

Automated Pneumonia Detection in Healthcare Using Efficient Deep Learning

Sadia Tariq¹, Malik Asim Rajwana¹, Muhammad Shoaib Rasheed¹, Tahir Abbas^{2*}, Nazir Ahmad¹, Zaiba Aziz¹, and Ehsan ul Haq¹

¹Department of Computer Science, National College of Business Administration & Economics Lahore, Multan Sub Campus, 60000, Pakistan.

²Department of Computer Science, Muhammad Nawaz Shareef University of Engineering & Technology, Multan, 60000, Pakistan.
Corresponding Author: Tahir Abbas. Email: tahir.abbas@bzu.edu.pk

Received: May 30, 2026 Accepted: September 03, 2026

Abstract: Accurate and timely pneumonia detection by using chest radiographs plays an important role in reducing death rates and also supports clinical decision-making. In recent years, medical image analysis has advanced considerably and deep learning techniques have made a significant contribution. However, finding a model that provides both high accuracy and low computational cost is still difficult. MobileNetV2 is adopted as the core architecture for binary classification of normal and pneumonia cases. This proposed approach is compared with the traditional Convolutional neural network and the well-known VGG16 model. To keep the evaluation consistent all models are trained and validated under the same settings. Metrics used for model performance evaluation are Accuracy, F1-score, confusion matrix and AUROC. Experimental results of MobileNetV2 are accuracy of 93.75%, F1-score of 93.33% and AUROC of 1.00, indicating that the proposed model outperforms VGG16 having 81.25% accuracy, 84.21% F1-score and an AUROC of 98.44%, and baseline CNN having 75.00% accuracy, 80% F1-score and an AUROC of 98.44%. MobileNetV2 offers an adequate compromise between the performance of classification and computational efficiency. The results confirm that lightweight deep learning architectures can deliver high diagnostic performance for pneumonia detection and may serve as practical solutions for real-world and resource-constrained clinical deployment.

Keywords: Pneumonia Detection; Chest X-ray Imaging/Chest Radiographs; Automated Deep Learning; Medical Image Classification and Computational Efficiency

1. Introduction

Worldwide, pneumonia is among the most frequent and severe respiratory tract infections that is mostly common in children, older adults and those having weak immune defenses [1]. According to global health statistics millions of pneumonia cases are reported annually which are leading to a substantial number of hospitalized patients and deaths [2, 3]. But early and accurate diagnosis plays an important role in improving patient recovery [4, 5]. Traditional diagnostic procedures take more time and depend on expert interpretation of chest radiographs [6, 7]. The limited availability of experienced radiologists in many healthcare settings especially in developing regions can be a reason for delays in timely diagnosis [8-10]. Chest radiographs imaging remains the most commonly utilized technique for the initial examinations of pneumonia due to its affordability, accessibility and rapid image acquisition [11, 12]. Even with these benefits, manual examination of chest X-rays can lead to differences of opinion among doctors and mistakes due to fatigue [13, 14]. The radiographic patterns linked with pneumonia are mild or indistinct and can be difficult to differentiate from normal anatomical changes or other lung diseases that can cause misdiagnosis [15-17]. The weaknesses of manual diagnosis have inspired the creation of automated computer-aided diagnostic systems (CAD) that offer clinicians reliable and standardized measures [18, 19]. Recent progress in deep learning especially CNNs has shown exceptional results in

medical image analysis [20-22]. CNN-based models are trained on raw inputs, thus eliminating the need to perform manual feature engineering to a great extent [23, 24]. Transfer learning can be used to use pretrained networks on medical imaging tasks with minimal data [25, 26]. This method not only reduce the training time but also enhances generalization [27, 28].

Architectures such as VGG, DenseNet and MobileNet can be used to detect pneumonia in the chest X-ray [29] but due to their size and high computational intensity they cannot be used in resource-constrained settings and when real-time performance is required [30-32]. MobileNetV2 along with other lightweight architectures is an attractive alternative since it offers a good balance between accuracy and efficiency [33-35]. This paper will introduce a MobileNetV2 a transfer learning model to identify pneumonia in the chest radiographs. The proposed model is trained and tested with a standard CNN and a VGG16-based transfer learning model so that the performance can be evaluated comprehensively [36, 37]. The results of the experiment indicate that the proposed MobileNetV2 model has an accuracy of 93.75, F1-score of 93.3% and area under the receiver operating characteristic curve (AUROC) of 1.00 a higher level of performance compared to the baseline CNN and VGG16 models without sacrificing computational efficiency.

The main findings of the current research are as follows:

1. A framework based on MobileNetV2 implemented using transfer learning to detect pneumonia efficiently.
2. A systematic comparison of the proposed framework with a traditional CNN model and a VGG16-based transfer learning model was tested on various measures of evaluation.
3. Experiments with detailed analysis prove the high accuracy, robustness and effective classification of the proposed method.

The remainder of the paper is organized in the following way: Section II will review the literature on the topic of pneumonia detection with the help of deep learning methods. Section III describes the proposed methodology and model architecture. Section IV describes the experimental results and performance evaluation while Section V provides a detailed discussion of the findings and Section VI concludes the paper and outlines directions for future research.

2. Related Work

[38] used a public dataset of 7,700 X-rays. Their model Improved BoxENet with transfer learning and achieved 98.6% accuracy in multiclass classification. A stated limitation is the high computational demand of comparable models.

[39] evaluated five CNN models trained on chest radiographs images for pneumonia detection. This study proposes a hybrid model that integrates VGG16 and DenseNet121 achieved the highest accuracy of 92.53% and outperformed VGG16 alone had 90.87% accuracy. The model validation on a single dataset may limit its generalizability.

[40] developed a pneumonia prediction model using a Deep CNN and InceptionV3 for feature extraction on chest radiographs by integrating an entropy-normalized feature reduction technique with an ensemble classifier.

[41] used an augmented Kaggle dataset of 18,200 chest radiographs. The proposed model achieved 98% accuracy by getting fused features from three lightweight CNNs and using XGBoost for classification. There exists a possible dependence upon data augmentation techniques that may limit the generalizability. VGG16 with data augmentation and class weighting achieved the highest accuracy of 91%. A major limitation is the dependence upon general ImageNet pre-training which may fail to capture features that are particular to pneumonia.

[42] developed LungXpert a VGG16-based model that attained 97.22% accuracy in detecting pneumonia from chest X-ray images. Although the results are notably strong, further validation on diverse datasets is still required.

[43] PELM an ensemble model that combines Inception V3, ResNet50 and a Vision Transformer. The model was trained on a collection of 50,000 chest radiographs images obtained from several public datasets. It achieved 96% accuracy, 99% precision, 91% recall, 95% F1-score, 91% specificity and an AUC of 0.91. However, clinical validation across a variety of healthcare environments is still needed.

[44] employed the ResNet50 model together with LRP-based interpretability for the analysis of chest X-ray images and achieving 91% accuracy and 0.85 interpretability score while other techniques showed performance trade-offs and may require additional validation.

[45] developed a custom CNN together with a pre-trained ResNet-50 model using a collection of 26,684 chest X-ray images obtained from the RSNA dataset. The custom CNN trained without the use of data augmentation and achieved an accuracy of 79 percent, along with a precision of 0.76, a recall of 0.73 and an F1-score of 0.74. However, a moderate level of overall accuracy indicates that additional optimization is still required.

[46] carried out a comparison of six convolutional neural network models using a collection of 6,939 chest X-ray images. Among these models, VGGNet achieved the highest performance with an accuracy of 97 percent. while the presence of dataset-specific biases and the continuing need for broader validation across more diverse data sources persist.

3. Materials and Methods

Pneumonia refers to an infection that affects the lungs and may result from bacterial, viral or environmental influences. The condition can lead to fluid accumulation in the lungs, making breathing difficult and is one of the major causes of death globally.

3.1. Dataset and Preprocessing

This study uses an open-access chest radiograph Pneumonia Dataset from the Kaggle repository [47] and originally released by Paul Mooney [48]. The data set consisted of 5,856 chest X-rays of which 1,583 were Normal and 4,273 were Pneumonia. The data have been split into a training set, a validation set and a testing set. There were 5216 images (1341 Normal and 3875 image Pneumonia) in the training set and 16 images (8 Normal and 8 Pneumonia) in the validation set. The test set was an independent set of 624 images, of which there were 234 Normal images and 390 Pneumonia images. A summary of the distribution of the data set is provided in Table 1.

The input size is maintained and compatible with deep convolutional neural network models by resizing all images in these groups to 160×160 pixels [49]. The range of pixel intensities to the [0,1] range is used in order to achieve a quicker and more stable training. Data augmentation is used to enhance the capability of the model to avoid overfitting. Random rotations, horizontal flips, zoom and minor geometric manipulations are all data augmentation techniques . This approach will make the dataset quite diverse and enable the model to identify meaningful aspects even with different image conditions. Figure 1 illustrates the dataset having two classes: normal (a, b) and pneumonia (c, d) chest images.

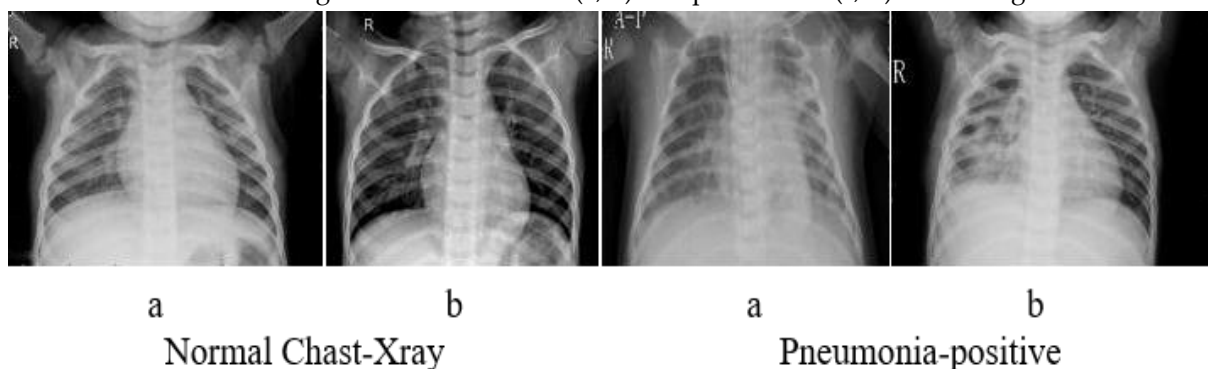


Figure 1. dataset having two classes: normal (a, b) and pneumonia (a, b) chest image.

Table 1. Details of classes, subfolders and the number of images in the Dataset

Dataset split	Normal	Pneumonia	Total
Training	1,341	3,875	5,216
Validation	8	8	16
Testing	234	390	624
Total	1,583	4,273	5,856

3.2. Data Training and Validation

After preprocessing, the dataset is organised into training and validation folders so that the proportions of each class remain unchanged. The proposed framework uses a light deep transfer-learning arrangement based on MobileNetV2 initialized with ImageNet pre-trained weights [50, 51]. A custom classification layer is introduced in place of the original head to perform binary pneumonia detection. At the same time two reference models VGG16 [52] and a traditional convolutional neural network (CNN) [53] are also employed so that a comparison of their performance may be carried out. Although the baseline models serve as comparative references but the proposed MobileNetV2 model is fully trained and optimized so that it may attain a higher level of performance. During training the parameters of the model are refined with the aid of the Adam optimization method together with a binary cross-entropy loss function. The progress of learning is observed by means of validation accuracy and loss curves to ensure stable convergence. Once the training stage is completed several performance measures evaluated namely accuracy, precision, recall, F1-score and AUROC to ensure a comprehensive assessment [54]. In addition to these measures confusion matrices and ROC curves are examined in order to evaluate performance for each class and to observe the model capacity for discrimination.

3.3. Proposed Model

The proposed framework employs MobileNetV2 as the backbone network because it provides an effective balance between classification accuracy and computational efficiency. MobileNetV2 is a lightweight CNN that uses computationally efficient depthwise and pointwise convolutions combined with inverted residual structures and linear bottlenecks to extract meaningful features from images. This property is very useful for healthcare-related applications with restricted computational resources, like image-based applications. This selection was also performed based on the performance of this CNN model with the baseline CNN and VGG16 model used in this study.

3.3.1. Input Representation

Let $X \in \mathbb{R}^{H \times W \times 3}$ represents an input chest radiographic image resized to a spatial resolution of $H \times W$ while $H = W = 160$ and normalized to the range $[0,1]$. The resulting tensor is then processed by MobileNetV2 backbone for hierarchical feature extraction.

3.3.2. Feature Extraction Using MobileNetV2

MobileNetV2 is composed of a sequence of inverted residual blocks. Each block consists of:

1. Pointwise expansion convolution (1×1)
2. Depthwise convolution (3×3)
3. Pointwise projection convolution (1×1)

Let X_l represent the input feature map to the l^{th} block. The expansion phase first increases the channel dimensionality:

$$Z_l = \sigma(W_l^{(1)} * X_l) \quad (1)$$

where $W_l^{(1)}$ denotes the weights of the 1×1 convolution, $*$ represents convolution and $\sigma(\cdot)$ is the ReLU6 activation function.

The next step is to apply a depthwise convolution to each channel independently:

$$D_l = \sigma(W_l^{(2)} \odot Z_l) \quad (2)$$

where $\odot$ denotes depthwise convolution and $W_l^{(2)}$ represents the depthwise filter.

Finally, a linear projection reduces the channel dimensionality:

$$P_l = W_l^{(3)} * D_l \quad (3)$$

$l \quad l \quad l$

If the spatial and channel dimensions are preserved, an inverted residual connection is added:

$$Y_l = X_l + P_l \quad (4)$$

Otherwise,

$$Y_l = P_l \quad (5)$$

This design preserves representational power while minimizing information loss and computational cost.

3.3.3. Transfer Learning Strategy

The MobileNetV2 backbone is initialized using ImageNet pre-trained weights. Let θ_b , θ_c represent backbone parameters and newly added classifier parameters respectively. During the training phase selected deeper layers of the backbone are fine-tuned while earlier block of layers remains frozen:

$$\theta = \{\theta_b^{(trainable)}, \theta_c\} \quad (6)$$

This strategy allows the network to adapt generic visual features to pneumonia-specific patterns such as lung opacities and consolidations.

3.3.4. Global Average Pooling (GAP)

The final convolutional feature map $F \in \mathbb{R}^{h \times w \times k}$ is processed using a Global Average Pooling layer to generate a feature vector by averaging across the spatial dimensions.

$$g_k = \frac{1}{h \times w} \sum_{a=1}^h \sum_{b=1}^w F_{a,b,k} \quad (7)$$

for each channel k . GAP significantly reduces the number of parameters and enforces spatial invariance.

3.3.5. Fully Connected Layer

The pooled feature vector g is mapped to a dense representation:

$$z = \sigma(W_f g + b_f) \quad (8)$$

where $W_f \in \mathbb{R}^{128 \times k}$, b_f is the bias vector and $\sigma(\cdot)$ denotes the ReLU6 activation.

3.3.6. Dropout Regularization

To reduce overfitting, dropout is applied:

$$\tilde{z} = z \odot m \quad (9)$$

where $m \sim \text{Bernoulli}(p)$ and $p = 0.5$

3.3.7. Output Layer and Prediction

The final Dense layer produces a single neuron output:

$$o = W_o \tilde{z} + b_o \quad (10)$$

The probability of pneumonia is computed using the sigmoid activation:

$$\hat{y} = \frac{1}{1+e^{-o}} \text{ where } \hat{y} \in [0,1] \quad (11)$$

A threshold of 0.5 is used for binary classification:

$$\hat{y} = \begin{cases} 1, & \hat{y} \geq 0.5 \quad (\text{Pneumonia}) \\ 0, & \hat{y} < 0.5 \quad (\text{Normal}) \end{cases} \quad (12)$$

3.4. Training Strategy

The model is trained using a single stage fine-tuning strategy with the earlier layers of MobileNetV2 partially frozen to learn generic visual features while the deeper layers and newly added layers are trainable to learn the medical imaging domain. Binary cross-entropy is used as the loss function for two classes. Adam optimizer was used with a learning rate of 1×10^{-4} which helps stabilize convergence during the training. The proposed model is trained for 15 epochs with a fixed batch size of 32 [55]. During the training to prevent overfitting all validation measures are observed. This overall training setup provides an appropriate balance between learning capacity and computational efficiency. Figure 2 represents the architecture of the proposed model based on MobileNetV2. Input image pass through convolutional and depth-wise separable layers for feature extraction and is classified by fully connected layer.

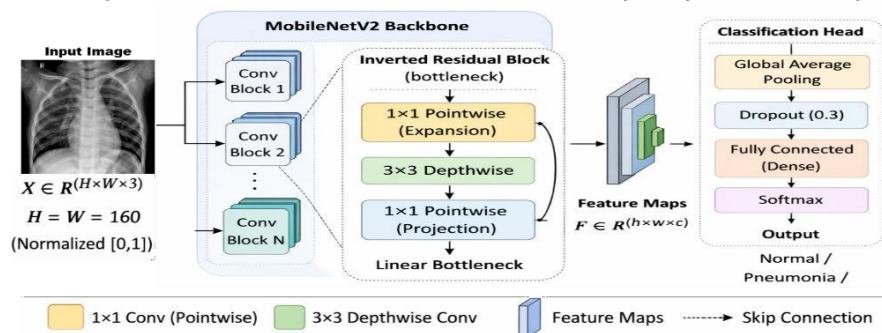


Figure 2. Proposed MobileNetV2-Based Model

3.5. Performance Evaluation

Quantitative assessment of the proposed model framework output is calculated using multiple metrics derived from the confusion matrix. The validation and test datasets are used for Performance evaluation and repeated across a series of independent trials to ensure that the proposed model is robust and has high statistical reliability. The confusion matrix for the r^{th} experimental run is represented as $O^{(r)}$ where $r = 1, 2, \dots, R$. The confusion matrix presents the Comparison of the ground-truth and predicted labels for the two classes Normal and Pneumonia. After preprocessing and partitioning the dataset the balanced class distribution within the validation set can be expressed as follows:

$$O_V(j) = \begin{bmatrix} \frac{|D^y|}{K} & 0 \\ 0 & \frac{|D^z|}{K} \end{bmatrix}, \quad \forall j \in \{1, 2, \dots, I\} \quad (13)$$

where $|D^y|$ represents total number of Normal images and $|D^z|$ represents the Pneumonia images while K represents the number of folds.

The predicted outcomes of the r^{th} run are summarized by the confusion matrix is represented as:

$$P^{(r)} = \begin{bmatrix} t_{+ive}^{(r)} & f_{+ive}^{(r)} \\ f_{-ive}^{(r)} & t_{-ive}^{(r)} \end{bmatrix}, \quad \forall r \in \{1, 2, \dots, R\} \quad (14)$$

where $t_{+ive}^{(r)}$, $f_{+ive}^{(r)}$, $f_{-ive}^{(r)}$ and $t_{-ive}^{(r)}$ represent true positive, false positive, false negative and true negative respectively.

To comprehensively assess the proposed model six standard performance metrics are computed for each run:

$$\{M_1^{(r)}, M_2^{(r)}, M_3^{(r)}, M_4^{(r)}, M_5^{(r)}, M_6^{(r)}\} \quad (15)$$

corresponding to accuracy, recall (sensitivity), specificity, precision, F1-score and Matthews Correlation Coefficient (MCC).

The accuracy is defined as:

$$S_{Acc}^{(r)} = \frac{t_{+ive}^{(r)} + t_{-ive}^{(r)}}{t_{+ive}^{(r)} + f_{+ive}^{(r)} + f_{-ive}^{(r)} + t_{-ive}^{(r)}} \quad (16)$$

The recall (sensitivity) is computed as:

$$S_{Rec}^{(r)} = \frac{t_{+ive}^{(r)}}{t_{+ive}^{(r)} + f_{-ive}^{(r)}} \quad (17)$$

The specificity measures the correct identification of normal cases and is expressed as:

$$S_{Spec}^{(r)} = \frac{t_{-ive}^{(r)}}{t_{-ive}^{(r)} + f_{+ive}^{(r)}} \quad (18)$$

Precision represented as:

$$S_{Prec}^{(r)} = \frac{t_{+ive}^{(r)}}{t_{+ive}^{(r)} + f_{+ive}^{(r)}} \quad (19)$$

The F1-score is calculated as:

$$S_{F1}^{(r)} = \frac{2 \times S_{Prec}^{(r)} \times S_{Rec}^{(r)}}{S_{Prec}^{(r)} + S_{Rec}^{(r)}} \quad (20)$$

The Matthews Correlation Coefficient (MCC) is formulated as:

$$S_{MCC}^{(r)} = \frac{(t_{+ive}^{(r)} - f_{+ive}^{(r)})(t_{-ive}^{(r)} - f_{-ive}^{(r)})}{\sqrt{(t_{+ive}^{(r)} + f_{+ive}^{(r)})(t_{-ive}^{(r)} + f_{-ive}^{(r)})(t_{+ive}^{(r)} + f_{-ive}^{(r)})(t_{-ive}^{(r)} + f_{+ive}^{(r)})}} \quad (21)$$

ROC curve is obtained by plotting the true positive rate (TPR) against the false positive rate (FPR):

$$TPR^{(r)} = \frac{t_{+ive}^{(r)}}{t_{+ive}^{(r)} + f_{-ive}^{(r)}} \quad (22)$$

$$FPR^{(r)} = \frac{f_{+ive}^{(r)}}{f_{+ive}^{(r)} + t_{-ive}^{(r)}} \quad (23)$$

So AUROC is calculated as:

$$AUROC = \int_0^1 TPR(FPR) d(FPR)$$

(24)

To enhance reproducibility and statistical reliability, all reported numerical results are averaged on R independent runs.

4. Results

This part will provide a quantitative and qualitative evaluation of the pneumonia detection system using the suggested model, MobileNetV2. The proposed model is compared to the Performance of two baseline models, a standard CNN and VGG16. The assessment is conducted on the validation dataset with Accuracy, F1-score and AUROC as the common metrics. Moreover, confusion matrices and ROC curves are examined to gain a deeper understanding of classification performance.

4.1. Quantitative Performance Comparison

Table 2 presents a summary of the performance obtained from all three models. The quantitative performance of the conventional CNN, VGG16 and the proposed MobileNetV2 model based on accuracy, F1-score, precision, recall and AUROC. The proposed MobileNetV2 was the best in all three-evaluation metrics. The CNN model obtained an accuracy of 0.7500, an F1-score of 0.8000, and an AUROC of 0.984. With the same AUROC as VGG16, it obtained a higher accuracy 0.8125 and a higher F1-score 0.8421. By contrast, the accuracy obtained by MobileNetV2 was 0.9375 the F1-score was 0.9333 and the AUROC was 1.000. Absolute improvement was obtained with MobileNetV2 obtaining an improvement of 18.75 percentage points compared with the traditional CNN and 12.50 percentage points compared with VGG16. Similarly, the F1-score of MobileNetV2 was better by 13.33 percentage points than CNN and 9.12 percentage points better than VGG16 model. Same as the Precision and recall of MobileNetV2 is better than CNN and VGG16 model. The improvements show that MobileNetV2 made more accurate class predictions and achieved a better precision/recall balance. The AUROC results also support the good discriminative power of the models evaluated. The AUROC value of both CNN and VGG16 was 0.984 and showed a good separation between pneumonia and normal class. MobileNetV2 yielded a perfect AUROC score of 1.000 on the validation subpopulation that was used to evaluate the model, corresponding to total separation within the classes. The AUROC improvement is smaller than the improvements on accuracy and F1-score, but it does show that MobileNetV2 maintained a very good discrimination while final classification results were more accurate. In general, the quantitative results show that the MobileNetV2 architecture outperforms the other two architectures tested. The accuracy, F1-score and AUROC values consistently improve as MobileNetV2 is trained for the chest X-ray pneumonia classification task and are evaluated for the test sets. The AUROC of 1.000, however, should be interpreted in the context of the current validation cohort and is not meant to be evidence of optimal generalization to independent clinical cohorts. The characteristics of the analyzed dataset, the small size of the test-set and relatively small validation set also restricts generalisation assessment.

Table 2. Evaluation of the models for performance and comparison.

Model	Precision	Recall	F1-score	Accuracy (%)	AUROC
CNN	0.8333	0.7500	0.7333	75.00	0.984
VGG16	0.8636	0.8125	0.8057	81.25	0.984
MobileNetV2 (Proposed)	0.9444	0.9375	0.9373	93.75	1.00

4.2. Performance in term of Accuracy

The accuracy of the three models assessed is shown in figure 3. The traditional CNN achieved an accuracy of 75.00%, the bottom accuracy in the classification accuracy of the architectures evaluated. The accuracy of the CNN model was improved by 6.25% by VGG16. The proposed MobileNetV2 gave the highest accuracy of 93.75% which is 18.75% higher than CNN and 12.50% higher than VGG16. The results show that the MobileNetV2 model achieved a significantly higher accuracy rate on the chest X-ray images that were evaluated compared to the accuracy rates of both of the comparison models. The higher accuracy indicates the usefulness of transfer-learning architecture for MobileNetV2 in the experimental conditions of this work for distinguishing pneumonia from normal cases. Thus, the accuracy results give quantitative evidence of the better classification performance of the proposed model.

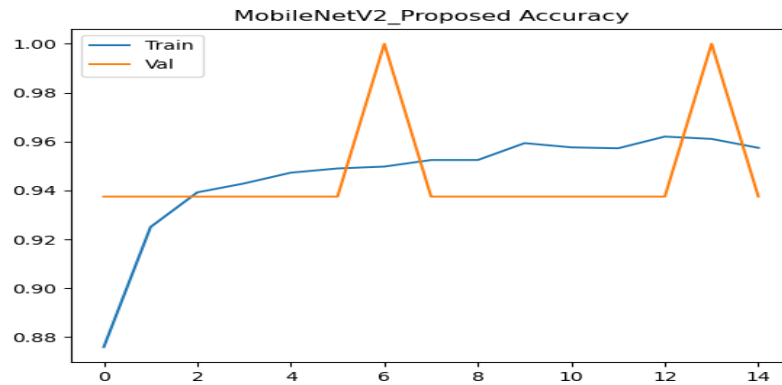


Figure 3. Accuracy comparison of the CNN, VGG16 and MobileNetV2 models.

4.3. Performance in terms of F1-Score

The results of the three models evaluated are compared based on the accuracy, F1-value and AUROC results in Figure 4. The conventional CNN obtained the F1 score of 0.8000 and the VGG16 obtained the F1 score 0.8421. The proposed MobileNetV2 delivered the best F1 score of 0.9333. It is an absolute improvement of 0.1333 (13.33 percentage points) when compared to CNN model and 0.0912 (9.12 percentage points) to the VGG16 model. The better F1 score for MobileNetV2, suggests that it outperformed the two baseline models in terms of precision and recall. This is especially important for the classification of pneumonia where a model should be able to classify pneumonia-positive patients accurately, and not classify normal patients as pneumonia. As the F1 score continues to improve consistently, along with its accuracy, it implies that the performance gain of MobileNetV2 was not only constrained to a single metric.

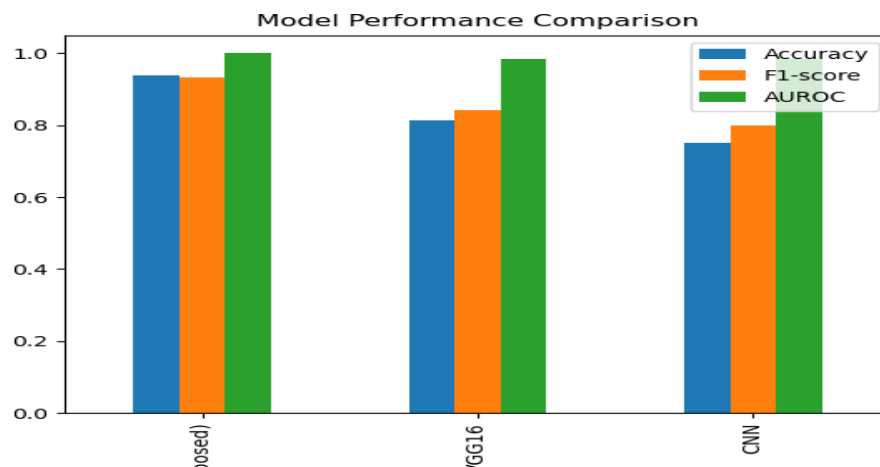


Figure 4. Comparison of accuracy, F1-score and AUROC for the evaluated models.

4.4. Performance in terms of Loss

The training and validation loss curve of the MobileNetV2 model are shown in Figure 5. Progressive improvement in the optimization process is reflected in the gradual decrease in the training loss from about 0.275 to 0.105. The loss for validation also decreased and achieved minimum value, around 0.115, and relatively small fluctuations in following training epochs. A reduction in the loss of both the training and validation sets shows that the model successfully learned discriminative representations from training set. The relatively close values of the training and validation losses at the end of training also indicate that the model may not have suffered a large gap between the training and validation loss. These observations are reflected in the high validation accuracy, F1-score, and AUROC as shown in Table 2. Hence, the loss curves are another evidence of successful convergence of the MobileNetV2 model with the training configuration used.

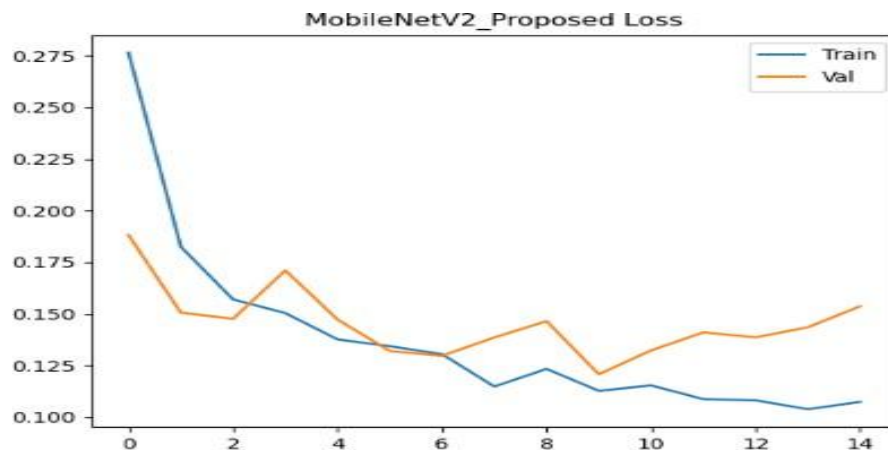
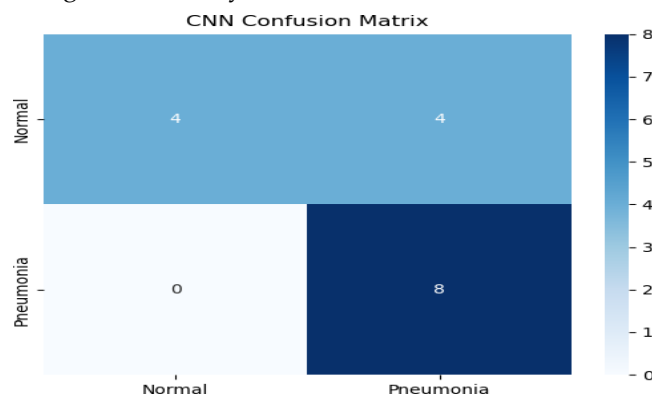


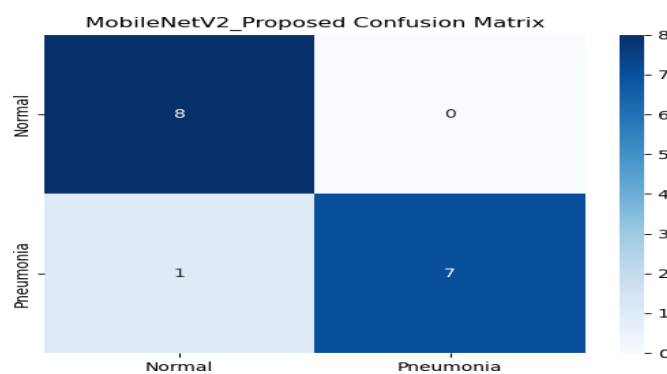
Figure 5. Training and validation loss curves of MobileNetV2.

4.5. Confusion Matrix Analysis

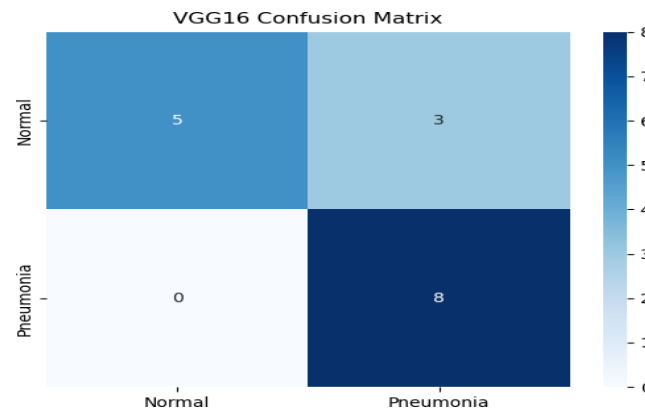
The confusion matrices shown in Figure 6 offer a class-wise evaluation of the predictions made by the CNN, VGG16, and MobileNetV2 models. The results demonstrate that the proposed MobileNetV2 resulted in fewer classification errors compared with the baseline architectures for the dataset used for validation. One in particular, the relatively low rate of false negative predictions, is significant for pneumonia detection because of the fact that pneumonia cases that were missed were misclassified as negative. The confusion matrix of the conventional CNN has a relatively large misclassification rate for normal and pneumonia cases. The VGG16 model shows an improvement over the CNN in terms of improving the class discrimination compared to the classification errors. MobileNetV2 has the highest accuracy and F1-score in Table 2, also yielding the best overall classification pattern among the three models. Thus, the results of the confusion matrix are qualitatively corroborating quantitative results. The lower level of misclassification found with MobileNetV2 suggests that the feature representations learned by this network were more useful at separation of radiographic patterns for pneumonia and normal cases when tested in the experimental settings of this study.



a: (CNN- confusion matrix)



b: (MobileNet-V2 Confusion matrix)

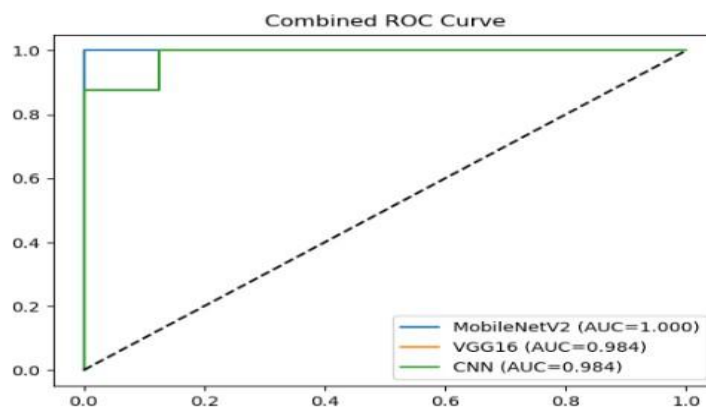


C: (VGG16- Confusion matrix)

Figure 6. Confusion matrices of (a) CNN, (b) VGG16 and (c) MobileNetV2.

4.6. ROC Curve Interpretation

The ROC curve of CNN model, VGG16 model and MobileNetV2 model are plotted in one graph, which is shown in figure 7. The MobileNetV2 ROC curve is the closest to the upper left corner of the ROC space with an AUROC of 1.000. This means that no differences exist in the pneumonia and normal classes in the evaluated validation set. Both the CNN and VGG16 models obtained an AUROC score of 0.984 and high discriminative capacity, as well. The AUROC difference between the two baseline models and MobileNetV2 is relatively small, but it matches the higher accuracy and F1-score of the MobileNetV2 model. In particular, it has excellent threshold-independent class discrimination and a larger rate of correct final class predictions, in the case of MobileNetV2. The performance observed may be due, in part, to efficient feature extraction characteristics of MobileNetV2 such as depth-wise separable convolutions and inverted residual structures. The AUROC of 1.000 should, however, be read with caution since it was calculated on the validation data set that was used for evaluation. It shows optimal separation between the evaluated samples but is not evidence of perfect diagnostic performance or extrapolation to independent samples or clinical populations.

**Figure 7.** Receiver operating characteristic (ROC) curves of the CNN, VGG16 and MobileNetV2 models.

5. Discussion

Timely and reliable detection of pneumonia in chest radiographic images is very important for clinical decisions. Still radiographic patterns analysis can be challenging due to overlapping features of normal anatomy with other pulmonary conditions. The proposed MobileNetV2-based framework addresses these challenges and outperforms conventional CNN and VGG16. Table 2 shows that the proposed model attains 93.75% accuracy, an F1-score of 0.9333 and an AUROC of 1.00 clearly showing that the proposed model outperforms the baseline models. Confusion matrix analysis shows that MobileNetV2 correctly identifies the majority of pneumonia-positive cases and a very low number of false negatives prevent missed diagnoses. Both the CNN and VGG16 base models resulted in higher misclassification rates. Table 2 shows

that VGG16 gave better performance than CNN but remained less effective than MobileNetV2. These improvements are mainly due to several factors:

- ❖ Transfer learning using a MobileNetV2 already pre-trained on ImageNet plays a key role in enhancing performance.
- ❖ The application of global average pooling helps limit overfitting by reducing model complexity.
- ❖ Data augmentation improves training diversity and supports better generalization.

From Figure 4, the ROC curve analysis shows an optimal sensitivity- specificity balance. The MobileNetV2 model has the potential to be a valuable assistant in screening for pneumonia, especially in low-resource areas. The results should not, however, be considered as proof of clinical readiness or as an alternative to the expert radiological diagnosis. Beyond validation in the small dataset, larger, more diverse, and multi-center datasets and prospective clinical validation is still necessary before deployment in the real-world. While The proposed model reduces parameter counts and lowers the overfitting risk which makes it suitable for clinical usage on limited-resource systems. The number of trainable and non-trainable parameters of the models evaluated were compared to further explore the complexity of the model. The proposed MobileNetV2 based model has 2,422,081 total parameters, out of which 164,097 are trainable and the remaining 2,257,984 are non-trainable parameters. In comparison, the custom CNN has 7,373,057 parameters, while VGG16 has 14,780,481 parameters. Therefore, the overall number of parameters of the proposed approach is significantly lower than both the comparison approaches. This parameter efficiency is obtained by leveraging the pre-trained MobileNetV2 backbone (with frozen weights) and a simple task-specific classification head. Overall, MobileNetV2 provides a solution that is more reliable and computationally efficient for automated pneumonia detection and outperforms traditional deep networks in accuracy, F1-score, AUROC and misclassification rates.

6. Conclusion and Future Work

This study presented an efficient deep learning framework MobileNetV2 for the automatic detection of pneumonia from chest radiographic images. The proposed model used transfer learning with a lightweight architectural design to attain high classification accuracy while maintaining computational complexity very low. To test the given approach and compare it with two reference models i.e. a standard CNN and VGG16, a set of extensive experiments was carried out. The experimental results indicate that the proposed MobileNetV2 model demonstrates significantly higher performance with an accuracy of 93.75 percent, F1-score of 0.9333 and the value of the AUROC of 1.00. These findings indicate a high classification and generalization of the proposed model. The suggested model performs better than CNN and VGG16 when it comes to classification error and demonstrates a reasonable balance between sensitivity and precision, which is a very important consideration in clinical decision-making. This near-optimal value of the AUROC is another strong reason that the proposed framework is more reliable for classifying pneumonia-positive and normal cases but at the same time it is also the limitation of this study is that splits of the dataset, pre-defined for the training, validation set and test sets were not rearranged any further. No independent verification was made of patient level separation, so the risk of patient level data leakage cannot be excluded, this being also a limitation of this study. Future work may focus on multi-institutional datasets with improved preprocessing strategies. Furthermore, explainable AI can be integrated to enhance generalization and interpretability in clinical use.

References

1. Kang, N., Subramanian, V.S., and Agrawal, A.: 'Influence of aging and immune alterations on susceptibility to pneumococcal pneumonia in the elderly', *Pathogens*, 2025, 14, (1), pp. 41
2. da Cunha Tasca, B.F., Machado, N.G., Lotufo Neto, N., dos Santos, L.O.F., and Biudes, M.S.: 'The respiratory health paradox in mato grosso seniors: environmental drivers of divergent hospitalization and mortality trends', *Air Quality, Atmosphere & Health*, 2026, 19, (3), pp. 59
3. Oterino-Moreira, I., Pérez-Encinas, M., Candel-González, F.J., and Lorenzo-Martínez, S.: 'Impact of Antimicrobial-Resistant Bacterial Pneumonia on In-Hospital Mortality and Length of Hospital Stay: A Retrospective Cohort Study in Spain', *Antibiotics*, 2025, 14, (10), pp. 1006
4. Li, F., Wang, S., Gao, Z., Qing, M., Pan, S., Liu, Y., and Hu, C.: 'Harnessing artificial intelligence in sepsis care: advances in early detection, personalized treatment, and real-time monitoring', *Frontiers in medicine*, 2025, 11, pp. 1510792
5. Ma, Y., Fan, S., and Xi, J.: 'Recent updates regarding the management and treatment of pneumonia in pediatric patients: a comprehensive review', *Infection*, 2025, 53, (6), pp. 2341-2359
6. Al Nufaei, Z.F., and Alshamrani, K.M.: 'Comparing ultrasound, chest X-ray, and CT scan for pneumonia detection', *Medical Devices: Evidence and Research*, 2025, pp. 149-159
7. Woznitza, N., Smith, L., Rawlinson, J., Au-Yong, I., George, B., Djearaman, M.G., Nair, A., Lee, R.W., Navani, N., and Ndwandwe, S.: 'AI-based chest X-ray prioritization in the lung cancer diagnostic pathway: the LungIMPACT randomized controlled trial', *Nature Medicine*, 2026, 32, (5), pp. 1737-1744
8. Al-Worafi, Y.M.: 'Quality of diseases/conditions diagnosis procedures in developing countries: status and future recommendations': 'Handbook of medical and health sciences in developing countries: education, practice, and research' (Springer, 2023), pp. 1-28
9. Kayembe, J.-M.N., and Kayembe, H.-C.N.: 'Challenges of Early Diagnosis and Effective Treatment of Pneumonia in Resource-Limited Settings': 'Viral Infectious Diseases Annual Volume 2025' (IntechOpen, 2025)
10. Ofori-Manteaw, B., Agyemang, P.N., Ofori, E.K., Angmorterh, S.K., and Dzefi-Tettey, K.: 'A snapshot of unreported radiographs, image interpretation turn-around times and implications for radiography practice in a resource-constrained setting', *Journal of Medical Imaging and Radiation Sciences*, 2025, 56, (6), pp. 102108
11. Irede, E.L., Aworinde, O.R., Lekan, O.K., Amienghemhen, O.D., Okonkwo, T.P., Onivefu, A.P., and Ifijen, I.H.: 'Medical imaging: a critical review on X-ray imaging for the detection of infection', *Biomedical Materials & Devices*, 2026, 4, (1), pp. 1-45
12. Siddiqi, R., and Javaid, S.: 'Deep learning for pneumonia detection in chest x-ray images: A comprehensive survey', *Journal of imaging*, 2024, 10, (8), pp. 176
13. Anderson, P.G., Tarder-Stoll, H., Alpaslan, M., Keathley, N., Levin, D.L., Venkatesh, S., Bartel, E., Sicular, S., Howell, S., and Lindsey, R.V.: 'Deep learning improves physician accuracy in the comprehensive detection of abnormalities on chest X-rays', *Scientific reports*, 2024, 14, (1), pp. 25151
14. Obajuluwa, A.M., Balogun, J.A., and Balogun, O.: 'Fundamentals of Radiology for Physical Therapists in Low-and Middle-Income Countries': 'Contemporary and Global Perspectives in Physical Therapy' (Springer, 2026), pp. 1813-1843
15. Michael, C.W., and Engels, M.: 'Benign Category in the WHO Reporting System for Lung Cytopathology: Diagnostic Difficulties and Differential Diagnosis', *Cytopathology*, 2025, 36, (5), pp. 443-457
16. Picchi, S.G., Minieri, A., Lassandro, F., Russo, G., and Lassandro, G.: 'Bacterial Granulomatous Lung Diseases: Radiological Findings and Differential Diagnosis', *Infectious Disease Reports*, 2026, 18, (3), pp. 53
17. Richie, R.C., and Engman, M.L.: 'Respiratory Disorders': 'Brackenridge's Medical Selection of Life Risks' (Springer, 2026), pp. 1-46
18. Ghai, D., Nanda, S.K., Mishra, S.P., Gadade, K., Dhir, K., and Dhande, A.: 'AI-driven diagnostic systems': 'Oral Cancer and Artificial Intelligence' (Elsevier, 2026), pp. 71-94
19. Kumar, A., and Bhattacharyya, A.: 'Introduction to computer-aided medical diagnosis (CAMD) systems': 'Non-Stationary and Nonlinear Data Processing for Automated Computer-Aided Medical Diagnosis' (Elsevier, 2026), pp. 1-14
20. Shobayo, O., and Saatchi, R.: 'Developments in deep learning artificial neural network techniques for medical image analysis and interpretation', *Diagnostics*, 2025, 15, (9), pp. 1072

21. Rehman, A.U., Sayyed, A., Abbas, T., Aslam, Z., Raza, R.A., and Saleh, N.I.M.: 'Uncertainty-Aware CSI Prediction and AI-Seeded Multi-Objective Optimization in RIS Assisted URLLC 6G Networks', *Computer Networks*, 2026, pp. 112625
22. Malik, M.H., abbas, T., Aoun, M., Janjua, J.I., Iqbal, S., and Sayyed, A.: 'Deep learning based feature extraction for liver tumour classification using computed tomography images', *Scientific Reports*, 2026
23. Ehtisham, R., Qayyum, W., Camp, C.V., Plevris, V., Mir, J., Khan, Q.-u.Z., and Ahmad, A.: 'Computing the characteristics of defects in wooden structures using image processing and CNN', *Automation in Construction*, 2024, 158, pp. 105211
24. L. C, M., and P. M, J.P.: 'An optimal deep learning approach for breast cancer detection and classification with pre-trained CNN-based feature learning mechanism', *Journal of Biomolecular Structure and Dynamics*, 2026, 44, (2), pp. 641-656
25. Doreen, A., Mwangi, W., and Muriithi, P.: 'Optimization of matching networks with transfer learning in few-shot pneumonia detection', *Discover Applied Sciences*, 2025, 7, (9), pp. 1024
26. Plested, J., Phiri, M., and Gedeon, T.: 'Deep transfer learning for image classification: a survey', *Artificial Intelligence Review*, 2026, 59, (3), pp. 100
27. Amin, M.I., Shen, M., Ishak, M.K., Manickam, S., and Karuppayah, S.: 'Enhancing generalization of cross-domain intrusion detection: a heterogeneous deep stacked ensemble approach', *Connection Science*, 2026, 38, (1), pp. 2599708
28. Liu, Y., Li, Y., Ma, Y., Wang, B., Guo, C., and Cao, J.: 'Improving Generalization in Rate of Penetration Prediction through Transfer Learning across Oil Fields', *Geoenergy Science and Engineering*, 2026, pp. 214466
29. Rashid, A.B., Asma, J., Barua, K., and Das, D.: 'An enhanced deep learning framework for pneumonia detection in chest x-rays', *SN Computer Science*, 2025, 6, (5), pp. 472
30. Anumula, S.K., Tamilmani, V., Singh, A.A., Rajendran, D., and Namburi, V.D.: 'Deep Learning-Based CNN Model for Automated Detection of Pneumonia from Chest X-Ray Images', in Editor (Ed.)^(Eds.): 'Book Deep Learning-Based CNN Model for Automated Detection of Pneumonia from Chest X-Ray Images' (Springer, 2026, edn.), pp. 360-377
31. Suppiah, R., Danthakanni, R., Nair, M., and Kafumbe, S.: 'Novel Hierarchical Edge AI Architecture for Resource-Constrained Embedded Platforms: A Comprehensive Framework for Distributed Intelligence', in Editor (Ed.)^(Eds.): 'Book Novel Hierarchical Edge AI Architecture for Resource-Constrained Embedded Platforms: A Comprehensive Framework for Distributed Intelligence' (IEEE, 2026, edn.), pp. 1-6
32. Rasheed, M.S., Abbas, T., Tariq, S., Ahmad, N., Haq, E.U., Sultan, N., Aziz, Z., and Arslan, M.: 'A DECENTRALIZED NETWORK INTRUSION DETECTION SYSTEM FOR DISTRIBUTED DENIAL OF SERVICE ATTACK PREVENTION IN VEHICULAR AD HOC NETWORKS', *Spectrum of Engineering Sciences*, 2026, 4, (3), pp. 2073-2095
33. Pavithra, A., Manjunath, K., Kusuma, H., Bhavana, B., and Shree, N.C.: 'Analysis of Chest X-Rays Based on MobileNetV2 Models for the Diagnosis of Symptoms of Pneumonia Using ML', in Editor (Ed.)^(Eds.): 'Book Analysis of Chest X-Rays Based on MobileNetV2 Models for the Diagnosis of Symptoms of Pneumonia Using ML' (Springer, 2025, edn.), pp. 535-547
34. Satrya, G.B., Kurniawan, F., Budiman, G., Pristisahida, A.O., Moesdradjad, B.K.P., Ramatryana, I.N.A., and Choutri, S.E.: 'Sustainable Maritime Applications with Lightweight Classifier Using Modified MobileNet', *Technologies*, 2026, 14, (3), pp. 161
35. Sattar, M.U., Alamro, M.A., Mihoub, A., Aljarboa, S., and Krichen, M.: 'Lightweight deep learning for medical imaging using MobileNetV2-based brain pathology classification with Grad-CAM interpretability', *Frontiers in Medicine*, 2026, 13, pp. 1810860
36. Kanwal, K., Shabir, A., Abbas, T., Hameed, M., and Khan, M.A.: 'Enhanced visual content retrieval via hierarchical deep fusion of multi-level CNN features: K. Kanwal et al', *The Visual Computer*, 2026, 42, (4), pp. 173
37. Tariq, S., Rajwana, M.A., Kashif, M.I., Malik, H., Bukhari, W.A., Khan, A., Ahmad, N., and ul Haq, E.: 'CNN-RL: A Deep Learning-Based Adaptive Crowdsourcing Model for Cybersecurity Education', *STATISTICS, COMPUTING AND INTERDISCIPLINARY RESEARCH*, 2025, 7, (2), pp. 301-317
38. Alharbi, A.H., and Hosni Mahmoud, H.A.: 'Pneumonia transfer learning deep learning model from segmented X-rays', in Editor (Ed.)^(Eds.): 'Book Pneumonia transfer learning deep learning model from segmented X-rays' (MDPI, 2022, edn.), pp. 987

39. Shrimali, S.: 'Optimizing chest X-ray analysis for pneumonia detection through comparative evaluation of transfer learning CNN Architectures', in Editor (Ed.)^(Eds.): 'Book Optimizing chest X-ray analysis for pneumonia detection through comparative evaluation of transfer learning CNN Architectures' (IEEE, 2024, edn.), pp. 1-5
40. Rajeashwari, S., and Arunesh, K.: 'Enhancing pneumonia diagnosis with ensemble-modified classifier and transfer learning in deep-CNN based classification of chest radiographs', *Biomedical Signal Processing and Control*, 2024, 93, pp. 106130
41. Chandola, Y., Uniyal, V., and Bachheti, Y.: 'Efficient chest x-ray feature extraction and feature fusion for pneumonia detection using lightweight pretrained deep learning models', *medRxiv*, 2025, pp. 2025.2006.2028.25330459
42. Hasan, M., Liu, H., Lakkamraju, V., Yasmin, F., Maurya, S., and Sinha, A.: 'Transfer Learning-Driven Deep Neural Network for Pneumonia Detection from Chest X-Rays', in Editor (Ed.)^(Eds.): 'Book Transfer Learning-Driven Deep Neural Network for Pneumonia Detection from Chest X-Rays' (IEEE, 2025, edn.), pp. 1-6
43. Yanar, E., Hardalaç, F., and Ayturan, K.: 'PELM: a deep learning model for early detection of pneumonia in chest radiography', *Applied Sciences*, 2025, 15, (12), pp. 6487
44. Colin, J., and Surantha, N.: 'Interpretable deep learning for pneumonia detection using chest x-ray images', *Information*, 2025, 16, (1), pp. 53
45. Usman, C., Rehman, S.U., Ali, A., Khan, A.M., and Ahmad, B.: 'Pneumonia disease detection using chest X-rays and machine learning', *Algorithms*, 2025, 18, (2), pp. 82
46. Aljuaid, H., Adlan, H., Alkebsi, B., Alfurhood, B.S., Liotta, A., and Cavallaro, L.: 'An experimental comparison of deep learning models for pneumonia classification from chest X-ray images', *Biomedical Signal Processing and Control*, 2026, 112, pp. 108742
47. Alam, M., Rahman, A., Yaser, S., Khan, M.A., and Mishu, A.M.: 'A Lightweight Dual-Branch MobileNet Architecture for Precise Pneumonia Detection From Chest X-Ray Images', in Editor (Ed.)^(Eds.): 'Book A Lightweight Dual-Branch MobileNet Architecture for Precise Pneumonia Detection From Chest X-Ray Images' (IEEE, 2026, edn.), pp. 1-6
48. Kermany, D.S., Goldbaum, M., Cai, W., Valentim, C.C., Liang, H., Baxter, S.L., McKeown, A., Yang, G., Wu, X., and Yan, F.: 'Identifying medical diagnoses and treatable diseases by image-based deep learning', *cell*, 2018, 172, (5), pp. 1122-1131. e1129
49. Li, M., and Yan, N.: 'IPD-YOLO: Person detection in infrared images from UAV perspective based on improved YOLO11', *Digital Signal Processing*, 2026, 168, pp. 105469
50. Gao, Z., Tian, Y., Lin, S.-C., and Lin, J.: 'A ct image classification network framework for lung tumors based on pre-trained mobilenetv2 model and transfer learning, and its application and market analysis in the medical field', *arXiv preprint arXiv:2501.04996*, 2025
51. Ul Haq, I., Husnain, G., Iqbal, A., and Alzahrani, A.S.: 'Attention-enhanced MobileNetV2 models for robust forest fire detection and classification', *Scientific Reports*, 2026, 16, (1), pp. 4805
52. Patil, Y.D., Tripathi, S.L., and Mahmud, M.: 'Deep Learning Architectures for Ophthalmic Diagnosis: Evaluating CNN, VGG16, ResNet50, and Inception V3 in Myopia Detection', in Editor (Ed.)^(Eds.): 'Book Deep Learning Architectures for Ophthalmic Diagnosis: Evaluating CNN, VGG16, ResNet50, and Inception V3 in Myopia Detection' (IEEE, 2026, edn.), pp. 1-6
53. Nadour, M., Rabehi, A., Hadroug, N., Guermoui, M., Tibermacine, I.E., Alanazi, A.K., Habib, M., and Rabehi, A.: 'Deep hybrid CNN-biLSTM model for accurate solar photovoltaic power forecasting: A comparative study with classical and neural models', *Energy Reports*, 2026, 15, pp. 109119
54. Yang, L., Shen, X., Liu, Y., Chen, J., Zou, Y., Xu, L., Ji, W., Zhang, Y., Liu, T., and Cao, Q.: 'Development and validation of KCPREDICT: a deep learning model for early detection of coronary artery lesions in Kawasaki disease patients', *Pediatric Cardiology*, 2026, 47, (1), pp. 266-275
55. Jeon, E., Jin, D., and Kim, Y.: 'Effects of Window and Batch Size on Autoencoder-LSTM Models for Remaining Useful Life Prediction', *Machines*, 2026, 14, (2), pp. 135