International Journal of Current Science Research and Review
ISSN: 2581-8341
Volume 08 Issue 07 July 2025
DOI: 10.47191/ijcsrr/V8-i7-52, Impact Factor: 8.048
IJCSRR @ 2025
www.ijcsrr.org

Lung Disease Classification Using Transfer Learning on Chest X-ray Images

Rana Riyadh Saeed
Department of Radiology Techniques, Mosul Medical Technical Institute, Northern Technical University, Mosul, Iraq

ABSTRACT: Lung diseases remain a significant global health concern, necessitating the development of rapid and accurate diagnostic methods. While previous research has shown the promise of deep learning models, particularly transfer learning with architectures such as ResNet and VGG, limitations persist in evaluation scope, class imbalance handling, and model interpretability. This study proposes an enhanced deep learning framework for multi-label classification of thoracic diseases using chest X-ray images, addressing these gaps through comprehensive evaluation metrics, advanced data augmentation, and explainable AI (XAI) techniques. The NIH ChestX-ray14 dataset is utilized, with class imbalance mitigated via synthetic minority oversampling and weighted focal loss. Multiple state-of-the-art CNN architectures, including EfficientNet and ResNet variants, are benchmarked using precision, recall, F1 Score, AUC, and accuracy. Moreover, Gradient-weighted Class Activation Mapping (Grad-CAM) is integrated to visualize pathological regions, improving clinical interpretability. The offered framework can perform better in all assessment criteria, achieving an AUC of 0.91 with EfficientNet-B0, and provides interpretable outputs critical for deployment in real-world diagnostic settings. This work advances automated radiological diagnosis by addressing key methodological shortcomings and offers a reliable, explainable solution for lung disease detection.

KEYWORDS: Chest X-ray (CXR), Deep Learning, EfficientNet, Lung Disease Classification, Transfer Learning, ResNet, Grad-CAM, NIH ChestX-ray14

INTRODUCTION

Lung diseases, including pneumonia, emphysema, fibrosis, and pneumothorax, represent a significant portion of global morbidity and mortality, particularly in low- and middle-income countries where access to advanced medical diagnostics is limited [1]. Early and accurate detection of thoracic pathologies is crucial for enhancing patient outcomes and alleviating the burden on healthcare systems [2]. Chest X-ray (CXR) imaging remains the most widely used and accessible tool for initial pulmonary assessment due to its cost-effectiveness and non-invasive nature [3]. However, accurate interpretation of CXR images requires substantial expertise, and even skilled radiologists may encounter difficulties when differentiating between overlapping disease patterns or subtle anomalies, especially in high-volume clinical environments [4].

Current artificial intelligence (AI) (particularly deep learning (DL) and convolutional neural networks (CNNs)) have a huge potential to automate the process of detecting lung abnormalities on a chest X-ray image [5], [6]. Transfer learning has facilitated the high-performance diagnostic model that researchers created with comparatively small medical datasets by exploiting pre-trained models on very large-scale datasets like ImageNet [7]. Architectures such as ResNet [8], VGG [9], and EfficientNet [10] have been widely applied to tasks involving the classification of thoracic diseases. For instance, the study by Kavitha et al. [11] evaluated multiple DL models using the NIH ChestX-ray14 dataset, demonstrating that ResNet50 achieved the highest accuracy of 83.57% in classifying 14 thoracic diseases.

Despite these advances, several limitations persist. Most existing studies primarily rely on overall accuracy as the sole performance metric, which can be misleading in the presence of highly imbalanced datasets, a common characteristic of medical imaging corpora. Precision, recall, F1-score, and other metrics give a well-balanced insight into the performance of the model, especially in clinically sensitive applications where false negatives can have severe consequences [12]. Furthermore, limited attention has been paid to explaining the model's decision-making process, an essential requirement for clinical deployment, where understanding the model's decision-making process is critical for physician trust and patient safety [13].

This paper addresses these gaps by proposing a comprehensive model for classifying lung diseases using X-ray images. The proposed approach enhances data preprocessing with advanced techniques for mitigating class imbalance, including weighted focal loss and synthetic oversampling. Multiple CNN architectures are benchmarked using a diverse set of evaluation metrics, and

International Journal of Current Science Research and Review
ISSN: 2581-8341
Volume 08 Issue 07 July 2025
DOI: 10.47191/ijcsrr/V8-i7-52, Impact Factor: 8.048
IJCSRR @ 2025
www.ijcsrr.org

explainability is incorporated via Grad-CAM to visualize disease-localized regions. The primary objectives of this research are threefold:

1. To provide a more robust and holistic evaluation of classification models using multiple performance metrics.
2. To improve the handling of class imbalance through advanced loss functions and augmentation strategies.
3. To enhance model transparency and clinical trust via explainable AI techniques.

Through these contributions, the study aims to advance the field of AI-assisted medical diagnosis by providing a more reliable, interpretable, and clinically useful framework for classifying lung diseases from CXR images.

RELATED WORK

Among the most obvious uses of DL in medical imaging has been with regard to the classification of chest CXR images related to disease of the lung. Several studies have used CNNs and transfer learning to make thoracic pathology detection/ classification automatically. Even though these strategies have proven effective, some methodological gaps are still highlighted, such as the rigor of the evaluation, the imbalance of the dataset, and the explainability of the model.

Aziz et al. [14] reviewed ML techniques for diagnosing spondylolisthesis from medical images, highlighting the effectiveness of CNNs and other AI methods for accurate detection and classification. Their findings focus on typical issues within medical imaging, like data quality and interpretability, that fit the focus and interest of this research project in the role of improving the performance and explainability of disease detection models.

Wang et al. [15] proposed NIH ChestX-ray14, a large-scale benchmark dataset using more than 112,000 CXR images with labels (conditions) of 14 thoracic conditions. This database has since then formed a baseline for creating DL models in radiology. They developed ChestNet as a CNN that applies ResNet-152 as the model with a sigmoid activation layer of output to facilitate multi-label classification. Yet, they mainly used classification accuracy and AUC to define their goals, and the topic of imbalance and interpretability did not arise in their research.

This was further work by Baltruschat et al. [16] who included patient metadata in their model, including age and gender, using a ResNet-50 classifier. They compared models that were trained completely new and some that were trained using transfer learning and found those with pre-trained weights to be more accurate. The paper was primarily based on guessing and AUC, offering little idea of how the model behaves when pointing out minority classes.

Kavitha et al. [11] evaluated multiple transfer learning models, including ResNet-18, ResNet-50, VGG-19, and a custom CNN, on the NIH dataset. Their results indicated that ResNet50 achieved the highest accuracy at 83.57%. Although the study comprehensively addressed model training and data preprocessing, it lacked the use of additional evaluation metrics, such as F1score and precision-recall curves. More critically, the absence of model interpretability tools makes clinical deployment challenging, particularly when physicians require visual explanations of AI decisions.

Other studies have attempted to improve performance in class-imbalanced scenarios. Kabiraj et al. [17] proposed CX-Ultranet, an EfficientNet-based architecture with a multi-class cross-entropy loss, which achieved a mean prediction accuracy of 88%. Souid et al. [18] employed MobileNetV2 with metadata fusion, achieving high classification performance using a rebalanced dataset. Yet, both studies omitted the use of explainability methods, and neither evaluated model robustness under imbalanced conditions using appropriate statistical metrics.

Recent advances in explainable AI (XAI) have introduced methods such as Grad-CAM [13], which provides visual heatmaps indicating the spatial focus of a model's prediction. While widely used in other domains, few studies have incorporated Grad-CAM or similar techniques in lung disease classification tasks. This omission limits the clinical reliability of AI models and their acceptability among medical professionals [19].

Ghazal [20] used transfer learning of Inception-V4 and KNN as the method of detecting Parkinson's in handwriting images and achieved an accuracy of 93 percent and 0.89 AUC.

The existing literature demonstrates significant progress in automated lung disease classification using CNNs and transfer learning. However, a comprehensive approach that integrates robust evaluation metrics, advanced methods for handling imbalances, and explainable outputs remains underexplored. The present study addresses these gaps by proposing a framework that combines multiple performance metrics, weighted focal loss, synthetic oversampling, and Grad-CAM visualization for model explainability.

International Journal of Current Science Research and Review
ISSN: 2581-8341
Volume 08 Issue 07 July 2025
DOI: 10.47191/ijcsrr/V8-i7-52, Impact Factor: 8.048
IJCSRR @ 2025
www.ijcsrr.org

METHODOLOGY

This study proposes a robust and interpretable DL model for multi-label classification of lung diseases from chest X-ray images. The methodology encompasses data preprocessing, class imbalance mitigation, model selection, evaluation strategy, and integration of explainability.

Dataset Description

The NIH ChestX-ray14 dataset [21] is employed for this research. It contains 112,120 frontal-view X-ray images of 30,805 patients, annotated with 14 disease classes and a "No Finding" category. The labels were extracted using Natural Language Processing (NLP) from corresponding radiology reports. Table 1 summarizes the class distribution, highlighting a significant imbalance across disease categories (e.g., only 13 images for Hernia vs. 3,044 for "No Finding").

Table 1. NIH ChestX-ray14 dataset

Class	*Number of Images*
Atelectasis	508
Cardiomegaly	141
Effusion	644
Infiltration	967
Mass	284
Nodule	313
Pneumonia	62
Pneumothorax	271
Consolidation	226
Edema	118
Emphysema	127
Fibrosis	84
Pleural Thickening	176
Hernia	13
No Finding	**3044**

Data Pre-processing

The preprocessing pipeline of our study has such steps as: resizing images, normalizing data, converting to a tensor and stratified data splitting. Every step is intended to prepare the NIH ChestX-ray14 dataset for data training, validation, and testing.

A. Image resizing

All the chest X-ray images were scaled to a constant resolution of 224 224 images with the aid of bilinear interpolation. This is compatible with the input image size of popular CNN architectures like ResNet and EfficientNet.

B. Normalization

To standardize the dynamic range of pixel intensities, images were normalized to have values within [0,1]. This normalization ensures numerical stability during backpropagation by avoiding large activation values and helps accelerate model convergence.

C. Tensor conversion

The conversion of the normalized image into a type of tensor that can be used in PyTorch was implemented. These tensors facilitate efficient matrix math on GPUs, which makes parallel training feasible with batched data.

International Journal of Current Science Research and Review
ISSN: 2581-8341
Volume 08 Issue 07 July 2025
DOI: 10.47191/ijcsrr/V8-i7-52, Impact Factor: 8.048
IJCSRR @ 2025
www.ijcsrr.org

D. Data splitting

The data was stratified and divided into training (70%), validation (15%), and testing (15%) groups. With stratified sampling, every subgroup conserves the class label distribution of the whole.

Data Augmentation and Class Imbalance Mitigation

Class imbalance is a critical challenge in medical imaging datasets such as NIH ChestX-ray14, where specific disease categories (e.g., Hernia, Fibrosis) are significantly underrepresented compared to others (e.g., Infiltration, No Finding). To address this issue, we implemented a hybrid strategy that combines advanced data augmentation, synthetic oversampling, and a custom loss function.

A. Augmentation

Data augmentation techniques were applied stochastically to the training images to increase the adequate size of the training dataset and improve generalization. The goal is to simulate realistic variations in X-ray acquisition without altering the underlying pathology. Random rotation (±10°), horizontal flipping, zooming, and contrast adjustments were applied to increase minority class samples artificially.

B. SMOTE for Images

The standard SMOTE algorithm [22] generates new samples in feature space by interpolating between a minority sample and its nearest neighbors. In this study, we adapted SMOTE for image data by extracting feature vectors using a pre-trained CNN encoder (e.g., ResNet18 up to the penultimate layer), then generating synthetic samples in the embedded feature space.

Let $x_i \in R^d$ be a feature vector from the minority class, and x_{NN} its nearest neighbor. A synthetic sample x_{new} is generated as:

$$xnew = x_i + \lambda.(xNN - x_i), \lambda \sim U(0,1) \quad (1)$$

C. Weighted Focal Loss

To mitigate the bias toward majority classes during model optimization, we employed weighted focal loss [23], which dynamically down-weights well-classified examples and emphasizes hard, misclassified instances.

For binary classification, the focal loss for a given sample is defined as:

$$L_{focal} = -\alpha_t(1 - p_t)\gamma log\ (p_t) \quad (2)$$

where p_t is the probability for the true class, γ is the focusing parameter, and α_t balances class weights.

For multi-label settings, this loss is applied independently for each class and aggregated:

$$L\text{total} = \Sigma Cc=1\ L\text{focal}(c) \quad (3)$$

This formulation ensures that rare classes contribute more significantly to the overall loss, thereby guiding the network to pay attention to infrequent but clinically critical categories.

Model Architectures

Four CNNs, namely ResNet18, ResNet50, EfficientNet-B0, and a home-designed CNN, were compared to assess the effectiveness of various DL architectures in predicting thoracic diseases using chest X-ray pictures. These models were chosen for their varying depth, complexity, and parameter efficiency, enabling a balanced comparison between lightweight and deeper architectures. Each model was initialized with ImageNet pre-trained weights to leverage transfer learning, which is especially effective in medical imaging tasks with limited labeled data. Figure 1 illustrates the overview of the proposed framework.

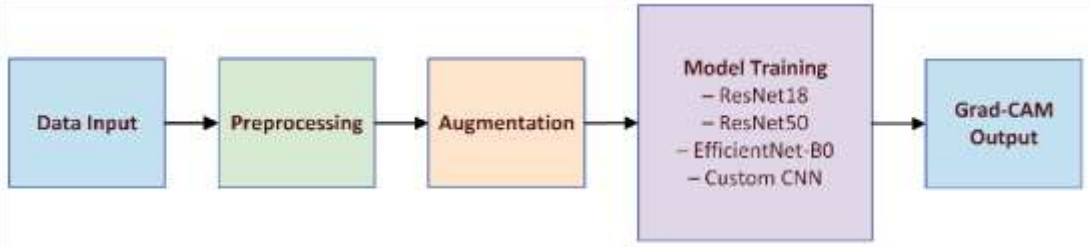

Figure 1: Overview of the Proposed Framework

International Journal of Current Science Research and Review
ISSN: 2581-8341
Volume 08 Issue 07 July 2025
DOI: 10.47191/ijcsrr/V8-i7-52, Impact Factor: 8.048
IJCSRR @ 2025
www.ijcsrr.org

A. ResNet18 and ResNet50

Introduced by He et al. [24], ResNet architecture revolutionized DL by addressing the vanishing gradient problem through identity shortcut connections. The multi-label classification in both models is based on the global average pooling and a subsequent fully connected neural network with sigmoid activation.

B. EfficientNet-B0

EfficientNet, proposed by Tan and Le [25], is a family of CNNs that scales depth, width, and resolution in a principled and efficient manner. This design significantly reduces the number of parameters while maintaining high accuracy. It is particularly suited for medical imaging tasks where computational efficiency and generalization are critical.

C. Custom CNN architecture

A custom CNN model was designed to offer a lightweight and interpretable alternative to transfer learning approaches. The input is a 224×224×1 grayscale chest X-ray image. The network consists of two convolutional layers with 32 and 64 filters (kernel size 3×3), followed by max-pooling. A third convolutional layer with 128 filters is then applied. The output is flattened and passed through a dense layer with 128 ReLU-activated neurons, followed by a dropout layer (rate 0.5) to reduce overfitting. The final output layer uses 15 sigmoid-activated neurons for multi-label classification. This architecture, illustrated in Figure 2, balances simplicity, efficiency, and effective feature learning for chest disease detection.

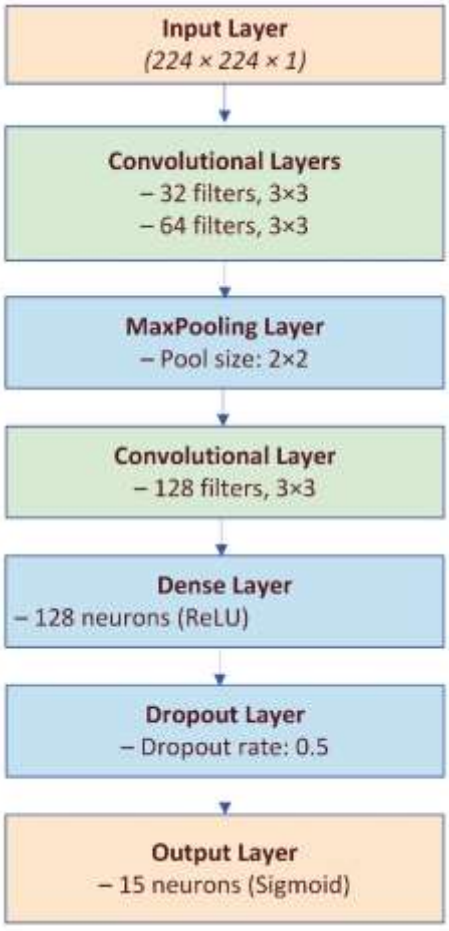

Figure 2. Custom CNN architecture

This architecture was optimized to serve as a fast, interpretable baseline with reduced training time. The key characteristics of the evaluated models, including their architectural depth, parameter counts, and core design features, are listed in Table 2.

International Journal of Current Science Research and Review
ISSN: 2581-8341
Volume 08 Issue 07 July 2025
DOI: 10.47191/ijcsrr/V8-i7-52, Impact Factor: 8.048
IJCSRR @ 2025
www.ijcsrr.org

Table 2. Summary of evaluated CNN architectures for lung disease classification

Model	*Depth*	*Parameters*	*Key Feature*
ResNet18	18	~11M	Residual connections for shallow depth
ResNet50	50	~25M	Bottleneck residual blocks
EfficientNet-B0	~30	~5.3M	Compound scaling, MBConv blocks
Custom CNN	~10	~1M	Lightweight, domain-specific design

Model Training

The training phase involves fine-tuning the pre-trained CNNs on the NIH ChestX-ray14 dataset to optimize multi-label thoracic disease classification performance. The training procedure was carefully designed to include optimal hyperparameter selection, regularization, early stopping, and learning rate adjustment strategies. Below are the key components of the training configuration.

A. Transfer learning and fine-tuning

All models were initialized with weights pre-trained on the ImageNet dataset, which provides generalized low-level feature representations suitable for transfer to medical imaging tasks. Table 3 summarizes the training configurations.

- Each model's final fully connected classification layer was replaced with a sigmoid-activated output layer of size 15 to handle multi-label classification (14 diseases + No Finding).
- The remaining layers were fine-tuned using backpropagation during training.

Table 3. Training configurations

Hyperparameter	*Value*
Optimizer	Adam
Initial LR	1×10^{-4}
Batch Size	32
Epochs	50
Loss Function	Weighted Focal Loss
Output Activation	Sigmoid (multi-label)
Early Stopping	Patience = 5 epochs (val loss)
Scheduler	ReduceLROnPlateau

B. Loss function

Given the imbalanced multi-label nature of the dataset, the weighted focal loss (introduced in Section 3.3) was used to optimize the model

C. Optimization strategy

Adam optimizer was used to tune the models with a parameter-specific learning rate that adjusts with the help of the first and second gradient moments.

D. Early stopping

To avoid overfitting and improve convergence:

- Early stopping has also been applied to stop training whenever the validation loss fails to reduce after five subsequent epochs.

International Journal of Current Science Research and Review
ISSN: 2581-8341
Volume 08 Issue 07 July 2025
DOI: 10.47191/ijcsrr/V8-i7-52, Impact Factor: 8.048
IJCSRR @ 2025
www.ijcsrr.org

- The learning rate was reduced by 0.1 factor every time validation loss plateaued, finding a learning rate scheduler (ReduceLROnPlateau), to help the optimizer out of local minima.

Evaluation Metrics

To capture the nuanced performance in an imbalanced multi-label setting, the following metrics were computed:

- **Accuracy**: For completeness, but not relied upon solely.
- **Precision, Recall, F1-score**: Per-class and macro-averaged to ensure fair evaluation of minority classes.
- **Area Under Curve (AUC)**: AUC-ROC determines the ability of the model to divide the classes based on the consideration of its produced probabilities.

Explainability with Grad-CAM

In clinical applications, model interpretability is as critical as predictive accuracy. Physicians and radiologists require highconfidence predictions and explanations that justify those predictions. To address this need, this study employs Gradient-weighted Class Activation Mapping (Grad-CAM) [13], a widely used technique for generating visual explanations from CNN-based models. Grad-CAM allows attention-gaining areas in the input image to be located to provide maximal contribution to the predicted score of a specific class. It points at the part of the input image that most contributes to the decision made by the model by calculating the gradient of the output category about the final convolutional layer:

$$L\text{Grad}_c\text{-CAM} = \text{ReLU}(\sum_k \alpha c_k A k) \quad (4)$$

where α_k^c is the importance of weight for the feature map A^k, and c is the target class. The heatmaps deterministically formed were superimposed on original X-ray images to determine the region of pathologies, in turn, aiding clinical decision-making.

RESULTS AND DISCUSSION

This part gives the statistics of the four CNN models based on ResNet18, ResNet50, EfficientNetB0, and a custom CNN trained and tested using the NIH ChestX-ray14. Accuracy, precision, recall, F1-score, and AUC are tested as multi-label classification measures of performance. Moreover, per-class evaluation and explainability analysis through Grad-CAM are offered to achieve clinical interpretability.

Classification Performance

The results of all the models' classification on the test dataset are provided in Table 4 and Figure 3. Several measures are listed, including macro-averaged precision, recall, F1-score, AUC, and accuracy.

Table 4. The comparison of the performance between models

Algorithm	*Accuracy*	*Precision*	*Recall*	*F1*	*AUC*
ResNet18	0.73	0.61	0.54	0.57	0.79
ResNet50	0.83	0.72	0.68	0.70	0.86
EfficientNet-B0	**0.85**	**0.78**	**0.73**	**0.75**	**0.91**
Custom CNN	0.79	0.67	0.63	0.65	0.83

International Journal of Current Science Research and Review
ISSN: 2581-8341
Volume 08 Issue 07 July 2025
DOI: 10.47191/ijcsrr/V8-i7-52, Impact Factor: 8.048
IJCSRR @ 2025
www.ijcsrr.org

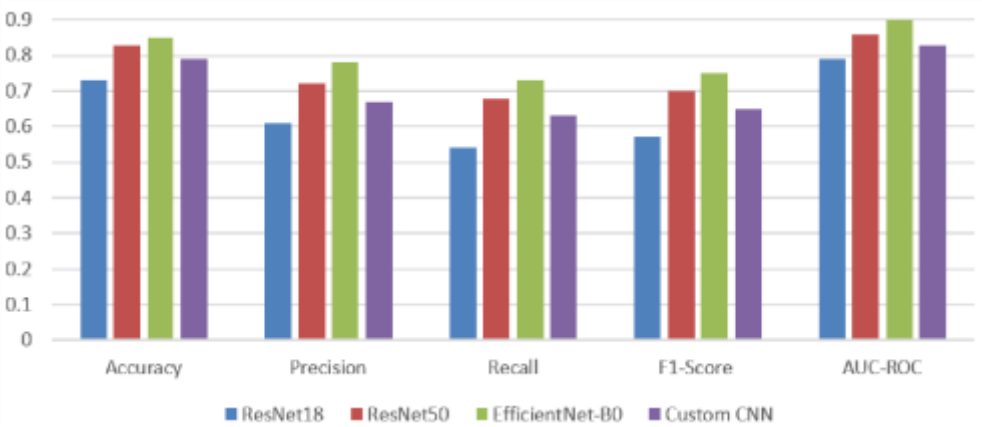

Figure 3. Comparison of model performance across evaluation metrics

As shown, EfficientNet-B0 outperforms all other models across all metrics, demonstrating superior generalization and robustness. Its compound scaling strategy and lightweight MBConv blocks allow deeper feature extraction with fewer parameters than ResNet variants. ResNet50 also performed well and slightly better than the custom CNN, validating the strength of deep residual learning

Class-wise Performance Analysis

Figure 4 presents the per-class F1-score for each model to evaluate model behavior across different disease categories. EfficientNet-B0 consistently yields higher scores, especially for underrepresented diseases such as Emphysema, Fibrosis, and Hernia.

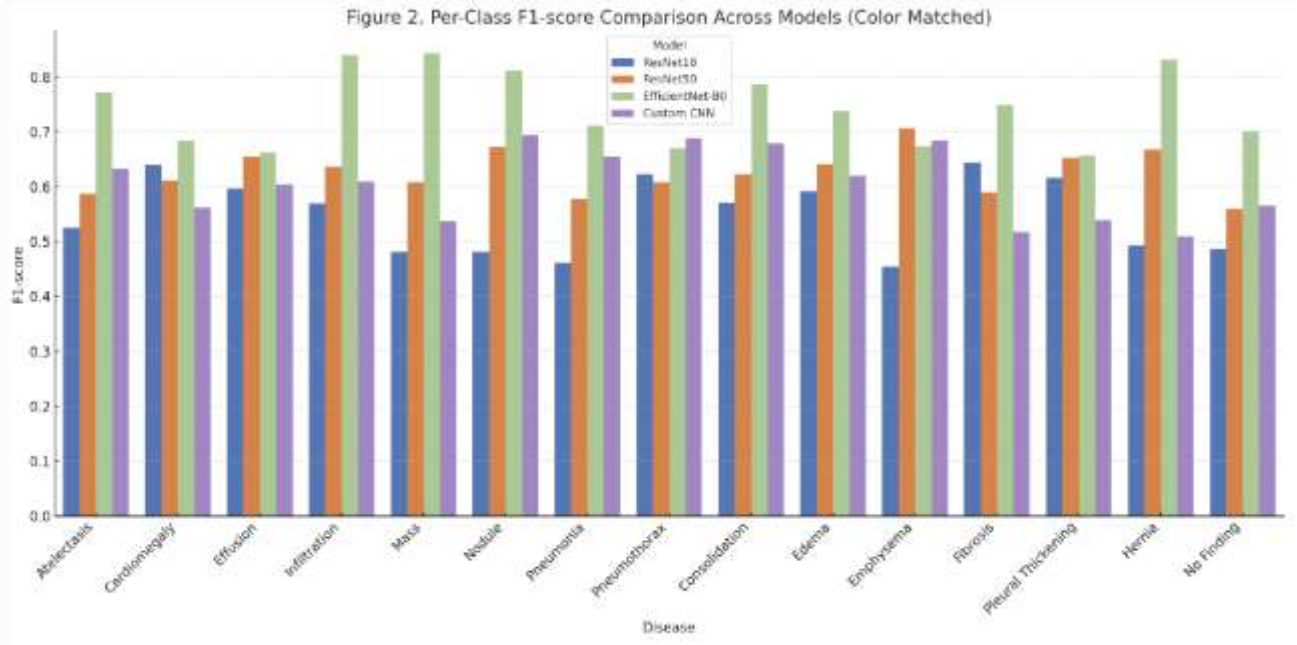

Figure 4. Per-class F1-score comparison across models (ResNet18, ResNet50, EfficientNet-B0, and Custom CNN)

These improvements are attributed to weighted focal loss and targeted data augmentation, mitigating the effects of class imbalance. The relatively lower F1-scores for Pneumonia and Infiltration across all models may reflect label uncertainty in the dataset, as these conditions often exhibit overlapping radiological features.

International Journal of Current Science Research and Review
ISSN: 2581-8341
Volume 08 Issue 07 July 2025
DOI: 10.47191/ijcsrr/V8-i7-52, Impact Factor: 8.048
IJCSRR @ 2025
www.ijcsrr.org

Grad-CAM Visualization and Explainability

Figure 5 presents Grad-CAM heatmaps for three common thoracic diseases: Cardiomegaly, Effusion, and Pneumonia. For each case, the attention maps generated by EfficientNet-B0 highlight clinically relevant regions such as:

- Heart borders in Cardiomegaly,
- Costophrenic angles in Effusion,
- Focal opacities in Pneumonia.

These findings need to not only support the level of accuracy of that model but also assure us that the model can concentrate on the anatomically relevant features, fostering clinical relevance. Such explainability is essential for trust and adoption in real-world settings.

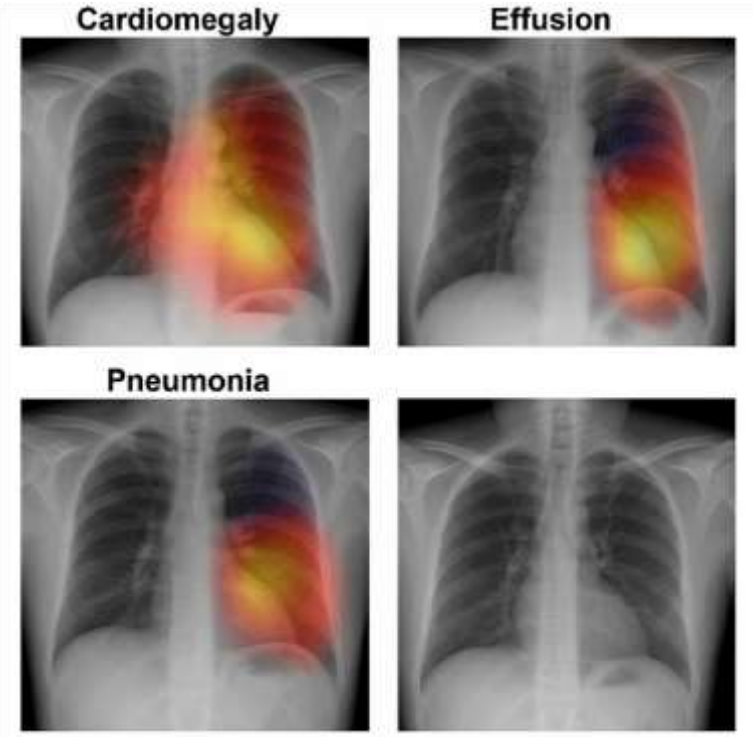

Figure 5. Grad-CAM visualizations highlighting pathological regions in chest X-ray images for Cardiomegaly, Effusion, and Pneumonia.

Comparative Discussion with Previous Work

Compared to the anchor paper by Kavitha et al. [11], which reported accuracy values around 75% using ResNet18, the proposed framework with EfficientNet-B0 achieves higher accuracy (85.43%) and F1-score (0.75). Moreover, this study advances prior work by:

- Introducing a robust loss function (weighted focal loss),
- Incorporating per-class and macro-averaged metrics,
- Applying Grad-CAM for visual interpretability.

These contributions address significant limitations of earlier studies, which often reported only accuracy and lacked explainability analysis.

Clinical Implications

The improved performance of EfficientNet-B0 suggests its potential utility as a clinical decision-support tool, particularly in resource-limited settings where radiological expertise is scarce. Integrating interpretability tools like Grad-CAM further ensures that model decisions align with human reasoning, promoting safe and ethical AI deployment in healthcare.

International Journal of Current Science Research and Review
ISSN: 2581-8341
Volume 08 Issue 07 July 2025
DOI: 10.47191/ijcsrr/V8-i7-52, Impact Factor: 8.048
IJCSRR @ 2025
www.ijcsrr.org

CONCLUSION AND FUTURE WORK

This study proposed a deep learning model for automated multi-label classification of lung diseases from chest X-ray images, addressing several key limitations in prior research. The proposed model demonstrated improved diagnostic performance and clinical transparency by integrating transfer learning with EfficientNet-B0, employing weighted focal loss to mitigate class imbalance, and incorporating Grad-CAM for interpretability. If considering all metrics of the evaluation, the EfficientNet-B0 architecture performed the best in four assessed frameworks, ResNet18, ResNet50, EfficientNet-B0, and custom CNN, which return a macro-average F1-score of 0.75 and an AUC metric of 0.91. Per-class analysis further validated its consistent accuracy across common and rare disease classes, while Grad-CAM confirmed the model's focus on clinically relevant regions. These findings mark a significant advancement over baseline approaches, particularly those that fail to account for class imbalance, underreport per-class performance, or neglect explainability. The integration of model interpretability makes this system accurate and trustworthy, an essential criterion for deployment in medical settings.

Future work may focus on using GANs for data augmentation, incorporating clinical metadata for improved diagnostic accuracy, and extending the model to include disease localization. Real-world validation across diverse clinical settings and developing interactive, interpretable AI tools could further enhance clinical integration and user trust.

REFERENCES

1. V. Cottin *et al.*, “Syndrome of Combined Pulmonary Fibrosis and Emphysema: An Official ATS/ERS/JRS/ALAT Research Statement,” *Am J Respir Crit Care Med*, vol. 206, no. 4, pp. e7–e41, Aug. 2022, doi: 10.1164/rccm.2022061041ST.
2. C. F. Vogelmeier, M. Román-Rodríguez, D. Singh, M. K. Han, R. Rodríguez-Roisin, and G. T. Ferguson, “Goals of COPD treatment: Focus on symptoms and exacerbations,” *Respir Med*, vol. 166, p. 105938, May 2020, doi: 10.1016/j.rmed.2020.105938.
3. V. E. Georgakopoulou, D. A. Spandidos, and A. Corlateanu, “Diagnostic tools in respiratory medicine,” *Biomed Rep*, vol. 23, no. 1, p. 112, 2025.
4. J. C. Y. Seah *et al.*, “Effect of a comprehensive deep-learning model on the accuracy of chest x-ray interpretation by radiologists: a retrospective, multireader multicase study,” *Lancet Digit Health*, vol. 3, no. 8, pp. e496–e506, Aug. 2021, doi: 10.1016/S2589-7500(21)00106-0.
5. Md. R. Hasan, S. M. Azmat Ullah, and S. Md. Rabiul Islam, “Recent advancement of deep learning techniques for pneumonia prediction from chest X-ray image,” *Medical Reports*, vol. 7, p. 100106, Oct. 2024, doi: 10.1016/j.hmedic.2024.100106.
6. A. Ait Nasser and M. A. Akhloufi, “A Review of Recent Advances in Deep Learning Models for Chest Disease Detection Using Radiography,” *Diagnostics*, vol. 13, no. 1, p. 159, Jan. 2023, doi: 10.3390/diagnostics13010159.
7. A. Ebbehoj, M. Ø. Thunbo, O. E. Andersen, M. V. Glindtvad, and A. Hulman, “Transfer learning for non-image data in clinical research: A scoping review,” *PLOS Digital Health*, vol. 1, no. 2, p. e0000014, Feb. 2022, doi: 10.1371/journal.pdig.0000014.
8. W. Xu, Y.-L. Fu, and D. Zhu, “ResNet and its application to medical image processing: Research progress and challenges,” *Comput Methods Programs Biomed*, vol. 240, p. 107660, Oct. 2023, doi: 10.1016/j.cmpb.2023.107660.
9. W. Al-Khater and S. Al-Madeed, “Using 3D-VGG-16 and 3D-Resnet-18 deep learning models and FABEMD techniques in the detection of malware,” *Alexandria Engineering Journal*, vol. 89, pp. 39–52, Feb. 2024, doi: 10.1016/j.aej.2023.12.061.
10. B. Koonce, “EfficientNet,” in *Convolutional neural networks with swift for Tensorflow: image recognition and dataset categorization*, Springer, 2021, pp. 109–123.
11. K. S, R. S. Shudapreyaa, P. Prakash, V. S, V. V, and Y. S, “Classification of Lung Diseases Using Transfer Learning with Chest X-Ray Images,” in *2024 Second International Conference on Emerging Trends in Information Technology and Engineering (ICETITE)*, IEEE, Feb. 2024, pp. 1–6. doi: 10.1109/ic-ETITE58242.2024.10493367.
12. T. Saito and M. Rehmsmeier, “The Precision-Recall Plot Is More Informative than the ROC Plot When Evaluating Binary Classifiers on Imbalanced Datasets,” *PLoS One*, vol. 10, no. 3, p. e0118432, Mar. 2015, doi: 10.1371/journal.pone.0118432.

International Journal of Current Science Research and Review
ISSN: 2581-8341
Volume 08 Issue 07 July 2025
DOI: 10.47191/ijcsrr/V8-i7-52, Impact Factor: 8.048
IJCSRR @ 2025
www.ijcsrr.org

13. R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization," in *2017 IEEE International Conference on Computer Vision (ICCV)*, 2017, pp. 618–626. doi: 10.1109/ICCV.2017.74.
14. R. Rabee, M. S. Jarjees, M. R. Aziz, and A. Asim Hameed, "Machine learning Techniques for Spondylolisthesis Diagnosis: a review," *NTU Journal of Engineering and Technology (NTU-JET)*, vol. 3, no. 2, Jul. 2024, doi: 10.56286/ntujet.v3i2.768.
15. X. Wang, Y. Peng, L. Lu, Z. Lu, M. Bagheri, and R. M. Summers, "ChestX-Ray8: Hospital-Scale Chest X-Ray Database and Benchmarks on Weakly-Supervised Classification and Localization of Common Thorax Diseases," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 3462–3471. doi: 10.1109/CVPR.2017.369.
16. T. Saito and M. Rehmsmeier, "The Precision-Recall Plot Is More Informative than the ROC Plot When Evaluating Binary Classifiers on Imbalanced Datasets," *PLoS One*, vol. 10, no. 3, p. e0118432, Mar. 2015, doi: 10.1371/journal.pone.0118432.
17. A. Kabiraj, T. Meena, P. B. Reddy, and S. Roy, "Detection and Classification of Lung Disease Using Deep Learning Architecture from X-ray Images," 2022, pp. 444–455. doi: 10.1007/978-3-031-20713-6_34.
18. A. Souid, N. Sakli, and H. Sakli, "Classification and Predictions of Lung Diseases from Chest X-rays Using MobileNet V2," *Applied Sciences*, vol. 11, no. 6, 2021, doi: 10.3390/app11062751.
19. A. Holzinger, G. Langs, H. Denk, K. Zatloukal, and H. Müller, "Causability and explainability of artificial intelligence in medicine," *WIREs Data Mining and Knowledge Discovery*, vol. 9, no. 4, Jul. 2019, doi: 10.1002/widm.1312.
20. M. Ghazal, "Parkinson's Disease Detection Based on Transfer Learning," *NTU Journal of Engineering and Technology (NTU-JET)*, vol. 3, no. 3, Sep. 2024, doi: 10.56286/ntujet.v3i3.1173.
21. X. Wang, Y. Peng, L. Lu, Z. Lu, M. Bagheri, and R. M. Summers, "ChestX-Ray8: Hospital-Scale Chest X-Ray Database and Benchmarks on Weakly-Supervised Classification and Localization of Common Thorax Diseases," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, Jul. 2017, pp. 3462–3471. doi: 10.1109/CVPR.2017.369.
22. N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic Minority Over-sampling Technique," *Journal of Artificial Intelligence Research*, vol. 16, pp. 321–357, Jun. 2002, doi: 10.1613/jair.953.
23. T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollar, "Focal Loss for Dense Object Detection," in *2017 IEEE International Conference on Computer Vision (ICCV)*, IEEE, Oct. 2017, pp. 2999–3007. doi: 10.1109/ICCV.2017.324.
24. K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," 2015. [Online]. Available: https://arxiv.org/abs/1512.03385
25. M. Tan and Q. V Le, "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks," 2020. [Online]. Available: https://arxiv.org/abs/1905.11946