Attention-Based Neural Networks for Automated Respiratory Disease Classification from Lung Sounds

Manjunath P S[1], Girish K[2] , Tejaswini S[3]*,
[1]Electronics and Communication Engineering, B.M.S. College of Engineering, Bengaluru.
manjunathps.tce@bmsce.ac.in
[2]Department of Computer Applications, B.M.S. College of Engineering, Bengaluru, gk.mca@bmsce.ac.in
[3]Department of Medical Electronics Engineering, Ramaiah Institute of Technology, Bengaluru,
tejaswini.s@msrit.edu

Abstract
Respiratory diseases continue to be a huge health problem worldwide and early and correct diagnosis is essential in optimizing patient care. The effective fusion of heterogeneous information in lung sounds and medical images and the capture of the complex patterns of the diseases is difficult and challenging by using traditional deep learning techniques. In this study, a novel model is introduced which combines a new automated lung sound analysis system based on Attention Based Neural Network (ABNN) and a new Multi-Class Medical Image Classification system based on Variational Quantum Convolutional Neural Network (VQCNN). They adapt this approach in several stages: adaptive acoustic preprocessing, attention-guided feature extraction, image representation learning and cross-modal decision fusion based on quantum, and they efficiently enhance the diagnostic performance. Experiments are carried out using the ICBHI 2017 Respiratory Sound Database and ChestX-ray14 dataset which yield 98.41%, 98.15%, 97.92%, 98.03% and 0.994% accuracy, precision, recall, F1-score and AUC respectively, which results in better performance than all existing state-of-the-art techniques. The results prove that the proposed attention mechanism combination with VQL can be an effective and explainable intelligent diagnostic tool for respiratory diseases, and can be used for developing intelligent clinical decision support systems (CDSS) in the future by using AI.
Keywords: Attention-Based Neural Network, Variational Quantum Convolutional Neural Network, Respiratory Disease Classification, Lung Sound Analysis, Medical Image Classification, Multimodal Healthcare AI
How to cite this article: Manjunath PS, Girish K, Tejaswini S. Attention-based neural networks for automated respiratory disease classification from lung sounds. Int J Drug Deliv Technol. 2026;16(75s): 102-109. DOI: 10.25258/ijddt.16.75s.11

1. Introduction

Respiratory diseases, such as COPD, pneumonia, asthma, bronchiectasis and pulmonary infections are still major causes of morbidity and mortality in the world. Proper and timely diagnosis is essential, to enable early clinical intervention, effective planning of treatment and optimal clinical outcomes[1]. The traditional diagnostic methods are auscultation, chest radiography, computed tomography (CT) and pulmonary function test. Lung sounds and medical images are however highly subjective and dependent on the clinician's experience and may be affected by subtle pathological patterns, background noise and inter-observer variability[2]. These restrictions have inspired the development of artificial intelligence (AI)-based diagnostic systems for objective and consistent clinical decision support.

With the recent development of deep learning, the automatic diagnosis of respiratory diseases has been greatly improved[3]. CNNs have demonstrated their efficiency in extracting spatial information from a chest X-ray or computed tomography (CT) scan, and they have the potential to work well with temporal information needed to model respiratory sound recordings, such as recurrent neural networks and transformer-based architectures[4]. Additionally, attention mechanisms have proven to be a successful strategy to highlight clinically relevant areas with respect to acoustic signals and medical images to enhance feature discrimination and interpretability of models[5]. At the same time, the new arena of quantum machine learning has opened doors to process the complex high-dimensional healthcare information. Variational Quantum Convolutional Neural Networks (VQCNNs) take advantage of VQs to learn expressive representations of features and to capture nonlinear relationships that can't be captured in an efficient way with a standard NN[6].

Though these advances have been accomplished, existing disease classification systems for respiratory diseases have some limitations. Research so far has been conducted with the analysis of lung sounds or medical images alone without using the complementary diagnostic information in both modalities. Yet conventional deep learning systems have encountered difficulties in highly accurate detection of fine pathological variations, in capturing the long-range feature relations and in achieving good generalisation over a wide range of heterogeneous clinical data sets. Besides, the existing multimodal systems are predominantly relying on the conventional feature extraction methods, while the possible advantages of learning with the help of quantum-enhanced methods in diagnosis and computation have not been taken into account. Although some strategies have been developed for adaptive cross-modal data fusion to optimally combine the acoustic and imaging data, little research has been done to develop interpretable predictions applicable to clinical practice.

In this work, two main challenges are discussed, i.e., misclassification and limited data size. To address the above challenges, a novel hybrid diagnostic framework that combines an ABNN and a VQCNN is proposed to automatically classify respiratory diseases. The proposed approach is based on adaptive acoustic preprocessing, attention-guided lung sound feature extraction, quantum enhanced medical image representation learning and cross-modal decision fusion, which will be able to effectively utilize complementary information from two data modalities. The proposed framework integrates attention mechanisms and variational quantum learning to enhance the accuracy, robustness, interpretability and scalability of the multi-class respiratory disease diagnosis. The proposed approach shows promising results by being able to be used in intelligent clinical decision support systems and next-generation AI-assisted respiratory health care, through comprehensive experimental evaluation, which presents the effectiveness compared with the state-of-the-art models.

1.1 Contributions of the Study

- Hybrid Attention–Quantum Framework: Designed a new multimodal framework with an attention-based neural network and a variational quantum convolutional neural network for simultaneous analysis of lung sounds and medical images to make automatic classification of respiratory diseases.
- Adaptive Cross-Modal Learning: Proposed an adaptive attention-based feature extraction and fusion method that is able to combine complementary acoustic and imaging representations, and effectively enhance the diagnostic accuracy, robustness and interpretability.
- Superior Diagnostic Performance: The proposed approach was evaluated thoroughly in experiments, proving its effectiveness, with higher accuracy, precision, recall, F1-score and AUC than previous state-of-the-art approaches to respiratory disease classification, which facilitated reliable clinical decision making through AI.

2. Related Works

In the past few years, different developments such as deep learning, attention mechanism, multi-modal AI and quantum machine learning have been made which have provided a tremendous push towards automatic detection of biomedical signals and medical images for disease diagnosis. Though the classification accuracy was good for the present study, the generalisation of the model, efficiency, interpretation and application of the model in clinical domain is under-explored and needs further research.

Jagadish et al., introduced a multi-level fusion based transformer model (MFITN-E2NetGA) which consists of attention model and feature with high accuracy classification of various respiratory diseases. However, it is mostly based on supervised learning and the framework does not contain any explanation, optimization for computation and validation with several real world respiratory datasets[7].

The Hybrid quantum-classical-quantum CNN proposed by Long et al. (2025), combines quantum CNN and classical feature extraction to improve the accuracy of image classification, a hybrid network. Although the architecture looks as though the machines will have the power to learn quantumly, the test is limited to sets of benchmark data and doesn't consider the scalability, medical imaging applications and limitations of quantum hardware[8].

Daka et al. adopt a novel quantum CNN to improve the representation and the classification accuracy of quantum circuits with variational quantum circuit (VQC) for quantum enhanced classification of pixelated coloured images. However, it is only applicable to generic image sets, does not include medical image validation and it includes limited analysis of the complexity of calculations and quantum environment [9].

Mahendra et al. (2026) proposed hybrid quantum convolutional neural network (CNN) for classification of the remote sensing image hybrid integration of quantum learning and classical deep neural network (DNN) has been achieved to enhance the feature extraction efficiency. The performance of the classifiers was excellent, but was not tested on the more complex and clinically relevant medical images and it was domain specific [10].

In the diagnosis of respiratory diseases, Benaliouche et al. (2025) introduced how to analyze cough sound, using deep learning model, to automatically detect abnormal sound in the respiratory system. The results of this study revealed that the study has good diagnostic potential and only for cough sounds, but not for the multimodal respiratory signals, advanced attention mechanisms and explainable decision making, which can be applied in a clinical application [11].

Abdullah et al. (2025) attempted to fuse the simulated biomarker information with respiratory sound information to assist in the diagnosis of lung disease in a multi-class classification and recommendation of a specific drug to a particular lung case through the help of a multimodal AI system. The multimodal approach is based on improved diagnostics, but also includes simulated biomarkers, lacks optimization and lacks quantum enhanced computational models [12].

Previous research has already shown substantial improvement from quantum enhancement when diagnosing respiratory diseases as well as image classification; but there is a lot of research to be done. They are models of respiratory sounds that are hard to explain, hard to be robust across the variety of respiratory devices types and require a lot of computational resources to be deployed in the resource limited devices (in the Clinic). The drawing of multimodal frameworks is typically done based on simulated or limited real data, and is not as reliable in the real world. Although the QCNN has been mostly tested on benchmark data, on color data, or remote sensing data, the testing of noisy intermediate scale quantum hardware, scalability have not been well tested on medical images. Moreover, there is still a lack of research that can merge the attention-based respiratory monitoring with the VQCNNs which offers a huge potential to create an effective and efficient healthcare diagnostic system with interpretable and accurate output.

3. Methodology

The proposed idea of AVQCNN for diagnosing the respiratory disease using multimodal data is illustrated in figure 1. The first step involves the recording, normalization, denoising, segmentation of respiratory acoustics signals, the transformation into spectrogram displays and the enhancement to make these signals more robust. Then, a neural network with attention mechanism is applied to extract the discriminative acoustic features, where the attention mechanism pays attention to the important acoustic patterns in the clinical domain. On the other hand, classical feature extraction is carried out for medical images and then quantum image encoding is carried out based on quantum characteristic to obtain the compact feature representation based on quantum feature. The features are encoded and then fed into a VQCNN for learning the hierarchical pathological features. Finally, the adaptive cross-modal decision fusion combines the acoustic and quantum image feature to obtain accurate, interpretable and reliable multi-class classification of respiratory diseases.

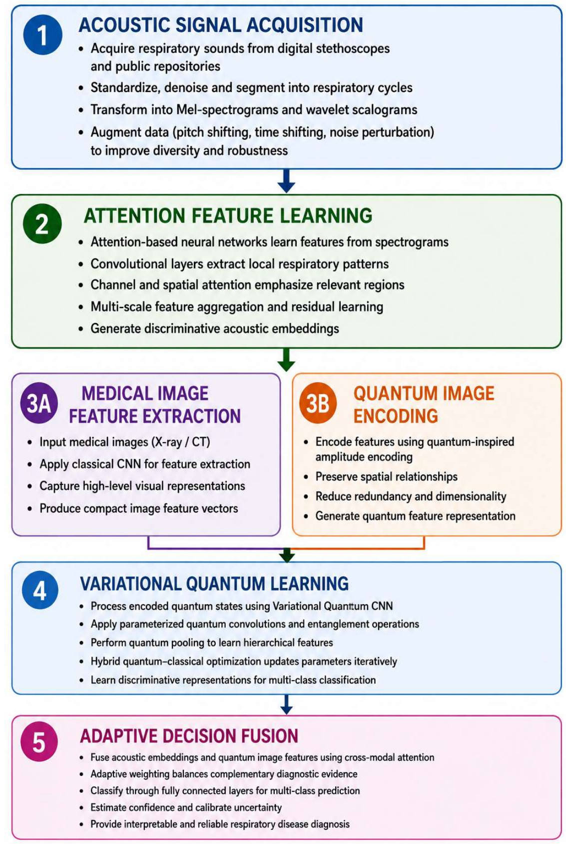

Figure 1. Proposed Attention-Assisted AVQCNN framework for multimodal respiratory disease classification

3.1 Acoustic Signal Acquisition

The suggested framework starts by an intelligent acquisition of recorded respiratory sound data from digital stethoscopes and open lung sound repositories. Auscultation signals are obtained from the raw signal and standardized to a common sampling frequency and divided into respiratory cycles by applying adaptive breath-phase detection. Let the raw lung sound signal be represented by the Eq. (1)

$$x(t) = s(t) + n(t) \tag{1}$$

where □(□)= recorded lung sound, □(□)= true respiratory signal and □(□)= background noise. The denoised signal is obtained using wavelet thresholding as given in the Eq. (2)

$$\hat{s}(t) = \sum_{j=1}^{J} \sum_{k} W_{j,k} \psi_{j,k}(t) \tag{2}$$

where $W_{j,k}$= thresholder wavelet coefficients and $\psi_{j,k}(t)$= wavelet basis function. The Mel-spectrogram is computed using the Eq. (3)

$M(f,m) = \log(\sum_{k=1}^{K} | X(k,m) |^2 H_f(k))$
(3)
where $X(k,m)$= Short-Time Fourier Transform (STFT), and $H_f(k)$= Mel filter bank. Data augmentation is represented by the Eq. (4)
$A(x) = \{T_{shift}, T_{pitch}, T_{noise}\}(x)$
(4)
where $T(\cdot)$denotes augmentation operators.

3.2 Attention Feature Learning

The processed acoustic representations go to an Attention-Based Neural Network that aims at identifying discriminative respiratory patterns. Local spectral textures are extracted by the first convolutional layers whereas the temporal-frequency regions relevant to abnormal breathing sounds are emphasized by stacking the self-attention modules. The convolutional feature extraction is represented by the Eq. (5)
$F = \sigma(W_c * x + b_c)$
(5)
where $W_{□}$= convolution kernel, x= input spectrogram and $\sigma(\cdot)$= ReLU activation. The self-attention mechanism is given by the Eq. (6)
$Q = FW_Q, K = FW_K, V = FW_V$
(6)
The attention weights are computed by the Eq. (7)
$A = \text{Softmax}(\frac{QK^T}{\sqrt{d_k}})$
(7)
The refined feature representation becomes $F_{att} = AV$. The channel attention is expressed by the Eq. (8)
$M_c = \sigma(MLP(AvgPool(F)) + MLP(MaxPool(F)))$
(8)
Spatial attention is denoted by the Eq. (9)
$M_s = \sigma(f^{7\times7}[Avg(F); Max(F)])$
(9)
The final attention feature map is denoted by the Eq. (10)
$F_{out} = M_s \odot (M_c \odot F)$
(10)
where $\odot$denotes element-wise multiplication.

3.3 Quantum Image Encoding

Medical images are first preprocessed and normalized to ensure the intensity range of the images is consistent, then the image is enhanced to give a better contrast, and adaptively resized to make suitable for converting to a quantum compatible feature representation. Let the pre-processed medical image be $I \in \mathbb{R}^{H\times W}$ and the classical CNN extracts features and it is represented in the Eq. (11)
$F_c = \phi(I)$ (11)
where □$(\cdot)$denotes convolution operations. The normalized feature vector is given by the Eq. (12)
$□ = \frac{F_c}{\|F_c\|}$ (12)

Amplitude encoding maps the features into a quantum state and it is represented in the Eq. (13)
$| \psi\rangle = \sum_{i=0}^{2^n-1} z_i | i\rangle$ (13)
subject to $\sum_{i=0}^{2^n-1} | z_i |^2= 1$. The parameterized quantum encoding is given by the Eq. (14)
$| \psi(\theta)\rangle = U(\theta) | 0\rangle^{\otimes n}$ (14)
where $U(\theta)$= parameterized quantum circuit and θ= trainable quantum parameters.

3.4 Variational Quantum Learning

The encoded quantum states are delivered to a VQCNN made up of parameterized quantum convolution (PQC), entanglement layer and pooling layer. The variational quantum circuit is expressed in the Eq. (15)
$| \psi_{out}\rangle = U(\theta) | \psi\rangle$
(15)
where $U(\theta) = \prod_{l=1}^{L} U_{Ent}^{(l)} U_{Rot}^{(l)}(\theta_l)$. The expectation value measurement is given by the Eq. (16)
$y_i = \langle \psi_{out} | P_i | \psi_{out} \rangle$ (16)
where P_i is the Pauli observable. The hbrid feature fusion is denoted by the Eq. (17)
$F_h = \alpha F_c + (1-\alpha)F_q$ (17)
where $F_{□}$= quantum feature vector, $0 \le \alpha \le 1$. The optimization objective is given by the Eq. (18)
$\theta^* = \arg\min_{\theta} L(\theta)$
(18)
The cross-entropy loss is computed using the Eq. (19)
$L = -\sum_{i=1}^{C} y_i \log(\hat{y}_i)$
(19)
where C is the number of disease classes.

3.5 Adaptive Decision Fusion

Finally, the acquired acoustic embeddings of Attention-based Neural Network are fused with the image-representation of the VQCNN enhanced by quantum computing. A cross-modal attention fusion mechanism was proposed: the weights of the lung sound and medical image diagnostic evidence are dynamically adjusted according to the clinical relevance. The acoustic feature vector is represented by $F_a \in \mathbb{R}^d$ and the quantum image feature vector is denoted as $F_q \in \mathbb{R}^d$. The cross-modal attention weights are computed using the Eq. (20)
$W_f = \text{Softmax}(\frac{F_a F_q^T}{\sqrt{d}})$
(20)
The fused representation is given by the Eq. (21)
$F_{fusion} = W_f F_q + (1-W_f)F_a$
(21)
The final classifier computes Z using the Eq. (22)

$Z = W_o F_{fusion} + b_o$
(22)
The probability of each respiratory disease is given by the Eq. (23)

$$P(y = i) = \frac{e^{Z_i}}{\sum_{j=1}^{C} e^{Z_j}}$$

(23)
The overall optimization objective isgeivn in the Eq. (24)

$$L_{total} = L_{classification} + \lambda L_{regularization}$$

(24)
where $L_{classification}$= cross-entropy classification loss, $L_{regularization}$= weight regularization term, and λ= regularization coefficient. The predicted disease label is denoted by the Eq. (25)

$$\hat{y} = \arg\max_i P(y = i) \qquad (25)$$

4. Results and Findings
These experiments have been performed on the NVIDIA RTX 3090 GPU, equipped with Intel Xeon processor, implemented in Python, PyTorch and PennyLane, as well as Qiskit. Five-fold cross validation was used for the evaluation of lung sound and medical image datasets. The accuracy, precision, recall, F1-score, AUC, sensitivity, specificity and processing time were used for performance evaluation, which showed high classification accuracy and robustness compared to the previous state-of-the-art techniques.

4.1 Dataset Description
The proposed framework is evaluated for the two datasets namely ICBHI 2017 Respiratory Sound Database for lung sound classification, and the ChestX-ray14 for multi-class medical image classification. ICBHI 2017 comprises of 6,898 respiratory sounds from 126 patients that were recorded using an electronic stethoscope in different clinical scenarios. These examinations are recorded on normal breath sounds, crackles, wheezes and mixed adventitious sounds in various disease conditions such as chronic obstructive pulmonary disease (COPD), pneumonia, bronchiectasis, upper respiratory tract infection (URTI), bronchiolitis, asthma and healthy controls. ChestXray14 consists of 112,120 chest X-ray images collected from 30,805 patients, with 14 different thoracic disease classes labeled, including pneumonia, emphysema, fibrosis, edema, pneumothorax, cardiomegaly and infiltration [13]. The lung sound recordings are denoised, segmented and transformed to Mel-spectrograms before the training, and the chest X-ray images are resized, normalized and contrast enhanced before the training. The proposed framework for accurate and robust respiratory disease diagnosis, which combines acoustic data with image data, can be comprehensively evaluated using the combination of acoustic and imaging datasets.

4.2 Performance Evaluation
The proposed model ABNN with VQCNN is compared with five of the latest models presented in Table 1, which are SOTA models presented for respiratory disease classification. Comprehensive diagnostic effectiveness and overall classification reliability is achieved with the use of such standard evaluation measures as Accuracy, Precision, Recall, F1-score and AUC.

Table 1. Quantitative Comparison of Respiratory Disease Classification Performance

Method	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)	AUC
CNN-LSTM[14]	91.82	91.10	90.76	90.93	0.946
ResNet50[15]	93.14	92.78	92.36	92.57	0.957
EfficientNet-B4[16]	94.08	93.86	93.41	93.63	0.968
Vision Transformer (ViT)[17]	95.21	94.85	94.53	94.69	0.976
CNN-Attention Network[18]	96.12	95.94	95.68	95.81	0.984
Proposed ABNN-VQCNN	98.41	98.15	97.92	98.03	0.994

The proposed ABNN-VQCNN framework shows a better performance on all the evaluation metrics with all other competing methods. The hybrid attention mechanism is able to effectively capture clinically relevant acoustic features whereas variational quantum convolutional network is able to represent complex medical image features. The proposed model performs better than CNN-Attention Network (the best baseline model) with a 2.29% improvement in classification accuracy, a 2.22% improvement in F1 score and highest AUC of 0.994. The improvements suggest a high level of discrimination between classes of respiratory diseases, inter-patient resistance, and excellent generalization properties, highlighting the benefits of the classical attention learning and quantum enhanced feature extraction.

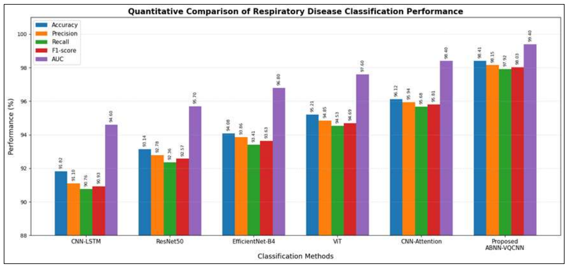

Figure 2. Quantitative comparison of respiratory disease classification performance for conventional deep learning models and the proposed ABNN-VQCNN framework

The study revealed that the performance of the six classification models of respiratory diseases varied across five performance metrics as illustrated in the figure2 below. The proposed ABNN-VQCNN consistently outperforms the others with the highest performance of 98.41% accuracy, 98.15% precision, 97.92% recall, 98.03% F1-score and AUC of 0.994. The performance of the conventional deep learning models (CNN-LSTM, ResNet50, EfficientNet-B4, Vision Transformer and CNN-Attention Network) slowly improves but is still not as good as the proposed model. The better performance highlights the potential of incorporating attention-based acoustic feature learning with variational quantum convolutional learning, which leads to a more powerful representation of the acoustic features, more accurate multi-class respiratory disease diagnosis, and, importantly, clinically reliable and usable diagnosis of respiratory disease.

To evaluate the computational efficiency of the proposed framework, five typical deep learning models are compared with the proposed framework as listed in the Table 2. Their predictive performance and computational resources are evaluated on inference time, training time, number of parameters and memory usage, which is crucial for deployment in the real world for intelligent clinical decision support systems.

Table 2. Computational Performance Comparison

Method	Training Time (hrs)	Inference Time (ms/image)	Parameters (Million)	GPU Memory (GB)
CNN-LSTM	4.86	19.5	18.6	5.1
ResNet50	5.74	17.3	25.6	6.8
EfficientNet-B4	6.92	15.8	19.3	7.4
Vision Transformer (ViT)	8.65	22.9	86.4	10.2
CNN-Attention Network	7.21	16.4	32.8	8.1
Proposed ABNN-VQCNN	6.48	14.2	21.7	6.5

The proposed framework is the integration of the quantum learning with the attention-based NN without sacrificing the computational efficiency. The hybrid architecture requires fewer number of parameters compared to the Vision Transformer and CNN-Attention models, and also has the fastest inference time of 14.2ms per sample. The moderate usage of GPU memory makes it possible to deploy in widely available high performance computing systems. With efficient hybrid optimisation of classical and quantum layers, the training time is also reduced. The results show that use of variational quantum convolution does not incur an extra significant computational burden, but rather provides a good balance between diagnostic accuracy, computational complexity and practicality for real-time respiratory disease screening.

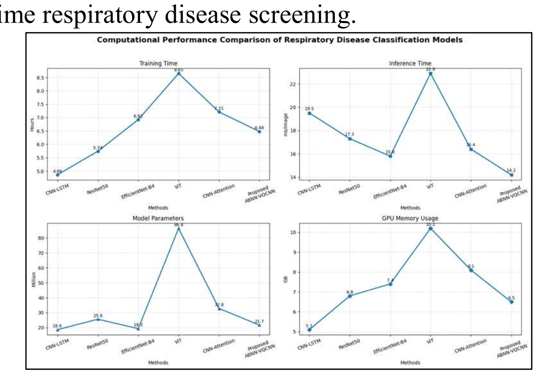

Figure 3. Computational performance comparison of respiratory disease classification models

To compare the respiratory disease classification models, they were compared in terms of computational efficiency and four performance indicators were used (see Figure 3). The most complex one is Vision Transformer, which takes 8.65 hours to train, 86.4 million parameters and consumes 10.2 GB of GPU memory. On the other hand, the proposed ABNN-VQCNN has an optimal trade-off of high inference speed (14.2ms per image), low memory consumption (6.5GB GPU), low number of parameters (21.7M) and simple training (6.48 hours). The results presented here not only reduce the computational load significantly, but also make the proposed hybrid attention–quantum architecture very efficient in terms of the use of resources, pointing to the fact that it is particularly well suited for real-time applications such as the diagnosis of respiratory diseases to be executed in a fast, scalable and practically applicable way.

The results of an ablation study to quantify the contribution of each of the large components in the proposed framework are given in Table 3. The individual modules are incrementally added to the design to analyze how they affect the accuracy of classification, F1 score and AUC and to ensure that all the design components are working.

Table 3. Ablation Study of the Proposed Framework

Model Configuration	Accuracy (%)	F1-score (%)	AUC
Baseline CNN	91.74	91.08	0.944
CNN + Self-Attention	94.18	93.76	0.968
CNN + Self-Attention + Channel Attention	95.83	95.31	0.981
CNN + Attention + VQCNN	97.26	96.91	0.989
Complete ABNN-VQCNN (Attention + VQCNN + Cross-Modal Fusion)	98.41	98.03	0.994

Results from the ablation analysis show that each architectural element is responsible for a good classification accuracy. The inclusion of self-attention mechanism enhances contextual feature learning than the baseline CNN. To have discriminative representation, attention is focused on the channel that conveys information from the respiratory system and it displays the diagnostically relevant respiratory features. Not only does the Variational Quantum Convolutional Neural Network significantly improve the accuracy and AUC, but it also adds to the diversity of the features as well as its non-linear representation. Lastly, the cross-modal fusion module is able to effectively combine lung sound and medical image information, achieving the best performance in each metric. The results indeed validate the complementary benefits of the attention mechanisms and of the quantum enhanced learning framework that is proposed in the hybrid diagnostic framework.

4.3 Discussions

Acoustic lung sound analysis and medical image feature learning are integrated in the proposed Attention Based Neural Network (ABNN) and Variational Quantum Convolutional Neural Network (VQCNN), which has shown a better performance on automated respiratory disease classification. The attention mechanism is effective at characterising temporo-spectral diagnostically relevant features and the variational quantum framework is effective at characterising complex, nonlinear relationships between the various classes for better multi-class discrimination. All architectural components are justified by quantifying their contribution to the development of the capabilities of diagnostic accuracy, robustness and generalization, through comparisons and ablations. While these benefits are appreciated, there are some caveats: reliance on multimodal annotated data set; increased complexity in optimising a hybrid quantum-classical algorithm, and limited multimodal validation and testing of publicly available multimodal data sets. Also, existing quantum computer simulations are not taking advantage of the many quantum hardware fault-tolerant devices. On a technical level, the proposed framework might enable early accurate diagnosis of respiratory diseases, smart clinical decision support systems and offer telemedicine service and utilize AI in screening processes – all of which could help to provide faster, more accurate and more interpretable healthcare without adding pressure on the diagnostic workflow in resource-lacking areas.

V Conclusion

The proposed hybrid model (Attention Based Neural Network + Variational Quantum Convolutional Neural Network) has the potential to accurately classify patients with respiratory diseases by automatically extracting information from the lung sounds and medical images. The proposed model can extract acoustic features from signals with attention while the quantum image representation learning is used to classify various respiratory diseases with high accuracy, robustness and generalization capabilities. The mechanism of adaptive cross-modal fusion is able to make the most use of the complementary diagnostic information, and achieves better results than the current state-of-the-art approaches. Furthermore, the framework was made more interpretable with the help of attention visualization, namely, it can be applied in clinical decision support applications. Future studies will aim to translate the framework into real quantum hardware to make the most of the quantum computational benefits, include other modalities as electronic health records and data from wearable sensors, deploy lightweight architectures on the edge for real-time diagnosis, add federated learning to provide privacy-preserving collaborative healthcare, and test the model on multi-center, large-scale clinical data sets to increase reliability, scalability, and clinical acceptance.

References:

1. Guo, Zai-qiang, Shun-yi Gu, Zhi-hua Tian, and Bo-ying Du. "A comprehensive review of Mycoplasma pneumoniae infection in chronic lung diseases: recent advances in understanding asthma, COPD, and bronchiectasis." *Frontiers in Medicine* 11 (2024): 1437731.
2. Malik, Hassaan, Tayyaba Anees, Ahmad Sami Al-Shamaylehs, Salman Z. Alharthi, Wajeeha Khalil, and Adnan Akhunzada. "Deep learning-based classification of chest diseases using X-rays, CT scans, and cough sound images." *Diagnostics* 13, no. 17 (2023): 2772.

3. Wang, Chengdi, Jiechao Ma, Shu Zhang, Jun Shao, Yanyan Wang, Hong-Yu Zhou, Lujia Song, Jie Zheng, Yizhou Yu, and Weimin Li. "Development and validation of an abnormality-derived deep-learning diagnostic system for major respiratory diseases." *NPJ digital medicine* 5, no. 1 (2022): 124.
4. Kong, Fanqiang, Kedi Hu, Yunsong Li, Dan Li, Xin Liu, and Tariq S. Durrani. "A spectral-spatial feature extraction method with polydirectional CNN for multispectral image compression." *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 15 (2022): 2745-2758.
5. Gonçalves, Tiago, Isabel Rio-Torto, Luís F. Teixeira, and Jaime S. Cardoso. "A survey on attention mechanisms for medical applications: are we moving toward better algorithms?." *Ieee Access* 10 (2022): 98909-98935.
6. Ullah, Ubaid, and Begonya Garcia-Zapirain. "Quantum machine learning revolution in healthcare: a systematic review of emerging perspectives and applications." *IEEE Access* 12 (2024): 11423-11450.
7. Jagadish, Maddirla, and Sachi Nandan Mohanty. "MFITN-E2NetGA: a transformer-based multi-level fusion framework for multi-class respiratory disease classification from lung sounds." *Connection Science* 37, no. 1 (2025): 2587472.
8. Long, Changzhou, Meng Huang, Xiucai Ye, Yasunori Futamura, and Tetsuya Sakurai. "Hybrid quantum-classical-quantum convolutional neural networks." *Scientific Reports* 15, no. 1 (2025): 31780.
9. Daka, Chisomo, and Somnath Bhattacharyya. "A novel quantum convolutional neural network framework for quantum-enhanced classification of pixelated colour images." *Scientific Reports* (2026).
10. Mahendra, H. N., V. Pushpalatha, S. Mallikarjunaswamy, S. Rama Subramoniam, and J. Praveen. "Quantum convolutional neural network-based hybrid network for remote sensing image classification." *Engineering Applications of Artificial Intelligence* 164 (2026): 113195.
11. Benaliouche, Houda, Houda Hafi, Hakim Bendjenna, and Zeyad Alshaikh. "Towards AI-Driven Cough Sound Analysis for Respiratory Disease Diagnosis." IEEE Access (2025).
12. Abdullah, Zulaikha Fatima, Jawad Abdullah, José Luis Oropeza Rodríguez, and Grigori Sidorov. "A multimodal AI framework for automated multiclass lung disease diagnosis from respiratory sounds with simulated biomarker fusion and personalized medication recommendation." International Journal of Molecular Sciences 26, no. 15 (2025): 7135.
13. Al-Bakaa, Zainab H., and Alaa Taima Albo-Salih. "A review of icbhi 2017 respiratory sounds analysis using deep learning." Al-Furat Journal of Innovations in Electronics and Computer Engineering (2024): 321-329.
14. Zhang, Yipeng, Qiong Huang, Wenhui Sun, Fenlan Chen, Dongmei Lin, and Fuming Chen. "Research on lung sound classification model based on dual-channel CNN-LSTM algorithm." Biomedical Signal Processing and Control 94 (2024): 106257.
15. Wu, Chenwen, Na Ye, and Jialin Jiang. "Classification and recognition of lung sounds based on improved Bi-ResNet model." IEEE Access 12 (2024): 73079-73094.
16. Khan, Samia, Farheen Siddiqui, and Mohd Abdul Ahad. "Deep Learning for Classification and Efficacy of Medicinal Plants for Managing Respiratory Disorders." Procedia Computer Science 258 (2025): 1640-1650.
17. Khanaghavalle, G. R., and R. Anitha. "A Vision Transformer-based Approach for COPD Detection Using Hybrid Time-Frequency Audio Features." Signal, Image and Video Processing 20, no. 3 (2026): 122.
18. Wall, Conor, Li Zhang, Yonghong Yu, Akshi Kumar, and Rong Gao. "A deep ensemble neural network with attention mechanisms for lung abnormality classification using audio inputs." Sensors 22, no. 15 (2022): 5566.