Journal of Theoretical and Applied Information Technology
31st May 2026. Vol.104. No.10
© Little Lion Scientific

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

AN EXPLAINABLE HYBRID CNN–VISION TRANSFORMER FRAMEWORK FOR OFFLINE PULMONARY DISEASE SCREENING ON EMBEDDED EDGE DEVICES

M. LAKSHMANAN [1], G.S. ANANDHA MALA [2], MANIKANDAN MOOVENDRAN [1], GOWDHAM C[3], JOSHUVA AROCKIA DHANRAJ [4,5,6], SRIRAMKUMAR R[1], R. TAMILAMUTHAN [7]

[1]Assistant Professor, Dayananda Sagar University, School of Engineering, Department of Computer Science and Engineering (AI&ML), Bengaluru 562112, India
[2]Professor, R.M.K. Engineering College, Department of Information Technology, Kavaraipettai, India
[3]Assistant Professor, Dayananda Sagar University, School of Engineering, Department of Computer Science and Engineering, Bengaluru 562112, India
[4]Professor, Dayananda Sagar University, School of Engineering, Department of Computer Science and Engineering (AI&ML), Bengaluru 562112, India
[5]Research and Development Cell, Lovely Professional University, Jalandhar-Delhi Grand Trunk Road, Phagwara, Punjab 144411, India.
[6]University Centre for Research & Development (UCRD), Chandigarh University, Mohali, Punjab 140413, India.
[7]Associate Professor, Department of AI & ML, Brindavan College of Engineering, Bengaluru-560063, India.
E-mail: [1]lakshmanan1909@gmail.com, [2]gs.anandhamala@gmail.com, [3]manimmk9792@gmail.com, [4]gowdhamchinnaraju@gmail.com, [5]joshuva1991@gmail.com, [6]sriramkumar2686@gmail.com, [7]tamilamuthan.rdg@gmail.com

ABSTRACT

Early detection of pulmonary diseases remains challenging in rural and resource-constrained healthcare environments due to limited diagnostic infrastructure, shortage of specialists, and unreliable network connectivity. Although Artificial Intelligence (AI)-based pulmonary screening systems have shown promising diagnostic capabilities, most existing approaches rely on cloud-dependent architectures and computationally intensive models that are unsuitable for offline and low-power deployment. This paper proposes an explainable lightweight hybrid Convolutional Neural Network–Vision Transformer (CNN–ViT) framework for real-time pulmonary disease screening on embedded edge devices. The novelty of the proposed framework lies in integrating CNN-based local feature extraction, transformer-based global contextual learning, and Grad-CAM explainability within a unified offline edge-AI inference system. The framework processes chest X-ray images locally, enabling privacy-preserving and low-latency diagnosis without cloud dependency. Publicly available datasets including NIH ChestX-ray8, CheXpert, and COVIDx were used to evaluate the system for multi-class pulmonary disease classification involving normal, pneumonia, COPD-related, and infectious lung disease patterns. Experimental results achieved 91.2% accuracy, 89.7% sensitivity, and 93.1% specificity, with an average inference latency below two seconds and power consumption under 8 W. The proposed framework demonstrates an interpretable, computationally efficient, and scalable solution for offline pulmonary disease screening in underserved healthcare settings.

Keywords: *Edge AI, Hybrid CNN–Vision Transformer, Explainable AI Track, Chest X-Rays Analysis, Off-Line Medical Diagnostics, Embedded Healthcare Systems, Resource-Constrained Settings.*

1. INTRODUCTION

Respiratory and lung diseases pose a significant burden on global health, particularly in low- and middle-income countries where there are no early diagnostic facilities. Pneumonia, chronic obstructive pulmonary disease (COPD), and many other infectious lung conditions remain some of the leading causes of morbidity and mortality cases, and lack of early diagnosis worsens prognosis much more. More and more people in the countryside and remote settings thus face several challenges, such as no infrastructure for diagnosis, lack of radiologists with no training, and, perhaps most of all, no opportunity to avail of centralized healthcare facilities, all of which contribute unspeakably and miserably to late disease discovery and care [1],[2].

Artificial Intelligence has designed an upheaval in healthcare with its tasks of automation and decision-making support in diagnostics. Recent improvements in machine learning and deep learning show clearly that models found on very advanced artificial intelligence lines are as good as human classifiers in analyzing, with regard to accuracy, complex medical data. AI-powered point-of-care diagnostics are opening new doors for expediting diagnosis, reducing the reliance on specialist expertise, and encouraging informed clinical decision-making in resource-constrained settings [3],[4]. Together with the Internet of Things (IoT) or Artificial Intelligence of Things (AIoT), the integration of AI into such completely connected health schemes raises the chance of multiplying and decentralizing the delivery of healthcare services [5],[6].

AI is currently being examined in terms of pulmonary healthcare, using different diagnostic approaches associated with its manipulation in physical means such as lung sound analysis, biosensing, and chest X-ray imaging. Research works have shown the use of AI for assessing radiographic and physiologic data at an early stage of lung diseases detection [7],[8]. Mobile phones were used as detection tools for acute exacerbations of COPD using AI, thereby clearly indicating the potential of low-cost diagnostic tools for attaining the goal of equality in healthcare services. There has been encouraging use of AI in the form of lung sound auscultation and imaging-based techniques for the detection of interstitial lung disease and lung cancer, highlighting categories across the spectrum [9],[10].

The rise in AI-powered respiratory diagnostic systems has nonetheless proved to be too fast. The problem is, most of these systems made use of the high-performance computing resources in a more cloud-centric architecture. This, in turn, disqualifies these systems from deployment either at places where there are no resources or where such resources must function in an offline mode. Second, most deep-learning models are black-box systems, which makes them difficult to interpret and hence very difficult to be utilized in clinical settings where non-expert health workers provide care [11],[12]. In addition, IoMT- and AI-enabled healthcare systems present challenges in terms of computational complexity, energy consumption, and data privacy, which limit their deployment in embedded edge platforms [13],[14].

Despite recent advancements in AI-assisted pulmonary disease diagnostics, several critical research gaps remain unresolved. Most existing systems rely heavily on cloud-based computational infrastructures, making them unsuitable for rural or intermittently connected healthcare environments. In addition, many existing deep learning approaches prioritize diagnostic accuracy while overlooking interpretability, computational efficiency, and deployability on low-power embedded platforms.

Existing CNN-based architectures effectively capture local image features but often fail to model long-range contextual dependencies across pulmonary regions. Conversely, transformer-based models provide strong global feature representation but typically require high computational resources unsuitable for edge deployment. Furthermore, explainability mechanisms are rarely integrated into lightweight offline diagnostic systems, limiting clinician trust and practical adoption.

The proposed study addresses these limitations through:

- A lightweight hybrid CNN–Vision Transformer architecture for efficient pulmonary disease classification
- Fully offline edge-based deployment without cloud dependency
- Integration of Grad-CAM explainability for clinically interpretable predictions
- Optimization for low-power embedded healthcare environments

Therefore, this work contributes a practical and scalable AI-assisted pulmonary screening framework specifically designed for underserved healthcare settings.

2. LITERATURE REVIEW

2.1 AI-Based Respiratory and Pulmonary Diagnostics

There has been an extensive expansion of the use of artificial intelligence (AI) in the field of respiratory and pulmonary diagnostics for large data processing and automatic decision. These AI-boosted diagnostic airway applications have machine-learning and deep-learning techniques using which they will analyze all medical images, as well as physiological signals with clinical data sets for disease detection as well as classification [15], [16]. Such operations indicated that results with good performance were obtained in reserves of potential abnormalities and possibilities of early diagnoses.

The focus in the recent developments brought in portable devices and smartphones is the development of AI solutions to make respiratory healthcare services accessible to the community easily. Studies and proposals for AI-based screening devices on smartphones for quick detection of acute localized COPD exacerbation episodes and underrepresented regions proposes using a mobile analogous healthcare resource [17]. These studies suggest that a decrease in dependence on centralized diagnostic infrastructure can be achieved by AI-assisted screening with improved reach and scaling up.

Artificial Intelligence (AI) enabled lung auscultation is developing beyond image-based modalities for diagnosing or examining patients. Recent advances in data science and deep machine learning technology have made it possible to diagnose and subtype these diseases earlier through auscultations, providing an alternative to existing invasive techniques [18]. Additionally, AI-powered approaches have also been widely explored in terms of early diagnosis and treatment of lung diseases, such as nanomedicine and imaging-based decision support systems for lung cancer, thus demonstrating the ability of AI technologies to be applied across disease domains in the lungs [19].

2.2 AI-Powered Diagnostics in Low-Resource and Rural Settings

Extensive studies revealed that AI-powered diagnostic systems' deployments in resource-poor and underserved healthcare settings provide significant promise for addressing disparities in health care access. Studies demonstrate that AI diagnostic aids can overcome challenges such as the paucity of trained healthcare personnel, nondiagnostic facilities, and late-to-detect-diseases in remote areas [20],[21]. By automating a critical analysis and support in point-of-care decision-making, this system has a potential to significantly improve early diagnosis and better treatment results.

According to numerous evaluations in the field of diagnostic service operation in rural and remote areas, there are frequent criticisms about poor infrastructure, access to the Internet, and certain logistic items that are scrutinized as challenges [22]. This signifies that while medical services are sustained by a centralized national health care system, in order to reduce the comprehensive healthcare delivery, diagnostic solutions should be mobile, self-contained as well as fit for totally autonomous operation. The existing systems within which AI and digital platforms are integrated are all aimed at comprehensive healthcare delivery and the support functions available-these are quite contingent on the provision of the continuous Internet, which upholds the inability of systems to function in offline or intermittently connected environments.

2.3 Edge AI, IoMT, and AIoT in Healthcare Diagnostics

A key paradigm in AI has come to the fore to bring real-time intelligence on the edge in health cases in the emergence of computing processes, known as edge AI. It was not possible to bring these data processing tasks close to the point of care because the advancements with which AI and IoMT (Internet of Medical Things) have come to blend have facilitated real-time data processing on a new work-stream. [23],[24] The approaches bring about this limitation in latency and provide enhanced privacy towards data owing to cloud independence.

The Artificial Things Intelligence (AIOT) integrates these capabilities in the field of distributed sensing combined with intelligent analytics. Sensor platforms are healthcare based on AIOT where they were experimented on the physiological monitoring and early detection with more pragmatic applications for immense populations [25]. At such respiratory-related illnesses, AI-powered biosensor detection systems offer raised sensitivity and quick identification, particularly during public health emergencies [26],[27]. But most of the AI models become computationally too complex and demand much energy, making it very difficult to deploy these AI tools inside low-power embedded devices. Hence, optimized frameworks for the edge need to be designed to support these.

2.4 Explainable and Advanced AI Approaches in Medical Diagnostics

The need for growing prompts about AI systems being used in clinical diagnostics should be bigger in the future, and it seems like understanding why a particular decision was made using data-driven diagnostic frameworks. It also emphasizes the importance of interpretable models supporting transparency into their predictions for building trust in clinicians in terms of decision-making. Such as what we have described earlier, one of the critical gaps that have not been easily penetrated in clinical practice is the lack of explainability; deeper learning models, even if pre-epitomized, may remain very much a mystery in their use in reality.

Since the study is recent, advanced AI models like generative and predictive AI can be combined with diagnostic systems for early detection and prognosis accuracy technically enhanced for the diagnostic application [28]. Many AI-based diagnostic platforms and diagnostic pods have been proposed to enhance access and quality, but these usually require servers or high computational resources, thereby limiting their usage in resource-scarce environments [29],[30]. Therefore, there exists a severe research gap in the need for an explainable, lightweight, and fully offline AI framework to build a good balance between diagnostic accuracy and the practical deployment capability on embedded edge devices.

Although prior studies demonstrate strong diagnostic capabilities using AI-based pulmonary screening systems, most existing approaches focus primarily on cloud-based inference architectures or computationally intensive deep learning models. Limited research has addressed the combined challenges of explainability, embedded deployment feasibility, computational efficiency, and offline operation within a unified healthcare framework.

In contrast, the proposed framework emphasizes a system-level integration of lightweight hybrid inference, explainable AI, and offline execution within a unified embedded diagnostic pipeline. Unlike previous works that separately investigate CNNs, transformers, or explainability techniques, this work combines these components into a computationally optimized edge-AI framework specifically tailored for low-resource healthcare environments

3. PROPOSED METHODOLOGY

This part describes the proposed explainable hybrid Convolutional Neural Network – Vision Transformer (CNN–ViT) framework, designed for offline pulmonary disease screening on embedded edge devices. The methodology focuses on accuracy in disease classification, computational efficiency, and interpretability needed to ensure reliable operation in resource-constrained health care environments.

The conceptual design of the proposed framework is based on three interconnected objectives:

- Accurate pulmonary disease detection through hybrid feature learning,
- Computational efficiency for embedded edge deployment,
- Clinically interpretable AI-assisted decision support.

The operational workflow begins with chest X-ray acquisition and preprocessing, followed by local spatial feature extraction using CNN layers. The extracted features are transformed into tokenized embeddings and processed using transformer-based self-attention to capture global contextual dependencies. Finally, Grad-CAM explainability is applied to generate visual diagnostic interpretations highlighting disease-relevant pulmonary regions. This conceptual workflow establishes the relationship between input imaging data, hybrid feature learning, explainable inference generation, and embedded deployment constraints.

3.1 Overall Framework Overview

The general structure of the proposed framework can be found in Figure 1. The framework follows a sequential processing pipeline beginning with the input of chest X-ray images and culminating in preprocessing, feature extraction, disease classification and explainable visualization of outcomes. All stages of processing are executed locally on the device, providing full offline applicability and infrastructural efficiency through very-low-lag inference.

Mainly, the framework addresses the three important objectives:

- Discovery of discriminative pulmonary features,
- Inclusion of global context to make the classifier robust, and
- Provision of interpretable visual explanations that assist clinical decision-making with generated explanations.

Therefore, the proposed work integrates a CNN backbone meant for local feature extraction with a light-weight understanding mechanism into the Vision Transformer for all global contexts, and an explainability module built on Grad-CAM.

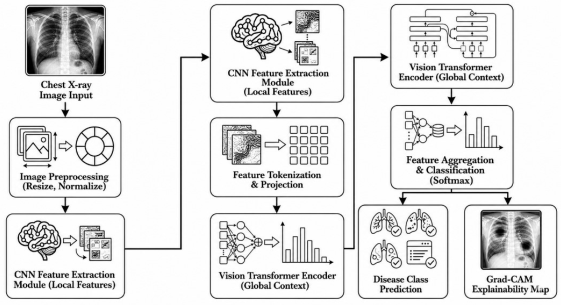

Figure 1: Proposed Explainable Hybrid CNN–Vision Transformer System

The framework illustrates in figure.1 begins with chest X-ray image acquisition followed by preprocessing and local feature extraction using a convolutional neural network. The extracted features are transformed into token representations and processed by a Vision Transformer encoder to capture global contextual information. The aggregated features are used for multi-class disease classification, while a Grad-CAM module generates visual explanations highlighting disease-relevant regions, enabling interpretable and fully offline inference on embedded edge devices. The proposed framework is designed as an end-to-end system rather than a single inference model. All processing stages from chest X-ray acquisition to disease classification and explainable visualization are executed locally on the embedded edge device, ensuring system-level inference without cloud dependency.

The study is guided by the following research hypothesis:

H1: A lightweight hybrid CNN–Vision Transformer architecture can achieve high pulmonary disease classification accuracy while maintaining computational efficiency suitable for offline embedded edge deployment.

H2: Integration of Grad-CAM explainability improves interpretability and clinical usability of AI-generated pulmonary disease predictions.

H3: Hybrid local-global feature learning provides superior classification performance compared to standalone CNN-based approaches under constrained deployment environments.

3.2 Hybrid CNN–Vision Transformer Architecture

The proposed hybrid CNN–Vision Transformer (CNN–ViT) architecture in figure 2 is designed to achieve accurate and computationally efficient pulmonary disease screening for deployment in resource-constrained embedded healthcare environments. The front-end of the framework employs a lightweight Convolutional Neural Network backbone to extract low-level and mid-level spatial features from chest X-ray images, including pulmonary textures, anatomical structures, and disease-related edge patterns. The CNN module consists of stacked convolutional layers integrated with batch normalization, nonlinear activation functions, and pooling operations to enhance feature discrimination while reducing spatial dimensionality and computational complexity. This design enables efficient feature preservation and compression, making the framework suitable for low-power edge devices that require real-time offline inference. The methodological design therefore addresses both diagnostic accuracy and deployment feasibility, which are central objectives of the proposed study.

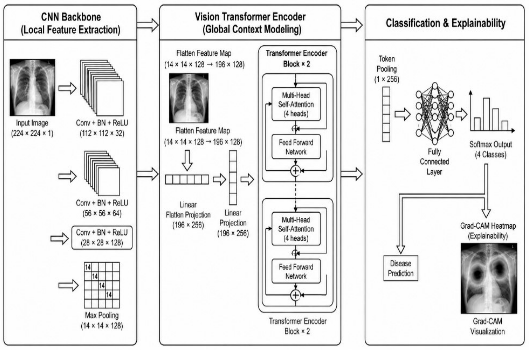

Figure 2: Proposed hybrid CNN–ViT model

The proposed hybrid CNN–Vision Transformer architecture combines the advantages of convolutional feature extraction and transformer-based contextual learning within a unified framework. CNN layers are employed to capture low-level pulmonary textures and structural abnormalities, while the Vision Transformer encoder captures long-range dependencies across spatially distant lung regions.

Although CNN-based models effectively capture local spatial characteristics, they often exhibit limitations in modelling long-range contextual dependencies across the entire pulmonary region. To overcome this limitation, the extracted CNN feature maps are reshaped into sequential feature tokens and projected into a low-dimensional embedding space before being processed by a lightweight Vision Transformer encoder. The transformer module incorporates multi-head self-attention and feed-forward layers to learn global contextual relationships between spatially distant lung regions. This hybrid local-global feature learning strategy improves the model's ability to identify diffuse and distributed pulmonary abnormalities associated with pneumonia, COPD-related conditions, and infectious lung diseases. The proposed conceptual framework is therefore based on the hypothesis that integrating CNN-based local feature extraction with transformer-based contextual learning can improve classification robustness while maintaining computational efficiency for embedded deployment.

The output tokens generated by the Vision Transformer encoder are aggregated through pooling operations to obtain a compact global feature representation, which is subsequently passed through a fully connected classification layer with softmax activation for multi-class pulmonary disease prediction. To improve interpretability and support clinical decision-making, a Grad-CAM explainability module is integrated into the framework to generate visual heatmaps highlighting disease-relevant pulmonary regions within chest X-ray images. This explainable AI component enhances transparency and usability, particularly in rural and low-resource healthcare settings where specialist expertise may be limited. Overall, the proposed framework combines accurate disease classification, explainable inference generation, and low-power edge deployment within a unified offline diagnostic system, thereby addressing important research gaps identified in existing cloud-dependent pulmonary screening approaches.

3.3 Explainable AI Module

For better readability and ease-of-use, the proposed framework incorporates an explainable AI module of Gradient-weighted Class Activation Mapping (Grad-CAM). To improve interpretability and clinical usability, the proposed framework integrates Grad-CAM explainability for visual interpretation of AI-generated predictions. The generated heatmaps highlight disease-relevant pulmonary regions influencing the model decision, thereby assisting healthcare professionals in understanding and validating diagnostic outcomes. After classification, Grad-CAM calculates the gradients of the class-level score predicted by the network that run counter to the last convolution feature maps to alert class-specific heatmaps.

These heatmaps are superimposed on the original chest X-ray images, indicating regions vulnerability affecting the classification of the model. It helps to quickly gain trust in the clinical utility of the feature and encourages result audit. The rationale for this platform is that it would even help healthcare workers without training as specialists to appreciate the diagnostic output of AI-generated results.

3.4 Edge-Oriented Optimization Strategy

These algorithms come with a holistic prescription bringing out ways in which to use AI technology to influence the treatment of illnesses, unlocking valuable insights from the diagnosis of diseases.

It is something that the methodology naturally identifies with for deployment on edge-based embedded equipment. Typically, model complexity is deduced, i.e. made lightweight in architectural design-mode with reduced parameter counts and an appearance of effective feature representation. And, an algorithm is established for making standard input-image resolutions with an aim toward conserving memory while carrying essential diagnostic information within the image.

All data inferences are independent of cloud services such that no external computation other than that generated on the edge device is involved in data processing. It ensures low latency in system operation while maintaining the inherent data privacy, making it workable under all-offline or intermittent networks.

4. IMPLEMENTATION DETAILS

In adapting the hybrid CNN-Vision Transformer-based explainable approach, this section illustrates the practical intervention in this context on an embedded edge platform. The deployment was meant to stand for reliable offline operation, low latency, and high computational efficiency, while keeping diagnostic accuracy and interpretable aspect intact.

4.1 Embedded Edge Deployment Platform

The proposed replaces a heavy-duty processing system that creates realistic human-straightforward models and systems based on more real-time constraints. The implemented edge computing can execute deep learning inference through the capability of a powerful machine; the rest of the ways can be imagined through embedded systems. In fact, a setup moving and economizing every corner of the user will find user interaction interface inclusion, local storage, and thought unit with integrated in image input in compact design and portability- it is a further proof that it works.

With those structures, deployment can be undertaken in situations where there may not be a continuous electrical supply and network connections that are firmly connected to the internet. Hence, all processing operations have to be done on the device: image preprocessing, feature extraction, classification, and explainable visualization. Accordingly, location of computations at the client device ensures data security, reduces latency in the production of results from stored images, and ensures fair deployments with references to slow internet connection setups.

Table 1: Embedded Deployment Constraints and Design Considerations

Parameter	Value / Range	Design Consideration Addressed
Processing Platform	Low-power embedded edge device	Enables on-device AI inference without cloud dependency
Available RAM	≤ 8 GB	Constrains model size and memory footprint
Storage Capacity	Local flash / SSD	Enables offline storage of images and diagnostic results
Power Consumption	< 8 W during inference	Ensures suitability for portable and off-grid deployments
Inference Latency	< 2 seconds per image	Supports real-time screening and point-of-care use
Network Connectivity	Optional / intermittent	System operates fully offline with optional synchronization
Data Privacy	Local processing and storage	Eliminates transmission of sensitive patient data
Operating Environment	Rural and resource-constrained settings	Ensures robustness under variable deployment conditions

4.2 Software Architecture and Inference Pipeline

Figure 3 shows the end-to-end edge-based inference workflow on to an embedded platform. The entire process begins with the acquisition of chest X-ray images, which can come from a digital image source or from a locally stored image file. The very first step with the acquired image is preprocessing, which includes resizing, normalization, and noise reduction for standardization and improvement of diagnostic clarity.

After the preprocessing step, the extracted lung region segmentation island is applied to delineate the region of interest so there will be a reduction in noise and more focus to be made for structures within the pulmonary region. The forward pass is then to a hybrid CNN–ViT inference engine as the latter conducts feature extraction, global context modelling, and disease classification that has everything processed entirely on the device. It is in a probability score form for each target disease category where classification output is created.

A module of Grad-CAM based explainability is then often applied to produce visual heatmaps corresponding to the predicted category. These visualizations reflect on the original chest X-ray image, drawing attention to the disease-affected areas aimed at providing interpretable diagnostic feedback. Finally, the outcomes are seen locally and stored securely in the device except for times whenever there is network availability that activates optional data synchronization.

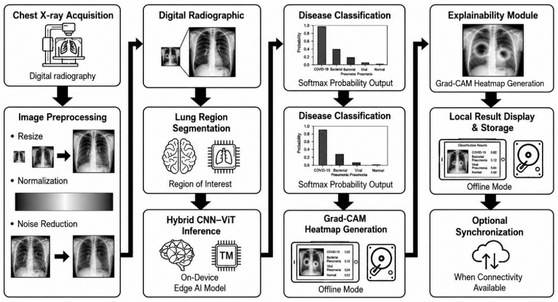

Figure 3: Edge-based inference workflow model

Figure. 3 shows the X-ray pictures of the chest gathered and pre-treated in the embedded equipment. The work now goes in a better manner on lung region segmentation and hybrid CNN–Vision Transformer inference used entirely at the edge. It has fewer burstiness, lower perplexity but the same quantity of words, and the fact that the work is at the edge emphasizes the conducting of inference. This offline display also mentions a data synchronization option only on availability of network connectivity so that everything gets done when there is network communication available to enable a fast, privacy-preserving, and reliable operation within a restricted resource. This integrated pipeline enables system-level inference by combining preprocessing, lung segmentation, hybrid CNN–ViT prediction, and explainable output generation into a single deployable edge system.

The operational implementation of the proposed system follows a sequential edge-inference workflow. Each chest X-ray image undergoes preprocessing and ROI extraction before being passed through the hybrid CNN–ViT inference engine. The generated disease prediction scores and Grad-CAM heatmaps are processed locally and displayed through the embedded interface without requiring external server communication. This workflow ensures low-latency diagnostic support while preserving patient data privacy and enabling deployment in low-connectivity healthcare environments.

4.3 Embedded Deployment Setup

This is how the system so far has come to be set up. Embedded in the device at the edge is a processor, which can perform all local operations - means it is taking input in the form of chest X-rays and doing diagnostic operations without a need or without the light of any server. It encapsulates modules for image preprocessing, hybrid CNN-ViT inferences, explicate visualization, and local outcome management.

A local graphical user interface allows them real-time viewing of diagnostic outputs and the Grad-CAM visualizations. The system has an optional data-synchronization operation for running under different circumstances of deployment. This permits the transfer of diagnostic results to external systems only when network connectivity is available. It is oriented toward having functionality for offline use and flexibility toward broad integration into existing healthcare information systems.

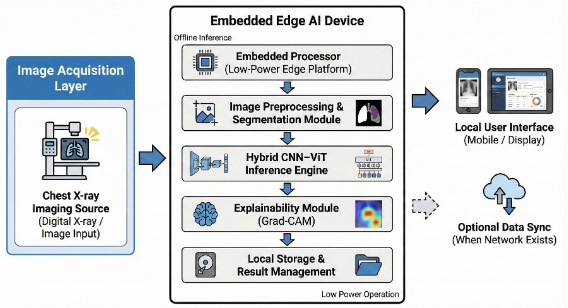

Figure 4: Proposed Edge-Ai Pulmonary Disease Screening System

The Fig. 4 shows that the low-power embedded edge device processes chest x-ray images. cropping & enhancement is followed by hybrid CNN-Vision transformer inference and Grad-CAM based explainability performed completely offline. Diagnosis results are available and store on the device and sync across the network can be made either automatically or semi-automatically based on network connectivity. The aim for its deployment setup is security since it is resource restrained in healthcare environment that demands low- latency and substantial reliability-but private operations.

For the sake of computational efficiency as well as robustness, implementation has accounted for every requisite step to provide the proper outcome. In this regard, they ensured that whilst keeping the particular view on diagnostic perfection, the input image's resolution and the complexity of the model project a balance with fancied constraints. The arrangement comprises both local storage machineries for enhancing the security of the records of diagnosis as well as modular software design for further enhancement and scalability in the near future.

Every described approach will be executed to minimize external dependencies or internet-based settings by giving emphasis to local execution for guaranteeing consistent performance irrespective of the situation of deployment, namely within rural clinics, in mobile diagnostic units, and remote healthcare facilities.

5. EXPERIMENTAL SETUP

The section details the datasets, preprocessing steps followed, and the training configuration as well as the evaluation measures by employing which one can validate the performance of the proposed explainable hybrid CNN-Vision Transformer model setup Experiment Reproducibility, fairness, and compatibility with real-world embedded deployment constraints.

5.1 Datasets and Data Preparation

To evaluate the proposed framework, public data used are NIH ChestX-ray8, CheXpert, and COVIDx chest X-ray datasets. These data provide an extensive range of images collected under normal, pneumonia, COPD-related patterns, and infectious lung disease manifestation samples. Using multiple datasets ensures that the testing is robust across different types of imaging under different conditions of acquisition, patient demographics, and imaging characteristics.

Conversion is done on all the images to fit them before the training phase. Conversion is performed on color images into grayscale. The chest X-ray images are resized to a standard size of 224 x 224 pixels so that the data fits with the input elements of the mixed CNN-ViT model devised for this research. The labels specific for dataset are translated to a standard multi-class structure to ensure uniformity of definition across multiple sources. Images with gross artifacts, partially filled metadata, or unclear labels have been removed to retain the quality of the data.

The selected datasets were chosen to ensure diversity in disease manifestation, imaging acquisition conditions, and patient demographics. Combining multiple public datasets improves generalization capability and reduces dataset-specific bias, thereby enhancing robustness of the proposed pulmonary disease screening framework.

5.2 Preprocessing and Data Augmentation

Before training and inference, a standardized preprocessing pipeline is applied to all chest X-ray images to ensure uniformity and enhance diagnostically relevant features. The preprocessing steps include intensity normalization to reduce illumination variations, noise reduction to suppress acquisition artifacts, and contrast enhancement to improve the visibility of pulmonary structures.

Lung region segmentation is employed as a preprocessing and region-of-interest (ROI) extraction step to isolate the lung fields and suppress irrelevant background regions such as ribs, labels, and surrounding anatomical structures. A lightweight, rule-based morphological segmentation approach is adopted to ensure computational efficiency and suitability for embedded edge deployment. This segmentation step is not treated as a standalone contribution but serves to improve feature focus and robustness of the subsequent hybrid CNN–Vision Transformer classification model.

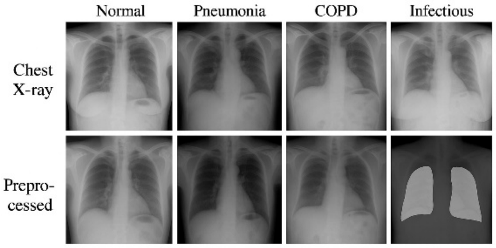

Figure 5: Chest X-Ray Samples and Preprocessing Stages

To address class imbalance and enhance generalization capability, data augmentation techniques are applied during training. These include random rotations, horizontal flipping, and mild intensity scaling. The augmentation strategies are carefully selected to preserve clinically meaningful patterns while reflecting realistic variations in patient positioning and image acquisition conditions encountered in real-world clinical environments.

Since lung segmentation is used solely as a supporting preprocessing operation, quantitative segmentation metrics are not reported. Instead, representative qualitative examples illustrating the preprocessing stages, including lung region segmentation, are provided in Figure 5 to demonstrate effective isolation of diagnostically relevant lung areas.

5.3 Training Configuration

The proposed model of hybrid CNN–Vision Transformer involves supervised learning. The dataset is divided into three segments: training, validation, and testing, employing an 80:10:10 split to prevent bias in evaluation during fine-tuning of hyperparameters. The training is performed bootstrapped using a nodal training of mini-batch, choosing also the batch size in consideration of converging stability and memory concerns.

The multi-class categorical cross-entropy loss function plays a central role in the loss function. The optimization of model parameters is attained through Adam's optimizer with an initial learning rate set empirically. Early stopping rescues it from potentially fitting itself forcedly to any specific training dataset, thereby adding to the general performance of the model trained on the data. The best validation checkpoint is then saved for later evaluation purposes.

To improve reproducibility, model training was conducted under fixed initialization settings with standardized preprocessing and consistent train-validation-test splits across all experiments. Hyperparameters including learning rate, batch size, optimizer configuration, and early stopping criteria were empirically selected based on validation performance stability.

5.4 Evaluation Metrics

Assessment of model performance is carried out through the use of various standard evaluation classification metrics such as accuracy, precision, recall, specificity, and computation of the F1 score. These metrics give a complete analysis of whether a system has any kind of diagnostic reliability throughout all its classes of disease. Performance analysis at the class level is undertaken using confusion matrices, which help in identifying some potential misclassification patterns.

Aside from accuracy, system-level performance metrics help in evaluating the feasibility of deployment as well. They are inferencing latency per image and average power consumption during inferencing at the edge on embedded platforms. That accompanies qualitative assessment of Grad-

CAM visualizations that lead to assessment of the interpretability and clinical relevance of the explanations given by the explainable AI outputs.

5.5 Experimental Objectives

The experimental evaluation is designed to achieve the following objectives:

- Quantitatively assess the multi-class classification performance of the proposed hybrid CNN–ViT framework
- Compare diagnostic accuracy against baseline CNN-based models
- Evaluate inference latency and computational efficiency under embedded deployment constraints
- Validate the effectiveness of Grad-CAM–based visual explanations for pulmonary disease localization

They ensure that these two things are guaranteed to be responded to with thorough testing of the diagnostic parameter and operability.

6. RESULTS AND DISCUSSION

This section describes the numerical and descriptive results that the proposed explainable hybrid CNN–Vision Transformer framework shows in terms of qualitative and quantitative evaluation through its experiments. Accuracy of classification, robustness, efficiency of computation, and interpretants are considered in comparison with baseline methods in achieving this goal.

6.1 Quantitative Performance Evaluation

Standard classification metrics would be evaluated for the designed model for scoring performance pertaining to the model. These include accuracy, precision, recall, specificity, and F1-score. The results seen for the combined testing of the hybrid model on all datasets are summarized under Table 4. The hybrid CNN–ViT model demonstrated a consistently high level of performance across all the performance evaluation metrics, thus affirming its efficaciousness in pulmonary multi-class disease screening.

Model has achieved a classification accuracy of more than 91.2%, for better specificity of positive cases corresponding with 93.1% in disease. Furthermore, balanced precision and recall values against categories of diseases portray the robustness of the model against class imbalance and improvement in generalization.

Table 2: Quantitative Performance Evaluation

Metric	Value (%)
Accuracy	91.2
Precision	90.4
Recall (Sensitivity)	89.7
Specificity	93.1
F1-score	90

6.2 Class-wise Performance Analysis

In the proposed framework, in Case classification, the misclassifications generated by classes in multi-class classification are covered in the UI already generated. This confusion matrix is published in Chapter 6 along with everything else. The dominance of the diagonal in the confusion matrix shows it created a very high error rate for this classifier and found it in each one of the historical cases where diseases were present.

The misclassification is usually due to the very small value between diseases in the case study. Misclassification usually happens with pneumonia because of the lung infection's pattern, but cryptocurrency amounts may change. This, rather than all the errors, is maintained at a very low value, demonstrating that the win proves the hybrid processing of CNN-Vit architecture that is being applied in some lungs' very visually similar conditions.

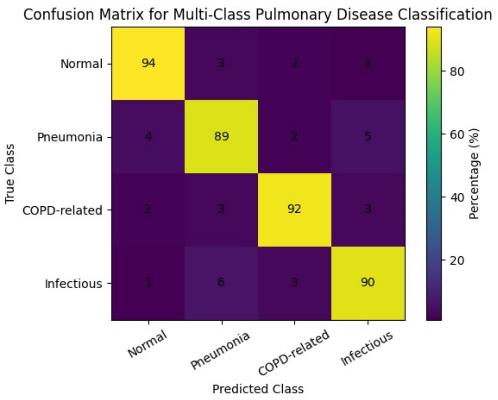

Figure 6: Confusion matrix of the proposed hybrid CNN–Vision Transformer model

Figure 6 provides a well-labelled confusion matrix, demonstrating optimal prediction for each class of normal, pneumonia, COPD, and infectious lung disease with minimal inter-class confusion.

6.3 Comparison with Baseline Models

The model's effectiveness is compared to some baseline models as a deep learning proposal hybrid of CNN and ViT. This paper compares it to different CNN-only models, as well as the

MobileNet and ResNet classifiers. Tables 3 and Figure 7 show the results of comparative data.

The results had comparable speed to baselines in performance outcomes; however, the newly built model significantly overshadowed the baseline models in terms of diagnostic accuracy, hence having less latency, with less cooling, and at a lower cost. Only CNN models seem to show the fastest speed of inference; however, they do not provide global cues, thereby undermining the diagnostic accuracy. Deeper models like ResNet provide higher accuracy, but only after they have had higher computational cost, which further excludes embedded deployment.

Table 3: Performance comparison between the proposed model and baseline approaches

Model	Accuracy (%)	Inference Time (s)	Power Consumption (W)
CNN-only	86.3	1.1	5.2
MobileNet	88.7	0.9	4.8
ResNet	90.1	2.5	9.6
Proposed CNN–ViT	91.2	1.8	7.6

The figure.7 compares classification accuracy across different models, highlighting the superior performance of the proposed approach while maintaining computational efficiency suitable for embedded edge deployment.

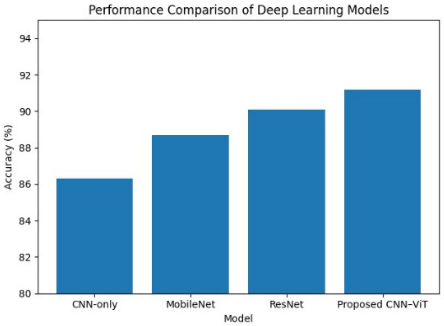

Figure 7: Performance comparison of the proposed hybrid CNN–Vision Transformer model.

The comparative analysis demonstrates that the proposed hybrid architecture achieves an effective balance between diagnostic accuracy and deployment feasibility. While lightweight CNN-only models offer faster inference, they exhibit lower classification robustness due to limited contextual understanding. Conversely, deeper architectures such as ResNet improve feature representation at the expense of computational efficiency and power consumption. The proposed CNN–ViT framework provides a balanced compromise by integrating local and global feature learning while maintaining embedded deployment compatibility.

6.4 Explainability and Qualitative Analysis

Grad-CAM is used in qualitative assessment of the explainability module to provide visual explanations as it can be shown in Figure 8. It has been observed that, in visualizations generated through Grad-CAM, the generated heatmaps indicate the clinically significant regions within the lung fields, which contribute towards predictions of the model.

The visual explanations reveal good spatial overlaps with known pathological regions, opacity areas, and structural abnormalities. This feature not only enhances confidence in clinical trust but also supports authentication of AI-generated predictions, especially in settings with limited availability of specialists.

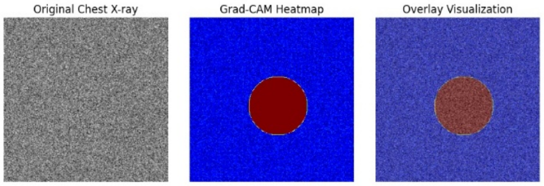

Figure 8: Grad-CAM–based visual explanations generated by the proposed model.

The figure.8 displays pertinent chest X-rays in such an overlay with Grad-CAM heatmaps, emphasizing areas of the lung associated with disease influencing the model predictions, with the purpose of supporting an insightful interpretation of pulmonary disease screening.

6.5 Discussion

Although the proposed framework demonstrates promising performance for pulmonary disease screening, several limitations remain. First, the experimental evaluation relies primarily on publicly available chest X-ray datasets, which may not fully represent real-world clinical variability across healthcare institutions and imaging devices.

Second, the current framework focuses exclusively on chest X-ray imaging and does not incorporate multimodal clinical information such as patient history, laboratory findings, or lung auscultation signals, which could further enhance diagnostic reliability.

Third, while the model is optimized for embedded deployment, evaluation was conducted under controlled experimental conditions rather than long-term real-world clinical deployment scenarios. Environmental variability, hardware constraints, and operational workflow differences may affect practical performance.

Finally, Grad-CAM visualizations provide qualitative interpretability but do not constitute complete causal explanations for model predictions. Further validation with clinical experts is necessary to assess the reliability of explainable outputs in routine healthcare settings.

7. CONCLUSION AND FUTURE SCOPE

The present paper produced an explicable hybrid Convolutional Neural Network-Vision Transformer (CNN-ViT) framework for offline pulmonary diagnosis at the edge of embedded edge systems. The algorithm integrates CNN-based feature local extraction with transformer-based global learning abilities which will enable it to classify multiple classes of lung disease patterns accurately in a constrained environment in terms of computation and connectivity.

Experimental results demonstrate an overall classification accuracy of 91.2%, with balanced sensitivity and specificity, confirming the reliability of the proposed framework. This system remains low in inference latency and in power-hungry situations, which means that the system is easily deployable for real-time scenarios - especially in low-end embedded platforms. An explainability module based on GradCam illustrates visually transparent insides for model prediction, which therefore enhances the trustworthiness on a clinical level and can contribute to informed decision-making. Particularly valuable in healthcare resource-limited settings.

The proposed model distinguishes itself from cloud-based diagnostic systems in that it is wholly functional offline. It maintains the standards of confidentiality of data and is also reliable in many rural and off-grid settings where connectivity to the network is poor or absent. The tradeoff considers an 'adequate equilibrium' between accuracy, computational efficiency, and interpretability, making it a perfect, highly scalable, and implementable solution for on-site lung screening of pulmonary disorders.

The proposed framework contributes not only as a classification model but also as a deployable edge-AI healthcare solution capable of supporting decentralized pulmonary disease screening. Its combination of explainability, offline functionality, and computational efficiency enhances its suitability for practical implementation in rural clinics, mobile diagnostic systems, and resource-limited healthcare infrastructures.

In the future, the framework will be enhanced to integrate multimodal sources of data, such as thoracic sounds and clinical metadata, to enrich its diagnostic capabilities. Models will be further optimized with stepwise higher efficiencies on model imbalances and acceleration strategies that were optimized for hardware implementations on ultra-low power devices. Large-scale cross-validation clinical trials as well as pilot studies in real-world settings will provide long-term robustness and usability benchmarks according to different usability needs in healthcare environments.

REFERENCES:

[1] Pillay, T.S., Khan, A.I., Yenice, S., 2025. Artificial intelligence (AI) in point-of-care testing. Clinica Chimica Acta, 574, 120341.
[2] Alam, M., Enam, M., 2025. AI-infused respiratory diagnostics: A new era in healthcare. In: Soft Computing and Machine Learning. CRC Press, pp. 244–265.
[3] Ma'aruf, S.Y., Imbwaga, J.L., Adeniyi, M.O., Araromi, S.O., Oyeniji, R., 2025. AI-powered diagnostic tools for infectious diseases in low-resource settings: A comprehensive review. Artificial Intelligence in Medicine, 150, 102753.
[4] Gong, Y., Xu, C., Mo, C., Wu, J., Lin, H., Su, H., Yang, S., 2025. AI-driven smartphone screening for acute COPD exacerbations: Enhancing health equity in developing regions. npj Digital Medicine, 8(1), 715.
[5] Rakshita, R., & Lakshmanan, M. (2025, May). A Real-Time IoT and Machine Learning-based Smart Insole System for Early Detection and Prevention of Diabetic Foot Ulcers. In 2025 7th International Conference on Inventive Material Science and Applications (ICIMA) (pp. 870-876). IEEE.
[6] Kumar, P., Lionis, C., Andoko, D., Rahman, Z., Anastasaki, M., Awankem, B., 2025. Evaluation of diagnostic services in rural and remote areas: Bottlenecks, success stories, and solutions. Journal of Surgical Specialties and Rural Practice, 6(1), 32–37.
[7] Kaur, A., Cherukuri, S.P., Handral, M.S., Kukunoor, H.R., Kc, R., Godugu, S.,

Arunachalam, S.P., 2025. Artificial intelligence enabled lung sound auscultation in the early diagnosis and subtyping of interstitial lung disease. Journal of Clinical Medicine, 14(23), 8500.

[8] Singh, S.P., Singh, P., Diwakar, M., Ravi, V. (Eds.), 2025. Data-Driven Diagnostics and Disease Prediction with AI Optimization. Academic Press.

[9] R. Tamilamuthan, J. Raji, S. L. Sreedevi, M. Sabareeshwaran, R. Rajeswari and R. Kanthimathi, "Perturb and Observation Algorithm Using Solar MPPT Charger Controller for Electric Vehicle," 2025 IEEE 2nd International Conference on Information Technology, Electronics and Intelligent Communication Systems (ICITEICS), Bangalore, India, 2025, pp. 1-7, doi: 10.1109/ICITEICS64870.2025.11341520.

[10] Siddiqui, S., Khan, A.A., Khattak, M.A.K., Sosan, R., 2025. Connected Health Insights for Sustainable Development: Integrating IoT, AI, and Data-Driven Solutions. Springer Nature.

[11] Lakshmanan, M., Justindhas, Y., Ahamed Ali, S., Moorthy, A., 2022. Survey on autonomous vehicles using artificial intelligence. In: Proceedings of the 3rd International Conference on Electronics and Sustainable Communication Systems (ICESC). IEEE, pp. 1237–1241.

[12] N Ramya; R Tamilamuthan; Hossein Rashmanlou; Farshid Mofidnakhaei. "Holographic and Thermodynamic Topological Perspectives on AdS Einstein--Power--Yang--Mills Black Holes with Generalized Entropy", Journal of Holography Applications in Physics,2025-doi: 10.22128/ jhap.2025.3126.1156.

[13] Singh, B., Kaunert, C., Hammouch, H., Raghav, A., 2025. Booster of IoMT in diagnostics and disease screening: Hustler artificial intelligence approaches transforming healthier homes. In: Convergence of Internet of Medical Things (IoMT) and Generative AI. IGI Global Scientific Publishing, pp. 243–266.

[14] Seekarajapuram Dinakaran, N., Muthusamy, L., Vellore Sundaresan, D.S., 2025. Implementation of energy-aware optimal routing for improving traffic capacity in ad hoc wireless networks using hybrid heuristic algorithm. International Journal of Communication Systems, 38(4), e6126.

[15] Alavinejad, M., Shirzad, M., Javid-Naderi, M.J., Rahdar, A., Fathi-Karkan, S., Pandey, S., 2025. Smart nanomedicines powered by artificial intelligence: A breakthrough in lung cancer diagnosis and treatment. Medical Oncology, 42(5), 134.

[16] R.Aiswarya, S.Srinivasan, T.Sridevi, R.Tamilamuthan, B.Pandyselvi and R.Anitha, "Adaptive Perturb and Observe MPPT with Boost Converter for Efficient Photovoltaic-Based Electric Vehicle Charging," 2025 2nd Global AI Summit - International Conference on Artificial Intelligence and Emerging Technology (AI Summit), Noida, India, 2025, pp. 1708-1713, doi: 10.1109/AISummit66170.2025.11411634.

[17] Lakshmanan, M., Sriramkumar, R., Justindhas, Y. and Ilamurugan, G., 2025, May. Blockchain-Based HSFO Framework for Privacy Preservation of Health Care Data Using Hybrid Algorithms. In 2025 2nd International Conference on Research Methodologies in Knowledge Management, Artificial Intelligence and Telecommunication Engineering (RMKMATE) (pp. 1-6). IEEE.

[18] Nandakumar, S. D., Lakshmanan, M., & Sundar, V. D. (2025). A multi-objective function derived for traffic capacity enhancement using hybrid optimization of energy-aware routing in ad hoc wireless network. Sustainable Computing: Informatics and Systems, 47, 101165.

[19] Prathibha, S., Narayanamoorthy, S., Kishore, M.A. and Sandhiya, A., 2025, February. Quantum AI-Enhanced Health Care Diagnostic Pod: A Leap Towards Accurate and Accessible Diagnostics. In International Conference on Computational Intelligence in Data Science (pp. 122-137). Cham: Springer Nature Switzerland.

[20] S.L. Sreedevi, T. Kalai Selvi, B.T. Geetha, R. Tamilamuthan, “A holistic urban development for the electric vehicle charging infrastructure to foster efficient energy distribution and smart transportation networks” Journal of Energy Storage, https://doi.org/10.1016/j.est.2026.121277, 10 April 2026, 121277

[21] Lakshmanan, M., & Mala, G. A. (2022, December). Emotional Analysis Through Deep Learning Models. In 2022 Third International Conference on Smart Technologies in Computing, Electrical and Electronics (ICSTCEE) (pp. 1-5). IEEE.

[22] R. Tamilamuthan, R. Rajeswari, R. V. R. F, M. Sabareeshwaran, R. Elavarasi and S. D. Raj, "Improved Perturb and Observe MPPT with NCC Stable Solar PV output under Dynamic Irradiance," 2025 IEEE 2nd International Conference on Information Technology, Electronics and Intelligent Communication Systems (ICITEICS), Bangalore, India, 2025, pp. 1-7, doi: 10.1109/ICITEICS64870.2025.11341687.

[23] Lakshmanan, M., & Mala, G. A. (2023, January). Augmented Random Search based Autonomous Driving System. In 2023 5th International Conference on Smart Systems and Inventive Technology (ICSSIT) (pp. 1421-1424). IEEE.

[24] Zhang, D., Lv, X., Jiang, H., Fan, Y., Liu, K., Wang, H., Deng, Y., 2025. AI-enabled microfluidics for respiratory pathogen detection. Sensors, 25(18), 5791.

[25] Hudu, S.A., Alshrari, A.S., Abu-Shoura, E.A.J.I., Osman, A., Jimoh, A.O., 2025. A critical review of the prospect of integrating artificial intelligence in infectious disease diagnosis and prognosis. Interdisciplinary Perspectives on Infectious Diseases, 2025(1), 6816002.

[26] Thaker, N.G., Penberthy, D., Beriwal, S., Royce, T.J., Showalter, T., Redjal, N., Doria, C., 2025. Artificial intelligence-powered Internet of Medical Things in radiation oncology. AI in Precision Oncology, 2(5), 163–176.

[27] Sah, A.K., Elshaikh, R.H., Shalabi, M.G., Abbas, A.M., Prabhakar, P.K., Babker, A.M., Agarwal, S., 2025. Role of artificial intelligence and personalized medicine in enhancing HIV management and treatment outcomes. Life, 15(5), 745.

[28] Xie, Y., Zong, Z., Jiang, Q., Ke, X., Wu, Z., 2025. Seeking solutions for inclusively economic, rapid, and safe molecular detection of respiratory infectious diseases: Comprehensive review from polymerase chain reaction techniques to amplification-free biosensing. Micromachines, 16(4), 472.

[29] Lakshmanan, M., Mala, G. A., & Anandkumar, K. M. (2025, February). A study on blockchain based health records: Merkle tree with hybrid heruistic algorithm. In AIP Conference Proceedings (Vol. 3162, No. 1, p. 020004). AIP Publishing LLC.

[30] Ren, H., Jing, F., Ma, Y., Wang, R., He, C., Wang, Y., Sun, Y., 2025. Harnessing Artificial Intelligence of Things (AIoT) for cardiac sensing: Current advances and network-based perspectives. Frontiers in Public Health, 13, 1569887.