Quantum-Enhanced Transformer Networks for Mammography Classification

Khin Thandar Kyaw, Usama Inam Paracha, Uman Khalid, and Hyundong Shin
Department of Electronics and Information Convergence Engineering, Kyung Hee University, Korea
Email: hshin@khu.ac.kr

***Abstract*—Conventional deep learning methodologies have been widely utilized for breast cancer classification. This paper proposes a quantum-aware layer in the classical transformer model. The model, trained with the quantum vision transformer (Q-ViT), achieves greater precision in classifying breast cancer images, with a precision of 80.77% and an area under the curve (AUC) of 0.7431. In contrast, the classical model attains only 77.56% precision with an AUC of 0.7335. Our results indicate that the quantum-enhanced model could provide improved ability to capture complex distributions, despite the presence of class imbalance.**

I. Introduction

The accurate and early diagnosis of breast cancer is critical for patients [1]. Mammography is the primary method for breast cancer screening, and deep learning aided diagnosis systems have become an active area of research. Vision transformers (ViTs) have demonstrated remarkable performance in various image classification tasks and excel in smaller datasets [2]. The field of quantum computing research is rapidly advancing, with increasing potential in a range of domains, including applications in vision-related tasks [3]–[5].

Recent advances in quantum machine learning (QML), particularly variational quantum algorithms (VQAs), have shown promise in efficiently extracting high-dimensional features from complex data using parameterized quantum circuits (PQCs) [6], [7]. Unlike classical networks, VQA exploit quantum parallelism and entanglement to represent richer feature spaces with fewer parameters, making them attractive for noisy intermediate-scale quantum devices (NISQ). Integrating such quantum layers into classical architectures, such as the ViT, opens the possibility of hybrid quantum-classical models that can potentially outperform purely classical methods in medical imaging tasks.

In this work, we develop a Q-ViT and evaluate its performance against a baseline classical ViT using the BreastMNIST dataset. Section II details the proposed methodology, including the integration of quantum layers into the classical transformer architecture. Section III presents the experimental results and performance comparison between the classical ViT and the proposed Q-ViT. Finally, Section IV provides a summary of the principal findings and offers recommendations for future work.

II. Methodology

We utilize the BreastMNIST dataset, a specialized subset of MedMNIST, which consists of 780 breast ultrasound images. The dataset is split into 70% for training, 10% for validation, and 20% for testing. The proposed model integrates a hybrid quantum-classical framework with the ViT for mammography classification. The model is formulated with three integral components: image patch-based feature extraction, transformer-based feature encoding, and a quantum-enhanced classification head. The Q-ViT deconstructs the input image, $\boldsymbol{x} \in \mathbb{R}^{H\times W\times C}$, into a sequence of fixed-size patches. The image is divided into a sequence of $N = (H/P)^2$ flattened patches, and each element in the sequence can be represented by x_p^i. These patches are linearly projected into a D dimensional latent space. To provide a global representation for classification, a learnable class token, $\boldsymbol{x}_{\text{cls}}$, is appended to this sequence. Following this, learnable positional embeddings, $\mathbf{E}_{\text{pos}}$, are added to each patch embedding to retain crucial spatial information to prevent permutation-invariant information.

The input to the Transformer encoder is the sequence, $\boldsymbol{z}_0 = (\boldsymbol{x}_{\text{cls}}, \boldsymbol{x}_{p1}^{\mathbf{E}}, \ldots, \boldsymbol{x}_{pN}^{\mathbf{E}}) + \mathbf{E}_{\text{pos}}$, where $\boldsymbol{z}_0 \in \mathbb{R}^{(N+1)\times D}$. Identical layers are included in the Transformer encoder. Each layer has a multi-head self-attention (MHSA) and a position-wise network. Residual connections are added with layer normalization (LN) to stabilize the network. The MHSA structure captures complex correlations between different areas of the image. The sequence $\boldsymbol{z}_0$ is processed through identical L transformer blocks. For each block $l \in 1, \ldots, L$, the rule is updated by MHSA and a multilayer perceptron (MLP) with LN and residual connections as

$$\boldsymbol{z}_l' = \text{MHSA}\Big(\text{LN}\big(\boldsymbol{z}_{l-1}\big)\Big) + \boldsymbol{z}_{l-1} \tag{1}$$

$$\boldsymbol{z}_l = \text{MLP}\Big(\text{LN}\big(\boldsymbol{z}_l'\big)\Big) + \boldsymbol{z}_l'. \tag{2}$$

The output of the class token of the final block, $\boldsymbol{z}_L^0$ represents the image. The final prediction of the classical Transformer model is obtained by

$$\hat{y}_c = \text{Softmax}\Big(\text{Linear}\big(\boldsymbol{z}_L^0\big)\Big). \tag{3}$$

To incorporate quantum feature extraction, we replace the classical classification head with a variational quantum circuit (VQC). The output of the final transformer block, $\boldsymbol{z}_l^0 \in \mathbb{R}^D$, is first linearly projected into a lower-dimensional space $\boldsymbol{z}_q \in \mathbb{R}^{d_q}$, where d_q equals the number of qubits. The VQC encodes $\boldsymbol{z}_q$ using parameterized single-qubit rotations followed by entangling CNOT layers across all qubits. Trainable parameters $\boldsymbol{\theta}$ are optimized via classical backpropagation, while the

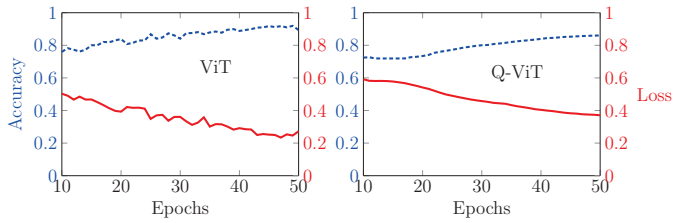

Figure 1. Comparison of accuracy and loss of the training dataset for the ViT and Q-ViT model.

Table I
PERFORMANCE COMPARISON OF VIT AND Q-VIT MODELS FOR BREAST CANCER IMAGE CLASSIFICATION.

Method	Accuracy	AUC
ViT (Classical Vision Transformer)	77.56	0.7335
Q-ViT (Quantum Vision Transformer)	**80.77**	**0.7431**

quantum circuit outputs measurement probabilities over the computational basis:

$$q(\boldsymbol{z}_q, \boldsymbol{\theta}) = \mathrm{VQC}(\boldsymbol{z}_q; \boldsymbol{\theta}) \in \mathbb{R}^{2^{d_q}}. \tag{4}$$

Finally, the quantum features $q(\boldsymbol{z}_q, \boldsymbol{\theta})$ are concatenated with the classical projection $\boldsymbol{z}_q$ and forwarded to a linear layer to obtain the final prediction:

$$\hat{y} = \mathrm{Softmax}\big(\mathrm{Linear}([\boldsymbol{z}_q \,\|\, q(\boldsymbol{z}_q, \boldsymbol{\theta})])\big). \tag{5}$$

III. RESULTS

We evaluate the performance of the ViT and Q-ViT. Training for both models was carried out over 50 epochs using the AdamW optimizer with 6 layers, 8 heads, a learning rate of $1e-3$, and a weight decay of 0.01. A learning rate scheduler was applied to change the learning rate during training, depending on the validation loss. The comparison of accuracy and loss of the training dataset for both models is shown in Figure 1. The line plots show that the application of quantum layers in the classical model gives almost similar training results.

The performance comparison between the ViT and the Q-ViT on the test set is summarized in their confusion matrices. The Q-ViT demonstrates improved classification accuracy by correctly identifying 104 malignant cases compared to 98 by the classical model in Figure 2. It also reduces false negatives to 10 from 16 and indicates fewer missed malignant detections. The quantum model correctly classifies 22 benign cases, whereas the classical one classifies 18 benign cases by misclassifying fewer benign samples. As described in Table I, the accuracy and AUC of the proposed Q-ViT outperform those of the ViT. These results suggest that the Q-ViT has the potential to exhibit superior discriminative capability and robustness to distinguish benign and malignant cases.

IV. CONCLUSION

This study explored Q-ViT in mammography classification. The model shows promise when dealing with unevenly distributed classes or limited data sizes. We observed that the proposed model outperformed the classical vision transformer on the test set. These findings demonstrate that Q-ViT is promising for the classification of binary medical images.

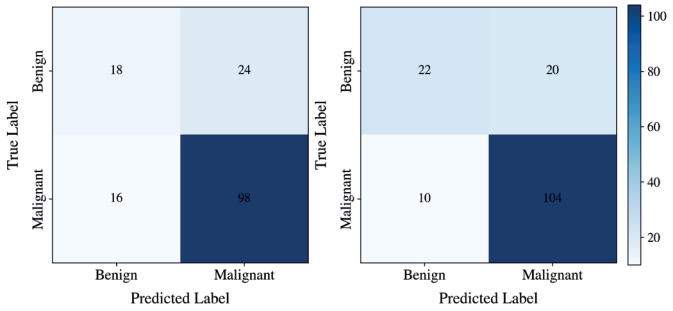

Figure 2. Confusion matrices comparing classification performance on test dataset: Left shows results from the ViT, and right shows results from the Q-ViT.

ACKNOWLEDGMENT

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korean government (MSIT) under RS-2025-00556064, by the MSIT (Ministry of Science and ICT), Korea, under the ITRC (Information Technology Research Center) support program (IITP-2025-RS-2021-II212046) supervised by the IITP (Institute for Information & Communications Technology Planning & Evaluation), and by a grant from Kyung Hee University in 2023 (KHU-20233663).

REFERENCES

[1] X. Liu, L. Sun, C. Li, B. Han, W. Jiang, T. Yuan, W. Liu, Z. Liu, Z. Yu, and B. Liu, "Lesion asymmetry screening assisted global awareness multi-view network for mammogram classification," *IEEE Trans. Med. Imag.*, vol. PP, pp. 1–1, Sep 2025.

[2] C. Wu and T. He, "A survey of applications of vision transformer and its variants," *2024 10th IEEE International Conference on Intelligent Data and Security (IDS)*, pp. 21–25, May 2024.

[3] S. M. A. Rizvi, U. I. Paracha, U. Khalid, K. Lee, and H. Shin, "Quantum machine learning: Towards hybrid quantum-classical vision models," *Mathematics*, vol. 13, no. 16, p. 2645, Aug. 2025.

[4] U. Khalid, M. S. Ulum, M. Z. Win, and H. Shin, "Integrated satellite-ground variational quantum sensing networks," *IEEE Commun. Mag.*, vol. 62, no. 10, pp. 20–27, Oct. 2024.

[5] U. Khalid, M. S. Ulum, A. Farooq, T. Q. Duong, O. A. Dobre, and H. Shin, "Quantum semantic communications for Metaverse: Principles and challenges," *IEEE Wireless Commun.*, vol. 30, no. 4, pp. 26–36, Aug. 2023.

[6] M. Cerezo, A. Arrasmith, R. Babbush, S. C. Benjamin, S. Endo, K. Fujii, J. R. McClean, K. Mitarai, X. Yuan, and L. Cincio, "Variational Quantum Algorithms," *Nat. Rev. Phys.*, vol. 3, no. 9, pp. 625–644, Sep 2021.

[7] U. Khalid, U. I. Paracha, Z. Naveed, T. Q. Duong, M. Z. Win, and H. Shin, "Quantum fusion intelligence for integrated satellite-ground remote sensing," *IEEE Wireless Commun.*, vol. 32, no. 3, pp. 46–55, Jun. 2025.