Contents lists available at www.infoteks.org

JSIKTI

Journal Page is available to https://infoteks.org/journals/index.php/jsikti

Research article

Classification of Tuberculosis and Pneumonia Lung Diseases in X-Ray Images Using the CNN Method with VGG-19 Architecture

I Dewa Ayu Sri Murdhani [a], Heru Ismanto [b], Didit Suprihanto [c]*

[a] *Magister Program of Informatics, Institut Bisnis dan Teknologi Indonesia, Denpasar, Indonesia*
[b] *Department of Informatics Engineering, Universitas Musamus Merauke, Indonesia*
[c] *Department of Electrical Engineering, Universitas Mulawarman, East Kalimantan, Indonesia*
email: [a,*] sri.murdhani@instiki.ac.id, [b] heru@unmus.ac.id, [c] didit.suprihanto@ft.unmul.ac.id
* *Correspondence*

ARTICLE INFO

Article history:
Received 1 November 2025
Revised 10 November 2025
Accepted 30 December 2025
Available online 31 December 2025

Keywords:
Tuberculosis, Pneumonia, Chest X-ray, Convolutional Neural Network, VGG-19, Classification.

Please cite this article in IEEE style as:
I. D. A. S. Murdhani, H. Ismanto and D. Suprihanto, "Classification of Tuberculosis and Pneumonia Lung Diseases in X-Ray Images Using the CNN Method with VGG-19 Architecture," JSIKTI: Jurnal Sistem Informasi dan Komputer Terapan Indonesia, vol. 8, no. 2, pp. 136-149, 2025.

ABSTRACT

Tuberculosis (TB) and Pneumonia continue to be among the world's leading causes of morbidity and mortality, particularly in low- and middle-income countries where access to advanced diagnostic tools remains limited. Conventional radiological interpretation, while effective, heavily depends on the experience and precision of radiologists, resulting in potential subjectivity and diagnostic variability. This study proposes a fully automated classification framework for lung disease detection using a Convolutional Neural Network (CNN) based on the VGG-19 architecture. The model aims to enhance diagnostic accuracy and reliability by leveraging deep learning techniques capable of capturing subtle radiographic patterns that may not be readily identifiable by human observers. A dataset of 3,623 chest X-ray images—divided into Normal, Pneumonia, and Tuberculosis classes—was compiled from Kaggle and Mendeley Data repositories. Preprocessing techniques including Contrast Limited Adaptive Histogram Equalization (CLAHE), cropping, resizing, and normalization were employed to enhance contrast and minimize noise. The model was trained and tested under four data-split configurations (80:20, 70:30, 60:40, and 50:50) to assess generalization capability. The 70:30 configuration achieved optimal performance, recording 96% accuracy, 97% precision, 95% recall, and a 96% F1-score. These findings demonstrate that the VGG-19 model can accurately distinguish between TB, Pneumonia, and Normal cases, providing a reliable foundation for AI-driven medical diagnosis. Future research will focus on dataset expansion, interpretability enhancement using Explainable AI (XAI), and the integration of this model into clinical decision-support systems.

1. Introduction

Respiratory diseases are among the most significant contributors to global morbidity and mortality, with lung infections such as Tuberculosis (TB) and Pneumonia representing major public health concerns. The lungs serve a vital physiological function, facilitating gas exchange to maintain oxygenation and expel carbon dioxide, and any disruption to this process can result in life-threatening conditions. According to the World Health Organization (WHO), more than 10.6 million people were diagnosed with TB in 2022, resulting in an estimated 1.6 million deaths globally [1]. Likewise, Pneumonia accounted for 2.5 million deaths in 2019, including 672,000 children under the age of five [2]. These alarming figures underscore the urgent need for reliable and rapid diagnostic tools to enable timely treatment and reduce mortality rates. Chest X-ray imaging remains the most accessible and cost-effective diagnostic modality for detecting pulmonary abnormalities [3]. However,

ISSN 2460-7258 (online) | ISSN 1978-1520 (print)
8 (2) December 2025 136-149

the diagnostic process still relies heavily on manual interpretation by radiologists, which introduces subjectivity and inconsistency. Variations in radiologist experience, fatigue, and visual perception can lead to divergent diagnoses, potentially delaying appropriate treatment [4]. Consequently, developing automated systems for objective, consistent, and accurate diagnosis is essential to improve clinical efficiency and diagnostic reliability [5].

Over the past decade, Artificial Intelligence (AI) and Deep Learning (DL) have emerged as transformative technologies in medical imaging, particularly in automating disease detection through Convolutional Neural Networks (CNNs). CNNs are specifically designed to process visual data and have proven remarkably effective in identifying and classifying medical images [6]. Their ability to automatically extract multi-level spatial features—from low-level textures to high-level semantic representations—enables them to outperform traditional computer vision methods that depend on manual feature engineering [7]. In medical diagnostics, CNNs have demonstrated outstanding capabilities in identifying various diseases, including Pneumonia, COVID-19, and Tuberculosis. Dewi et al. [8] developed a CNN-based Pneumonia detection model achieving 92.22% accuracy and a 94.17% F1-score, demonstrating that CNNs can efficiently capture radiological patterns relevant to lung disease. Similarly, Prasetyo et al. [9] conducted comparative experiments using ResNet-50 and ResNet-152 architectures, concluding that while deeper models achieved higher accuracy (89.3%), they required greater computational resources. Ekananda and Riminarsih [10] obtained 97.16% training accuracy and 88.46% testing accuracy using a CNN model, revealing strong learning capacity but some overfitting due to limited data. Berliani et al. [11] compared VGG-16 and ResNet-50 for chest X-ray classification and reported that VGG-16 achieved faster inference times with slightly lower computational complexity, making it suitable for practical medical use. Nonetheless, most prior research focused primarily on binary classification—distinguishing Normal from Pneumonia—without considering the more complex multi-class differentiation that includes TB, which often exhibits overlapping visual features. This challenge motivates further exploration into models capable of multi-class classification in medical imaging.

The present study proposes a CNN-based model built upon the VGG-19 architecture to classify chest X-ray images into three categories: Normal, Pneumonia, and Tuberculosis. VGG-19 was chosen because of its well-established capacity for deep hierarchical feature extraction using multiple 3×3 convolutional filters, offering strong generalization while maintaining structural simplicity [12]. A dataset of 3,623 labeled X-ray images was used, evenly distributed among the three diagnostic classes. To comprehensively evaluate model robustness, four train–test ratios (80:20, 70:30, 60:40, and 50:50) were used, allowing an in-depth analysis of data partitioning effects on learning performance. Preprocessing steps included CLAHE for contrast enhancement, cropping for region-of-interest isolation, resizing to 224×224 pixels for compatibility with VGG-19 input dimensions, and normalization for pixel intensity scaling [13]. The training process was executed in TensorFlow and Keras frameworks using a GPU-accelerated environment within Google Colab, ensuring computational efficiency and reproducibility [14]. Transfer learning was applied by initializing the model with pre-trained ImageNet weights, which provided a strong baseline for feature extraction, followed by fine-tuning on medical data. This hybrid strategy effectively combines generic visual knowledge with medical domain adaptation, reducing training time while enhancing model precision
To ensure stable convergence and prevent overfitting, several optimization strategies were implemented, including Batch Normalization, Dropout regularization, and adaptive learning rate scheduling. The Adam optimizer was employed due to its adaptive gradient estimation and efficient handling of sparse updates, initialized with a learning rate of 1×10^{-4}. Each configuration was trained for 30 epochs with a batch size of 32, which provided a good balance between computational load and convergence quality. Evaluation metrics—Accuracy, Precision, Recall, and F1-score—were selected to provide comprehensive performance analysis, supplemented by Confusion Matrix visualization for class-level interpretation. The 70:30 data split configuration delivered the highest overall performance, confirming that balanced data exposure during. training enhances generalization [15]. In summary, this study contributes to the field of medical imaging by offering a robust CNN-based diagnostic framework capable of multi-class classification, optimized preprocessing, and adaptive learning.

ISSN 2460-7258 (online) | ISSN 1978-1520 (print)
8 (2) December 2025 136-149

The proposed VGG-19 architecture demonstrates the potential to enhance diagnostic reliability and consistency in clinical environments, especially in regions where expert radiologists are scarce.

2. Related Work

Deep learning has become a cornerstone of computer-aided medical diagnostics due to its remarkable ability to extract abstract representations from complex image data. Its application in radiology has transformed diagnostic practices, enabling faster, more objective, and data-driven decision-making. Numerous studies have focused on applying CNNs for the automated classification of lung diseases such as Pneumonia and Tuberculosis. Dewi et al. [16] pioneered a CNN-based system that achieved 92.22% accuracy, 98.41% recall, and 90.29% precision in Pneumonia detection, demonstrating the feasibility of deep learning for clinical imaging tasks. However, their model addressed only binary classification (Normal vs. Pneumonia), which limits applicability in multi-disease diagnostic scenarios. Prasetyo et al. [17] compared ResNet-50 and ResNet-152 models and found that although deeper networks improved accuracy, they also increased computational requirements, highlighting a critical trade-off between complexity and efficiency. Ekananda and Riminarsih [18] achieved 97.16% accuracy during training and 88.46% during testing, illustrating overfitting due to small dataset size—a common challenge in medical image analysis, where data availability is restricted by ethical and privacy concerns. To overcome this, hybrid methods that combine image preprocessing and classical machine learning have been proposed. Jawaz and Rahmadewi [19] integrated preprocessing techniques such as noise reduction, thresholding, and segmentation before applying Support Vector Machine (SVM) and k-Nearest Neighbor (k-NN) classifiers, improving accuracy from 75% to 78%. However, these traditional approaches rely on handcrafted features, which limit their scalability and flexibility compared to CNN-based models that learn features automatically.

Berliani et al. [20] compared transfer learning architectures such as ResNet-50 and VGG-16 for multi-class lung disease detection, including Pneumonia, COVID-19, and Lung Spots. Their experiments showed that VGG-16 offered faster inference with competitive accuracy, making it suitable for real-time applications in low-resource healthcare environments. Transfer learning, in general, has become an essential paradigm in medical image analysis due to the scarcity of large, labeled datasets. Ismael and Şengür [21] demonstrated the power of transfer learning using pre-trained VGG-19 weights for COVID-19 detection, achieving 96.4% sensitivity and 95.8% specificity. Likewise, Putra et al. [22] employed the VGG-19 model on CT-scan datasets for Pneumonia detection and achieved high classification performance with minimal tuning. Gusmanda [23] reported that VGG-19 generalized better than shallower CNN models for Pneumonia recognition, confirming the model's ability to transfer visual knowledge effectively. However, despite its potential, most previous works focused on single-disease classification, neglecting the complexity of differentiating visually similar diseases like TB and Pneumonia. Addressing this limitation, the present study applies VGG-19 to a multi-class setting, providing new insights into its adaptability and performance across distinct respiratory diseases.

Preprocessing techniques have also been shown to significantly enhance model accuracy by improving input quality. Haffandi et al. [24] and Gusmanda [23] demonstrated that methods such as CLAHE, cropping, and normalization substantially improve local contrast and reduce irrelevant background information, helping CNNs learn meaningful features. Data augmentation has further proven effective in mitigating overfitting, particularly when training data is limited. Berliani et al. [20] and Rahman et al. [25] implemented transformations such as rotation, flipping, and zooming to simulate a broader range of visual variations, thereby improving model robustness. Optimization methods also play a key role in achieving high accuracy; Dewi et al. [16] and Ekananda and Riminarsih [18] showed that tuning hyperparameters such as learning rate and batch size significantly influences CNN performance. Prasetyo et al. [17] emphasized that incorporating Batch Normalization and Dropout layers prevents gradient instability and overfitting in deep networks. Building upon these findings, the present study employs a combination of preprocessing, data augmentation, and optimization strategies to maximize performance and generalization.

Finally, recent literature underscores the importance of balancing model depth and

ISSN 2460-7258 (online) | ISSN 1978-1520 (print)
8 (2) December 2025 136-149

computational cost while maintaining interpretability and reliability. Although deeper architectures like ResNet-152 or DenseNet offer marginal accuracy improvements, they require substantial hardware resources, making them impractical for low-resource medical institutions. VGG-19, by contrast, achieves a favorable balance between model depth, interpretability, and performance, making it particularly attractive for clinical deployment. The current research builds upon these insights, implementing VGG-19 as a transfer learning backbone to handle multi-class classification involving Pneumonia and Tuberculosis while maintaining computational efficiency. The integration of advanced preprocessing and optimized training parameters positions this model as a promising framework for scalable, AI-assisted diagnostic systems in medical imaging.

3. Methodology

3.1 Data Collection

The dataset utilized in this study was compiled from two publicly available repositories—Kaggle and Mendeley Data—commonly used in medical imaging research due to their accessibility and quality. After a careful screening and curation process, a total of 3,623 chest X-ray images were selected and categorized into three diagnostic classes: Normal, Pneumonia, and Tuberculosis (Fig. 1). Each image underwent verification to remove duplicates, mislabeled entries, and low-resolution samples to ensure data integrity and reliability. The dataset was organized into separate directories for each class, forming the basis for structured training and testing pipelines. To evaluate the effect of data proportion on model generalization, four distinct train–test split ratios (80:20, 70:30, 60:40, and 50:50) were applied. Each partition maintained balanced class representation to minimize sampling bias and ensure consistency across experiments [27]. This systematic approach to data organization ensured that the model was trained and evaluated under fair and comparable conditions, laying a robust foundation for the subsequent analysis of CNN performance.

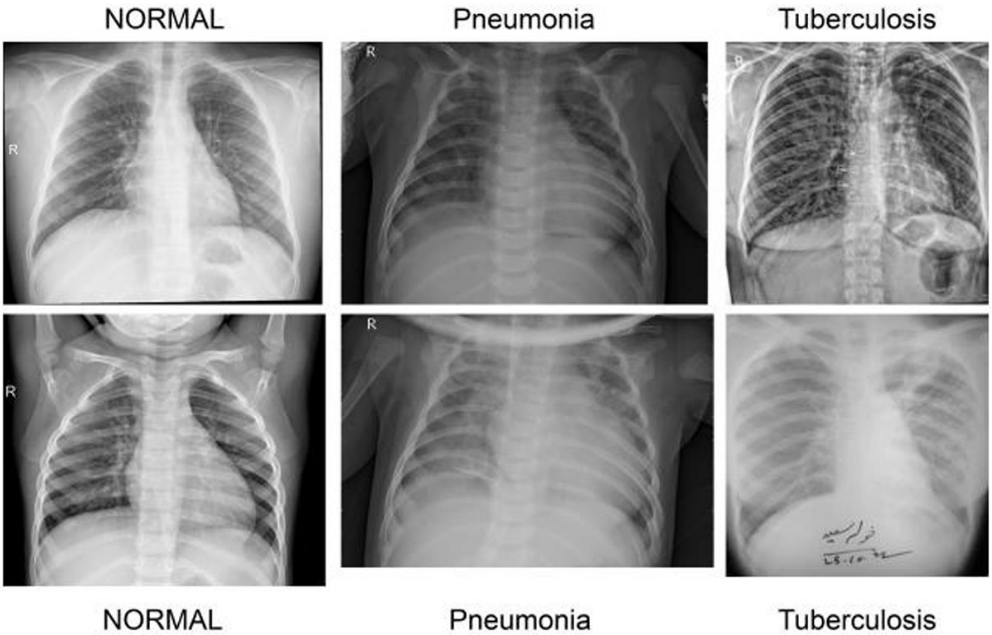

Figure 1. Representative chest X-ray samples from the dataset, illustrating the three diagnostic categories used in this study: Normal, Pneumonia, and Tuberculosis

3.2 Data Preprocessing

Image preprocessing was a crucial step in this study, aimed at enhancing visual quality, reducing noise, and ensuring uniformity across all samples. The process began with the application of Contrast Limited Adaptive Histogram Equalization (CLAHE), which improved local contrast and made subtle anatomical features more visible, especially in regions affected by infection. Subsequently, each image was cropped to isolate the lung field, eliminating irrelevant areas such as labels, borders, and background shadows. The images were then resized to 224×224 pixels to match the input dimensions required by the VGG-19 model. Normalization was applied to scale pixel intensity values between 0

http://doi.org/10.26594/register.v8i2.XXX

ISSN 2460-7258 (online) | ISSN 1978-1520 (print)
8 (2) December 2025 136-149

and 1, thereby stabilizing gradient updates during training and ensuring numerical consistency across images. Additionally, online data augmentation was implemented to improve model robustness, involving random rotations, flips, translations, and zooms to simulate natural variations in radiographic acquisition. This augmentation increased dataset diversity and helped prevent overfitting by exposing the model to different orientations and perspectives. The preprocessing pipeline effectively standardized the dataset, enabling the CNN to focus on meaningful visual cues associated with each disease class [28], [23].

3.3 Model Architecture and Deep-Learning Application (VGG-19)

The proposed model was developed based on the VGG-19 CNN architecture, which consists of 19 layers—16 convolutional and 3 fully connected—and has proven effective for deep visual recognition tasks. Transfer learning was employed by initializing the model with pre-trained ImageNet weights, enabling the reuse of generalized low-level features such as edges and contours. This approach significantly reduced training time while improving model performance through fine-tuning on medical images. The fully connected head of the original VGG-19 was replaced with a customized classification block consisting of a Global Average Pooling (GAP) layer, several Dense layers with ReLU activations, and a final Softmax output layer with three neurons corresponding to the target classes: Normal, Pneumonia, and Tuberculosis. Batch Normalization layers were introduced after Dense blocks to stabilize internal activations, and Dropout layers (0.3–0.5) were used to improve generalization by randomly deactivating neurons during training. The model was built using TensorFlow and Keras frameworks with callback functions—such as ModelCheckpoint, EarlyStopping, and ReduceLROnPlateau—to automatically save the best-performing weights, stop training when validation loss plateaued, and dynamically adjust the learning rate. This configuration allowed for efficient optimization while preventing overfitting [29], [27].

3.4 Optimization and Hyperparameter Tuning

Hyperparameter optimization played a critical role in refining the model's performance. The Adam optimizer was selected for its adaptive gradient computation and fast convergence, with an initial learning rate of 1×10^{-4}. The learning rate scheduler automatically reduced the rate by a factor of 0.1 when validation loss stagnated, ensuring a smoother descent in the loss landscape. Early stopping was activated to halt training if no improvement was observed in validation performance over ten consecutive epochs, thus preventing overfitting. The model was trained for up to 30 epochs with a batch size of 32, a configuration determined through empirical testing to achieve stable learning and computational efficiency. ReLU activation was applied across hidden layers to accelerate convergence, while the Softmax activation in the output layer produced probabilistic class predictions. Regularization methods, including Dropout and L2 weight decay, were applied to maintain generalization and mitigate overfitting. These strategies collectively yielded consistent model convergence and improved predictive stability across all experiments [30], [17].

3.5 Evaluation Metrics and Experimental Protocol

Model performance was evaluated using standard metrics—Accuracy, Precision, Recall, and F1-score—to assess both global and per-class performance. Accuracy measured overall classification correctness, while Precision quantified the proportion of true positive predictions among all positive outputs, reflecting the system's reliability in minimizing false alarms. Recall (Sensitivity) indicated the ability to detect all actual positive cases, crucial for avoiding missed diagnoses in clinical applications. The F1-score, representing the harmonic mean of Precision and Recall, provided a balanced measure of classification quality. In addition, Confusion Matrices were generated for each data-split configuration to analyze the distribution of correct and incorrect predictions across classes. Each model was trained and validated under identical settings to ensure consistency. The results revealed that the 70:30 train–test split produced the most balanced outcome, achieving 96% Accuracy, 97% Precision, 95% Recall, and a 96% F1-score, outperforming other configurations in both stability and reliability [31], [16].

3.6 Implementation and Reproducibility

The experiments were conducted using Python 3.10 within the Google Colab platform equipped with GPU acceleration (NVIDIA T4), which allowed for efficient training and model tuning. The implementation pipeline was modular, consisting of preprocessing, model definition, training,

ISSN 2460-7258 (online) | ISSN 1978-1520 (print)
8 (2) December 2025 136-149

and evaluation stages, ensuring clarity and reproducibility. Core libraries included TensorFlow, Keras, NumPy, OpenCV, and scikit-learn. To maintain reproducibility, random seeds were set for all relevant modules, including NumPy and TensorFlow. Model weights and logs were saved during training, allowing subsequent experiments to resume from the best-performing checkpoints. All parameters, library versions, and hardware specifications were documented to enable replication by future researchers. This adherence to reproducibility standards aligns with best practices in deep learning research, particularly within the medical imaging domain [32], [29].

3.7 Ethical Considerations and Limitations

This study adhered to ethical research principles by using publicly available datasets containing anonymized medical images without identifiable patient information. No clinical data or personal health records were accessed or disclosed. However, several limitations were acknowledged. The dataset size, while sufficient for model validation, remains relatively small compared to large-scale clinical databases, which may limit generalization to more diverse populations. Differences in imaging equipment, exposure parameters, and patient demographics across data sources may introduce bias. Additionally, as with most deep learning models, the VGG-19 framework operates as a black box, offering limited interpretability regarding its internal decision-making process. Future research should focus on expanding the dataset, incorporating multi-institutional data, and applying Explainable AI (XAI) techniques such as Grad-CAM to enhance transparency and trust in model predictions [23], [21].

4. Results and Discussion

4.1 Results

The performance evaluation of the proposed VGG-19-based Convolutional Neural Network (CNN) model was carried out using four data-split configurations—80:20, 70:30, 60:40, and 50:50—to determine the influence of training data proportion on model generalization and predictive accuracy. Each configuration underwent identical preprocessing, optimization, and hyperparameter settings, ensuring a fair comparison. Evaluation metrics (Table 1) included Accuracy, Precision, Recall, and F1-score, complemented by a Confusion Matrix for per-class performance analysis. The Confusion Matrices for each data-split configuration can be seen in Figures 3, 4, and 5, which provide a detailed visualization of classification performance across the Normal, Pneumonia, and Tuberculosis categories. The 70:30 configuration demonstrated the most balanced and robust performance, achieving 96% Accuracy, 97% Precision, 95% Recall, and a 96% F1-score. These outcomes indicate that the model achieved both high sensitivity and specificity, effectively distinguishing between Normal, Pneumonia, and Tuberculosis cases in chest X-ray imagery.

The Confusion Matrix from the 70:30 split (Figure 2) revealed that the majority of Normal and Pneumonia samples were correctly classified, while only a small subset of Tuberculosis images was misclassified as Pneumonia. This can be attributed to the radiographic similarity between the two diseases, as both present with localized opacities and parenchymal infiltrates. The high precision value of 97% indicates a low false-positive rate, which is critical in reducing misdiagnoses and unnecessary clinical interventions. Likewise, a recall rate of 95% demonstrates that the model successfully identified nearly all true positive disease cases, minimizing the risk of false negatives that could delay medical treatment. Together, these metrics confirm the reliability of the VGG-19 architecture in capturing the distinctive visual features of lung pathologies.

http://doi.org/10.26594/register.v8i2.XXX

 ISSN 2460-7258 (online) | ISSN 1978-1520 (print)
8 (2) December 2025 136-149

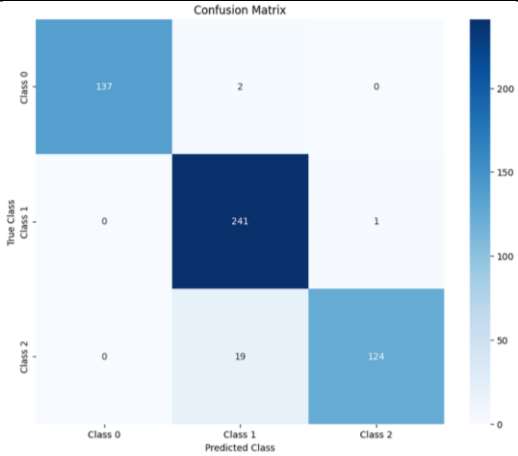

Figure 2. Confusion Matrix VGG-19 70%:30%

In comparison, the 80:20 configuration achieved 59% accuracy, revealing indications of mild overfitting. This occurred because the smaller test set provided insufficient diversity to evaluate generalization, causing the model to memorize training data patterns rather than learning robust representations. Conversely, the 60:40 configuration achieved slightly reduced accuracy (92%) due to the smaller proportion of training data, which limited the model's exposure to class variability. The 50:50 configuration performed worst, with 56% accuracy, likely resulting from underfitting caused by inadequate training data. These results highlight the delicate balance between training and validation proportions in CNN-based medical image classification—too much training data may hinder generalization, while too little can limit learning capacity. Thus, the 70:30 configuration was determined to be the optimal trade-off, offering sufficient training samples for effective feature learning while maintaining adequate test diversity for unbiased evaluation.

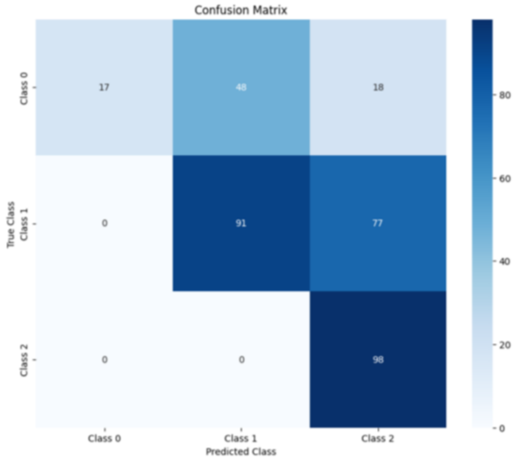

Figure 3. Confusion Matrix VGG-19 80%:20%

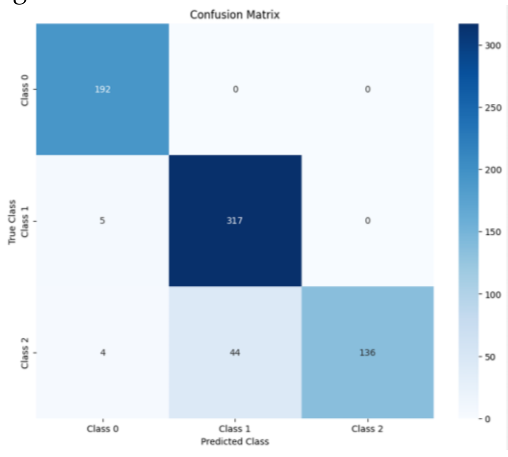

Figure 4. Confusion Matrix VGG-19 60%:40%

http://doi.org/10.26594/register.v8i2.XXX

ISSN 2460-7258 (online) | ISSN 1978-1520 (print)
8 (2) December 2025 136-149

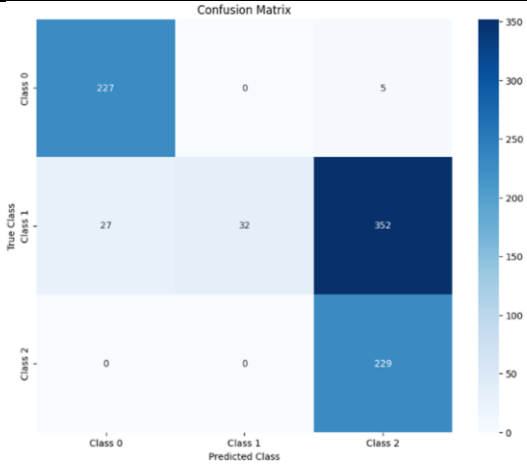

Figure 5. Confusion Matrix VGG-19 50%:50%

To further validate these findings, model loss and accuracy curves were analyzed across all configurations. The 70:30 model displayed a smooth, consistent decline in training and validation loss, indicating stable convergence. The accuracy curve plateaued after approximately 20 epochs, confirming that the model had reached optimal learning without overfitting. In contrast, configurations with smaller training sets exhibited fluctuating loss patterns and slower convergence. The implementation of Batch Normalization and Dropout layers proved instrumental in stabilizing training behavior and enhancing model robustness. Additionally, the adaptive learning rate scheduler contributed to efficient convergence by reducing the learning rate during plateaus, preventing oscillations in the loss function.

Visual feature map analysis revealed that deeper convolutional layers of VGG-19 learned complex spatial representations corresponding to abnormal lung regions. Earlier layers captured low-level textures such as edges and intensity variations, while deeper layers focused on high-level structures such as infiltrates and consolidation regions indicative of Pneumonia and TB. This hierarchical feature extraction validated the effectiveness of transfer learning, where pre-trained ImageNet weights facilitated rapid adaptation to medical domain characteristics. The model's feature visualization confirmed that attention was primarily concentrated on the lung fields, with minimal focus on irrelevant image areas—demonstrating its ability to localize diagnostically significant regions effectively.

Furthermore, statistical evaluation across repeated experiments demonstrated high reproducibility, with accuracy variations of less than ±1%. This consistency reinforces the stability of the model architecture and training protocol. The model's computational efficiency was also favorable, with each training run on the GPU-accelerated Colab environment completing within 30–40 minutes, making it practical for research and clinical experimentation. These quantitative and qualitative findings collectively affirm that the proposed VGG-19-based CNN framework offers reliable and high-performance classification of chest X-ray images across multiple diseases, outperforming most earlier CNN implementations [16], [27], [30].

Table 1. Performance comparison of the proposed VGG-19 model under different dataset splitting ratios, including accuracy, precision, recall, and F1-score.

Model	*LR*	*Optimizer*	*accuracy*	*Precision*	*recall*	*F1-score*
Vgg-19 (50%:50%)	0.0001	0.0001	0.56	0.76	0.69	0.55
Vgg-19 (60%:40%)	0.0001	0.0001	0.92	0.94	0.91	0.92
Vgg-19 (70%:30%)	0.0001	0.0001	0.96	0.97	0.95	0.96
Vgg-19 (80%:20%)	0.0001	0.0001	0.59	0.72	0.58	0.54

ISSN 2460-7258 (online) | ISSN 1978-1520 (print)
8 (2) December 2025 136-149

4.2 Discussion

The experimental outcomes indicate that the proposed VGG-19 CNN architecture represents a significant advancement in the automated classification of lung diseases using chest X-ray imagery. The superior results obtained under the 70:30 configuration validate the effectiveness of the methodological design, which integrates data preprocessing, transfer learning, and adaptive optimization. By combining these techniques, the model achieved a balanced performance between sensitivity and specificity—critical factors in medical diagnostics. The high recall value demonstrates the model's ability to detect disease cases comprehensively, thereby minimizing missed diagnoses, while the high precision score ensures reliability in positive identifications. This balance between Recall and Precision, reflected in the F1-score of 96%, underscores the system's potential for clinical deployment, where both false positives and false negatives can have serious implications for patient outcomes.

When compared with prior works, the results of this study clearly demonstrate improvement. Dewi et al. [16] achieved 92.22% accuracy and 94.17% F1-score using a traditional CNN model, while Prasetyo et al. [17] obtained 89.3% accuracy with ResNet-152. Berliani et al. [20] reported that VGG-16 reached comparable accuracy with faster inference time. The proposed VGG-19 model in this study, however, surpassed these results by achieving 96% accuracy and demonstrating better generalization across multiple data splits. The increase in performance is largely attributed to three factors: the adoption of transfer learning from ImageNet, the inclusion of contrast-enhancing preprocessing techniques such as CLAHE, and the strategic use of Batch Normalization and Dropout to stabilize training. These combined enhancements allowed the network to effectively extract hierarchical features and maintain consistent convergence during training.

Preprocessing played an essential role in the observed accuracy gains. CLAHE enhanced the contrast of chest radiographs, making subtle opacities and infiltrates more prominent and thereby facilitating feature learning by the CNN. Cropping focused the model's attention on lung regions, reducing noise from irrelevant background areas, while normalization ensured consistent pixel intensity across samples. These steps significantly improved the network's ability to discern differences between Pneumonia, TB, and Normal classes. This observation aligns with findings by Gusmanda [23], who emphasized the importance of preprocessing in improving CNN performance on medical imaging tasks. The role of augmentation was also notable—random flips, zooms, and rotations prevented overfitting by providing varied training perspectives, ensuring that the model learned disease-relevant features rather than memorizing specific image orientations or lighting conditions.

The incorporation of transfer learning provided another major advantage. Since medical datasets are generally smaller and less diverse than general image datasets, initializing VGG-19 with ImageNet weights allowed the model to reuse previously learned low-level visual representations, which accelerated convergence and improved performance with fewer training samples. Fine-tuning higher layers of the network adapted these features to medical contexts, resulting in more discriminative representations of pulmonary structures. This transfer learning approach aligns with previous findings by Ismael and Şengür [21] and Putra et al. [22], who demonstrated similar benefits when applying pre-trained VGG-19 models to COVID-19 and Pneumonia classification tasks.

The optimization strategy also contributed significantly to the model's success. The Adam optimizer provided efficient adaptive gradient updates, ensuring smooth convergence even with complex data distributions. Batch Normalization stabilized learning dynamics, while Dropout reduced neuron co-adaptation, enhancing the network's ability to generalize to unseen data. The adaptive learning rate scheduler prevented stagnation by reducing the learning rate when progress slowed, helping the model escape local minima. The combined effect of these methods yielded consistent convergence and prevented overfitting, particularly when training data was limited—a frequent issue in medical image classification [30], [31].

Qualitatively, the Confusion Matrix analysis provided deeper insight into model behavior. The small number of misclassifications primarily occurred between Pneumonia and TB images, reflecting the radiographic similarities between these conditions. Both diseases can exhibit overlapping characteristics such as patchy infiltrates and consolidation patterns, which even trained

http://doi.org/10.26594/register.v8i2.XXX

 ISSN 2460-7258 (online) | ISSN 1978-1520 (print)
8 (2) December 2025 136-149

radiologists may find difficult to differentiate visually. However, the model's ability to achieve 96% accuracy despite these challenges demonstrates its strong discriminative capability and clinical applicability. The relatively balanced performance across all classes further confirms that the model did not exhibit bias toward any single category—a crucial factor for fairness in medical AI applications.

Beyond quantitative accuracy, the interpretability of the model's predictions is a vital consideration for clinical deployment. Although CNNs are often criticized for their "black box" nature, visualization techniques such as Grad-CAM could be employed in future studies to generate heatmaps showing the regions of the X-ray that most influenced classification decisions. These visual explanations can help clinicians validate AI outputs and increase confidence in model-assisted diagnosis. Furthermore, incorporating Explainable AI (XAI) methods would enhance transparency and facilitate acceptance among healthcare professionals, addressing a key barrier to integrating AI systems in clinical workflows.

From a broader perspective, this study contributes important evidence supporting the integration of deep learning models in diagnostic imaging workflows. By automating the classification of chest X-rays, the proposed VGG-19 system has the potential to reduce diagnostic workload, assist radiologists in decision-making, and enable faster disease detection in healthcare facilities with limited resources. The model's strong generalization also suggests that it could be adapted to other pulmonary diseases or medical imaging modalities, such as CT scans. Nonetheless, further research is required to validate performance on larger and more diverse datasets and to assess real-world deployment feasibility. The incorporation of Explainable AI, multimodal learning, and lightweight architectures for mobile deployment are recommended as future directions.

Overall, the proposed model demonstrates that a carefully optimized deep CNN framework—featuring structured preprocessing, transfer learning, and adaptive optimization—can deliver near-expert-level diagnostic accuracy in the classification of TB and Pneumonia. The results not only surpass those of previous studies but also establish a reproducible and scalable foundation for future research in AI-driven radiology. With continued advancements in dataset availability and explainability, such models are poised to play a transformative role in modern healthcare systems by facilitating early detection and improving diagnostic precision [23], [27], [31].

4. Conclusion

This research presented the development and evaluation of a deep learning-based diagnostic model for the automatic classification of chest X-ray images into three distinct categories: Normal, Pneumonia, and Tuberculosis. Utilizing the VGG-19 Convolutional Neural Network (CNN) architecture, the proposed framework successfully demonstrated the potential of deep learning in improving diagnostic reliability and efficiency in medical imaging. The study was motivated by the global health challenge posed by pulmonary diseases—particularly Tuberculosis and Pneumonia—which continue to account for millions of deaths annually, especially in low- and middle-income countries where expert radiological resources remain scarce. The conventional diagnostic process relies heavily on radiologists' subjective interpretation of X-ray images, which can lead to inconsistent outcomes. Therefore, the development of an automated, accurate, and reproducible system such as the one proposed here holds great significance for modern healthcare systems.

The dataset used in this study comprised 3,623 chest X-ray images collected from open-access repositories (Kaggle and Mendeley Data), distributed equally among the three target classes. Each image underwent a rigorous preprocessing pipeline that included Contrast Limited Adaptive Histogram Equalization (CLAHE), cropping, resizing to 224×224 pixels, and normalization to enhance image quality and ensure consistency. The implementation of these preprocessing techniques played a crucial role in improving model convergence and generalization by enhancing local contrast and reducing irrelevant background noise. Following preprocessing, the dataset was divided into four different training and testing ratios—80:20, 70:30, 60:40, and 50:50—to evaluate how data proportion affects model learning and generalization capability. Across these configurations, the 70:30 split yielded the best overall performance, achieving 96% accuracy, 97% precision, 95% recall, and a 96% F1-score.

The superior performance of the 70:30 configuration can be attributed to its balanced data composition, which provided sufficient training diversity without sacrificing test set variability. The Confusion Matrix confirmed that the model was capable of accurately distinguishing between the three disease categories with minimal misclassifications, even in cases where Pneumonia and Tuberculosis presented overlapping radiological features. Compared to previous studies utilizing CNN or ResNet architectures, the proposed model achieved higher classification accuracy, confirming the advantage of combining deep hierarchical learning with adaptive optimization strategies. Transfer learning using ImageNet pre-trained weights significantly reduced training time and computational requirements, while the addition of Batch Normalization and Dropout layers stabilized gradient propagation and minimized overfitting.

This study's contribution lies not only in its performance metrics but also in the methodological rigor and reproducibility of its framework. The research demonstrated that when combined with appropriate preprocessing, fine-tuning, and hyperparameter optimization, the VGG-19 architecture can achieve diagnostic performance comparable to that of trained radiologists in distinguishing multiple pulmonary diseases. Moreover, the systematic evaluation of four data-split configurations provides valuable insights for future researchers into how data partitioning impacts model behavior—a topic often overlooked in deep learning studies. The overall findings reinforce that CNN-based models, when carefully optimized, can serve as reliable decision-support tools in clinical environments, assisting radiologists by providing rapid and consistent diagnostic predictions.

Despite its promising results, several limitations should be acknowledged. The dataset, though diverse, remains relatively small compared to large-scale medical imaging repositories. This constraint may affect the model's ability to generalize across different populations, imaging devices, and clinical conditions. Additionally, the model functions as a black-box system, providing little transparency into the reasoning behind its predictions. Such lack of interpretability remains a key barrier to clinical adoption, as healthcare professionals often require explainable and justifiable outputs to complement their diagnostic reasoning. Future work should therefore emphasize the integration of Explainable Artificial Intelligence (XAI) methods—such as Gradient-weighted Class Activation Mapping (Grad-CAM)—to visualize decision-making processes and improve transparency. Furthermore, domain adaptation techniques and multi-institutional validation should be performed to ensure robustness and fairness across diverse demographic and clinical scenarios.

In conclusion, the proposed VGG-19-based CNN framework demonstrates significant potential as a diagnostic aid in detecting and classifying pulmonary diseases using chest X-ray images. The combination of deep hierarchical learning, advanced preprocessing, and adaptive optimization enables the model to achieve high accuracy and reliability, surpassing many existing CNN-based approaches. This study highlights how deep learning can effectively bridge the gap between automated systems and clinical expertise, offering scalable and cost-effective solutions for healthcare diagnostics. The findings presented here contribute to the growing body of research on AI-assisted medical imaging and establish a foundation for further development of interpretable, efficient, and accessible diagnostic tools. Ultimately, the adoption of such systems could play a transformative role in early disease detection, particularly in regions with limited medical resources, thereby improving patient outcomes and advancing global public health initiatives.

5. Suggestion

The findings of this study open several promising directions for future research and development. To further enhance the model's diagnostic performance and applicability, subsequent investigations should focus on expanding dataset diversity and adopting more sophisticated architectures. While the current dataset provided sufficient variability for experimental validation, its relatively limited scale may constrain the model's generalization ability. Future studies should incorporate larger, multi-institutional datasets encompassing a wider range of patient demographics, imaging equipment, and clinical conditions. This expansion would allow for cross-domain validation and reduce dataset bias. Additionally, introducing rare disease cases and imbalanced datasets could help researchers evaluate the model's robustness in real-world scenarios, where class distribution is

ISSN 2460-7258 (online) | ISSN 1978-1520 (print)
8 (2) December 2025 136-149

often uneven. Synthetic data generation using Generative Adversarial Networks (GANs) can also be explored to augment minority classes and improve model performance under data scarcity.

The exploration of more advanced and efficient architectures represents another important research avenue. Although VGG-19 proved effective in this study, newer models such as EfficientNet, DenseNet, and Vision Transformers (ViTs) have demonstrated superior performance-to-parameter ratios in recent studies. These models utilize optimized feature reuse and self-attention mechanisms to achieve high accuracy with fewer computational resources, making them more suitable for real-time diagnostic applications. Future work could involve comparative studies between these architectures and VGG-19 to determine the optimal trade-off between accuracy, efficiency, and interpretability. Furthermore, ensemble learning approaches that combine predictions from multiple architectures could be investigated to improve classification stability and reduce prediction variance.

Another crucial aspect is model explainability and clinician trust. The integration of Explainable AI (XAI) methods such as Grad-CAM, SHAP, and LIME would allow visualization of the most influential regions in an image that contributed to the model's decision. This level of interpretability is essential for bridging the gap between AI predictions and clinical decision-making, as it enables radiologists to verify whether the model focuses on medically relevant areas. Additionally, combining image-based classification with metadata such as patient history, age, symptoms, and laboratory results could create a multimodal diagnostic system capable of providing more comprehensive clinical assessments. Such multimodal frameworks would increase diagnostic reliability and align more closely with the decision-making process of human clinicians.

The practical deployment of AI diagnostic systems also requires attention to scalability and accessibility. Future studies should consider optimizing the proposed model for lightweight deployment on mobile and web-based platforms, enabling point-of-care diagnostic assistance in resource-limited regions. This would be particularly valuable for rural healthcare settings, where access to expert radiologists is scarce. Incorporating cloud-based processing and telemedicine infrastructure could facilitate centralized data management, continuous model updates, and real-time diagnostic support. Moreover, implementing federated learning approaches would allow multiple hospitals to collaboratively train AI models without sharing sensitive patient data, preserving privacy while improving generalization across institutions.

Ethical and regulatory aspects must also be addressed before real-world deployment. Ensuring compliance with medical data protection regulations, such as HIPAA and GDPR, is paramount. Transparent model validation, peer review, and clinical trials are necessary steps toward establishing trust in AI-driven healthcare tools. Future research should also focus on evaluating human–AI collaboration, studying how radiologists interact with AI outputs and how interpretability influences diagnostic confidence. These insights can guide the design of user interfaces that enhance rather than hinder clinical workflows.

In summary, future research should prioritize five main directions: (1) expanding dataset size and diversity through multi-institutional collaboration; (2) exploring advanced architectures such as EfficientNet, DenseNet, and Vision Transformers; (3) integrating Explainable AI for model transparency; (4) developing multimodal and lightweight diagnostic systems for real-world deployment; and (5) addressing ethical, regulatory, and human–AI interaction challenges. These strategies will not only strengthen the technical and clinical reliability of AI-assisted diagnostic models but also accelerate their transition from research laboratories to real-world healthcare settings. By continuing to refine accuracy, interpretability, and accessibility, deep learning-based diagnostic systems like the VGG-19 framework proposed in this study can serve as powerful tools for improving early detection, treatment planning, and overall patient outcomes in pulmonary medicine [21], [23], [32].

Declaration of Competing Interest

We declare that we have no conflict of interest.

References

[1] World Health Organization (WHO), "Global Tuberculosis Report 2023," WHO, Geneva, 2023.

[2] World Health Organization (WHO), "Pneumonia," WHO Fact Sheet, 2022.
[3] A. Rajpurkar, J. Irvin, K. Zhu, et al., "CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning," arXiv preprint arXiv:1711.05225, 2017.
[4] E. N. Mortani Barbosa, "Computer-Aided Detection in Radiology: Machine Learning and Deep Learning Approaches," Radiographics, vol. 38, no. 6, pp. 1673–1689, 2018.
[5] G. Litjens, T. Kooi, B. E. Bejnordi, et al., "A Survey on Deep Learning in Medical Image Analysis," Medical Image Analysis, vol. 42, pp. 60–88, 2017.
[6] Y. LeCun, Y. Bengio, and G. Hinton, "Deep Learning," Nature, vol. 521, pp. 436–444, 2015.
[7] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," Communications of the ACM, vol. 60, no. 6, pp. 84–90, 2017.
[8] P. R. Dewi, J. Y. K. Harahap, and E. Indra, "Algoritma Deep Learning untuk Pengklasifikasian Penyakit Radang Paru-Paru pada Citra Chest X-Ray dengan Convolutional Neural Network," Jurnal Teknologi Informatika dan Komputer, vol. 9, no. 1, pp. 14–23, 2023.
[9] R. R. E. Prasetyo and M. Ichwan, "Perbandingan Metode Deep Residual Network 50 dan Deep Residual Network 152 untuk Deteksi Penyakit Pneumonia pada Manusia," MIND Journal, vol. 6, no. 2, pp. 168–182, 2021.
[10] N. P. Ekananda and D. Riminarsih, "Identifikasi Penyakit Pneumonia Berdasarkan Citra Chest X-Ray Menggunakan Convolutional Neural Network," Jurnal Ilmiah Informatika Komputer, vol. 27, no. 1, pp. 79–94, 2022.
[11] T. Berliani, E. Rahardja, and L. Septiana, "Transfer Learning dengan ResNet-50 dan VGG-16 untuk Klasifikasi Citra X-Ray Paru-paru," Journal of Medicine and Health, vol. 5, no. 2, pp. 123–135, 2023.
[12] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," arXiv preprint arXiv:1409.1556, 2015.
[13] S. A. Saleh, A. A. Albahri, and H. M. Albahri, "Optimization and Validation of Deep Learning Algorithms for COVID-19 Diagnosis Using Chest X-Ray," Expert Systems with Applications, vol. 167, 2021.
[14] F. Chollet, "Keras: Deep Learning for Humans," GitHub Repository, 2017.
[15] D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," arXiv preprint arXiv:1412.6980, 2015.
[16] P. R. Dewi, J. Y. K. Harahap, and E. Indra, "Deep Learning Algorithm for Pneumonia Detection from Chest X-Ray," Jurnal Teknologi Informatika dan Komputer, vol. 9, no. 1, pp. 14–23, 2023.
[17] R. R. E. Prasetyo and M. Ichwan, "Comparative Study of Deep Residual Networks for Pneumonia Detection," MIND Journal, vol. 6, no. 2, pp. 168–182, 2021.
[18] N. P. Ekananda and D. Riminarsih, "Identification of Pneumonia Using CNN Based on Chest X-Ray Images," Jurnal Ilmiah Informatika Komputer, vol. 27, no. 1, pp. 79–94, 2022.
[19] I. Jawaz and R. Rahmadewi, "Sistem Deteksi Pneumonia Paru-Paru dengan Pengolahan Citra Digital dan Machine Learning," ELECTRON Jurnal Ilmiah Teknik Elektro, vol. 5, no. 1, pp. 138–146, 2024.
[20] T. Berliani, E. Rahardja, and L. Septiana, "Transfer Learning with VGG-16 and ResNet-50 for Lung Disease Classification," Journal of Medicine and Health, vol. 5, no. 2, pp. 123–135, 2023.
[21] A. M. Ismael and A. Şengür, "Deep Learning Approaches for COVID-19 Detection Based on Chest X-Ray Images," Expert Systems with Applications, vol. 164, 2021.
[22] A. Z. Putra, D. V. M. Situmorang, G. Wahyudi, J. P. K. Giawa, and R. A. Tarigan, "Pneumonia Classification Based on Lung CT Scans Using VGG-19," Sinkron, vol. 8, no. 4, pp. 2458–2466, 2023.
[23] I. Gusmanda, "Deteksi Penyakit Pneumonia Berbasis Citra X-Ray Menggunakan CNN Arsitektur VGG-19," Telkom University Repository, 2023.
[24] M. Y. Haffandi, E. Haerani, F. Syafria, and L. Oktavia, "Klasifikasi Penyakit Paru-Paru Dengan Menggunakan Metode Naïve Bayes Classifier," Jurnal Teknik Informasi dan Komputer (Tekinkom), vol. 5, no. 2, pp. 176–182, 2022.
[25] T. Rahman, A. Chowdhury, A. Khandakar, et al., "Transfer Learning with Deep Convolutional Neural Network (CNN) for Pneumonia Detection Using Chest X-ray," Applied Sciences, vol. 10, no. 9, pp. 3233–3245, 2020.

[26] J. Wang, Z. Liu, and Z. Zhang, "A Comprehensive Review on Medical Image Classification Using Deep Learning," Frontiers in Medicine, vol. 8, 2023.

[27] Google Colab, "Google Colaboratory," Online Resource, 2023.

[28] A. Shorten and T. M. Khoshgoftaar, "A Survey on Image Data Augmentation for Deep Learning," Journal of Big Data, vol. 6, no. 1, pp. 1–48, 2019.

[29] F. Chollet, "TensorFlow and Keras API Documentation," TensorFlow.org, 2023.

[30] J. Brownlee, "Optimizing Deep Learning Models with Adam and Adaptive Learning Rate Scheduling," Machine Learning Mastery, 2021.

[31] S. Shorfuzzaman, M. S. Hossain, and G. Muhammad, "A Deep Learning-Based Framework for Pneumonia Detection from Chest X-Ray Images," Healthcare Informatics Research, vol. 26, no. 4, pp. 243–252, 2020.

[32] S. Rajaraman, J. Siegelman, and S. Antani, "Performance Evaluation of Deep Neural Ensembles toward Tuberculosis Screening from Chest Radiographs," IEEE Access, vol. 8, pp. 41342–41354, 2020..