use Gaussian Error Linear Unit (GeLU) [12], defined as:

$$\text{GELU}(x) = x\Phi(x),$$

Apart from MLP, we leverage the main building block of a transformer architecture [13] by taking a matrix $X \in \mathbb{R}^{N\times D}$ and transforming it. Each of these layers has two sub-layers: a multi-head self-attention mechanism (MHA), the core of the transformer, and a simple MLP:

$$Z = X + \text{LayerNorm}(\text{MHA}(X, X, X)),$$

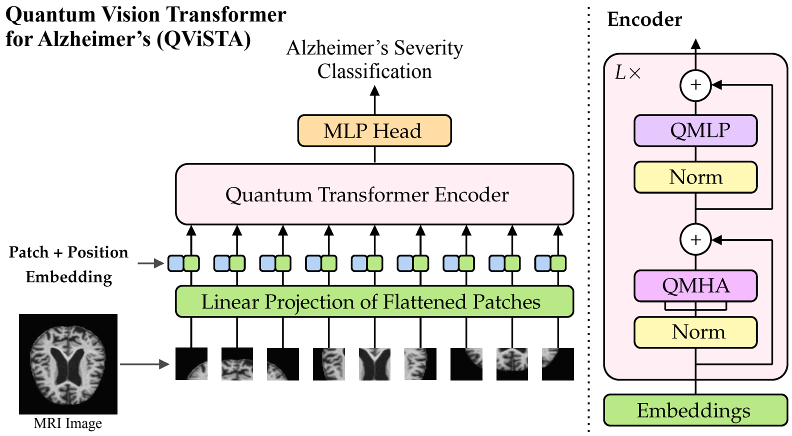

Figure 2: QViSTA architecture overview for Alzheimer's classification.

$$X' = Z + \text{LayerNorm}(\text{MLP}(Z)).$$

The attention function is vital, allowing the transformer to focus on specific input patches. The attention function is defined as [13]:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{D_k}}\right)V,$$

where D_k is the dimension of the keys. The baseline vision transformer [14] divides the image into patches given by $N = \frac{HW}{P^2}$ and then transforms it into patch embeddings:

$$z_i^0 = Ex_i' + p_i$$

In quantum computing, the fundamental unit of information is the qubit which can exist in a superposition state to represent non-binary states. Qubits can be defined with the unit vector $|\psi\rangle$ in the Hilbert space $\mathbb{C}^{2^n}$. A quantum circuit is a series of "gates" to change a qubit state represented by $U|\psi\rangle$ where U is a $2^n \times 2^n$. For QViSTA, we use an R_x gate, which performs a single qubit rotation along the x-axis, and the CNOT gate, which operates over two qubits and flips the target qubit only if the first qubit is $|\psi\rangle$, represented by the following matrices:

$$R_X(\theta) = \begin{bmatrix} \cos(\theta/2) & -i\sin(\theta/2) \\ -i\sin(\theta/2) & \cos(\theta/2) \end{bmatrix}$$

$$\text{CNOT} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix}.$$

As with the classical ViT, the image is split into patches linearly embedded with position embeddings defined by the patch size. For QViSTA, however, these patches are fed to the Quantum Transformer Encoder, which employs VQCs in the multi-head attention (MHA) and multi-layer perceptron (MLP) components. An overview of QViSTA's architecture is depicted in Figure 2[1]. The configuration of the VQC we use is depicted in Figure 3[2]. Initially, each feature of the vector $\mathbf{x} = (x_0, \ldots, x_{n-1})$ is converted into rotation angles and embedded into the qubits. Subsequently, a layer of single-qubit rotations, parameterized by $\theta = (\theta_0, \ldots, \theta_{n-1})$, operates on each qubit. These parameters are optimized alongside the other model parameters. Following this, a ring of CNOT gates is applied to entangle the qubit states, emulating the effect of matrix multiplication. Finally, each qubit is measured, and the output proceeds to the subsequent component of the encoder. We use Ray Tune [16] to tune the hyperparameters and employ its advanced algorithms, such as Population Based Training (PBT) and HyperBand/ASHA [17], to optimize QViSTA for maximum robustness and efficiency. Both

[1]The figure is inspired by [14], but has been modified to reflect the architecture for QViSTA.

[2]The configuration is inspired by [15], but has been modified to reflect the configuration for QViSTA.

Table 1: Tuned hyperparameters used to define QViSTA's network.

Hyperparameter	Value
Batch Size	16
Epochs	30
Patch Size	64
Hidden Size	6
Hidden MLP Size	5
Number of Transformer Blocks	6
Number of Attention Heads	3
Optimizer	AdamW
Gradient Clipping	Norm 1
Learning Rate Scheduler	Linear warmup (9K steps: 0 to 10^{-3}); cosine decay (70K steps)
Total number of hyperparameters θ: 25,390 for quantum; 26,224 for classical	

QViSTA and the classical ViT are trained with the same hyperparameters for consistent comparison. We use the AdamW optimizer with gradient clipping to ensure stability and robustness by preventing large gradients from hindering optimization. A cosine annealing learning rate scheduler with warm-up and cosine decay is employed for smooth convergence, particularly beneficial for transformer models [18]. A detailed breakdown of the model hyperparameters is shown in Table 1. To evaluate the performance of QViSTA, we use the Receiver Operating Characteristic (ROC) curve. For AD classification, this curve represents the model's ability to correctly predict a scan (TPR: true positive rate) versus its ability to incorrectly predict a scan (FPR: false positive rate). For each epoch of each model configuration, we compute the area under the ROC curve (AUC). After all epochs are run, we select the parameters from the epoch with the highest validation AUC and re-evaluate them on a separate test batch to obtain the final test AUC. We use Google's JAX [19] and Flax [20] libraries to implement and train the classical components of QViSTA and the classical baseline (ViT). In addition, we use TensorCircuit [21] to implement, train, and execute the VQCs through mathematical simulations on an Intel CPU. TensorCircuit enables rapid training of the quantum model, achieving approximately two minutes per epoch.

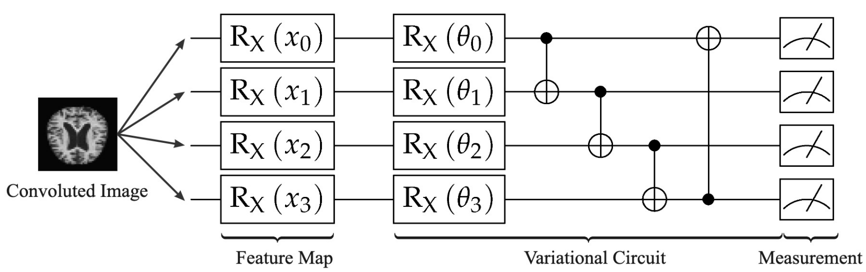

Figure 3: VQC configuration where Rx denotes rotations around the X-axis.

3 Results and Discussion

QViSTA and the baseline ViT's AUC scores and confusion matrices are depicted in Figure 4. We find that QViSTA achieved a validation AUC of 87.86% and a test AUC of 86.67%. The baseline ViT had a validation AUC of 88.39% and a test AUC of 88.39%. The ROC curve for QViSTA indicates that it performs best in classifying Moderate Demented cases with an AUC of 0.96 and worst in classifying Very Mild Demented cases with an AUC of 0.70. The ViT follows a similar performance pattern, performing best for Moderate Demented cases with an AUC of 0.97 and worst for Very Mild Demented cases with an AUC of 0.74. Observing the confusion matrices, QViSTA achieves the highest TPR for Moderate Demented cases, with 861 correctly identified out of 900. Very Mild Demented cases demonstrate the highest misclassification rates, with only 483 correctly identified. In comparison, the baseline ViT also shows strong performance in identifying Moderate Demented cases, with 880 correct classifications. Similar to QViSTA, the Very Mild Demented cases perform

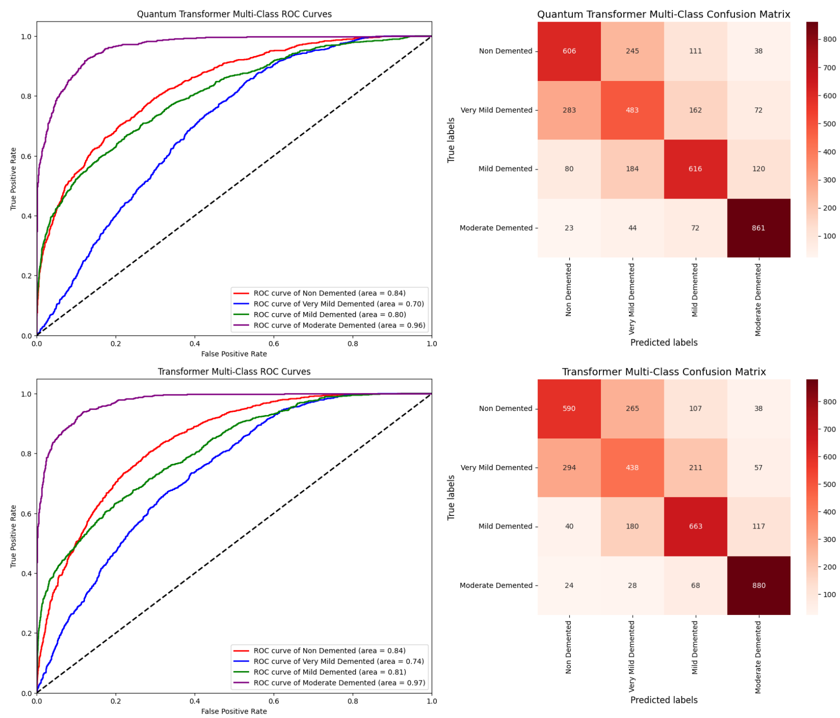

Figure 4: Images on the right: ROC curves for QViSTA and baseline ViT. The black dashed line represents the performance of a random classifier. Images on the left: Multiclass label confusion matrices for QViSTA and baseline ViT.

the worst, with only 438 correctly identified instances. We believe the models performed better on the Moderate Demented cases as they present more pronounced symptoms, leading to higher classification accuracy as the models can more easily identify the more significant deviations in the data. Conversely, the models performed poorer on Non Demented and Very Mild Demented cases as the subtle differences in symptoms and features between these stages make it challenging for the models to differentiate them accurately. Both models peak at epoch 30, suggesting an equal rate of convergence. We observed that QViSTA performed very similarly to ViT, with a slight difference in test accuracy and slightly higher ROC areas for ViT. However, parameter usage seemed to favor QViSTA, placing it as the lighter and potentially more efficient of the two. This may imply better use on hospital computers. We believe that the simulation of qubits resulted in significant memory consumption and reduced accuracy. While accuracy ended up being slightly lower for QViSTA, the simulation of qubits seemed to play a prominent role in the difference. We hypothesize that it is harder for the optimizer to find good parameters for these mathematically simulated VQCs, resulting in a slightly lower accuracy score for QViSTA. In addition, these simulated VQCs are not able to truly exploit quantum parallelism, resulting in naturally inferior robustness compared to a true quantum computer.

4 Conclusions and Future Work

In this paper, we introduced QViSTA, a novel QViT architecture designed for multi-stage early diagnosis of Alzheimer's. The novelty comes from applying this architecture for multi-stage early Alzheimer's diagnosis and introducing optimized VQCs designed for this task. QViSTA was benchmarked against a classical ViT and used a smaller feature space to achieve comparable performance. We aim to advance QViSTA by implementing multimodality with PET scans and genetic data. Furthermore, we aim to include other quantum-inspired optimization algorithms, such as Quantum Annealing [22]. Finally, we hope to leverage actual quantum hardware[3], for QViSTA and investigate its performance.

[3] `https://www.ibm.com/quantum`

References

[1] Dementia, Mar 2023. URL https://www.who.int/news-room/fact-sheets/detail/dementia.

[2] Dementia statistics, Jun 2020. URL https://www.alzint.org/about/dementia-facts-figures/dementia-statistics/.

[3] It's not always dementia: Top 5 misdiagnoses, Jun 2024. URL https://www.humangood.org/resources/senior-living-blog/top-five-dementia-misdiagnoses.

[4] Thousands are misdiagnosed with dementia every year, Jun 2024. URL https://news.umiamihealth.org/en/its-not-always-dementia-heres-what-to-know/.

[5] Xia-an Bi, Xi Hu, Hao Wu, and Yang Wang. Multimodal data analysis of alzheimer's disease based on clustering evolutionary random forest. *IEEE Journal of Biomedical and Health Informatics*, 24(10):2973–2983, 2020. doi: 10.1109/JBHI.2020.2973324.

[6] Z. Zhao, J. H. Chuah, K. W. Lai, C. O. Chow, M. Gochoo, S. Dhanalakshmi, N. Wang, W. Bao, and X. Wu. Conventional machine learning and deep learning in alzheimer's disease diagnosis using neuroimaging: A review. *Frontiers in Computational Neuroscience*, 17:1038636, 2023. doi: 10.3389/fncom.2023.1038636. URL https://doi.org/10.3389/fncom.2023.1038636.

[7] Ryan Kim. Implementing a hybrid quantum-classical neural network by utilizing a variational quantum circuit for detection of dementia, 2023. URL https://arxiv.org/abs/2301.12505.

[8] Kamila Zaman, Tasnim Ahmed, Muhammad Abdullah Hanif, Alberto Marchisio, and Muhammad Shafique. A comparative analysis of hybrid-quantum classical neural networks, 2024. URL https://arxiv.org/abs/2402.10540.

[9] José Maurício, Inês Domingues, and Jorge Bernardino. Comparing vision transformers and convolutional neural networks for image classification: A literature review. *Applied Sciences*, 13(9):5521, 2023. doi: 10.3390/app13095521. URL https://www.mdpi.com/2076-3417/13/9/5521.

[10] uraninjo. Augmented alzheimer mri dataset v2, 2020. URL https://www.kaggle.com/datasets/uraninjo/augmented-alzheimer-mri-dataset-v2. Accessed: 2024-06-25.

[11] Juergen Schmidhuber. Annotated history of modern ai and deep learning. *arXiv preprint arXiv:2212.11279*, 2022. URL https://arxiv.org/abs/2212.11279. Technical Report IDSIA-22-22.

[12] Dan Hendrycks and Kevin Gimpel. Gaussian error linear units (gelus), 2023. URL https://arxiv.org/abs/1606.08415.

[13] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *Advances in Neural Information Processing Systems 30 (NIPS 2017)*, 2017. URL https://proceedings.neurips.cc/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html.

[14] 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. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*, 2021. URL https://arxiv.org/abs/2010.11929.

[15] Marçal Comajoan Cara, Gopal Ramesh Dahale, Zhongtian Dong, Roy T. Forestano, Sergei Gleyzer, Daniel Justice, Kyoungchul Kong, Tom Magorsch, Konstantin T. Matchev, Katia Matcheva, and Eyup B. Unlu. Quantum vision transformers for quark–gluon classification. *Axioms*, 13(5):323, May 2024. ISSN 2075-1680. doi: 10.3390/axioms13050323. URL http://dx.doi.org/10.3390/axioms13050323.

[16] Richard Liaw, Eric Liang, Robert Nishihara, Philipp Moritz, Joseph E. Gonzalez, and Ion Stoica. Tune: A research platform for distributed model selection and training, 2018. URL https://arxiv.org/abs/1807.05118.

[17] Richard Liaw, Eric Liang, Robert Nishihara, Philipp Moritz, Joseph E. Gonzalez, and Ion Stoica. Tune: A research platform for distributed model selection and training. *arXiv preprint arXiv:1807.05118*, 2018. URL https://arxiv.org/pdf/1807.05118.

[18] Ilya Loshchilov and Frank Hutter. Sgdr: Stochastic gradient descent with warm restarts. *arXiv preprint arXiv:1608.03983*, 2016. URL https://arxiv.org/pdf/1608.03983. University of Freiburg.

[19] Google Research. Jax: Autograd and xla. https://github.com/google/jax, 2023. Accessed: 2023-06-26.

[20] Google Research. Flax: A neural network library and ecosystem for jax. https://github.com/google/flax, 2023. Accessed: 2023-06-26.

[21] Shi-Xin Zhang, Jonathan Allcock, Zhou-Quan Wan, Shuo Liu, Jiace Sun, Hao Yu, Xing-Han Yang, Jiezhong Qiu, Zhaofeng Ye, Yu-Qin Chen, Chee-Kong Lee, Yi-Cong Zheng, Shao-Kai Jian, Hong Yao, Chang-Yu Hsieh, and Shengyu Zhang. Tensorcircuit: a quantum software framework for the nisq era. *Quantum*, 7:912, February 2023. ISSN 2521-327X. doi: 10.22331/q-2023-02-02-912. URL http://dx.doi.org/10.22331/q-2023-02-02-912.

[22] Hadi Salloum, Hamza Shafee Aldaghstany, Osama Orabi, Ahmad Haidar, Mohammad Reza Bahrami, and Manuel Mazzara. Integration of machine learning with quantum annealing. *Advanced Information Networking and Applications*, pages 338–348, 2024. doi: 10.1007/978-3-031-57870-0_30. URL https://www.researchgate.net/publication/379781610_Integration_of_Machine_Learning_with_Quantum_Annealing.