A Hybrid Deep Learning Approach for Pneumonia Detection from Chest X-Ray Images Using Custom CNN and Transfer Learning

Ashoktaru Pal[1]
Indrajit Pal[2]
Madhurima Das[3]

Computer Application dept.
Future Institute of Engineering and Management
Kolkata-700150- India
[1]pal.ashoktaru@gmail.com
[2]indrajit.pal@teamfuture.in

Abstract. Pneumonia remains an important worldwide health concern, significantly affecting vulnerable populations such as children and the elderly. The timely and precise diagnosis of pneumonia using chest X-ray (CXR) imaging is essential; however, manual interpretation often faces challenges due to complex visual characteristics and inherent diagnostic subjectivity. This study introduces a hybrid deep learning framework that merges a customized Convolutional Neural Network (CNN) with the MobileNetV2 architecture, aimed at improving automated pneumonia classification in CXR images. The approach is structured into five primary phases: data preprocessing and increase, baseline CNN design, transfer learning utilizing VGG16 and MobileNetV2, hybrid model integration, and thorough evaluation, including Grad-CAM visualizations. The dataset was meticulously separated into training, validation, and testing sets, employing augmentation techniques to enhance model robustness. Individual models, namely the standalone CNN, VGG16, and MobileNetV2, recorded accuracies of 87.98%, 90.06%, and 85.26%, respectively, while the hybrid model achieved a superior test accuracy of 90.06% with minimal loss. Notably, the hybrid architecture evidenced high F1-scores of 0.86 for NORMAL and 0.92 for PNEUMONIA classes, demonstrating an effective balance of precision and recall, further validated by Grad-CAM visualizations that illustrate clinically relevant areas of focus. Despite certain limitations such as dataset size and potential class imbalances, the proposed model effectively enhances pneumonia detection, positioning itself as a valuable asset for clinical decision-making, particularly in resource-limited surroundings.

Keywords: Pneumonia Detection, Chest X-ray, Deep Learning, Convolutional Neural Network (CNN), Transfer Learning, MobileNetV2, VGG16, Hybrid Model, Medical Image Classification, Grad-CAM, Image Augmentation, Model Interpretability.

(Received / Accepted)

1 Introduction

Pneumonia remains one of the most common and potentially life-threatening breathing infections, affecting millions globally each year. It is particularly dangerous for vulnerable populations, including children, the elderly, and individuals with weak-

ened immune systems. Early and accurate diagnosis is crucial in ensuring timely medical intervention, which can significantly reduce mortality and morbidity rates. Clinicians can clearly identify infection symptoms like lung consolidation with chest X-rays, infiltrates, or pleural effusion[20]. Convolutional neural networks (CNNs), a type of deep learning technique, have demonstrated promise in medical picture classification applications in recent years. Radiologists may benefit from the use of artificial intelligence in medical diagnostics, reduce diagnostic delays, and enhance accuracy. This study seeks to address these issues by creating and assessing deep learning- based models for binary classification of chest X-rays, aiming to detect pneumonia while maintaining interpretability through visualization techniques such as Grad-CAM.

The aim of this study is to develop an efficient and accurate hybrid deep learning model by combining a custom CNN and MobileNetV2 for automated classification of pneumonia from chest X-ray images, and to compare its performance against traditional CNN and pre-trained transfer learning models.

This article presents an investigation into pneumonia classification using chest X-ray images, employing systematic preprocessing, model development, and comparative evaluation. Image augmentation techniques are applied to enhance generalization and mitigate over fitting during training[36]. As baseline models, a custom convolutional neural network (CNN), VGG16, and MobileNetV2 are implemented and evaluated. Building on these architectures, a hybrid model is developed by integrating a custom CNN branch with MobileNetV2 to exploit their complementary feature extraction capabilities. Experimental results are comparatively analyzed, with performance visualizations highlighting the superior classification precision and resilience of the hybrid model over standalone approaches.

- **Data Preparation and Augmentation:** The dataset is structured into train, validation, and test folders, and image data generator is used for preprocessing. Training data is augmented with random rotations, zooms, and horizontal flips to improve generalization. Images are rescaled to normalize pixel values between 0 and 1. The generator flow loads images in batches, reducing memory usage. Validation data is split from the training set, ensuring unbiased evaluation. The test set remains untouched for final performance metrics. This preprocessing phase ensures the models learn from diverse and normalized inputs, which is crucial for effective convergence and better generalization to unseen data during inference.

- **Custom CNN Model Design and Training:** A sequential CNN is built with three convolutional layers followed by max pooling, capturing hierarchical spatial features. After flattening, it connects to a dense layer with dropout for regularization. The model is compiled using binary cross-entropy loss and the Adam optimizer. It is trained for 10 epochs on the chest X-ray training dataset, with validation at each step to monitor overfitting. To illustrate learning progress, accuracy and loss are plotted for both training and validation sets. This baseline model serves as a foundation for comparison against transfer learning and hybrid approaches, providing a custom architecture for pneumonia detection.

- **Transfer Learning with VGG16 and MobileNetV2:** Pretrained VGG16 and MobileNetV2 models are used as fixed feature extractors by freezing their convolutional base layers[1]. Custom classification heads are appended, including global average pooling, dense layers, dropout, and sigmoid output for binary classification. Each model is trained on the same dataset configuration for fairness. The performance is evaluated on test data, including accuracy, classification reports, and confusion matrices. The feature reuse from large-scale ImageNet training allows faster convergence and improved accuracy. Transfer learning enables leveraging deep semantic features learned from diverse datasets, which significantly enhances pneumonia detection performance compared to the baseline custom CNN model. Hybrid Model Integration: A dual-branch hybrid architecture combines a shallow custom CNN with the feature-rich MobileNetV2. Each branch independently extracts features from the same input image. The CNN focuses on handcrafted features, while MobileNetV2 captures high-level representations. Global average pooling layers reduce dimensionality before feature fusion via concatenation. The combined vector is passed through dense and dropout layers be-

fore final classification. The model is compiled and trained similarly to previous models. This fusion approach leverages complementary strengths of traditional CNNs and transfer learning, improving the model's resilience and identifying local and global characteristics to diagnose pneumonia in chest X-rays more precisely.

- **Evaluation, Visualization, and Grad-CAM Interpretation:** Each model is evaluated using classification metrics, confusion matrices, and visualizations. Accuracy and loss trends are plotted to assess learning behavior. Predictions on sample images are shown with actual and predicted labels to inspect inference quality. Grad-CAM heatmaps highlight regions of interest in pneumonia-positive X-rays, explaining model attention and improving interpretability. Finally, a bar chart comparison summarizes model performances across accuracy and loss, helping identify the best-performing architecture. These evaluation techniques provide transparency, validate results, and ensure the models not only achieve high metrics but also learn meaningful visual patterns for real-world clinical decision support.

This study holds considerable significance in both the clinical and academic domains. From a clinical perspective, the integration of deep learning into pneumonia detection can substantially enhance early diagnosis, particularly in areas where access to radiologists is limited. Accurate automated systems can assist healthcare professionals by reducing diagnostic time and offering second opinions, thereby improving patient care and outcomes. The use of Grad-CAM further enables clinicians to understand model decisions, fostering trust and facilitating adoption in clinical workflows.

This research contributes to the growing body of research in medical image analysis using artificial intelligence. It provides a comprehensive comparison between a custom CNN and two widely adopted transfer learning models VGG16 and MobileNetV2 using the same dataset and evaluation metrics. The incorporation of Grad-CAM bridges the gap between performance and interpretability, addressing a major challenge in deep learning applications in healthcare.

This research provides several significant contributions:

- **Development of a Hybrid Deep Learning Model:** Designed a novel two-branch hybrid architecture that combines a custom CNN with MobileNetV2. The custom CNN captures domain-specific features, while MobileNetV2 contributes high-level pre-trained features, enhancing feature diversity and improving diagnostic performance for pneumonia detection.
- **Effective Feature Fusion Strategy:** Implemented a feature fusion mechanism using concatenation of outputs from both branches, followed by dense layers and dropout. This integration of handcrafted and transfer-learned features boosts classification accuracy while controlling overfitting.
- **Comprehensive Evaluation and Visualization:** Conducted thorough evaluation using classification metrics (accuracy, precision, recall, F1-score), confusion matrix, and prediction visualization. This ensures transparent and interpretable performance analysis of the hybrid model on unseen test data.
- **Real-Time Generalization Through Data Augmentation:** Applied real-time data augmentation (rotation, zoom, flips) during training, enabling the hybrid model to generalize well on varied chest X-ray images and reducing the risk of overfitting on limited medical data.

In Section 2, relevant work on image caption generation is reviewed. The materials and procedures are discussed in Section 3, with particular attention to the different tools and methods utilized for image encoding and decoding. The experimental setup, result analysis, and outcome comparison are covered in Section 4. The paper is finally concluded in Section 5 with a review of the results and some directions for further research.

2 Literature Review

In recent years, the application of custom Convolutional Neural Networks (CNNs) has gained significant momentum in the field of medical imaging, particularly for the diagnosis of pneumonia from chest X-ray (CXR) images. Unlike transfer learning models, which adapt pretrained weights from large datasets, custom CNN architectures are designed and optimized specifically for the task and data at hand, allowing for a more tailored solution that captures domain-specific features.

It is presented a robust 50-layer CNN architecture trained on a dataset comprising 5,852 CXR images to classify cases into viral pneumonia, bacterial pneumonia, and normal categories. Their architecture was rigorously validated using five-fold cross-validation and achieved an outstanding accuracy of $99.7\% \pm 0.2$, sensitivity of $99.74\% \pm 0.1$, and an AUC of 0.9812 [7]. The study underlined how crucial early and precise diagnosis is, particularly for vulnerable groups like children and the elderly. Their proposed model demonstrates not only high diagnostic performance but also practical viability for deployment in remote or resource-constrained environments, where access to radiological expertise is limited.

It is proposed an innovative hybrid approach by combining deep CNNs with hand-crafted features extracted through SURF and ORB techniques[22]. Their method incorporated segmentation via InfoMGAN, followed by feature fusion and classification using machine learning models like SVM and random forest. The model achieved a testing accuracy of 98.20% and was statistically validated through ANOVA and McNemarâs tests. Although the model was designed for the classification of ten chest diseases including pneumonia it highlights the evolving role of feature fusion in enhancing CNN performance and interpretability. The use of handcrafted features alongside CNNs provides better feature localization, particularly valuable for diseases with subtle radiographic signatures.

It is conducted a comparative analysis of twelve widely known ImageNet pre-trained models, including VGG16 and MobileNetV2, to evaluate their effectiveness in distinguishing pneumonia cases of bacterial and viral origin, including those caused by SARS-CoV-2[8].

Using a combined dataset of 6,330 chest X-ray images, they fine-tuned these architectures and reported promising results, with models reaching an average F1-score of up to 84.46%. The study further integrated interpretability techniques like Grad-CAM and evaluated model robustness by training on reduced dataset sizes (50%, 20%, 10%). This analysis underlined the adaptability of transfer learning models to real-world challenges in medical diagnostics, especially during outbreaks of novel pathogens.

Similarly, RadoÄaj and MartinoviÄ (2025) focused on architectural variations and performance improvements using models like InceptionV3, InceptionResNetV2, DenseNet201, and MobileNetV2[27]. The researchers employed different convolution strategies, including multi-scale and stride convolutions, combined with the Mish activation function. Their results indicated that MobileNetV2, when enhanced with multi-scale convolutions, reached an accuracy of 94.37%. Grad-CAM visualizations revealed that such configurations captured diffuse pneumonia patterns effectively, particularly in paediatric cases. These findings stress the importance of choosing the right convolutional structure and activation function when adapting pre-trained models for medical imaging.

It is proposed a customized MobileNetV2 architecture (MobileNetV2L2), incorporating L2 regularization to improve generalization[37]. Their model outperformed other state-of-the-art approaches, achieving a test accuracy of 95.51%, while a secondary custom CNN model attained 99.26%. These results highlight the flexibility and efficiency of MobileNetV2, particularly in environments with computational constraints. The authors demonstrated that model fine-tuning, especially with architectural innovation,s can significantly enhance the performance of standard transfer learning models.

This study sets a new benchmark for transfer learning in medical diagnostics and encourages further exploration into domain-adaptive pre-training techniques[18].

In another notable study[33], applied DenseNet121 and ResNet50 to a large dataset of over 100,000 chest X-ray images and demonstrated superior performance over expert radiologists in detecting conditions such as pneumothorax and oedema. The integration of explainable AI tools like SHAP, LIME, and occlusion sensitivity analysis not only improved transparency but also built trust in automated systems among healthcare professionals.

It is emphasized the crucial role of interpretability and visualization techniques in enhancing the clinical reliability of deep learning models in medical image analysis[33]. Their comprehensive survey identifies a range of methodologies such as saliency maps, Class Activation Maps (CAMs), and gradient-based visualizations that contribute to a deeper understanding of model behavior. These techniques not only improve transparency but also enable clinicians to verify whether the model is attending to medically relevant regions in an image, which is essential for adoption in real-world healthcare settings.

Authors	Method Used	Result	Research Gap
Colin and Surantha (2025)	LRP, Adversarial training, CAM, Spatial Attention	LRP best trade-off between interpretability and accuracy (0.91 accuracy, 0.85 MRS)	Performance-interpretability trade-offs require further optimization
RadoÄaj and MartinoviÄ (2025)	CNN architectures with Mish activation, Grad-CAM	InceptionResNetV2 best accuracy (97.18%); Grad-CAM shows distinct pathology patterns	Need for further studies on real-time application and multi-modal data integration
Gu and Lee (2024)	Transfer learning from ImageNet pretrained models	Transfer learning outperforms training from scratch; effective for grayscale-to-RGB CXR images	Domain adaptation for medical images needs further theoretical development
Tripathi et al. (2024)	Customized MobileNetV2 + CNN with L2 regularization	Proposed models outperform EfficientNet, ResNet variants; high accuracy and validation results	Limited generalization beyond datasets used; interpretability not deeply explored
Sufian et al. (2024)	DenseNet121, ResNet50, NLP, SHAP, LIME	High AUC (94%), outperform radiologists in some tasks; enhanced interpretability with SHAP and LIME	Complexity of AI integration in clinical workflow; needs usability studies
Bhati et al. (2024)	Review of interpretability methods (CAM, saliency)	Visualization improves trust and clinical relevance of models	Lack of standardization in interpretability metrics and clinical validation
Neshat et al. (2024)	Hybrid Inception-ResNet, Grad-CAM	High accuracy in pneumonia classification; interpretability via Grad-CAM; improved clinical relevance	Exploration of larger datasets and deployment in clinical settings
Malik et al. (2023)	Hybrid DCNN + feature extraction + ML classifiers	Achieved 98.2% accuracy for 10 chest diseases; robust statistical validation	Complexity of pipeline; needs simplification for clinical use
Ahmed et al. (2022)	Survey of XAI techniques in medical image analysis	Comprehensive XAI methods classification; benefits for automated reporting	Need for real-world clinical testing and integration of explanations in AI workflows
Avola et al. (2022)	Fine-tuned 12 ImageNet pretrained models	Models reached up to 84.46% F1-score on viral/bacterial pneumonia; Grad-CAM visualizations support interpretability	Performance drops with reduced training data; needs robustness improvement
Alsharif et al. (2021)	Custom CNN, 5-fold cross-validation	High accuracy (99.7%), sensitive viral/bacterial/normal classification; cost-effective for remote areas	Limited exploration of model interpretability and real-time deployment
Hosseinzadeh Taher et al. (2021)	Pretrained ImageNet, iNat2021, self-supervised models	Self-supervised and continual pretraining boost performance on medical tasks	Transferability of models across all medical image types remains uncertain

Table 1: Review of recent research work in pneumonia detection

A review table has been describe above in table1 with research gap op those author. The be-

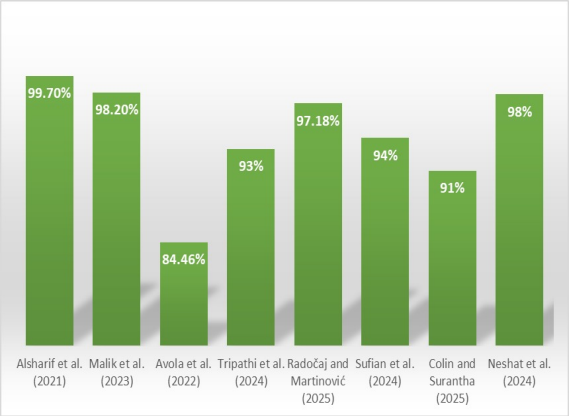


Figure 1: Bar diagram of Accuracy over the different models of Journals.

low figure 1 stated the accuracy of the chosen literatures based on the Chest X-Ray for Pneumonia as discussed in the articles of the different authors that is highly mentioned in the X axis of this chart. It is further expand the discussion by exploring explainable AI within the context of automated medical report generation[3]. Their review classifies XAI techniques used in both image interpretation and natural language explanation, highlighting how these methods contribute to model transparency and clinician trust.

Despite significant advances in applying deep learning to medical imaging, several critical gaps remain that hinder the full clinical adoption of these technologies. Firstly, while transfer learning using pre-trained models such as VGG16 and MobileNetV2 has demonstrated promising results in pneumonia and lung disease classification ([8],[37]), challenges persist in effectively bridging the domain gap between natural images used for pretraining and medical images[18]. Continual pretraining and fine-tuning strategies have been proposed but require further validation across diverse datasets and modalities.

Secondly, although many studies achieve high accuracy in disease classification tasks ([24], [33]), they often do so using imbalanced datasets or limited image sources, which limits generalizability. Semi-supervised learning and data augmentation techniques show potential to mitigate these issues, but these methods are still underexplored for broader pneumonia subtypes and related lung diseases.

This review study has examined the current landscape of deep learning applications in pneumonia diagnosis from chest radiographs, focusing on four primary themes critical to advancing the field. First, the development of custom convolutional neural networks (CNNs) has demonstrated promising accuracy in detecting pneumonia by addressing domain-specific challenges such as data imbalance and subtle pathological features. Models with tailored architectures can outperform standard networks, providing cost-effective solutions, especially in resource-limited settings.

Transfer learning using state-of-the-art architectures like VGG16 and MobileNetV2 has been widely adopted to improve classification performance. Fine-tuning pretrained models on large datasets enables efficient feature extraction, making these approaches valuable for medical imaging tasks where annotated data is often scarce. The integration of such models offers robustness and adaptability across diverse pneumonia types, including viral and bacterial infections. Comparative performance evaluations reveal that while many deep learning models achieve high accuracy, their effectiveness varies depending on the dataset characteristics and model complexity. The significance of thorough benchmarking and the requirement for models that strike a balance between accuracy, computational efficiency, and generalizability are highlighted by these studies.

3 Materials and Methods

The implementation of the pneumonia classification and visualization pipeline leverages several essential Python libraries and tools that streamline deep learning, image processing, and visualization tasks. TensorFlow and its high-level API Keras serve as the primary frameworks for building, training, and evaluating convolutional neural networks (CNNs), including custom models and pre-trained architectures like VGG16 and MobileNetV2. NumPy facilitates efficient array manipulation and numerical operations, crucial for image preprocessing and tensor transformations.

For image handling and enhancement, OpenCV is utilized especially for resizing, colour conversion, and overlay operations in the Grad-CAM visualization process. Matplotlib provides a flexible interface for plotting model performance metrics and displaying Grad-CAM outputs. Addition-

ally, Seaborn is used to create aesthetically pleasing and informative visualizations such as heatmaps and confusion matrices. Our method has a defined workflow in below figure2.

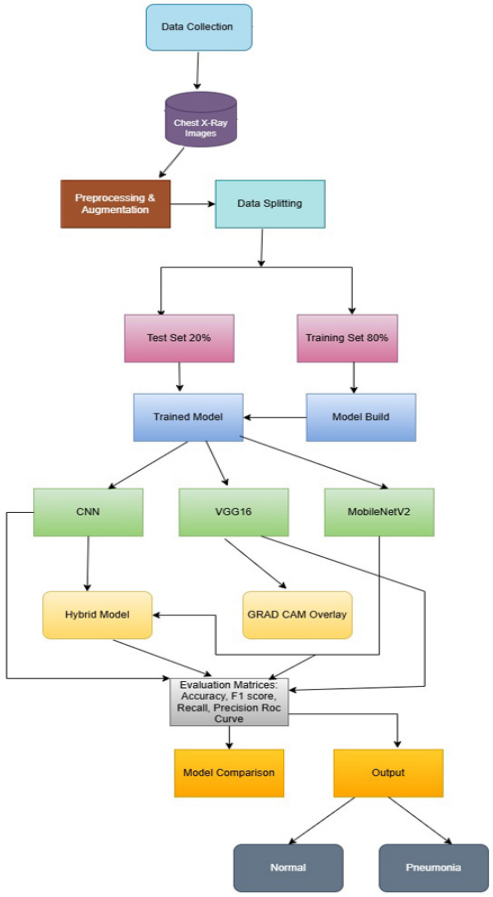


Figure 2: Workflow diagram for Pneumonia detection.

3.1 Baseline Convolutional Neural Network (CNN)

The first model implemented for pneumonia detection was a basic Convolutional Neural Network (CNN), designed using TensorFlowâs Keras Sequential API. The architecture comprised three convolutional layers with 32, 64, and 128 filters respectively, each using a 3×3 kernel and ReLU activation:

$$\sigma(z) = \max(0, z)$$

For an input image

$$X \in \mathbb{R}^{224\times224\times3},$$

each convolutional layer computed its feature maps as:

$$a_{ij}^{(k)} = \sigma\left(\sum_{c=1}^{C}\sum_{m=1}^{h}\sum_{n=1}^{w} W_{mn}^{(k,c)} \cdot X_{(i+m),(j+n),c} + b^{(k)}\right),$$

where $W^{(k,c)}$ and $b^{(k)}$ are the trainable kernel weights and bias.

To progressively reduce spatial dimensions while preserving salient features, each convolution was followed by a max-pooling operation:

$$P_{ij}^{(k)} = \max\left\{a_{2i,2j}^{(k)},\ a_{2i+1,2j}^{(k)},\ a_{2i,2j+1}^{(k)},\ a_{2i+1,2j+1}^{(k)}\right\}.$$

The pooled output was then flattened into a 1D vector $z \in \mathbb{R}^n$, which served as input to a fully connected dense layer with 128 units:

$$h = \sigma\left(W^{(1)}z + b^{(1)}\right), \quad W^{(1)} \in \mathbb{R}^{128\times n},\ b^{(1)} \in \mathbb{R}^{128}.$$

To reduce overfitting, a Dropout layer ($p = 0.5$) was added before the final classification stage. The output layer consisted of a single sigmoid unit:

$$\hat{y} = \sigma\left(W^{(2)}h + b^{(2)}\right) = \frac{1}{1 + e^{-\left(W^{(2)}h + b^{(2)}\right)}},$$

where $\hat{y} \in [0, 1]$ represents the probability of pneumonia.

Training was guided by the binary cross-entropy loss function:

$$L(y, \hat{y}) = -\left[y \log(\hat{y}) + (1 - y)\log(1 - \hat{y})\right],$$

and optimized using the Adam optimizer, which updates parameters as:

$$\theta \leftarrow \theta - \eta\frac{\partial L}{\partial \theta}.$$

This baseline CNN provided a computationally efficient and interpretable framework, effectively learning hierarchical features (edges, textures, lung patterns) critical for distinguishing pneumonia from normal cases. Despite its simplicity, it established a strong baseline before progressing to deeper architectures such as VGG16 and MobileNetV2.

3.2 VGG16 (Transfer Learning)

The second model used in this study is built on the **VGG16** architecture â a deep convolutional neural network originally trained on the ImageNet dataset. In this approach, we applied transfer learning, using the pre-trained VGG16 as a feature extractor. To achieve this, the fully connected classification layers at the top of the network were removed (`include_top=False`), keeping only the convolutional base. All layers in this base were frozen (`layer.trainable=False`) to retain the valuable hierarchical features learned from ImageNet.

On top of this base, a custom classification head was added. A GlobalAveragePooling2D (GAP) layer compressed each 7×7 feature map into a single value, yielding a vector:

$$g_k = \frac{1}{49} \sum_{i=1}^{7} \sum_{j=1}^{7} f_{i,j,k}, \quad k = 1, \ldots, 512 \tag{1}$$

resulting in $g \in \mathbb{R}^{512}$. This was followed by a dense layer with 128 units and ReLU activation:

$$h = \text{ReLU}(W_1 g + b_1), \quad h \in \mathbb{R}^{128}, \tag{2}$$

then a dropout layer with rate $p = 0.5$, and finally a dense output layer with sigmoid activation:

$$\hat{y} = \sigma(W_2 h' + b_2), \quad \hat{y} \in (0, 1), \tag{3}$$

where the sigmoid function is defined as:

$$\sigma(z) = \frac{1}{1 + e^{-z}}. \tag{4}$$

The model was optimized using the binary cross-entropy loss function:

$$L(y, \hat{y}) = -\big[y \log(\hat{y}) + (1-y) \log(1-\hat{y})\big], \quad y \in \{0, 1\}. \tag{5}$$

The overall architecture can be summarized as:

$$x \in \mathbb{R}^{224\times224\times3} \xrightarrow{\text{VGG16}} f \in \mathbb{R}^{7\times7\times512} \xrightarrow{\text{GAP}}$$
$$\text{g} \in \mathbb{R}^{512} \xrightarrow{\text{Dense+ReLU}} h \in \mathbb{R}^{128} \xrightarrow{\text{Dropout}}$$
$$\text{h'} \xrightarrow{\text{Dense+Sigmoid}} \hat{y} \in (0, 1).$$

The model was compiled with the Adam optimizer and trained for 10 epochs. This VGG16-based model effectively transfers robust low-level features from natural images to chest X-ray analysis, achieving improved performance for pneumonia detection, particularly in scenarios with limited medical datasets.

3.3 MobileNetV2 (Transfer Learning)

The third model developed in this study is based on **MobileNetV2**, a lightweight and efficient convolutional neural network specifically optimized for mobile and embedded vision applications. In this work, transfer learning was applied by using a pre-trained MobileNetV2 model as a fixed feature extractor. To achieve this, the modelâs top layers were removed (`include_top=False`), and all convolutional layers were frozen to preserve the valuable feature representations learned from the ImageNet dataset.

To tailor the network for binary classification of chest X-ray images, a custom classification head was added. The output from the MobileNetV2 base passed through a Global Average Pooling (GAP) layer to reduce feature dimensionality, followed by a fully connected Dense layer with 128 ReLU-activated neurons to introduce non-linearity. A Dropout layer with a rate of 0.5 was included to minimize overfitting, and the final layer consisted of a single sigmoid neuron to produce probabilities distinguishing between NORMAL and PNEUMONIA cases. The model was compiled using the Adam optimizer and trained for 10 epochs with the pre-defined training and validation generators.

The lightweight nature of MobileNetV2 allows it to perform efficiently even on systems with limited computational power. Despite its compact design, it maintains strong predictive performance through the use of depthwise separable convolutions, which significantly reduce the number of parameters and computational operations without compromising accuracy.

Input Each chest X-ray image is resized to:

$$x \in \mathbb{R}^{224\times224\times3}.$$

Feature Extraction (Frozen) The MobileNetV2 base processes the input through convolutional and bottleneck layers, producing:

$$f = \text{MobileNetV2}(x), \quad f \in \mathbb{R}^{7\times7\times1280}.$$

Global Average Pooling (GAP) Each feature map is reduced to a single scalar:

$$g_k = \frac{1}{49} \sum_{i=1}^{7} \sum_{j=1}^{7} f_{i,j,k}, \quad k = 1, \ldots, 1280,$$

resulting in:

$$g \in \mathbb{R}^{1280}.$$

Fully Connected Layer (Dense + ReLU) A linear transformation followed by ReLU activation:

$$h = \text{ReLU}(W_1 g + b_1), \quad W_1 \in \mathbb{R}^{128\times 1280},\ b_1 \in \mathbb{R}^{128},$$

where:

$$h \in \mathbb{R}^{128}.$$

Dropout To reduce overfitting, a dropout mask is applied:

$$h' = h \odot m, \quad m \sim \text{Bernoulli}(0.5).$$

Output Layer (Sigmoid) The final probability for binary classification is computed as:

$$\hat{y} = \sigma(W_2 h' + b_2), \quad W_2 \in \mathbb{R}^{1\times 128},\ b_2 \in \mathbb{R},$$

where the sigmoid function is defined as:

$$\sigma(z) = \frac{1}{1+e^{-z}}, \quad \hat{y} \in (0,1).$$

Loss Function Binary cross-entropy is used as the loss function:

$$L(y,\hat{y}) = -\big[y\log(\hat{y}) + (1-y)\log(1-\hat{y})\big], \quad y \in \{0,1\}.$$

Optimization The parameters $\{W_1, W_2, b_1, b_2\}$ are updated using the Adam optimizer:

$$\theta \leftarrow \theta - \eta \cdot \text{AdamGradient}(L,\theta).$$

Final Architecture Summary

$$x \in \mathbb{R}^{224\times224\times3} \xrightarrow{\text{MobileNetV2}} f \in \mathbb{R}^{7\times7\times1280} \xrightarrow{\text{GAP}} g \in \mathbb{R}^{1280} \xrightarrow{\text{Dense+ReLU}} h \in \mathbb{R}^{128} \xrightarrow{\text{Dropout}} h' \in \mathbb{R}^{128} \xrightarrow{\text{Dense+Sigmoid}} \hat{y} \in (0,1).$$

3.4 Algorithm

The Pneumonia Classification algorithm 3.4 presents an organized pipeline for generating explanatory captions from input images based on a CNN, VGG16 and MobileNetV2 model. It starts by choosing and initializing a pretrained CNN (e.g., VGG16, MobileNetV2) to extract high-level visual features.

PneumoniaClassifier$x \in \mathbb{R}^{224\times224\times3}$
Step 1: Initialize datasets (training, validation, test) using ImageDataGenerator
Step 2: Normalize images (rescale = 1/255) and apply data augmentation
each epoch
Step 3: Train CNN Forward propagate training samples through Conv2D, MaxPooling, GAP, Dense, Dropout, Sigmoid Compute Binary Crossentropy Loss and update CNN weights (Adam Optimizer)
Step 4: Train VGG16 (transfer learning, top excluded) Forward propagate samples through pretrained VGG16 base, GAP, Dense, Dropout, Sigmoid Compute Binary Crossentropy Loss and update only top layers (Adam Optimizer)
Step 5: Train MobileNetV2 (transfer learning, top excluded) Forward propagate samples through pretrained MobileNetV2 base, GAP, Dense, Dropout, Sigmoid Compute Binary Crossentropy Loss and update only top layers (Adam Optimizer)
Step 6: Evaluate CNN, VGG16, MobileNetV2 on test set Compute Accuracy, Precision, Recall, F1-score Generate confusion matrix and classification report $\hat{y}_{CNN}, \hat{y}_{VGG}, \hat{y}_{MobileNet} \in [0,1] \rightarrow$ {NORMAL, PNEUMONIA}

3.5 Hybrid Model (Custom CNN + MobileNetV2):

The hybrid model for pneumonia detection is designed using a two-branch deep learning framework that integrates a custom Convolutional Neural Network (CNN) with the pretrained MobileNetV2 architecture. The input consists of chest X-ray images resized to 224×224 pixels, and preprocessing is performed using `ImageDataGenerator` for augmentation and normalization.

3.5.1 Branch 1: Custom CNN Feature Extraction

The first branch (custom CNN) extracts handcrafted low- to mid-level features through stacked convolution and pooling layers. Formally, given an input image $x \in \mathbb{R}^{224\times224\times3}$, convolution and activation operations can be expressed as:

$$a_1 = \text{ReLU}(W_1 * x + b_1), \quad a_1 \in \mathbb{R}^{224\times224\times32} \tag{6}$$

$$a_2 = \text{MaxPool}(a_1), \quad a_2 \in \mathbb{R}^{112\times112\times32} \tag{7}$$

$$a_3 = \text{ReLU}(W_2 * a_2 + b_2), \quad a_3 \in \mathbb{R}^{112\times112\times64} \tag{8}$$

$$a_4 = \text{MaxPool}(a_3), \quad a_4 \in \mathbb{R}^{56\times56\times64} \tag{9}$$

$$a_5 = \text{ReLU}(W_3 * a_4 + b_3), \quad a_5 \in \mathbb{R}^{56\times56\times128} \tag{10}$$

Global Average Pooling summarizes spatial information into a compact feature vector:

$$f_{CNN} = \frac{1}{56 \times 56} \sum_{i=1}^{56} \sum_{j=1}^{56} a_5[i, j, :], \quad f_{CNN} \in \mathbb{R}^{128} \tag{11}$$

3.6 Branch 2: MobileNetV2 Feature Extraction

The second branch (MobileNetV2) leverages transfer learning, where pretrained ImageNet weights are frozen. Its output after global average pooling is:

$$f_{MobileNetV2} = \text{GAP}(\text{MobileNetV2}(x)), \quad f_{MobileNetV2} \in \mathbb{R}^{1280} \tag{12}$$

3.6.1 Fusion and Classification

Both branches are fused via concatenation:

$$f_{concat} = \text{Concat}(f_{CNN}, f_{MobileNetV2}), \quad f_{concat} \in \mathbb{R}^{1408} \tag{13}$$

This fused feature vector is passed through a fully connected layer with ReLU activation and dropout regularization:

$$h = \text{ReLU}(W_d f_{concat} + b_d), \quad h \in \mathbb{R}^{256} \tag{14}$$

$$h' = h \odot m, \quad m \sim \text{Bernoulli}(0.5) \tag{15}$$

Finally, a sigmoid output layer produces the pneumonia classification:

$$\hat{y} = \sigma(W_0 h' + b_0), \quad \hat{y} \in (0, 1) \tag{16}$$

with the sigmoid function defined as:

$$\sigma(z) = \frac{1}{1 + e^{-z}} \tag{17}$$

3.6.2 Loss Function

The model is trained with Binary Cross-Entropy Loss:

$$L(y, \hat{y}) = -\Big[y \log(\hat{y}) + (1 - y) \log(1 - \hat{y})\Big] \tag{18}$$

This design allows the CNN to capture spatially localized features while MobileNetV2 contributes semantically rich representations. The below figure3 architecture supports modular evaluation of their fusion significantly enhances classification accuracy and generalization compared to standalone models.

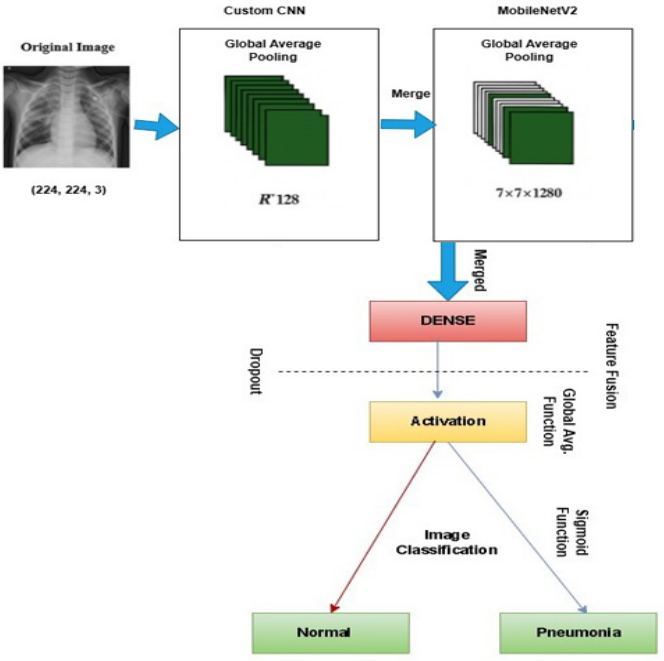


Figure 3: Hybrid Model Diagram for Pneumonia Detection

3.7 Algorithm

The Pneumonia Classification algorithm 3.7 presents an organized pipeline for generating explanatory captions from input images based on a CNN, VGG16 and MobileNetV2 model. It starts by choosing and initializing a pretrained CNN (e.g., VGG16, MobileNetV2) to extract high-level visual features.

Output: Predicted probability of Pneumonia $\hat{y} \in [0, 1]$

Step 1: Initialize input shape $x \in \mathbb{R}^{224\times224\times3}$ and parameters (IMG_SIZE, input_shape)

Step 2: Load dataset and create data generators (train, validation, test) with rescaling and augmentation

Step 3: Define input layer for the model

Step 4: Construct Custom CNN branch: Apply Conv2D (32 filters, 3x3, ReLU) → MaxPooling2D (2x2) Apply Conv2D (64 filters, 3x3, ReLU) → MaxPooling2D (2x2) Apply Conv2D (128 filters, 3x3, ReLU) → GlobalAveragePooling2D

Step 5: Construct MobileNetV2 branch: Load pre-trained MobileNetV2 (include_top=False, weights=imagenet, trainable=False) Apply GlobalAveragePooling2D

Step 6: Fuse features from CNN and MobileNetV2 branches using Concatenate

Step 7: Apply Dense layer (256 units, ReLU) → Dropout (0.5) → Dense (1 unit, Sigmoid)

Step 8: Compile the model with Adam optimizer, binary cross-entropy loss, and accuracy metric

Step 9: Train the model on training and validation data for E epochs

Step 10: Evaluate the model on test data and obtain predictions

Step 11: Compute classification report and confusion matrix; visualize predictions

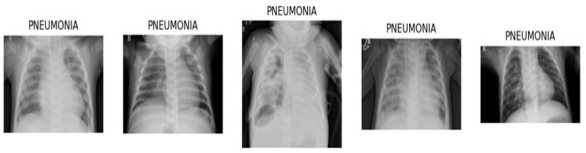
Step 12: Return predicted probability $\hat{y}$

4 Experiment and Result

4.1 Data Description

The dataset used in this study is the Chest X-Ray Images (Pneumonia) dataset, sourced from Kaggle, comprising a total of 5,863 anterior-posterior chest X-ray images divided into two categories: NORMAL and PNEUMONIA. The Guangzhou Women and Children's Medical Center's paediatric patients, ages one to five, provided these pictures. Each image was carefully reviewed for quality, with unreadable scans removed. Two skilled doctors first labelled the diagnoses, and a third expert verified the evaluation set to guarantee high labelling accuracy.

The dataset is organized into three main folders: train, validation (val), and test, each containing subfolders for the two categories. The training set includes 1,342 NORMAL and 3,876 PNEUMONIA images that has been shown in below figure4 as a few samples. The validation set has 9 images per class, and the test set includes 234 NORMAL and 390 PNEUMONIA images. This structure enables proper model training, hyperparameter tuning, and unbiased performance evaluation. For implementation, the dataset directory was accessed using a defined base path. A Python function was developed to count images in each category and subfolder using os.listdir(). This systematic approach ensured a clear understanding of dataset distribution, which is crucial for model training and evaluation balance.



(a) Pneumonia X-ray images

(b) Normal X-ray images

Figure 4: Visualizing Sample Chest X-ray Images from NORMAL and PNEUMONIA Classes (Training Set)

4.2 Data Preprocessing

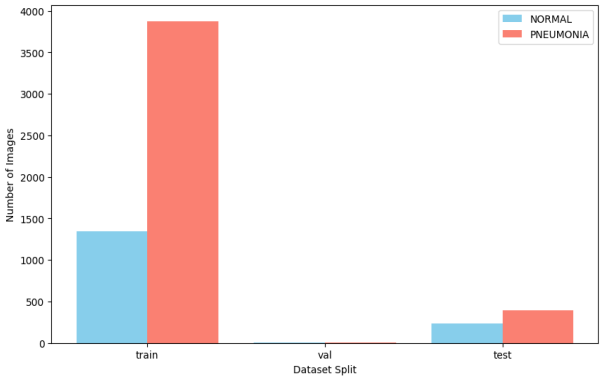
To prepare the chest X-ray dataset for deep learning model training, several preprocessing steps were applied. Firstly, all images were resized to a uniform dimension of 150×150 pixels using the `target_size` parameter in the `ImageDataGenerator`, ensuring consistency in input shape for convolutional neural networks.

Next, normalization was performed by rescaling pixel values from the original range $[0, 255]$ to $[0, 1]$ using the `rescale=1./255` setting. This helps accelerate training convergence and improves model stability. The training set was subjected to data augmentation in order to decrease overfitting and increase dataset diversity. The `ImageDataGenerator` class was used to implement random transformations including rotation ($\pm 15°$), zoom ($\pm 10\%$), and horizontal flipping. These augmentations simulate real-world variations in X-ray images without changing their labels. For the validation and test sets, only normalization was applied to maintain data integrity. The images were then loaded in batches using the `flow_from_directory` method, ready for model

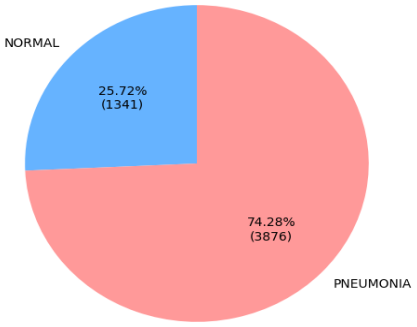
training and evaluation.

4.3 Visualizations

To better understand the structure and characteristics of the chest X-ray dataset, several visualizations were implemented using Python libraries such as Matplotlib, Seaborn, and OpenCV. These visualizations provided insights into data distribution, sample diversity, image resolution, and pixel intensity, all of which are crucial for guiding preprocessing and model design.



(a) Bar Plot for Count per Class in Each Dataset Split



(b) Pie Chart for Class Distribution in the Training Set (NORMAL vs PNEUMONIA)

Figure 5: Data visualization of the split dataset

Firstly, a bar plot in figure5(a) was created to display the number of images in each dataset split (train, validation, test) across the two categories: NORMAL and PNEUMONIA. The data was extracted by counting the images in respective subdirectories. This plot was constructed using plt.bar() and highlighted the class imbalance, especially in the training set, where PNEUMONIA images significantly outnumber NORMAL ones.

Secondly, a pie chart in above figure5(b) was used to show the class distribution within the training set. This helped visualize the proportion of images per class, making the imbalance more apparent. The chart was generated using plt.pie() with custom percentage labels and absolute counts for better interpretability.

For qualitative analysis, sample X-ray images from both classes were visualized. A custom function utilized OpenCV and matplotlib.pyplot to randomly display five images from each class within the training set. This allowed for a visual comparison of pathological versus normal features.

4.4 Results

The convolutional neural network (CNN) designed for pneumonia classification was trained and evaluated using chest X-ray images. Based on the results, the CNN achieved a test accuracy of 87.98%, indicating that the model performs effectively on unseen data.

The below figure7 suggest the confusion matrix further reveals detailed performance across both classes. Out of 234 actual NORMAL cases, 164 were correctly classified while 70 were misclassified as PNEUMONIA. Conversely, out of 390 PNEUMONIA cases, 385 were accurately identified, and only 5 were wrongly predicted as NORMAL. This suggests the model is highly sensitive to detecting pneumonia, exhibiting high recall (0.99) for the PNEUMONIA class but a slightly lower recall (0.70) for the NORMAL class.

The below mentioned table2 represents the classification comparison report that supports this observation. The precision for NORMAL is 0.97, and for PNEUMONIA is 0.85. The F1-scoreâbalancing precision and recallâis 0.81 for NORMAL and 0.91 for PNEUMONIA, resulting in an overall F1-score of 0.87 (weighted average). This indicates the model is more reliable at identifying pneumonia cases than ruling them out.

The training curves show a consistent increase in training accuracy, reaching over 94%, while the validation accuracy fluctuates between 64% and 88%, indicating possible overfitting. The training loss decreases steadily, whereas the validation loss is unstable, peaking around the 5th epoch. This divergence highlights the need for regularization or early stopping.

The binary classification of chest X-ray images into NORMAL and PNEUMONIA categories was accomplished with encouraging results by the VGG16-based transfer learning model. The final test accuracy reached 90.06%, which is a notable

Model	Class	Precision	Recall	F1-score	Accuracy
CNN	NORMAL	0.97	0.70	0.81	0.88
	PNEUMONIA	0.89	0.99	0.91	
	Macro Avg	0.91	0.84	0.86	
	Weighted Avg	0.89	0.88	0.87	
VGG16	NORMAL	0.91	0.81	0.86	0.90
	PNEUMONIA	0.89	0.95	0.92	
	Macro Avg	0.90	0.88	0.89	
	Weighted Avg	0.90	0.90	0.90	
MobileNetV2	NORMAL	0.93	0.66	0.77	0.85
	PNEUMONIA	0.83	0.87	0.89	
	Macro Avg	0.88	0.81	0.83	
	Weighted Avg	0.86	0.85	0.85	

Table 2: Performance comparison of different models on chest X-ray classification

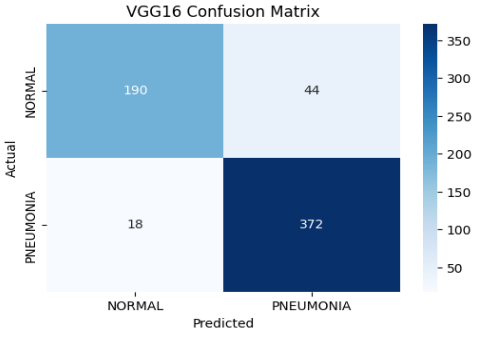
performance considering the modelâs reliance on pre-trained weights from ImageNet and fine-tuning of top layers for pneumonia detection.

The confusion matrix indicates in the below figure6(a) that out of 234 actual NORMAL cases, 190 were correctly classified, and 44 were misclassified as PNEUMONIA. Conversely, the model accurately identified 372 out of 390 PNEUMONIA cases, with only 18 false negatives.

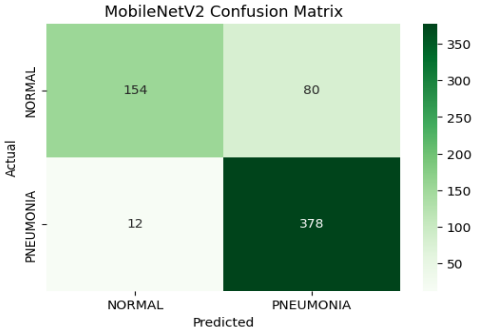
This classification report suggests a high sensitivity in detecting pneumonia (recall of 0.95) and acceptable performance on normal cases (recall of 0.81). The overall precision and F1-scores for both classes further validate the modelâs robustness: NORMAL($F1 = 0.86$), PNEUMONIA ($F1 = 0.92$).

The training and validation curves reveal additional insights in the below figure7,8. The training accuracy steadily increases, nearing 0.95, while validation accuracy shows fluctuations but ultimately converges close to the training line, indicating a good fit by the final epochs. Likewise, training loss continually declines, reaching 0.14, while validation loss shows some oscillations before decreasing, suggesting initial overfitting was reduced with continued training and regularization strategies like dropout.

Chest X-ray pictures were binary classified using the MobileNetV2 model, a lightweight and effective convolutional neural network, in order to identify pneumonia. At 85.26% test accuracy, the model demonstrated a respectably high level of generalisation to unknown data. From the confusion matrix, the model correctly predicted 154 out of 234 NORMAL cases and 378 out of 390 PNEUMONIA cases in below figure6(b). This demonstrates a strong ability to detect pneumonia, though it struggles more with correctly identifying NORMAL cases, as shown by 80 misclassifications.



(a) Confusion matrix of VGG16 model



(b) Confusion matrix of MobileNetV2 model

Figure 6: All models confusion matrix

In terms of classification metrics, the model achieved precision, recall, and F1-scores of 0.93, 0.66, and 0.77 respectively for the NORMAL class, and 0.83, 0.97, and 0.89 for the PNEUMONIA class. The high recall for PNEUMONIA (0.97) is a critical success, as it reflects the modelâs strong ability to minimize false negativesâvital in medical diagnostics. However, the relatively low recall for NORMAL indicates that a notable portion of healthy patients were misclassified as having pneumonia, which could lead to unnecessary concern or treatment. The training and validation accuracy and loss graphs further highlight overfitting.

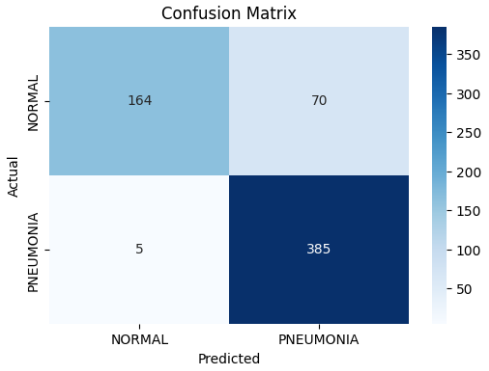
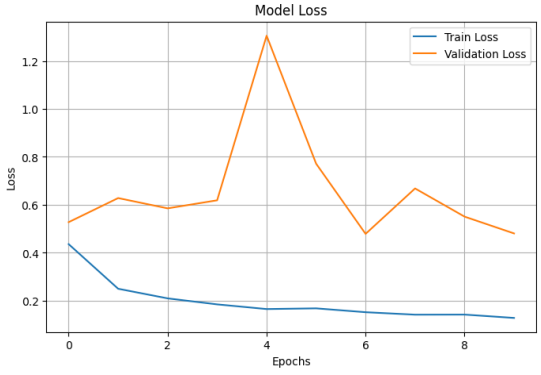
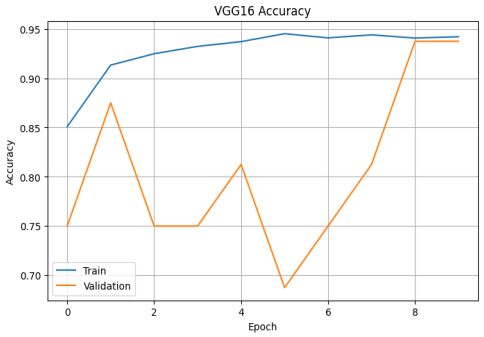


Figure 7: Confusion matrix of CNN model

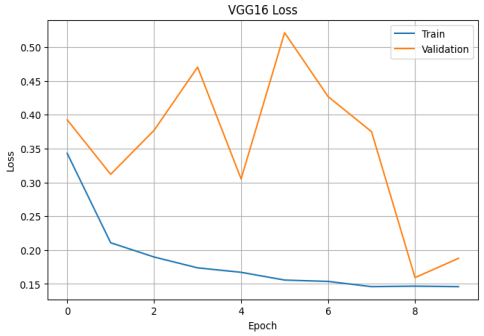
The training accuracy increased steadily, nearing 96%, while validation accuracy showed fluctuation across epochs and remained significantly lower. Similarly, training loss decreased consistently, while validation loss fluctuated, reflecting instability in model generalization.



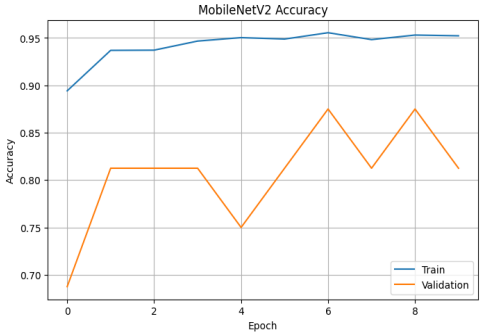
(a) CNN model loss training and validation curves



(b) VGG16 model accuracy training and validation curves



(c) VGG16 model loss training and validation curves



(d) MobileNetV2 model accuracy training and validation curves

Figure 8: All model comparisons on accuracy and loss training and validation curves

A Grad-CAM visualization was applied using the VGG16 model to interpret model decisions on pneumonia detection. The overlay heatmap revealed that the model accurately focuses on relevant thoracic regions, particularly areas around the lungs, which exhibit infection indicators. This supports the modelâs interpretability, confirming that the high performance is not due to spurious correlations but clinically meaningful patterns. The color

intensity in the lungs correlates with areas of abnormal opacity, validating the modelâs attention. This explainability technique enhances trust in the modelâs decisions, especially in sensitive medical applications such as pneumonia diagnosis from chest X-rays shows in figure9.

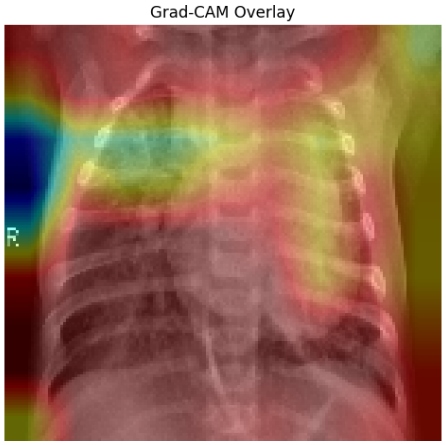
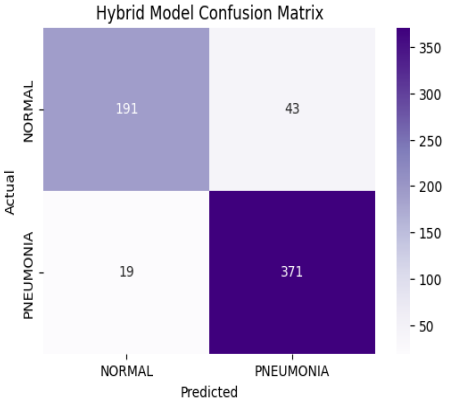


Figure 9: Grad-CAM visualization highlighting pneumonia regions.

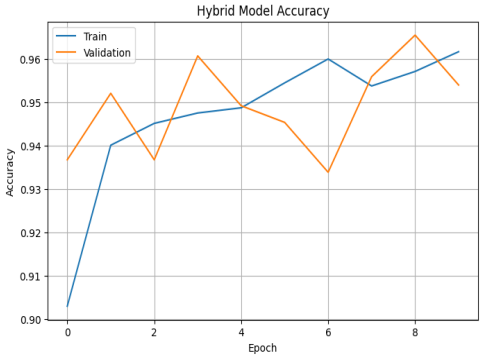
The Hybrid Model, a fusion of traditional Convolutional Neural Network (CNN) and MobileNetV2 architectures, demonstrates robust performance in pneumonia detection tasks.

The training and validation accuracy curves indicate consistent and high classification performance, with validation accuracy peaking at around 96.7% and training accuracy exceeding 96% shown figure10(b) accuracy and figure10(c) shown loss training and validation curves. Furthermore, the loss graphs depict a decreasing trend for both training and validation, converging below 0.15, which reflects good model generalization and minimized overfitting.

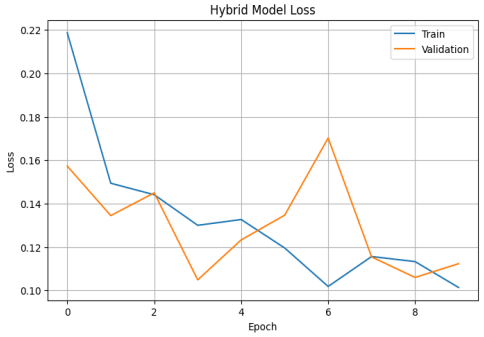
The model achieved a test accuracy of 90.06%, confirming its reliability on unseen data. This result is consistent with the classification report, where NORMAL class achieved a precision of 0.91 and recall of 0.82, while PNEUMONIA class attained a precision of 0.90 and recall of 0.95. The overall F1-scores of 0.86 (NORMAL) and 0.92 (PNEUMONIA) shows below in table3, suggest the model maintains a strong balance between precision and recall across both categories.



(a) Confusion matrix of hybrid model



(b) Accuracy validation curve of hybrid model



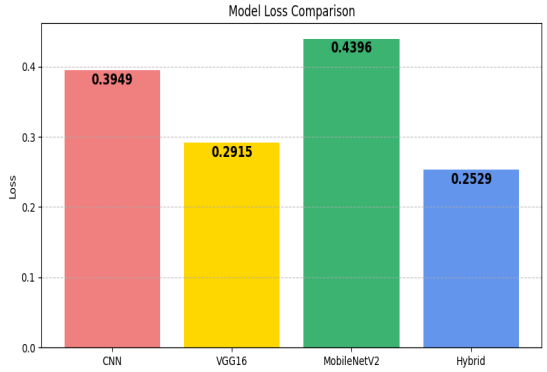
(c) Loss validation curve of hybrid model

Figure 10: Confusion matrix, Accuracy and Loss validation curve of hybrid model

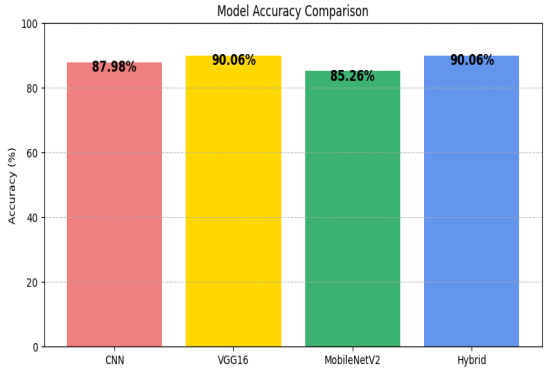
The confusion matrix above figure10(a) supports these findings, with 191 true positives and 43 false negatives for the NORMAL class, and 371 true positives and only 19 false negatives for PNEUMONIA. This highlights the model's particular strength in detecting pneumonia cases correctly, which is critical in clinical diagnostics to reduce the risk of missed diagnoses.

Class	Precision	Recall	F1-score	Support
NORMAL	0.91	0.82	0.86	234
PNEUMONIA	0.90	0.95	0.92	390
Accuracy	0.90			624
Macro Avg	0.90	0.88	0.89	624
Weighted Avg	0.90	0.90	0.90	624

Table 3: Performance Metrics



(a) All models training and validation loss comparison



(b) All models training and validation accuracy comparison

Figure 11: Training and validation loss and accuracy comparison bar graph of all models

The above bar chart figure11 presents a comparative analysis of the test loss values across four models: CNN, VGG16, MobileNetV2, and a Hybrid model combining CNN and MobileNetV2.

From the comparison, the Hybrid model exhibits the lowest loss (0.2529), indicating better generalization. In contrast, MobileNetV2 shows the highest loss (0.4396), suggesting relatively lower performance. VGG16 and CNN models perform moderately, with losses of 0.2915 and 0.3949, respectively.

5 Conclusion

This study evaluated the performance of four deep learning architecturesâCNN, VGG16, MobileNetV2, and a Hybrid model combining CNN and MobileNetV2âfor the detection of pneumonia from chest X-ray images. Comprehensive assessment was conducted using metrics such as accuracy, loss, confusion matrices, and classification reports. Among the models, the Hybrid architecture outperformed others with a test accuracy of 90.06%, the lowest observed loss at 0.2529, and high precision and recall across both classes. VGG16 demonstrated competitive performance, though it slightly lagged in recall and exhibited a marginally higher loss. The CNN model achieved strong recall for pneumonia cases but struggled in classifying normal images accurately. MobileNetV2, while computationally efficient, showed the highest test loss and variability during validation, indicating reduced generalization capability.

To support model interpretability, Grad-CAM visualization was implemented using VGG16, which confirmed that the model accurately focused on key thoracic regions, particularly the lungs, where pneumonia indicators typically appear. This reinforces confidence in the modelâs predictions and adds explainability for potential clinical integration.

However, limitations include the moderate dataset size and possible class imbalance, which may affect model robustness in broader clinical environments. Despite this, the Hybrid model presents a well-rounded, effective solution for automated pneumonia diagnosis.

References

[1] Abdou, M. A. Literature review: Efficient deep neural networks techniques for medical

image analysis. *Neural Computing and Applications*, 34(8):5791–5812, 2022.

[2] Aggarwal, R., Sounderajah, V., Martin, G., Ting, D. S., Karthikesalingam, A., King, D., Ashrafian, H., and Darzi, A. Diagnostic accuracy of deep learning in medical imaging: a systematic review and meta-analysis. *NPJ digital medicine*, 4(1):65, 2021.

[3] Ahmed, S. B., Solis-Oba, R., and Ilie, L. Explainable-ai in automated medical report generation using chest x-ray images. *Applied Sciences*, 12(22):11750, 2022.

[4] Alharbi, A. H. and Hosni Mahmoud, H. A. Pneumonia transfer learning deep learning model from segmented x-rays. In *Healthcare*, volume 10, page 987. MDPI, 2022.

[5] Ali, M., Shahroz, M., Akram, U., Mushtaq, M. F., Altamiranda, S. C., Obregon, S. A., Díez, I. D. L. T., and Ashraf, I. Pneumonia detection using chest radiographs with novel efficientnetv2l model. *IEEE Access*, 2024.

[6] Alshanketi, F., Alharbi, A., Kuruvilla, M., Mahzoon, V., Siddiqui, S. T., Rana, N., and Tahir, A. Pneumonia detection from chest x-ray images using deep learning and transfer learning for imbalanced datasets. *Journal of Imaging Informatics in Medicine*, pages 1–20, 2024.

[7] Alsharif, R., Al-Issa, Y., Alqudah, A. M., Qasmieh, I. A., Mustafa, W. A., and Alquran, H. Pneumonianet: Automated detection and classification of pediatric pneumonia using chest x-ray images and cnn approach. *Electronics*, 10(23):2949, 2021.

[8] Avola, D., Bacciu, A., Cinque, L., Fagioli, A., Marini, M. R., and Taiello, R. Study on transfer learning capabilities for pneumonia classification in chest-x-rays images. *Computer Methods and Programs in Biomedicine*, 221:106833, 2022.

[9] Barakat, N., Awad, M., and Abu-Nabah, B. A. A machine learning approach on chest x-rays for pediatric pneumonia detection. *Digital Health*, 9:20552076231180008, 2023.

[10] Barragán-Montero, A., Javaid, U., Valdés, G., Nguyen, D., Desbordes, P., Macq, B., Willems, S., Vandewinckele, L., Holmström, M., Löfman, F., et al. Artificial intelligence and machine learning for medical imaging: A technology review. *Physica Medica*, 83:242–256, 2021.

[11] Bhati, D., Neha, F., and Amiruzzaman, M. A survey on explainable artificial intelligence (xai) techniques for visualizing deep learning models in medical imaging. *Journal of Imaging*, 10(10):239, 2024.

[12] Canbek, G., Taskaya Temizel, T., and Sagiroglu, S. Ptopi: A comprehensive review, analysis, and knowledge representation of binary classification performance measures/metrics. *SN Computer Science*, 4(1):13, 2022.

[13] Chiwariro, R. and Wosowei, J. B. Comparative analysis of deep learning convolutional neural networks based on transfer learning for pneumonia detection. *Int. J. Res. Appl. Sci. Eng. Technol*, 11(1):1161–1170, 2023.

[14] Colin, J. and Surantha, N. Interpretable deep learning for pneumonia detection using chest x-ray images. *Information*, 16(1):53, 2025.

[15] Gu, C. and Lee, M. Deep transfer learning using real-world image features for medical image classification, with a case study on pneumonia x-ray images. *Bioengineering*, 11(4):406, 2024.

[16] Guan, H. and Liu, M. Domain adaptation for medical image analysis: a survey. *IEEE Transactions on Biomedical Engineering*, 69(3):1173–1185, 2021.

[17] Hasani, N., Farhadi, F., Morris, M. A., Nikpanah, M., Rhamim, A., Xu, Y., Pariser, A., Collins, M. T., Summers, R. M., Jones, E., et al. Artificial intelligence in medical imaging and its impact on the rare disease community: threats, challenges and opportunities. *PET clinics*, 17(1):13, 2022.

[18] Hosseinzadeh Taher, M. R., Haghighi, F., Feng, R., Gotway, M. B., and Liang, J. A systematic benchmarking analysis of transfer learning for medical image analysis. In *Domain Adaptation and Representation Transfer, and Affordable Healthcare and AI for Resource Diverse Global Health: Third MICCAI Workshop, DART 2021, and First MIC-*

CAI Workshop, FAIR 2021, Held in Conjunction with MICCAI 2021, Strasbourg, France, September 27 and October 1, 2021, Proceedings 3, pages 3–13. Springer, 2021.

[19] Ibrahim, A. U., Ozsoz, M., Serte, S., Al-Turjman, F., and Yakoi, P. S. Pneumonia classification using deep learning from chest x-ray images during covid-19. *Cognitive computation*, 16(4):1589–1601, 2024.

[20] 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*, pages 1–45, 2024.

[21] Lavazza, L. and Morasca, S. Common problems with the usage of f-measure and accuracy metrics in medical research. *IEEE Access*, 11:51515–51526, 2023.

[22] Malik, H., Anees, T., Chaudhry, M. U., Gono, R., Jasiński, M., Leonowicz, Z., and Bernat, P. A novel fusion model of hand-crafted features with deep convolutional neural networks for classification of several chest diseases using x-ray images. *IEEE Access*, 11:39243–39268, 2023.

[23] Müller, D., Soto-Rey, I., and Kramer, F. Towards a guideline for evaluation metrics in medical image segmentation. *BMC Research Notes*, 15(1):210, 2022.

[24] Musha, A., Al Mamun, A., Tahabilder, A., Hossen, M. J., Jahan, B., and Ranjbari, S. A deep learning approach for covid-19 and pneumonia detection from chest x-ray images. *International Journal of Electrical & Computer Engineering (2088-8708)*, 12(4), 2022.

[25] Neshat, M., Ahmed, M., Askari, H., Thilakaratne, M., and Mirjalili, S. Hybrid inception architecture with residual connection: fine-tuned inception-resnet deep learning model for lung inflammation diagnosis from chest radiographs. *Procedia Computer Science*, 235:1841–1850, 2024.

[26] Nirthika, R., Manivannan, S., Ramanan, A., and Wang, R. Pooling in convolutional neural networks for medical image analysis: a survey and an empirical study. *Neural Computing and Applications*, 34(7):5321–5347, 2022.

[27] Radočaj, P. and Martinović, G. Interpretable deep learning for pediatric pneumonia diagnosis through multi-phase feature learning and activation patterns. *Electronics*, 14(9):1899, 2025.

[28] 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*, 93:106130, 2024.

[29] Sathyanarayanan, S. and Tantri, B. R. Confusion matrix-based performance evaluation metrics. *African Journal of Biomedical Research*, pages 4023–4031, 2024.

[30] Sharma, S. and Guleria, K. A deep learning based model for the detection of pneumonia from chest x-ray images using vgg-16 and neural networks. *Procedia Computer Science*, 218:357–366, 2023.

[31] Shilpa, N., Banu, W. A., and Metre, P. B. Revolutionizing pneumonia diagnosis: Ai-driven deep learning framework for automated detection from chest x-rays. *IEEE Access*, 2024.

[32] Souid, A., Sakli, N., and Sakli, H. Classification and predictions of lung diseases from chest x-rays using mobilenet v2. *Applied Sciences*, 11(6):2751, 2021.

[33] Sufian, M. A., Hamzi, W., Sharifi, T., Zaman, S., Alsadder, L., Lee, E., Hakim, A., and Hamzi, B. Ai-driven thoracic x-ray diagnostics: Transformative transfer learning for clinical validation in pulmonary radiography. *Journal of Personalized Medicine*, 14(8):856, 2024.

[34] Suganyadevi, S., Seethalakshmi, V., and Balasamy, K. A review on deep learning in medical image analysis. *International Journal of Multimedia Information Retrieval*, 11(1):19–38, 2022.

[35] Thakur, P. S., Sheorey, T., and Ojha, A. Vgg-icnn: A lightweight cnn model for crop disease identification. *Multimedia Tools and Applications*, 82(1):497–520, 2023.

[36] Thanapol, P., Lavangnananda, K., Bouvry, P., Pinel, F., and Leprévost, F. Reducing overfitting and improving generalization in training convolutional neural network (cnn) under

limited sample sizes in image recognition. In *2020-5th International Conference on Information Technology (InCIT)*, pages 300–305. IEEE, 2020.

[37] Tripathi, A., Singh, T., Nair, R. R., and Duraisamy, P. Improving early detection and classification of lung diseases with innovative mobilenetv2 framework. *IEEE Access*, 2024.

[38] Tripathi, M. Analysis of convolutional neural network based image classification techniques. *Journal of Innovative Image Processing (JIIP)*, 3(02):100–117, 2021.

[39] Tsuneki, M. Deep learning models in medical image analysis. *Journal of Oral Biosciences*, 64(3):312–320, 2022.

[40] Wang, R., Lei, T., Cui, R., Zhang, B., Meng, H., and Nandi, A. K. Medical image segmentation using deep learning: A survey. *IET image processing*, 16(5):1243–1267, 2022.

[41] Zhu, W., Qiu, P., Chen, X., Li, H., Wang, H., Lepore, N., Dumitrascu, O. M., and Wang, Y. Beyond mobilenet: An improved mobilenet for retinal diseases. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 56–65. Springer, 2023.