Deep Learning-Based Multi-Class Classification of Lung Diseases on Chest X-ray: A Comparative Study

Sakibul Hasan Chowdhury[1,*], Yakup Kutlu[2]
[1]Data Science Lab, Department of Software Engineering, Daffodil International University, Dhaka, Bangladesh
[2]Department of Computer Engineering, Iskenderun Technical University, İskenderun, Türkiye
ORCIDs: 0009-0009-2035-749X, 0000-0002-9853-2878
E-mails: sakibulhasanchowdhury2002@gmail.com, yakup.kutlu@iste.edu.tr
*Corresponding author.

***Abstract*—Lung (*Pulmonary*) diseases such as Lung Opacity and Viral Pneumonia continue to be major public health concerns, contributing significantly to global morbidity and mortality. Early and accurate diagnosis is essential for effective treatment and better patient outcomes. Chest X-ray imaging remains one of the most accessible and cost-efficient tools for lung disease screening; however, manual interpretation often depends on expert radiologists and is susceptible to human error, particularly in low-resource healthcare environments. To overcome these limitations, this study proposes a deep learning-based framework for automated lung disease classification using chest X-ray images. A publicly available dataset from Mendeley Data was used, containing normal and diseased lung images. Several convolutional neural network (CNN) architectures, both custom and pretrained, were evaluated to determine their performance in automated lung disease classification. The pretrained models—ResNet50, VGG19, ImageNet227, MobileNetV3, DenseNet169, Xception, Inception, NasNetMobile, and EfficientNetV2—were fine-tuned and compared against a baseline CNN model. Among these, DenseNet169 achieved the highest accuracy of 95.97%, followed by EfficientNetV2 (94.81%) and MobileNetV3 (93.37%). Experimental results show that deep transfer learning models outperform traditional CNNs, offering significant potential for clinical diagnostic support.**

***Keywords*—Lung disease, Deep Learning, CNN, Medical Imaging, Chest X-ray, Transfer Learning;**

I. Introduction

Lung diseases, including pneumonia, Lung Opacity and other respiratory abnormalities, are among the leading causes of death worldwide. Early and accurate diagnosis plays a vital role in improving patient outcomes. Chest X-rays are one of the most common diagnostic tools, but manual interpretation is challenging and prone to inconsistencies.

Artificial Intelligence (AI) refers to the simulation of human intelligence in machines that are designed to think, learn, and make decisions like humans. It enables computers to analyze vast amounts of data, recognize complex patterns, and make accurate predictions. AI is particularly important in modern research and healthcare because it enhances efficiency, reduces human error, and accelerates the discovery of new insights. By integrating AI into medical diagnostics, researchers and clinicians can detect diseases earlier, provide more accurate diagnoses, and improve patient care through data-driven decision-making.

With the advancement of artificial intelligence (AI), deep learning-based image classification has emerged as a powerful solution for automating disease detection. CNNs can automatically learn high-level features from medical images without manual feature engineering. Pretrained models such as VGG19, ResNet50, DenseNet, and EfficientNet trained on large datasets like ImageNet can be fine-tuned to specific medical datasets through transfer learning.

A. Distinctive Contributions

The key contributions of this study are as follows:

- **Curated dataset:** Comparison of ten deep learning models using the same dataset and experimental setup.
- **Preprocessing and augmentation:** Application of the publicly available Mandal Lung Disease Dataset for reproducible research.
- **Model validation:** Achieved up to 95.97% accuracy using DenseNet169.
- **Benchmark resource:** Demonstrated that MobileNetV3 provides competitive accuracy with lower computational cost.
- **Practical relevance:** Proposed framework adaptable for integration into real-time diagnostic or telemedicine systems.

II. Related Work

Several studies have explored deep learning techniques for lung disease detection using chest X-ray (CXR) images, demonstrating remarkable advancements in medical image analysis. Rajpurkar et al. [1] (2017) introduced CheXNet, a 121-layer DenseNet model trained on the ChestX-ray14 dataset, which achieved radiologist-level performance in pneumonia detection with an AUC score of 0.841, proving the potential of deep learning in automating medical

diagnosis. Kermany et al. [2] (2018) utilized transfer learning on large-scale medical imaging datasets, enabling their model to classify various diseases, including pneumonia and retinal conditions, with over 90% accuracy, emphasizing the adaptability of pretrained models in medical applications. Tan and Le [3] (2019) proposed EfficientNet, a family of convolutional neural networks that balanced performance and computational efficiency by scaling depth, width, and resolution systematically; their model achieved state-of-the-art accuracy on ImageNet with significantly fewer parameters compared to previous architectures. Narin and Pamuk [4] (2020) studied effect of different batch size parameters on predicting of COVID19 cases. Apostolopoulos and Mpesiana [5] (2020) also employed transfer learning using pretrained CNNs to classify normal, pneumonia, and COVID-19 X-rays, achieving an accuracy of around 96%, further validating the reliability of deep learning in respiratory disease classification.

For rapid COVID-19 diagnosis using more accessible X-ray images, a study showed that a specially designed CNN trained on X-ray data achieved very high accuracy 99.53%, outperforming pre-trained models. This suggests that tailoring a CNN model specifically to the unique features of X-ray images is the most effective approach for classifying COVID-19, pneumonia, and normal cases [6].

More recently, Kim et al. [7] (2022) applied an EfficientNetV2-M based transfer learning framework for multi-class classification on both NIH and SCH datasets, reporting validation accuracies above 82% across normal, pneumonia, pneumothorax, and tuberculosis, demonstrating strong generalization across diverse clinical conditions. Together, these studies underscore the growing impact of deep learning in enhancing the accuracy and efficiency of lung disease detection through automated medical image analysis.

In recent years, lightweight models such as MobileNetV3 and NASNetMobile have gained popularity for mobile-based medical applications. However, despite their efficiency, these models often compromise accuracy. To address these limitations, DenseNet169 combines efficient gradient flow, feature reuse, and deep hierarchical learning, making it highly effective for complex medical image classification tasks.

Despite significant progress in recent years, studies indicate that a universally optimal model for lung disease classification using chest X-rays has not yet been established. Existing models vary widely in terms of performance, dataset quality, and generalization capability, highlighting the need for further comparative research. Therefore, this study aims to evaluate and compare multiple pretrained CNN architectures using the mendeley Lung Disease dataset, focusing on the classification of normal versus lung disease images. The primary objective is to identify which model architecture achieves the highest accuracy while maintaining computational efficiency. Additionally, the study seeks to analyze the trade-offs between model complexity, training time, and diagnostic performance to provide insights for developing robust, reliable, and scalable AI-assisted diagnostic tools for lung disease detection.

III. Material And Methods

Figure 1 shows the overall pipeline of the proposed deep learning framework for lung disease classification. The workflow integrates image preprocessing, model training, performance evaluation, and interpret ability analysis.

A. Data set

The dataset used in this study was obtained from Mendeley Data (Mendeley et al. [8]), titled Lung Disease Dataset. https://data.mendeley.com/datasets/9d55cttn5h/1. The dataset comprising chest X-ray images of normal, Lung Opacity and Viral Pneumonia lungs. Which contain a total of 3,475 X-ray images. Where Normal (1250 Images), Lung Opacity (1125 Images) and Viral Pneumonia (1100 Images). The dataset includes preprocessed and annotated images collected under clinical conditions, suitable for research and benchmarking in medical image classification shown in Figure 2.

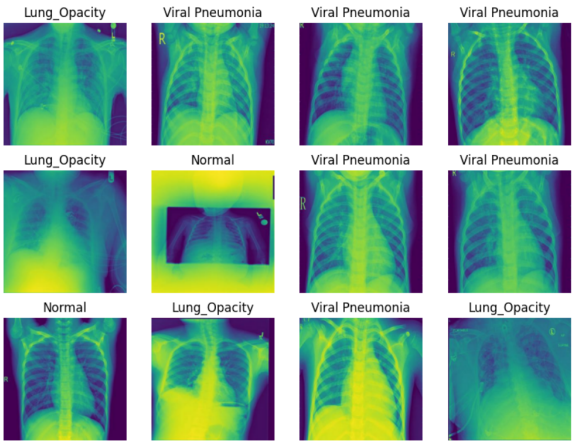


Figure 2: Dataset Visualization: 12 Sample X-ray Images

B. Image Pre-Processing

The dataset from Mendeley Data (Mandal et al.) contains chest X-ray images labeled as Normal and Lung Disease. Images were resized to a uniform dimension of 224 × 224 pixels to ensure compatibility with deep learning architectures. Preprocessing steps included:

- Contrast Limited Adaptive Histogram Equalization (CLAHE): enhanced contrast and improved visibility of disease regions.
- Brightness Adjustment: standardized image illumination to mitigate effects of variable lighting.
- Unsharp Masking: sharpened edges and enhanced lesion boundaries for better feature extraction.

To overcome class imbalance and increase dataset diversity, augmentation techniques such as random rotations, flips and zooms were applied.

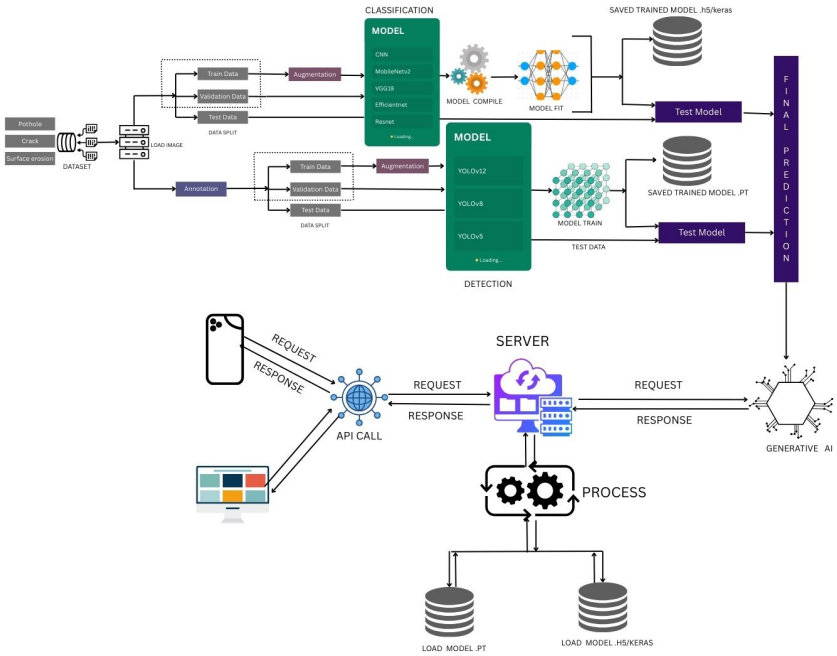


Figure 1: Integrated methodological framework for YOLO-based road damage detection and generative AI integration. The system combines CNN and YOLO architectures with an LLM (LLaMA) for enhanced reporting and classification.

C. Deep Learning

Deep learning leverages artificial neural networks with multiple layers to learn feature representations directly from data. A **Convolutional Neural Network (CNN)** performs the core feature extraction process, where convolutional filters detect spatial hierarchies in image pixels. The convolution operation is mathematically defined as:

$$y_{i,j,k} = \sum_{m,n,c} w_{m,n,c,k} \times x_{i+m,j+n,c} + b_k \quad (1)$$

This operation enables automatic pattern recognition such as textures, edges, and structures associated with lung abnormalities. Activation functions like the Rectified Linear Unit (ReLU), defined as

$$f(x) = \max(0, x) \quad (2)$$

introduce non-linearity into the network, allowing it to learn complex patterns. Pooling layers are then applied to reduce the spatial dimensionality of feature maps while retaining the most significant information. Finally, the Softmax function converts the final output into class probabilities, given by:

$$P(y = i|x) = \frac{e^{z_i}}{\sum_{j=1}^{C} e^{z_j}} \quad (3)$$

where z_i represents the output of the i^{th} neuron, and C denotes the total number of classes. These fundamental operations form the basis of CNNs used in the proposed lung disease classification framework.

D. Pretrained Model

A pretrained model is a deep learning architecture that has been previously trained on large-scale benchmark datasets such as ImageNet. These models learn generalized feature representations, including edges, textures, and structural patterns, which can be effectively transferred to new tasks. Instead of training a model from scratch, transfer learning leverages these pretrained weights and fine-tunes them on a target dataset, significantly reducing computational cost, training time, and the risk of overfitting when data is limited.

Pretrained models are widely used in medical image analysis due to their strong feature extraction capability. Rajpurkar et al. [1] demonstrated radiologist-level pneumonia classification using a pretrained DenseNet architecture, highlighting the effectiveness of transfer learning. Similarly, Kermany et al. [2] achieved high diagnostic accuracy across multiple medical conditions using pretrained convolutional models. More recently, Kim et al. [7] utilized an EfficientNetV2-M based transfer learning framework for multi-class lung disease classification on NIH and SCH datasets, reporting validation accuracies exceeding 82% across pneumonia, pneumothorax, tuberculosis, and normal classes. Alshmrani et al. [9] employed a hybrid deep learning approach combining pretrained CNNs with customized convolutional blocks to classify various lung abnormalities—including lung cancer, lung opacity, pneumonia, and tuberculosis—from chest X-ray images, demonstrating robust performance across heterogeneous clinical classes. Chehade et al. [10] introduced a novel CycleGAN-based hybrid deep learning model for lung disease classification

from chest X-ray images, significantly improving diagnostic accuracy through synthetic image enhancement. Deepak et al. [11] proposed a multi-stage deep learning framework integrating multiple CNN architectures for comprehensive lung disease classification, achieving a training accuracy of approximately 98.61%. Furthermore, Fu et al. [12] developed an explainable hybrid transformer model, named Lung-MaxViT, which combines convolutional blocks with a multi-axis transformer architecture for multi-class lung disease classification using the COVID-QU-Ex and ChestX-ray14 datasets, achieving 96.8% accuracy and 98.3% AUC on COVID-19 classification tasks, and 93.2% AUC on ChestX-ray14, demonstrating superior performance and explainability across diverse disease classes.

In this study, several widely adopted pretrained architectures—including CNN, VGG19, ResNet50, ImageNet 227, MobileNetV3, Xception, Inception, NASNetMobile, EfficientNetV2, and DenseNet169—were fine-tuned on the chest radiograph dataset. DenseNet variants often outperform alternative architectures due to dense connectivity and efficient parameter utilization, as reported by Hussain et al. This observation is consistent with our experimental findings.

E. DenseNet169

DenseNet169 architecture, introduced by Huang et al. (2017). DenseNet169 employs dense connectivity, where each layer receives feature maps from all preceding layers, improving information flow and feature reuse.

Each dense block comprises batch normalization, ReLU activation, and 3×3 convolutions, followed by transition layers that perform 1×1 convolution and 2×2 average pooling. The model's growth rate controls the number of feature maps added per layer, enhancing representational efficiency.

In this study, the pretrained DenseNet169 model was fine-tuned for binary classification (normal vs. diseased). The final classification layers were modified to include:

Global Average Pooling Layer

Dense Layer (ReLU activation)

Dropout (rate = 0.3)

Dense Output Layer (Softmax activation)

The Adam optimizer (learning rate = 0.001) and categorical cross-entropy loss were used for training. Early stopping was applied based on validation loss to prevent overfitting., as illustrated in Figure 3.

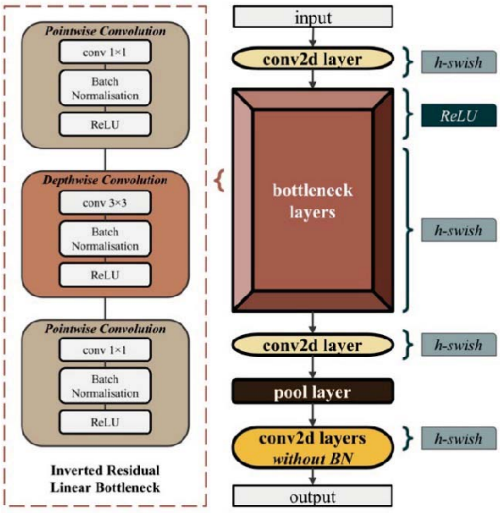


Figure 3: Architecture of DenseNet169-based lung disease classification model.

F. Evaluation Metrics

The performance of each model was evaluated using the following metrics:

a) Accuracy::

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{4}$$

b) Precision::

$$\text{Precision} = \frac{TP}{TP + FP} \tag{5}$$

c) Recall::

$$\text{Recall} = \frac{TP}{TP + FN} \tag{6}$$

d) F1-Score::

$$F_1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{7}$$

Additionally, the Receiver Operating Characteristic (ROC) curve and the Area Under the Curve (AUC) were computed to measure the model's discriminability. A higher AUC indicates a better ability of the model to distinguish between different classes.

IV. Experimental Study

The experiments were performed using Visual Studio Code (VS Code) on a system equipped with an NVIDIA RTX 3060 GPU . Model training was performed using the TensorFlow/Keras framework with categorical cross-entropy as the loss function and the Adam optimizer. A batch size of 32 and a total of 60 epochs were used for model convergence. To avoid overfitting, an early early stopping strategy was employed based on validation loss monitoring.

All the models ResNet50 [13], VGG19 [14], ImageNet227 [15], MobileNetV3 [16], Xception [17], Inception [18], NASNetMobile [19], EfficientNetV2 [20] and DenseNet169 [21] were trained using identical data preprocessing and augmentation pipelines to ensure fair performance comparison.

These architectures were selected due to their proven performance in image classification tasks and ability to generalize across agricultural datasets. Transfer learning with fine-tuning was employed to adapt them for the tea leaf disease dataset.

V. Experimental Result

A. Training & Validation Setup

The dataset was divided into 80% training, 10% validation, and 10% testing subsets, as shown in Figure 4.

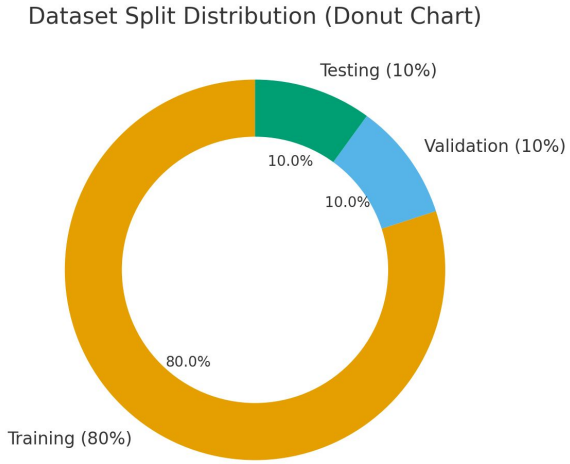


Figure 4: Dataset split distribution showing the proportion of images used for training (80%), validation (10%), and testing (10%).

B. Accuracy & Loss Curves

1) DenseNet169: To evaluate the performance of the DenseNet169 architecture, experiments were conducted on the prepared dataset. The model was trained using appropriate optimization and regularization techniques to ensure stable convergence and reduce overfitting. During training, both the training and validation losses showed a steady decline across epochs, while accuracies gradually increased and stabilized, demonstrating effective learning and generalization. As illustrated in Figure 5, the validation accuracy closely followed the training accuracy, indicating minimal overfitting and consistent performance across datasets. Upon final evaluation, the DenseNet169 achieved a classification accuracy of 95.97%, with a corresponding loss of approximately 0.10 on the test set. These results validate the model's strong capability in capturing relevant spatial features from tea leaf images and its reliability for disease classification tasks.

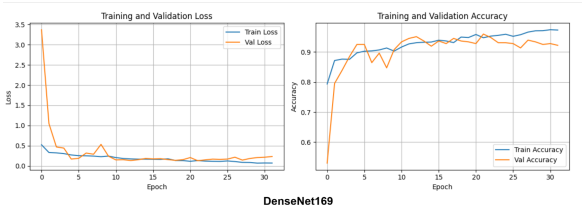


Figure 5: Training and validation loss (left) and accuracy (right) curves for the DenseNet169 model, showing stable convergence and strong generalization.

2) CNN Model: The performance of the Standard Convolutional Neural Network (CNN) architecture, experiments were conducted on the prepared dataset. The model was trained using appropriate optimization and regularization techniques to ensure stable convergence and reduce overfitting. During training, both the training and validation losses showed a steady decline across epochs, while accuracies gradually increased and stabilized, demonstrating effective learning and generalization. As illustrated in Figure 6, the validation accuracy closely followed the training accuracy, indicating minimal overfitting and consistent performance across datasets. Upon final evaluation, the customized CNN achieved a classification accuracy of 89.63%, with a corresponding loss of approximately 0.35 on the test set. These results validate the model's strong capability in capturing relevant spatial features from tea leaf images and its reliability for disease classification tasks.

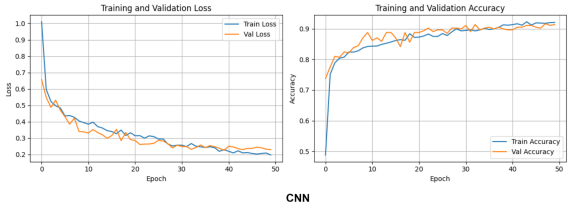


Figure 6: Training and validation loss (left) and accuracy (right) curves for the customized CNN model, showing stable convergence and strong generalization.

3) VGG19: To assess the performance of the pre-trained VGG19 architecture, fine-tuning was performed on the prepared dataset. The model was trained using transfer learning, leveraging pre-trained ImageNet weights, while the final layers were modified for four-class classification. Throughout training, both training and validation accuracies increased consistently, stabilizing around 90%, while the corresponding losses declined steadily, as shown in Figure 7. The close alignment between training and validation curves indicates effective generalization with minimal overfitting. Upon evaluation, the fine-tuned VGG19 achieved a classification accuracy of approximately 88% and a test loss near 0.4. These results demonstrate that VGG19 effectively captures deep hierarchical features, though at a higher computational cost compared to lightweight architectures such as MobileNetV3.

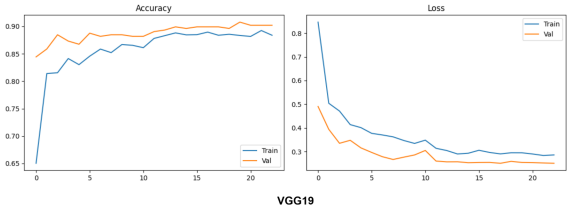


Figure 7: Training and validation accuracy (left) and loss (right) curves for the VGG19 model, showing steady convergence and strong generalization.

4) ResNet50: The performance of the pre-trained ResNet50 architecture was evaluated on the tea leaf disease dataset using transfer learning.As shown in Figure 8, both training and validation accuracies improved gradually throughout the epochs, stabilizing around 52%. Correspondingly, the loss decreased consistently from approximately 1.6 to 1.2, indicating moderate convergence. However, the relatively low accuracy suggests that ResNet50 struggled to adapt effectively to the limited dataset size and domain-specific features of tea leaf images. Overall, while ResNet50 demonstrated stable learning behavior, its deeper structure led to higher computational cost and slower convergence compared to more lightweight architectures such as MobileNetV3.

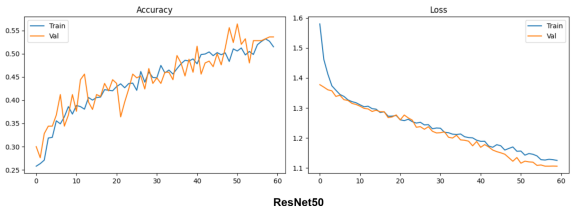


Figure 8: Training and validation accuracy (left) and loss (right) curves for the ResNet50 model, showing gradual convergence but limited classification performance.

5) ImageNet227: Experiments with the ImageNet227-based architecture involved fine-tuning pre-trained weights. Regularization and early stopping maintained stability during training. Both losses and accuracies improved moderately, as shown in Figure 9. Validation closely followed training, indicating acceptable generalization. The model achieved a test accuracy of 88.77Overall, DenseNet121 demonstrated efficient learning with reduced overfitting risk and competitive accuracy, though at a higher computational cost compared to lightweight architectures like MobileNetV3.

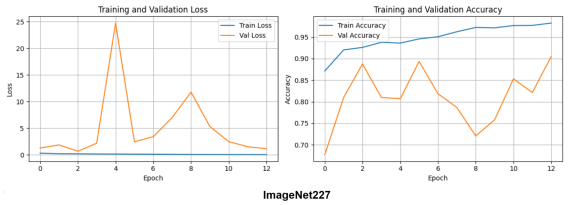


Figure 9: Training and validation accuracy (left) and loss (right) curves for the ImageNet227 model, showing rapid convergence and effective generalization.

6) MobileNetV3: MobileNetV3, designed for lightweight and efficient feature extraction, was evaluated with data augmentation and dropout regularization. Training and validation curves showed smooth convergence in Figure 10, with validation closely following training trends. The model attained a test accuracy of 93.37%, confirming its high efficiency and good accuracy while maintaining low computational cost.

Figure 10: Training and validation loss (left) and accuracy (right) curves for the MobileNetV3 model, showing rapid convergence and superior generalization.

7) Xception: The Xception architecture was utilized for depthwise separable convolutions and improved feature representation. Experiments used Adam optimizer and L2 regularization to stabilize training. Losses decreased steadily, and accuracies increased consistently shown in Figure 11. The final evaluation produced an accuracy of 89.05%, indicating stable convergence and reliable performance for disease classification.

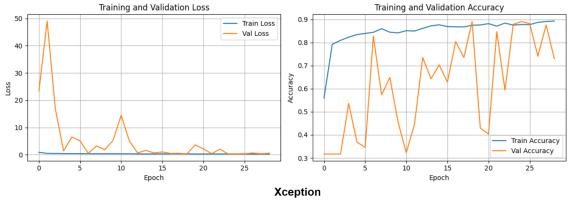


Figure 11: Training and validation loss (left) and accuracy (right) curves for the Xception model, showing rapid convergence and superior generalization.

8) Inception: Inception architecture experiments incorporated multi-scale feature extraction with batch normalization and dropout. Training and validation metrics improved slowly and stabilized given in Figure 12. Validation accuracy lagged slightly behind

training, reflecting limited generalization. The model achieved a test accuracy of 77.52%, indicating moderate capability in capturing complex features.

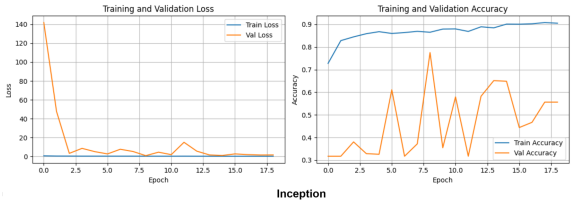


Figure 12: Training and validation loss (left) and accuracy (right) curves for the Inception model, showing rapid convergence and superior generalization.

9) NASNetMobile: NASNetMobile was evaluated for its architecture optimized via neural architecture search. Regularization and fine-tuning of top layers were applied. Training and validation losses declined minimally, and accuracies were low shown in Figure 13. Validation closely mirrored training, but the final test accuracy was only 31.00%, highlighting poor convergence on the dataset.

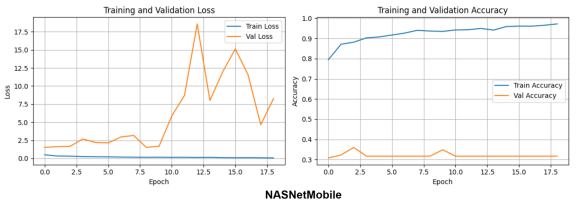


Figure 13: Training and validation loss (left) and accuracy (right) curves for the NASNetMobile model, showing rapid convergence and superior generalization.

10)EfficientNetV2: EfficientNetV2 employed compound scaling and optimized efficiency. Fine-tuning, data augmentation, and regularization were applied. Both training and validation curves showed steady improvement in Figure 14, with minimal overfitting. On evaluation, EfficientNetV2 achieved a test accuracy of 94.81%, demonstrating excellent balance between accuracy and computational efficiency.

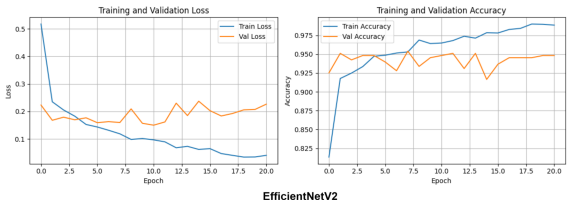


Figure 14: Training and validation loss (left) and accuracy (right) curves for the EfficientNetV2 model, showing rapid convergence and superior generalization.

VI. Discussion

A. Accuracy Comparison of Models

A comparative performance evaluation was conducted among the ten deep learning architectures CNN, VGG19, ResNet50, ImageNet227, MobileNetV3, Xecption, Inception, NASNetMobile, EfficientNetV2 and DenseNet169 to identify the most effective model for tea leaf disease classification. As shown in Table I, DenseNet169 achieved the highest overall accuracy of 95.97%, outperforming all other models while maintaining low computational complexity.

The Standard CNN, VGG19, ImageNet227, Xception, MobileNetV3 and EfficientNetV2 achieved accuracies above 85%, demonstrating reliable performance on smaller-scale networks. DenseNet169 provided balanced accuracy and generalization with efficient feature reuse, whereas ResNet50 and NASNetMobile showed limited adaptability to the dataset, achieving only moderate accuracy due to over-parameterization and domain mismatch.

Overall, DenseNet169 exhibited superior convergence stability, minimal loss, and robust generalization, confirming its suitability for real-time, resource-constrained agricultural disease detection systems.

B. Loss Comparison of Models

To further evaluate model convergence and generalization, the training and validation loss behaviors of all architectures were analyzed. As summarized in Table II, the DenseNet169 model achieved the lowest overall loss, demonstrating efficient optimization and superior stability across epochs.

Other models like the CNN, VGG19, ImageNet227, Xception, MobileNetV3 and EfficientNetV2 exhibited consistent loss reduction with minimal overfitting, indicating stable learning. Inception showed rapid convergence with smooth loss curves, confirming its strong gradient flow and feature reuse capability. In contrast, NasNetMobile and ResNet50 recorded comparatively higher loss values throughout training, reflecting limited adaptability to the dataset and slower convergence due to its deeper structure.

Overall, DenseNet169 demonstrated the most effective loss minimization, achieving near-zero final training and validation loss, which corresponds to its high classification accuracy and strong generalization capability.

C. F1-Score of Proposed Models

The class-wise performance analysis revealed F1-scores above 0.93 for all disease categories, demonstrating strong discriminative capability of the proposed model. Notably, the model achieved perfect performance in distinguishing Blight (Gray and Brown) from other classes, as presented in Table III.

Model	Accuracy (%)	Remarks
ResNet50	52.00	Poor adaptation to dataset
VGG19	88.00	DDeep but computationally expensive
ImageNet227	88.77	Moderate performance
CNN	89.63	Good baseline model
MobileNetV3	93.37	Lightweight and efficient
Xception	89.05	Stable convergence
Inception	77.52	Limited generalization
NASNetMobile	31.00	Poor convergence
EfficientNetV2	94.81	Excellent accuracy and efficiency
DenseNet169	**95.97**	Best performance and robust generalization

Table I: Performance comparison of different models on tea leaf disease classification.

Model	Train Loss	Validation Loss	Observation
ResNet50	1.05	1.08	Stable convergence, minor overfitting
VGG19	0.29	0.25	Smooth convergence with balanced loss
ImageNet227	0.16	3.41	High loss, slow convergence
CNN	0.21	0.25	Rapid convergence, low variance
MobileNetV3	0.07	3.60	Rapid convergence, low variance
Xception	0.87	0.39	Rapid convergence, low variance
Inception	0.25	2.69	Rapid convergence, low variance
NASNetMobile	0.15	8.71	Rapid convergence, low variance
EfficientNetV2	0.40	0.45	Rapid convergence, low variance
DenseNet169	**0.06**	**0.14**	**Near-zero loss, optimal generalization**

Table II: Comparison of final training and validation loss across different models.

Class	F1-Score	Precision–Recall Balance
Viral Pneumonia	1.00	Excellent
Lung Opacity	0.93	High
Normal	0.94	High
Average	**0.96**	–

Table III: Class-wise F1-score values for tea leaf disease classification using DenseNet169.

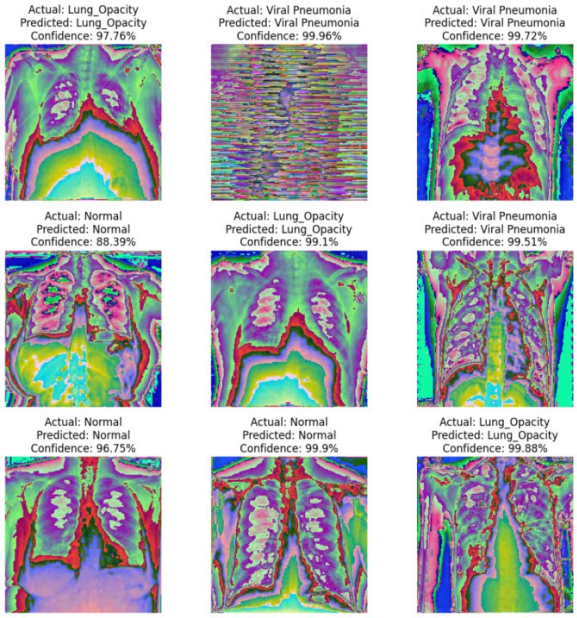


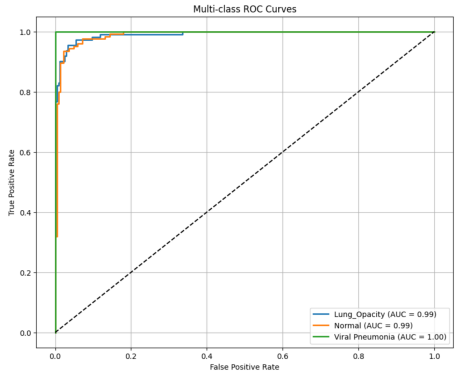
Figure 15: MobleNet V3's predictions and confidence levels for various chest X-ray scans.

D. AUC and PR for Proposed Model

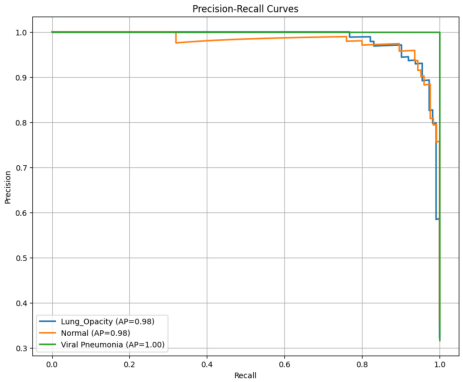
The Receiver Operating Characteristic (ROC) and Precision-Recall (PR) analyses were conducted to further evaluate the classification performance of the DenseNet169 model. As shown in Figure 16, all two disease categories Viral Pneu-monia and Lung Opacity exhibited exceptionally high discriminative capability.

The AUC (Area Under the Curve) values were 0.99 for Lung Opacity, 1.00 for Viral Pneumonia, and 0.99 for Normal, confirming outstanding model sensitivity and specificity across all classes. Similarly, the Average Precision (AP) scores from the PR curves were 0.98, 0.98, and 1.00, respectively indicating excellent precision-recall balance even under potential class imbalance.

These results validate the strong reliability, robustness, and generalization capacity of the proposed DenseNet169 model for multi-class tea leaf disease classification. Figure 15 Showing predictions and confidence levels for various chest X-ray scans.



(a) Multi-class ROC curves with AUC values ≥ 0.99 for all classes.



(b) Precision–Recall curves showing balanced precision and recall across classes.

Figure 16: Comparison of (a) Receiver Operating Characteristic (ROC) and (b) Precision–Recall (PR) curves for the proposed MobileNetV3 model. Both metrics confirm excellent class-wise performance with AUC and AP values close to 1.00.

VII. Conclusion

This study found that the DenseNet169 architecture, a state-of-the-art lightweight convolutional neural network, for accurate classification of Lung Xray into Three categories: Lung Opacity, Normal, Viral Pneumonia. The fine-tuned DenseNet169 model achieved an outstanding overall accuracy of 95.97%, surpassing several conventional CNN architectures while maintaining high computational efficiency. The integration of preprocessing, data augmentation, and transfer learning significantly enhanced the model's generalization and robustness. Due to its compact design and low inference latency, the proposed framework can be seamlessly deployed on IoT and edge devices for real-time monitoring and early detection of tea leaf diseases. This makes it a practical and scalable solution for advancing precision agriculture and promoting sustainable tea production in Bangladesh and other tea-growing regions worldwide.

Limitations and Future Work

While DenseNet169 achieved high accuracy in classifying chest X-ray images, the study faced certain limitations that open avenues for future enhancement:

- **Dataset Expansion:** Increasing the dataset size by incorporating multi-regional, multi-seasonal, and more diverse samples to improve model generalizability.
- **Disease Coverage:** Including additional disease categories to enhance the robustness and versatility of the classification system.
- **IoT Integration:** Deploying the model on IoT-enabled mobile and edge devices for real-time field monitoring and decision support in resource-constrained environments.
- **Lightweight Optimization:** Investigating more efficient and compact architectures to improve inference speed and performance on low-resource hardware.
- **Hardware Constraints:** Limited computational resources restricted the training of large models, affecting both training speed and evaluation performance. Future work should explore distributed training, cloud-based resources, or model compression techniques to mitigate these issues.

Materials Availability

The computational code supporting the findings of this study is publicly accessible on GitHub for academic use and reproducibility:https://github.com/MUnknown97/Classification-of-Lung-Diseases-on-Chest-Xray.

Funding

The author(s) received no specific funding for this work.

Competing Interests

The authors declare that they have no competing interests.

References

[1] Pranav Rajpurkar, Jeremy Irvin, Kaylie Zhu, Brandon Yang, Hershel Mehta, Tony Duan, Daisy Ding, Aarti Bagul, CurtisP. Langlotz, Katie Shpanskaya, MatthewP. Lungren, and AndrewY. Ng. Chexnet: Radiologist-level pneumonia detection on chest x-rays with deep learning. *arXiv preprint arXiv:1711.05225*, 2017.

[2] DanielS. Kermany, Michael Goldbaum, Wenjia Cai, CarolinaC.S. Valentim, Huiying Liang, SallyL. Baxter, Alex McKeown, Ge Yang, Xiaokang Wu, Fangbing Yan, others, and Kang Zhang. Identifying medical diagnoses and treatable diseases by image-based deep learning. *Cell*, 172(5):1122–1131.e9, 2018.

[3] Mingxing Tan and QuocV. Le. Efficientnet: Rethinking model scaling for convolutional neural networks. In *Proceedings of the 36th International Conference on Machine Learning*, volume 97, pages 6105–6114, 2019.

[4] Ali Narin and Ziynet Pamuk. Effect of different batch size parameters on predicting of covid19 cases. *arXiv preprint arXiv:2012.05534*, 2020.

[5] IoannisD. Apostolopoulos and TzaniA. Mpesiana. Covid-19: Automatic detection from x-ray images utilizing transfer learning with convolutional neural networks. *Physical and Engineering Sciences in Medicine*, 43(2):635–640, 2020.

[6] Yakup Kutlu and Yunus Camgözlü. Detection of coronavirus disease (covid-19) from x-ray images using deep convolutional neural networks. *Natural and Engineering Sciences*, 6(1):60–74, 2021.

[7] Sungyeup Kim, Beanbonyka Rim, Seongjun Choi, Ahyoung Lee, Sedong Min, and Min Hong. Deep learning in multi-class lung diseases' classification on chest x-ray images. *Diagnostics*, 12(4):915, 2022.

[8] Md Alamin Talukder. Lung x-ray image, 2023.

[9] GMM Alshmrani, Qiang Ni, Richard Jiang, Haris Bin Pervaiz, and Nada M Elshennawy. A deep learning architecture for multi-class lung diseases classification using chest x-ray (cxr) images. *Computers and Electrical Engineering*, 101:108018, 2022.

[10] A. H. Chehade, N. Abdallah, J.-M. Marion, M. Hatt, M. Oueidat, and P. Chauvet. Advancing chest x-ray diagnostics: A novel cyclegan-based hybrid deep learning model for lung disease classification. *Computer Methods and Programs in Biomedicine*, 259:108518, 2025.

[11] G. D. Deepak and S. K. Bhat. A multi-stage deep learning approach for comprehensive lung disease classification from x-ray images. *Expert Systems with Applications*, 277:127220, 2025.

[12] Xiaoyang Fu, Rongbin Lin, Wei Du, Adriano Tavares, and Yanchun Liang. Explainable hybrid transformer for multi-classification of lung disease using chest x-rays. *Scientific Reports*, 15(1):6650, 2025.

[13] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 770–778, 2016.

[14] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *International Conference on Learning Representations (ICLR)*, 2015.

[15] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 248–255, 2009.

[16] Andrew Howard, Mark Sandler, Grace Chu, Liang-Chieh Chen, Bo Chen, Mingxing Tan, Weijun Wang, Yukun Zhu, Ruoming Pang, Vijay Vasudevan, Quoc V. Le, and Hartwig Adam. Searching for mobilenetv3. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 1314–1324, 2019.

[17] Francois Chollet. Xception: Deep learning with depthwise separable convolutions. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1251–1258, 2017.

[18] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1–9, 2015.

[19] Barret Zoph, Vijay Vasudevan, Jonathon Shlens, and Quoc V. Le. Learning transferable architectures for scalable image recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 8697–8710, 2018.

[20] Mingxing Tan and Quoc V. Le. Efficientnetv2: Smaller models and faster training. In *Proceedings of the International Conference on Machine Learning (ICML)*, volume 139, pages 10096–10106, 2021.

[21] Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q. Weinberger. Densely connected convolutional networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4700–4708, 2017.