BRAIN. Broad Research in Artificial Intelligence and Neuroscience
e-ISSN: 2067-3957 | p-ISSN: 2068-0473
Covered in: Web of Science (ESCI); EBSCO; JERIH PLUS (hkdir.no); IndexCopernicus; Google Scholar; SHERPA/RoMEO; ArticleReach Direct; WorldCat; CrossRef; Peeref; Bridge of Knowledge (mostwiedzy.pl); abcdindex.com; Editage; Ingenta Connect Publication; OALib; scite.ai; Scholar9; Scientific and Technical Information Portal; FID Move; ADVANCED SCIENCES INDEX (European Science Evaluation Centre, neredataltics.org); ivySCI; exaly.com; Journal Selector Tool (letpub.com); Citefactor.org; fatcat!; ZDB catalogue; Catalogue SUDOC (abes.fr); OpenAlex; Wikidata; The ISSN Portal; Socolar; KVK-Volltitel (kit.edu)
2026, Volume 17, Issue 3, pages: 236-257
Submitted: May 15th, 2026 | *Accepted for publication:* July 21st, 2026

An Explainable Attention-Guided Quantum Vision Transformer for Lung Disease Classification using Chest X-Ray Images

Nyni K. A.
Research Scholar, Department of ECE,
Karunya Institute of Technology & Sciences,
Coimbatore, Tamil Nadu, India; Assistant Professor,
Department of Mechatronics Engineering,
Jyothi Engineering College, Thrissur, Kerala, India.
nynibilal@gmail.com

J. Anitha*
Professor, Department of ECE,
Karunya Institute of Technology & Sciences,
Coimbatore, Tamil Nadu, India.
anithaj@karunya.edu,
https://orcid.org/0000-0001-7977-8410

Jarin T.
Department of CSE, Jyothi Engineering College,
APJ Abdul Kalam Technological University,
Kerala. India. jarint@ieee.org,
https://orcid.org/0000-0002-1974-0688

* *Corresponding author*

Abstract: *Lung diseases remain a severe world health problem, which contributes to mortality worldwide. Early and accurate diagnosis of pulmonary abnormalities using chest radiographs is essential to improve clinical management and patient outcomes. Although deep learning (DL) models have achieved considerable success in automated lung disease classification, convolutional networks involve challenges in capturing local pathological patterns and global contextual dependencies. Furthermore, growing interest in quantum machine learning has provided new paths for improving feature representation and classification efficiency. To overcome these challenges, the present study proposes a Quantum-Enhanced Vision Transformer Network (QEViT) for automated classification of lung disease from chest X-ray images (CXR). The proposed work integrates EfficientNetB4 for hierarchical feature extraction, Convolutional Block Attention Module (CBAM) for attention-guided feature refinement, Vision Transformer (ViT) for global contextual modelling, and a Variational Quantum Circuit (VQC) for quantum feature generation. The effectiveness of the proposed QEViT was evaluated on the Lung Disease Dataset and the COVID-19 Radiography Database, achieving classification accuracies of 99.8% and 99.7%, respectively. Experimental outcomes proved that integrating DL, attention mechanisms, transformer-based contextual learning, and quantum feature generation improves classification performance. Finally, explainable artificial intelligence (XAI) using Grad-CAM and Score-CAM was incorporated to provide visual interpretations of the model's predictions. The proposed QEViT network may support automated CXR image analysis for lung disease classification.*

Keywords: *lung disease classification; chest X-ray; variational quantum circuit; vision transformer; EfficientNetB4; CBAM; explainable artificial intelligence.*

How to cite: Nyni, K. A., Anitha, J., & Jarin, T. (2026). An explainable attention-guided quantum vision transformer for lung disease classification using chest X-ray images. *BRAIN. Broad Research in Artificial Intelligence and Neuroscience, 17*(3), 236–257. https://doi.org/10.70594/brain/17.3/15

1. Introduction

Lung diseases are considered one of the major causes of death worldwide, posing a major public health challenge (Hroub et al., 2024). According to the World Health Organisation (WHO), chronic respiratory diseases and lung cancer contribute to millions of deaths yearly. The recent statistics related to global cancer incidences show that 2.5 million cases of lung cancer were diagnosed in 2022 (Bray et al., 2024). Early and accurate diagnosis of lung disease is essential for improving patient survival rates, which minimises treatment costs and enables proper medical intervention. Medical imaging modalities such as chest X-rays (CXR) and computed tomography (CT) are used to detect lung disease (Askr et al., 2025). Computer-Aided Diagnosis (CAD) systems are applied widely nowadays to aid in the diagnosis of lung disorders (Joyson et al., 2026).

Recently, artificial intelligence (AI) models, such as deep learning (DL) models, have shown remarkable achievements in clinical image analysis. Convolutional Neural Networks (CNNs) and transformer approaches are utilised in automated lung disease detection and classification tasks (Yi et al., 2025). These models capture discriminative features from large-scale imaging data and reduce the dependency on manual feature engineering (Alghadhban et al., 2025). Several studies have reported encouraging performance in identifying lung cancer, pneumonia, tuberculosis, and other respiratory diseases. Generally, existing DL models require substantial computational resources and may fail to capture complex feature interactions in high-dimensional medical data (Al-Qaness et al., 2024).

To address the limitations of machine learning (ML) and DL models, researchers have extensively examined quantum-computing-based solutions (Alshmrani et al., 2023). The principles used in Quantum computing are superposition and entanglement, which process information in various ways from conventional computing systems (Xie et al., 2026). The integration of quantum ML and DL with healthcare applications has garnered significant attention because of its potential to capture complex data distributions and enhance learning efficiency (Poloju & Rajaram, 2025). Recent studies have shown that hybrid quantum-classical models can improve feature representation and classification performance in different medical imaging tasks, which creates paths for smart disease diagnosis (Iqbal et al., 2022).

Despite the superior performance achieved by advanced AI systems, their black-box nature often restricts medical acceptance and reliance. Medical experts need transparent and interpretable decision-making analysis to validate automated predictions and support diagnostic decisions (Iqbal et al., 2022). Explainable Artificial Intelligence (XAI) models have emerged as efficient tools for visualising model behaviour and highlighting important regions clinically in medical images. By improving interpretability and reliability, XAI can aid the practical deployment of intelligent diagnostic systems in healthcare environments (Hiremath et al., 2026). Hence, making a precise, efficient, and XAI-based model for lung disease classification is a main study direction in medical image analysis.

In contrast to previous CNN-based, Transformer-based, or hybrid quantum-classical models, QEViT integrates EfficientNetB4 to capture local features, CBAM to refine features through attention mechanisms, ViT to capture global contextual information, and VQC to represent quantum-enhanced features within a unified framework. Moreover, Grad-CAM and Score-CAM mechanisms have been applied in the model to enhance the explainability of the decision-making process. This integrated design enables effective learning of complementary local, global, and quantum features for multiclass lung disease classification using chest X-ray images. The foremost contributions are:

- To develop a QEViT network that integrates EfficientNetB4, CBAM, ViT, and VQC for automated classification of lung disease from CXR images.
- To enhance pulmonary feature representation through the combination of local feature extraction, attention-guided feature refinement, and global contextual learning, thereby improving the identification of pathological abnormalities.

- To incorporate VQC-based quantum feature generation to capture complex non-linear interactions and feature representations.

The remainder of the paper is organised as follows. Section 2 reviews recent studies on lung disease classification. Section 3 presents the proposed QEViT model. Section 4 reports the experimental results and provides a comparative performance analysis. Finally, Section 5 concludes the paper.

2. Related Works

An XAI-based model was presented for lung adenocarcinoma classification (Lakshmi et al., 2026). To address data imbalance and improve model generalisation, image augmentation was performed using a Medical Generative Adversarial Network (MGAN). For classification, an Optimum NAS-MoE model was developed, where a Mixture-of-Experts (MoE) Transformer module was placed after the NASNet-encoded model to capture complex feature representations. Finally, the NASNet backbone was optimised using the Addax Optimisation Algorithm (AOA).

An XAI model based on enhanced lightweight CNN architectures (Khan et al., 2025) was proposed for the accurate and interpretable classification of several pulmonary diseases. Lightweight CNN models were designed to reduce computational demands and maintain strong predictive performance. A hybrid AI-based model (Shivwanshi & Nirala, 2025) was presented for lung cancer classification from CT images. The methodology combined CNN-based lung nodule segmentation with handcrafted radiomic feature extraction, followed by an optimised CatBoost classifier for multiclass lung nodule classification. Then, SHAP-based XAI techniques were employed to interpret the contributions of individual features to classification outcomes.

To address the feature selection challenge, a Hybrid Spotted Hyena Optimiser-Seagull Algorithm (SHO-SA) (Prasad et al., 2024) was introduced to identify informative features. Image quality was enhanced by bicubic interpolation to reduce noise and improve visual quality. Then, a Deep Convolutional Generative Adversarial Network (DCGAN) was employed to increase sample diversity. The selected lung features were analysed using a hybrid CNN-LSTM model.

An intelligent CAD-based model was presented for automated pneumonia detection using CXR images (Parthasarathy & Saravanan, 2024). Initially, image quality was improved using median filtering to suppress noise and improve image clarity. ResNet50 was then employed for deep feature representations. Finally, a Long Short-Term Memory (LSTM) with Harris Hawks Optimiser (HHO) was applied to pneumonia classification.

An advanced lung disease prediction model (Indumathi & Siva, 2023) that integrated a hybrid Bidirectional LSTM (Bi-LSTM) with a Mask Region Based CNN (M-RCNN) was proposed. The Crystal algorithm was used to optimise the M-RCNN, while the Bi-LSTM was employed to capture temporal and long-range feature dependencies. The output features of the optimised M-RCNN were subsequently processed by the Bi-LSTM and the fully connected layer to perform lung disease prediction.

An improved lung disease classification model to identify COVID-19, viral pneumonia, and normal cases using CXR and CT images was presented (Khate & Neelima, 2026). Initially, discriminative features were extracted using pre-trained DenseNet-121 and ResNet-50 networks. The extracted representations were combined to form a unified feature vector containing complementary information. The Honey Badger Algorithm (HBA) was employed to improve classification performance. Finally, a Support Vector Machine (SVM) was trained using the resulting feature vector, enabling reliable lung disease classification. An advanced lung disease classification model (Chutia et al., 2025) based on enhanced CXR images was presented. Initially, the image quality was enhanced. Subsequently, a three-channel CNN was presented for feature extraction, and a Genetic Algorithm (GA) was employed for feature optimisation.

An effective DL model was proposed for classifying lung CXR images into six categories (Zhu et al., 2025). The hybrid feature extraction model integrated features obtained from VGG16

and a CNN using a feature-fusion mechanism. For enhancing the representational capability of the extracted features, Squeeze-and-Excitation (SE) blocks and self-attention modules were incorporated. Through channel-wise and spatial feature refinement, the model achieved reliable identification of lung diseases. The SE-ResNeXt-50 with CNN (Priya & Bharathi, 2025) was presented for lung cancer classification. Initially, Dynamic Quadri-Histogram Equalisation (QDHE) was used to improve image quality, and data augmentation was performed to enhance model generalisation. Hyperparameter optimisation was employed to refine the learning process and enhance predictive performance.

3. Proposed Methodology

The proposed framework is designed to achieve accurate and robust lung disease classification. Initially, the input CXR images undergo pre-processing, and the pre-processed images are passed through QEViT to extract deep features. To enhance feature learning, the globally contextualised features are encoded into a VQC to generate quantum features. Finally, the generated quantum features are fed into a classification head to predict the lung disease category. The suggested QEViT model is described in Figure 1.

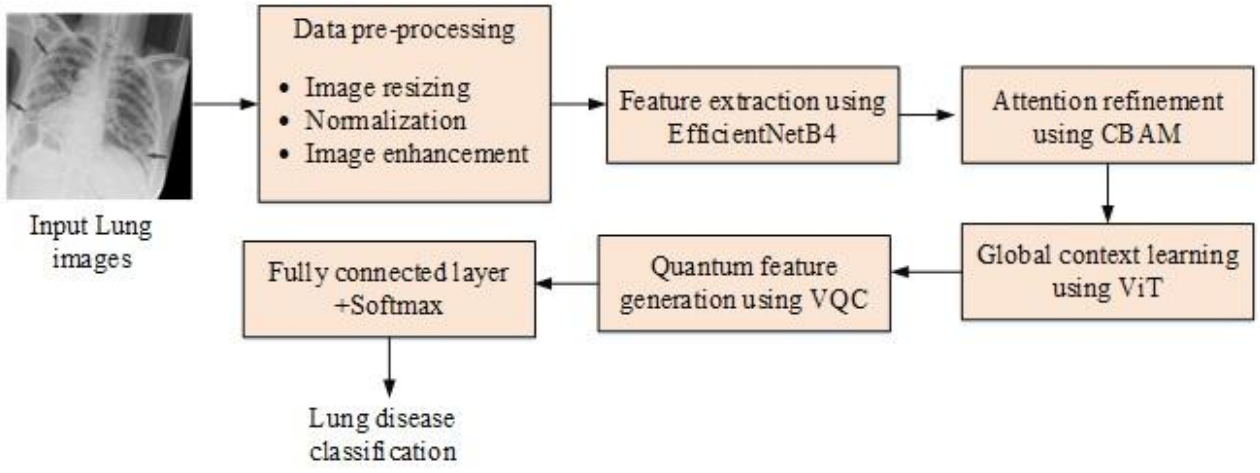


Figure 1. Framework of the suggested lung disease classification model

3.1. Pre-Processing

Medical images exhibit variations in spatial resolution, illumination, and noise characteristics. Therefore, pre-processing is performed to improve image quality and ensure consistency. The input lung images are resized to a fixed dimension to maintain uniformity across the dataset and reduce computational complexity. Let denote the original image and denote the resized image. The resizing operation is expressed as follows:

$$I_r(y,z) = \mathrm{Re}\,size\left(I(y,z), H, W\right) \tag{1}$$

where H and W are the image height and width.

Normalisation is applied to standardise pixel intensity values and enhance network convergence. The value of normalised pixel is computed as follows:

$$I_n = \frac{I - I_{\min}}{I_{\max} - I_{\min}} \tag{2}$$

where $I_{\min}$ and $I_{\max}$ are the minimum and maximum pixel values in the image.

The Contrast-Limited Adaptive Histogram Equalisation (CLAHE) is applied to improve the visibility of pulmonary structures. The enhanced image is computed as follows:

$$I_{en} = CLAHE(I_n) \tag{3}$$

The local histogram equalisation process (Lai et al., 2017) is given as follows:

$$I_{en}(y,z) = \frac{CDF(I_{en}(y,z)) - CDF_{\min}}{1 - CDF_{\min}} \tag{4}$$

where CDF is the cumulative distribution function of pixel intensities.

3.2. QEViT for Deep Feature Extraction

The proposed QEViT is a hybrid DL model designed for accurate lung disease classification. As shown in Figure 2, the network has EfficientNetB4, CBAM, and ViT modules to capture local and global pulmonary characteristics. Initially, the input CXR images are pre-processed and fed into EfficientNetB4 for feature extraction. The extracted features are refined by CBAM and further processed by ViT for capturing global contextual information. The resulting feature representation is then projected into a compact form and encoded into the VQC for quantum feature generation. Finally, the generated quantum features are passed to the fully connected classification layer for lung disease prediction.

Local feature extraction using EfficientNetB4: EfficientNetB4 is presented as the backbone network to extract discriminative local features from lung images. The network utilises a compound scaling model that uniformly scales the input resolution, network depth and width to achieve superior classification performance while maintaining computational efficiency. The architecture uses depthwise separable convolutions, inverted residual layers, and squeeze-and-excitation modules, enabling efficient feature reuse and preservation of major spatial information. In lung disease classification, EfficientNetB4 captures fine-grained pulmonary characteristics. By advanced convolutional operations, the network transforms low-resolution pixels into high-level semantic representations suitable for attention and transformer-based processing. The local feature extraction process is expressed as follows:

$$F_L = f_\theta(I_{en}(y,z)) \tag{5}$$

where f_θ is the learnable parameter.

EfficientNetB4 produces a high-dimensional feature representation that preserves essential pulmonary patterns and serves as the basis for subsequent attention-based refinement.

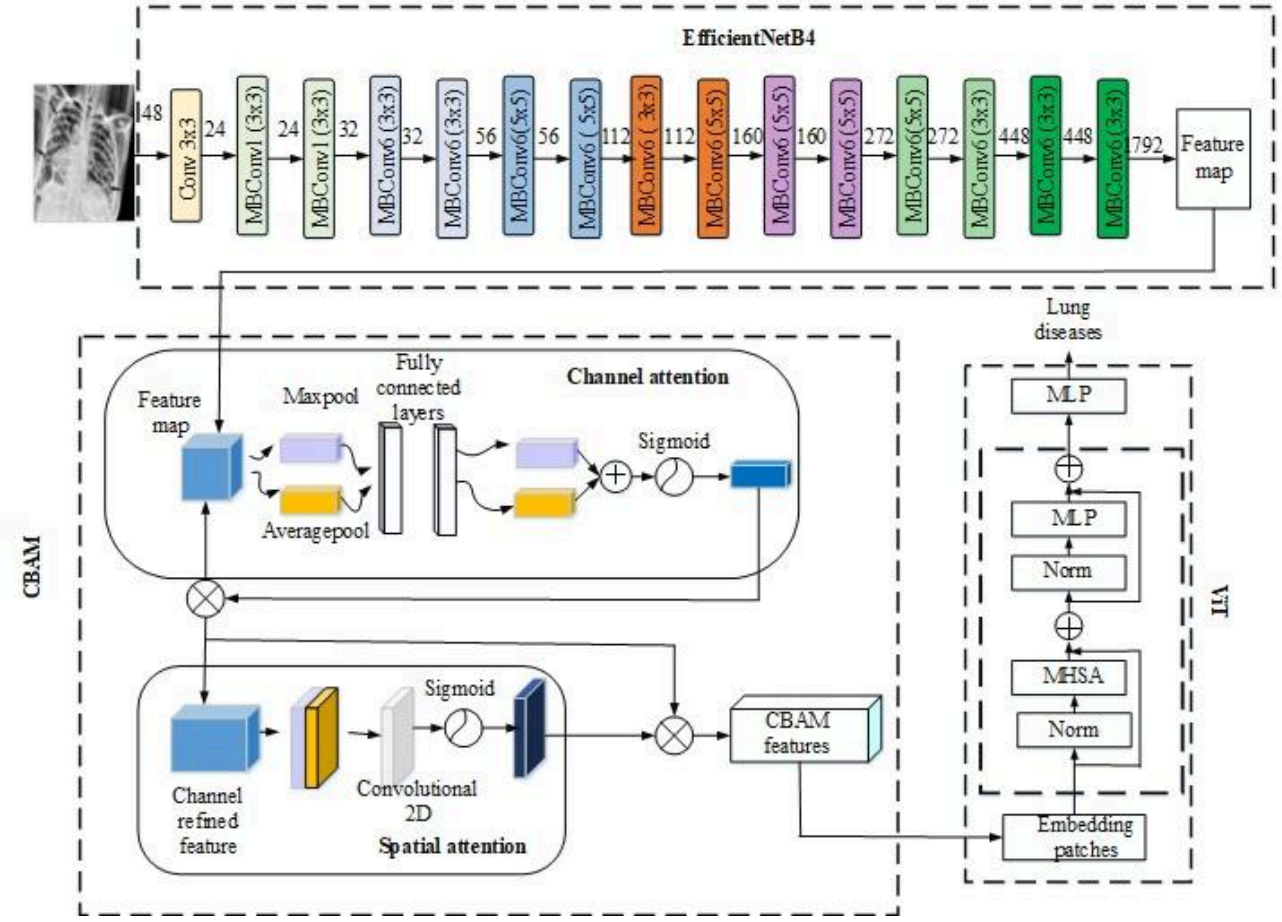


Figure 2. Architecture of the QEViT for deep feature extraction

Attention-based feature refinement using CBAM: CBAM is incorporated to enhance the discriminative ability of the extracted features. CBAM sequentially applies channel and spatial attention to highlight disease-relevant regions and suppress unwanted background information. The channel attention module assigns higher weights to feature channels which have clinically important pulmonary patterns. The operation of channel attention is presented as follows:

$$M_C(F) = \sigma\left(MLP\left(Averagepool(F)\right) + MLP\left(\max pool(F)\right)\right) \tag{6}$$

where F is the input feature map, is the sigmoid activation function, is the average pooling operation and is the pooling operation.

The spatial attention module minimises the influence of background structures and highlights disease-relevant spatial regions within the lung image, enabling precise localisation of pathological abnormalities. The operation of spatial attention $M_s(F')$ is expressed as follows:

$$M_s(F') = \sigma\left(f_{7\times7}\left[Averagepool(F'); \max pool(F')\right]\right) \tag{7}$$

where $f_{7\times7}$ is a convolutional operation with kernel size 7×7, and $[;]$ is channel-wise concatenation.

The CBAM-enhanced feature representation highlights clinically significant lung abnormalities and reduces background interference.

Global feature extraction using ViT: ViT is used to capture global contextual and long-range features in lung images. Conventional CNNs focus only on local receptive fields; ViT processes image representations as a sequence of patches and captures their relations through self-attention mechanisms. This enables the network to capture complex relations among spatially distant regions. The ViT uses a patch size of 16 × 16 pixels to divide the input feature map into a sequence of non-overlapping patches. All patches are linearly embedded and combined with positional encodings before being processed by the transformer encoder. The ViT encoder has 12 multi-head self-attention heads, which makes effective modelling of long-range spatial dependencies and global contextual information across the CXR image.

Initially, the input feature representation is divided into a series of pre-determined patches. All patches are flattened and mapped into a latent embedding space by a learnable linear transformation. Positional embeddings are integrated for preserving the spatial arrangement of image patches. The embedding process of a patch X_0 is represented as follows:

$$X_0 = \left[y_p^1 E; y_p^2 E; .....; y_p^M E \right] + E_{pos} \tag{8}$$

where y_p^j is the j^{th} image patch, y_p^j is the embedding matrix, E_{pos} is the positional encoding, and N is the total number of image patches.

The embedded patch sequence is preceded by multi-transformer encoder layers, which have Multi-Head Self-Attention (MHSA) and feed-forward networks (FFNs). The self-attention mechanism is computed as follows:

$$At(Q,K,V) = soft\max\left(\frac{K^T Q}{\sqrt{d_k}} \right) V \tag{9}$$

where $Q = YW_Q, \quad K = YW_K, \quad V = YW_V$; Q, K and V are query, key, and value matrices. Here, Y is the input patch embeddings, W_Q, W_K and W_V are learnable projection matrices and d_k is the key vector's dimensionality.

The self-attention operation enables every patch to establish relations with all other patches in the image, capturing global contextual dependencies. This ability is highly advantageous in lung disease analysis, where pathological changes may appear over several spatial locations and show complex inter-regional correlations. The transformer encoder output is expressed as follows:

$$F_{global} = ViT(X_0) \tag{10}$$

where F_{global} is the representation of the global contextual feature generated by the ViT. These features contain semantic information and are used as input for subsequent classification and quantum feature learning stages.

3.3.VQC for Quantum Feature Generation

To enhance the discriminative ability of the extracted deep features, a VQC is integrated into the proposed QEViT model; its structure is given in Figure 3. The VQC exploits quantum mechanical principles such as superposition and entanglement for learning high-level feature representations. The integration of quantum computing enables the network to capture complex non-linear interactions among pulmonary characteristics and enhances disease discrimination. Initially, the globally contextualised feature vector generated by the Vision Transformer is transformed into a quantum-compatible representation.

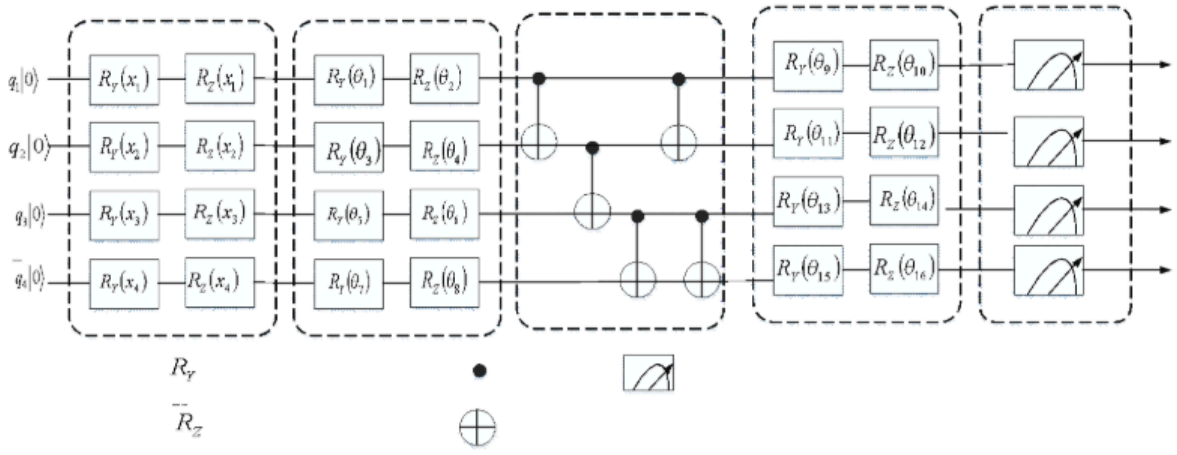


Figure 3. VQC for quantum feature generation

The ViT creates a feature vector of size 768. Before quantum encoding, a dense layer projects the feature size down to four elements for the four qubits. The normalised feature values are in the range $[-\pi,\pi]$, and then quantum angle encoding is applied on them.

Let F_{global} is the feature vector obtained from the transformer module. A subset of these features is encoded into quantum states using angle encoding, where all feature values manage the rotation angle of a respective qubit. The quantum state preparation process is represented as follows:

$$|\psi\rangle = U_{enc}\left(F_{global}\right)|0\rangle^{\otimes n} \tag{11}$$

where U_{enc} is the encoding operation, n is the number of qubits, and is the initial quantum state.

After the feature-encoding process, the quantum state is processed through a sequence of trainable quantum gates. In the proposed work, parameterised rotation gates around the Y-axis and Z-axis are presented for capturing quantum representations. The respective gate operations are defined as follows:

$$R_Y\left(\theta_j\right)=\begin{bmatrix}\cos\left(\theta_j/2\right) & -\sin\left(\theta_j/2\right)\\ \sin\left(\theta_j/2\right) & \cos\left(\theta_j/2\right)\end{bmatrix} \quad R_z\left(\theta_j\right)=\begin{bmatrix}e^{-k\theta_j/2} & 0\\ 0 & e^{k\theta_j/2}\end{bmatrix} \tag{12}$$

where θ_j is the trainable parameter with the j^{th} quantum gate. These variational gates transform the encoded quantum state and capture complex non-linear feature representations. The trainable parameters are represented as follows:

$$=\{\theta_1,\theta_2,\theta_3,....,\theta_m\} \tag{13}$$

For capturing interactions among encoded features, entanglement layers are introduced through controlled-NOT (CNOT) gates. These operations allow information exchange among qubits and assist the learning of complex feature dependencies. The transformed quantum state is given as follows:

$$|\phi\rangle = U(\Theta)|\psi\rangle \tag{14}$$

where $U(\Theta)$ do learnable parameters parameterise the trainable quantum circuit and $|\phi\rangle$ is the transformed quantum state.

The VQC has 4 qubits along with 2 variational layers. The operations performed by each of the layers will be as follows:

Angle Encoding → RY Rotation → RZ Rotation → Linear CNOT Entanglement

CNOT operations are used between $(q0 \to q1)$, [ ] and $(q2 \to q3)$. As each qubit has one trainable RY parameter and one trainable RZ parameter, each layer has 8 trainable parameters, and hence a total of 16 trainable quantum parameters in the entire circuit.

The resultant quantum features are obtained by computing the Pauli-Z operator's expectation values.

$$Q_j = \langle \phi | Z_j | \phi \rangle \tag{15}$$

where Q_j is the output of the j^{th} qubit and Z_j is the respective Pauli-Z observable. The generated quantum feature vector is given as follows:

$$Q = [Q_1, Q_2, Q_3, Q_n] \tag{16}$$

where n is the total number of qubits.

The quantum circuit was implemented using the PennyLane framework with the default qubit simulator integrated with TensorFlow. All experiments are conducted using the simulator, and no experiments are conducted on physical quantum hardware. The model parameters are optimised by automatic differentiation (backpropagation) provided by PennyLane and TensorFlow, enabling end-to-end training of the hybrid quantum-classical architecture.

3.4. Classification Head and Training Strategy

The VQC generates the quantum feature vector and is fed to a fully connected classification layer for final lung disease detection. The classification head is used for transforming the quantum representations into class-specific logits with respect to various disease classes. The classification process is expressed as:

$$Y = WQ + b \tag{17}$$

where W and b are the learnable parameters, Y is the output logits.

The Softmax activation is used to obtain the final class probabilities, and the loss function used is categorical cross-entropy, which optimises the model.

$$Loss = \sum_{j=1}^{C} y_j \log\left(P\left(y_j\right)\right) \tag{18}$$

where y_j is the ground-truth label and $P\left(y_j\right)$ is the predicted class probability.

During training, the Adam optimiser is used to update classical and quantum trainable parameters by gradient-based optimisation. The hybrid quantum-classical learning process enables the network to capture discriminative representations for precise lung disease classification.

3.5. Visual interpretability via Grad-CAM and Score-CAM

Interpretability is an important need in CAD systems, as it enables clinicians to analyse the decision-making of DL approaches. To improve the transparency and reliability of the proposed QEViT model, Gradient-weighted Class Activation Mapping (Grad-CAM) and Score-Weighted Class Activation Mapping (Score-CAM) are presented for generating visual details for lung disease predictions. These techniques show the image areas that contribute most to the classification outcome, providing insight into the diagnostic reasoning.

4. Results Analysis

The suggested QEViT model is implemented using Python with TensorFlow and PennyLane libraries on a Windows 11 platform. The experiment analysis is evaluated on a workstation equipped with an Intel Core i9 and 32 GB RAM to ensure robust model training and inference. The hyperparameters used for training the proposed QEViT are presented in Table 1.

Table 1. Hyperparameters

Parameters	Values
Size of Input Image	224 × 224
Size of Batch	32
Epochs	100
Optimizer	Adam
Learning Rate	0.0001
Dropout Rate	0.5
Activation	ReLU
Output Activation	SoftMax
Loss Function	Categorical Cross-Entropy
Number of Qubits	4
Quantum Layers	2
Rotation Gates	RY, RZ
Entanglement Gates	CNOT

4.1. Datasets

The proposed study utilises two publicly available CXR datasets, namely the Lungs Disease Dataset (Dalvi, 2022) and the COVID-19 Radiography Database (Chowdhury et al., 2020). The utilisation of these datasets enables comprehensive evaluation of the proposed QEViT model across various pulmonary disease classes and diverse imaging cases.

Lung Disease Dataset consists of CXR images and includes classes such as COVID-19, Viral Pneumonia, Bacterial Pneumonia, Tuberculosis, and Normal classes. The dataset was obtained by combining various benchmark accessible repositories and augmented to nearly 10,000 images after removing duplicate samples.

The COVID-19 Radiography dataset has four classes of CXR images, which include 3,616 COVID-19 images, 10,192 Normal images, 6,012 Lung Opacity images, and 1,345 Viral Pneumonia images, along with respective lung masks. This provides a large-scale benchmark for analysis of pulmonary disease. Table 2 shows the dataset partitioning strategy and the number of images allocated to the training, validation, and testing subsets for each disease category.

Table 2: Dataset partitioning strategy

Class	Total	Train (70%)	Validation (10%)	Test (20%)
COVID-19	3,616	2,531	362	723
Normal	10,192	7,134	1,019	2,039
Lung Opacity	6,012	4,208	601	1,203
Viral Pneumonia	1,345	942	134	269

Before training, duplicates in the dataset were filtered out using image hash-matching to eliminate repeated samples and prevent data leakage. After preprocessing, each dataset was split into training (70%), validation (10%) and testing (20%) sets while keeping the distribution among classes intact through stratified sampling. Dataset splitting was done after removing duplicates, such that no example would be present in both training/validation/testing datasets. Furthermore, data augmentation models are applied only to the training set, and the validation and testing sets remain unchanged to ensure an unbiased evaluation of the proposed QEViT model.

4.2. Evaluation Metrics

For validating the robustness of the proposed QEViT, metrics like Accuracy, Precision, Recall (True Positive Rate (TPR)), F1-Score, Specificity, False Positive Rate (FPR), confusion matrix and Receiver Operating Characteristic-Area Under the Curve (ROC-AUC) are used.

As lung disease classification is a multi-class problem, precision, Recall, F1-Score, Specificity, and FPR are first computed independently for each class in a one-vs-rest approach. The final reported values correspond to the macro-average obtained by averaging the individual class metrics with equal weight. ROC-AUC values are also computed using the one-versus-rest approach and macro averaging. These metrics demonstrate the classification ability of the proposed model by using true positives T_{pox}, true negatives T_{nex}, false positives F_{pox}, and false negatives F_{nex}. Accuracy is the ratio of the correctly classified images among all evaluated samples, and it is expressed as follows:

$$Accuracy = \frac{T_{pox} + T_{nex}}{T_{pox} + T_{nex} + F_{pox} + F_{nex}} \tag{19}$$

Precision is the ratio of correctly identified disease-positive cases among all cases predicted as positive.

$$\Pr e = \frac{T_{pox}}{T_{pox} + F_{pox}} \tag{20}$$

Recall computes the model's potential to identify patients with lung disease accurately.

$$\mathrm{Re} = \frac{T_{pox}}{T_{pox} + F_{nex}} \tag{21}$$

The F1-Score provides a balanced analysis of $\Pr e$ and Re by computing the harmonic mean.

$$Accuracy = \frac{2 \times \Pr e \times \mathrm{Re}}{\Pr e + \mathrm{Re}} \tag{22}$$

Specificity computes the ability of the model to identify healthy or disease-negative cases correctly.

$$Specificity = \frac{T_{nex}}{T_{pox} + F_{nex}} \tag{23}$$

FPR measures the rate of disease-negative cases that are incorrectly classified as positive.

$$FPR = \frac{F_{pox}}{T_{nex} + F_{pox}} \tag{24}$$

4.3. Qualitative Analysis

The qualitative assessment of the suggested QEViT is provided in Figure 4 using the Lung Disease and COVID-19 Radiography datasets. Figure 4(a) represents the input CXR images, and Figure 4(b) presents the processed images after resizing and normalisation. Figures 4(c) and 4(d) represent the Grad-CAM and Score-CAM visualisation techniques, which are obtained from the suggested QEViT. The analysis shows that the highlighted regions mostly focus on abnormal lungs. Compared with the original images, the activation maps provide clear localisation of disease-related regions and show the interpretability and reliability of the suggested QEViT for lung disease diagnosis.

The activation maps generated using Grad-CAM and Score-CAM reveal the areas of the image that play an essential role in driving the prediction of the model and help us understand the learned feature representations. These activation maps highlight the relevant regions in the lungs associated with the diseases, which show that the suggested QEViT considers significant regions of the image while classifying the disease. Therefore, the Grad-CAM and Score-CAM results are intended to support the analysis of the model's predictions rather than to validate the clinical accuracy of the diagnosis. Although the proposed model correctly identifies the disease category, Figures 4(c) and 4(d) show that certain activation regions extend beyond the primary lung abnormalities. This indicates that the explainability methods highlight image regions influencing the prediction but should not be interpreted as precise lesion segmentation or definitive clinical localisation.

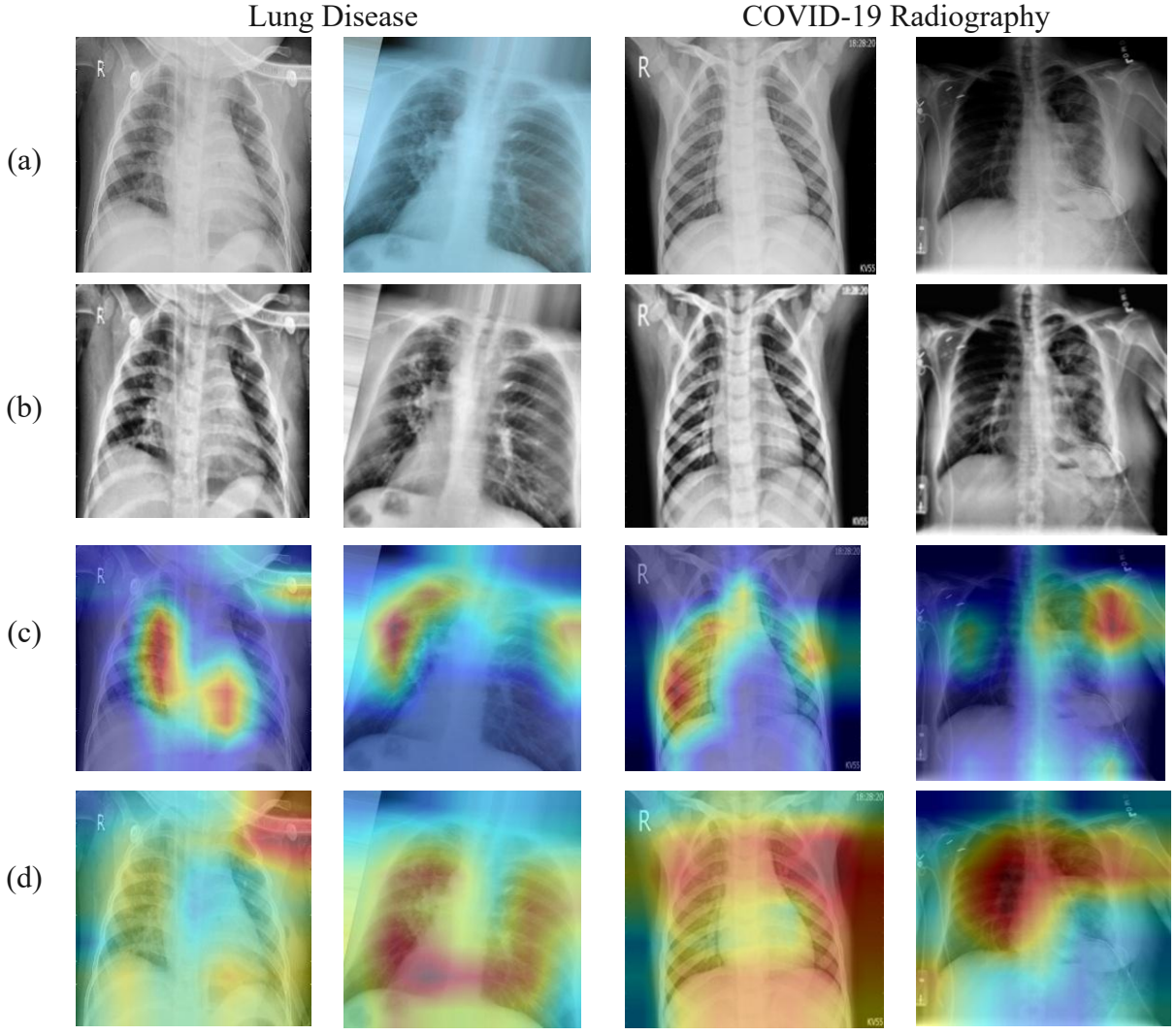


Figure 4. Visual analysis of (a) Input images, (b) Pre-processed images, (c) GradCAM and (d) ScoreCAM images of the QEViT model

4.4. Quantitative Analysis

The quantitative analysis evaluates the robustness of the suggested QEViT by various evaluation metrics on the Lung Disease and COVID-19 Radiography datasets. The suggested QEViT model is compared with EfficientNetB4, ResNet50, DenseNet121, and ViT models to validate its efficacy. The obtained results show the improvement of QEViT across various lung disease categories.

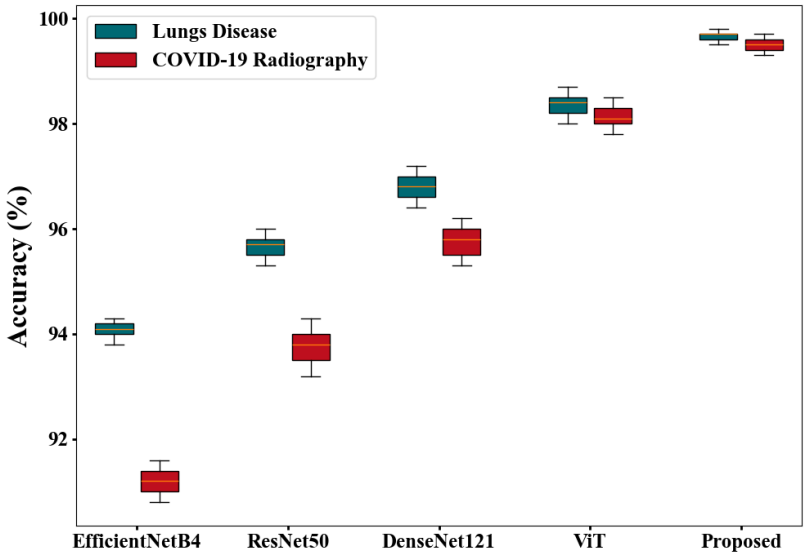


Figure 5. Accuracy comparison

Figure 5 illustrates the accuracy analysis on the Lung Disease and COVID-19 Radiography datasets. For the Lung Disease dataset, QEViT attained an accuracy of 99.8%, and it is superior to EfficientNetB4 (91.2%), ResNet50 (93.8%), DenseNet121 (95.8%), and ViT (98.1%). Similarly, for the COVID-19 Radiography dataset, QEViT achieved 99.7% accuracy, and models like EfficientNetB4, ResNet50, DenseNet121, and ViT achieved 94.1%, 95.7%, 96.8%, and 98.4%, respectively. The superior accuracy of QEViT is because of the integration of transformer-based global feature learning and quantum-enhanced representation generation.

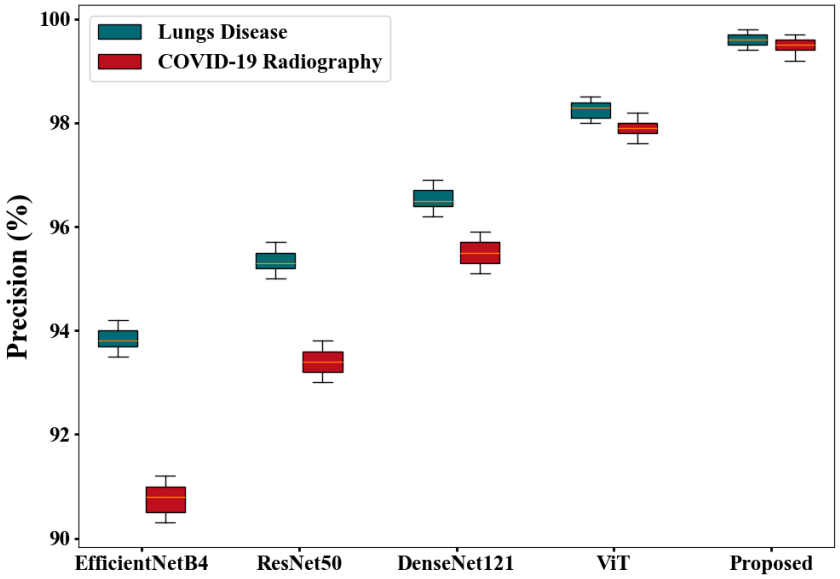


Figure 6. Precision comparison

Figure 6 delineates the precision analysis for two datasets. On the Lung Disease dataset, QEViT achieved a precision value of 99.5%, and EfficientNetB4, ResNet50, DenseNet121, and ViT models attained precision values of 90.8%, 93.4%, 95.5%, and 97.9%. Then, on the COVID-19 Radiography dataset, QEViT attained 99.6% precision and was superior to other existing models. This performance proved that QEViT was an efficient tool for classifying diseases.

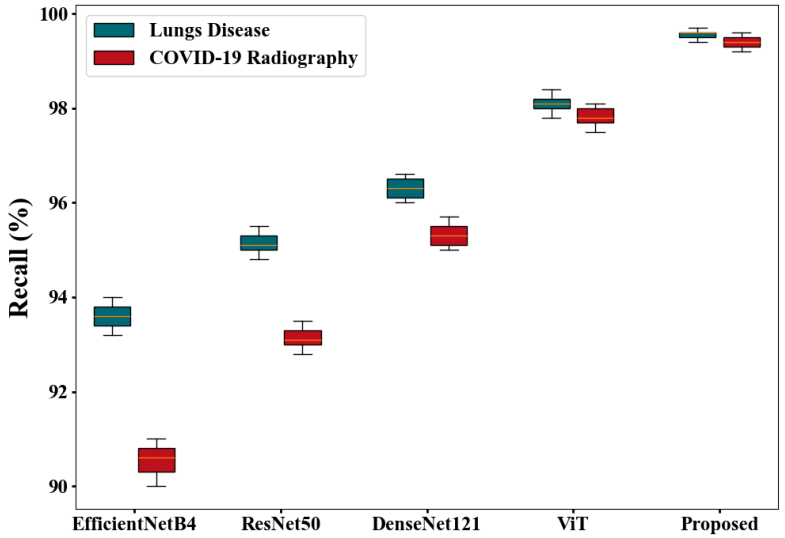


Figure 7. Recall comparison

Figure 7 illustrates the recall comparison on the two datasets. This comparison demonstrates the strong disease detection ability of the proposed QEViT. For the Lung Disease dataset, QEViT achieved a recall of 99.4%; then, EfficientNetB4, ResNet50, DenseNet121, and ViT obtained 90.6%, 93.1%, 95.3%, and 97.8%, respectively. Similarly, on the COVID-19 Radiography dataset, QEViT reached 99.6% recall, which is superior to other methods.

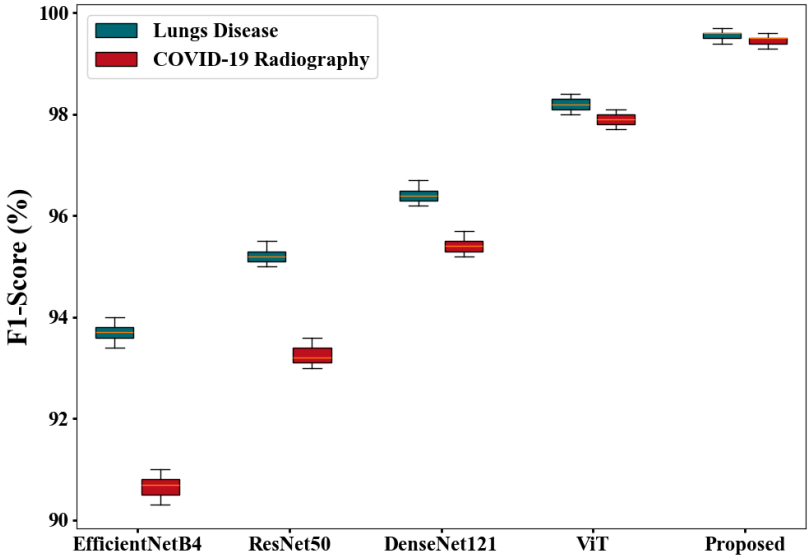


Figure 8. F1-score comparison

Figure 8 delineates the F1-score comparison on the two datasets. The F1-score results confirmed the balanced classification ability of QEViT. On the Lung Disease dataset, QEViT achieved an F1-score of 99.5%, compared with 90.7%, 93.2%, 95.4%, and 97.9% for EfficientNetB4, ResNet50, DenseNet121, and ViT, respectively. Then, QEViT achieved an F1-score of 99.6% on the COVID-19 Radiography dataset.

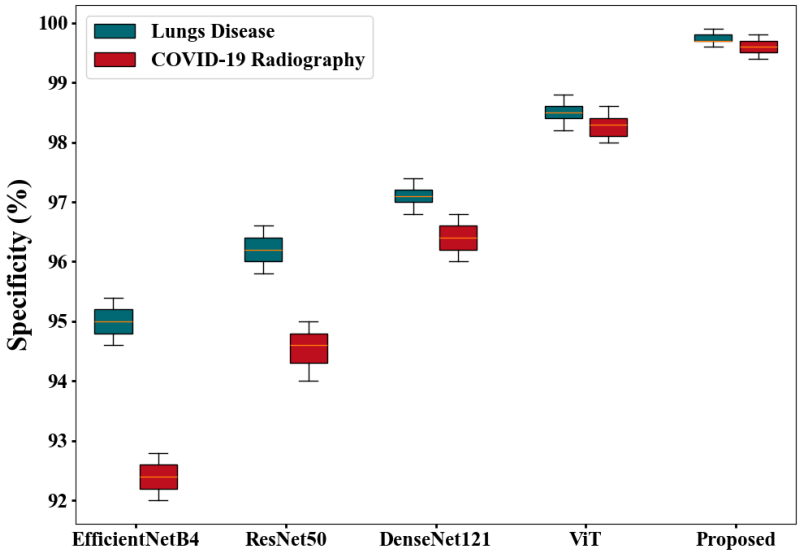


Figure 9. Specificity comparison

Figure 9 shows the specificity comparison on the two datasets. For the Lung Disease dataset, models like EfficientNetB4, ResNet50, DenseNet121, ViT and QEViT achieved a specificity of 92.4%, 94.6%, 96.4%, 98.3% and 99.6%, respectively. Similarly, for the COVID-19 Radiography dataset, QEViT attained a specificity of 99.7%, which is superior to EfficientNetB4 (95.0%), ResNet50 (96.2%), DenseNet121 (97.1%), and ViT (98.5%). The higher specificity indicates that the proposed model preserves healthy samples. This improvement is because of the enhanced feature separability ensured by the quantum-enhanced representation learning process.

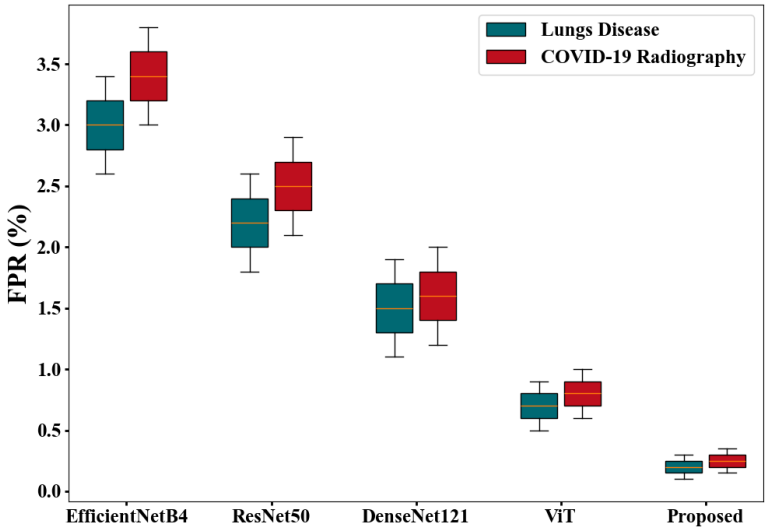


Figure 10. FPR comparison

Figure 10 delineates the FPR comparison on the two datasets. A lower FPR indicates a lower false alarm rate and higher prediction reliability. For the Lung Disease dataset, QEViT attained the lowest FPR of 0.25%, compared with 3.40%, 2.50%, 1.60%, and 0.80% for EfficientNetB4, ResNet50, DenseNet121, and ViT models. Then, for the COVID-19 Radiography dataset, QEViT achieved an FPR of only 0.20%, whereas EfficientNetB4, ResNet50, DenseNet121, and ViT models achieved 3.00%, 2.20%, 1.50%, and 0.70%, respectively. The significant reduction in FPR shows that QEViT generates lower false alarms than existing methods.

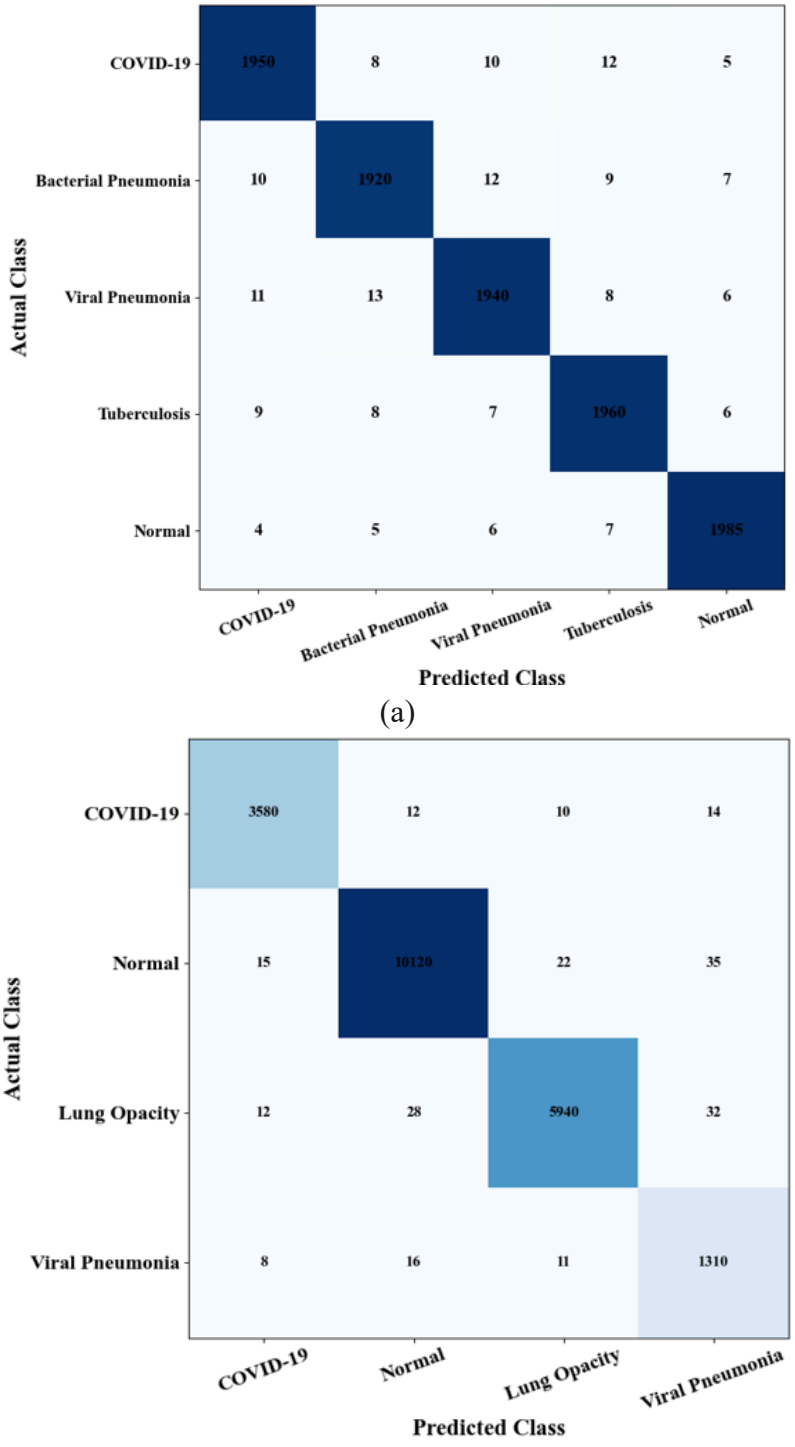


Figure 11. Confusion matrix of the proposed QEViT on (a) Lung Disease dataset and (b) COVID-19 Radiography dataset

The confusion matrices shown in Figure 11 are generated using only the independent test set after model training. Each matrix represents the classification results of all test samples, and the reported accuracy values are calculated from these test-set predictions. All matrices illustrate the number of correctly and incorrectly categorised instances for the respective disease categories. In Figure 11 (a), for the Lung Disease dataset, the model classified 1,950 COVID-19, 1,920 Bacterial Pneumonia, 1,940 Viral Pneumonia, 1,960 Tuberculosis, and 1,985 Normal samples accurately. In Figure 11 (b), for the COVID-19 Radiography dataset, QEViT determined 3,580 COVID-19,

10,120 Normal, 5,940 Lung Opacity, and 1,310 Viral Pneumonia images correctly. This analysis confirmed the superior classification ability and robustness of the proposed QEViT model.

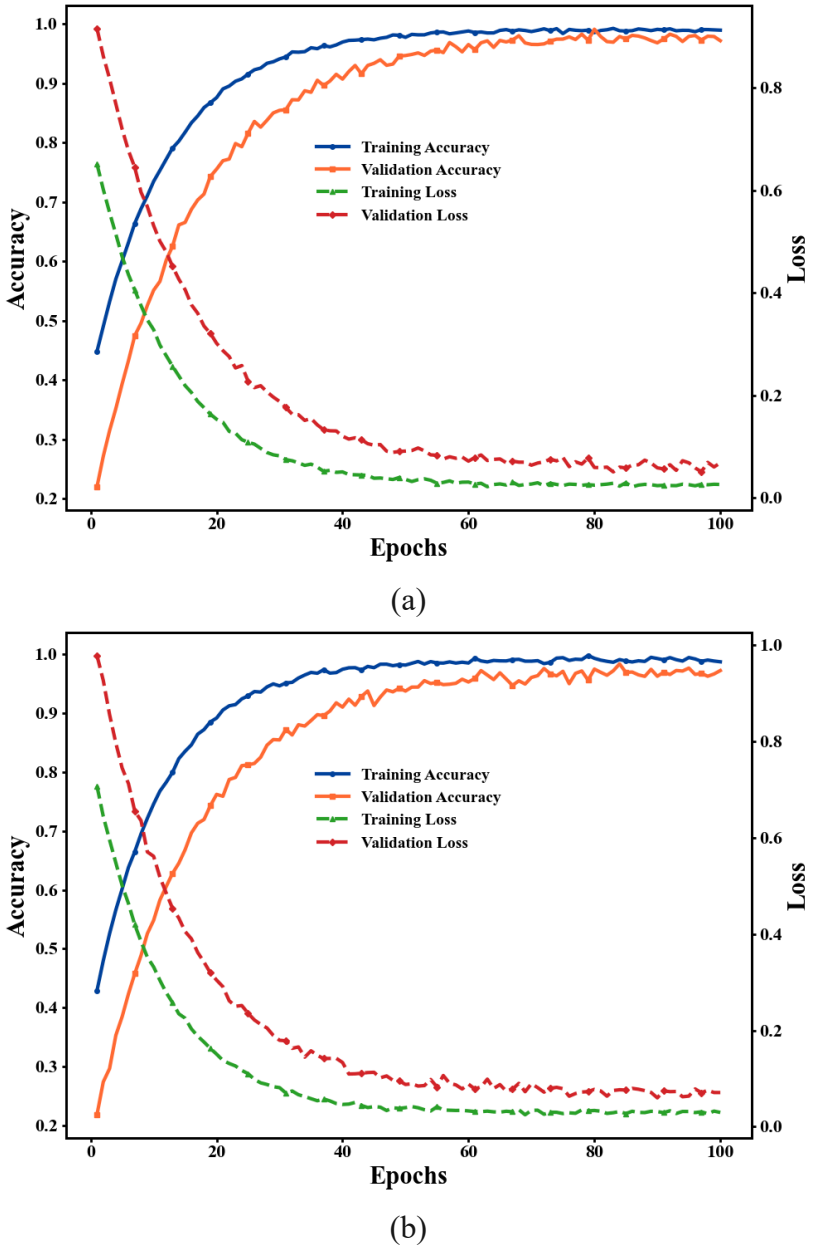
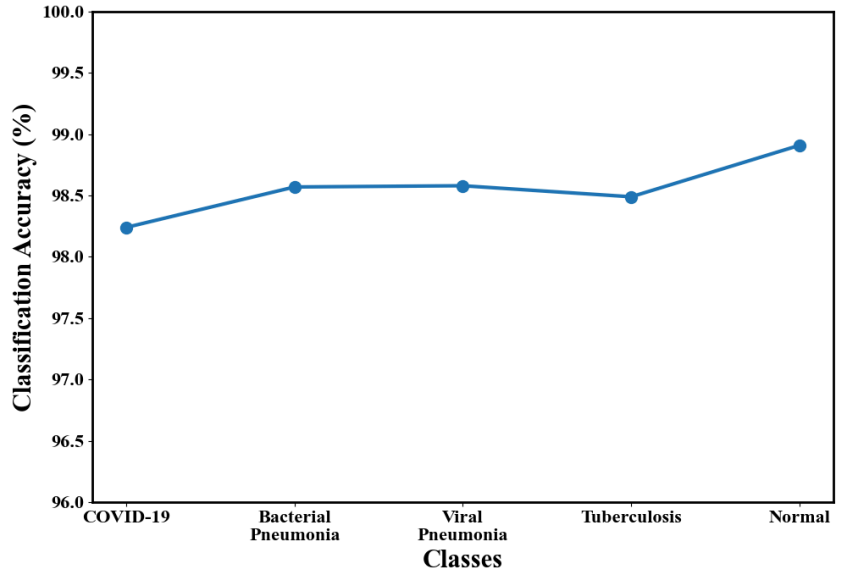
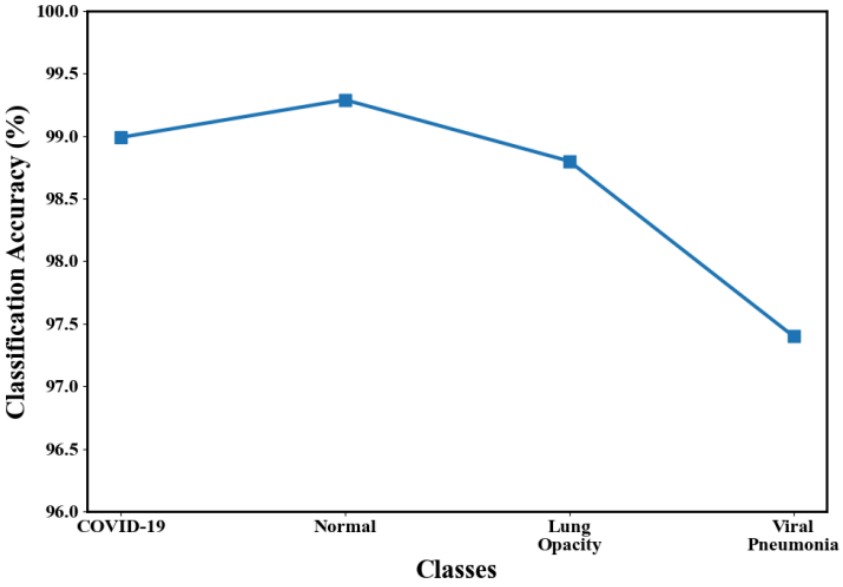


Figure 12. Accuracy-Loss analysis of the proposed QEViT on (a) Lung Disease dataset and (b) COVID-19 Radiography dataset

Figure 12 shows the training and validation accuracy-loss curves of the proposed QEViT model. For both datasets, training and validation accuracies rapidly increase in the initial epochs and gradually converge to values which exceed 99%; this shows strong generalisation ability. Then, the training and validation losses decrease gradually and stabilise at very low values near the final epochs.



(a)



(b)

Figure 13. Class-wise performance analysis on (a) Lung Disease dataset and (b) COVID-19 Radiography dataset

The class-wise performance in Figure 13 shows the classification ability of the proposed QEViT model. In Figure 13 (a), for the Lung Disease dataset, the proposed QEViT model achieved accuracies of 98.24%, 98.57%, 98.58%, 98.49%, and 98.91% for COVID-19, Bacterial Pneumonia, Viral Pneumonia, Tuberculosis, and Normal classes.

Then, in Figure 13 (a), for the COVID-19 Radiography dataset, accuracy values of 98.99%, 99.29%, 98.80%, and 97.40% are obtained for COVID-19, Normal, Lung Opacity, and Viral Pneumonia classes. The high values of accuracy across every class show that QEViT captures pulmonary features and maintains reliable performance.

Table 3. Ablation study

Lung Disease dataset						
Models	**Accuracy (%)**	**Precision (%)**	**Recall (%)**	**F1-Score (%)**	**Specificity (%)**	**FPR (%)**
EfficientNetB4	91.2	90.8	90.6	90.7	92.4	3.4
EfficientNetB4 + CBAM	94.6	94.1	93.8	93.9	95.1	2.2
EfficientNetB4 + CBAM + ViT	98.1	97.9	97.8	97.9	98.3	0.8
EfficientNetB4 + CBAM + ViT + Dense	99.0	98.9	98.8	98.8	99.0	0.55
EfficientNetB4 + CBAM + ViT + MLP	98.9	99.1	98.7	99.0	99.2	0.42
QEViT (EfficientNetB4 + CBAM + ViT + VQC)	99.8	99.5	99.4	99.5	99.6	0.25
COVID-19 Radiography dataset						
EfficientNetB4	94.10	93.8	93.6	93.7	95.1	3.0
EfficientNetB4 + CBAM	96.5	96.2	96.0	96.1	96.9	1.9
EfficientNetB4 + CBAM + ViT	98.4	98.3	98.1	98.2	98.5	0.7
EfficientNetB4 + CBAM + ViT + Dense	99.0	98.8	98.7	98.7	98.9	0.48
EfficientNetB4 + CBAM + ViT + MLP	99.1	98.7	98.9	98.9	99.1	0.35
QEViT (EfficientNetB4 + CBAM + ViT + VQC)	99.7	99.6	99.6	99.6	99.7	0.20

The ablation analysis in Table 3 validates the contribution of every component within the suggested QEViT model. The EfficientNetB4 model achieves the lowest performance on both datasets. Integration of the CBAM module improves the performance, and the addition of the ViT enhances performance by capturing long-range contextual dependencies that are complex to model by convolutional operations. To further evaluate the contribution of the proposed VQC, two the VQC is replaced with a fully connected Dense layer and a two-layer Multilayer Perceptron (MLP). Finally, incorporation of the VQC provided the best results. These findings show that every module contributes to the entire network, with the combined model achieving better classification performance.

Table 4. Comparative analysis with literature works

References	**Accuracy (%)**	**Precision (%)**	**Recall (%)**	**F1-Score (%)**
[8]	96.48	97.56	93.75	95.62
[9]	84.7	86.2	85.6	85.9
[11]	95.10	95.67	95.10	95.38
[15]	96.74	93.68	96.74	95.15
[20]	99.33	98	97	98
[22]	99.15	99.5	97.58	98.54
Proposed	**99.8**	**99.5**	**99.4**	**99.5**

Table 4 delineates a comparative study of the suggested QEViT model with respect to recent papers based on lung disease classification approaches. The comparison is performed using different measures. As observed, the suggested QEViT model achieved better performance. Although the proposed QEViT achieves competitive performance compared with literature works, direct numerical comparison must be interpreted carefully because these studies used different datasets, pre-processing procedures, class distributions, and train-test partitioning strategies. Hence, the comparison is primarily considered as a qualitative reference rather than a strictly controlled benchmark.

Table 5. Wilcoxon Signed-Rank Test with Holm–Bonferroni Correction

Models	$\rho-$ **value**	**Adjusted** $\rho-$ **value**	**Effect Size (** r **)**
QEViT vs EfficientNetB4	0.0008	0.0032	0.93
QEViT vs ResNet50	0.0011	0.0033	0.90
QEViT vs DenseNet121	0.0018	0.0036	0.86
QEViT vs ViT	0.0045	0.0045	0.81

To conduct another comparison between the performance of the proposed QEViT and the baseline methods, the Wilcoxon Signed-Rank Test, adjusted by the Holm-Bonferroni method, is applied. The analysis is carried out based on the performance measures for each model obtained by running experiments several times using the independent test set. According to the null hypothesis, no statistically significant difference exists between the compared models and the alternative hypothesis assumes that a statistically significant difference exists among them. A significance level of 0.05 is chosen for the test, as well as an effect size measure.

A Wilcoxon signed-rank test with Holm–Bonferroni correction is used to evaluate the statistical significance of the performance improvements achieved by QEViT. As illustrated in Table 5, the adjusted p-values are below 0.05, indicating significant differences between QEViT and the existing models. The effect size values ranged from 0.81 to 0.93, indicating a high practical effect. These findings show that the performance achieved by QEViT is superior and statistically significant on two datasets.

5. Conclusion

This work presented a QEViT model for automated classification of lung disease from CXR images. The approach integrated EfficientNetB4 for deep feature extraction, CBAM for attention-based feature refinement, ViT for global contextual learning, and VQC for quantum-enhanced feature representation. Experimentation on two datasets demonstrated the promising performance of the proposed approach.

The explainable visualisations provided insight into the image regions influencing the model predictions. The obtained results suggest that the proposed QEViT may support automated CXR analysis for lung disease classification. However, further validation using larger independent clinical datasets is required before practical clinical deployment. Although the QEViT framework seems promising based on the results obtained, the study has several drawbacks. First, the experiments were conducted using publicly available datasets that might not completely reflect the variety present in real clinical practice.

Therefore, external validation needs to be done by testing the proposed algorithm with independent multi-centre datasets. Moreover, while the VQC provided better results, more research is needed by conducting the study with large amounts of data using other quantum computing systems. Further work will focus on extending the proposed QEViT model to explore multimodal learning by combining CXR with clinical metadata and electronic health records to support personalised disease diagnosis.

References

Alghadhban, A., Ramadan, R. A., & Alazmi, M. (2025). Advancing respiratory disease diagnosis: A deep learning and vision transformer-based approach with a novel X-ray dataset. *Computers in Biology and Medicine, 194*, Article 110501. https://doi.org/10.1016/j.compbiomed.2025.110501

Al-Qaness, M. A. A., Zhu, J., Al-Alimi, D., Dahou, A., Alsamhi, S. H., Abd Elaziz, M., & Ewees, A. A. (2024). Chest X-ray images for lung disease detection using deep learning techniques: A comprehensive survey. *Archives of Computational Methods in Engineering, 31*(6), 3267–3301. https://doi.org/10.1007/s11831-024-10081-y

Alshmrani, G. M. M., Ni, Q., Jiang, R., Pervaiz, H., & Elshennawy, N. M. (2023). A deep learning architecture for multi-class lung diseases classification using chest X-ray (CXR) images. *Alexandria Engineering Journal, 64*, 923–935. https://doi.org/10.1016/j.aej.2022.10.053Askr, H., Moawad, M., Darwish, A., & Hassanien, A. E. (2025). Multiclass deep learning model for predicting lung diseases based on honey badger algorithm. *International Journal of Information Technology, 17*(2), 1147–1154. https://doi.org/10.1007/s41870-024-02046-y

Bray, F., Laversanne, M., Sung, H., Ferlay, J., Siegel, R. L., Soerjomataram, I., & Jemal, A. (2024). Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians, 74*(3), 229–263. https://doi.org/10.3322/caac.21834

Chowdhury, M. E. H., Rahman, T., Khandakar, A., Mazhar, R., Kadir, M. A., Mahbub, Z. B., Islam, K. R., Khan, M. S., Iqbal, A., Al Emadi, N., Reaz, M. B. I., & Islam, M. T. (2020). Can AI help in screening viral and COVID-19 pneumonia? *IEEE Access, 8*, 132665–132676. https://doi.org/10.1109/ACCESS.2020.3010287

Chutia, U., Tewari, A. S., & Singh, J. P. (2025). Lung disease classification using deep learning and genetic algorithm. *Neural Computing and Applications, 37*(25), 20519–20534. https://doi.org/10.1007/s00521-024-10527-1

Dalvi, O. M. (2022). *Lung disease dataset (4 types)* [Data set]. Kaggle. https://www.kaggle.com/datasets/omkarmanohardalvi/lungs-disease-dataset-4-types

Hiremath, T. C., Shwetha, K. S., Ujwal, U. J., Tejaswini, M. R., Narayanapur, A., & Shivayogi, U. (2026). Efficient lung abnormality detection using deep convolutional neural networks. *SN Computer Science, 7*(5), Article 409. https://doi.org/10.1007/s42979-026-05025-1

Hroub, N. A., Alsannaa, A. N., Alowaifeer, M., Alfarraj, M., & Okafor, E. (2024). Explainable deep learning diagnostic system for prediction of lung disease from medical images. *Computers in Biology and Medicine, 170*, Article 108012. https://doi.org/10.1016/j.compbiomed.2024.108012

Indumathi, V., & Siva, R. (2023). An efficient lung disease classification from X-ray images using hybrid Mask-RCNN and BiDLSTM. *Biomedical Signal Processing and Control, 81*, Article 104340. https://doi.org/10.1016/j.bspc.2022.104340

Iqbal, A., Usman, M., & Ahmed, Z. (2022). An efficient deep learning-based framework for tuberculosis detection using chest X-ray images. *Tuberculosis, 136*, Article 102234. https://doi.org/10.1016/j.tube.2022.102234

Joyson, J. L., Soundar, K. R., & Nancy, P. (2026). A multi-modal fusion-based deep learning with finetuned LLaMA 3 for lung disease diagnosis using PACS radiology reports. *Biomedical Signal Processing and Control, 117*, Article 109466.

Khan, S., Siddiqui, F., & Ahad, M. A. (2025). Bridging efficiency and interpretability: Explainable AI for multi-classification of pulmonary diseases utilising modified lightweight CNNs. *Image and Vision Computing, 158*, Article 105553. https://doi.org/10.1016/j.imavis.2025.105553

Khate, K., & Neelima, A. (2026). Deep feature fusion and selection for lung disease classification. *Soft Computing, 30*(4), 2609–2624. https://doi.org/10.1007/s00500-025-11040-w

Lai, Y. R., Tsai, P. C., Yao, C. Y., & Ruan, S. J. (2017). Improved local histogram equalisation with gradient-based weighting process for edge preservation. *Multimedia Tools and Applications, 76*(1), 1585–1613. https://doi.org/10.1007/s11042-015-3147-7

Lakshmi, G. G., Chinnasamy, P., & Nagaraj, P. (2026). Optimal NAS-MoE: Optimised NASNet and MoE model for lung adenocarcinoma classification with explainable AI. *Biomedical Signal Processing and Control, 113*, Article 109128. https://doi.org/10.1016/j.bspc.2025.109128

Parthasarathy, V., & Saravanan, S. (2024). Computer-aided diagnosis using Harris Hawks optimiser with deep learning for pneumonia detection on chest X-ray images. *International Journal of Information Technology, 16*(3), 1677–1683. https://doi.org/10.1007/s41870-023-01700-1

Poloju, N., & Rajaram, A. (2025). Hybrid technique for lung disease classification based on machine learning and optimisation using X-ray images. *Multimedia Tools and Applications, 84*(21), 23531–23553. https://doi.org/10.1007/s11042-024-19959-2

Prasad, U., Chakravarty, S., & Mahto, G. (2024). Lung cancer detection and classification using deep neural network based on hybrid metaheuristic algorithm. *Soft Computing, 28*(15–16), 8579–8602. https://doi.org/10.1007/s00500-023-08845-y

Priya, A., & Bharathi, P. S. (2025). SE-ResNeXt-50-CNN: A deep learning model for lung cancer classification. *Applied Soft Computing, 171*, Article 112696. https://doi.org/10.1016/j.asoc.2025.112696

Shivwanshi, R. R., & Nirala, N. S. (2025). A hybrid AI method for lung cancer classification using explainable AI techniques. *Physica Medica, 134*, Article 104985. https://doi.org/10.1016/j.ejmp.2025.104985

Yi, S., Qin, S., She, F., & Shao, D. (2025). BSD: A multi-task framework for pulmonary disease classification using deep learning. *Expert Systems with Applications, 259*, Article 125355. https://doi.org/10.1016/j.eswa.2024.125355

Xie, J., Xie, L., Xiao, T., Zhu, R., Hu, C., Qian, K., & Zhang, C. (2026). DiemaNet: An automated detection model for three typical respiratory diseases. *Biomedical Signal Processing and Control, 112*, Article 108504. https://doi.org/10.1016/j.bspc.2025.108504

Zhu, J., Al-Qaness, M. A. A., Al-Alimi, D., Tao, H. Z., & Alsamhi, S. H. (2025). Multiple lung diseases detection using advanced deep learning model with attention mechanisms and upsampling features. *Engineering Applications of Artificial Intelligence, 156*, Article 111038. https://doi.org/10.1016/j.engappai.2025.111038