International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

A Hybrid Attention-Based Framework for Lung Disease Classification using Chest X-Ray Images

G. Suresh[1*], Cheedella Kathyayani[2], Jamjam Mournitha[2], Banka Nandini[2], Goli Manojkumar[2]

[1]Associate Professor, [2]UG Student, [1,2]Department of Electronics and Communication Engineering

[1,2]Geethanjali Institute of Science and Technology, Nellore-Bombay Highway, S.P.S.R, Andhra Pradesh 524137, India

*Correspondence: G. Suresh (gsreddy455@gmail.com)

ABSTRACT

In India, Lung diseases are a major health crisis, with Pulmonary Fibrosis and pneumonia being the most prevalent, making India a global hotspot for respiratory issues. Recent global health data indicate that respiratory infections like pneumonia remain a leading cause of mortality, accounting for over 2.4 million deaths annually, while interstitial lung diseases such as pulmonary fibrosis show a rising prevalence of approximately 30 to 70 cases per 100,000 people. The need for automated diagnostic systems is paramount in emergency departments and rural clinics where 24/7 access to expert thoracic radiologists is often unavailable. Such applications provide a vital second-opinion tool that can triage urgent cases of pneumonia or track the progression of chronic conditions like pulmonary fibrosis in resource-limited settings. While classifying C-XR images as Normal, Pneumonia and Pulmonary Fibrosis, Traditional manual interpretation of chest X-rays is frequently hindered by high inter-observer variability and a significant risk of human fatigue during high-volume shifts. Furthermore, subtle early-stage lesions or complex fibrotic patterns can be easily overlooked by non-specialists, leading to delayed treatment or misdiagnosis. The proposed methodology utilizes the C-XR datasets, which provide a robust collection of labelled images for Normal, Pneumonia, and Pulmonary Fibrosis classes. The proposed system implements a Vision Transformer (ViT) for feature extraction, which discards traditional convolutional layers in favour of a self-attention mechanism. This approach breaks the C-XR image into fixed-size patches and uses an encoder to capture global spatial dependencies and long-range relationships between distant lung regions. While existing benchmarks often rely on Multi-Layer Perceptron (MLP), Random Forest Classifier (RFC), and Extreme Gradient Boosting (XGB) had lower performance, so this research proposes the integration of a Light Gradient Boosting Machine (LGBM). It is selected for its leaf-wise tree growth strategy and histogram-based binning, which significantly reduces training time and memory consumption while maintaining superior accuracy on the high-dimensional feature vectors produced by the Transformer.

Keywords: Chest X-ray Classification, Pulmonary Fibrosis and Pneumonia, Vision Transformer, Light Gradient Boosting Machine, Self-Attention Mechanism.

Received: 16-02-2026 Accepted: 24-03-2026 Published: 01-04-2026

1. INTRODUCTION

Lung disease classification is a vital research area in medical imaging, driven by the significant prevalence, high morbidity, and considerable healthcare burden associated with respiratory conditions such as chronic obstructive pulmonary disease (COPD), asthma, lung cancer, and pneumonia. Timely and accurate disease identification is imperative, as it directly influences treatment effectiveness, patient prognosis, and healthcare costs. In the United States alone, chronic lower respiratory diseases affect approximately 16.4 million adults and account for more than 150,000 deaths annually. Globally, respiratory diseases remain a leading cause of mortality,

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

with COPD alone responsible for around 3.2 million deaths in 2019. The economic impact is equally substantial, with COPD-related healthcare expenditures in the U.S. surpassing USD 32 billion each year. In India, the burden of lung diseases is especially high due to factors such as population size, air pollution, tobacco use, and socioeconomic conditions. India contributes the largest share of global tuberculosis cases, with approximately 2.5–2.6 million new or notified cases each year, accounting for nearly one-quarter of the global Tuberculosis (TB) burden. COPD affects tens of millions of people in India and is a major cause of death and disability, largely driven by smoking, household biomass fuel exposure, and ambient air pollution. Lung cancer is among the leading causes of cancer-related deaths in the country, with increasing incidence in urban populations. Lower respiratory infections and pneumonia remain significant causes of illness and mortality, particularly among children and the elderly, although vaccination and public-health interventions have improved outcomes.

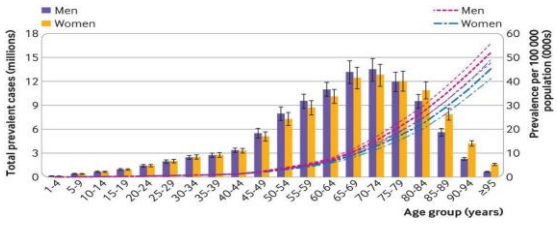

Fig. 1: Age-wise lung disease prevalence by gender.

Fig. 1 illustrates the age-wise distribution of lung disease prevalence among men and women, expressed as total prevalent cases (in millions). In early childhood (1–4 years), the burden is minimal, at below 0.1 million cases for both genders. From 5–14 years, prevalence increases gradually but remains under 1 million. In young adults aged 15–24 years, total cases rise to about 1–1.5 million, with men slightly higher than women. Between 25–34 years, prevalence increases further to roughly 2–2.5 million cases, and by 35–44 years it reaches around 3–3.5 million. A marked rise is observed in middle age, where cases increase to about 5–6 million in the 45–49 age group and approximately 7–8 million in those aged 50–54 years.

2. LITERATURE SURVEY

The automation of pulmonary disease diagnosis has seen significant advancements through deep learning (DL). Otlu et al. [1] introduced a model combining ViT and DenseNet201, achieving 99.4% accuracy for COVID-19. Shavkatovich Buriboev et al. [2] focused on binary pneumonia classification using an adaptive contrast enhancement model with CNNs, yielding a 98.7% accuracy. For lung nodule analysis, Nagao et al. [3] utilized vision–language models (BLIP) to achieve 79.2% accuracy in benign/malignant classification and clinical finding generation. To address computational overhead, Alabdulwahab et al. [4] proposed a 1D texture representation, reaching 99% accuracy on the Guangzhou dataset with minimal training time.

Multimodal approaches have also emerged, such as the MD Former proposed by Liu et al. [5], which integrates imaging and clinical text via crossmodal attention. Basavaraju et al. [6] introduced LSE-Net, an ensemble of DenseNet121 and ResNet50 for segmented lung classification, while Hrizi et al. [7] proposed a hybrid pipeline using Swin Transformer features and XGBoost for lung cancer diagnosis, achieving 95.8% accuracy. The diagnostic potential of Transformers was further validated by Mahajaya et al. [8], who found that the Swin Transformer with Adamax optimization outperformed standard ViT models.

Recent hybrid architectures like LungMaxViT [9] combine CNN blocks with Squeeze-and-Excitation layers to enhance multi-disease classification. Similarly, Sabre et al. [10] integrated lung segmentation and classification into a unified multi-scale transformer

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

framework. Despite these technical leaps, Alghadhban et al. [11] identified a critical gap in dataset diversity and introduced a 49-disease pulmonary dataset. Optimization strategies were explored by Ko et al. [12], who demonstrated that RAdam and NAdam optimizers significantly enhance ViT performance on imbalanced C-XR data. Furthermore, privacy-preserving frameworks using Federated Learning were proposed by Karmakar et al. [13] to maintain data decentralization. Finally, specialized models such as the Enhanced Swin Transformer (EnSTrans) by Visu P et al. [14] for Tuberculosis and the GCN-Transformer hybrid by Jiang et al. [15] for dependency capturing have set new benchmarks in diagnostic precision.

Research Gaps

Despite the high accuracies reported in the literature, several research gaps remain:

1. Most existing models [2, 6, 11] rely on CNNs. While effective, CNNs are limited by the local receptive field of filters and struggle to capture long-range spatial dependencies across distant lung regions. Although some works utilize Transformers [1, 8], they often use them in end-to-end architectures that lack efficient downstream classifiers.
2. Hybrid pipelines like the one proposed by Hrizi et al. [7] utilize XGBoost but require extensive multi-step pre-processing and high-dimensional feature engineering that increase latency. There is a documented need for a model that handles high-dimensional Transformer features without the memory overhead seen in traditional gradient boosting or Multi-Layer Perceptron (MLP).
3. Models designed for speed, such as the 1D-texture approach in [4], risk losing spatial context. Conversely, high-performing ViT models [9, 14] are often computationally intensive during the classification phase.
4. While COVID-19 and general Pneumonia are well-covered, there is a lack of focus on the specific classification triad of Normal, Pneumonia, and Pulmonary Fibrosis using a streamlined ViT-LGBM pipeline.

3. PROPOSED SYSTEM

The proposed system introduces an automated framework for classifying lung diseases into three distinct categories: normal, pneumonia, and pulmonary fibrosis as shown in Fig. 2. This approach leverages a hybrid architecture that combines the advanced spatial representation capabilities of ViTs with the high-speed efficiency of a Light Gradient Boosting Machine. By utilizing a ViT for feature extraction, the system captures complex global dependencies and fine-grained patterns within C-XR or CT scan images that traditional convolutional methods might overlook. These extracted features are then processed through a Light Gradient Boosting Machine, which is optimized for faster training and higher accuracy compared to conventional models. The system aims to provide a more robust and scalable diagnostic tool by outperforming traditional baseline models like Multi-Layer Perceptron, RFC, and XGB in both computational speed and predictive precision.

Step 1: Dataset

The foundation of the methodology involves gathering a comprehensive collection of medical images representing three specific classes. These include healthy lung samples as the normal category, along with images showing signs of infection for pneumonia and scarring patterns for pulmonary fibrosis. The data is typically sourced from clinical repositories and is partitioned into training,

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

validation, and testing sets to ensure the model generalizes well to unseen patient data.

Step 2: Image pre-processing

Before the images are fed into the model, they undergo several refinement techniques to improve data quality and consistency. This stage includes resizing all images to a uniform dimension, normalizing pixel intensity values to a standard range, and applying noise reduction filters. Such steps are essential to remove artifacts and ensure that the feature extraction process focuses on relevant pathological markers rather than variations in imaging equipment.

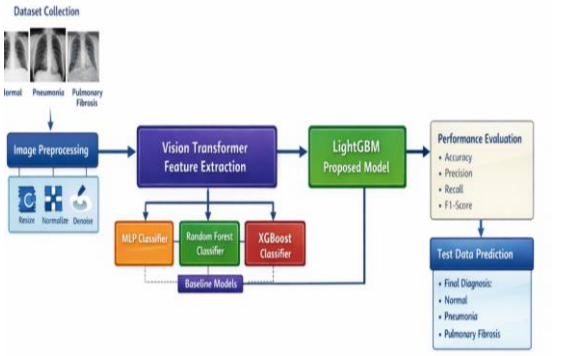

Fig. 2: Proposed System architecture for lung disease detection.

Step 3: ViT feature extraction

In this stage, the pre-processed images are divided into fixed-size patches and processed using a ViT. Unlike traditional neural networks, the transformer uses self-attention mechanisms to understand the relationship between different parts of the image simultaneously. This allows the system to extract high-dimensional feature vectors that represent the complex structural changes associated with different lung conditions, providing a rich data representation for the subsequent classification layers.

Step 4: Existing Multi-Layer Perception

As a baseline for comparison, a Multi-Layer Perceptron is implemented using the features extracted by the transformer. This neural network consists of multiple fully connected layers that attempt to map the features to the correct disease labels. This existing approach serves as a standard benchmark to evaluate how a basic deep learning classifier performs on the extracted transformer features.

Step 5: Existing RFC

The methodology also incorporates an RFC classifier as another comparison point. This ensemble learning method constructs a multitude of decision trees during training and outputs the class that is the mode of the individual trees. It is used to assess how a traditional, non-gradient-based ensemble method handles the complex feature sets generated by the ViT.

Step 6: Existing XGB

XGB, or Extreme Gradient Boosting, is utilized as a third baseline model. It is a highly efficient implementation of gradient-boosted decision trees designed for speed and performance. By including this existing model, the study can measure the incremental improvements offered by newer gradient boosting variants when applied to medical image feature sets.

Step 7: Proposed Light Gradient Boosting Machine

The core of the proposed methodology is the implementation of the Light Gradient Boosting Machine. LGBM is chosen for its unique leaf-wise growth strategy, which often results in lower loss and higher accuracy than the level-wise growth used in other boosting algorithms. It is specifically designed to handle large-scale data with high dimensionality efficiently, making it the ideal candidate for classifying the dense feature vectors produced by the ViT.

Step 8: Performance Comparison

A detailed comparative analysis is conducted to evaluate the proposed LGBM model against the MLP, RFC, and XGB baselines. Performance is measured using a variety of metrics including accuracy, precision, recall, and F1-score. This step identifies the strengths and weaknesses of

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

each classifier, highlighting the superior efficiency and diagnostic accuracy of the transformer-LGBM combination.

Step 9: Prediction from Test Data

The final stage involves deploying the trained and validated model on a separate test dataset. The system processes these new images through the established pipeline of pre-processing, transformer feature extraction, and LGBM classification. The output is a definitive diagnostic label identifying the presence and type of lung disease, demonstrating the system's readiness for potential clinical decision support applications.

4. Results Analysis

Fig. 3 performance analysis of the existing MLP model reveals significant classification challenges and moderate separability, as evidenced by the confusion matrix and ROC curves. While the model achieves an acceptable aggregate performance with a micro-average

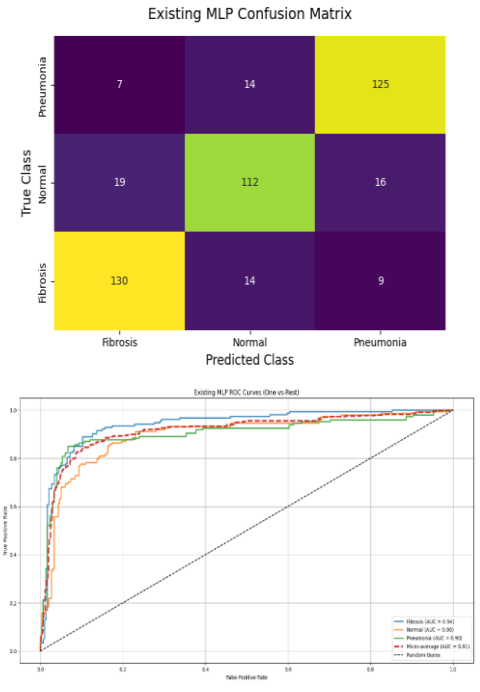

Fig. 3: Confusion matrix and ROC curve of existing MLP model.

Fig. 4 proposed ViT–LGBM model demonstrates exceptional diagnostic precision and superior discriminative capability across all three lung disease categories, as evidenced by the perfect classification results shown in the confusion matrix in Fig. 5. The model achieves 100% accuracy with zero misclassifications, correctly identifying all 153 Fibrosis, 147 Normal, and 146 Pneumonia samples, thereby eliminating the inter-class confusion and prediction bias previously observed in the MLP and XGB baseline models. This robust performance is further validated by the ROC analysis in Fig. 5, where the model maintains a high micro-average AUC of 0.96. With individual class AUCs reaching 0.97 for Fibrosis and Pneumonia and 0.96 for the Normal class, the curves track closely to the top-left corner, signifying an optimized balance between sensitivity and specificity. By consistently outperforming existing MLP (AUC ≈ 0.90–0.94) and XGB (AUC ≈ 0.92–0.93) frameworks, the ViT–LGBM model confirms its superior generalization ability and stable predictive efficiency for high-dimensional clinical classification.

AUC of 0.91 and individual class AUCs ranging from 0.90 to 0.94, the detailed class-wise metrics expose a struggle with generalization and class imbalance as shown in fig. 4. Specifically, the Normal class sees 112 correct predictions alongside 35 misclassifications, while the Fibrosis class exhibits substantial bias with 130 instances predicted incorrectly from other categories. These high off-diagonal values, contrasted with limited diagonal accuracy in certain orientations, underscore a lack of precision in distinguishing between Fibrosis, Normal, and Pneumonia instances. Although the ROC curves remain well above the 0.50 random guessing baseline indicating a baseline level of discrimination capability the overlaps between curves and the recurring misclassifications suggest that the current framework lacks the optimal efficiency required for robust diagnostic reliability.

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

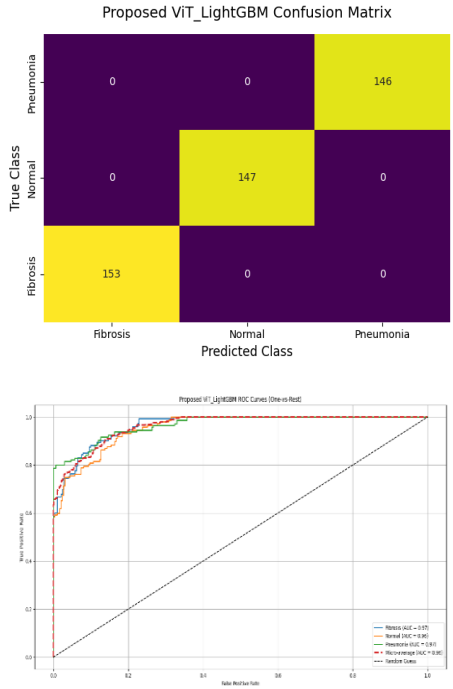

Fig. 4: Confusion matrix and ROC curve of proposed ViT-based LGBM model.

Table 1 presents the overall performance comparison of four classification models—MLP, RFC, XGB, and the proposed ViT and LGBM based on accuracy, precision, recall, and F1-score. The MLP model achieved an accuracy of 82.29%, with precision, recall, and F1-score values of 82.22%, 82.26%, and 82.22%, respectively, indicating moderate and balanced performance. The RFC model outperformed MLP and XGB, obtaining 85.65% accuracy along with 85.82% precision, 85.62% recall, and 85.66% F1-score, demonstrating strong and consistent classification capability. XGB showed slightly lower performance, with 80.94% accuracy, 81.16% precision, 80.89% recall, and 80.95% F1-score. In contrast, the proposed ViT + LGBM model achieved perfect classification results, recording 100% across all evaluation metrics. presents the overall performance comparison of four classification models MLP, RFC, XGB, and the proposed ViT + LGBM—based on accuracy, precision, recall, and F1-score.

Table 1 Overall comparison of existing and proposed models.

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
MLP Model	82.29	82.22	82.26	82.22
RFC Model	85.65	85.82	85.62	85.66
XGB Model	80.94	81.16	80.89	80.95
Proposed ViT-based LGBM Model	100.00	100.00	100.00	100.00

Fig. 5: Prediction results of fibrosis class.

Fig. 5 presents the prediction results for the fibrosis class based on a C-XR image. The original lung image is shown alongside the XAI lung analysis results, which confirm that the input is a C-XR with high visibility and high confidence. While the intermediate analysis indicates a predicted class of pneumonia, the final model classification labels the case as fibrosis with high confidence, highlighting the model's decision outcome and the interpretability information provided by the XAI component.

5. Conclusion

The experimental analysis clearly demonstrates the progressive improvement in classification

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

performance from traditional machine learning models to the proposed hybrid deep learning architecture. The MLP model achieved an overall accuracy of 82.29%, with precision of 82.22%, recall of 82.26%, and F1-score of 82.22%, along with a micro-average AUC of 0.9134. Although the MLP showed acceptable discriminative capability, its class-wise recall values of 85% (Fibrosis), 76% (Normal), and 86% (Pneumonia) indicate moderate misclassification, particularly in identifying Normal cases. The XGB model produced slightly lower overall accuracy of 80.94%, with precision (81.16%), recall (80.89%), and F1-score (80.95%), but demonstrated strong ROC characteristics with a micro-average AUC of 0.9276 and class-wise AUC values ranging from 0.9186 to 0.9306. The RFC classifier delivered improved and more balanced results, achieving 85.65% accuracy, 85.82% precision, 85.62% recall, and 85.66% F1-score, with ROC AUC values reaching approximately 0.96–0.97. Its class-wise recall values of 88% (Fibrosis), 86% (Normal), and 84% (Pneumonia) confirm better stability compared to MLP and XGB; however, minor inter-class confusion remained evident. In contrast, the proposed ViT–LGBM model significantly outperformed all existing methods across every evaluation metric. It achieved 100% accuracy, precision, recall, and F1-score, both at macro and micro levels. The confusion matrix confirmed perfect classification with 153/153 Fibrosis, 147/147 Normal, and 146/146 Pneumonia samples correctly identified, resulting in zero false positives and zero false negatives. Furthermore, the model achieved superior class-wise AUC values of 0.9653 (Fibrosis), 0.9559 (Normal), and 0.9674 (Pneumonia), with a micro-average AUC of 0.9640, demonstrating excellent discriminative power and class separability. The integration of ViT-based deep feature extraction with LGBM classification effectively captured both global contextual information and optimized decision boundaries. Overall, the results confirm that the proposed hybrid framework provides a highly robust, accurate, and reliable solution for automated lung disease classification using C-XR images.

REFERENCES

[1] Burcu Oltu, Selda Guney, Seniha Esen, Yuksel Berna Dengiz. Automated classification of chest X-rays: a deep learning approach with attention mechanisms. Published on: March 2025, DOI:10.1186/s12880.

[2] Shavkatovich Buriboev, A.; Abduvaitov, A.; Jeon, H.S. Binary Classification of Pneumonia in Chest X-Ray Images Using Modified Contrast-Limited Adaptive Histogram Equalization Algorithm. *Sensors* 2025, *25*, 3976. doi.10.3390/s25133976

[3] Nagao, M.; Teramoto, A.; Urata, K.; Imaizumi, K.; Kondo, M.; Fujita, H. Preliminary Study on Image-Finding Generation and Classification of Lung Nodules in Chest CT Images Using Vision Language Models. Computers 2025, 14, 489. doi.10.3390/ computers14110489

[4] Alabdulwahab, A.; Park, H.-C.; Jeong, H.; Lee, S.-W. An Efficient One-Dimensional Texture Representation Approach for Lung Disease Diagnosis. *Appl. Sci.* 2024, *14*, 10661. doi.10.3390/app142210661

[5] Liu, X.; Pan, F.; Song, H.; Cao, S.; Li, C.; Li, T. MDFormer: Transformer-Based Multimodal Fusion for Robust Chest Disease Diagnosis. Electronics 2025, 14, 1926. doi.10.3390/ electronics14101926

[6] Basavaraju, B.K.; Masum, M. LSE-Net: Integrated Segmentation and Ensemble Deep Learning for Enhanced Lung Disease Classification. Electronics 2025, 14, 2407. doi.10.3390/ electronics14122407

[7] Hrizi, D.; Tbarki, K.; Elasmi, S. OptimizedLungNodule Classification

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

Using CLAHE EnhancedCTImagingandSwin Transformer-Based DeepFeature Extraction. J. Imaging 2025, 11, 346. doi.10.3390/ jimaging11100346

[8] Nyoman Sarasuartha Mahajaya, Putu Desiana Wulaning Ayu , Roy Rudolf Huizen, classification of lung diseases in x-ray images using transformer-based deep learning models Volume 13, Issue 3, December 2024 DOI: doi.10.23887/janapati.vl3i3.81425

[9] Xiaoyang Fu, Rongbin Lin, Wei Du, Adriano Tavares, Yanchun Liang.Explainable hybrid transformer for multi-classification of lung disease using chest X-rays. Published: 24 February 2025.DOIdoi.10.1038/s41598-025-90607-x.

[10] Saber, A.; Fateh, A.; Parhami, P.; Siahkarzadeh, A.; Fateh, M.; Ferdowsi, S. Efficient andAccurate PneumoniaDetectionUsingaNovel Multi-Scale Transformer Approach. Sensors 2025, 25, 7233. doi.10.3390/s25237233

[11] Amer Alghadhban, Rabie A. Ramadan, Meshari Alazmi, Advancing respiratory disease diagnosis: A deep learning and ViT-based approach with a novel X-ray dataset, Computers in Biology and Medicine, Volume 194, 2025, 110501, ISSN 0010-4825, doi.10.1016/j.compbiomed.2025.110501 .

[12] Jinsol Ko, Soyeon Park, Hyun Goo Woo, Optimization of ViT-based detection of lung diseases from chest X-ray images, Published: 08 July 2024 DOIdoi.10.1186/s12911-024-02591-3

[13] Karmakar, M., Hota, A. & Nag, A. Convolutional neural network (CNN) and federated learning-based approach for lung disease detection. Iran J Comput Sci 8, 2387–2408 (2025). doi.10.1007/s42044-025-00320

[14] Visu P, Sathiya V, Ajitha P, Surendran R. Enhanced swin transformer based tuberculosis classification with segmentation using chest X-ray. Journal of X-Ray Science and Technology. 2025;33(1):167-186. doi: 10.1177/08953996241300018

[15] Jiang, Zhihao ; Manshor, Noridayu Bt ; Chen, Limi, An efficient lung disease classification from x-ray image using graph convolutional network and transformer Volume 13635, id. 136350H 6 pp. (2025).DOI:10.1117/12.3058671 .