Building of Informatics, Technology and Science (BITS)
Volume 7, No 2, September 2025 Page: 1069−1077
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i2.7840

Comparative Analysis of Random Forest and Convolutional Neural Network (CNN) Algorithms for Pneumonia Detection in Chest X-ray Images: Accuracy, Interpretability, and Computational Efficiency

Siffa Zaena, Kemas Muslim Lhaksmana[*]

Fakultas Informatika, Teknik Informatika, Telkom University, Bandung, Indonesia
Email: [1]siffazaena@student.telkomuniversity.ac.id, [2,*]kemasmuslim@telkomuniversity.ac.id
Email Penulis Korespondensi: kemasmuslim@telkomuniversity.ac.id

Submitted: **01/07/2025**; Accepted: **01/09/2025**; Published: **02/09/2025**

Abstract−Pneumonia is a lung infection that can be detected through chest X-ray images. Manual diagnosis requires radiological expertise and time, thus an accurate automated method is needed. This study aims to compare the performance of two image classification algorithms, Convolutional Neural Network (CNN) and Random Forest (RF), in detecting pneumonia. The dataset used was obtained from Kaggle, consisting of 5,863 X-ray images categorized into three classes: bacterial pneumonia, viral pneumonia, and normal. Preprocessing steps include image resizing, normalization, and data augmentation. The CNN model was built using multiple convolutional and pooling layers, while RF utilized numerical features derived from histograms and texture. The CNN model demonstrated superior performance, achieving 92.4% accuracy, 93.1% precision, 91.6% recall, and 92.3% F1-score, compared to 82.7%, 80.3%, 85.1%, and 82.6% for Random Forest, respectively. Although CNN offers better accuracy, RF excels in interpretability. In conclusion, CNN is more effective for image-based pneumonia classification, yet RF remains relevant in applications requiring transparent decision-making. Potential biases, such as class imbalance and limited demographic representation in the dataset, could influence model performance and generalizability across different patient populations.

Keywords: Convolutional Neural Network; Pneumonia; Image Classification; Random Forest; X-ray

1. INTRODUCTION

Pneumonia is one of the leading infectious respiratory diseases that contributes significantly to morbidity and mortality, particularly among children and the elderly. According to the World Health Organization (WHO), pneumonia accounts for approximately 14% of all deaths among children under five globally, with over 700,000 fatalities reported annually [1]. In Indonesia, pneumonia remains a major public health issue, especially in regions with limited access to healthcare facilities. Data from the 2018 National Basic Health Research (Riskesdas) highlighted that the prevalence of pneumonia among children under five reached 4.2%, a relatively high figure compared to other ASEAN countries [2].

Effective pneumonia management hinges on rapid and accurate diagnosis, which is often challenging due to the shortage of medical personnel and advanced diagnostic tools, particularly at the primary healthcare level. Chest X-rays are frequently used to assist in diagnosing pneumonia by detecting signs of lung infection. However, interpreting these X-rays requires significant expertise, and results can vary among radiologists, depending on their experience and subjective judgment. Furthermore, in remote areas, the lack of radiology experts exacerbates this problem, leading to potential misdiagnoses.

Diagnosing an issue quickly and accurately is the key to avoiding any adverse results. A chest X-ray is one of the most basic diagnostic tools. Through it, we can see significantly infiltrated lungs and opacities that are a mark of lung infections. However, these interpretations are sometimes quite difficult because the experts are either not present or are very far away, so that even the most knowledgeable radiologists can be mistaken. Therefore, the possibility of the development of AI-based methods specifically image classification algorithms that can be used to help in an automatic and genuinely accurate process of pneumonia detection. Both Convolutional Neural Networks (CNN) and Random Forest (RF) algorithms have shown promise in creating pneumonia detection systems from medical images.

This challenge calls for the development of automated systems using artificial intelligence (AI), specifically image classification algorithms, to assist in the accurate and efficient diagnosis of pneumonia. Both Convolutional Neural Networks (CNN) and Random Forest (RF) have shown promise in detecting pneumonia from medical images. The visual differences in chest X-rays, such as infiltrates, consolidations, and healthy lung textures, provide essential features for developing classification models using these algorithms. While CNNs excel at extracting complex patterns from medical images, RF is a robust algorithm known for its efficiency in handling high-dimensional, imbalanced data and providing transparent decision-making..

Artificial Intelligence driven diagnostic tools offer fast, accurate, and consistent pneumonia detection. Recent research highlights the effectiveness of machine learning (ML) and deep learning (DL) techniques in medical image classification [3]. For instance, [4] demonstrated how deep neural networks can classify skin cancer with dermatologist-level accuracy, highlighting the potential of AI in healthcare applications. Among ML methods, Random Forest (RF) is widely used; it builds multiple decision trees via bagging to improve prediction accuracy and reduce overfitting risks [5]. RF is notably robust when handling high-dimensional, imbalanced data and competes well against other algorithms like Support Vector Machines (SVM) and K-Nearest Neighbors (KNN).

Meanwhile, CNNs specialized DL models for image analysis use convolutional layers to extract key features through filtering operations. CNNs have achieved outstanding results in pulmonary disease classification, including pneumonia detection. For example, Rajpurkar et al. developed CheXNet, a CNN trained on over 100,000 chest X-rays, outperforming average radiologist accuracy for pneumonia [3]. Likewise, Maysanjaya reported CNN reaching 92% accuracy, surpassing conventional diagnostic methods [6].

Nonetheless, CNNs have drawbacks such as requiring large datasets, lengthy training times, and high computational resources (often needing GPUs). Additionally, CNNs are often regarded as "black-box" models, complicating clinical interpretation and auditing. Conversely, RF models train faster and offer better interpretability but struggle to capture spatial features inherent in images [7]. Therefore, comparing RF and CNN for pneumonia classification is essential to understand each method’s advantages and limitations. While several studies have investigated RF and CNN separately for pneumonia detection on chest X-rays, direct comparisons of these approaches within the same framework are limited. Furthermore, evaluations often lack comprehensive metrics such as accuracy, precision, recall, and computational efficiency. This study aims to fill these gaps by systematically comparing RF and CNN in classifying pneumonia from medical images.

2. RESEARCH METHODOLOGY

2.1 Research Stages

Figure 1. Dataset sample

Figure 1 illustrates representative chest X-ray images used in this study, showing three different conditions: normal lungs (left), bacterial pneumonia (center), and viral pneumonia (right). The normal chest X-ray shows clear lung fields with consistent radiolucency, no abnormal opacities, and clearly visible anatomical structures such as the diaphragm and heart borders. In contrast, the bacterial pneumonia image demonstrates focal lobar consolidation, which appears as dense white areas due to the accumulation of pus or fluid in the alveoli, typically affecting one specific lobe of the lung. This pattern is commonly associated with bacterial infections. On the other hand, the viral pneumonia image shows diffuse interstitial infiltrates, which manifest as widespread haziness or reticular patterns across both lungs. These changes reflect inflammation of the lung interstitium and are characteristic of viral infections, which tend to be more scattered and bilateral than bacterial infections.

This study adopts a comparative quantitative research design with an experimental approach to evaluate the effectiveness of two image classification algorithms Random Forest (RF) and Convolutional Neural Network (CNN) in detecting pneumonia from chest X-ray images. The experimental quantitative method is chosen to facilitate an objective assessment using measurable metrics such as accuracy, precision, recall, and F1-score. These metrics allow for a detailed comparison of the models’ capabilities in classifying the X-ray images correctly. Publicly available datasets, such as those hosted on Kaggle, are utilized in this study. These datasets are commonly used in medical imaging research due to their large sample sizes, diverse patient data, and consistent annotations, which make them suitable for training and testing machine learning models [6]. Furthermore, the use of open-access datasets supports reproducibility and validation by other researchers, which is essential to ensure that the models developed are reliable and generalizable for clinical application [8].

The overall process of image classification in this research involves several sequential stages, beginning with data collection and extensive preprocessing. First, the dataset is cleaned by identifying and removing any mislabeled entries to ensure data integrity. The images are then resized to a consistent dimension of 150x150 pixels to standardize the input and facilitate the model training process. Any unnecessary borders surrounding the lungs in the X-ray images are cropped to focus solely on the relevant areas, thus improving the model's performance by reducing irrelevant background information. The pixel values are normalized by scaling them to a range of 0 to 1, which helps accelerate convergence during training and ensures uniformity across the dataset.

Additionally, data augmentation techniques are applied to artificially expand the dataset and improve the model's ability to generalize. These techniques include random horizontal flipping, rotation, zooming, and slight adjustments to brightness, which help the model become more robust to variations in image orientation and lighting conditions. After preprocessing, the dataset is divided into two sets: 80% for training and 20% for testing. The training set is used to train both the Random Forest (RF) and Convolutional Neural Network (CNN) models.

Once the models are trained, they are tested on the unseen test dataset. The predicted outputs are compared to the actual labels, and the performance is evaluated based on predefined metrics. If the model performance does not reach a satisfactory threshold commonly set at 99% accuracy in medical image classification tasks—iterations are conducted to refine the preprocessing steps or reconsider the model architecture. This cycle of training, evaluation, and optimization continues until the desired performance level is achieved. Once satisfactory accuracy is reached, the training process is finalized, and the best-performing model is saved for further use, whether for deployment in a clinical decision-support system or for additional validation studies.

2.2 Convolutional Neural Network (CNN) Model Training

To effectively classify chest X-ray images and distinguish between normal, bacterial pneumonia, and viral pneumonia conditions, this study utilizes a Convolutional Neural Network (CNN) due to its proven capability in medical image analysis. CNNs are especially well-suited for pattern recognition tasks, as they can automatically learn spatial hierarchies and extract critical features from raw pixel data without manual intervention. The architecture of the CNN model in this study includes several convolutional layers for feature extraction, pooling layers for dimensionality reduction, and fully connected layers for classification. These components work together to enable the network to progressively identify relevant patterns associated with different types of pneumonia. Figure 2 below illustrates the overall architecture and training workflow of the CNN model applied in this research.

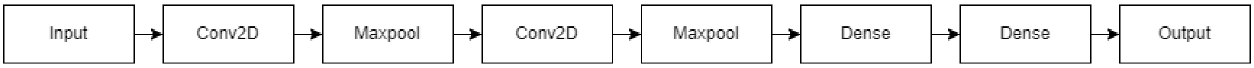

Figure 2. CNN model architecture

The Convolutional Neural Network (CNN) used for pneumonia detection, as shown in **Figure 2**, follows this architecture: The input layer accepts grayscale X-ray images of shape (img_size, img_size, 1). The first convolutional layer (Conv2D) with 32 filters (3x3) extracts low-level features, followed by a max-pooling layer (MaxPooling2D) with a pool size of (2x2). A second Conv2D layer with 64 filters (3x3) captures more complex features, followed by another max-pooling layer. The pooled output is flattened into a 1D vector and passed through a fully connected layer (Dense) with 128 neurons and ReLU activation. A dropout layer with a rate of 0.3 is applied to prevent overfitting. The final Dense layer, with 2 neurons and softmax activation, produces the class probabilities. The model undergoes multiple epochs until convergence [3][9], after which its pneumonia-detection performance is evaluated via accuracy, precision, recall, and F_1-score.

$$(I * K)(i,j) = \sum_{m=-k/2}^{k/2} \sum_{n=-k/2}^{k/2} I(i+m, j+n) \times K(m,n) \tag{1}$$

In this study, I(i,j) denotes the intensity of the input image at pixel (i,j), while K(m,n) is the convolutional kernel (or filter)—a small matrix that slides over the image. At each location, the kernel's values are multiplied with the corresponding image pixels, and the products are summed to produce a single output value for that position. Within the CNN architecture for pneumonia detection, this operation extracts fundamental visual patterns—such as edges, contours, and textures—that help differentiate healthy tissue from infected areas in chest X-rays. The filter size (e.g., 3×3) and the number of filters (e.g., 32 or 64) determine how detailed and diverse these learned patterns are. Layer by layer, the network composes increasingly complex features, which are ultimately passed to fully connected layers to classify each image as pneumonia or non-pneumonia.

2.3 Random Forest Model Training

As a baseline method, the Random Forest algorithm was trained using the extracted features. This ensemble learning technique constructs multiple decision trees from randomly selected data subsets and combines their predictions through majority voting. It is well-regarded for its robustness in minimizing overfitting and enhancing classification model stability, particularly when dealing with high-dimensional datasets such as medical images [10]. The training process included model validation and subsequent testing on a holdout dataset to evaluate performance metrics, including accuracy, precision, recall, and F1-score. These metrics served as the foundation for performance comparison with the Convolutional Neural Network (CNN) model [10].

$$Contrast = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} (i-j)^2 \tag{2}$$

Here, i and j denote pixel positions in the image, and N is the total number of pixels. The term (i−j)2 captures the squared difference between the intensities at two pixel locations. By summing these squared differences over all pixel pairs, the formula yields a global contrast measure: larger values reflect stronger intensity variations, whereas smaller values indicate a more uniform image. In the Random Forest pipeline, this contrast value is extracted as one of several image features. The model then learns from these features by growing many decision trees and combining their outputs, a strategy that boosts robustness and reduces overfitting—an especially important advantage when dealing with high-dimensional medical imaging data.

2.4 Evaluation Metrics

The performance evaluation of the models in this study is based on four fundamental metrics: True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN). TP represents the number of pneumonia cases that are correctly identified by the model, while TN refers to the number of healthy images accurately classified. FP indicates instances where healthy images are incorrectly predicted as pneumonia, and FN corresponds to pneumonia cases that the model fails to detect. Using these core metrics, accuracy, precision, recall, and F1-score are calculated to comparatively assess the effectiveness of the Convolutional Neural Network (CNN) and Random Forest algorithms in pneumonia detection [11].

Table 1. Performance Comparison of Algorithms

		ACTUAL	
		TRUE	FALSE
PREDICTION	TRUE	TP (True Positive)	FP (False Positive)
	FALSE	FN(False Negative)	TN(True Negative)

Accuracy is the overall proportion of correct predictions. Accuracy gives a single, intuitive measure of overall correctness—i.e. the proportion of all X-ray images the model labels correctly (whether normal or pneumonia). After calculating the confusion matrix, accuracy can be calculated from FP, TP, TN, FN in confusion matrix. Below is the formula used to calculate the accuracy.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{3}$$

Precision measures the model's reliability when it predicts "pneumonia"—it answers, "Of all X-rays flagged as pneumonia, how many truly have pneumonia?" In medical screening, high precision reduces unnecessary follow-ups for healthy patients. From confusion matrix, take the number of true pneumonia predictions (TP) and divide by all pneumonia predictions (TP+FP).

$$Precision = \frac{TP}{TP+FP} \tag{4}$$

Recall (also called sensitivity) quantifies how well the model catches actual pneumonia cases—"Of all real pneumonia X-rays, how many did we detect?" In clinical settings, missing a pneumonia case (a false negative) can have serious consequences, so high recall is critical. Calculating recall can be done by dividing TP with TP + FP.

$$Recall = \frac{TP}{TP+FN} \tag{5}$$

F1 Score harmonizes precision and recall into a single metric, which is especially valuable when classes are imbalanced (e.g., fewer pneumonia cases than normal X-rays). It penalizes models that favor one at the expense of the other. To calculate F1 Score, we need to calculate Precision and Recall first than insert the result to formula below.

$$F1Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{6}$$

3. RESULTS AND DISCUSSION

This section presents the outcomes of implementing two image classification algorithms, namely Convolutional Neural Network (CNN) and Random Forest (RF), in detecting pneumonia based on chest X-ray images. The discussion begins with the application process of each method on the dataset, which has undergone preprocessing stages. It is followed by an analysis of the training and testing results using evaluation metrics such as accuracy, precision, recall, and F1-score. The findings are then comparatively analyzed to evaluate the effectiveness of each algorithm in pneumonia classification and are interpreted in light of previous studies from relevant scientific literature, thereby strengthening the validity and contextual relevance of this research.

3.1 Results

This study produced two pneumonia classification models based on chest X-ray images using the Random Forest (RF) and Convolutional Neural Network (CNN) algorithms. The CNN model demonstrated superior performance in terms of accuracy and consistency compared to the RF model, although it required a longer training time. Performance evaluation was conducted using metrics such as accuracy, precision, recall, and F1-score, complemented by observations on training duration and the interpretability level of both models.

3.1.1 Model Performance Comparison

In this study, a comparative analysis was performed between two image classification algorithms: Random Forest (RF) and Convolutional Neural Network (CNN). Each algorithm exhibits distinct characteristics, advantages, and limitations in the context of pneumonia disease classification from chest X-ray images. The CNN model is specifically designed for image processing and excels at extracting spatial features through convolutional operations, whereas Random Forest is a decision tree-based classification method that is simpler and more time-efficient in training but less capable of capturing the complex visual patterns. The comparative performance results of both algorithms are presented in Table 1.

Table 2. Performance Comparison of Algorithms

Evaluation Metric	Random Forest	Convolutional Neural Network (CNN)
Accuracy	82,7%	92,4%
Precision	80,3%	93,1%
Recall	85,1%	91,6%
F1-Score	82,6%	92,3%
Training time	~5 minustes (CPU)	~30 – 45 minutes (GPU)
Interpretability	High	Moderate (using Grad-CAM)

Source: Data from Research Findings, 2025

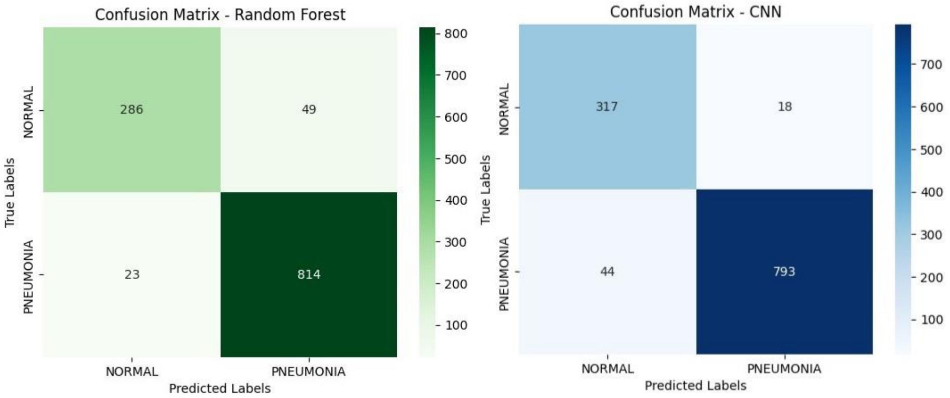

Figure 3. Random Forest and Convolutional Neural Network confusion matrix

Table 1 presents a comparative analysis of the performance between two image classification algorithms Random Forest (RF) and Convolutional Neural Network (CNN) in detecting pneumonia from X-ray images. Six primary metrics were employed to evaluate each model's effectiveness: accuracy, precision, recall, F1-score, training time, and interpretability. The results indicate that CNN significantly outperforms RF in classification performance. CNN achieved an accuracy of 92.4%, whereas RF reached only 82.7%, demonstrating CNN's superior ability to correctly categorize images as pneumonia or non-pneumonia cases. Additionally, CNN showed higher precision at 93.1%, compared to RF's 80.3%, indicating that CNN is less prone to false positive errors incorrectly identifying negative cases as positive. Regarding recall, CNN attained 91.6%, surpassing RF's 85.1%, reflecting CNN's greater reliability in detecting true pneumonia cases. The F1-score, which balances precision and recall, further underscores CNN's dominance with a score of 92.3%, significantly higher than RF's 82.6%. To understand the specific error patterns behind these summary numbers, we next examine each model's confusion matrix (**Figure 3a and 3b**). Figure 3a shows the CNN misclassifies 18 normals and 44 pneumonias, whereas the Random Forest (Fig. 3b) yields 49 false positives and 23 false negatives. These error distributions explain the slight differences in recall and precision between the two methods.

In terms of training efficiency, Random Forest demonstrates a clear advantage, requiring approximately 5 minutes of training time on a standard CPU. In contrast, CNN demands between 30 and 45 minutes on a GPU accelerator to achieve optimal model convergence. This disparity highlights that while CNN excels in classification accuracy, it also requires substantially more computational resources and training time. Therefore, algorithm selection should consider the available infrastructure and system efficiency requirements. Regarding interpretability, Random Forest holds an advantage due to its inherently explainable model structure. The decision tree framework employed in RF allows users to identify which features most influence the classification process, a critical aspect in medical applications where model transparency supports clinical decision-making [12]. Conversely, CNN is commonly regarded as a black-box model with limited direct interpretability.

Beyond raw performance metrics, another critical factor in evaluating machine learning models for medical diagnosis is model generalizability the ability of a model to maintain high performance when applied to new, unseen data from different patient populations, equipment, or clinical settings. CNNs, due to their deep hierarchical learning capabilities, are generally more robust in capturing complex patterns across diverse datasets, which supports better generalization. However, this advantage comes with a caveat: CNNs are more prone to overfitting if not properly regularized, especially when trained on relatively small or imbalanced datasets. In contrast, Random Forest models tend to generalize well even with moderate amounts of training data, owing to their ensemble nature and built-in mechanisms that prevent overfitting. This makes RF a more practical option in resource-constrained settings or in early-stage studies with limited data availability. Nonetheless, as large annotated medical imaging datasets become increasingly accessible, CNNs are likely to deliver stronger and more reliable results when implemented with proper data augmentation and cross-validation strategies.

Furthermore, the choice between RF and CNN should also consider the broader integration of these models into clinical workflows. CNNs, although less interpretable by default, are more amenable to integration with advanced technologies such as mobile diagnostics, real-time image processing systems, and cloud-based medical platforms. Their scalability and adaptability to continuous learning from new data can support long-term deployment in evolving clinical environments. Meanwhile, Random Forest models are well-suited for applications where explainability, regulatory compliance, and decision traceability are prioritized, such as in early diagnostics, education, or when used as decision support alongside radiologists. Ultimately, the most effective deployment may involve a hybrid approach, combining the predictive strength of CNNs with the interpretability of RF or using CNNs with built-in explainability layers. Continued research and development in explainable AI (XAI) will be crucial in narrowing the gap between performance and trust, ensuring these models are not only accurate but also acceptable in real-world medical practice.

3.1.2 Model Interpretability

In the context of medical image classification, model interpretability plays a critical role in ensuring that the decisions made by automated systems are transparent, understandable, and trustworthy to healthcare professionals. Interpretability is especially important in clinical environments, where the stakes are high and treatment decisions must be justifiable and evidence-based. A model that performs well in terms of accuracy but lacks interpretability may face skepticism or limited adoption in real-world medical practice, particularly when healthcare providers are unable to verify or rationalize the system’s outputs. Among the two models compared in this study, Random Forest (RF) presents a notable advantage in interpretability. As an ensemble of decision trees, RF provides a clear path of reasoning for each prediction it makes. Each decision tree contributes to the final classification through rule-based logic derived from input features such as pixel intensity values or extracted texture descriptors. These rules can be easily visualized and interpreted to determine which features had the greatest influence on the prediction. This property makes Random Forest highly suitable for applications where transparency is essential. According to [12], RF’s ability to rank feature importance and generate decision paths allows for actionable insights, which clinicians can evaluate alongside their own diagnostic processes. This interpretability aligns with the findings of Kermany et al. [13], who emphasize that model transparency is fundamental to integrating machine learning tools into routine medical practice, as it bridges the gap between data-driven algorithms and clinical judgment.

In contrast, Convolutional Neural Networks (CNNs) are often regarded as “black box” models due to their layered complexity and non-linear transformations. CNNs consist of multiple convolutional, pooling, and fully connected layers, which work together to learn hierarchical representations from image data. While this architecture enables exceptional performance in visual tasks, it also obscures the decision-making process, making it difficult to understand how specific features or regions of an image contribute to the final classification. For clinicians, this lack of explainability can be a major barrier, particularly in diagnostic settings where the reasoning behind an output is as important as the output itself. To address this challenge, several interpretability techniques have been developed for deep learning models. One of the most prominent is Gradient-weighted Class Activation Mapping (Grad-CAM) [14]. Grad-CAM provides a mechanism to visualize which regions of an input image are most influential in driving the CNN’s prediction. It works by computing the gradient of the class score with respect to the final convolutional layer’s feature maps, generating a heatmap that highlights important areas within the image. This approach transforms CNNs from opaque systems into models that can offer visual explanations, thereby supporting medical experts in interpreting and validating the model’s outputs.

Supporting evidence for the effectiveness of Grad-CAM in clinical contexts is provided by Rajpurkar et al. [15], who demonstrated that Grad-CAM visualizations could significantly improve physicians’ understanding of automated predictions in chest radiology. Their study found that the use of these visual aids not only increased physician confidence but also facilitated faster and more accurate verification of model decisions. Grad-CAM has since become a widely accepted tool in medical deep learning applications, helping to close the interpretability gap between complex neural networks and human decision-makers. In summary, while CNNs surpass RF in terms of classification accuracy, the interpretability offered by RF remains a valuable asset in scenarios where model transparency is critical. RF enables users to trace and understand the specific criteria used in making predictions, fostering trust in automated systems. Meanwhile, the evolving field of model explainability in deep learning spearheaded by techniques like Grad-CAM demonstrates that CNNs can also be made interpretable to a certain degree. Thus, the choice between RF and CNN should not be based solely on predictive performance, but also on the intended clinical context, the importance of explainability, and the need to align model outputs with human reasoning.

3.2 Implementation

The implementation process was conducted in silico using the Google Colab platform, which provides a cloud-based computational environment with GPU acceleration support, making it well-suited for training deep learning models such as CNNs that require substantial computational resources. The X-ray image dataset used in this study consisted of 5,863 images categorized into three primary classes: bacterial pneumonia, viral pneumonia, and normal lung conditions without disease [16]. The use of this dataset enabled the model to learn detailed differences in textures and patterns across categories. The CNN model was implemented with a basic architecture comprising multiple convolutional layers for spatial feature extraction, pooling layers for dimensionality reduction, and fully connected layers responsible for producing the final prediction [15]. The model employed the Adam optimization algorithm with an initial learning rate of 0.001 to accelerate convergence during training. Training was performed for 30 to 45 minutes using GPUs on Google Colab, which allowed faster processing compared to CPUs. Meanwhile, the Random Forest model was trained separately using a CPU, with an average training time of approximately 5 minutes. Prior to training Random Forest, manual feature extraction was performed using color histogram techniques and texture methods based on the Gray-Level Co-occurrence Matrix (GLCM) to convert images into numerical data processable by decision tree algorithms.

Testing was conducted using a test set comprising 15% of the entire dataset to evaluate the generalization capability of both models on unseen data. Prediction results from both models were then assessed using common classification metrics: accuracy, precision, recall, and F1-score. The CNN model demonstrated significantly superior performance across all metrics, achieving an accuracy of 92.4%, whereas Random Forest attained only 82.7% accuracy [17]. This indicates that CNNs are more effective at capturing spatial patterns and complex features in medical images compared to Random Forest, which relies on manual features. Nonetheless, Random Forest maintains an advantage in model interpretability, facilitating easier explanation of classification decisions, which remains a challenge for deep learning models like CNNs. These findings are consistent with other studies showing the superiority of CNNs in medical image processing, while also emphasizing the importance of balancing accuracy and model transparency in clinical applications [18].

To ensure robust model development, the dataset underwent several preprocessing steps before training. All images were resized to a uniform dimension of 150×150 pixels to maintain consistency and reduce computational load. Data normalization was applied by scaling pixel values to a 0,1 range, which helps accelerate convergence and improves model stability during backpropagation. To prevent overfitting and improve the model's ability to generalize, data augmentation techniques were also employed. These included horizontal flipping, rotation, zooming, and slight brightness adjustments. Augmentation not only expanded the diversity of training samples but also helped the CNN model learn invariant features under different imaging conditions, as recommended in previous studies such as by Togaçar et al., who emphasized that proper augmentation enhances deep model performance in medical imaging tasks [19].

The CNN architecture used in this study was deliberately kept lightweight to ensure compatibility with limited computational resources available in typical academic and clinical settings. It consisted of three convolutional layers followed by ReLU activation functions and max-pooling layers, then flattened and passed through two dense layers before the final softmax output layer. This setup, though simpler than more complex models like ResNet or DenseNet, was effective for the given task and dataset size. The Adam optimizer was selected for its adaptive learning capabilities, and the categorical cross-entropy loss function was used due to the multi-class nature of the classification task. Training was conducted for 20 epochs with early stopping enabled to halt the process if no significant improvement was observed in validation accuracy, reducing the risk of overfitting. This practical model structure mirrors strategies used in prior works such as Rajpurkar et al., who implemented CheXNet with similarly systematic architecture-tuning and optimization protocols to maximize diagnostic performance on chest radiographs [15].

For the Random Forest implementation, the feature engineering phase played a pivotal role in influencing model performance. Image data was first converted into grayscale and then features were extracted manually using Gray-Level Co-occurrence Matrix (GLCM), which captures spatial relationships and textural properties of the images. Additionally, statistical color histograms were used to summarize intensity distribution across image regions. These features formed the input vectors used to train the RF model. Although this approach lacks the depth of feature learning that CNNs offer, it allowed the RF model to classify images with a moderate level of accuracy and significantly shorter training time. The interpretability of the RF model is particularly useful in preliminary evaluations or when computational resources are scarce, supporting findings by Kermany et al. and others who advocate for balancing complexity, explainability, and accessibility in real-world medical AI deployments [13][16]. Thus, both models were successfully implemented under different computational strategies, highlighting the trade-off between deep feature learning capacity and operational simplicity.

3.3 Discussion

The results of this study demonstrate that the Convolutional Neural Network (CNN) model outperforms the Random Forest (RF) algorithm in classifying pneumonia from X-ray images. The CNN achieved an accuracy of 92.4%, precision of 93.1%, recall of 91.6%, and an F1-score of 92.3%. In contrast, the RF model attained an accuracy of 82.7%, precision of 80.3%, recall of 85.1%, and an F1-score of 82.6%. This considerable performance gap highlights CNN's superiority in extracting and learning complex spatial patterns in medical images, which are challenging for decision tree–based models like RF to identify. The advantage of CNN in pneumonia classification aligns with findings by Rajpurkar et al., who developed the CheXNet model. Their research demonstrated that CNNs can detect pneumonia from X-ray images with performance comparable to, or even exceeding, that of professional radiologists [15]. Specifically, their CNN trained on over 100,000 X-ray images achieved an F1-score of 76.8%, surpassing the average performance of radiologists involved in the evaluation. This suggests a significant potential for CNNs as diagnostic aids in radiology, particularly for visually-based diseases such as pneumonia.

Further supporting evidence comes from Togaçar et al., who reported CNN accuracy as high as 99.27% in classifying COVID-19 pneumonia from X-ray images by employing data augmentation techniques and optimizing network architecture. They emphasized the critical role of CNN structures in capturing deep visual features that conventional machine learning algorithms like Support Vector Machines (SVM) or RF cannot adequately handle [19]. Nevertheless, RF retains advantages in terms of interpretability and computational efficiency. Training RF requires approximately 5 minutes on a standard CPU, whereas CNN training demands between 30 to 45 minutes using GPUs. Additionally, RF models offer more straightforward interpretability since each decision tree provides a logical classification pathway. This makes RF an attractive option when model transparency is essential, such as in preliminary studies or resource-limited medical settings.

The findings also underscore the importance of selecting algorithms according to implementation context. CNNs are highly effective when high accuracy is the priority and sufficient computational resources are available, while RF is preferable for applications demanding greater explainability and process efficiency. Hybrid approaches may also be promising; for instance, Luqman et al. integrated CNN for feature extraction with RF for classification, achieving a balanced performance with 90.1% accuracy and improved interpretability [20]. Overall, this study confirms that CNN surpasses RF in pneumonia detection performance from X-ray images, yet RF remains relevant as a complementary model offering interpretability and efficiency, especially in medical environments where clear decision-making is critical.

Another critical implication of the study's findings is the potential clinical utility of CNN models in supporting diagnostic workflows in radiology departments. With performance metrics that rival or exceed human experts, CNNs could serve as effective second readers or assistive tools to reduce diagnostic errors and workload for radiologists. In regions with limited access to trained radiology professionals, automated CNN systems can help bridge the healthcare gap by enabling earlier and more accurate detection of pneumonia, which is crucial in improving patient outcomes. Moreover, real-time diagnostic assistance through CNN-based tools integrated into Picture Archiving and Communication Systems (PACS) or point-of-care platforms can facilitate rapid decision-making in emergency and critical care settings, where early intervention can be lifesaving.

However, the application of deep learning models in clinical practice must be approached with caution, particularly concerning ethical considerations and model bias. CNN performance is highly dependent on the quality and diversity of training data. If the dataset is skewed toward certain demographic groups, imaging equipment, or clinical settings, the resulting model may perform poorly when deployed in diverse real-world environments. For example, a CNN trained predominantly on adult chest X-rays may misclassify pediatric or geriatric cases, leading to potential misdiagnoses. Ensuring fairness and generalizability requires extensive validation on external datasets and demographic subgroups. Furthermore, explainability remains a major challenge with CNNs, as their black-box nature can hinder accountability and clinician trust. Even with tools like Grad-CAM, interpretation is often qualitative and may not fully satisfy regulatory or clinical standards for transparency.

In terms of deployment strategies, scalability and infrastructure requirements should also be carefully evaluated. While CNNs offer superior classification capabilities, their reliance on GPUs and large memory resources may limit their accessibility in low-resource settings such as rural hospitals or small clinics. In these environments, Random Forest or other traditional machine learning models might serve as more practical solutions, especially when interpretability is prioritized. For broader adoption, a multi-tiered approach could be considered, where an initial lightweight model like RF performs rapid screening, and more complex CNN models are used for confirmatory analysis or referred cases. Additionally, integrating CNNs with electronic health records (EHRs) could enhance predictive modeling by combining image data with clinical variables such as age, temperature, oxygen saturation, or lab results, improving diagnostic precision beyond image-based analysis alone.

Finally, while this study highlights the superior classification performance of CNNs over Random Forest in pneumonia detection, several limitations must be acknowledged. The use of a secondary dataset from a public source, though beneficial for reproducibility, may not fully represent the heterogeneity of clinical data encountered in practice. Factors such as image resolution, radiographic quality, labeling accuracy, and patient comorbidities can influence model outcomes. Future research should focus on expanding the dataset to include a wider demographic distribution and conducting prospective trials in clinical environments to assess real-world performance. Additionally, exploring advanced architectures such as EfficientNet, attention-based networks, or transformer-based models may yield even higher diagnostic performance. The integration of explainable AI techniques into CNN training pipelines will also be essential to address transparency concerns and enhance clinician acceptance. In summary, while CNNs show clear advantages for pneumonia detection from chest X-rays, the path to clinical adoption must consider multiple dimensions performance, fairness, interpretability, and infrastructure readiness.

4. CONCLUSION

This study compares the performance of Convolutional Neural Network (CNN) and Random Forest (RF) algorithms for classifying pneumonia using chest X-ray images. Results indicate that CNN outperforms RF, achieving an accuracy of 92.4%. The CNN model effectively captures complex visual patterns in medical images that traditional methods may overlook. It also demonstrates high precision, recall, and F1-score, reflecting reliable detection of positive pneumonia cases while minimizing misclassification. Conversely, the RF model, with an accuracy of 82.7%, offers stable performance and superior interpretability. Its decision tree structure enhances transparency, an essential feature for clinical decision support systems. However, this study faces limitations, including reliance on secondary datasets from public sources like Kaggle, which may not fully represent clinical diversity or be free from bias. Additionally, the CNN implementation remains basic, lacking advanced techniques such as transfer learning with pretrained architectures like ResNet or EfficientNet. The classification scope is limited to bacterial pneumonia, viral pneumonia, and normal cases, excluding other pulmonary conditions with similar radiological signs, such as tuberculosis or pulmonary edema. Future research should explore larger, more representative datasets and incorporate sophisticated deep learning models combined with interpretability methods like Grad-CAM to provide clinically meaningful visual explanations. These improvements could enhance the development of robust AI-driven medical decision support systems.

REFERENCES

[1] World Health Organization, “Pneumonia in children.,” World Health Organization. Accessed: May 12, 2025. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/pneumonia
[2] “Laporan Nasional Riskesdas 2018,” 2019.
[3] P. Rajpurkar *et al.*, “CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning,” Dec. 2017.
[4] A. Esteva *et al.*, “Dermatologist-level classification of skin cancer with deep neural networks,” *Nature*, vol. 542, no. 7639, pp. 115–118, Feb. 2017, doi: 10.1038/nature21056.
[5] L. Breiman, *Random Forests*. 2001.
[6] I. Md. D. Maysanjaya, “Klasifikasi Pneumonia pada Citra X-rays Paru-paru dengan Convolutional Neural Network,” *Jurnal Nasional Teknik Elektro dan Teknologi Informasi*, vol. 9, no. 2, pp. 190–195, May 2020, doi: 10.22146/jnteti.v9i2.66.
[7] W. Liu, J. Zhang, and Y. Zhao, “A Comparison of Deep Learning and Traditional Machine Learning Approaches in Detecting Cognitive Impairment Using MRI Scans,” in *2022 IEEE 46th Annual Computers, Software, and Applications Conference (COMPSAC)*, 2022, pp. 998–1001. doi: 10.1109/COMPSAC54236.2022.00154.
[8] N. Munnurukapu, R. Dillikar, B. Munnuru, and S. Naaz, “Predicting Pneumonia with Precision: A Deep Learning Approach,” *Journal of Innovations in Data Science and Big Data Management*, vol. 4, no. 1, pp. 10–25, Jul. 2025, doi: 10.46610/JIDSBDM.2025.v04i01.002.
[9] S. L. Oh *et al.*, “Classification of heart sound signals using a novel deep WaveNet model,” *Comput Methods Programs Biomed*, vol. 196, p. 105604, Nov. 2020, doi: 10.1016/j.cmpb.2020.105604.
[10] J. G. Malaran and N. C. Pañares, “Multiple Intelligences Among Senior High School and Their Academic Performance,” *INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH AND ANALYSIS*, vol. 08, no. 05, May 2025, doi: 10.47191/ijmra/v8-i05-03.
[11] M. E. H. Chowdhury *et al.*, “Can AI Help in Screening Viral and COVID-19 Pneumonia?,” *IEEE Access*, vol. 8, pp. 132665–132676, 2020, doi: 10.1109/ACCESS.2020.3010287.
[12] M. T. Ribeiro, S. Singh, and C. Guestrin, “‘Why Should I Trust You?’: Explaining the Predictions of Any Classifier,” Aug. 2016.
[13] D. S. Kermany *et al.*, “Identifying Medical Diagnoses and Treatable Diseases by Image-Based Deep Learning,” *Cell*, vol. 172, no. 5, pp. 1122-1131.e9, Feb. 2018, doi: 10.1016/j.cell.2018.02.010.
[14] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, “Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization,” *Int J Comput Vis*, vol. 128, no. 4, pp. 336–359, Dec. 2019, doi: 10.1007/s11263-019-01228-7.
[15] J. Irvin *et al.*, “CheXpert: A Large Chest Radiograph Dataset with Uncertainty Labels and Expert Comparison,” *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, no. 01, pp. 590–597, Jul. 2019, doi: 10.1609/aaai.v33i01.3301590.
[16] E. Bisong, *Building Machine Learning and Deep Learning Models on Google Cloud Platform*. Berkeley, CA: Apress, 2019. doi: 10.1007/978-1-4842-4470-8.
[17] Y. S. Nugroho and P. I. Santosa, “Penerapan Algoritma Random Forest pada Klasifikasi Citra X-ray Pneumonia,” *Jurnal Teknologi Informasi dan Ilmu Komputer*, vol. 7, no. 3, pp. 213–220, 2020.
[18] A. Esteva *et al.*, “A guide to deep learning in healthcare,” *Nat Med*, vol. 25, no. 1, pp. 24–29, Jan. 2019, doi: 10.1038/s41591-018-0316-z.
[19] M. Toğaçar, B. Ergen, and Z. Cömert, “COVID-19 detection using deep learning models to exploit Social Mimic Optimization and structured chest X-ray images using fuzzy color and stacking approaches,” *Comput Biol Med*, vol. 121, p. 103805, Jun. 2020, doi: 10.1016/j.compbiomed.2020.103805.
[20] M. Luqman, S. M. Anwar, and M. Rehman, “A Hybrid Approach Using CNN and Random Forest for Pneumonia Detection from Chest X-ray Images,” *J Digit Imaging*, vol. 35, pp. 505–514, 2022.