MobileNet-Driven detection of bacterial and viral pneumonia with Grad-CAM heatmap insights

Yuri Pamungkas [a,1,*], Adrian Jaleco Forca [b,2], Muhammad Nur Afnan Uda [c,3], Uda Hashim [d,4]

[a] Department of Medical Technology, Institut Teknologi Sepuluh Nopember, Surabaya 60111, Indonesia
[b] College of Science and Technology, Guimaras State University, Buenavista 5044, Philippines
[c] Department of Electronic and Computer Engineering, Universiti Malaysia Sabah, Kinabalu 88400, Malaysia
[d] Department of Electrical and Electronics Engineering, Universiti Malaysia Sabah, Kinabalu 88400, Malaysia
[1] yuri@its.ac.id; [2] adrian.forca@gsu.edu.ph; [3] nurafnan@ums.edu.my; [4] uda@ums.edu.my
* corresponding author

ARTICLE INFO

Article history
Received September 8, 2025
Revised December 23, 2025
Accepted January 11, 2026
Available online May 31, 2026

Keywords
MobileNet
MobileNetV2
Pneumonia Classification
Chest X-ray
Grad-CAM

ABSTRACT

Pneumonia remains a major cause of morbidity and mortality in children, and chest X-ray imaging is the most widely available diagnostic modality. Differentiating bacterial from viral pneumonia remains challenging due to overlapping radiographic features. This study investigates lightweight MobileNet architectures combined with Gradient-weighted Class Activation Mapping (Grad-CAM) to enable efficient and interpretable pneumonia classification. The contribution of this work is to emphasize a balanced trade-off between diagnostic performance, computational efficiency, and explainability rather than state-of-the-art accuracy. A dataset of 5,842 pediatric chest X-ray images from Guangzhou Women and Children's Medical Center was used, including bacterial, viral, and normal cases. MobileNet and MobileNetV2 were trained using stochastic gradient descent with categorical cross-entropy for 20 epochs and a batch size of 32, and evaluated using 10-fold cross-validation. Grad-CAM was applied to visualize class-discriminative lung regions contributing to model predictions. MobileNet outperformed MobileNetV2 across most metrics, achieving 79.32% accuracy, 81.02% precision, 78.15% recall, 77.82% F1-score, 89.49% specificity, and AUC-ROC values of 94.80% (macro) and 90.52% (micro). MobileNetV2 achieved 76.44% accuracy and a macro AUC-ROC of 93.70%. Grad-CAM visualizations showed that both models predominantly focused on clinically relevant lung regions, supporting transparent decision-making. In conclusion, MobileNet with Grad-CAM provides an efficient and interpretable framework for pneumonia screening. Although its performance does not exceed larger architectures, its favorable efficiency–interpretability trade-off makes it suitable for resource-limited clinical settings.

1. Introduction

Pneumonia remains a major global health challenge, causing significant morbidity and mortality, particularly among children and elderly populations [1]. Differentiating between bacterial and viral pneumonia is essential for guiding appropriate clinical management and antibiotic stewardship [2]. However, chest X-ray (CXR), the most widely available and cost-effective diagnostic imaging modality, often presents overlapping radiographic features between bacterial and viral cases [3]. This overlap leads to diagnostic uncertainty and potential delays in treatment decisions, especially in resource-constrained healthcare settings where expert radiologists may be limited [4]. Such diagnostic ambiguity may

contribute to inappropriate antibiotic use or delayed targeted therapy, reinforcing the need for reliable decision-support systems. The research problem addressed in this study is therefore the difficulty in accurately distinguishing bacterial from viral pneumonia using conventional radiological interpretation alone, particularly in pediatric cases [5].

In recent years, deep learning methods have shown strong performance in pneumonia detection from CXR images. High-capacity architectures such as Inception v3 with Integrated Gradients achieved excellent accuracy and explainability [6], while attention-based models [7], Neural Architecture Search [8], Vision Transformers [9], ensemble learning [10]–[12], and feature optimization techniques [13] have further pushed accuracy beyond 97–99%. However, most of these approaches rely on computationally intensive architectures that may limit practical deployment in real-world clinical settings. Consequently, high numerical performance alone does not guarantee clinical usability or adoption.

Lightweight architectures such as MobileNet have gained attention due to their efficiency, low computational cost, and scalability in low-resource environments [14]. In many clinical scenarios, particularly primary care and resource-limited settings, deployability and transparency are equally critical as accuracy. While prior studies have incorporated explainability, they often depend on complex backbone models [15]. Without clear interpretability, clinical trust in AI-based systems remains limited [16]. Therefore, this study integrates MobileNet with Grad-CAM to generate visual explanations that highlight diagnostically relevant lung regions [17], [18]. This design prioritizes a balanced trade-off between performance, interpretability, and real-world feasibility rather than maximum accuracy alone.

Although this study utilizes a well-known public pediatric CXR dataset from a single hospital, this characteristic reflects a realistic clinical context rather than a limitation. Single-center pediatric datasets represent consistent imaging protocols commonly encountered in routine practice, making them suitable for evaluating clinically applicable AI frameworks. Moreover, using a widely adopted dataset enables direct comparison with prior studies, allowing the contribution of model interpretability and architectural efficiency to be more clearly assessed [19], [20]. Thus, the dataset is treated as a controlled benchmark rather than the primary source of novelty. The novelty of this research lies not in achieving state-of-the-art accuracy but in emphasizing interpretability and clinical relevance. By prioritizing transparent decision-making through Grad-CAM visualization on a lightweight MobileNet architecture, this study addresses practical barriers to clinical AI adoption. The contribution is a pneumonia classification framework that balances diagnostic performance, explainability, and deployment feasibility, thereby supporting more trustworthy and applicable AI-assisted diagnosis in everyday healthcare environments.

2. Method

2.1. Dataset

This study made use of a publicly available chest X-ray dataset that was obtained from pediatric patients between one and five years old [21]. The images were collected retrospectively at the Guangzhou Women and Children's Medical Center in Guangzhou, China. All chest X-ray examinations had been performed as part of routine clinical care, and the data were anonymized before being made available for research. The dataset specifically consists of anterior–posterior chest X-rays, which are the preferred view in pediatric practice because children in this age group are often unable to cooperate for standard posterior–anterior positioning. The dataset includes three categories of diagnosis, namely bacterial pneumonia, viral pneumonia, and normal cases. A total of 5,842 chest X-ray images were included. Among them, 2,773 images were identified as bacterial pneumonia, 1,494 images as viral pneumonia, and 1,575 images as normal. To evaluate the model fairly, the dataset was divided into a training set and a testing set. The training set contained 5,217 images, consisting of 2,531 bacterial pneumonia, 1,345 viral pneumonia, and 1,341 normal. The testing set contained 625 images, consisting of 242 bacterial pneumonia, 149 viral pneumonia, and 234 normal.

This dataset has been widely used in previous studies on deep learning approaches for pneumonia detection because it provides a relatively balanced distribution between bacterial and viral pneumonia and is based on clinically relevant pediatric populations [22], [23]. The inclusion of both bacterial and viral categories reflects the real-world diagnostic challenge in distinguishing between these two conditions [24], while the presence of normal chest radiographs ensures that the classification task remains clinically meaningful [25]. The use of this dataset in the present study ensures both clinical relevance and comparability with existing research. In addition, the focus on anterior–posterior pediatric chest X-rays highlights the importance of developing reliable diagnostic support systems for vulnerable populations where timely and accurate diagnosis plays a decisive role in treatment outcomes.

2.2. Pre-Processing

Before training the MobileNet architecture, all chest X-ray images underwent a standardized pre-processing pipeline to enhance data quality and to increase the robustness of the model [26]. Each image was rescaled so that the pixel intensity values ranged between 0.0 and 1.0, ensuring normalization of the RGB channels and facilitating more stable training convergence (Fig. 1). This step minimized the effects of varying image brightness and contrast across the dataset [27]. To improve the generalization ability of the model and reduce the risk of overfitting, data augmentation techniques were applied [28]. Random rotations were introduced within a range of minus fifty to plus fifty degrees, which allowed the model to adapt to variations in patient positioning during X-ray acquisition. Zooming was also applied, with random scaling between one minus 0.5 and one plus 0.5, enabling the model to learn from both magnified and reduced visual perspectives.

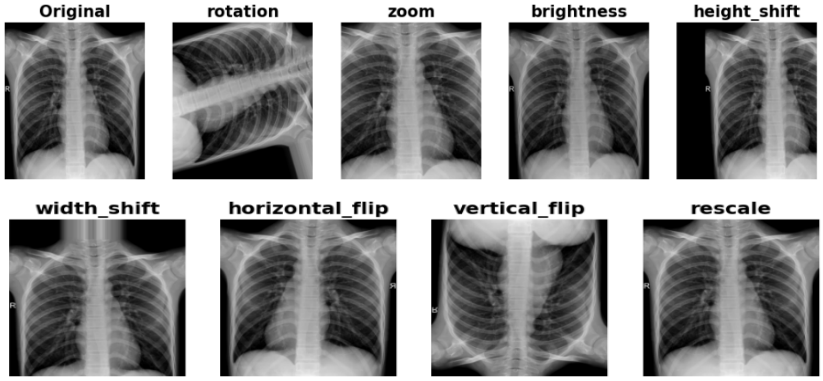


Fig. 1. Pre-processing of CXR images

Brightness augmentation was included by adjusting the illumination levels randomly within the interval of 0.3 to 1.0. This accounted for differences in exposure and imaging conditions that are often encountered in real-world clinical practice. In addition, images were randomly flipped both horizontally and vertically. Horizontal flipping simulated left–right variations in anatomical presentation, while vertical flipping introduced robustness against orientation inconsistencies that may occur during pediatric imaging procedures [29]. Spatial translation was incorporated through random shifts along both vertical and horizontal axes. The images were shifted up or down within a range of negative thirty percent to positive thirty percent of the image height, and left or right within the same proportion of the image width. These transformations exposed the model to diverse positional variations of lung fields within the radiographs, thereby reinforcing its ability to detect pneumonia-related patterns regardless of image alignment [30], [31].

2.3. MobileNet Architectures

MobileNet is a lightweight CNN architecture designed for efficient computation while maintaining competitive performance in image classification and recognition tasks [32]. Unlike conventional deep CNNs that rely on computationally expensive convolutional operations, MobileNet introduces a streamlined design that reduces the number of parameters and operations [33], [34]. The MobileNet architecture begins with an input layer that processes chest X-ray images of size 224×224×3 (Fig. 2). These images, although grayscale in practice, are formatted with three channels to align with standard

pre-trained convolutional networks. The first layer is a 3×3 convolution with 32 filters and a stride of 2, which reduces the spatial resolution to 112×112×32 while extracting basic low-level features. After the initial convolution, the network applies a sequence of depthwise separable convolutions, the key innovation of MobileNet. Each block consists of a depthwise convolution followed by a pointwise convolution. The depthwise convolution applies one filter per input channel, focusing on spatial filtering, while the pointwise convolution (1×1 kernel) combines these outputs across channels. This design drastically reduces computational complexity compared to standard convolution. For example, at the first stage, the feature map is expanded from 32 to 64 channels while preserving the 112×112 resolution.

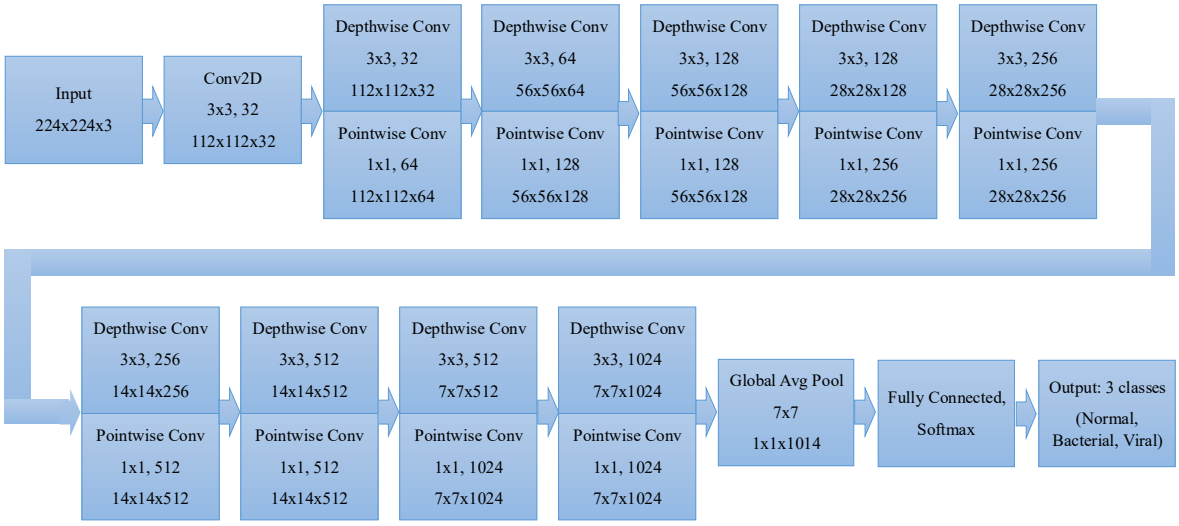


Fig. 2. MobileNet Architecture

The network then continues to increase representational capacity by stacking additional depthwise separable convolution blocks. With strides set at specific stages, the spatial resolution of the feature maps is gradually reduced from 112×112 to 56×56, then to 28×28, and eventually to 14×14 and 7×7 (Fig. 3). Simultaneously, the number of filters increases, moving from 128 and 256 channels up to 512 and 1024. This progression allows the network to capture both fine-grained local features and broader semantic structures, which are essential for differentiating bacterial pneumonia, viral pneumonia, and normal chest radiographs. A particularly important stage occurs at the 14×14 resolution, where multiple depthwise separable convolution layers with 512 filters are repeated. These repeated operations strengthen the model's ability to recognize mid-level patterns in the radiographs, such as opacities and consolidation areas. As the network approaches the final layers, depthwise and pointwise convolutions with 1024 filters at 7×7 resolution consolidate high-level abstract features relevant for classification. To transition from convolutional feature maps to class predictions, the architecture employs global average pooling. This layer reduces each 7×7×1024 feature map into a compact 1×1×1024 vector by averaging spatial information. The resulting representation is both efficient and robust against overfitting, as it eliminates the need for fully connected layers with large parameter counts. Finally, a fully connected layer with Softmax activation produces the output, classifying each input chest X-ray into one of three categories: normal, bacterial pneumonia, or viral pneumonia.

MobileNetV2 was designed as an improvement over MobileNetV1, addressing both efficiency and representational capacity [35]. Its architecture emphasizes the use of inverted residual structures combined with linear bottlenecks, allowing the model to capture complex image features without a significant increase in computational cost [36]. The MobileNetV2 architecture begins with an input of 224×224×3, representing resized chest X-ray images. The first operation is a 3×3 convolution layer with 32 filters, which reduces the resolution to 112×112×32 and extracts low-level features from the input images. This stage prepares the data for the series of inverted residual bottleneck blocks that form the core of MobileNetV2. Each bottleneck block is defined by four parameters, the expansion factor (t), the number of output channels (c), the number of repetitions (n), and the stride (s). In the first stage, the

network applies a bottleneck block with t=1 and c=16, maintaining the resolution at 112×112×16. Subsequent bottlenecks gradually reduce the spatial resolution while increasing the depth of feature maps. For example, blocks with t=6 and c=24 reduce the feature maps to 56×56×24, while blocks with t=6 and c=32 generate 28×28×32 outputs. This progression continues with blocks producing 14×14×64 and 14×14×96 feature maps, capturing increasingly abstract visual patterns relevant for pneumonia detection.

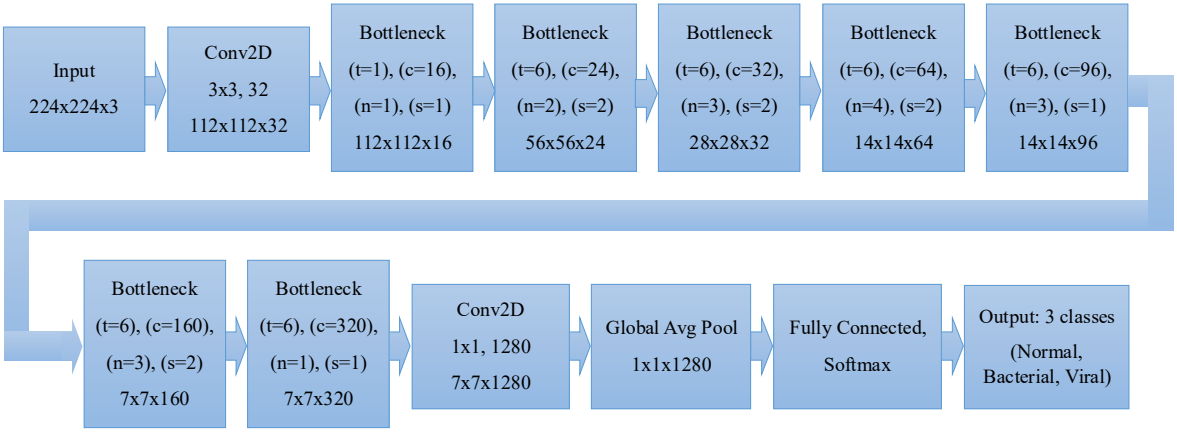


Fig. 3. MobileNetV2 Architecture

Deeper layers expand the network's representational capacity further. Bottleneck blocks with t=6 and c=160 downsample the features to 7×7×160, and a subsequent block projects these into 7×7×320. These stages allow the model to learn high-level semantic features, consolidating information about lung opacity patterns, structural abnormalities, and other diagnostic cues. A 1×1 convolution with 1280 filters then transforms the features into a dense representation of size 7×7×1280, ensuring sufficient discriminative power before classification. To move from feature extraction to prediction, the network employs global average pooling, which reduces the feature maps into a compact 1×1×1280 vector. This step minimizes the risk of overfitting by reducing parameters while retaining essential information. Finally, a fully connected layer with a Softmax activation outputs the probabilities for three clinically relevant classes (normal, bacterial pneumonia, and viral pneumonia).

2.4. Grad-CAM

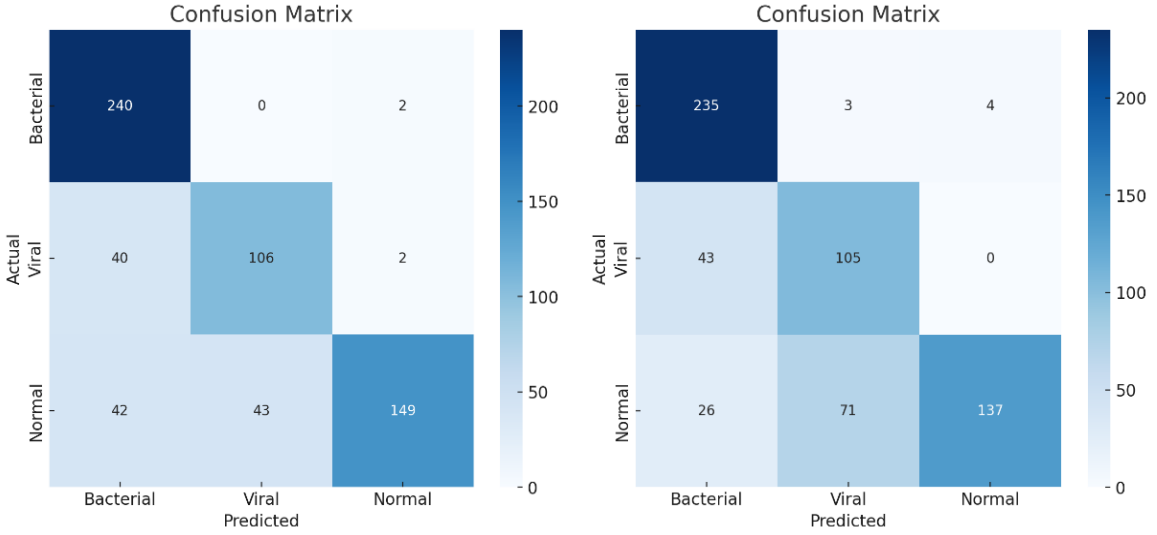
Grad-CAM is an interpretability technique designed to visualize the regions of an image that contribute most to a model's prediction [37]. Instead of treating deep neural networks as black-box classifiers, Grad-CAM highlights the discriminative areas in the input that strongly influence the decision [38], [39]. By backpropagating the gradients of the target class through the final convolutional layers, Grad-CAM produces a localization map that identifies important spatial features while preserving contextual information [40], [41]. In the context of pneumonia detection from chest X-ray images, Grad-CAM provides valuable insights into whether the model is focusing on clinically relevant lung regions such as areas of opacity, consolidation, or infiltration [42], [43]. This interpretability is crucial for bridging the gap between artificial intelligence models and clinical practice, since medical professionals require transparency in decision-making tools [44].

3. Results and Discussion

The experiments were conducted using standardized hyperparameters to ensure consistency and comparability between the models. Both MobileNet and MobileNetV2 were trained using the SGD optimizer with categorical cross-entropy as the loss function. Softmax activation was applied at the output layer to enable probabilistic classification across the three classes, which include normal, bacterial pneumonia, and viral pneumonia. The models were trained for 20 epochs with a batch size of 32. The selection of a limited number of training epochs was intended to reduce the risk of overfitting and to reflect realistic computational constraints in clinical deployment, and preliminary experiments showed that extending training beyond this range did not lead to stable performance gains.

Model robustness was validated using 10-fold cross-validation. While confusion matrices and detailed class-wise results are presented for a representative fold to support interpretability, the overall performance trends were consistent across folds, indicating stable model behavior. Evaluation was carried out using multiple performance metrics including accuracy, precision, recall, specificity, F1-score, and AUC-ROC to provide a comprehensive assessment of classification effectiveness. This multi-metric evaluation approach is particularly important when working with imbalanced clinical datasets, where overall accuracy alone may not adequately reflect class-specific performance.

The evaluation of MobileNet demonstrated strong performance in distinguishing bacterial pneumonia from the other two classes, as reflected in the confusion matrix. Out of 242 bacterial pneumonia cases, 240 were correctly classified, with only minimal misclassification into the normal class (Fig. 4). Viral pneumonia cases were moderately well identified, with 106 out of 149 correctly classified, while some were misclassified as bacterial or normal. For the normal class, 149 out of 234 samples were correctly predicted, while a considerable proportion was misclassified as either bacterial or viral. This behavior can be partly explained by class imbalance and the substantial overlap of radiographic features between viral pneumonia and normal lung appearances. These findings indicate that MobileNet performs particularly well in detecting bacterial pneumonia but encounters greater difficulty when separating viral pneumonia from normal cases, which often exhibit subtle and nonspecific imaging patterns.



(a) MobileNet (b) MobileNetV2

Fig. 4. Confusion Metrics of Pneumonia Detection

MobileNetV2 exhibited a comparable classification profile with a slightly different distribution of errors. The confusion matrix showed that 235 out of 242 bacterial pneumonia cases were correctly identified, while viral pneumonia resulted in 105 correct predictions. The normal class achieved 137 correct classifications out of 234 samples, with fewer instances misclassified as bacterial but more misclassified as viral compared to MobileNet. The consistency of these patterns across both architectures suggests that the observed limitations are more strongly related to dataset characteristics and class imbalance rather than model instability or insufficient training.

The evaluation of MobileNet and MobileNetV2 models provides comprehensive insights into their comparative performance for the classification of bacterial pneumonia, viral pneumonia, and normal chest X-ray images (Fig. 5). As summarized in Table 1, MobileNet achieved a higher overall accuracy of 79.32% compared to 76.44% obtained by MobileNetV2. Although this headline accuracy is lower than that reported by more complex architectures, it should be interpreted in the context of the lightweight design and deployment-oriented objective of the proposed model. This finding indicates that MobileNet was able to provide more consistent predictions across the test dataset, supporting its role as an efficient baseline model rather than a state-of-the-art performance benchmark.

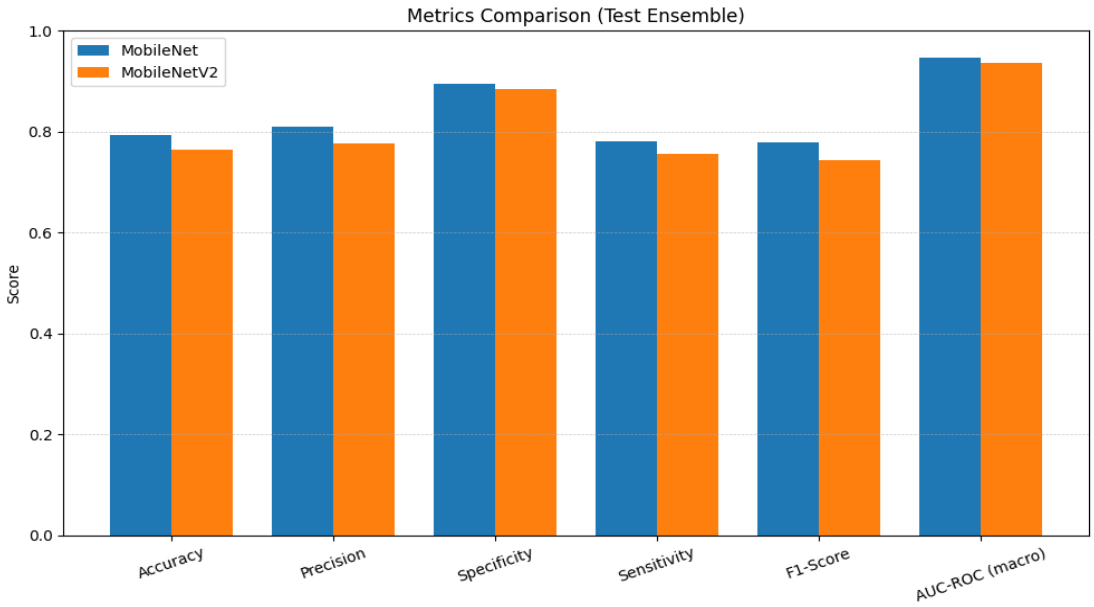


Fig. 5. Metrics Comparison of Pneumonia Detection using MobileNet and MobileNetV2

Table 1. Evaluation Metrics (Test Ensemble)

Metric	MobileNet (%)	MobileNetV2 (%)
Accuracy	79.32	76.44
Precision	81.02	77.70
Recall	78.15	75.53
Specificity	89.49	88.45
F1-Score	77.82	74.45
AUC-ROC (Macro)	94.80	93.70
AUC-ROC (Micro)	90.52	88.78

A closer examination of precision and recall reveals further distinctions between the two architectures. MobileNet demonstrated a precision of 81.02%, outperforming MobileNetV2 at 77.70%. This indicates that when MobileNet predicts pneumonia, particularly bacterial pneumonia, it does so with relatively high confidence, reducing the likelihood of incorrect positive predictions for severe cases. MobileNet also achieved a recall of 78.15% compared to 75.53% for MobileNetV2, suggesting a slightly stronger ability to capture true pneumonia cases. These trends are reflected in the F1-score, where MobileNet recorded 77.82% versus 74.45% for MobileNetV2, illustrating a more balanced trade-off between sensitivity and specificity.

Despite these strengths, both models exhibited weaker performance in the Normal class, with MobileNet correctly identifying 149 out of 234 normal cases. This limitation can be attributed to the conservative decision behavior of the model, which tends to favor pneumonia detection in ambiguous cases in order to minimize missed diagnoses. From a clinical perspective, this bias toward sensitivity may increase false positive rates for healthy patients but reduces the risk of failing to detect pneumonia, which is often considered a more critical error in early screening scenarios. Such behavior is consistent with clinical decision-support systems designed for triage, where suspected cases are flagged for further expert review rather than final diagnosis.

Both models demonstrated high specificity, with MobileNet achieving 89.49% and MobileNetV2 achieving 88.45%, indicating a strong overall capacity to identify negative cases across the dataset. The discrepancy between class-wise performance highlights the inherent difficulty of distinguishing normal chest X-rays from mild or early-stage viral pneumonia, which often share subtle and overlapping radiographic features. The discriminative capability of the models was further supported by AUC-ROC analysis (Fig. 6). MobileNet achieved a macro-averaged AUC-ROC of 94.80% and a micro-averaged AUC-ROC of 90.52%, outperforming MobileNetV2 in both measures. These high AUC values suggest

that, despite moderate accuracy at a fixed decision threshold, the underlying feature representations learned by MobileNet remain robust and separable across classes.

Taken together, the results indicate that MobileNet, despite its relatively simpler architecture, surpassed MobileNetV2 in several critical evaluation metrics. The superior performance of MobileNet suggests that depthwise separable convolutions provided sufficient representational capacity for this task, whereas the additional complexity introduced in MobileNetV2 through inverted residual bottlenecks did not yield substantial improvements. From an academic perspective, this underscores the importance of aligning model complexity with the characteristics of the dataset, and highlights the relevance of lightweight models such as MobileNet for clinically oriented applications where computational efficiency and interpretability are paramount [45].

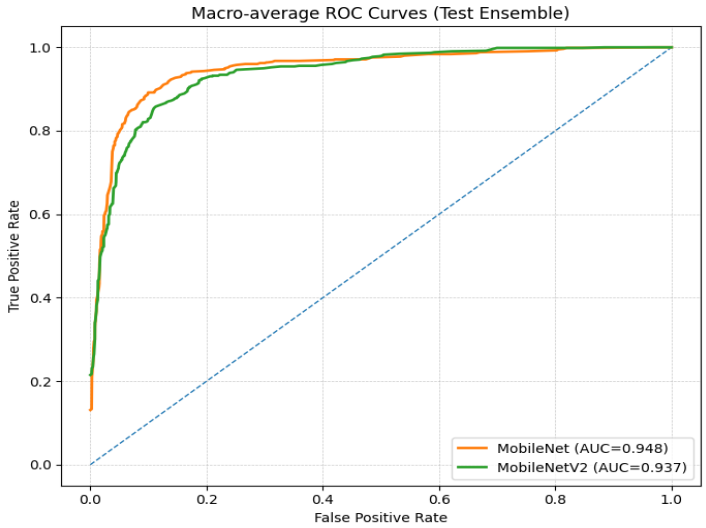


Fig. 6. ROC Curves Comparison of Pneumonia Detection using MobileNet and MobileNetV2

The performance comparison demonstrates that MobileNet and MobileNetV2 offer a favorable balance between classification performance and computational efficiency. MobileNet achieved an accuracy of 79.32% with a precision of 81.02%, recall of 78.15%, and an AUC-ROC of 94.80% (Table 2). Although these values are lower than those reported by larger architectures such as DenseNet201, ResNet50, and InceptionV3, they remain competitive when interpreted in relation to the substantially lower model complexity. Rather than outperforming deep architectures in absolute terms, MobileNet demonstrates that reasonable diagnostic performance can be achieved using a lightweight design. This distinction is important, as the goal of the proposed approach is not to replace high-capacity models but to provide a more deployable alternative under constrained computational conditions.

Table 2. Performance comparison with other architectures on the same dataset

Models	Accuracy (%)	Precision (%)	Recall (%)	AUC-ROC (Macro)	Parameter	Model Size (MB)
DenseNet121	77.88	78.32	76.83	93.89	8.1M	33
DenseNet169	80.45	79.82	79.21	94.78	14.3M	57
DenseNet201	77.88	78.33	77	94.3	20.2M	80
Xception	81.2	80.5	79.9	95.2	22.9M	88
VGG16	78.1	77.2	75.6	93.1	138M	528
VGG19	78.5	77.5	76	93.4	144M	549
ResNet50	80.1	79.8	78.5	94.6	25.6M	98
InceptionV3	81	80.9	79.8	95.1	23.8M	92
EfficientNetB0	80.8	80.1	79.3	94.8	5.3M	20
EfficientNetB4	**82.4**	**81.7**	**80.8**	**95.7**	**19M**	**77**
MobileNet	79.32	81.02	78.15	94.80	4.3M	16
MobileNetV2	76.44	77.7	75.53	93.70	3.5M	14

MobileNet requires only 4.3 million parameters and approximately 16 MB of storage, which is considerably smaller than DenseNet201 and ResNet50. MobileNetV2 further reduces the computational footprint to 3.5 million parameters and 14 MB, albeit with a modest reduction in accuracy and AUC-ROC. This trade-off highlights that performance gains obtained by larger models come with substantial increases in memory and computational demands, which may limit their applicability in real-world clinical environments. Notably, architectures such as VGG16 and VGG19, despite their very large parameter counts and storage requirements, did not consistently outperform MobileNet. These findings indicate that architectural efficiency rather than sheer model size plays an important role in achieving practical performance for pneumonia classification.

The Grad-CAM visualizations shown in Fig. 7 and Fig. 8 provide qualitative insight into how MobileNet and MobileNetV2 attend to chest X-ray regions during classification. In both models, activation maps are primarily concentrated within lung fields, often highlighting regions associated with consolidation and infiltration that correspond to known radiographic features of pneumonia. This alignment indicates that the models rely on anatomically and clinically relevant regions rather than background artifacts. MobileNet frequently produces more localized activation patterns, whereas MobileNetV2 exhibits broader and more distributed attention across lung regions. These observations are descriptive in nature and are consistent with explainability behaviors reported in prior studies on lightweight and residual-based architectures [46].

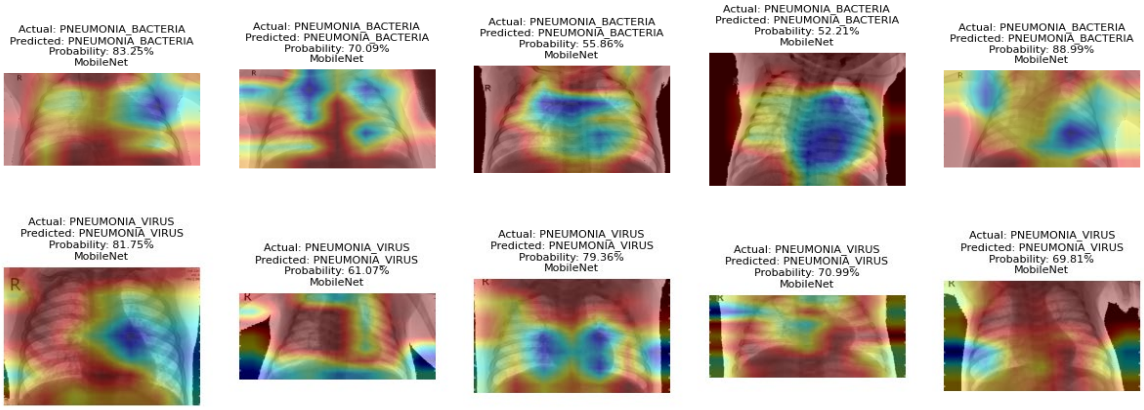


Fig. 7. Results of Grad-CAM Based Pneumonia Detection Testing using the MobileNet model

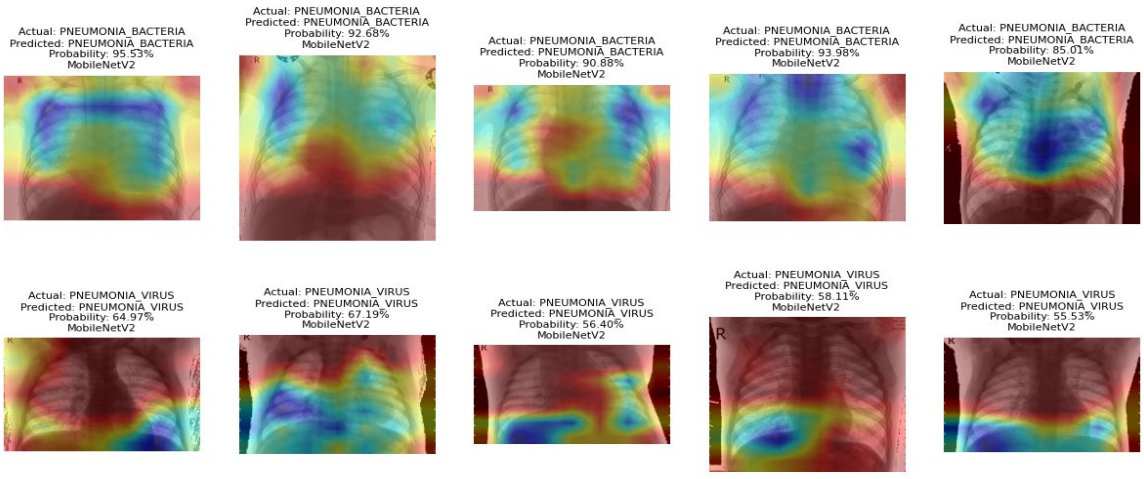


Fig. 8. Results of Grad-CAM Based Pneumonia Detection Testing using the MobileNetV2 model

However, the interpretation of these attention patterns remains qualitative. The present study does not include quantitative validation of heatmap quality, such as overlap analysis with expert-annotated segmentation masks or formal radiologist scoring, and therefore no definitive claims are made regarding

class-specific diagnostic superiority. Instead, the Grad-CAM results are presented to demonstrate transparency and plausibility of model decision-making. The observed correspondence between high-confidence predictions and activation over clinically meaningful lung regions supports the interpretability of both models and aligns with prior work emphasizing the role of Grad-CAM in chest radiograph analysis [47]. At the same time, occasional activation outside lung regions highlights the limitations of gradient-based explainability methods and reinforces the need for cautious clinical interpretation. Overall, these findings suggest that while MobileNet and MobileNetV2 do not outperform larger architectures in absolute performance metrics, they offer a compelling efficiency-interpretability trade-off. By framing Grad-CAM as supportive qualitative evidence rather than definitive diagnostic validation, this study avoids overstating its claims while reinforcing the value of lightweight and explainable models for screening and decision-support applications.

4. Conclusion

This study addressed the challenge of distinguishing bacterial and viral pneumonia in pediatric chest X-ray images by employing lightweight convolutional neural networks combined with explainability techniques. The experimental results showed that MobileNet consistently outperformed MobileNetV2 across most evaluation metrics, achieving an accuracy of 79.32%, precision of 81.02%, recall of 78.15%, F1-score of 77.82%, specificity of 89.49%, and strong AUC-ROC values of 94.80% (macro) and 90.52% (micro). MobileNetV2 demonstrated slightly lower performance, with an accuracy of 76.44%, F1-score of 74.45%, specificity of 88.45%, and AUC-ROC values of 93.70% (macro) and 88.78% (micro). These results indicate that MobileNet offers a more favorable balance between classification performance and architectural simplicity within the scope of this study. In addition to quantitative evaluation, Grad-CAM visualizations provided qualitative insight into model behavior by highlighting lung regions associated with pneumonia-related abnormalities. Both MobileNet and MobileNetV2 predominantly focused on clinically relevant lung areas, supporting the transparency and plausibility of their predictions. While differences in activation patterns between the two architectures were observed, these findings are intended to illustrate model interpretability rather than to assert definitive class-specific diagnostic advantages. The visual explanations reinforce the role of explainable AI as a complementary tool for increasing clinician trust rather than as a substitute for clinical judgment. Overall, this study demonstrates that MobileNet combined with Grad-CAM constitutes an efficient and interpretable framework for pediatric pneumonia classification. Although its performance does not exceed that of larger and more complex architectures, the trade-off between accuracy, explainability, and computational efficiency makes MobileNet a suitable candidate for assistive screening and decision-support applications, particularly in resource-limited clinical settings. Future work should focus on validation using larger and more diverse multi-center datasets, quantitative assessment of explainability through expert annotation, and evaluation in real-world deployment scenarios to further establish clinical relevance and robustness.

Acknowledgment

The authors would like to acknowledge the Department of Medical Technology, Institut Teknologi Sepuluh Nopember, for the facilities and support in this research. The authors also gratefully acknowledge financial support from the Institut Teknologi Sepuluh Nopember for this work, under project scheme of the Publication Writing and IPR Incentive Program (PPHKI) 2025.

Declarations

Author contribution. All authors contributed equally to the main contributor to this paper. All authors read and approved the final paper

Funding statement. This research was funded by the Research Department of the Institut Teknologi Sepuluh Nopember (ITS) 2025, grant number 2366/PKS/ITS/2025

Conflict of interest. The authors declare no conflict of interest.

Additional information. No additional information is available for this paper.

Data Availability Statements

The data that support the findings of this study are openly available in Kaggle Repository at https://www.kaggle.com/datasets/paultimothymooney/chest-xray-pneumonia

References

[1] Y. Ma, S. Fan, and J. Xi, "Recent updates regarding the management and treatment of pneumonia in pediatric patients: a comprehensive review," Infection, vol. 53, no. 6, pp. 2341–2359, Dec. 2025, doi: 10.1007/s15010-025-02605-w.

[2] M. Zhu et al., "Antibiotic de-escalation patterns and outcomes in critically ill patients with suspected pneumonia as informed by bronchoalveolar lavage results," Eur. J. Clin. Microbiol. Infect. Dis., vol. 44, no. 8, pp. 1861–1871, Aug. 2025, doi: 10.1007/s10096-025-05144-2.

[3] K. R. Balmuri, S. Konda, K. kumar Mamidala, M. Gunda, and S. Rani B, "Automated and reliable detection of multi-diseases on chest X-ray images using optimized ensemble transfer learning," Expert Syst. Appl., vol. 246, p. 122810, Jul. 2024, doi: 10.1016/j.eswa.2023.122810.

[4] Y. Akhter, R. Singh, and M. Vatsa, "AI-based radiodiagnosis using chest X-rays: A review," Front. Big Data, vol. 6, pp. 1–27, Apr. 2023, doi: 10.3389/fdata.2023.1120989.

[5] R. R. Watkins, "Using Precision Medicine for the Diagnosis and Treatment of Viral Pneumonia," Adv. Ther., vol. 39, no. 7, pp. 3061–3071, Jul. 2022, doi: 10.1007/s12325-022-02180-8.

[6] J. Rabbah, M. Ridouani, and L. Hassouni, "Improving pneumonia diagnosis with high-accuracy CNN-Based chest X-ray image classification and integrated gradient," Biomed. Signal Process. Control, vol. 101, p. 107239, Mar. 2025, doi: 10.1016/j.bspc.2024.107239.

[7] S. Singh, M. Kumar, A. Kumar, B. K. Verma, and S. Shitharth, "Pneumonia detection with QCSA network on chest X-ray," Sci. Rep., vol. 13, no. 1, p. 9025, Jun. 2023, doi: 10.1038/s41598-023-35922-x.

[8] A. Gupta, P. Sheth, and P. Xie, "Neural architecture search for pneumonia diagnosis from chest X-rays," Sci. Rep., vol. 12, no. 1, p. 11309, Jul. 2022, doi: 10.1038/s41598-022-15341-0.

[9] S. Singh, M. Kumar, A. Kumar, B. K. Verma, K. Abhishek, and S. Selvarajan, "Efficient pneumonia detection using Vision Transformers on chest X-rays," Sci. Rep., vol. 14, no. 1, p. 2487, Jan. 2024, doi: 10.1038/s41598-024-52703-2.

[10] D. J. Alapat, M. V. Menon, and S. Ashok, "A Review on Detection of Pneumonia in Chest X-ray Images Using Neural Networks," J. Biomed. Phys. Eng., vol. 12, no. 6, pp. 551–558, Dec. 2022, doi: 10.31661/jbpe.v0i0.2202-1461.

[11] A. Shavkatovich Buriboev, A. Abduvaitov, and H. S. Jeon, "Binary Classification of Pneumonia in Chest X-Ray Images Using Modified Contrast-Limited Adaptive Histogram Equalization Algorithm," Sensors, vol. 25, no. 13, p. 3976, Jun. 2025, doi: 10.3390/s25133976.

[12] M. Mujahid, F. Rustam, R. Álvarez, J. Luis Vidal Mazón, I. de la T. Díez, and I. Ashraf, "Pneumonia Classification from X-ray Images with Inception-V3 and Convolutional Neural Network," Diagnostics, vol. 12, no. 5, p. 1280, May 2022, doi: 10.3390/diagnostics12051280.

[13] K. Vijayakumar, M. N. H. Maziz, S. Ramadasan, S. Kadry, and S. Arunmozhi, "Classification of chest radiographs into healthy/pneumonia using Harris-Hawks Algorithm optimized deep-features," Discov. Comput., vol. 28, no. 1, p. 115, Jun. 2025, doi: 10.1007/s10791-025-09629-8.

[14] R. Ochoa-Ornelas, A. Gudiño-Ochoa, J. A. García-Rodríguez, and S. Uribe-Toscano, "Enhancing early lung cancer detection with MobileNet: A comprehensive transfer learning approach," Franklin Open, vol. 10, p. 100222, Mar. 2025, doi: 10.1016/j.fraope.2025.100222.

[15] A. Jovic, N. Frid, K. Brkic, and M. Cifrek, "Interpretability and accuracy of machine learning algorithms for biomedical time series analysis – a scoping review," Biomed. Signal Process. Control, vol. 110, p. 108153, Dec. 2025, doi: 10.1016/j.bspc.2025.108153.

[16] M. S. Vani, R. V. Sudhakar, A. Mahendar, S. Ledalla, M. Radha, and M. Sunitha, "Personalized health monitoring using explainable AI: bridging trust in predictive healthcare," Sci. Rep., vol. 15, no. 1, p. 31892, Aug. 2025, doi: 10.1038/s41598-025-15867-z.

[17] B. Priego-Torres, D. Sanchez-Morillo, E. Khalili, M. Á. Conde-Sánchez, A. García-Gámez, and A. León-Jiménez, "Automated engineered-stone silicosis screening and staging using Deep Learning with X-rays," Comput. Biol. Med., vol. 191, p. 110153, Jun. 2025, doi: 10.1016/j.compbiomed.2025.110153.

[18] P. Dusza, T. Banzato, S. Burti, M. Bendazzoli, H. Müller, and M. Wodzinski, "Comparative evaluation of CAM methods for enhancing explainability in veterinary radiography," Sci. Rep., vol. 15, no. 1, p. 29690, Aug. 2025, doi: 10.1038/s41598-025-14060-6.

[19] P. Radočaj and G. Martinović, "Interpretable Deep Learning for Pediatric Pneumonia Diagnosis Through Multi-Phase Feature Learning and Activation Patterns," Electronics, vol. 14, no. 9, p. 1899, May 2025, doi: 10.3390/electronics14091899.

[20] G. A. P. Rodrigues, A. L. M. Serrano, G. D. Bispo, G. P. R. Filho, V. P. Gonçalves, and R. I. Meneguette, "IHRAS: Automated Medical Report Generation from Chest X-Rays via Classification, Segmentation, and LLMs," Bioengineering, vol. 12, no. 8, p. 795, Jul. 2025, doi: 10.3390/bioengineering12080795.

[21] D. Kermany, K. Zhang, and M. Goldbaum, "Labeled Optical Coherence Tomography (OCT) and Chest X-Ray Images for Classification," Mendeley Data, 2018. [Online]. Available at: https://data.mendeley.com/datasets/rscbjbr9sj/2.

[22] A. Shah and M. Shah, "Advancement of deep learning in pneumonia/Covid-19 classification and localization: A systematic review with qualitative and quantitative analysis," Chronic Dis. Transl. Med., vol. 8, no. 3, pp. 154–171, Sep. 2022, doi: 10.1002/cdt3.17.

[23] E. Yanar, F. Hardalaç, and K. Ayturan, "PELM: A Deep Learning Model for Early Detection of Pneumonia in Chest Radiography," Appl. Sci., vol. 15, no. 12, p. 6487, Jun. 2025, doi: 10.3390/app15126487.

[24] B. E. Dedeoglu, A. R. Tanner, N. J. Brendish, H. E. Moyses, and T. W. Clark, "Comparison of two rapid host-response tests for distinguishing bacterial and viral infection in adults with acute respiratory infection," J. Infect., vol. 89, no. 6, p. 106360, Dec. 2024, doi: 10.1016/j.jinf.2024.106360.

[25] D. Rickard, M. A. Kabir, and N. Homaira, "Machine learning-based approaches for distinguishing viral and bacterial pneumonia in paediatrics: A scoping review," Comput. Methods Programs Biomed., vol. 268, p. 108802, Aug. 2025, doi: 10.1016/j.cmpb.2025.108802.

[26] S. A. El-Ghany, M. Elmogy, M. A. Mahmood, and A. A. Abd El-Aziz, "A Robust Tuberculosis Diagnosis Using Chest X-Rays Based on a Hybrid Vision Transformer and Principal Component Analysis," Diagnostics, vol. 14, no. 23, p. 2736, Dec. 2024, doi: 10.3390/diagnostics14232736.

[27] M. Z. Hoque, A. Keskinarkaus, P. Nyberg, and T. Seppänen, "Stain normalization methods for histopathology image analysis: A comprehensive review and experimental comparison," Inf. Fusion, vol. 102, p. 101997, Feb. 2024, doi: 10.1016/j.inffus.2023.101997.

[28] T. Islam, M. S. Hafiz, J. R. Jim, M. M. Kabir, and M. F. Mridha, "A systematic review of deep learning data augmentation in medical imaging: Recent advances and future research directions," Healthc. Anal., vol. 5, p. 100340, Jun. 2024, doi: 10.1016/j.health.2024.100340.

[29] P. R, G. Gajendran, S. Boulaaras, and S. S. Tantawy, "PediaPulmoDx: Harnessing cutting edge preprocessing and explainable AI for pediatric chest X-ray classification with DenseNet121," Results Eng., vol. 25, p. 104320, Mar. 2025, doi: 10.1016/j.rineng.2025.104320.

[30] S. C. Pereira, J. Rocha, A. Campilho, and A. M. Mendonça, "Distribution-based detection of radiographic changes in pneumonia patterns: A COVID-19 case study," Heliyon, vol. 10, no. 16, p. e35677, Aug. 2024, doi: 10.1016/j.heliyon.2024.e35677.

[31] H. Thunold, M. Riegler, A. Yazidi, and H. Hammer, "A Deep Diagnostic Framework Using Explainable Artificial Intelligence and Clustering," Diagnostics, vol. 13, no. 22, p. 3413, Nov. 2023, doi: 10.3390/diagnostics13223413.

[32] M. A. Khan, H. Park, and J. Chae, "A Lightweight Convolutional Neural Network (CNN) Architecture for Traffic Sign Recognition in Urban Road Networks," Electronics, vol. 12, no. 8, p. 1802, Apr. 2023, doi: 10.3390/electronics12081802.

[33] D. Ngo, H.-C. Park, and B. Kang, "Edge Intelligence: A Review of Deep Neural Network Inference in Resource-Limited Environments," Electronics, vol. 14, no. 12, p. 2495, Jun. 2025, doi: 10.3390/electronics14122495.

[34] M. Li, Y. Jiang, Y. Zhang, and H. Zhu, "Medical image analysis using deep learning algorithms," Front. Public Heal., vol. 11, no. November, pp. 1–28, 2023, doi: 10.3389/fpubh.2023.1273253.

[35] T. Öznacar and N. Varol Kayapunar, "Advanced skin cancer prediction with medical image data using MobileNetV2 deep learning and optimized techniques," Sci. Rep., vol. 15, no. 1, p. 28962, Aug. 2025, doi: 10.1038/s41598-025-14963-4.

[36] J. Lee, J. Park, and Y. Lee, "Toward Efficient Cancer Detection on Mobile Devices," IEEE Access, vol. 13, pp. 34613–34626, 2025, doi: 10.1109/ACCESS.2025.3543838.

[37] M. Ennab and H. Mcheick, "Advancing AI Interpretability in Medical Imaging: A Comparative Analysis of Pixel-Level Interpretability and Grad-CAM Models," Mach. Learn. Knowl. Extr., vol. 7, no. 1, p. 12, Feb. 2025, doi: 10.3390/make7010012.

[38] D. Muhammad and M. Bendechache, "Unveiling the black box: A systematic review of Explainable Artificial Intelligence in medical image analysis," Comput. Struct. Biotechnol. J., vol. 24, pp. 542–560, Dec. 2024, doi: 10.1016/j.csbj.2024.08.005.

[39] J. Manoharan and Y. Sivagnanam, "A Novel Human Action Recognition Model by Grad-CAM Visualization with Multi-level Feature Extraction Using Global Average Pooling with Sequence Modeling by Bidirectional Gated Recurrent Units," Int. J. Comput. Intell. Syst., vol. 18, no. 1, p. 118, May 2025, doi: 10.1007/s44196-025-00848-x.

[40] M. R. Hasan, S. M. Azmat Ullah, and S. M. Rabiul Islam, "Recent advancement of deep learning techniques for pneumonia prediction from chest X-ray image," Med. Reports, vol. 7, p. 100106, Oct. 2024, doi: 10.1016/j.hmedic.2024.100106.

[41] E. Hildt, "What Is the Role of Explainability in Medical Artificial Intelligence? A Case-Based Approach," Bioengineering, vol. 12, no. 4, p. 375, Apr. 2025, doi: 10.3390/bioengineering12040375.

[42] T. Huang, X. Huang, and H. Yin, "Deep learning methods for improving the accuracy and efficiency of pathological image analysis," Sci. Prog., vol. 108, no. 1, pp. 1–34, Jan. 2025, doi: 10.1177/00368504241306830.

[43] S. Kundu, S. Mukhopadhyay, R. Talukdar, D. Kaplun, A. Voznesensky, and R. Sarkar, "Deep learning model for grading carcinoma with Gini-based feature selection and linear production-inspired feature fusion," Sci. Rep., vol. 15, no. 1, p. 21225, Jul. 2025, doi: 10.1038/s41598-025-00217-w.

[44] E. Mahamud, N. Fahad, M. Assaduzzaman, S. M. Zain, K. O. M. Goh, and M. K. Morol, "An explainable artificial intelligence model for multiple lung diseases classification from chest X-ray images using fine-tuned transfer learning," Decis. Anal. J., vol. 12, p. 100499, Sep. 2024, doi: 10.1016/j.dajour.2024.100499.

[45] R. Ochoa-Ornelas, A. Gudiño-Ochoa, J. A. García-Rodríguez, and S. Uribe-Toscano, "A robust transfer learning approach with histopathological images for lung and colon cancer detection using EfficientNetB3," Healthc. Anal., vol. 7, p. 100391, Jun. 2025, doi: 10.1016/j.health.2025.100391.

[46] M. Aasem and M. Javed Iqbal, "Toward explainable AI in radiology: Ensemble-CAM for effective thoracic disease localization in chest X-ray images using weak supervised learning," Front. Big Data, vol. 7, pp. 1–14, May 2024, doi: 10.3389/fdata.2024.1366415.

[47] K. D. Reddy and A. Patil, "CXR-MultiTaskNet a unified deep learning framework for joint disease localization and classification in chest radiographs," Sci. Rep., vol. 15, no. 1, p. 32022, Aug. 2025, doi: 10.1038/s41598-025-16669-z.