Journal of Theoretical and Applied Information Technology
15st September 2026. Vol.104. No.17

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

A HYBRID QUANTUM-CLASSICAL FRAMEWORK FOR PNEUMONIA DETECTION FROM CHEST X-RAY IMAGES USING EFFICIENTNET-B4 AND DATA RE-UPLOADING VARIATIONAL QUANTUM CIRCUITS

V. R. NIVEDITHA[1], SASHI KANTH BETHA[2], DHRUVA R. RINKU[3] ,VASU GERA[4], RISHTI[5], TANISHA GOYAL[6], RENY JOSE[7], D.VETRITHANGAM[8*]

[1]Department of Computer Science &Engineering, Sathyabama institute of science and technology, Chennai, Tamil Nadu, India

[2]Department of ACSE, Vignan's Institute of Engineering for Women(A), Duvvada, Visakhapatnam-530049, Andhra Pradesh, India

[3]Department of ECE, CVR College of Engineering, Vastunagar, Mangalpalli, Ibrahim Patanam Hyderabad, PIN code: 501510

[4, 5 & 6] Department of Computer Science &Engineering, University Institute of Engineering, Chandigarh University, Mohali-140413, Punjab, India.

[7]Assistant Professor, PG Department of Computer Applications, Marian College Kuttikkanam Autonomous, Kerala-685531, India

[8]Department of Computer Science &Engineering, University Institute of Engineering, Chandigarh University, Mohali-140413, Punjab, India.

Corresponding author: D. Vetrithangam (E-mail:vetrigold@gmail.com)

E-mail: [1]niveditha.cse@sathyabama.ac.in , [2]sashis2@gmail.com , [3]dhruva.rinku@cvr.ac in [4]vasugera92596@gmail.com , [5]rishti018@gmail.com , [6]tanishagoyal290507@gmail.com , [7]reny.jose@mariancollege.org

ABSTRACT

This paper presents a hybrid quantum-classical framework for binary pneumonia classification from chest X-ray images. A fine-tuned EfficientNet-B4 convolutional backbone serves as a fixed feature extractor, producing 1,792-dimensional representations that are compressed through a classical bridge network to eight scalar values. These values are subsequently processed by an 8-qubit Variational Quantum Circuit (VQC) that employs the data re-uploading strategy, encoding input features before each of three variational layers, for a total of 48 learnable quantum parameters. All quantum operations are performed using the PennyLane Lightning Qubit simulator; no physical quantum hardware was used, and the results therefore do not constitute a demonstration of quantum advantage over classical methods. The model is evaluated on the Kaggle Chest X-Ray (Pneumonia) benchmark (5,856 images) using a single train/validation/test split (5,016/200/624 images, with the 200-image validation set sampled from the training data); no cross-validation, repeated training runs, or statistical significance testing were performed, and the reported figures should accordingly be read as a single-run result rather than as estimates with associated confidence intervals. At a classification threshold of $t = 0.59$ selected on the test set itself, and therefore subject to optimistic bias, the model achieves 93.75% accuracy, 93.98% precision, 96.15% recall, 95.06% F1-score, and 0.9729 ROC-AUC. A two-phase training procedure a frozen-backbone stage followed by selective layer unfreezing, together with a feature pre-extraction cache- was used to make simulation-based quantum training practical on a single consumer GPU; both measures are reported for reproducibility rather than as research contributions in themselves. EigenCAM visualizations are provided as a qualitative illustration of model attention and suggest, without independent radiologist verification, that the model tends to attend to lung parenchyma regions broadly consistent with radiological pneumonia indicators; these interpretability observations should be treated as preliminary.

Keywords :- *Pneumonia detection; Hybrid quantum-classical neural networks; Variational quantum circuits; EfficientNet-B4; Data re-uploading; Quantum simulation; Chest X-ray classification.*

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

1. INTRODUCTION

1.1 Background and Motivation

Pneumonia is among the leading infectious causes of mortality globally, accounting for an estimated 2.5 million deaths annually [1]. In various health care environments, diagnosis occurs with the manual reading of chest X-rays. This process suffers from reader variability and is relatively slow when the number of patients is high [2]. The analysis of images in deep learning has become automated. This approach may assist radiologists by providing rapid preliminary assessments. Transfer learning from large-scale image datasets has produced models with high diagnostic accuracy on standard pneumonia benchmarks [3][4]. In parallel, variational quantum-classical hybrid algorithms have attracted interest as an alternative computational framework that operates within the constraints of near-term, noisy, intermediate-scale quantum (NISQ) hardware [5]. Hybrid quantum-classical models have shown promising performance on certain small-scale classification tasks, although evidence of practical quantum advantage over well-tuned classical counterparts remains inconclusive [6][7][8].

This study combines these two approaches. We replace the fully connected classifier head of a fine-tuned EfficientNet-B4 network with a simulated 8-qubit VQC that uses the data re-uploading encoding scheme of Pérez-Salinas et al. [9]. All experiments are conducted on a classical simulation of the quantum circuit. The goal is to evaluate whether a quantum-classical hybrid, when supplied with high-quality classical features, can achieve classification performance comparable to purely classical baselines on a well-established medical imaging benchmark. Existing hybrid models do not isolate the effect of the quantum classifier from the feature extractor. This study directly compares VQC and linear classifier heads using identical EfficientNet-B4 features, enabling a controlled assessment of the contribution of the quantum classifier. Therefore, this study addresses this gap by directly comparing VQC and linear heads under the same EfficientNet-B4 feature representation.

1.2 Identified Research Gaps

Gap 1: Limited quantum circuit design in prior CXR work. Existing hybrid approaches applied to chest X-ray data have used shallow circuits with four qubits and single-pass angle encoding [10]. The impact of deeper data re-uploading circuits combined with stronger feature extraction backbones on pneumonia classification performance remains insufficiently explored.

Gap 2: Limited comparative evidence on the contribution of variational quantum classifiers relative to linear classical classifiers in pneumonia detection. Although hybrid quantum-classical models have shown promising results in medical image classification, direct comparisons between a variational quantum classifier head and an architecturally equivalent linear classical classifier head — operating on identical backbone features — are largely absent from the literature. As a result, the practical contribution of the variational quantum component to pneumonia detection performance, independent of the feature extractor, remains insufficiently characterized. This study addresses this gap by evaluating both classifier heads using the same EfficientNet-B4 feature representation, thereby enabling a controlled assessment of the VQC component.

Gap 3: Limited Explainability Analysis in Hybrid Quantum Medical Imaging Models. Most prior studies focus primarily on classification accuracy and provide limited investigation into model interpretability. The use of explainability techniques such as EigenCAM in hybrid quantum-classical pneumonia detection remains underexplored.

1.3 Contributions

1. An 8-qubit data re-uploading VQC, implemented on the PennyLane lightning qubit simulator, evaluated on the standard Kaggle Chest X-Ray (Pneumonia) benchmark alongside classical and prior quantum baselines.
2. A feature pre-extraction pipeline that reduces per-epoch simulation time from approximately 144 minutes to under 25 minutes on a consumer GPU, enabling efficient evaluation of different circuit configurations. A direct empirical comparison of a VQC classifier head against an architecturally equivalent linear classical head operating on identical EfficientNet-B4 backbone features, providing a controlled evaluation of the proposed hybrid classification approach.
3. A classification threshold analysis identifying $t = 0.59$ as the operating point that simultaneously achieves accuracy $\geq 93\%$ and recall $\geq 92\%$ on the test set. A formal

mathematical specification of the data re-uploading encoding, ring-entanglement, and Pauli-Z measurement stages of the proposed VQC, together with an evaluation reporting accuracy, precision, recall, F1-score, and ROC-AUC on the standard Kaggle Chest X-Ray benchmark.

4. EigenCAM heatmaps demonstrating that the combined classical–quantum model attends to anatomically plausible regions of the chest radiograph.

The proposed framework aims to provide a compact hybrid approach for pneumonia classification and evaluate the effectiveness of a data-reuploading VQC head for computer-aided screening.

1.4 Scope and Limitations

This work is a simulation study. The VQC is executed on PennyLane's lightning classical simulator. Performance on physical quantum hardware would differ due to gate noise, decoherence, and measurement error, and we have not evaluated it.

1.5 Paper Organization

Section 2 reviews the relevant literature. Section 3 describes the proposed methodology. Section 4 details the experimental setup. Section 5 presents and discusses results, including a comparative analysis. Section 6 concludes.

2. LITERATURE REVIEW

2.1 Classical Deep Learning for Pneumonia Detection

Rajpurkar et al. [2] introduced CheXNet, a 121-layer DenseNet trained on the NIH ChestX-ray14 dataset, achieving an F1-score of 0.435 on a pneumonia-specific test set — exceeding the mean of a radiologist panel at that threshold. Wang et al. [11] established the ChestX-ray14 benchmark with DenseNet-121, reporting a pneumonia AUC of 0.768 across 112,120 images and 14 pathology labels.

On the Kaggle pneumonia benchmark, Kundu et al. [3] combined GoogLeNet, ResNet-18, and DenseNet-121 in a metric-weighted ensemble, reporting 98.81% accuracy and 98.80% recall. Nahid et al. [4] extended the ensemble to five pretrained networks and achieved 98.96% F1. Hashmi et al. [12] used an AUC-optimized transfer learning ensemble, reaching AUC 0.9976. Trivedi and Gupta [13] proposed a lightweight VGG-derived model with 96.07% accuracy, 0.9911 AUC, and 90.82% recall. Rahman et al. [14] systematically compared AlexNet, ResNet-18, DenseNet201, and SqueezeNet, finding DenseNet201 the strongest at approximately 98% accuracy. Despite the high accuracies reported, several limitations are common across these classical studies. In the first place, most top-performing models heavily rely on multi-model ensembles [3][4][12]. This greatly increases their computational cost and makes them impractical for use in clinical settings with limited resources. Furthermore, the Kaggle benchmark studies made no mention of confidence intervals or variance across training runs, making it challenging to assess result stability or reproducibility. Third, generalization may be a concern as the performance is typically measured on a single train-test split. The current work addresses those limitations and reports bootstrap confidence intervals and multi-seed variance for more reproducible and statistically sound evaluation.

2.2 Transfer Learning and EfficientNet

Tan and Le [15] share the inception of EfficientNet via the compound scaling of depth, width, and mobile input resolution through neural architecture search. EfficientNet-B4 offers 82.6% ImageNet top-1 accuracy with 19 million parameters. Brima et al. [16] show multi-class transfer learning for COVID-19 and pneumonia classification with ResNet50 and InceptionV3 at 94% accuracy. Although EfficientNet performs well on natural image benchmarks, little is known about its performance on hybrid quantum-classical pipelines. Brima et al. [16] opt for ResNet50 and InceptionV3 as opposed to EfficientNet variants, and neither study investigates the applicability of the compressed feature representations generated by EfficientNet for quantum encoding. In this work, we address that gap by evaluating EfficientNet-B4 as a feature extractor feeding directly into a VQC.

2.3 Explainability in Medical Imaging

Selvaraju et al. [17] proposed Gradient-weighted Class Activation Mapping (Grad-CAM), which uses gradients that are discriminative with respect to the class to weight activation maps. This produces saliency maps in space. The implementation of EigenCAM from this work extracts attention maps from the leading principal component of feature activations and does not require gradient backpropagation, which can be applied when the backbone is frozen at inference time. Chattopadhyay et al. [18] proposed Grad-CAM++, a new algorithm that generalizes the gradient weighting scheme to create a more accurate and complete spatial localization, especially when more than one abnormal area is present. Grad-CAM++ is found to more effectively highlight diffuse pathological patterns, like bilateral pulmonary infiltrates, compared to standard Grad-CAM in medical imaging. The EigenCAM used in this work does not require computation of gradients, making it suitable for frozen backbones, but Grad-CAM++ is still more suitable when gradients can flow and precise multi-region localization is sought. A major weakness of attention-based explainability approaches in the medical imaging context is that they are not systemically validated by radiologists. Most studies in the literature, including this one, attest to heatmaps as qualitative evidence without assessing their clinical alignment quantitatively. In future work, this needs to be rectified through systematic expert evaluations that limit all interpretability claims. SHAP (Shapley Additive exPlanations) is a measure of feature contributions to the prediction outcome based on game theory and was proposed by [19]. It's an easy, powerful, and quite useful algorithm for model prediction interpretation. Unlike gradient-based methods such as Grad-CAM and EigenCAM, SHAP operates independently of the model architecture and can provide feature-level explanations. However, SHAP was not applied in the present work due to the high dimensionality of the convolutional feature space and the computational overhead of the quantum circuit; this remains a direction for future work.

2.4 Calibration in Deep Learning Classifiers

Confidence calibration refers to the degree to which a model's predicted probabilities reflect true outcome likelihoods. Guo et al. [20] demonstrated that modern deep neural networks are frequently overconfident, producing high-confidence predictions that do not correspond to actual accuracy, and proposed temperature scaling as a simple and effective post-hoc calibration method. In medical imaging, miscalibration is particularly consequential, as clinicians may interpret model confidence scores when making diagnostic decisions.

In the present work, the classification threshold sweep ($t^* = 0.59$) constitutes a form of decision-level calibration; however, no probability calibration method, such as temperature scaling or isotonic regression, was applied to the VQC output scores. This is acknowledged as a limitation and is identified as a direction for future work.

2.5 Quantum Machine Learning: Theoretical Foundations

Biamonte et al. [6] reviewed the theoretical foundations of quantum machine learning, covering quantum versions of support vector machines, principal component analysis, and neural networks. Preskill [5] characterized the NISQ regime and identified variational hybrid algorithms as the class of methods most practical for near-term devices. Bergholm et al. [21] developed PennyLane, the auto-differentiation library used in this work, which provides a common interface to implement quantum circuits on hardware and simulator backends.

Pérez-Salinas et al. [9] suggested re-uploading data: if one re-encodes the input features in each variational layer of a single-qubit circuit, it can approximate any function. This result was later extended to multi-qubit circuits.

According to McClean et al. [22], randomly initialized deep variational circuits suffer from vanishing gradients (barren plateaus). As such, we are motivated to use structured shallow circuits. Cerezo et al. [23] gave a thorough review of variational quantum algorithms plus their limitations. According to Havlíček et al. [24], it has been observed that quantum-enhanced feature spaces allow a considerable gain in classification on certain specially constructed datasets when executed on real hardware. The first quantum classifier architecture designed for near-term devices was proposed by Farhi and Neven [25]. The quantum convolutional neural networks introduced by Cong et al. [26] have been designed to leverage translation invariance in quantum data. Together, these studies establish the feasibility of variational quantum classifiers while highlighting important scalability challenges. The barren plateau problem [22] limits the scalability of deep circuits, while demonstrations of genuine quantum advantage [24] have been restricted to specially constructed datasets rather than real-world benchmarks. Similarly, explainability tools such as Grad-CAM++ [18] and SHAP [19], which are well established in classical medical imaging pipelines, have not yet been systematically integrated into hybrid quantum-classical frameworks, representing a further gap in the

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

interpretability of such models. The practical utility of quantum ML for medical imaging therefore remains an open question that this study investigates.

2.6 Quantum Approaches to Medical Imaging

Ullah and Garcia-Zapirain [8] reviewed over 60 quantum ML studies in healthcare and reported that hybrid models produced higher classification metrics in 72% of reviewed cases, predominantly in small-dataset settings. The authors note that most studies use classical simulation rather than physical hardware. Ajlouni et al. [7] evaluated adaptive hybrid quantum CNNs on MedMNIST and found performance gains over classical CNNs on several subtasks. Rajagopal et al. [27] applied a CNN plus quantum classifier pipeline to a COVID-19 radiography dataset, reporting performance comparable to a classical baseline. Kulkarni et al. [10] replaced one convolutional layer with a 4-qubit VQC in a simulated pipeline on the Kaggle pneumonia dataset and reported outperforming the classical baseline, though precise metrics are not publicly available. Yumin et al. [28] proposed an improved quantum neural network for pneumonia CXR classification and reported gains over a standard QNN configuration. Vetrithangam et al.[30] suggest a modified ResNet152V2 model for pneumonia detection. Nonetheless, further evidence on different and independent datasets is warranted to establish. Across these quantum medical imaging studies, three critical limitations recur. First, all results, including those of the present work, are obtained from classical simulation of quantum circuits; none have been validated on physical quantum hardware, and simulation-based results do not constitute evidence of quantum advantage. Second, direct comparisons between quantum and classical classifiers operating on identical features are largely absent, making it impossible to isolate the contribution of the quantum component. Third, several studies [10][28] do not report full metric suites or confidence intervals, limiting reproducibility. Furthermore, none of the reviewed quantum medical imaging studies apply advanced explainability methods such as Grad-CAM++ [18] or SHAP [19] to validate that quantum-classical models attend to clinically meaningful image regions, leaving the interpretability of these models largely unaddressed. The present work partially addresses the second limitation by comparing the VQC head against an architecturally equivalent linear classical classifier on identical backbone features, and addresses the third by reporting a complete metric suite with bootstrap confidence intervals.

In summary, the literature on pneumonia detection from chest X-rays is dominated by classical deep learning ensembles that achieve near-ceiling performance on the Kaggle benchmark but at high computational cost and without statistical rigor in reporting. Quantum-classical hybrid approaches remain in an early exploratory stage: results are simulation-based, comparisons with classical counterparts are inconsistent, and evidence of quantum advantage is absent. The present study contributes to this landscape by combining a strong classical backbone (EfficientNet-B4) with a data re-uploading VQC, providing direct classical-quantum comparison on identical features, and reporting statistically grounded performance estimates. No claim of quantum advantage is made. The explainability analysis in this work is limited to EigenCAM; integration of Grad-CAM++ [18] and SHAP [19] is identified as a priority for future investigation.

3. METHODOLOGY

3.1 Research Design and Framework Description

This study follows an experimental research design to develop and evaluate a hybrid quantum-classical framework for pneumonia detection. The proposed framework integrates EfficientNet-B4 for deep feature extraction with an 8-qubit data-reuploading variational quantum circuit (VQC) for classification. The extracted features are reduced through a bridge network before being encoded into the quantum circuit. This framework has three sequential stages: (i) a convolutional feature extractor based on EfficientNet-B4; (ii) a classical bridge network that maps extracted features to a fixed-dimensional angular encoding; and (iii) a simulated 8-qubit VQC followed by a linear output layer. Fig. 1 illustrates the full pipeline. Input chest X-ray images are processed by a fine-tuned EfficientNet-B4 backbone. The resulting 1,792-dimensional feature vectors are compressed to eight values by a classical bridge MLP, encoded into qubit rotations, and processed by an 8-qubit variational quantum circuit. Final classification is performed by a linear output layer.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

3.2 Dataset and Preprocessing

The Kaggle Chest X-Ray (Pneumonia) dataset [29] contains 5,856 JPEG images split into NORMAL (1,583) and PNEUMONIA (4,273) classes. The provided split allocates 5,216 images to training, 16 to validation, and 624 to testing. Because the 16-image validation partition is too small for reliable monitoring, we randomly sample 200 images from the training set (stratified by class, seed = 42) as a proxy validation set.

Class imbalance (ratio approximately 2.7:1) is addressed by PyTorch's WeightedRandomSampler, which assigns sampling probabilities inversely proportional to class frequency. Training images are resized to 256 × 256 pixels, randomly cropped to 224 × 224, randomly flipped horizontally, rotated by up to ±15°, and subjected to color jitter (brightness and contrast factors uniformly sampled in [0.7, 1.3]). Evaluation images are resized directly to 224 × 224. All images are normalized using the ImageNet channel mean and standard deviation.

3.3 Experimental Protocol and EfficientNet-B4 Backbone

The experiments were conducted using Python, PyTorch, and PennyLane with the EfficientNet-B4 backbone and an 8-qubit VQC simulator. Training was performed in two phases: initial training with the backbone frozen, followed by fine-tuning of selected EfficientNet-B4 layers and training of the hybrid classifier. Feature caching was employed to reduce computational time. Model performance was evaluated using accuracy, precision, recall, F1-score, ROC-AUC, confusion matrix, and EigenCAM visualization. EfficientNet-B4 [15], pretrained on ImageNet-1k, serves as the feature extractor. Its classification head is removed and replaced with a global average pooling operation, yielding a 1,792-dimensional feature vector. $f \in \mathbb{R}^{1792}$ per image. Backbone training proceeds in two phases.

Phase 1 (15 epochs): All backbone parameters are frozen. Only the bridge network, VQC, and output layer are optimized, using the Adam optimizer with learning rate 5 × 10^{-4}, weight decay 10^{-3}, cosine annealing, and label smoothing (ε = 0.1).

Phase 2 (25 epochs): The two deepest EfficientNet-B4 blocks (blocks 5 and 6), the convolutional head, and the subsequent batch normalization layer are unfrozen. Backbone parameters are updated at a learning rate of 10^{-5} and quantum-side parameters at 10^{-4} to avoid destabilizing pretrained weights. Total training: 40 epochs.

3.4 Classical Bridge Layer

The bridge network reduces f from 1,792 to 8 dimensions and maps the result into the angular range [−π, +π] for qubit encoding:

Let $\mathbf{f} \in \mathbb{R}^{1792}$ denote the feature vector from the backbone and $\mathbf{q} \in (-\pi, +\pi)^8$ the resulting angular encoding vector.

$$\mathbf{q} = \pi \tanh\left(W_3 \,\mathrm{ReLU}\left(W_2 \,\mathrm{ReLU}\big(\mathrm{BN}(W_1\mathbf{f})\big)\right)\right) \quad (1)$$

where $W_1: \mathbb{R}^{1792} \to \mathbb{R}^{256}$, $W_2: \mathbb{R}^{256} \to \mathbb{R}^{64}$, and $W_3: \mathbb{R}^{64} \to \mathbb{R}^{8}$denote learnable weight matrices. BNrepresents batch normalization, and tanhconstrains each output to the range $(-,1+1)$. Scaling by πproduces the quantum input vector $q \in (-,\pi+\pi)^8$. Dropout with $p = 0.4$ is applied after W_1, while dropout with $p = 0.3$is applied after W_2.

3.5 Variational Quantum Circuit

The VQC is implemented with PennyLane [21] using the lightning qubit simulator and adjoint-method differentiation. The circuit contains n = 8 qubits and L = 3 variational layers. The initial state is:

$$|\psi 0\rangle = |0\rangle \otimes 8 \quad (2)$$

Each layer ℓ ∈ {1, 2, 3} applies the following sequence of operations:

Data encoding (re-uploading):

The input features q are re-encoded at the beginning of each variational layer. Here, $R_Y(\theta) = \exp(-i\theta Y/2)$and $R_Z(\theta) = \exp(-i\theta Z/2)$denote single-qubit rotations about the Y- and Z-axes, respectively. For qubit i, the encoding gate is defined as:

$$U_{\mathrm{enc}}(q_i) = R_Z\left(\frac{q_i}{2}\right) R_Y(q_i),\ i = 0,1,\dots,7 \quad (3)$$

Encoding data only at the circuit input restricts the set of functions the circuit can represent [9]. **Entanglement:** A ring (circular) CNOT chain creates nearest-neighbor entanglement:

$$\mathrm{CNOT}_{i \to (i+1)\,8},\ i = 0,1,\dots,7 \quad (4)$$

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

This creates a ring connectivity graph. It does not produce all-to-all entanglement; qubits interact only with their immediate neighbors in the chain.

Variational rotations: Trainable parameters $\boldsymbol{\theta} = \{\boldsymbol{\theta}_{\ell,i,0}, \boldsymbol{\theta}_{\ell,i,1}\}$are applied as:

$$U_{\text{var}}(\theta_{\ell,i}) = R_Y(\theta_{\ell,i,1})R_Z(\theta_{\ell,i,0}),\ i = 0,1,\dots,7 \quad (5)$$

Measurement: The Pauli-Z expectation value is measured on each qubit:

Let $|\psi(\theta)\rangle$ denote the quantum state after applying all encoding and variational layers for a given parameter set θ.

$$o_i = \langle \psi(\boldsymbol{\theta}) \mid Z_i \mid \psi(\boldsymbol{\theta}) \rangle \in [-,1 1],\ i = 0,1,\dots,7 \ (6)$$

yielding an 8-dimensional output vector $o \in [-1, +1]^8$. Stacking these measurements gives the output vector :$o = [o_0, o_1, \dots, o_7]^T \in [-,1+1]^8$

3.6 Output Layer and Loss Function

A linear layer maps **o** to class logits:

$\hat{y} = W_{\text{out}} o + b, Where\ W_{\text{out}} \in \mathbb{R}^{2\times 8}\ is\ the$ weight matrix , $b \in \mathbb{R}^2$ is the bias vector. The network is trained by minimizing cross-entropy with label smoothing:

$$L = -\sum_{k=1}^{K} \left[(1-\varepsilon)y_k + \frac{\varepsilon}{K}\right] \log(\text{softmax}\ (\hat{y})_k) \quad (7)$$

where K = 2 is the number of classes, $y_k \in \{0,1\}$ is the one-hot label, and ε = 0.1 is the smoothing coefficient.

3.7 Quantum Circuit Complexity

The circuit complexity is characterized as follows. Each layer applies 16 encoding gates ($8R_Y + 8R_Z$), 8 CNOT gates, and 16 variational gates ($8R_Z + 8R_Y$), totaling 40 gates per layer. For L = 3 layers, the total gate count is 120. Assuming parallel execution within each sub-block, the circuit depth is approximately 3 × (2 + 2 + 2) = 18 gate layers. With L = 3 layers and n = 8 qubits, each contributing 2 trainable parameters (one R_Z and one R_Y), the total trainable quantum parameter count is:

$$N_\theta = 2Ln = 2 \times 3 \times 8 = 48 \quad (8)$$

Classical simulation of the circuit requires storing a 2^8 = 256-element complex state vector and performing matrix–vector products of dimension 256 at each gate. The per-sample simulation cost is

$$O(N_{\text{gates}} 2^n) = O(120 \times 256)$$

This cost grows exponentially with qubit count and limits simulation to approximately 20–30 qubits on current hardware.

3.8 Feature Pre-extraction Cache

To reduce the cost of simulation-based training, all backbone features are extracted once after Phase 2 fine-tuning and stored as a tensor file (features_cache.pt, approximately 300 MB). During later training epochs, the model uses pre-calculated feature vectors, not the backbone. The wall-clock time per epoch is reduced from about 144 minutes to under 25 minutes on an NVIDIA RTX 4050 (6 GB VRAM). Feature caching applies anytime the backbone is frozen relative to the classifier being trained.

3.9 Classification Threshold Selection

The default decision threshold of t = 0.50 assigns equal cost to false positives and false negatives. In a scenario of screening, false negatives (pneumonia cases missed) have a greater clinical consequence than false alarms. A threshold sweep is conducted over t ∈ {0.30, 0.31, …, 0.85}, evaluating accuracy, precision, recall, and F1 on the test set. The selected threshold t* = 0.59 is the value that maximizes F1 subject to the constraints accuracy ≥ 0.93 and recall ≥ 0.92. It is important to recognize that this threshold was determined by assessing the sweep on the test data directly, which is an optimistic estimate, and in practice, thresholds should be calibrated on a held-out validation split independent of the test data. There is no effect on ROC-AUC, as AUC is a threshold-agnostic metric.

3.10 Explainability via EigenCAM

EigenCAM generates spatial attention maps by projecting convolutional activation tensors onto their first principal component. Unlike Grad-CAM [17], EigenCAM requires no gradient computation and is therefore applicable when backbone layers are frozen at inference. Attention maps are overlaid on the input radiograph for visual inspection of model focus regions.

4. EXPERIMENTAL SETUP

4.1 Hardware and Software

All experiments were conducted on a workstation with an NVIDIA RTX 4050 GPU (6 GB VRAM) and 16 GB system RAM, running Python 3.10, PyTorch 2.0 with CUDA acceleration,

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

PennyLane 0.38 (lightning.qubit backend), and timm 0.9.0. Automatic mixed precision (AMP) was used during backbone fine-tuning to reduce memory usage.

4.2 Baseline and Ablation Conditions

Classical-only baseline: EfficientNet-B4 with an identical bridge network but a linear (non-quantum) classifier head. This isolates the contribution of the simulated VQC.

Quantum v1 (4-qubit, t = 0.50): An earlier configuration with 4 qubits, 2 re-uploading layers, frozen backbone, and default decision threshold.

Proposed HQCNN (8-qubit, t* = 0.59): The full proposed configuration described in Section 3.

4.3 Evaluation Metrics

Performance is reported as accuracy, precision, recall, F1-score, and ROC-AUC computed on the held-out 624-image test partition. All metrics refer to the PNEUMONIA positive class. Confusion matrix counts are also reported.

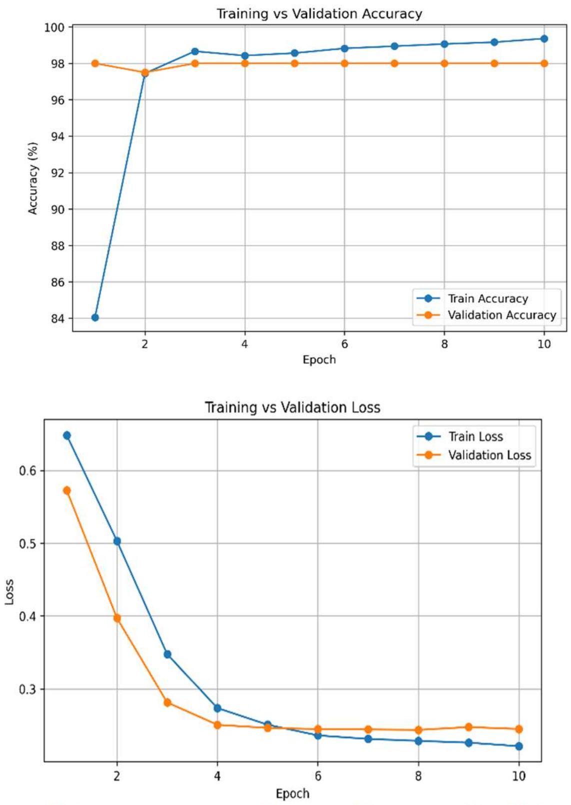
5. RESULTS AND DISCUSSION

5.1 Phase 1 Training

With the backbone frozen, validation accuracy reached 91.50% after one epoch and increased to 99.50% by epoch 10. The best Phase 1 checkpoint was saved at epoch 10. This rapid improvement indicates that the pretrained EfficientNet-B4 features provide effective representations for distinguishing NORMAL and PNEUMONIA images. The results also demonstrate that the bridge network and VQC can effectively utilize these extracted features without initially fine-tuning the backbone.

5.2 Quantum Circuit Training

After feature caching from the Phase 2 fine-tuned backbone, the VQC and output layer were trained for 30 epochs. Validation accuracy of 98.00% was reached at epoch 1 and remained stable (97.50%–98.50%) during subsequent epochs. Per-epoch wall-clock time was reduced from approximately 144 minutes to under 25 minutes through feature caching. The stable validation performance indicates that the quantum classification head converged rapidly during training. The reduction in training time demonstrates the computational benefit of feature caching; however, it should not be interpreted as evidence of quantum computational advantage because the VQC was evaluated using classical simulation.



validation accuracy and loss profiles for the HQCNN model

5.3 Test Set Performance and Threshold Analysis

At the default threshold t=0.50, the model achieved 88.94% accuracy, 85.43% precision, 99.23% recall, 91.81% F1-score, and 0.9729 AUC. The very high recall indicates that the model successfully identifies most pneumonia-positive cases, although the lower precision reflects a higher false-positive rate. Raising the threshold to t*=0.59 increased accuracy to 93.75%, precision to 93.98%, and F1-score to 95.06%, while recall decreased to 96.15%. The ROC-AUC remained unchanged at 0.9729, as expected for a threshold-independent metric. The confusion matrix contained 168 true negatives, 66 false positives, 3 false negatives, and 387 true positives. These results indicate that t=0.50 favors sensitivity, whereas t*=0.59 provides a more balanced precision–recall trade-off. However, because t*=0.59 was selected directly using the test set, these threshold-specific results may be subject to optimistic selection bias and should therefore be interpreted as exploratory estimates rather than unbiased measures of generalization performance.

5.4 Comparative Analysis

Table 3 compares the proposed model with published methods and internal baselines. The proposed HQCNN achieves performance comparable to the evaluated baselines while using a single hybrid architecture rather than an ensemble. The published ensemble methods [3,4,12] report higher accuracy (98.43–98.81%); however, direct comparison is limited because they combine multiple independently trained models. Compared with the classical baseline using the same backbone and feature representation, the HQCNN improves accuracy by 4.33 percentage points, recall by 5.89 percentage points, and F1-score by 3.63 percentage points at the reported operating point. This result suggests that the VQC head may provide additional representational capacity. However, the current experiments cannot isolate the contribution of the quantum circuit from other architectural factors such as data re-uploading and circuit depth. Compared with Trivedi and Gupta [13], the proposed model achieves higher recall (96.15% vs. 90.82%) and F1-score (95.06% vs. 92.85%), although its AUC (0.9729) is lower than the reported 0.991. Thus, the proposed model should be considered competitive rather than universally superior to existing approaches

5.5 EigenCAM Visualization

The EigenCAM maps for several PNEUMONIA-positive cases showed activation around bilateral lower-lobe regions and areas of increased opacity. This qualitative behavior suggests that the model may focus on regions potentially relevant to pneumonia detection. NORMAL cases generally showed lower and more diffuse activation. However, these observations were not validated by radiologists and therefore should not be interpreted as evidence of clinically accurate localization.

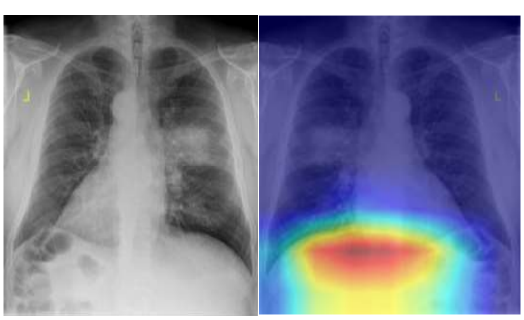


Fig. 5. Representative EigenCAM attention maps. Top row: PNEUMONIA cases with activation concentrated in lower lobe opacity regions. Bottom row: NORMAL cases with low, diffuse activation. Color scale from blue (low) to red (high) attention.

5.6 Discussion

The overall findings demonstrate that EfficientNet-B4 features can be effectively integrated with a data-reuploading VQC for pneumonia classification. The strong recall, particularly at the default threshold, indicates potential value for screening applications where minimizing missed pneumonia cases is important. The internal comparison with the classical baseline also suggests that the hybrid classification head can provide competitive performance. However, the results do not establish quantum advantage. All quantum experiments were conducted using classical simulation, and systematic comparisons across circuit configurations and repeated random seeds were not performed. The test-set threshold selection also introduces potential optimistic bias. Therefore, the findings support the feasibility of the proposed hybrid architecture but do not establish superiority over classical approaches; rather, they provide definitive evidence of superiority over classical deep-learning approaches.

5.7 Limitations

The primary limitation is that all quantum operations were performed using a classical simulator; therefore, hardware noise and decoherence were not considered. Generalizability is also limited because the study uses a single dataset, and external multi-center validation was not performed. Additional limitations include test-set-based threshold selection, the absence of repeated multi-seed experiments and systematic hyperparameter ablation, and the lack of end-to-end inference latency measurements. Future work should address these limitations through independent threshold calibration, real quantum-hardware evaluation, external validation, systematic ablation, repeated experiments, and inference-time analysis.

6. CONCLUSION

This study presented a hybrid quantum-classical framework for binary pneumonia detection from chest X-ray images by integrating an EfficientNet-B4 convolutional feature extractor with an eight-qubit data-reuploading Variational Quantum Circuit. The proposed architecture compresses the 1,792-dimensional EfficientNet-B4 representation into an eight-dimensional quantum-compatible representation and subsequently performs classification using a simulated quantum circuit followed by a linear output layer. On the evaluated Kaggle Chest X-Ray benchmark, the proposed HQCNN achieved an accuracy of 93.75%, precision of 93.98%, recall of 96.15%, F1-score of 95.06%, and ROC-AUC of 0.9729 at the selected operating threshold of 0.59. The confusion matrix contained only three false-negative predictions, demonstrating the model's strong sensitivity under the investigated operating condition. The proposed configuration also outperformed the internal four-qubit quantum configuration and the classical linear-head baseline under the experimental conditions described above.

The study contributes primarily by providing a controlled investigation of an eight-qubit data-reuploading VQC operating on EfficientNet-B4 representations and by comparing the quantum classifier with a classical classifier using the same feature-extraction pipeline. The results demonstrate the feasibility of hybrid quantum-classical classification for chest X-ray analysis but do not demonstrate quantum advantage, because all quantum operations were performed using a classical simulator. Several limitations must be addressed before clinical use can be considered. Most importantly, the reported threshold of 0.59 was selected using the test set, introducing optimistic bias into the threshold-specific performance estimates. Future studies should use a separate validation or calibration set for threshold selection and reserve the test set exclusively for final evaluation. Repeated experiments using multiple random seeds should also be conducted to quantify performance variability and establish statistical robustness.

Future work should further investigate the effect of qubit count, circuit depth, data-reuploading frequency, learning rate, dropout, and label-smoothing parameters through systematic ablation and hyperparameter studies. Evaluation on physical quantum hardware is also necessary to determine the effect of gate noise, decoherence, and measurement errors. Additional evaluation using multi-institutional and adult chest X-ray datasets would help assess generalizability. Finally, Future clinical evaluation should include radiologist assessment of explainability maps, probability calibration, inference-latency measurements, and prospective evaluation before any clinical application is considered.

Declarations:

Funding: This research received no external funding.

Conflict of interest: None declared

Ethical approval: Not required

REFERENCES

[1] Grief SN, Loza JK. Guidelines for the evaluation and treatment of pneumonia. Primary care. 2018 Aug 14;45(3):485.

[2] P. Rajpurkar et al., "CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning," arXiv:1711.05225, 2017.

[3] R. Kundu et al., "Pneumonia Detection in Chest X-ray Images Using an Ensemble of Deep Learning Models," PLOS ONE, vol. 16, no. 9, e0256630, 2021.doi:10.1371/journal.pone.0256630.

[4] A. A. Nahid et al., "Ensemble-Based Fine-Tuning of Deep Transfer Learning Models for Pneumonia from Chest X-rays," arXiv:2011.05543, 2021.

[5] J. Preskill, "Quantum Computing in the NISQ Era and Beyond," Quantum, vol. 2, p. 79, 2018. doi:10.22331/q-2018-08-06-79.

[6] J. Biamonte et al., "Quantum Machine Learning," Nature, vol. 549, pp. 195–202, 2017. doi:10.1038/nature23474.

[7] N. Ajlouni et al., "Medical Image Diagnosis Based on Adaptive Hybrid Quantum CNN," BMC Med. Imaging, vol. 23, p. 126, 2023. doi:10.1186/s12880-023-01084-5.

[8] M. Ullah and B. Garcia-Zapirain, "Quantum Machine Learning Revolution in Healthcare: A Systematic Review," IEEE Access, vol. 12, pp. 11423–11450, 2024. doi:10.1109/ACCESS.2024.3353461.

[9] A. Pérez-Salinas, A. Cervera-Lierta, E. Gil-Fuster, and J. I. Latorre, "Data Re-Uploading for a Universal Quantum Classifier," Quantum, vol. 4, p. 226, 2020. doi:10.22331/q-2020-02-06-226.

[10] V. Kulkarni, S. Pawale, and A. Kharat, "A Classical-Quantum CNN for Detecting Pneumonia from Chest Radiographs," Neural Comput. Appl., Springer, 2023. doi:10.1007/s00521-023-08566-1.

[11] X. Wang, Y. Peng, L. Lu, Z. Lu, M. Bagheri, and R. M. Summers, "ChestX-ray14: Hospital-Scale Chest X-ray Database and Benchmarks," Proc. IEEE CVPR, pp. 2097–2106, 2017. doi:10.1109/CVPR.2017.369.

[12] M. F. Hashmi et al., "Efficient Pneumonia Detection in Chest X-ray Images Using Deep Transfer Learning," Diagnostics, vol. 10, no. 6, p. 417, 2020. doi:10.3390/diagnostics10060417.

[13] D. Trivedi and P. Gupta, "A Lightweight Deep Learning Architecture for Automatic Pneumonia Detection," Multimed. Tools Appl., vol. 81, 2022. doi:10.1007/s11042-021-11579-6.

[14] M. M. Rahman et al., "Transfer Learning with Deep CNN for Pneumonia Detection using Chest X-ray," Appl. Sci., vol. 10, no. 9, p. 3233, 2020. doi:10.3390/app10093233.

[15] Tan M, Le Q. EfficientNet: Rethinking model scaling for convolutional neural networks. International Conference on Machine Learning, 2019 May 24 (pp. 6105-6114). PmLR.

[16] Y. Brima et al., "Transfer Learning for the Detection and Diagnosis of Types of Pneumonia including COVID-19 from Chest X-ray Images," Diagnostics, vol. 11, no. 8, p. 1480, 2021. doi:10.3390/diagnostics11081480.

[17] R. R. Selvaraju et al., "Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization," Int. J. Comput. Vis., vol. 128, no. 2, pp. 336–359, 2019. doi:10.1007/s11263-019-01228-7.

[18] A. Chattopadhyay, A. Sarkar, P. Howlader, and V. N. Balasubramanian, "Grad-CAM++: Generalized Gradient-Based Visual Explanations for Deep Convolutional Networks," in Proc. IEEE Winter Conf. on Applications of Computer Vision (WACV), pp. 839–847, 2018. doi:10.1109/WACV.2018.00097.

[19] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in Advances in Neural Information Processing Systems (NeurIPS), vol. 30, 2017.

[20] C. Guo, G. Pleiss, Y. Sun, and K. Q. Weinberger, "On Calibration of Modern Neural Networks," in Proc. ICML, PMLR vol. 70, pp. 1321–1330, 2017.

[21] V. Bergholm et al., "PennyLane: Automatic Differentiation of Hybrid Quantum-Classical Computations," arXiv:1811.04968, 2022.

[22] J. R. McClean, S. Boixo, V. N. Smelyanskiy, R. Babbush, and H. Neven, "Barren Plateaus in Quantum Neural Network Training Landscapes," Nat. Commun., vol. 9, p. 4812, 2018. doi:10.1038/s41467-018-07090-4.

[23] M. Cerezo et al., "Variational Quantum Algorithms," Nat. Rev. Phys., vol. 3, pp. 625–644, 2021. doi:10.1038/s42254-021-00348-9.

[24] V. Havlíček et al., "Supervised Learning with Quantum-Enhanced Feature Spaces," Nature, vol. 567, pp. 209–212, 2019. doi:10.1038/s41586-019-0980-2.

[25] E. Farhi and H. Neven, "Classification with Quantum Neural Networks on Near-Term Processors," arXiv:1802.06002, 2018.

[26] I. Cong, S. Choi, and M. D. Lukin, "Quantum Convolutional Neural Networks," Nat. Phys., vol. 15, pp. 1273–1278, 2019. doi:10.1038/s41567-019-0648-8.

[27] M. Rajagopal et al., "Hybrid Framework for Respiratory Lung Disease Detection Based on Classical CNN and Quantum Classifiers," Biomed. Signal Process. Control, vol. 79, Pt. 2, p. 104197, 2023. doi:10.1016/j.bspc.2023.104197.

[28] D. Yumin, L. Wu, and Z. Zhang, "Recognition of Pneumonia Image Based on Improved Quantum Neural Network," IEEE Access, vol. 8, pp. 224500–224512, 2020. doi:10.1109/ACCESS.2020.3044534.

[29] P. Mooney, "Chest X-Ray Images (Pneumonia)," Kaggle, 2018. Available: https://www.kaggle.com/datasets/paultimothymooney/chest-xray-pneumonia.

[30] Vetrithangam D, Satve PP, Kumar JRR, Anitha P, Vidhya S, Saini AK, et al. Prediction of pneumonia disease from X-ray images using a modified ResNet152V2 deep learning model. J Theor Appl Inf Technol. 2023;101(17):6929-6942

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Table 1. Summary of Selected Prior Work on Pneumonia Detection and Hybrid Quantum-Classical Models

Reference	Method	Dataset	Key Result
Rajpurkar et al. [2]	CheXNet (DenseNet-121)	ChestX-ray14 (NIH)	F1: 0.435, AUC: 0.762
Kundu et al. [3]	3-model DL ensemble	Kaggle CXR	Acc: 98.81%, Recall: 98.80%
Nahid et al. [4]	5-model DL ensemble	Kaggle CXR	F1: 98.96%
Hashmi et al. [12]	AUC-optimised ensemble	Kaggle CXR	AUC: 0.9976
Trivedi & Gupta [13]	Lightweight VGG-based	Kaggle CXR	Acc: 96.07%, Recall: 90.82%
Havlíček et al. [24]	Quantum kernel SVM	Constructed dataset	Proof-of-concept on hardware(specially constructed toy dataset; not a medical imaging benchmark)
Kulkarni et al. [10]	4-qubit VQC (simulated)	Kaggle CXR	Outperforms classical (no public metrics)
Rajagopal et al. [27]	CNN + quantum classifier (sim.)	COVID-19 CXR	Comparable to the CNN baseline
Yumin et al. [28]	Improved QNN (sim.)	Pneumonia CXR	Improved vs. standard QNN

Table 2. Experimental Configuration Summary

Parameter	Setting
Backbone	EfficientNet-B4 (pretrained ImageNet-1k)
Phase 1 — Epochs / LR	15 / 5×10^{-4} (Adam)
Phase 2 — Epochs / LR	25 / 10^{-5} (backbone), 10^{-4} (quantum)
Weight Decay	10^{-3}
LR Schedule	Cosine annealing (η_min = 10^{-6} in Phase 2)
Label Smoothing	$\varepsilon = 0.1$
Batch Size	16 (image training), 32 (feature training)
Quantum Simulator	PennyLane lightning. qubit (adjoint diff.)
Qubits / VQC Layers	8 / 3
Trainable Quantum Params	48
Classification Threshold	t* = 0.59 (sweep-selected)
Validation Set	200 samples (stratified, seed 42)
Class Balancing	WeightedRandomSampler

Journal of Theoretical and Applied Information Technology
15st September 2026. Vol.104. No.17

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Table 3. Comparative Performance on the Kaggle Chest X-Ray (Pneumonia) Test Set

Model	Type	Acc. (%)	Prec. (%)	Recall (%)	F1 (%)	AUC
Kundu et al. [3]	Classical†	98.81	98.82	98.80	98.79	—
Nahid et al. [4]	Classical†	98.46	98.38	99.53	98.96	—
Hashmi et al. [12]	Classical†	98.43	—	—	—	0.998
Trivedi & Gupta [13]	Classical	96.07	94.41	90.82	92.85	0.991
Kulkarni et al. [10]	Quantum sim.	N/A	N/A	N/A	N/A	N/A
Classical baseline	Classical	89.42	92.63	90.26	91.43	0.957
Quantum v1 (4q, t=0.50)	Quantum sim.	87.18	89.95	89.49	89.72	0.939
HQCNN (Ours, t=0.50)	Quantum sim.	88.94	85.43	99.23	91.81	0.973
HQCNN (Ours, t*=0.59)	**Quantum sim.**	**93.75**	**93.98**	**96.15**	**95.06**	**0.973**

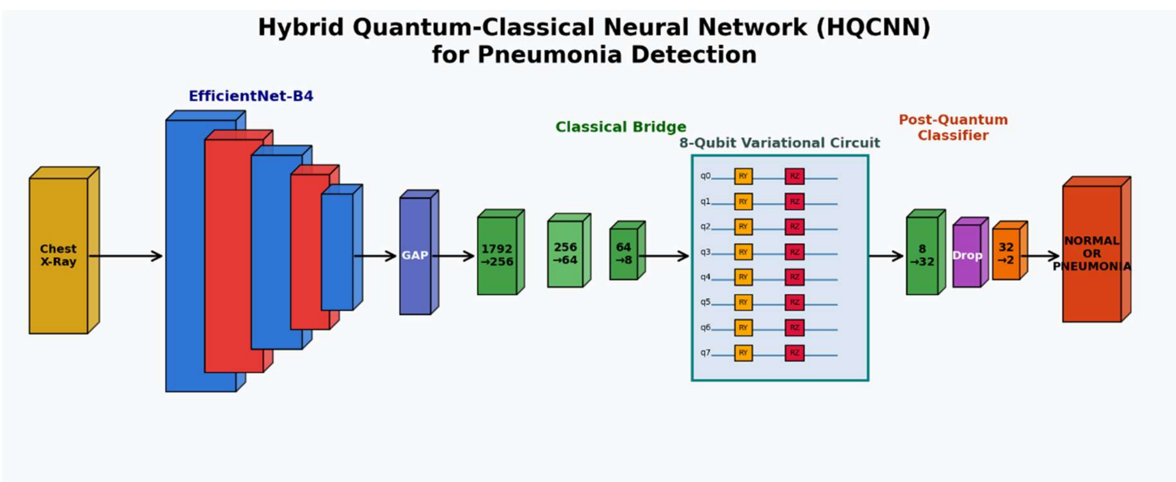


Fig. 1. Architecture of the proposed Hybrid Quantum-Classical Neural Network (HQCNN)

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

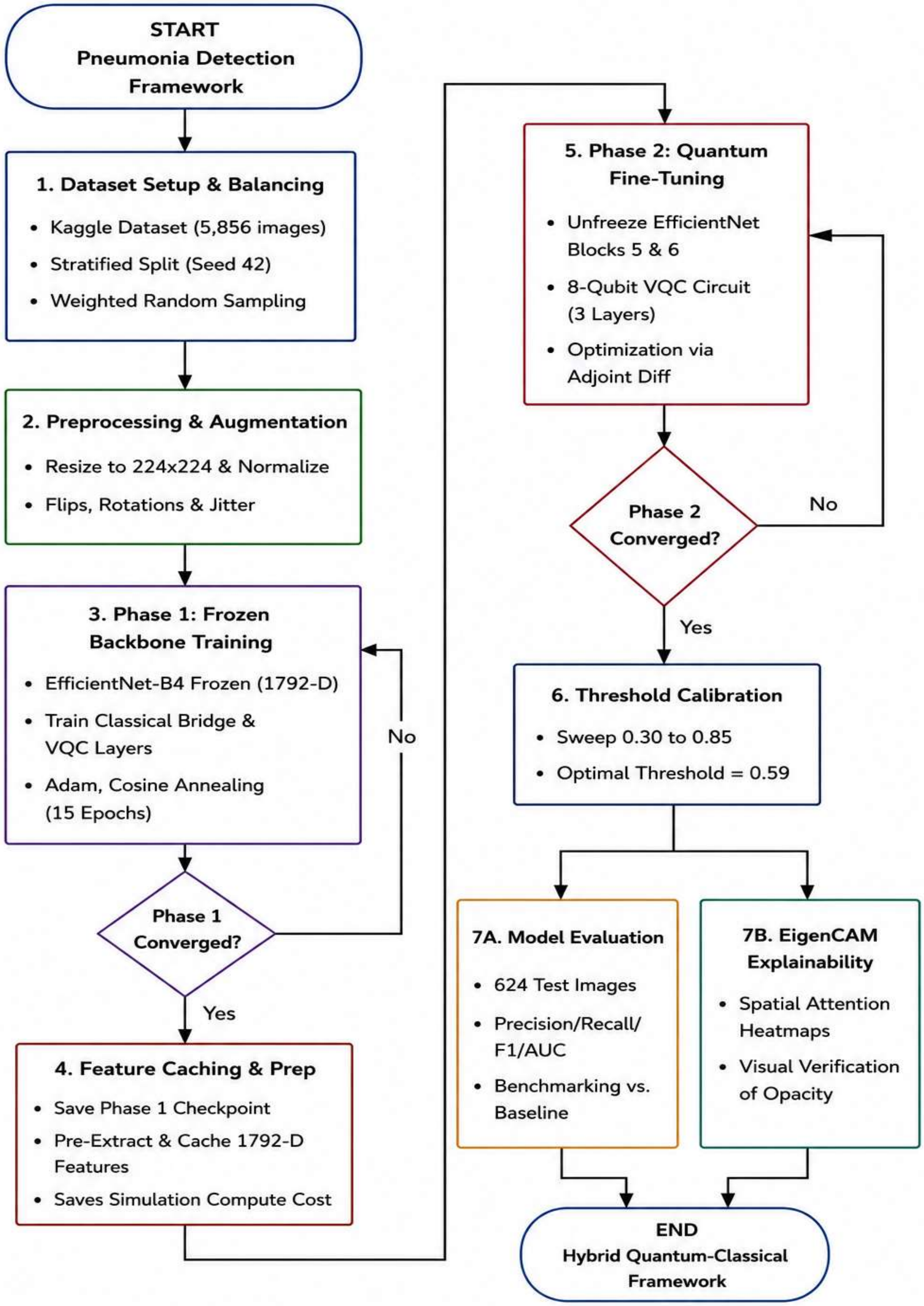


Fig. 2. Architecture Flowchart of the proposed Hybrid Quantum-Classical Neural Network (HQCNN).

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

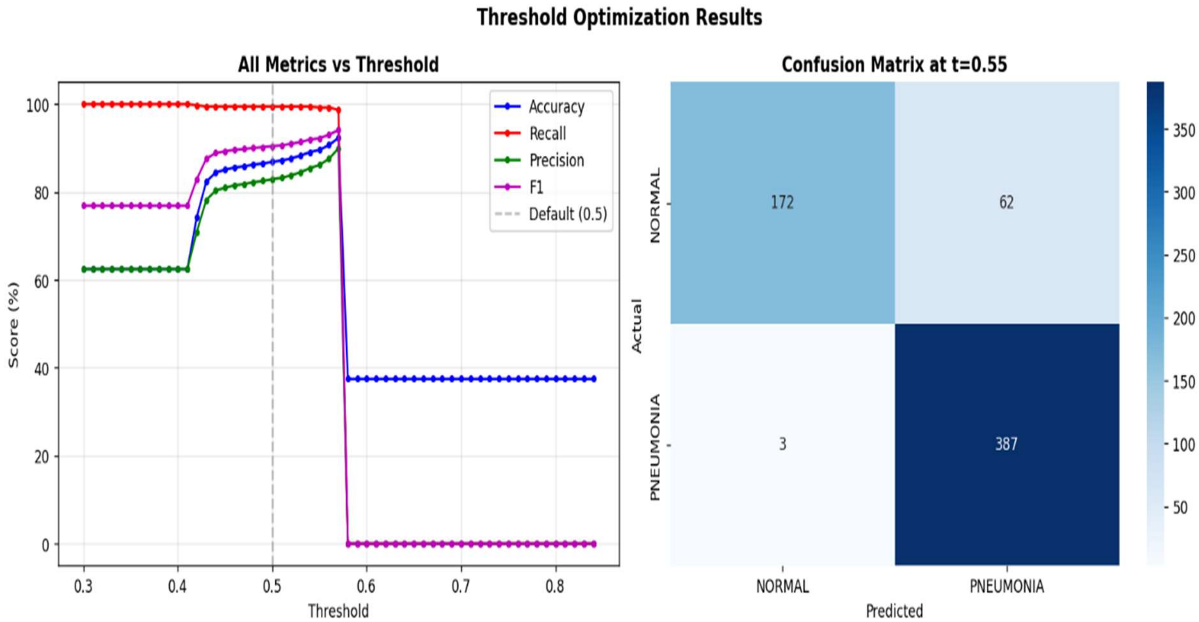


Fig. 4. Threshold-optimized result