Journal of Informatics Engineering (JUTIF) Vol. 7, No. 2, April 2026, Page. 793-810
P-ISSN: 2723-3863 https://jutif.if.unsoed.ac.id
E-ISSN: 2723-3871 DOI: https://doi.org/10.52436/1.jutif.2026.7.2.5016

Exploring Ensemble Architectures on Lung X-Ray Multi-Class Image for Classification Using Convolutional Neural Network and Random Forest

Devin Garmenta Nuriansyah*[1], Putu Desiana Wulaning Ayu[2], Dandy Pramana Hostiadi[3]

[1]Magister Program, Department of Magister Information Systems, Institut Teknologi dan Bisnis STIKOM Bali, Jln. Raya Puputan No. 86, Denpasar 80234, Indonesia
[2,3]Department of Magister Information Systems, Institut Teknologi dan Bisnis STIKOM Bali, Jln. Raya Puputan No. 86, Denpasar 80234, Indonesia

Email: [1]222012021@stikom-bali.ac.id, [2]wulaning_ayu@stikom-bali.ac.id, [3]dandy@stikom-bali.ac.id

Received : Jul 3, 2025; Revised : Oct 9, 2025; Accepted : Nov 9, 2025; Published : Apr 15, 2026

Abstract

The lungs are vital organs that play an important role in the respiratory and circulatory systems. Early detection of lung diseases through medical images, especially *Chest X-Ray* (CXR), is still a challenge due to the limited amount of data and complexity in image interpretation. This research aims to develop an effective image classification approach for lung disease detection by comparing two main methods: direct training using *Convolutional Neural Network* (CNN) and a *hybrid* method involving feature extraction from CNN model, feature selection using *Chi-Square* method, and classification using *Random Forest* algorithm.

To overcome data imbalance and increase variation, *data augmentation* techniques such as rotation, vertical and horizontal flipping, and zooming are used. Four popular CNN architectures are used in training, namely VGG16, ResNet-50, InceptionV3, and MobileNet. After training, features are extracted and stored in .csv format. Next, feature selection using the *Chi-Square* method and classification with *Random Forest* are performed.

The experimental results show that direct CNN training achieves high accuracy, with MobileNet reaching the highest performance at 98.83%. However, this approach requires significant computational resources and longer training time. In contrast, the hybrid method offers competitive accuracy with lower computational demands. The findings highlight the potential of combining deep learning and traditional machine learning to create efficient, accurate, and resource-friendly medical image classification systems. This research has significant implications for supporting early diagnosis of lung diseases, reducing diagnostic workload for medical professionals, and enabling the development of deployable AI-assisted healthcare solutions in resource-limited settings.

Keywords : *Augmentation, Chest X-Ray, Convolutional Neural Network, Feature Extraction, MobileNet, Random Forest.*

1. INTRODUCTION

The lungs are one of the vital organs in the human respiratory system that play a crucial role in the exchange of oxygen and carbon dioxide. Optimal lung function is essential for maintaining both survival and quality of life. However, various pulmonary disorders such as COVID-19, pneumonia, and tuberculosis remain major global health concerns [1], [2].

Lung diseases can lead to serious complications, including respiratory failure, severe infections, and even death. According to data from the World Health Organization (WHO), the prevalence of chronic lung diseases continues to rise, making them one of the leading causes of mortality worldwide. The impact of these diseases extends beyond health, creating substantial social and economic burdens [3], [4].

One of the major challenges in managing lung diseases is the need for accurate and rapid early detection. Traditional diagnostic methods often rely on invasive or high-cost procedures and require

interpretation by experienced medical professionals. Furthermore, the complex and diverse radiological patterns of lung diseases frequently complicate effective diagnosis and monitoring [5], [6], [7], [8].

In recent years, technological advances particularly in medical imaging and artificial intelligence (AI) have opened new opportunities in the early detection of lung diseases. Machine learning approaches, especially Convolutional Neural Networks (CNNs), have shown considerable potential in classifying medical images with high accuracy. Nevertheless, challenges such as limited data availability and complex disease patterns persist and must be addressed [13], [14], [15].

Several studies have proposed the use of CNN architectures such as VGG16, ResNet-50, InceptionV3, and MobileNet for chest X-ray classification with promising results [16], [20]. Other works have explored hybrid methods combining CNN feature extraction with machine learning classifiers such as Support Vector Machine (SVM) or Random Forest (RF) to enhance diagnostic accuracy [21], [23]. However, most previous research still faces limitations related to computational efficiency and the ability to generalize across different disease classes, especially when deployed in resource-constrained environments [24], [26].

Novelty of this study lies in the integration of CNN-based feature extraction with Chi-Square feature selection and Random Forest classification, forming an ensemble framework that balances accuracy and computational efficiency. Unlike previous works that focused only on single CNN models or direct transfer learning, this research emphasizes a hybrid approach capable of addressing class imbalance, improving feature relevance, and enabling deployment on low-resource systems such as mobile or edge devices [27], [28].

Therefore, the objective of this research is to develop an improved early diagnosis model for lung diseases based on chest X-ray images by employing multiple CNN architectures (VGG16, ResNet-50, InceptionV3, and MobileNet), optimized with adaptive learning strategies, and further enhanced through feature selection and Random Forest classification. This study is expected to contribute both methodologically and practically to advancing automated lung disease diagnosis and supporting clinical decision-making.

2. METHOD

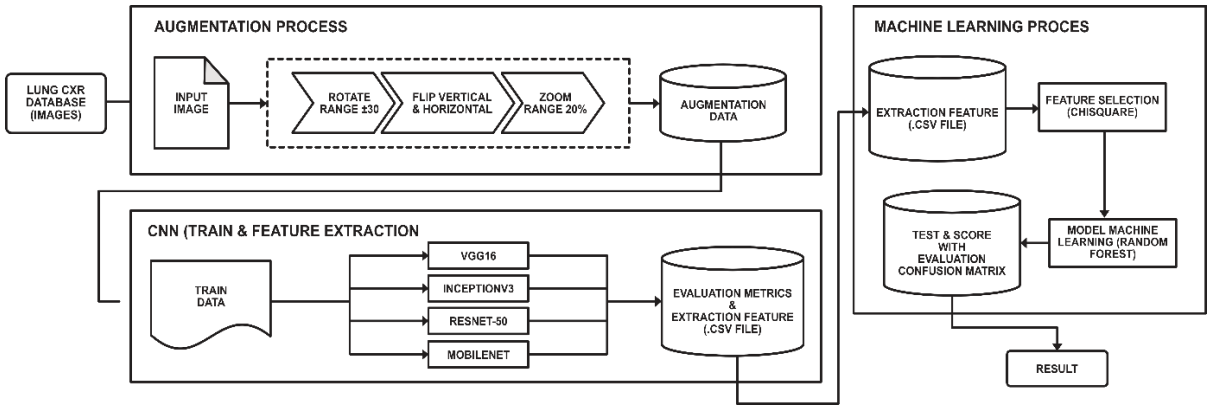

Figure 1. Proposed Method of Classification of Lung Diseases

In this section, the stages of the research implementation are explained in detail, which includes the research design in the form of the proposed method, implementation procedures, techniques used, and testing steps applied. This research is designed to analyze Chest X-Ray images as a non-invasive diagnostic tool in classifying lung diseases. The research methodology involves data processing through augmentation techniques and the implementation of Convolutional Neural Network (CNN) architectures, including VGG16, ResNet-50, InceptionV3, and MobileNet. Additionally, the Chi-Square

feature selection method is employed in conjunction with the Random Forest (RF) machine learning algorithm.

The detail and Workflow proposed model in this research, showed in figure 1, then each stage will be explained in detail and systematically**.** Then each stage will be explained in detail and systematically.

2.1. Chest X-Ray Image

The data used in this study is a public dataset obtained from Kaggle where in this dataset as a whole has a total image of 7097 which is divided from Chest X-ray covid-19, pneumonia, tuberculosis, and normal images. Has a variety of formats ranging from .jpg, .jpeg, and .png. The dataset access link used is as follows (https://www.kaggle.com/datasets/jtiptj/chest-xray-pneumoniacovid19tuberculosis). Chest X-Ray images have advantages in terms of relatively low cost, wide availability, and a fast and non-invasive acquisition process. In the context of digital image processing, Chest X-Ray images are usually represented in two-dimensional grayscale format with a certain level of resolution, depending on the radiographic device used [9]. Chest X-Ray images can be seen in figure 2.

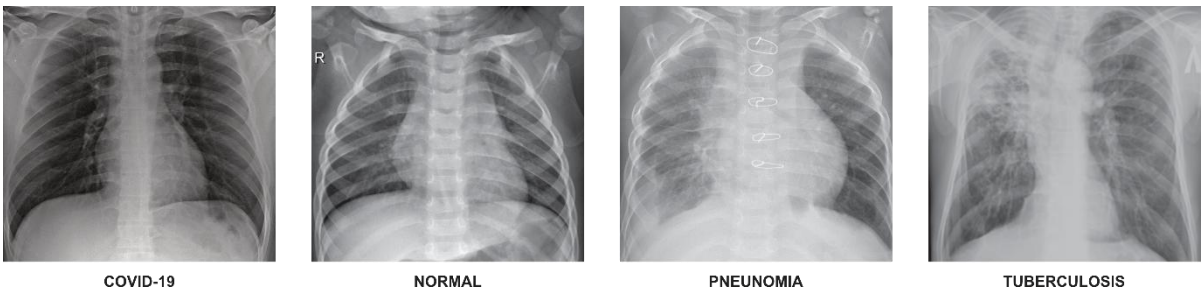

Figure 2. Chest X-Ray Image

Chest X-Ray images also have their own challenges, such as overlapping anatomical structures, variations in image quality, and unbalanced label distribution, all of which can affect the performance of classification models. Therefore, in this study, pre-processing, data augmentation, and the use of pre-trained Convolutional Neural Networks (CNN) models are performed in order to maximize feature extraction and improve classification accuracy.

2.2. Data Augmentation

At the initial stage of this study, after loading the dataset from the chest X-ray image directory, image augmentation was applied as a crucial preprocessing step to balance the data distribution across all classes. Image augmentation is not merely aimed at increasing the volume of data, but it also plays a significant role in enhancing the diversity of image features and improving the generalization capability of deep learning models, especially when addressing the common issue of class imbalance in medical datasets [10], [11].

In the original dataset, the Pneumonia class contained a significantly higher number of images compared to other classes such as Normal or other pulmonary conditions. This imbalance could potentially cause the model to become biased toward the majority class, reducing its ability to learn meaningful patterns from minority classes. To mitigate this, a dual approach was employed: firstly, random undersampling was applied to the Pneumonia class to reduce its dominance; secondly, augmentation techniques were used to increase the number of images in the minority classes, with a target of approximately 2,000 images per class.

The augmentation process involved a variety of transformation techniques, including random rotation, horizontal flipping, zooming, and adjustments to brightness and contrast. These operations

were selected because they introduce realistic variations in the images without altering their semantic meaning, allowing the model to learn a broader set of feature representations. Augmentation was applied exclusively to the training set, while the validation and test sets remained unaltered to ensure objective model evaluation.

Several recent studies have confirmed the importance of augmentation in deep learning-based chest X-ray classification. For instance, Ait Nasser and Akhloufi [10] emphasized the use of data augmentation and image enhancement to address class imbalance and improve model performance across a wide range of chest diseases. Kundu et al. [11] demonstrated that combining data augmentation with an ensemble of deep learning models yielded superior performance in pneumonia detection compared to single-model baselines. Similarly, Garstka and Strzelecki [12] analyzed the influence of augmentation on CNN performance and found that it significantly improved classification accuracy and helped prevent overfitting. In the context of COVID-19 detection, similar observations were made by Singh et al. [13], who showed that transfer learning combined with aggressive augmentation improved sensitivity and robustness of CNN models on CXR images.

Therefore, the augmentation process implemented in this study was not merely a quantitative data enhancement strategy, but a critical method for improving model robustness, reducing overfitting, and achieving higher classification performance in chest X-ray image analysis. A detailed breakdown of the dataset before and after augmentation is presented in table 1.

Table 1. Details of the Number of Images Before and After Augmentation Process

Chest X-ray Image Dataset	Augmentation Process	
	Before	After
Covid-19 Image	460	2003
Normal Image	1341	2004
Pneumonia Image	3875	2003
Tuberculosis Image	650	2002

The augmentation process involves three main stages:

1. Image size adjustment to 224×24 pixels with padding where necessary,
2. Image transformation using ImageDataGenerator (30 degree rotation, 20% zoom, horizontal and vertical flip, and fill mode),
3. Save the results into the "Augmentation Results" directory.

To prevent duplication, a hashing method is applied to filter out identical images. The estimated amount of per-image augmentation can be seen in table 2.

Table 2. Estimated Augmentation per Image

Image Class	**Original Image Count**	**Process**	**Augmented Image**
Covid-19	460	2000/460 = 4.35	4-5 Images
Normal	1341	2000/1341=1.49	1-2 Images
Pneumonia	3875	*Random Sampling*	Multiple Image Removal
Tuberculosis	650	2000/650=3.08	3 Image

2.2.1. CNN Architecture

CNN is one of the *deep learning* architectures specifically designed to process visual data, such as images and videos. Its advantage lies in its ability to recognize complex patterns in image data through a learning process that resembles the way the human brain works in identifying objects [14], [15].

Journal of Informatics Engineering (JUTIF) Vol. 7, No. 2, April 2026, Page. 793-810
P-ISSN: 2723-3863 https://jutif.if.unsoed.ac.id
E-ISSN: 2723-3871 DOI: https://doi.org/10.52436/1.jutif.2026.7.2.5016

Through a series of convolution, pooling, and activation layers, CNNs can extract important features from images automatically without the need for a complicated manual extraction process. This approach has proven to be effective in various fields, especially in image recognition, object classification, and medical image analysis such as X-rays and MRIs. Thus, understanding how CNN works is an important foundation before we go further into an in-depth discussion of its architecture and application. The CNNs used in this research are VGG16, ResNet50, InceptionV3, and MobileNet [14], [16].

2.2.2. VGGNET Architecture

VGGNet is an architecture built by Karen Simonyan and Andrew Zisserman. This architecture states that the depth of a network is an important component for high performance [17]. The deeper the CNN network, the higher the performance. VGGNet has many parameters that require a lot of memory. Most of these parameters are in the *fully connected layer* at the beginning, because if they are removed it does not significantly degrade performance. Simonyan et al. proposed a *deep convolutional network* that has a depth of between 16-19 layers and consists of very small convolution filters [18]. The architecture of VGGNet can be seen in figure 3.

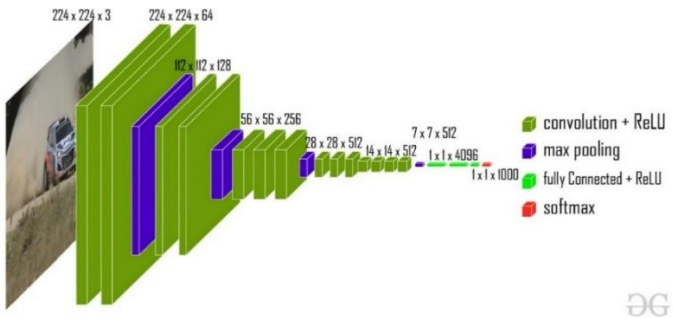

Figure 3. VGG16 Model Architecture [src. Geeksforgeeks]

2.2.3. ResNet50 Architecture

He *et al*. reformulated the layer as a residual learning function by referring to the *i* :r rather than learning an unreferenced function and proposed ResNet which has a maximum depth of 152 layers. This means that ResNet is eight times deeper than VGGNet, but still has lower complexity. In 2015, ResNet won the ILSVRC classification *challenge* [19]. The architecture of ResNet50 can be seen in figure 4.

Broadly speaking, *ResNet (Residual Network)* solves the *vanishing gradient* problem by using *shortcut connections.* The basic equation (1) in *ResNet* is:

$$y = x + F(x, \{W_i\}) \quad (1)$$

Where y is the *output*,x is the *input*, and$F(x, \{W_i\})$ is the *residual* function.

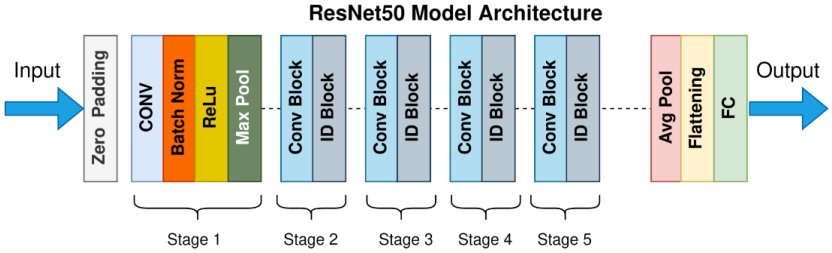

Figure 4. ResNet50 Model Architecture [src. Towardsdatascience]

2.2.4. InceptionV3 Architecture

InceptionV3 is one of the deep neural network (CNN) architectures developed by Google as part of the *Inception* family. *InceptionV3* is designed to be efficient in terms of computation and memory while still maintaining high accuracy in image classification tasks. The architecture utilizes the *Inception* module, which combines multiple convolution filters of different sizes in a single layer, allowing the network to capture different scales of features in the input image [20], [21]. The architecture of InceptionV3 can be seen in figure 5.

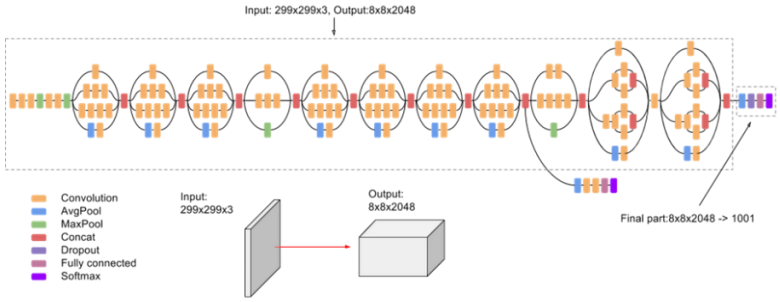

Figure 5. InceptionV3 Model Architecture [src. Digitalocean]

InceptionV3 consists of 48 layers and uses techniques such as *factorized convolutions*, label *smoothing*, and *auxiliary classifiers* to improve performance and stability during training. The *Inception* module allows this architecture to reduce the number of parameters and computational operations, making it more efficient compared to traditional architectures that use standard convolution layers. Equation (2) in the *inception* module is a combination of various convolutions and *pooling*:

$$y = Concat\big(Conv1(x), Con2(x), Conv3(x), Pool(x)\big) \quad (2)$$

Where,$Concat$ is the concatenation operation, and , ,$Conv1(x)Con2(x)Conv3(x)$, and$Pool(x)$ are the results of various *convolutions* and *pooling* applied to the input$x$. *InceptionV3* has shown excellent performance in various image recognition competitions and has been widely applied in research and practical applications, including medical image analysis and object detection [22].

2.2.5. MobileNet Architecture

MobileNet is a lightweight Convolutional Neural Network (CNN) architecture designed to address the high computational demands of traditional CNNs, making it especially suitable for mobile and embedded devices. The core innovation of this architecture lies in replacing standard convolution with depthwise separable convolution, a method that breaks the convolution process into two sequential stages. First, depthwise convolution performs spatial filtering independently on each input channel. Then, the resulting feature maps are linearly combined using pointwise convolution with 1×1 kernels. This approach significantly reduces both the overall model size and computational cost while maintaining effective feature extraction capabilities [23], [24].

Moreover, Prasetyo et al. demonstrated that augmenting the MobileNetV1 architecture with bottleneck and expansion layers—especially tailored for resource-constrained applications (e.g., freshness detection of fish eyes)—can further enhance efficiency while maintaining classification accuracy; their model outperformed standard MobileNet in Scopus-indexed experiments [25]. This highlights not only the flexibility but also the adaptability of MobileNet for domain-specific tasks.

Recent work by Xing et al. (2024) introduced L-MobileNet, a hardware-optimized variation designed for FPGA-based embedded platforms. It achieved approximately 3× speed-up and reduced parameter count by 3.7× relative to MobileNetV2—while preserving comparable accuracy on CIFAR-10 and CIFAR-100 datasets—demonstrating the continued relevance and adaptability of depthwise–pointwise CNN designs in embedded vision systems [26]. The architecture of MobileNet can be seen in figure 6.

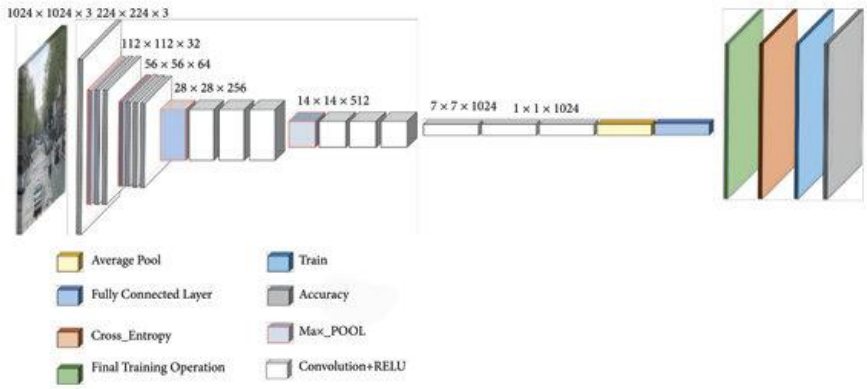

Figure 6. MobileNet Model Architecture [src. Researchgate]

2.2.6. Evaluation Model

The performance of the ensemble model was evaluated using the following metrics: accuracy Equation (3), which gives an overview about the model's performance across all classes and recall Equation 4), which is better suited for imbalanced datasets by capturing all the instances of the minority classes. This metric is important when the cost of missing positive instance is high. The precision Equation (5) metric ensures that the positive predictions made by the model are generally correct. We also used the F1-score metric Equation (6), providing a balance between the precision and recall which is useful when both false positives and false negatives are critical.

$$Accuracy\ (acc) = (TP + TN)/(TP + TN + FP + FN) \quad (3)$$

$$Recall = TP/(TP + FN) \quad (4)$$

$$Precision = TP/(TP + FP) \quad (5)$$

$$F1 - score = (2 \times Precision \times Recall)/(Precision + Recall) \quad (6)$$

3. RESULT

In this study, we systematically integrated several scenarios to perform chest X-ray image classification using deep learning and machine learning approaches.

3.1. Training and Feature Extraction Process Using CNN (Scenario 1)

The first scenario focuses on the classification process utilizing four Pretrained CNN architectures, namely VGG16, ResNet50, InceptionV3, and MobileNet. Each model is trained on the prepared dataset and evaluated using five primary performance metrics: accuracy, AUC, recall, F1-score, and the confusion matrix. In addition to evaluating classification performance, this stage also includes feature extraction from each CNN model, with the extracted features exported in .csv format for subsequent processing.

3.1.1 VGG16

The VGG model is known for its simple multilevel convolutional network structure. In this study, a *pre-trained* variant of VGG16 was used on ImageNet and *fine-tuned* for X-ray image classification. Feature extraction is performed on the *fully connected* layer before the final classification layer. In the training process using VGG16 there were 30 epochs that took place out of the targeted 50 epochs. The best result of VGG16 is at Epoch 27 with an accuracy value of 97.83%, Validation Accuracy 96.50%, Loss of 6.34%, and Validation Loss of . The learning rate used in this study has decreased from the initial learning rate of 0.001, at epoch 24 it drops to 0.005 until epoch 30. The decrease in learning rate is done to help the model achieve optimal performance. The results of using the VGG16 model in this study can be seen in table 3.

The final performance evaluation was conducted using the best VGG16 model obtained at epoch 27. The classification results on the test dataset are presented in table 4, with an overall accuracy of 96.67%, along with high precision, recall, and F1-score values across all four classes.

Table 4. Classification Report of VGG16 Model

Class	Precision (%)	Recall (%)	F1-Score (%)	Support
Tuberculosis (Class 0)	96%	98%	97%	600
Covid-19 (Class 1)	96%	97%	96%	608
Normal (Class 2)	96%	94%	95%	597
Pneunomia (Class 3)	99%	98%	98%	598
Overall Accuracy		96.67%		

After training and evaluating the model, the classification performance was visualized using confusion matrix to understand the ability of the model to distinguish each class, as shown in figure 7.

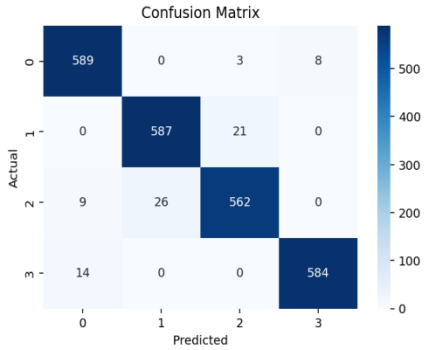

Figure 7. Confusion Matrix Result by VGG16

The classification model was evaluated using a confusion matrix with four classes: Tuberculosis (0), Covid-19 (1), Normal (2), and Pneumonia (3). Results demonstrate high accuracy across all categories, with recall values exceeding 94%.

For Tuberculosis, 589 of 600 images were correctly classified (recall: 98.17%), with most errors involving Pneumonia and Normal. Covid-19 achieved a recall of 96.55%, though 21 cases were misclassified as Normal, reflecting visual similarities between the two conditions. The Normal class reached 94.13% recall, with most errors misclassified as Covid-19, further highlighting this overlap. Pneumonia achieved 584 correct predictions out of 598 (recall: 97.66%), with misclassifications mainly toward Tuberculosis.

Overall, the model demonstrates strong reliability in distinguishing chest X-ray images, even for diseases with complex visual characteristics such as Pneumonia and Tuberculosis. Remaining errors,

particularly between Covid-19 and Normal, could be reduced through enhanced data augmentation and optimized fine-tuning of the CNN architecture.

3.1.2 ResNet50

The ResNet50 model, based on a deep residual architecture, enables effective training of very deep networks without degradation. In this study, a pre-trained ImageNet variant was fine-tuned for chest X-ray classification, with feature extraction performed on the layer preceding the final classifier to leverage the representational strength of residual blocks. Training was conducted for 50 epochs, with the best performance achieved at epoch 33, yielding 99.69% accuracy, 97.63% validation accuracy, 1.01% loss, and 7.67% validation loss. The learning rate was progressively reduced from 0.001 to 0.0005 (epoch 18) and to 0.00025 (epoch 27 onwards), improving convergence and final validation performance. Overall, ResNet50 demonstrated strong stability and high reliability in medical image recognition, as reflected in the results presented in Table 5.

Table 5. Training and Validation Performance of ResNet50 Model per Epoch

Epoch	Train Accuracy (%)	Train Loss (%)	Val_Accuracy (%)	Val_Loss (%)	Learning Rate
Epoch 1	80,27%	61,56%	95,17%	13,54%	0,001
…	…	…	…	…	…
Epoch 18	98,61%	4,83%	97,13%	10,18%	0,0005
…	…	…	…	…	…
Epoch 27	99,50%	1,52%	97,71%	8,12%	0,00025
…	…	…	…	…	…
Epoch 33	99,69%	1,01%	97,63%	7,67%	0,00025

The final performance evaluation was conducted using the best ResNet50 model obtained at epoch 33. The classification results on the test dataset are presented in table 6, with an overall accuracy of 99.69%, along with high precision, recall, and F1-score values across all four classes.

Table 6. Classification Report of ResNet50 Model

Class	Precision (%)	Recall (%)	F1-Score (%)	Support
Tuberculosis (Class 0)	98%	99%	99%	600
Covid-19 (Class 1)	96%	98%	97%	608
Normal (Class 2)	98%	95%	96%	597
Pneunomia (Class 3)	100%	100%	100%	598
Overall Accuracy		98%		

After training and evaluating the model, we visualize the classification performance using confusion matrix to understand the model's ability to distinguish each class, as shown in figure 8, Confusion Matrix Result by ResNet50.

The ResNet-50 model achieved high performance in chest X-ray classification across four classes: Tuberculosis (0), Covid-19 (1), Normal (2), and Pneumonia (3). The confusion matrix shows minimal misclassification with recall values exceeding 94% in all categories.

For Tuberculosis, 595 of 600 images were correctly classified (recall: 99.17%), with only minor errors distributed across other classes. Covid-19 reached a recall of 97.86%, with limited misclassification into Tuberculosis and Normal. The Normal class achieved 94.80% recall, though some overlap with Covid-19 images remains a challenge due to visual similarities. Pneumonia demonstrated

the best performance, with 596 of 598 images correctly identified (recall: 99.67%), and only two misclassified as Tuberculosis.

Overall, ResNet-50 shows excellent generalization and reliability, with particularly strong performance in distinguishing Tuberculosis and Pneumonia, and minor challenges in separating Covid-19 from Normal cases.

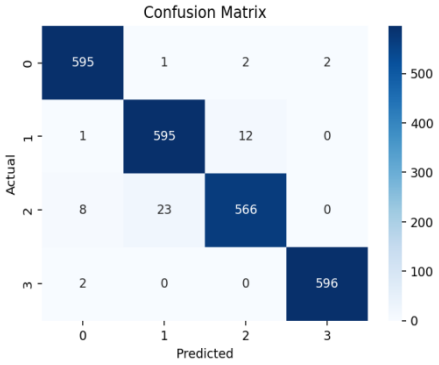

Figure 8. Confusion Matrix Result by ResNet50

3.1.3 InceptionV3

Table 7. Training and Validation Performance of InceptionV3 Model per Epoch

Epoch	Train Accuracy (%)	Train Loss (%)	Val_Accuracy	Val_Loss	Learning Rate
Epoch 1	78.16%	0,6559	0,9159	0,2262	0,001
…	…	…	…	…	…
Epoch 8	96.94%	0,0788	0,9592	0,1412	0,0005
…	…	…	…	…	…
Epoch 19	98.51%	0,0445	0,9634	0,1194	0,00025
…	…	…	…	…	…
Epoch 25	99.14%	0,0267	0,9705	0,1266	0,000125
…	…	…	…	…	…
Epoch 30	99.65%	0,0146	0,9692	0,1201	0,000125
Epoch 31	99.56%	0,018	0,968	0,1149	0,000125

The InceptionV3 model, with its inception modules combining multiple kernel sizes to capture features at different scales, was fine-tuned from a pre-trained ImageNet version for chest X-ray classification. Feature extraction was performed on the fully connected layer preceding the final classifier.

Training was conducted for 50 epochs, with the best performance achieved at epoch 30, yielding 99.65% training accuracy, 96.92% validation accuracy, 1.46% loss, and 12.01% validation loss. The learning rate was initially set at 0.001, reduced to 0.0005 at epoch 8, further decreased to 0.00025 at epoch 19, and finally lowered to 0.000125 from epoch 25 until the end of training. This stepwise adjustment effectively stabilized training and improved convergence in later stages.

Overall, InceptionV3 demonstrated consistent and strong performance, with stable validation results, confirming its suitability for complex medical image classification tasks. The results are presented in Table 7.

The final performance evaluation was conducted using the best InceptionV3 model obtained at epoch 30. The classification results on the test dataset are presented in table 8, with an overall accuracy of 99.65%, along with high precision, recall, and F1-score values across all four classes.

Table 8. Classification Report of InceptionV3 Model

Class	Precision (%)	Recall (%)	F1-Score (%)	Support
Tuberculosis (Class 0)	97%	97%	97%	600
Covid-19 (Class 1)	95%	98%	96%	608
Normal (Class 2)	97%	94%	95%	597
Pneunomia (Class 3)	98%	98%	98%	598
Overall Accuracy		96.67%		

After training and evaluating the model, the classification performance was visualized using confusion matrix to understand the ability of the model to distinguish each class, as shown in figure 9, Confusion Matrix Result by InceptionV3.

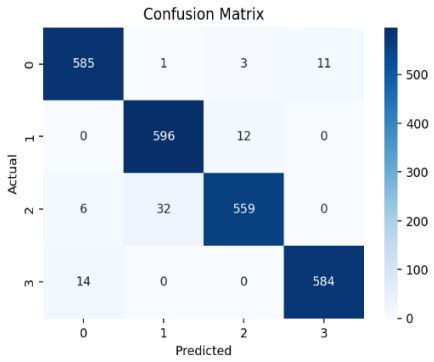

Figure 9. Confusion Matrix Result by InceptionV3

The InceptionV3 model achieved excellent performance in classifying chest X-ray images across four categories: Tuberculosis (0), Covid-19 (1), Normal (2), and Pneumonia (3). The confusion matrix results indicate high accuracy with limited misclassification.

For Tuberculosis, 585 of 600 images were correctly classified (recall: 97.5%), with most errors directed toward Pneumonia. Covid-19 achieved a recall of 98.03% (596/608), with only 12 images misclassified as Normal, showing reliable distinction despite visual similarities. The Normal class reached 93.63% recall (559/597), with most errors against Covid-19, reflecting the challenge of differentiating mild Covid-19 from healthy cases. Pneumonia achieved 97.66% recall (584/598), with errors mainly misclassified as Tuberculosis, suggesting overlapping radiographic features.

Overall, InceptionV3 delivered strong classification performance with recall values above 93% for all classes. While some misclassification occurred between Normal and Covid-19 as well as between Pneumonia and Tuberculosis, the model demonstrated high reliability and suitability for automated lung disease detection using chest X-rays

3.1.4 MobileNet

The MobileNet architecture, designed for computational efficiency through depthwise separable convolutions, was fine-tuned from an ImageNet pre-trained model for chest X-ray classification. Feature extraction was performed on the layer before the final classifier to maintain both accuracy and efficiency.

Training was carried out for 50 epochs using the Adam optimizer, with the best results achieved at epoch 28, yielding 99.62% training accuracy, 98.50% validation accuracy, 1.49% training loss, and 4.21% validation loss. The learning rate started at 0.001, reduced to 0.0005 from epoch 12 to 30, and further lowered to 0.00025 from epoch 31 onwards, effectively stabilizing convergence and enhancing final performance.

Overall, MobileNet demonstrated strong accuracy with high computational efficiency, making it a promising candidate for deployment in edge or mobile device-based medical systems. Results are summarized in Table 9.

Table 9. Training and Validation Performance of MobileNet Model per Epoch (Adam Optimizer)

Epoch	Train Accuracy (%)	Train Loss (%)	Val_Accuracy (%)	Val_Loss (%)	Learning Rate
Epoch 1	78,86%	58,50%	96,75%	10,17%	0,001
…	…	…	…	…	…
Epoch 13	98,69%	3,88%	98,21%	5,36%	0,0005
…	…	…	…	…	…
Epoch 27	99,38%	2,01%	98,46%	3,85%	0,0005
Epoch 28	99,62%	1,49%	98,50%	4,21%	0,0005
Epoch 29	99,54%	1,50%	98,67%	3,99%	0,0005
Epoch 30	99,33%	2,48%	98,83%	4,15%	0,0005
Epoch 31	99,22%	2,26%	98,54%	4,04%	0,00025
…	…	…	…	…	…
Epoch 35	99,31%	1,47%	98,63%	4,17%	0,00025

The final performance evaluation was conducted using the best MobileNet model obtained at epoch 28. The classification results on the test dataset are presented in table 10, with an overall accuracy of 98.83%, along with high precision, recall, and F1-score values across all four classes.

Table 10. Classification Report of MobileNet Model

Class	Precision (%)	Recall (%)	F1-Score (%)	Support
Tuberculosis (Class 0)	99%	99%	99%	600
Covid-19 (Class 1)	98%	99%	99%	608
Normal (Class 2)	98%	97%	98%	597
Pneunomia (Class 3)	100%	100%	100%	598
Overall Accuracy		98.83%		

After training and evaluating the model, we visualize the classification performance using confusion matrix to understand the ability of the model to distinguish each class, as shown in figure 10, Confusion Matrix Result by MobileNet with Adam optimizer.

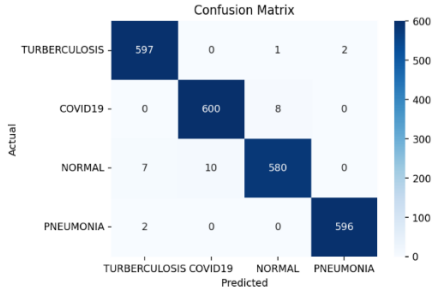

Figure 10. Confusion Matrix Result by MobileNet (Adam Optimizer)

The MobileNet model demonstrated excellent performance in classifying chest X-ray images into Tuberculosis, Covid-19, Normal, and Pneumonia, with recall values exceeding 97% in all categories. For Tuberculosis, 597 of 600 images were correctly classified (recall: 99.5%), with minimal errors directed to Normal and Pneumonia. Covid-19 achieved 98.68% recall (600/608), with only 8 cases

misclassified as Normal, reflecting visual similarity between mild Covid-19 and healthy cases. The Normal class obtained 97.15% recall (580/597), with most errors assigned to Covid-19 and Tuberculosis, highlighting the challenge of differentiating early disease from normal conditions. Pneumonia showed near-perfect performance, with 596 of 598 images correctly identified (recall: 99.66%) and only two misclassified as Tuberculosis.

Overall, MobileNet achieved high stability and generalization with low misclassification rates across all categories. Its lightweight design combined with strong accuracy makes it highly suitable for medical image based diagnostic support, particularly on resource-constrained devices.

3.2. Training Process of Feature Extraction Results on Machine Learning (Scenario 2)

In the second scenario, extracted features were used as input for machine learning-based classification with two testing schemes: with and without feature selection. A ranking-based Chi-Square method was applied to identify the most relevant features prior to classification, after which both selected and non-selected features were classified using the Random Forest algorithm. Evaluation used the same metrics as in the CNN stage, namely accuracy, precision, recall, and F1-score, to allow direct comparison.

Table 11. Evaluation Result

Model	BR (Best Ranked)	AUC	Accuracy	Recall	Precision	F1-Score
VGG16 + CS + RF	250	98.84%	91.46%	91.46%	91.58%	91.51%
VGG16+RF	-	98.71%	91.16%	91.07%	91.30%	91.10%
ResNet50 + CS + RF	900	99.14%	94.03%	94.03%	94.19%	94.10%
ResNet50+RF	-	99.12%	93.31%	93.23%	93.44%	93.28%
InceptionV3 + CS + RF	1650	98.62%	91.81%	91.81%	91.98%	91.87%
InceptionV3+RF	-	98.60%	90.46%	90.36%	90.61%	90.42%
MobileNet + CS + RF	800	99.29%	95.16%	95.16%	95.25%	95.25%
MobileNet + RF	-	99.30%	94.53%	94.52%	94.54%	94.53%

Feature extraction utilized ImageNet-pretrained CNN architectures, where the fully connected layers were removed and features were obtained from the final layer before classification. The resulting high-dimensional numerical vectors were stored in *.csv* format for further processing. These feature sets were then used in Random Forest, both with and without Chi-Square selection, to assess the contribution of feature selection to classification efficiency and accuracy.

The comparative results of each machine learning process based on CNN feature extraction are presented in Table 11.

Four Convolutional Neural Network (CNN) architectures, namely VGG16, ResNet50, InceptionV3, and MobileNet, were evaluated to perform feature extraction from chest X-ray images. The extracted features were stored in *.csv* format, containing numerical representations of the image features. These features were then classified using the Random Forest (RF) algorithm, either directly or after feature selection with the Chi-Square (CS) method.

Based on the evaluation results, the combination of MobileNet + CS + RF achieved the best overall performance, with an AUC value of 99.29%, accuracy and recall of 95.16%, precision of 95.26%, and F1-Score of 95.25%. This indicates that MobileNet was able to extract highly representative features, while the Chi-Square method enhanced the effectiveness of Random Forest classification.

The ResNet50 + CS + RF combination also showed excellent results, obtaining an AUC of 99.14% and an accuracy of 94.03%. Compared with ResNet50 + RF without feature selection, which reached an accuracy of 93.31%, the application of Chi-Square clearly improved accuracy as well as other evaluation metrics.

For VGG16, a slight improvement was observed with Chi-Square, where accuracy increased from 91.16% (without CS) to 91.46% (with CS), and F1-Score increased from 91.10% to 91.51%. A similar trend was also found in InceptionV3, with accuracy improving from 90.46% to 91.81% and F1-Score increasing from 90.42% to 91.87%.

In terms of Best Ranked (BR), which reflects the number of optimal features selected by Chi-Square, InceptionV3 required the largest number of features (1650) to achieve optimal performance. In contrast, VGG16 required only 250 features, while MobileNet and ResNet50 required 800 and 900 features respectively. These findings indicate that the complexity and number of relevant features differ across CNN architectures.

Overall, the results demonstrate that Chi-Square feature selection contributes significantly to the improvement of classification performance. Among the evaluated CNN models, MobileNet and ResNet50 show the most stable and superior performance in classification tasks after feature extraction. To provide a clearer overview of performance, the evaluation results presented in Table 11 are visualized in curve form. These curves illustrate the comparison of AUC, accuracy, recall, precision, and F1-Score across different combinations, making it easier to observe the impact of Chi-Square feature selection on classification results.

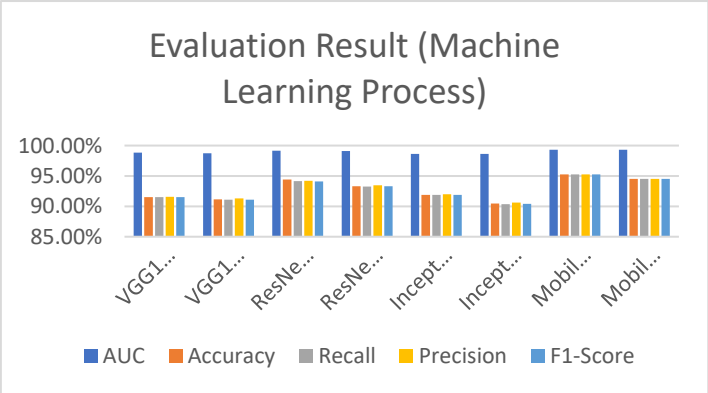

Figure 11. Evaluation Result (Machine Learning Process)

The curves also help highlight the most consistently superior models across various metrics, as well as show the positive contribution of the feature selection process to model accuracy and stability. This visualization supports intuitive comparative analysis and strengthens the understanding of the superiority of certain method combinations over others.

3.3. Final Outcome of the Comparison of Both Processes (Scenario 3)

The third scenario compares CNN-based classification with a machine learning approach using Random Forest. The goal is to identify which method provides better performance in terms of accuracy and evaluation metrics, while also assessing efficiency.

Direct CNN training produced the highest accuracy but required significant computational resources and complex tuning. In contrast, the feature extraction approach combined with Chi-Square selection and Random Forest achieved competitive performance with greater efficiency, making it more practical for real-world applications.

Vol. 7, No. 2, April 2026, Page. 793-810
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.2.5016

Thus, the second approach is a viable alternative for medical image classification systems that prioritize speed, portability, and scalability without major loss in predictive quality. The detailed results are shown in Table 12 and Figure 12.

Table 12. Comparison of CNN Accuracy and CS + RF Accuracy

Model	CNN Accuracy	Ekstraksi + CS + Random Forest Accuracy
VGG16	96.67%	91.46%
ResNet50	98.00%	94.03%
InceptionV3	96.67%	91.81%
MobileNet	**98.83%**	95.16%

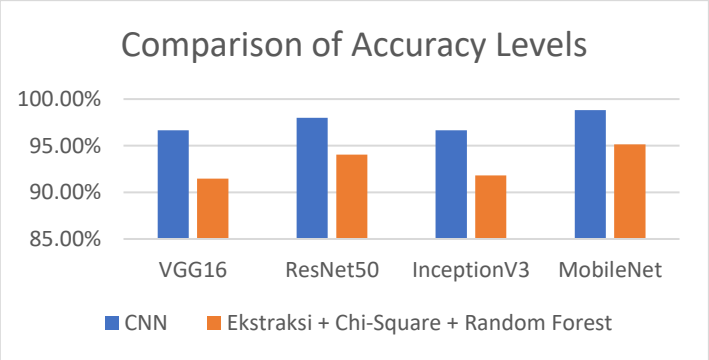

Figure 12. Final Result - Comparison of Accuracy Levels

As an effort to position the findings of this study within a broader context, a comparison was conducted with several previous studies focusing on lung disease classification using chest X-ray images. Table 13 presents a summary of the methods, model architectures, optimization techniques, and classification accuracies achieved by various approaches in earlier research. This comparison aims to evaluate how well the proposed model in this study performs relative to existing methods, including pure deep learning approaches, transfer learning, and hybrid techniques.

Table 13. Comparison with Previous Studies

Researcher	Method / Architecture	Optimization Technique	Accuracy
Narin et al. (2020)	ResNet50 + SVM (Hybrid)	Manual parameter tuning for SVM	98.00%
Apostolopoulos & Mpesiana (2020)	VGG19 (Transfer Learning)	Pre-trained model	98.66%
Apostolopoulos & Mpesiana (2020)	MobileNetV2 (Transfer Learning)	Pre-trained model	96.78%
Hadhoud et al (2025)	Two-Step Hybrid ResNet-50 + Vision Transformer (ViT-B/16) for chest X-ray classification (TB vs Pneumonia)	Transfer learning + fine-tuning (optimizer Adam/SGD)	96.18%
This Study (2025)	Extraction Feature MobileNet + Chi-Square + Random Forest (Hybrid)	Chi-square feature selection + Random Forest Classifier	95.16%
This Study (2025)	**MobileNet (Deep Learning)**	**Adam optimizer**	**98.83%**

Narin et al. (2020) [27] reported an accuracy of 98.00% using a hybrid method involving ResNet50 combined with Support Vector Machine (SVM), applied on a multi-class dataset that included COVID-19, Pneumonia, and Normal categories. Similarly, Apostolopoulos and Mpesiana (2020) [28] achieved accuracies of 98.66% using VGG19 and 96.78% with MobileNetV2 through a transfer learning approach.

Meanwhile, Hadhoud et al. (2024) [29] proposed a two-step hybrid CNN–Transformer approach, using ResNet-50 for local feature extraction and ViT-B/16 for global context. Applied to chest X-rays for Tuberculosis and Pneumonia detection, the model achieved 96.18% accuracy, highlighting the effectiveness of CNN–Transformer integration in pulmonary disease classification.. Compared to these studies, the present research achieved comparable and in some cases superior results. The best performance from direct CNN training was obtained using MobileNet with an accuracy of 98.83%, while the hybrid method combining MobileNet, Chi-Square feature selection, and Random Forest achieved 95.16%. These findings affirm that the proposed methods are not only competitive in terms of accuracy but also offer practical advantages in terms of computational efficiency, particularly when deployed in resource-constrained environments.

The comparative insights from these studies reinforce the contribution of this research in providing a balanced approach between performance and efficiency. Furthermore, the results support the relevance of hybrid methods as a viable alternative for medical image classification, especially in real-world applications where computational resources may be limited.

4. DISCUSSIONS

The results show that the use of Convolutional Neural Network (CNN) architectures such as MobileNet, ResNet-50, InceptionV3, and VGG16 can provide excellent classification performance in detecting lung diseases based on Chest X-Ray images. MobileNet, in particular, proved to excel in terms of efficiency and accuracy, achieving the highest accuracy value of 98.83% in the direct training process, as well as 95.16% in the hybrid approach using Random Forest with Chi-Square feature selection.

The comparison between the pure deep learning approach and the hybrid method illustrates that while direct training with CNN yields very high accuracy, it requires large computational resources. On the other hand, the hybrid approach that combines CNN feature extraction, Chi-Square feature selection, and classification with Random Forest is able to provide competitive accuracy with better computational efficiency. This approach is relevant for implementation on resource-constrained systems, such as edge or mobile devices, thereby opening opportunities for broader and more practical clinical applications.

Performance evaluation based on the confusion matrix shows that all four CNN architectures have high classification capabilities for all classes, namely Tuberculosis, Covid-19, Normal, and Pneumonia. Although there were some misclassifications, most of them were caused by visual similarities between classes, especially between the Covid-19 and Normal classes, and between Pneumonia and Tuberculosis. This is a common challenge in medical image analysis, where the boundaries between classes are not always visually clear.

The addition of Chi-Square feature selection was shown to provide improved performance across the CNN architectures used. The optimal number of features selected varied depending on the architecture, with MobileNet and ResNet-50 showing the most consistent and superior performance in classification after feature selection. This finding is in line with previous literature which states that dimensionality reduction through feature selection can improve the generalization and efficiency of classification algorithms.

Beyond technical performance, the findings of this study highlight an important urgency in the medical field. The high accuracy achieved indicates that automated classification systems based on CNN and Random Forest can play a critical role in supporting radiologists to accelerate early detection of

lung diseases, especially in developing countries where medical experts and advanced diagnostic facilities are limited. The hybrid approach also demonstrates the feasibility of deploying such systems in low-resource environments by reducing computational costs without sacrificing accuracy. This makes the proposed model not only academically significant but also practically impactful in improving public health outcomes through faster, more accessible, and more reliable diagnostic support.

5. CONCLUSION

This study shows that CNN and machine learning can classify lung diseases well. MobileNet achieved the best accuracy at 98.83%, higher than ResNet50 (98.00%), InceptionV3 (96.67%), and VGG16 (96.67%). MobileNet also outperformed a previous study using VGG19 with 98.66%.

Direct CNN training gave high accuracy but needed high resources. To reduce this, a hybrid method was used. It combined CNN feature extraction, Chi-Square selection, and Random Forest. The best hybrid result came from MobileNet + CS + RF with 95.16% accuracy. ResNet50 reached 94.03%, InceptionV3 91.81%, and VGG16 91.46%.

Misclassifications were few and mostly between similar classes. For example, COVID-19 and Normal, or Pneumonia and Tuberculosis.

In conclusion, CNN training gave the highest accuracy, while the hybrid method balanced accuracy and efficiency. MobileNet proved the most effective and consistent. This model is suitable for medical imaging, especially in limited-resource settings.

REFERENCES

[1] E. Alqaissi, "A novel and ultralight convolutional neural network model for real-time detection of infectious lung diseases," *Digit Health*, vol. 11, Jan. 2025, doi: 10.1177/20552076251318155.

[2] I. Mwendo, K. Gikunda, and A. Maina, "Deep transfer learning for detecting Covid-19, Pneumonia and Tuberculosis using CXR images -- A Review," Mar. 2023, [Online]. Available: http://arxiv.org/abs/2303.16754

[3] D. Colombi *et al.*, "Computer-Aided Evaluation of Interstitial Lung Diseases," Apr. 01, 2025, *Multidisciplinary Digital Publishing Institute (MDPI)*. doi: 10.3390/diagnostics15070943.

[4] M. M. Kabir, M. F. Mridha, A. Rahman, M. A. Hamid, and M. M. Monowar, "Detection of COVID-19, pneumonia, and tuberculosis from radiographs using AI-driven knowledge distillation," *Heliyon*, vol. 10, no. 5, Mar. 2024, doi: 10.1016/j.heliyon.2024.e26801.

[5] P. K. Saha, S. A. Nadeem, and A. P. Comellas, "A survey on artificial intelligence in pulmonary imaging," Nov. 01, 2023, *John Wiley and Sons Inc*. doi: 10.1002/widm.1510.

[6] M. Mamalakis *et al.*, "DenResCov-19: A deep transfer learning network for robust automatic classification of COVID-19, pneumonia, and tuberculosis from X-rays," Apr. 2021, doi: 10.1016/j.compmedimag.2021.102008.

[7] A. Al-Kababji, F. Bensaali, and S. P. Dakua, "Scheduling Techniques for Liver Segmentation: ReduceLROnPlateau Vs OneCycleLR," Feb. 2022, [Online]. Available: http://arxiv.org/abs/2202.06373

[8] K. Qazi Waqas, "A Machine Learning-based Method for COVID-19 and Pneumonia Detection," *IgMin Research*, vol. 2, no. 7, pp. 518–523, Jul. 2024, doi: 10.61927/igmin211.

[9] M. S. Sunarjo, H.-S. Gan, and D. R. I. M. Setiadi, "High-Performance Convolutional Neural Network Model to Identify COVID-19 in Medical Images," *Journal of Computing Theories and Applications*, vol. 1, no. 1, pp. 19–30, Aug. 2023, doi: 10.33633/jcta.v1i1.8936.

[10] A. Ait Nasser and M. A. Akhloufi, "A Review of Recent Advances in Deep Learning Models for Chest Disease Detection Using Radiography," Jan. 01, 2023, *Multidisciplinary Digital Publishing Institute (MDPI)*. doi: 10.3390/diagnostics13010159.

[11] J. Garstka and M. Strzelecki, "Pneumonia detection in X-ray chest images based on convolutional neural networks and data augmentation methods," in *Signal Processing - Algorithms, Architectures, Arrangements, and Applications Conference Proceedings, SPA*, IEEE Computer Society, Sep. 2020, pp. 18–23. doi: 10.23919/spa50552.2020.9241305.

[12] S. K. Tiwari, S. Pratap Tomar, S. Swami, and S. Singh Parihar, "Covid-19 Diagnosis from Chest X-Ray Images using Transfer Learning and Data Augmentation," 2024.
[13] R. Kundu, R. Das, Z. W. Geem, G. T. Han, and R. Sarkar, "Pneumonia detection in chest X-ray images using an ensemble of deep learning models," *PLoS One*, vol. 16, no. 9 September, Sep. 2021, doi: 10.1371/journal.pone.0256630.
[14] S. Pendhari, N. Pendhari, and S. Shroff, "Benchmarking Deep Learning Models for Automated MRI-based Brain Tumor Detection: In-Depth Analysis of CNN, VGG16, VGG19, ResNet-50, MobileNet, and InceptionV3," 2024. [Online]. Available: https://www.kaggle.com/datasets/sartajbhuvaji/brain-tumor-
[15] I. D. Mienye, T. G. Swart, G. Obaido, M. Jordan, and P. Ilono, "Deep Convolutional Neural Networks in Medical Image Analysis: A Review," Mar. 01, 2025, *Multidisciplinary Digital Publishing Institute (MDPI)*. doi: 10.3390/info16030195.
[16] M. Caldwell, "Research on Medical Image Diagnosis Models Based on Convolutional Neural Networks," 2025.
[17] M. S. Al Reshan *et al.*, "Detection of Pneumonia from Chest X-ray Images Utilizing MobileNet Model," *Healthcare (Switzerland)*, vol. 11, no. 11, Jun. 2023, doi: 10.3390/healthcare11111561.
[18] Farhana Akter Sunny, S. S. Nakshi, M. Parbhez, and M. A. Rahaman, "COVID-19 Identification System from X-Ray Images of Chest using Deep Neural Network with Transfer Learning," *GUB Journal of Science and Engineering*, vol. 10, no. 1, pp. 53–67, Jul. 2024, doi: 10.3329/gubjse.v10i1.74945.
[19] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," Dec. 2015, [Online]. Available: http://arxiv.org/abs/1512.03385
[20] A. H. M. Linkon, M. M. Labib, T. Hasan, M. Hossain, and M. E. Jannat, "Deep learning in prostate cancer diagnosis and Gleason grading in histopathology images: An extensive study," Jan. 01, 2021, *Elsevier Ltd*. doi: 10.1016/j.imu.2021.100582.
[21] Y. Kumaran S, J. J. Jeya, R. Mahesh T, S. B. Khan, S. Alzahrani, and M. Alojail, "Explainable lung cancer classification with ensemble transfer learning of VGG16, Resnet50 and InceptionV3 using grad-cam," *BMC Med Imaging*, vol. 24, no. 1, Dec. 2024, doi: 10.1186/s12880-024-01345-x.
[22] A. Chanda, "An In-Depth Analysis of CIFAR-100 Using Inception v3," 2025, doi: 10.13140/RG.2.2.30629.20969.
[23] V. S. K. Tangudu, J. Kakarla, and I. B. Venkateswarlu, "COVID-19 detection from chest x-ray using MobileNet and residual separable convolution block," *Soft comput*, vol. 26, no. 5, pp. 2197–2208, Mar. 2022, doi: 10.1007/s00500-021-06579-3.
[24] A. G. Howard *et al.*, "MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications," Apr. 2017, [Online]. Available: http://arxiv.org/abs/1704.04861
[25] E. Prasetyo, R. Purbaningtyas, R. D. Adityo, N. Suciati, and C. Fatichah, "Combining MobileNetV1 and Depthwise Separable convolution bottleneck with Expansion for classifying the freshness of fish eyes," *Information Processing in Agriculture*, vol. 9, no. 4, pp. 485–496, Dec. 2022, doi: 10.1016/j.inpa.2022.01.002.
[26] S. Velu, "An efficient, lightweight MobileNetV2-based fine-tuned model for COVID-19 detection using chest X-ray images," *Mathematical Biosciences and Engineering*, vol. 20, no. 5, pp. 8400–8427, 2023, doi: 10.3934/mbe.2023368.
[27] A. Narin, C. Kaya, and Z. Pamuk, "Automatic detection of coronavirus disease (COVID-19) using X-ray images and deep convolutional neural networks," *Pattern Analysis and Applications*, vol. 24, no. 3, pp. 1207–1220, Aug. 2021, doi: 10.1007/s10044-021-00984-y.
[28] I. D. Apostolopoulos and T. A. Mpesiana, "Covid-19: automatic detection from X-ray images utilizing transfer learning with convolutional neural networks," *Phys Eng Sci Med*, vol. 43, no. 2, pp. 635–640, Jun. 2020, doi: 10.1007/s13246-020-00865-4.
[29] Y. Hadhoud *et al.*, "From Binary to Multi-Class Classification: A Two-Step Hybrid CNN-ViT Model for Chest Disease Classification Based on X-Ray Images," *Diagnostics*, vol. 14, no. 23, Dec. 2024, doi: 10.3390/diagnostics14232754.