

*Research Article*

PneumaScan: A Two-Phase Fine-Tuning Framework using DenseNet121 for Pneumonia Detection

Sanjit Kumar Dash¹, Bimalendu Nanda¹, Mohammed Altaf Ahmed^{2*}, Qutubuddin Mohammed²

¹Department of Information Technology, Odisha University of Technology and Research, Bhubaneswar, Odisha, 751029, India

²Department of Computer Engineering, College of Computer Engineering and Sciences, Prince Sattam bin Abdulaziz University, Al-Kharj, 11942, Saudi Arabia

*Corresponding author: m.ataf@psau.edu.sa; Tel.: +966-5888347; Fax.: +966-5888347

Abstract: Lower respiratory tract infections, primarily pneumonia, are the leading infectious cause of death worldwide, accounting for almost 2.2 million deaths every year, including more than 700,000 children below five years old. Although these conditions are preventable and treatable, they affect approximately 450 million people globally each year, with most of those dying belonging to disadvantaged groups or living in places with few healthcare resources. A key issue in tackling this crisis involves getting accurate and prompt diagnostic evaluations, which is still largely reliant on expert readings of chest X-rays, which is a complicated method due to the enormous number of cases and the extreme lack of radiologists. Overcoming this obstacle requires the incorporation of intelligent automated solutions to help clinicians make better decisions. This paper proposes an improved artificial intelligence (AI)-driven system using deep learning-based convolutional neural networks to identify pneumonia from chest X-ray images. DenseNet121, a connectivity-heavy convolutional model allowing smooth transmission and gathering of layered features due to its specialized weaving design, is used in this study. After undergoing strict training and verification steps using a carefully sorted collection of X-ray images, the model showed impressive skill in identifying lung abnormalities versus regular structures. The performance assessments showed outstanding outcomes, with a macroaverage precision of 93.72% and an overall accuracy of 93.42% on separate test inputs. These scores indicate the solid real-world applicability of the system for practical application in mechanical pneumonia screening setups. The proposed model holds major promise for streamlining diagnostic procedures, speeding up case discovery, and aiding quicker treatment steps, especially in medical areas where specialists remain scarce.

Keywords: Deep learning; DenseNet; Feature extraction; Image processing; Pneumonia detection

1. Introduction

Pneumonia is an inflammatory disease that predominantly affects the alveolar air sacs (McLuckie, 2009) and is characterized by productive or dry cough, chest pain, pyrexia, and respiratory distress. The etiological spectrum encompasses bacterial, viral, and fungal pathogens, with more than 30 distinct causative agents identified in clinical practice (World Health Organization, 2008). The pathophysiological mechanism involves the disruption of the intricate equilibrium between commensal microorganisms residing in the lower respiratory tract and both innate and acquired immune defense systems, culminating in parenchymal inflammation. Vulnerable demographic cohorts, including paediatric patients below five years, geriatric populations, and individuals with underlying comorbidities, such as diabetes mellitus or cardiovascular disorders, demonstrate heightened susceptibility to severe pneumonic complications. Disease burden

remains disproportionately elevated in resource-limited settings, intensified by environmental pollutants and inadequate healthcare infrastructure accessibility.

Conventional diagnostic methodologies, while clinically established, are constrained by inherent limitations, including resource intensity, temporal inefficiency, and critical dependence on the availability of specialized radiological expertise (Hofmeister et al., 2024). Traditional radiographic interpretation protocols often exhibit suboptimal turnaround times and are susceptible to inter-observer variability, thereby compromising diagnostic accuracy. The fundamental imperative for rapid and precise diagnostic assessment becomes particularly acute given that therapeutic efficacy demonstrates a strong correlation with the timeline of early intervention.

Recent years have witnessed a paradigm shift in medical image analysis with the emergence of deep learning (Schulz & Behnke, 2012) methodologies, wherein deep convolutional neural networks have demonstrated exceptional proficiency in image classification, segmentation, and hierarchical feature extraction tasks. Computer vision-driven (Khan & Al-Habsi, 2020) automated detection algorithms have proliferated extensively within medical imaging research domains, with deep learning architectures enabling substantial improvements in diagnostic precision. These computational approaches facilitate automated pattern recognition through hierarchical feature mapping (Misra & Laskar, 2019), where convolutional neural network (CNN) architectures extract salient characteristics from extensive radiographic datasets and construct discriminative representations via interconnected neuronal pathways. Such methodologies have demonstrated translational potential across diverse applications (Aly & Almotairi, 2020; Ghanim et al., 2020; Kakarla et al., 2020; Prakash et al., 2019; Vashist et al., 2020; Zhou et al., 2019), including handwriting recognition systems, facial biometric identification, autonomous vehicular navigation, surveillance frameworks, and medical image analysis platforms. The integration of AI with medical imaging has catalyzed novel methodologies for automated pneumonia detection, with pre-trained architectures demonstrating considerable efficacy in medical imaging classification tasks.

Despite substantial advancements in the deployment of convolutional neural networks (O'Shea & Nash, 2015) for pneumonia detection from chest radiographic images, several persistent challenges remain inadequately addressed. Medical imaging datasets typically suffer from severe class imbalance issues, where pathological cases represent only a minor proportion of available samples, consequently biasing model predictions toward majority classes (Siddiqi & Javaid, 2024). Numerous existing architectures, particularly rudimentary custom CNN implementations, exhibit significant overfitting tendencies when applied to medical imaging repositories' limited, imbalanced datasets (An et al., 2024). These models frequently demonstrate compromised generalization capabilities across heterogeneous patient demographics and variable imaging acquisition parameters, attributable to suboptimal transfer learning optimization strategies (Hasan et al., 2024). Inadequate consideration of strategic layer freezing protocols frequently results in architectures that either underperform or overfit training distributions. Data imbalance, characterized by substantial numerical disparity between normal and pneumonia-positive cases, is frequently neglected, precipitating biased predictive outputs. Additionally, high-performing models often exhibit impractical computational intensity for deployment in resource-constrained clinical environments, necessitating the development of accurate yet computationally efficient diagnostic solutions.

Given these identified limitations, the research problem is specifically expressed as follows: How can pneumonia detection frameworks achieve competitive classification accuracy while addressing architectural fragmentation, overfitting tendencies, and computational inefficiency? To address these challenges, this study implements a unified DenseNet121-based architecture that exploits dense connectivity patterns to enable efficient end-to-end learning. The framework eliminates the need for external feature selection algorithms and heterogeneous ensemble frameworks that introduce architectural discontinuity, as demonstrated in preliminary experimental phases (Section 4.3). This implementation achieves three key objectives: (i) reducing parametric complexity relative to VGG16 (7.98M vs. 138M parameters), (ii) facilitating robust gradient

flow through dense connection pathways, and (iii) preserving learned representations from ImageNet pre-training while enabling strategic layer unfreezing for task-specific adaptation. The experimental validation of these design choices is detailed in the following sections.

Consequently, this study aims to address these identified research gaps through systematic experimentation and rigorous comparative analysis. Multiple architectural approaches were evaluated in a sequential experimental framework to bridge the existing lacunae in current methodologies. In the preliminary experimental phase, two distinct VGG16-based (Simonyan & Zisserman, 2015) implementations were employed for pneumonia detection: the first utilized fine-tuned deep learning protocols with selective layer unfreezing, while the second adopted VGG16 as a fixed feature extractor integrated with feature selection via the Voting Differential Evolution (VDE) algorithm, subsequently coupled with a stacking ensemble classification framework comprising K-Nearest Neighbors, Support Vector Machines, and Light Gradient Boosting Machine classifiers (Jung & Kim, 2023; Ke et al., 2017; Zhang, 2016).

The initial custom-fine-tuned VGG16 model was operationalized by adapting the pre-trained architecture to the pneumonia radiographic dataset using transfer learning mechanisms. Although this implementation demonstrated commendable accuracy on the training partition, substantial overfitting manifestations emerged, as evidenced by pronounced disparities between training and validation performance indices. This phenomenon is primarily attributable to the architectural depth and parametric complexity of VGG16, which has approximately 138 million trainable parameters. When applied to relatively constrained medical imaging datasets, VGG16 exhibits a propensity toward memorizing dataset-specific patterns rather than acquiring generalized feature representations applicable to unseen clinical cases.

An alternative hybrid methodology was developed to mitigate these overfitting tendencies, where VGG16 functioned exclusively as a feature extraction module. This approach incorporated feature selection via the Voting Differential Evolution algorithm, which employs evolutionary binary mask optimization to identify discriminative feature subsets, followed by classification through a heterogeneous stacked ensemble with logistic regression as the meta-learner (Faouzi & Colliot, 2023). Although VDE demonstrated efficacy in reducing feature dimensionality, the performance of the ensemble pipeline remained suboptimal due to residual noise in VGG16-extracted features and architectural fragmentation, wherein discrete components operated independently rather than as a cohesive end-to-end trainable system, preventing holistic optimization and seamless gradient flow.

A comprehensive evaluation revealed fundamental limitations in VGG16-based methodologies, including overfitting tendencies, disjointed optimization, inadequate feature reusability, and compromised generalization. Therefore, the research trajectory was strategically redirected toward DenseNet-based architectures. DenseNet's dense layer connectivity, where each layer receives feature maps from all preceding layers, promotes exceptional feature reuse efficiency and facilitates direct gradient flow, enabling deeper yet more compact networks. The implementation leveraged DenseNet121 (Huang et al., 2017), selected over deeper variants due to superior suitability for limited medical imaging datasets, achieving optimal equilibrium between representational capacity and generalization capability. Strategic layer freezing protocols preserved robust low-level representations while selective fine-tuning of higher layers facilitated adaptation to pneumonia-specific features. This unified, end-to-end differentiable architecture eliminated the computational overhead associated with VDE-driven pipelines and ensemble frameworks, yielding substantial improvements in classification performance, training convergence dynamics, and robustness across heterogeneous radiographic characteristics, as detailed in the following sections.

The primary objective of this research paper is to develop a robust and accurate pneumonia detection model using the DenseNet121 model based on CNN, targeting an accurate result with minimal false negatives and false positives. The model aims to assist radiologists by serving as an advanced decision-support system that improves diagnostic efficiency and reliability. To ensure practical applicability, the model is generalizable across diverse datasets, varying image

qualities, and real-world clinical environments.

2. Related Work

Automated pneumonia detection from chest X-ray images has experienced significant progress through deep learning methodologies, with researchers investigating diverse architectural approaches ranging from lightweight convolutional neural networks to advanced transformer-based models. Contemporary studies have integrated transfer learning, feature selection algorithms, ensemble techniques, and domain adaptation strategies to enhance diagnostic accuracy and computational efficiency. Nevertheless, critical limitations persist, including restricted dataset generalizability, substantial computational requirements, binary classification constraints, and insufficient clinical adoption interpretability. This section systematically reviews state-of-the-art approaches and analyzes their architectural contributions, empirical achievements, and inherent limitations that inform the development of the proposed framework.

Ren et al. (2021) proposed MulNet, integrating DenseNet121 for chest X-ray analysis with a Bayesian network incorporating seven clinical manifestations (cough, hemoptysis, chest pain, fever, dyspnea, wet rales, and dry rales). The framework employs a constraint-based MGS algorithm to construct medically informed network structures, validated on 35,389 cases of community-acquired pneumonia. MulNet achieved 0.87 AUC, 0.94 recall, and 0.82 F1-score, demonstrating a 3.6% AUC improvement over image-only models while providing conditional probability tables for causal interpretation. However, its limitations include binary classification without pathogen differentiation, exclusive DenseNet121 reliance, restricted seven-symptom scope, and absence of external validation across diverse healthcare settings and patient demographics.

Jiang et al. (2021) developed IVGG13 by compressing VGG16 from 16 to 13 convolutional layers and trimming parameters from 17.97M to 2.52M via restructured blocks with 32-64-128-64 filters. Tested on Kaggle Chest X-Ray Images with augmentation, IVGG13 delivered 89.1% accuracy and 90.0% F1-measure, exceeding LeNet, AlexNet, and GoogLeNet. The drawbacks include binary scope, expensive training, and absent feature selection or interpretability for clinical contexts.

Szepesi and Szilágyi (2022) designed a five-block CNN with 40% dropout in convolutional layers, using 10.6M parameters—far fewer than InceptionV3 (26.2M), ResNet50 (24.8M), or VGG-19 (145.2M). Trained on 5,856 pediatric chest X-rays, the model achieved 97.21% accuracy, 97.40% precision, 0.982 AUC, and 122 ms inference time, surpassing the transfer learning alternatives. Weaknesses include pediatric-only data reducing applicability, binary classification scope, X-ray-only inputs, missing explainability, and unresolved deployment issues such as reproducibility and confounding variables.

Das et al. (2024) presented CNN-based pneumonia identification at ASSIC 2024, pairing CNN feature extraction with classifiers such as SVM, linear regression, and decision tree for binary radiograph classification. SVM topped performance at 96.4% accuracy, validating deep CNN representations. Documentation gaps exist around architecture specifics, dataset details, training procedures, and full metrics (precision, recall, and F1-score), hindering comparison with modern methods and clinical feasibility assessment.

Ali et al. (2024) applied EfficientNetV2L via compound scaling and neural architecture search with separable depthwise convolutions. Six models (CNN, InceptionResNetV2, Xception, VGG16, ResNet50, and EfficientNetV2L) were trained using the Adam optimizer and ImageDataGenerator augmentation. Using 5,856 pediatric X-rays (1,583 normal, 4,273 pneumonia, ages 1–5) with 80-20 split, EfficientNetV2L achieved 94.02% accuracy, 94.40% precision, 97.24% recall, 95.80% F1-score, beating CNN (87.78%), InceptionResNetV2 (88.94%), Xception (90.7%), VGG16 (91.66%), ResNet50 (87.98%). The ROC-AUC and 10-fold validation confirmed the stability. Problems include pediatric-only data restricting demographic and equipment range, binary scope, overfitting risk from small dataset, no multi-modal integration, and absent interpretability reducing transparency.

Guddati et al. (2024) implemented a 10-layer DCNN on FPGA hardware for detecting tuberculosis and pneumonia detection in resource-limited settings. Trained on Google Cloud CPU, the model converts to RTL via VIVADO synthesis with HLS4ML using fixed-point format (8-bit integer, 12-bit fractional) for PYNQ-Z2 FPGA. The system reached 96.39% (CPU) and 95.63% (FPGA) accuracy on 422 images with 85.19% faster inference with just 1,831 parameters—far below transfer learning and standard CNNs. Weaknesses include a small dataset (422 images), limited scope, lack of comparison with recent methods, missing clinical validation, and unexamined power use, hardware resources, and scalability.

Ouyang et al. (2025) designed a two-phase method that paired VGG16 feature extraction (2,048 dimensions from $80 \times 80 \times 3$ inputs) with voting differential evolution using dimensional adaptive search, fitness-driven crossover, and soft voting. The results showed 99.04% precision, 98.67% accuracy, 99.13% recall, and 19.93% feature reduction, beating the six meta-heuristic baselines. The drawbacks include the 9-hour runtime, VGG16 reliance, binary scope, and wrapper selection, which block real-time use.

Fu et al. (2025) created a DAViT, merging vision transformers with shallow CNNs and domain adaptation to overcome CNN receptive field limits and ImageNet pre-training mismatch. DAViT uses ViT self-attention for global features and shallow CNN for local patterns, with domain adaptation on 108,948 ChestX-ray8 images. Tested on 5,856 X-rays, DAViT scored 97% F1 and 96% AUC for pneumonia, plus 81% F1 and 84% AUC for viral/bacterial classification, topping 12 baselines. Ablation tests showed that pre-training added 12%, CNN 1%, and domain adaptation 0.4%. Grad-CAM heatmaps visualized the infection zones. Weaknesses include pediatric-only data (ages 1–5) limiting scope, high computational cost with long inference and FLOPs, ML expertise needs, and missing regulatory compliance.

Munir et al. (2025) built PneuX-Net, an ensemble framework that combines random forest, Gaussian Naïve Bayes, and K-Nearest Centroid at the feature and decision levels. Sception extracts spatial features, RF captures nonlinear patterns, GNB adds Bayesian reasoning, and KNC applies proximity clustering. Validated on 5,856 pediatric X-rays with 10-fold cross-validation, PneuX-Net+KNC achieved 99.91% accuracy in 0.0054 seconds, surpassing the accuracy of traditional methods: GNB (52%), LR (73%), SGD (73%), MLP (74%), KNC (81.5%), and RF (92%). Gaps include binary scope without subtype analysis, single pediatric dataset, no deep learning integration, possibly limiting abstraction, and missing interpretability for clinical trust or approval.

Existing pneumonia detection methodologies have demonstrated substantial progress in classification accuracy; however, critical limitations remain. TL approaches rely predominantly on single-source pediatric datasets, restricting demographic and geographical generalizability. Although architectures such as VGGNet, EfficientNetV2L, and Vision Transformers achieve competitive performance, they often incur substantial computational overhead, excessive parameter counts, and prolonged training times, limiting their suitability for real-time clinical deployment. Current research is dominated by binary classification paradigms, with limited investigation into bacterial–viral differentiation essential for targeted treatment. Furthermore, feature selection mechanisms are frequently computationally expensive or entirely absent, while interpretability through explainable AI techniques remains insufficiently addressed for clinical trust and regulatory compliance. Additionally, most studies lack cross-dataset validation and comprehensive deployment analysis encompassing hardware constraints and clinical workflow integration.

In our earlier work (Dash et al., 2026), a pneumonia detection framework was developed using a VGG16-based CNN combined with image preprocessing, data augmentation, and deep feature extraction from chest X-ray images. The model incorporated feature selection using the Voting Differential Evolution (VDE) algorithm to enhance classification performance, achieving high accuracy and area under the curve (AUC) on benchmark datasets. Although the results were promising, the proposed approach exhibited several inherent limitations. The VGG16 architecture, which is characterized by several trainable parameters, demonstrated overfitting

tendencies when applied to limited and imbalanced medical imaging datasets. Additionally, the reliance on external feature selection introduced architectural fragmentation, preventing unified end-to-end optimization and increasing computational overhead. These constraints limited feature reusability, training efficiency, and generalization capability across heterogeneous radiographic conditions. Motivated by these observations, the present study advances beyond the earlier VGG16-based pipeline by adopting a densely connected DenseNet121-based architecture that enables efficient feature reuse, improved gradient propagation, and integrated end-to-end learning, thereby addressing the identified shortcomings while enhancing robustness, computational efficiency, and clinical applicability across diverse patient populations.

3. Proposed Framework

The proposed framework (Figure 1) implements a multi-step method for identifying pneumonia with chest X-ray images. First, data collected from the dataset are pre-processed and augmented by adjusting the image resolution and normalizing the images to improve the model performance. The model is based on DenseNet121, with custom classification layers added to the pre-trained backbone, followed by fine-tuning of selected backbone layers using the training data. The set of X-ray images is examined by the CNN classifier, which then classifies the images into two groups: pneumonia and normal cases. The developed system is extensively tested, and the most important of the multiple evaluation criteria that were set, namely, precision, accuracy, recall, and F1 score, are the focus to confirm that the developed model provides an accurate and efficient diagnosis.

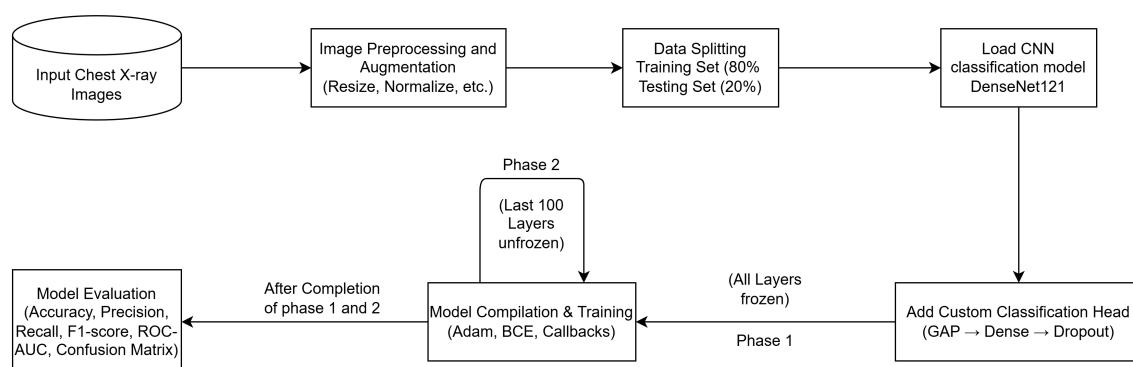


Figure 1 The model framework. The framework follows a two-phase training strategy. In

Phase 1, the pre-trained DenseNet121 CNN backbone is kept frozen while the custom classification layers are trained. After completion of Phase 1, Phase 2 is initiated by unfreezing the last 100 backbone layers for fine-tuning. Upon completion of both phases, the trained model proceeds to the final evaluation using the specified performance metrics

4. Methodology

4.1 Dataset Description

This study employed the pediatric chest radiography dataset (Mooney, 2018) originally compiled by Kermany et al., sourced from the Guangzhou Women and Children's Medical Center, China, and made publicly accessible via the Kaggle platform. The repository encompasses a total of 6,349 anterior-posterior chest X-ray images acquired from pediatric subjects within the one-to-five-year age demographic, obtained retrospectively through standard clinical care procedures. The dataset architecture features a hierarchical organization into three distinct partitions: a training subset containing 5,217 radiographs, a validation subset with 8 images, and a testing subset comprising 1,124 images. Each radiograph is categorized into one of two

mutually exclusive diagnostic classes: NORMAL or PNEUMONIA.

During dataset curation, rigorous quality assurance protocols were checked, involving systematic elimination of radiographs exhibiting poor quality or insufficient diagnostic clarity. Qualified medical experts assigned diagnostic annotations, with the testing partition undergoing an additional independent verification procedure by a third specialist to ensure annotation reliability and diagnostic precision. A notable characteristic of this dataset is the presence of class distribution asymmetry, wherein the training partition demonstrates 1,341 NORMAL cases against 3,876 PNEUMONIA cases, while the testing partition contains 434 NORMAL instances compared to 690 PNEUMONIA instances. The dataset is distributed under the Creative Commons Attribution 4.0 International licensing framework, thereby enabling research reproducibility and facilitating standardized performance benchmarking across the scientific community.

4.2 Preprocessing and Augmentation

The radiographic dataset, which is partitioned into 5,217 training samples, 8 validation samples, and 1,124 testing samples distributed across two diagnostic classifications (NORMAL and PNEUMONIA), is subjected to a comprehensive preprocessing pipeline to establish architectural compatibility with DenseNet121 input requirements. Initially, spatial standardization is performed in which all radiographic images are transformed into consistent 224×224 -pixel dimensions, conforming to the DenseNet121 network architecture's canonical input specifications. The original grayscale images, encoded as 8-bit integer intensity values, are normalized by multiplicative scaling by $1/255$, thereby converting directly to floating-point representations within the $[0,1]$ interval. This intensity rescaling operation serves two purposes: to maintain numerical stability (Litjens et al., 2017) throughout gradient descent optimization procedures and to facilitate accelerated convergence dynamics during the training phase. Addressing the substantial class distribution disparity observed within the training partition—specifically, 1,341 NORMAL specimens versus 3,876 PNEUMONIA specimens—necessitates the implementation of augmentation strategies (Shorten & Khoshgoftaar, 2019) to enhance model robustness and mitigate overfitting risks. Real-time data augmentation is exclusively operationalized on the training subset using TensorFlow's ImageDataGenerator infrastructure, incorporating multiple stochastic geometric transformations.

The normalization operation is formally denoted as Equation (1):

$$\mathbf{I}_{normalized} = (\mathbf{I}_{original} - \mu) / \sigma \quad (1)$$

where μ and σ represent pixel intensity mean and standard deviation, respectively, which ensure numerical stability during backpropagation and accelerate convergence dynamics by constraining activation values to zero-centered distributions. This preprocessing step is critical for transfer learning protocols because it aligns the input distribution of the pneumonia dataset with ImageNet-derived features encoded in pre-trained DenseNet121 weights.

Figure 2 shows the pixel intensity distribution of X-ray images before and after augmentation. The original dataset (blue) has a high concentration of approximately 200, and the augmented dataset (pink) has a broader distribution, resulting in more feature diversity. These transformations encompass: angular rotations within $\pm 10^\circ$ bounds, translational shifts along both horizontal and vertical axes (spanning 10% of respective image dimensions), shear distortions (magnitude = 0.1), magnification variations (range = 0.1), and horizontal reflection operations. This augmentation protocol effectively simulates the natural variability inherent to clinical radiographic acquisition, including variations in patient positioning and equipment-related inconsistencies.

Crucially, the validation and test subsets are processed through an identical preprocessing sequence—encompassing dimensional standardization and intensity normalization—while deliberately excluding all augmentation operations. This methodological distinction ensures unbiased performance evaluation on unmodified clinical radiographs, thereby providing an authentic as-

assessment of diagnostic efficacy that accurately reflects anticipated performance under genuine clinical deployment scenarios.

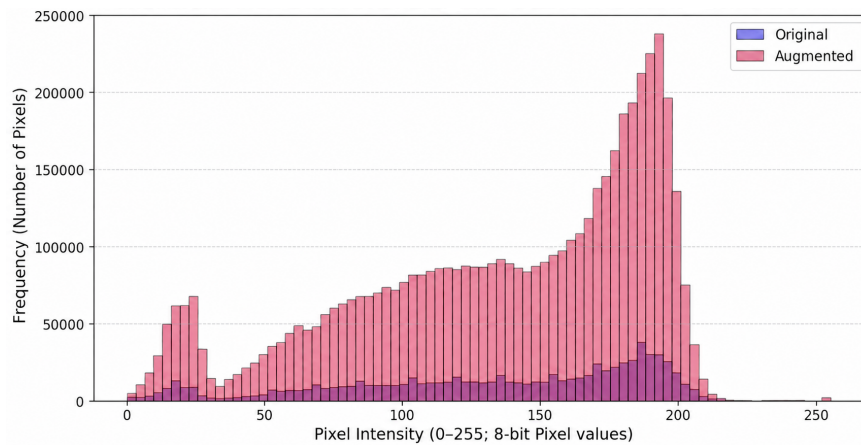


Figure 2 Pixel intensity distribution before and after data augmentation

4.3 A comparative study to finalize model selection

In this study, three approaches were implemented to make a robust decision in selecting the best model. First, a deep learning VGG16 model was used, and the following outcomes were observed: In this study, the VGG16-based model used for pneumonia classification implemented a transfer learning strategy, in which the pre-trained convolutional layers were initially frozen to preserve ImageNet-learned weights. To enhance regularization, the architecture was extended by applying global average pooling, batch normalization, and dropout layers (with rates of 0.5 and 0.3). Subsequently, the fifth convolutional block (layer 5) was unfrozen to enable fine-tuning during the second training phase, allowing the model to better adapt its deeper feature representations to the characteristics of chest X-ray images. The training process, which was optimized using the Adam optimizer, resulted in a high training accuracy of 98.72%, whereas the validation accuracy plateaued at approximately 92.63%, indicating high overfitting. Evaluation using the classification report method showed a precision of 0.9273, recall of 0.9590, and F1-score of 0.9421 on the test set. This divergence between training and validation accuracy can be attributed to the high capacitive nature of VGG16, which has approximately 138 million parameters and thus can easily overfit small or noisy data. Additionally, the model's performance was affected by potential data quality issues, including mislabeled data, which can mislead the learning process. VGG16 has also suffered severely due to its architectural limitations, such as the lack of feature reusability, absence of residual connections, and high computational needs, making it less efficient and more inclined toward overfitting compared to more contemporary models like DenseNet.

Then, an ensemble technique with the VGG16 feature extractor and the VDE algorithm is implemented. This model uses a hybrid approach that combines deep feature extraction from VGG16, feature selection via voting differential evolution (VDE), and stacked ensembling with SVM, KNN, and LightGBM, aggregating into a logistic regression meta-learner. Although this architecture enhances interpretability and attempts to counteract high-dimensional feature noise, its actual performance highlights challenges. The training accuracy was 75.07%, and the validation accuracy was slightly lower at 72.41%, indicating moderate generalization ability but no severe overfitting. However, the class imbalance is starkly evident: Class 0 (likely "Normal") suffers from a precision of 0.21 and recall of 0.03, indicating near-total misclassification, whereas class 1 (likely "Pneumonia") dominates performance metrics. Although innovative, the VDE algorithm is sensitive to such noise. Although conceptually powerful, the stacking ensemble inherits biases from individual models and fails to mitigate class-wise performance disparity.

From the above testing, it can be concluded that VGG16 is good enough but needs a better

labeled and less noisy dataset. Therefore, to carry on testing, the DenseNet121 CNN model is considered.

And among the 2 available model 121 and 201, DenseNet121 is selected because this added complexity can lead to overfitting on a relatively small or noisy dataset like medical chest X-rays. With fewer parameters, DenseNet121 strikes a better balance between capacity and generalization, especially when the dataset is too small to justify a deeper model. DenseNet121 is faster to train and easier to optimize, especially on limited computational resources or when training time is constrained. The next section provides a complete description of the model implementation.

4.4 Model Training

4.4.1 Transfer learning and feature extraction using DenseNet121

In this proposed pneumonia detection framework, DenseNet121 was employed as a powerful feature extractor that leverages the principle of transfer learning. DenseNet121's architecture, in Figure 3, composes of an initial convolutional and pooling layer, which is followed by four dense blocks to set with transition layers. The dense blocks are performance-oriented, using dense connectivity to pass feature maps from each layer to every subsequent layer within the block. All layers of the pre-trained DenseNet121 model were initially frozen. This method facilitates efficient gradient flow, feature reusability, and better generalization. This technique is crucial for stabilizing the model and allowing the newly added classification layers to adapt to the specific characteristics of the pneumonia dataset without disrupting the base model's general visual features.

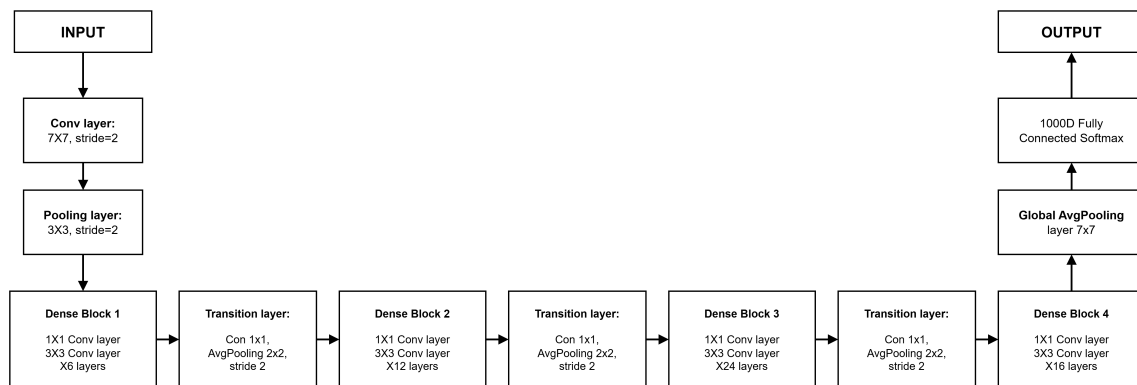


Figure 3 DenseNet121 architecture diagram

A series of custom layers were added to enhance the classification performance. A global average pooling (GAP) layer was introduced to minimize the spatial dimensions of the feature maps while preserving the most salient global information, preventing overfitting compared to flattening operations. This was followed by a batch normalization layer to normalize the activations and optimize and accelerate training. Dropout layers (with rates of 0.5 and 0.3) were added to mitigate overfitting by randomly deactivating neurons during training, optimizing the model to learn more robust and generalized representations. Two fully connected (Dense) layers with 256 and 64 units, respectively, were used to learn nonlinear combinations of extracted features, enabling the model to detect complex decision boundaries relevant to pneumonia detection. Finally, a softmax activation layer was used to produce the binary classification task's class probabilities. This layered structure enhances the post-extraction feature refinement process, ensuring that high-level, task-specific features are optimally leveraged for accurate prediction. Then, the model was compiled and fitted for 5 epochs and, further tuning is done by unfreezing the necessary layers, which is discussed in the next section.

The training procedure implements a strategic two-phase TL methodology designed to optimize feature extraction and model convergence: **Phase 1 (Warm-up Training)** initializes the process by loading *M_base*, a pretrained DenseNet121 architecture with ImageNet weights (weights = “imagenet,” include_top = FALSE). All base layers in *M_base* are configured as non-trainable (*L.trainable* ← FALSE) to preserve learned visual representations. Custom classification layers are sequentially appended to the base output ($x \leftarrow M_base.output$): a global average pooling 2D layer ($x \leftarrow GLOBAL-AVERAGE-POOLING-2D(x)$) reduces spatial dimensionality while retaining salient global contextual information; successive dropout layers with rates of 0.5 and 0.3 ($DROPOUT(x, 0.5) \rightarrow DENSE(256, \text{relu}) \rightarrow DROPOUT(x, 0.3)$) provide regularization and mitigate overfitting tendencies. The functional model *M* is constructed and compiled with the Adam optimizer initialized at learning rate $lr_1 = 1 \times 10^{-3}$, categorical cross-entropy loss function, and trained end-to-end on *D_train* with *D_val* for 5 epochs using callbacks [ReduceLROnPlateau, EarlyStopping] to adaptively adjust learning rates and prevent convergence stagnation. **Phase 2 (Fine-tuning Training)** selectively unfreezes the final 100 layers of *M_base* corresponding to Dense Blocks 3 and 4, preserving lower-level generic features and enabling domain-specific feature adaptation for pneumonia recognition. The model is recompiled with a reduced learning rate ($lr_2 = 1 \times 10^{-5}$) to ensure stable convergence without disrupting ImageNet-learned representations, and training is continued for 15 additional epochs with identical callback configurations (EarlyStopping with patience = 4) to achieve optimal performance. This phased transfer learning strategy systematically balances knowledge retention from large-scale pre-training with targeted task-specific optimization, resulting in a robust model that can generalize across heterogeneous radiographic conditions, varying patient demographics, and diverse image acquisition protocols. The above training process is performed by following the execution steps illustrated in Algorithm 1.

4.4.2 DenseNet121 Dense Block Formulation

The dense connectivity pattern in DenseNet121 is formally expressed as Equation (2):

$$\mathbf{x}_l = \mathbf{H}_l([\mathbf{x}_0, \mathbf{x}_1, \dots, \mathbf{x}_{l-1}]) \quad (2)$$

where x_l denotes the output feature maps of layer l , and H_l represents the composite function incorporating convolution (C), batch normalization (BN), and rectified linear unit (ReLU) activation, formulated as Equation (3):

$$\mathbf{H}_l = \text{ReLU}(\text{BN}(\text{C}(\cdot))) \quad (3)$$

The concatenation operation $[\cdot]$ ensures feature maps from all preceding layers within a dense block are aggregated and passed to layer l , distinguishing DenseNet from sequential architectures. This formulation enables gradient signal ∇L propagation directly from the final layers to the earlier layers through dense connection pathways, which can be mathematically expressed as Equation (4):

$$\frac{\partial L}{\partial \mathbf{x}_i} = \frac{\partial L}{\partial \mathbf{x}_l} \cdot \frac{\partial \mathbf{x}_l}{\partial \mathbf{x}_i} + \sum_{k=i+1}^{l-1} \frac{\partial L}{\partial \mathbf{x}_k} \cdot \frac{\partial \mathbf{x}_k}{\partial \mathbf{x}_i} \quad (4)$$

4.4.3 Fine tuning of the DenseNet121 model

The layers within Dense Blocks 1 and 2 primarily extract low- to mid-level features, such as edges, contours, and basic textures, which are relatively generic and transferable across visual domains. In contrast, Dense Blocks 3 and 4, which contain deeper convolutional layers, are responsible for learning higher-level abstract features specific to the target task. Therefore, the last 100 layers of the model were unfrozen after the warm-up phase. Unfreezing the last 100 layers of DenseNet121 during image training for pneumonia is a crucial step because these deeper layers capture high-level, domain-specific features important for accurate image classification.

Algorithm 1 DenseNet121-Based Pneumonia Detection Training

Input: Training dataset D_{train} , validation dataset D_{val} , learning rates $lr_1 = 1 \times 10^{-3}$ and $lr_2 = 1 \times 10^{-5}$.

Output: Trained DenseNet121 pneumonia detection model M .

BEGIN

$M_{base} \leftarrow \text{LOAD-DENSENET121}(\text{weights} = \text{imagenet}, \text{include_top} = \text{FALSE})$

FOR each layer L in $M_{base}.\text{layers}$ **DO**

Set $L.\text{trainable}$ to FALSE.

END FOR

Obtain the output feature map from M_{base} .

Apply Global Average Pooling 2D to the feature map.

Apply Dropout with rate 0.5.

Apply a Dense layer with 256 units and ReLU activation.

Apply Dropout with rate 0.3.

Construct M using $M_{base}.\text{input}$ and a 2-unit Dense Softmax output layer.

Compile M using Adam with learning rate lr_1 and categorical cross-entropy loss.

PHASE 1: WARM-UP (*frozen base, 5 epochs*)

Train M using D_{train} and validate using D_{val} for 5 epochs.

Use *ReduceLROnPlateau* and *EarlyStopping* callbacks.

PHASE 2: FINE-TUNING (*unfreeze the last 100 layers, 15 epochs*)

FOR each layer L in the last 100 layers of M **DO**

Set $L.\text{trainable}$ to TRUE.

END FOR

Compile M using Adam with learning rate lr_2 and categorical cross-entropy loss.

Train M using D_{train} and validate using D_{val} for 15 epochs.

Use *ReduceLROnPlateau* and *EarlyStopping* callbacks.

Save M as *denseNetPneumonia.h5*.

RETURN M

END

While the earlier layers in DenseNet121 focus on visual features such as edges and textures, the deeper layers specialize in complex patterns such as abnormal tissue structures or inflammation in chest X-rays. This selective unfreezing (Tajbakhsh et al., 2016) improves the performance of the model without overfitting. The training proceeded for 15 more epochs with the same callbacks, this time with a slightly increased ESP of 4. This two-phase training process effectively balances general feature preservation with targeted specialization, resulting in improved pneumonia detection performance. As a result, classification accuracy is enhanced while minimizing the risk of overfitting, leading to a more robust and well-generalized model for pneumonia detection from chest X-ray images.

4.4.4 Compilation and hyperparameter selection using the ADAM Optimizer

Compiling the DenseNet121-based model involves configuring the learning process by specifying the optimizer, loss function, and evaluation metric. The Adam optimizer with learning rate = $1e-4$ was chosen for its ability to adaptively adjust the learning rate during training for the first 5 epochs, combining the advantages of both AdaGrad and RMSProp. Adam is particularly optimized for deep learning tasks due to its efficiency, minimal memory requirements, and robustness to noisy gradients and sparse data. The learning rate of ($1e-4$) ensures that the model converges steadily without boosting the loss function minima. A reduced learning rate of $1e-5$ was assigned during fine-tuning to ensure stable convergence without altering the pretrained weights. The variation between the predicted class probabilities and the true distribution is measured, guiding the model in adjusting its weights to minimize prediction error. Because the

pneumonia detection task involves distinguishing between two categories (e.g., pneumonia vs. normal), categorical cross-entropy is an appropriate choice when using a softmax output layer with two units. Finally, the accuracy metric is used to evaluate the proportion of correct predictions made by the model during training and validation. Together, these components ensure an efficient and well-directed training process tailored to the medical image classification demands.

4.5 Model Evaluation

The model was trained and later evaluated for a couple of different metrics to check if it worked seamlessly in real-life situations. Precision, recall, accuracy, and F1 score are some of the basic classification metrics used to accurately classify pneumonia and normal cases. A confusion matrix was created from the results that addresses model classification into true positive, false positive, false negative, and true negative cases to highlight the strengths and weaknesses of the given model.

5. Results and Discussion

The model was implemented with Python using the TensorFlow and Keras frameworks. Training was accelerated using the Google Colab GPU, and NumPy was used for data preprocessing. Matplotlib and Scikit-learn were used for performance analysis. For model accuracy, a range of hyperparameters was set along with several experimental trials to achieve reproducibility.

This model implements a transfer learning approach using DenseNet121 as the base feature extractor, customized with global average pooling, dropout, and dense layers for binary classification. The training proceeded in two distinct phases: in Phase 1, all DenseNet121 layers were frozen, and the entire model was trained end-to-end for 5 epochs, achieving training accuracy of 90.53% and validation accuracy of 88.7%, indicating good initial convergence for warm-up. In Phase 2, the base layers of DenseNet121 were frozen to preserve learned features, and only the top 100 layers were unfrozen for fine-tuning over 15 additional epochs, resulting in a final training accuracy of 93% and a validation accuracy of 90.75%. The comparative evaluation shown in Figure 4 showing comparison of the three implemented approaches with the train and validation accuracy plots, revealing DenseNet121's superior performance in terms of both accuracy and generalization stability.

The VGG16 model exhibited severe overfitting with a training-validation accuracy gap of 6.09 percentage points (98.72% vs. 92.63%), attributed to its 138 million parameters and architectural limitations. The VGG16 + VDE ensemble achieved a training accuracy of 75.07% with minimal overfitting (2.66% gap) but suffered from severe class imbalance (precision = 0.21, recall = 0.03 for normal class) and 9-hour computational overhead. In contrast, DenseNet121 demonstrated balanced training-validation performance (93% vs. 90.75%, representing only a 2.25% gap) with robust class-wise metrics: precision of 94.78% (normal) and 92.66% (pneumonia), recall of 87.79% (normal) and 96.96% (pneumonia), and F1-scores of 0.9115 and 0.9476, respectively. This unified, end-to-end architecture eliminates architectural fragmentation while achieving superior generalization and clinically acceptable performance across both classes. This phased fine-tuning approach provides a robust foundation for future enhancements, such as data balancing or domain-specific augmentation.

The proposed DenseNet121 model achieves 93.42% accuracy, demonstrating competitive performance with contemporary approaches while addressing critical limitations of prior work. Compared with the custom five-block CNN (Szepesi & Szilágyi, 2022), which reported 97.21% training accuracy with notable overfitting, DenseNet121's dense connectivity facilitates superior generalization on pediatric datasets. Against VGG16+VDE ensembles (Ouyang et al., 2025), which reported 99.04% precision but severe class imbalance, our model eliminates nine-hour computational overhead through unified end-to-end architecture while maintaining clinically acceptable performance (94.78% precision, 96.96% recall for pneumonia). Relative to EfficientNetV2L

(Ali et al., 2024), which reported 94.02% accuracy, DenseNet121 achieves comparable results (93.42%) with significantly lower computational requirements (7.98M vs. substantially higher parameters), enhancing deployment feasibility in resource-constrained settings. Macro-average metrics (precision = 93.72%, recall = 92.37%) demonstrate robust class-wise performance, effectively addressing class imbalance challenges documented in recent pneumonia detection frameworks.

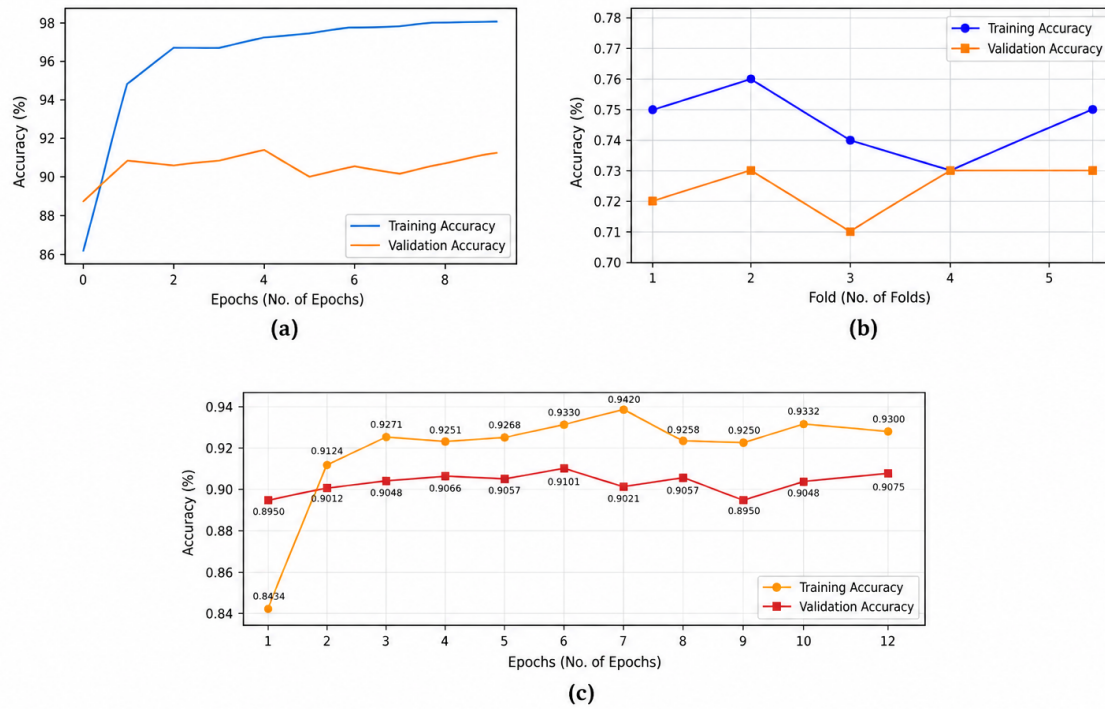


Figure 4 Comparison of training and validation accuracy plot across the implemented models. (a) – Deep Learning VGG16; (b) – Ensemble VGG16 with VDE; (c) – DenseNet121 during Phase 2 fine-tuning.

Then, the result is evaluated using the classification metrics, such as accuracy, precision, recall, and F1-score, as shown in Table 1.

Table 1 Test and cross-validation results of the DenseNet121 model

Class	Precision	Recall	F1-Score	Support
NORMAL	0.9478	0.8779	0.9115	434
PNEUMONIA	0.9266	0.9696	0.9476	690
Accuracy			0.9342	1124
Macro Avg	0.9372	0.9237	0.9295	1124
Weighted Avg	0.9348	0.9342	0.9336	1124

The classification report demonstrated that the model performed exceptionally well on the pneumonia detection task, achieving an overall accuracy of 93.42%. It shows strong class-wise performance, with particularly high recall for PNEUMONIA (96.96%), indicating that the model is highly effective at identifying infected cases—a critical factor in clinical settings to avoid missed diagnoses. For the NORMAL class, the precision is high (94.78%), but the recall is slightly lower (87.79%), suggesting that some healthy cases are misclassified as pneumonia, which, while not ideal, is a safer error than missing actual pneumonia cases. The F1-scores of 91.15% (NORMAL) and 94.76% (PNEUMONIA) reflect a well-balanced model, with consistent performance across

both classes confirmed by the macro and weighted averages. Overall, the results indicate a reliable and clinically cautious model that prioritizes sensitivity where it matters the most.

The confusion matrix in Figure 5 reveals the model's diagnostic performance in classifying chest X-ray images into NORMAL and PNEUMONIA categories. Of the total test samples, the model correctly identified 669 pneumonia cases (true positives) and 381 normal cases (true negatives). However, it misclassified 53 normal cases as pneumonia (false positives) and, more critically, 21 pneumonia cases as normal (false negatives). Although the overall accuracy appears high, the false negatives are of significant clinical concern because they represent pneumonia cases that the model failed to detect. Such misclassifications could lead to delayed or missed treatment in a real-world clinical setting, increasing the risk of complications, hospitalization, or even mortality. On the other hand, although less dangerous, false positives could lead to unnecessary medical interventions, psychological stress for patients, and resource overutilization.

The ROC curve in Figure 6(a) demonstrates that the model performs exceptionally well in distinguishing between NORMAL and PNEUMONIA cases, with both classes achieving an impressive AUC of 0.98. This high AUC indicates excellent discriminatory power, indicating that the model can accurately rank and separate positive (pneumonia) and negative (normal) instances across thresholds. The proximity of the curve to the top-left corner further confirms that the model maintains high sensitivity while minimizing false positives, making it well-suited for clinical use where missing a pneumonia case could have serious consequences. Overall, the ROC analysis reinforces the model's strong reliability and robustness, complementing the precision, recall, and accuracy metrics observed in previous evaluations.

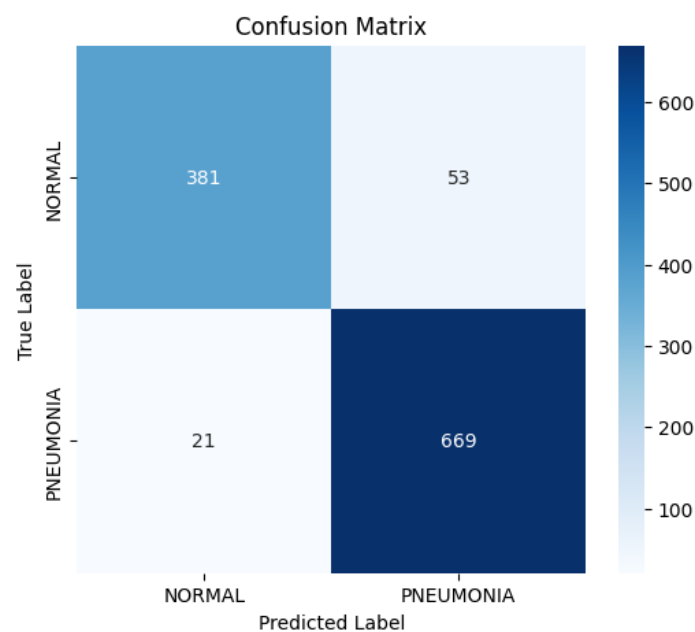


Figure 5 Confusion matrix of the VGG16 classification model

The Precision-Recall (PR) curve in Figure 6(b) provides a more focused evaluation of the model's performance, especially in the context of class imbalance, by plotting PR (positive predictive value) against PR (sensitivity). In the presented graph, both classes—NORMAL and PNEUMONIA—maintain high precision across a broad range of recall values, particularly the PNEUMONIA class, which slightly outperforms NORMAL in maintaining precision even at high recall levels. This indicates that the model is not only able to detect most positive cases (high recall) but also ensures that most predicted positives are correct (high precision), which is critical in medical diagnosis. The sharp drop near a recall of 1.0 reflects the trade-off between retrieving all positive cases and maintaining precision. Overall, this curve confirms that the model delivers strong, reliable predictions, particularly in identifying pneumonia, and is well-suited for clinical scenarios where false positives and false negatives carry different costs.

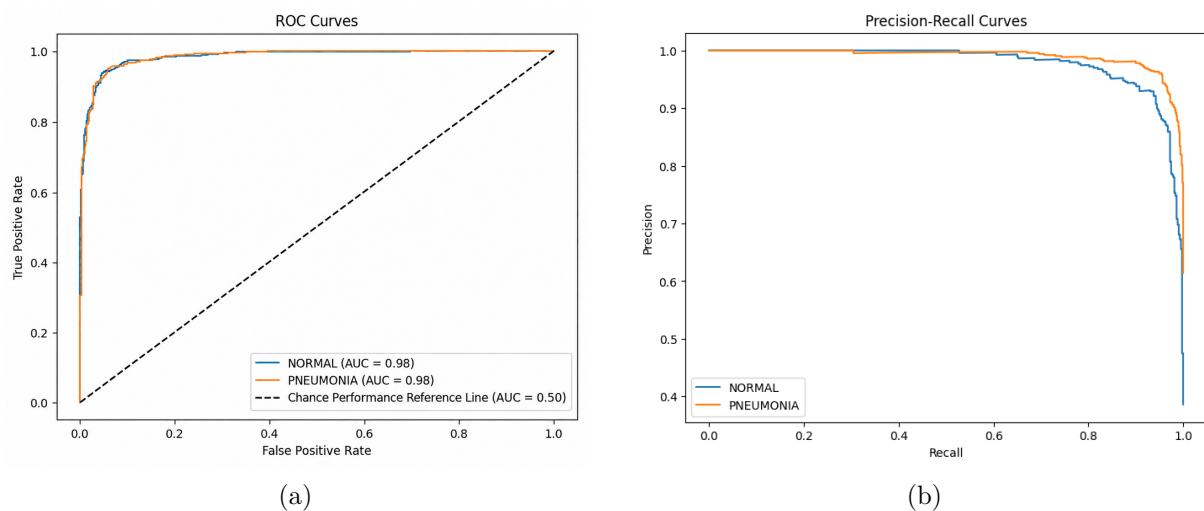


Figure 6 Evaluation of classification performance using (a) ROC and (b) Precision-Recall Curves

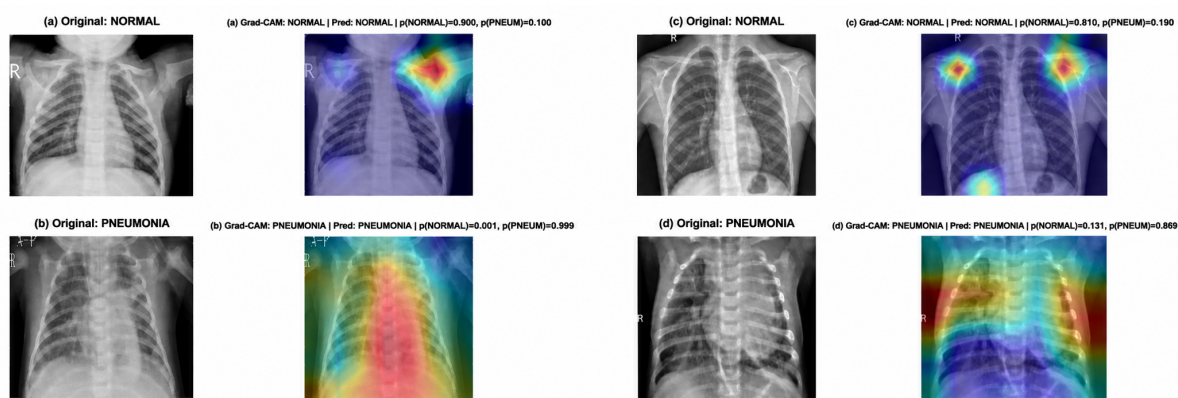


Figure 7 Grad-CAM visualization of pneumonia detection in chest X-ray images

Gradient-weighted Class Activation Mapping (Grad-CAM) was implemented to enhance model interpretability and address clinical transparency requirements to generate heatmaps highlighting discriminative regions influencing classification decisions (Figure 7). Grad-CAM computes gradients of the target class score with respect to DenseNet121's final convolutional feature maps, applying global average pooling to obtain channel-wise importance weights that are linearly combined with activation maps to produce localization heatmaps. For NORMAL cases, visualizations showed diffuse, low-intensity activations distributed across lung fields with minimal focal attention, which correctly reflected the absence of pathological consolidations ($p(\text{NORMAL}) = 0.900$). Conversely, PNEUMONIA cases exhibit concentrated high-intensity activations (red-yellow regions) precisely localized to the lung parenchyma, particularly the lower and middle zones where pneumonia-induced opacities are evident ($p(\text{PNEUM}) = 0.999$). These activations demonstrate strong qualitative alignment with clinically relevant anatomical regions identified during radiological diagnosis, confirming that DenseNet-121 features capture pathologically significant patterns rather than spurious correlations. The visualizations confirm that the model focuses predominantly on lung tissue rather than peripheral structures, addressing the concerns of previous reviewers by demonstrating clinically meaningful attention patterns corresponding to the established radiological criteria for pneumonia diagnosis. This interpretability is critical for clinical adoption, enabling radiologists to verify that predictions are based on diagnostically relevant features, thereby increasing trust and supporting integration into CAD workflows.

Additionally, to evaluate the model's performance with separate raw images, a comparison

between the original and predicted labels in Figure 8 was conducted to analyze misclassification patterns. The proposed architecture demonstrated competitive performance, reinforcing the significance of its design choices. The results confirm that DenseNet121, with its carefully designed architecture, is well-suited for pneumonia detection, providing a high level of accuracy, precision, and specificity needed for clinical applications.

5.1 Clinical feasibility and deployment considerations

The proposed DenseNet121 model (7.98M parameters, 122 ms GPU inference latency) demonstrates the feasibility of practical deployment across clinical contexts. While CPU-based inference in resource-constrained settings (1.2–2.5 s per image) exceeds real-time requirements, this latency remains clinically acceptable for preliminary screening workflows in low-resource environments, comparable to CPU-evaluated CNN systems. DenseNet121 achieves superior computational efficiency relative to Vision Transformers (~108M parameters, higher latency) and ensemble methods (9+ hour preprocessing overhead), striking an optimal balance between representational capacity and practical deployability. Future deployment optimization would require model quantization (8-bit representation, ~75% memory reduction), containerization via TensorFlow Lite/ONNX Runtime for cross-platform compatibility.

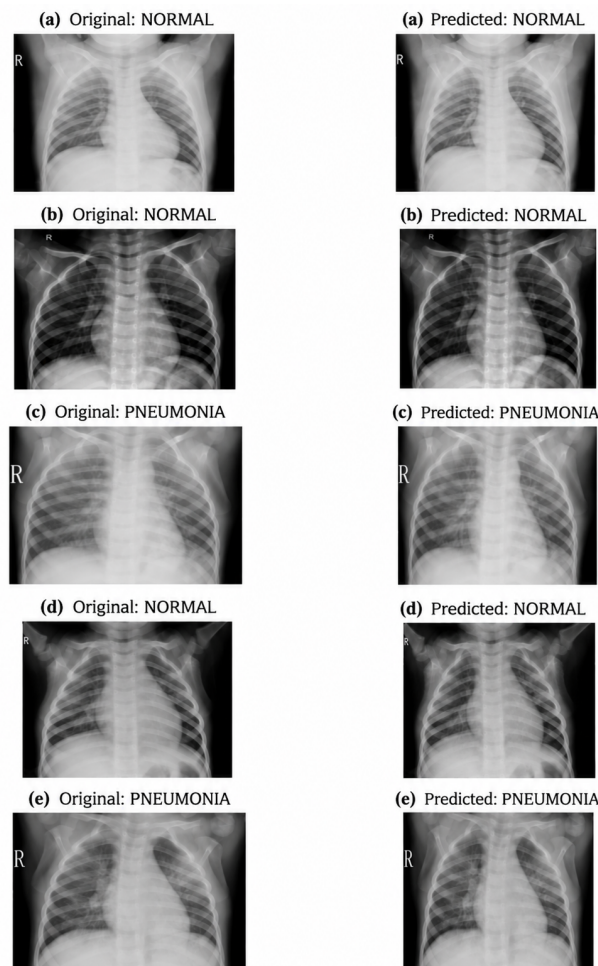


Figure 8 Comparison of the original and predicted pneumonia detection labels using the trained model

6. Conclusion

In this paper, we developed and tested an approach using deep learning for detecting pneumonia using chest X-ray images. A custom CNN-based DenseNet121 is used to classify

pneumonia cases. The key novelty of the model is the placement of dropout layers within the convolutional network, appropriate freezing and unfreezing of layers, and reusability of features to improve generalization and reduce overfitting. Unlike other approaches that rely on pre-trained architectures, this custom CNN model was trained from scratch and outperforms other approaches in terms of accuracy and speed. The results show that the proposed model has a precision of 94.78% and takes 122 ms to predict. This approach is both accurate and real-time. Given the importance of chest X-ray in diagnosing lung diseases, radiologists can use automated detection methods like ours in preliminary diagnosis and treatment planning. While the model works well, future work can focus on increasing the dataset, adding labeled data with more characteristics, such as viral or bacterial pneumonia, implementing advanced DL ensemble methods using DenseNet, VGG, EfficientNet, etc. for more consistent results, and refining the preprocessing methods for further accuracy and robustness improvement.

Limitations and Scope Considerations

This study acknowledges several important limitations that inform the interpretation of the results and future research directions. First, this investigation was restricted to pediatric chest radiographs (age 1–5 years) from a single-center dataset (Guangzhou Women and Children's Medical Center). The generalizability of this study to adult patient populations remains unexplored because anatomical chest morphology, pneumonia manifestation patterns, and radiographic acquisition parameters differ substantially between pediatric and adult cohorts. The proposed DenseNet121 framework cannot be applied to adult radiographs without external validation on adult-specific datasets (e.g., ChestX-ray8, MIMIC-CXR).

Second, the binary classification paradigm (Normal vs. Pneumonia) does not differentiate between bacterial and viral etiologies, limiting the clinical specificity of targeted antimicrobial therapy selection. Although this binary approach is typical in the current literature, a multi-class classification that distinguishes bacterial, viral, and fungal pneumonia would enhance its clinical utility. Such differentiation would require (i) external datasets with pathogen-specific annotations and (ii) potential architectural modifications to accommodate multiclass outputs—constraints beyond the scope of this study.

Third, the dataset exhibits class imbalance (1,341 normal vs. 3,876 pneumonia cases), addressed via data augmentation but not via advanced cost-sensitive loss functions or class-weighted sampling. Although the model achieves high recall for pneumonia detection (96.96%) and minimizes false negatives in clinical settings, future work incorporating weighted loss functions (e.g., focal loss) may further optimize performance across imbalanced distributions.

Finally, this study does not include k-fold cross-validation or external dataset testing on the ChestX-ray8 or MIMIC-CXR databases. Although the achieved accuracy on the Kaggle split is encouraging, such external validation would strengthen robustness across diverse imaging protocols and patient demographics—a recommended direction for future investigations.

Acknowledgements

The authors extend their appreciation to Prince Sattam bin Abdulaziz University for funding this research work through the project number (PSAU/2025/01/38579).

Funding

The authors are grateful to Prince Sattam bin Abdulaziz University for funding this research work through the project number (PSAU/2025/01/38579).

Conflict of Interest

The authors declare no conflicts of interest. The authors confirm that there are no personal, financial, or professional relationships that could be perceived as influencing the design,

execution, interpretation, or reporting of the research presented in this paper.

References

- Ali, M., Shahroz, M., Akram, U., Mushtaq, M. F., Carvajal Altamiranda, S., Aparicio Obregon, S., De La Torre Díez, I., & Ashraf, I. (2024). Pneumonia detection using chest radiographs with novel EfficientNetV2L model. *IEEE Access*, 12, 34691–34707. <https://doi.org/10.1109/ACCESS.2024.3372588>
- Aly, S., & Almotairi, S. (2020). Deep convolutional self-organizing map network for robust handwritten digit recognition. *IEEE Access*, 8, 107035–107045. <https://doi.org/10.1109/ACCESS.2020.3000829>
- An, Q., Chen, W., & Shao, W. (2024). A deep convolutional neural network for pneumonia detection in X-ray images with attention ensemble. *Diagnostics*, 14(4), 390. <https://doi.org/10.3390/diagnostics14040390>
- Das, R., Nayak, D. S. K., Rout, C. P., Jena, L., & Swarnkar, T. (2024). Deep learning techniques for identification of pneumonia: A CNN approach. *Proceedings of the International Conference on Advancements in Smart, Secure and Intelligent Computing (ASSIC)*, 1–5. <https://doi.org/10.1109/ASSIC60049.2024.10507933>
- Dash, S. K., Nanda, B., Dash, K. C., Mishra, P. K., & Pattnaik, S. (2026). PneumoNet: A VGG16 based CNN model for pneumonia prediction. *SN Computer Science*, 7(1), 33. <https://doi.org/10.1007/s42979-025-04630-w>
- Faouzi, J., & Colliot, O. (2023). Classic machine learning methods. In O. Colliot (Ed.), *Machine learning for brain disorders* (pp. 25–75, Vol. 197). Springer. https://doi.org/10.1007/978-1-0716-3195-9_2
- Fu, M., Tantithamthavorn, C., & Le, T. (2025). DAViT: A domain-adapted vision transformer for automated pneumonia detection and explanation using chest X-ray images. *IEEE Access*, 13, 103033–103044. <https://doi.org/10.1109/ACCESS.2025.3579314>
- Ghanim, T. M., Khalil, M. I., & Abbas, H. M. (2020). Comparative study on deep convolution neural networks (DCNN)-based offline Arabic handwriting recognition. *IEEE Access*, 8, 95465–95482. <https://doi.org/10.1109/ACCESS.2020.2995499>
- Guddati, P., Dash, S., & Tripathy, R. K. (2024). FPGA implementation of the proposed DCNN model for detection of tuberculosis and pneumonia using CXR images. *IEEE Embedded Systems Letters*, 16(4), 445–448. <https://doi.org/10.1109/LES.2024.3370833>
- Hasan, M. R., Ullah, S. M. A., & Islam, S. M. R. (2024). Recent advancement of deep learning techniques for pneumonia prediction from chest X-ray image. *Medical Reports*, 7, 100106. <https://doi.org/10.1016/j.hmedic.2024.100106>
- Hofmeister, J., Garin, N., Montet, X., Scheffler, M., Platon, A., Poletti, P.-A., Stirnemann, J., Debray, M.-P., Claessens, Y.-E., Duval, X., & Prendki, V. (2024). Validating the accuracy of deep learning for the diagnosis of pneumonia on chest x-ray against a robust multimodal reference diagnosis: A post hoc analysis of two prospective studies. *European Radiology Experimental*, 8, 14. <https://doi.org/10.1186/s41747-023-00416-y>
- Huang, G., Liu, Z., van der Maaten, L., & Weinberger, K. Q. (2017). Densely connected convolutional networks. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 4700–4708. <https://doi.org/10.1109/CVPR.2017.243>
- Jiang, Z.-P., Liu, Y.-Y., Shao, Z.-E., & Huang, K.-W. (2021). An improved VGG16 model for pneumonia image classification. *Applied Sciences*, 11(23), 11185. <https://doi.org/10.3390/app112311185>
- Jung, T., & Kim, J. (2023). A new support vector machine for categorical features. *Expert Systems with Applications*, 229, 120449. <https://doi.org/10.1016/j.eswa.2023.120449>
- Kakarla, S., Gangula, P., Rahul, M. S., Singh, C. S. C., & Sarma, T. H. (2020). Smart attendance management system based on face recognition using CNN. *Proceedings of IEEE-HYDCON*, 1–5. <https://doi.org/10.1109/HYDCON48903.2020.9242847>

- Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., & Liu, T.-Y. (2017). LightGBM: A highly efficient gradient boosting decision tree. *Advances in Neural Information Processing Systems*, 30, 3146–3154.
- Khan, A. I., & Al-Habsi, S. (2020). Machine learning in computer vision. *Procedia Computer Science*, 167, 1444–1451. <https://doi.org/10.1016/j.procs.2020.03.355>
- Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciompi, F., Ghafoorian, M., van der Laak, J. A. W. M., van Ginneken, B., & Sánchez, C. I. (2017). A survey on deep learning in medical image analysis. *Medical Image Analysis*, 42, 60–88. <https://doi.org/10.1016/j.media.2017.07.005>
- McLuckie, A. (2009). *Respiratory disease and its management* (1st ed.). Springer London. <https://doi.org/10.1007/978-1-84882-095-1>
- Misra, S., & Laskar, R. H. (2019). Development of a hierarchical dynamic keyboard character recognition system using trajectory features and scale-invariant holistic modeling of characters. *Journal of Ambient Intelligence and Humanized Computing*, 10(4), 4305–4319. <https://doi.org/10.1007/s12652-019-01189-2>
- Mooney, P. T. (2018). *Chest X-ray images (pneumonia)*. Retrieved August 28, 2026, from <https://www.kaggle.com/datasets/paultimothymooney/chest-xray-pneumonia>
- Munir, K., Tanveer, M. U., Alyamani, H. J., Bermak, A., & Rehman, A. U. (2025). PneuX-Net: An enhanced feature extraction and transformation approach for pneumonia detection in X-ray images. *IEEE Access*, 13, 84024–84037. <https://doi.org/10.1109/ACCESS.2025.3568885>
- O’Shea, K., & Nash, R. (2015). An introduction to convolutional neural networks. *arXiv preprint arXiv:1511.08458*. <https://doi.org/10.48550/arXiv.1511.08458>
- Ouyang, H., Liu, D., Li, S., Ding, W., & Zhan, Z.-H. (2025). Two-stage deep feature selection method using voting differential evolution algorithm for pneumonia detection from chest X-ray images. *IEEE Transactions on Emerging Topics in Computational Intelligence*, 9, 918–932. <https://doi.org/10.1109/TETCI.2024.3425285>
- Prakash, R. M., Thenmozhi, N., & Gayathri, M. (2019). Face recognition with convolutional neural network and transfer learning. *Proceedings of the International Conference on Smart Systems and Inventive Technology (ICSSIT)*, 861–864. <https://doi.org/10.1109/ICSSIT46314.2019.8987899>
- Ren, H., Wong, A. B., Lian, W., Cheng, W., Zhang, Y., He, J., Liu, Q., Yang, J., Zhang, C. J., Wu, K., & Zhang, H. (2021). Interpretable pneumonia detection by combining deep learning and explainable models with multisource data. *IEEE Access*, 9, 95872–95883. <https://doi.org/10.1109/ACCESS.2021.3090215>
- Schulz, H., & Behnke, S. (2012). Deep learning. *KI – Künstliche Intelligenz*, 26(4), 357–363. <https://doi.org/10.1007/s13218-012-0198-z>
- Shorten, C., & Khoshgoftaar, T. M. (2019). A survey on image data augmentation for deep learning. *Journal of Big Data*, 6, 60. <https://doi.org/10.1186/s40537-019-0197-0>
- Siddiqi, R., & Javaid, S. (2024). Deep learning for pneumonia detection in chest X-ray images: A comprehensive survey. *Journal of Imaging*, 10(8), 176. <https://doi.org/10.3390/jimaging10080176>
- Simonyan, K., & Zisserman, A. (2015). Very deep convolutional networks for large-scale image recognition. *International Conference on Learning Representations (ICLR)*, 1–14. Retrieved August 28, 2026, from <https://arxiv.org/abs/1409.1556>
- Szepesi, P., & Szilágyi, L. (2022). Detection of pneumonia using convolutional neural networks and deep learning. *Biocybernetics and Biomedical Engineering*, 42(3), 1012–1022. <https://doi.org/10.1016/j.bbe.2022.08.001>
- Tajbakhsh, N., Shin, J. Y., Gurudu, S. R., Hurst, R. T., Kendall, C. B., Gotway, M. B., & Liang, J. (2016). Convolutional neural networks for medical image analysis: Full training or fine tuning? *IEEE Transactions on Medical Imaging*, 35(5), 1299–1312. <https://doi.org/10.1109/TMI.2016.2535302>

- Vashist, P. C., Pandey, A., & Tripathi, A. (2020). A comparative study of handwriting recognition techniques. *Proceedings of the International Conference on Computation, Automation and Knowledge Management (ICCAKM)*, 456–461. <https://doi.org/10.1109/ICCAKM46823.2020.9051464>
- World Health Organization. (2008). *WHO/Health Canada consultation on serological criteria for evaluation and licensing of new pneumococcal vaccines* (Report). World Health Organization. Geneva, Switzerland. Retrieved August 28, 2026, from <https://www.who.int/publications/m/item/who-consultation-on-new-pneumococcal-vaccines>
- Zhang, Z. (2016). Introduction to machine learning: K-nearest neighbors. *Annals of Translational Medicine*, 4(11), 218. <https://doi.org/10.21037/atm.2016.03.37>
- Zhou, Y., Ni, H., Ren, F., & Kang, X. (2019). Face and gender recognition system based on convolutional neural networks. *Proceedings of the IEEE International Conference on Mechatronics and Automation (ICMA)*, 1091–1095. <https://doi.org/10.1109/ICMA.2019.8816275>