Dicle University
Journal of Engineering
https://dergipark.org.tr/tr/pub/dumf
duje.dicle.edu.tr

(Research Article)

The Role of Weight Initialization and Preprocessing Techniques in Analyzing Chest X-ray Images with Deep Neural Networks: A Comparative Study

Ömer Faruk SÖYLEMEZ[1*]

[1] Dicle University, Department of Computer Engineering, osoylemez@dicle.edu.tr, Orcid No: 0000-0002-4076-5230

ARTICLE INFO

Article history:

Received 6 November 2025
Received in revised form 16 February 2026
Accepted 23 March 2026
Available online 25 March 2026

Keywords:

Chest X-rays, deep learning, medical image analysis, preprocessing

Doi: 10.24012/dumf.1818553

* Corresponding author

ABSTRACT

Accurate interpretation of chest X-rays (CXRs) using machine learning techniques plays a vital role in improving diagnostic practices within healthcare. This study advances CXR analysis by exploring various weight initialization and preprocessing methods with deep neural networks. Specifically, (i) an object detection model was trained to identify regions of interest (ROIs) within CXRs, enabling focused image analysis; (ii) weight initialization strategies, including partial fine-tuning, complete fine-tuning, and random initialization, were assessed using the EfficientNet-B1 model on both original and cropped images; and (iii) preprocessing techniques such as histogram equalization, bilateral filtering, Gaussian filtering, and Contrast Limited Adaptive Histogram Equalization (CLAHE) were applied to enhance image quality. Results show that ROI selection significantly improves model performance by focusing on relevant image areas. Pre-trained ImageNet weights with complete fine-tuning outperformed random initialization, demonstrating the advantages of transfer learning, particularly with limited datasets. Among preprocessing methods, the combination of CLAHE with bilateral filtering achieved the highest accuracy (94.60%), suggesting that advanced preprocessing methods substantially enhance model performance. These results underscore the importance of effective weight initialization and preprocessing in optimizing diagnostic accuracy and efficiency in CXR analysis.

Introduction

Chest diseases, particularly those affecting the lungs, represent a significant global health concern. According to the World Health Organization (WHO), chronic respiratory diseases are among the top three leading causes of death worldwide [1]. For instance, in 2019, approximately 3.23 million people succumbed to chronic obstructive pulmonary disease alone [2]. Additionally, lung cancer, which is frequently detected through chest X-rays (CXRs), remains a leading cause of cancer-related deaths, accounting for around 1.8 million deaths in 2020 [3].

Given the prevalence and severe impact of these diseases, early and accurate detection is crucial. CXR imaging is a primary diagnostic tool used in the detection and monitoring of various lung conditions [4]. They are utilized to detect abnormalities in the lungs, but the analysis process is often time-consuming and requires specialized expertise. Consequently, the potential for human error and delays in the analysis process can lead to significant issues in clinical practice [5].

While advanced imaging modalities, such as X-ray computed tomography or nuclear magnetic resonance imaging, provide high-resolution three-dimensional images, they are associated with higher radiation doses, increased costs, and limited accessibility in routine clinical settings. CXRs, in contrast, offer rapid, low-dose, and widely available imaging, making them a practical first-line tool for large-scale screening and initial diagnosis. The ALARA (As Low As Reasonably Achievable) principle emphasizes minimizing patient exposure to radiation, further reinforcing the preference for CXR in many clinical scenarios [6], [7], [8]

Recent advancements in artificial intelligence and machine learning have offered substantial potential for automating and accelerating medical image analysis. Deep learning techniques have made significant strides in this area, providing enhanced capabilities for image processing [9], [10]. Specifically, deep learning models, such as convolutional neural networks (CNNs) and other advanced architectures, have emerged as promising approaches for analyzing CXRs [11]. Unlike traditional methods, which often rely on manual feature extraction and heuristic approaches, deep learning models automatically learn complex patterns and long-range dependencies directly from the data. This capability allows them to outperform conventional techniques in terms of accuracy and efficiency, presenting a strong foundation for improving the diagnostic process of CXRs.

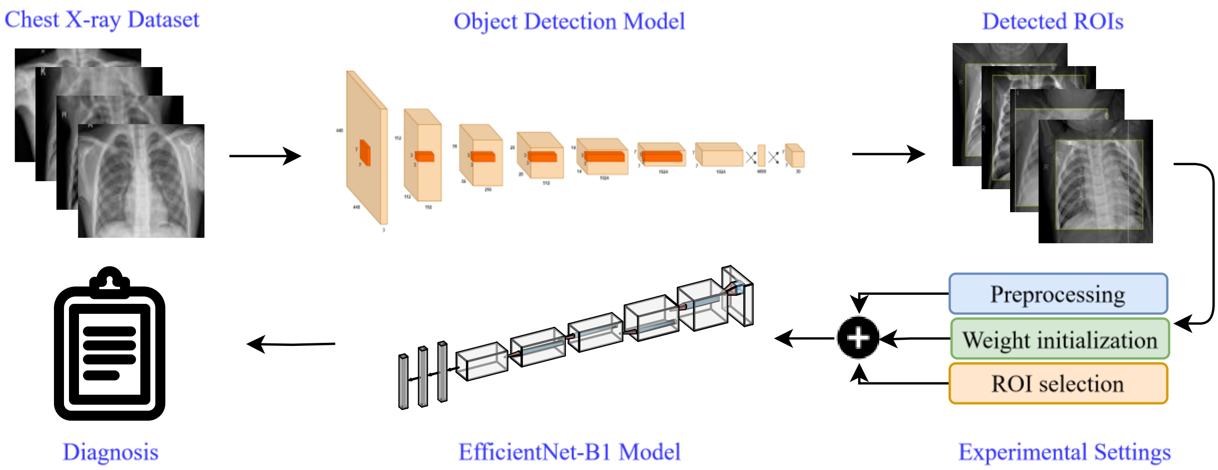

Figure 1. The workflow of the proposed study

Studies on CXR classification have leveraged a diverse array of deep learning architectures to enhance diagnostic performance and accuracy [12]. CNNs, including widely adopted models such as VGG, ResNet, and DenseNet, have been extensively utilized due to their strong performance in feature extraction and hierarchical learning. In addition to CNNs, transformer-based models have recently gained prominence in medical image analysis [13]. Transformers, originally designed for natural language processing tasks, have been adapted for vision tasks through architectures like Vision Transformers (ViTs) [14]. These models offer the advantage of capturing long-range dependencies and global context, which can be particularly beneficial for understanding complex patterns and relationships in CXR images. Hybrid approaches that combine CNNs with transformers or other techniques have also been explored [15], [16]. These methods aim to leverage the strengths of both architectures, integrating the spatial feature extraction capabilities of CNNs with the global context modeling of transformers.

Deep learning techniques offer a comprehensive approach to feature extraction from CXR images. However, this broad approach can make it challenging to eliminate irrelevant features, as the models process the entire image. To address this issue, Region of Interest (ROI) selection is crucial. ROI selection enables the model to focus on specific areas of the image that are most relevant to the diagnosis, thereby enhancing the accuracy of feature extraction and reducing the impact of non-relevant information. Studies have shown that incorporating ROI selection significantly improves accuracy by narrowing the focus to critical image regions [17].

Image processing techniques further refine this process by improving image quality. Methods such as histogram equalization and bilateral filtering enhance the clarity of features but can also introduce non-linearity, which may affect the performance of certain deep learning models. Advanced image processing techniques are essential for preparing high-quality data that facilitates more effective training and better model performance [18]. Proper weight initialization is also critical for effective training, as it helps the model converge more efficiently and avoids common issues such as vanishing or exploding gradients. Optimal weight initialization strategies can significantly enhance model convergence and overall performance [19]. By integrating ROI selection, advanced image processing techniques, and optimal weight initialization, deep learning models can achieve more precise and reliable diagnostic outcomes.

Although numerous studies have proposed novel architectures or ensemble strategies for CXR classification, limited attention has been given to systematically evaluating the combined impact of ROI selection, weight initialization strategies, and preprocessing techniques under a controlled experimental setup. Most prior works focus primarily on architectural innovation, whereas our study isolates and quantitatively analyzes these methodological components using a unified backbone network. This enables a clearer understanding of how preprocessing, transfer learning strategies, and region-focused training influence performance independently of architectural modifications.

The contributions of this study, in which we aimed to enhance CXR analysis through different weight initialization and preprocessing techniques with deep neural networks, are as follows:

- We trained a state-of-the-art object detection model to identify rib cage-centered ROIs within the CXR dataset and demonstrated that training on cropped regions yields measurable performance improvements compared to full-image training.
- We conducted a controlled comparison of weight initialization strategies (random initialization, partial fine-tuning, and complete fine-tuning) using EfficientNet-B1, isolating their impact on performance and computational efficiency.
- We systematically evaluated multiple preprocessing techniques and their combinations, showing that different strategies produce significantly different performance behaviors, highlighting the importance of preprocessing design choices in CXR analysis. Results

indicate that various preprocessing techniques have distinct effects on CXR analysis.

The workflow of the proposed study is given in Figure 1. The remainder of the study is organized as follows: Section 2 reviews related work on CXR analysis, covering studies conducted before and after the advent of deep learning. Section 3 describes the methodology of the study. Section 4 presents the experimental work, detailing the setup and execution. Section 5 reports the results of the experiments. Section 6 discusses the interpretation of the results, challenges encountered, and limitations of the study. Finally, Section 7 provides a general assessment of the study, summarizing the key contributions, overall impact, and potential directions for future research.

Related work

Before the advent of deep learning, CXR analysis primarily relied on traditional image processing techniques and heuristic methods [20], [21]. Early approaches often involved manual feature extraction and rule-based classification [22], [23], [24]. These methods typically used handcrafted features such as texture, shape, and edge information to identify abnormalities in CXR images [25]. For example, techniques such as template matching and histogram-based methods were employed to detect features indicative of conditions like pneumonia or lung tumors [26], [27]. Despite their utility, these methods were limited by their reliance on expert knowledge and their inability to adapt to new or unseen data without significant manual intervention.

The advent of deep learning significantly changed the ways of analyzing and classifying CXR images. Deep learning models, particularly those based on CNNs and transformers, have revolutionized the field by automating feature extraction and learning complex patterns directly from raw image data. These models offer the advantage of adapting to diverse and large datasets, leading to substantial improvements in diagnostic accuracy and efficiency.

CNNs have become a cornerstone of medical image analysis, including CXR classification [11], [28]. Wang et al. [29] propose an end-to-end deep learning framework for the segmentation and diagnosis of pneumothorax on CXRs. The framework utilizes a fully convolutional DenseNet integrated with multi-scale modules and spatial and channel squeezes and excitation modules. To enhance boundary segmentation precision, a spatial weighted cross-entropy loss function is introduced. The study involves 11051 front-view CXR images, including 5566 cases of pneumothorax and 5485 cases of non-pneumothorax. The experimental results show a mean pixel-wise accuracy of 93.45%, a dice similarity coefficient of 92.97%, a diagnostic accuracy of 93.45%, and an F1-score of 92.97%. Owais et al. [30] present a comprehensive computer aided diagnosis framework for the diagnosis of tuberculosis using CXR images. The study proposes a fusion-based deep classification network for decisions and a multilevel similarity measure algorithm to retrieve the best-matched cases from a database of previous patients. The framework was evaluated on two well-known CXR datasets from the US National Library of Medicine and the National Institutes of Health. The results show an F1 score of 0.929, an average precision of 0.937, an average recall of 0.921, an accuracy of 0.928, and AUC of 0.965. Rajaraman et al. [31] propose a systematic approach incorporating pretraining, ensemble methods, and statistical validation for improving COVID-19 detection from CXRs. Their study utilizes several prominent datasets. A custom U-Net with dropout layers is employed to segment the lung ROI from the background, using Gaussian dropouts in the encoder to mitigate overfitting and provide restrictive regularization. Results demonstrate that the top-3 weighted averaging ensemble method achieved an accuracy of 90.97%, an AUC of 0.9508, a precision of 93.94%, and an F1-score of 90.91%. Wang et al. [32] investigate the potential of deep learning for assessing pneumoconiosis on CXRs and compare its performance with certified radiologists. They collected a dataset of 1,881 CXR images, comprising 923 images from subjects with pneumoconiosis and 958 from normal subjects, all acquired from a screening setting. The study employed the Inception-V3 deep CNN for classification, achieving an AUC of 0.878 (95% CI 0.811 to 0.946).

Transformer-based models have recently gained traction in medical image analysis, providing the advantage of capturing long-range dependencies and global context within images. Okolo et al. [33] proposed the Input Enhanced Vision Transformer, an enhanced deep learning model for CXR image classification. They evaluate on four datasets containing various pathologies, including tuberculosis, pneumonia, and COVID-19. The model achieves F1-scores between 96.39% and 100%, with improvements of up to +5.82% over Vision Transformers (ViT). Additionally, model shows higher sensitivity (93.50% to 100%) and precision (97.96% to 100%) compared to ViT. Chen et al. [34] utilize machine learning to improve the diagnostic accuracy of COVID-19 using CXRs. Their study evaluates various architectures, including EfficientNet, multiscale vision transformers, efficient vision transformers, and ViTs, against an extensive open-source dataset consisting of 3616 COVID-19, 6012 lung opacity, 10192 normal, and 1345 viral pneumonia images. Multiscale models, such as MViT and EfficientNet, are noted for their tendency to overfit. In contrast, the fine-tuned ViT model achieves an accuracy of 95.79% for four-class classification, 99.57% for three-class classification, a recall of 98.58%, precision of 98.87%, F1 score of 98.73%, specificity of 99.76%, and an AUC of 0.9993.

Hybrid models that integrate both CNNs and transformer-based approaches have emerged as a promising direction for CXR analysis. These models aim to leverage the strengths of both architectures, combining the local feature extraction capabilities of CNNs with the global context modeling of transformers. Ukwuoma et al. [16] propose a hybrid explainable deep learning framework for the identification of pneumonia using CXR images. The model combines ensemble CNNs and the transformer encoder mechanism. The ensemble learning backbone extracts robust features from raw input X-ray images using two different ensembles. The transformer encoder utilizes the self-

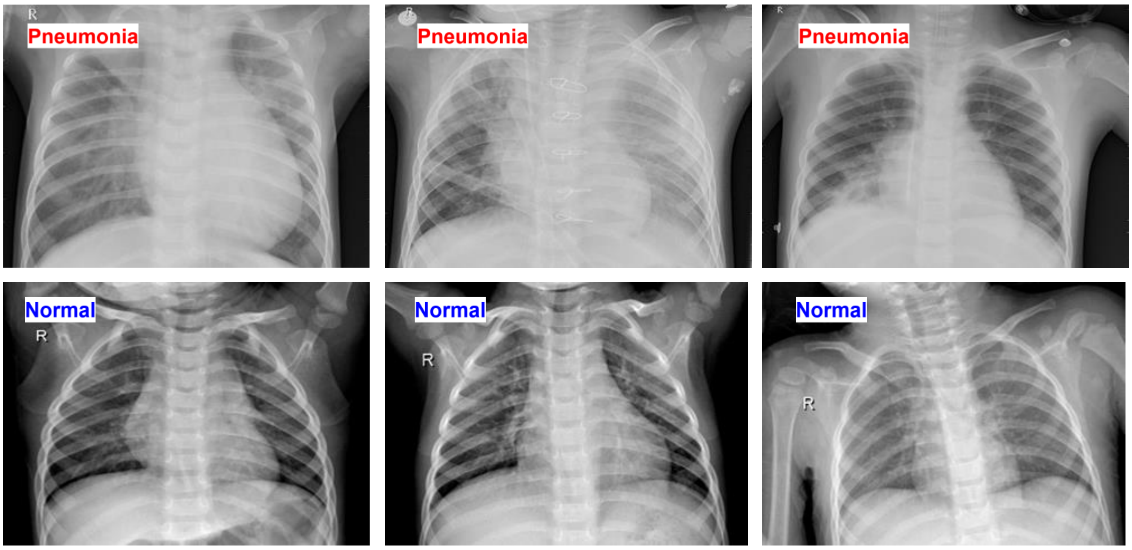

Figure 2. Examples of the normal and pneumonia classes from the dataset used in the study.

attention mechanism with a multilayer perceptron for disease identification. The proposed hybrid deep learning model achieved a classification performance of 99.21% in both overall accuracy and F1-score for binary classification tasks, and it reached 98.19% accuracy and 97.29% F1-score for multi-classification tasks. Wu et al. [15] propose CheXNet, a hybrid deep learning network designed to enhance multi-label CXR classification. This approach addresses the limitations of conventional CNNs and their integration with transformers, such as difficulties in capturing label dependencies and the lack of direct interaction between CNN and transformer models. CheXNet integrates three key components: the Label Embedding and Multi-Scale Pooling module for capturing label dependencies and fusing features across scales, the Inner Branch module for incorporating detailed local features, and the Information Interaction module to facilitate interaction between the CNN and Transformer. This hybrid structure enables CNNs to provide richer inputs to transformers and allows transformers to enhance feature extraction in CNNs. The proposed method demonstrates effectiveness with average AUCs of 82.56% and 76.80% on the CXR11 and CXR14 datasets, respectively.

Methodology

Dataset preparation

We used "validated OCT and Chest X-Ray images dataset" presented in [35]. The dataset consists of anterior-posterior CXR images from pediatric patients aged one to five years, sourced from the retrospective cohorts at Guangzhou Women and Children's Medical Center, Guangzhou. These images were obtained during routine clinical care. Ethical approval was secured from the Institutional Review Board (IRB)/Ethics Committee. The study was conducted in compliance with the Health Insurance Portability and Accountability Act (HIPAA) of the United States and in accordance with the principles of the Declaration of Helsinki [36].

To prepare the CXR images for analysis, an initial quality control screening was performed to discard low-quality or unreadable scans. Expert diagnoses were provided by two physicians, ensuring that only accurately graded images were used for training the AI system. Additionally, a third expert reviewed the evaluation set to verify and correct any grading errors [36].

The dataset contains 5856 CXR images categorized into two classes: pneumonia and normal. The image sizes range from 384×127 to 2916×2583. The dataset is divided into two parts: training and testing. There are 5232 training samples and 624 test samples, with approximately 9% of the images used for testing and 91% for training. The number of samples for each part is provided in Table 1. Examples of the normal and pneumonia classes from the dataset used in the study are given in Figure 2.

The dataset split used in this study follows the original predefined training and test partitions. Although the dataset exhibits class imbalance, this structure was intentionally preserved to ensure comparability with prior studies conducted on the same dataset. Altering the class distribution or redefining the splits could compromise consistency and potentially introduce data leakage, especially if images belonging to the same patient were distributed across different splits. Therefore, the original dataset configuration was maintained.

The original dataset includes a validation subset consisting of only 16 images. Due to the extremely limited sample size, this subset was not considered statistically sufficient to support a reliable validation process. Therefore, the predefined validation split was not used as a standalone validation set. Instead, a pseudo-validation strategy was employed by partitioning the training data into training and validation subsets. This internal split was used exclusively for hyperparameter tuning, early stopping, and model selection. The predefined test set was strictly held out and was not involved in any stage of model development. Final model evaluation was performed exclusively on this unseen test set, ensuring a more robust, unbiased, and comparable assessment of the model's generalization performance.

Table 1. Distribution of CXR Images in Training and Testing Sets.

	Normal	Pneumonia	Total
Train	1079 (27.0%)	3106 (74.2%)	4185 (71.5%)
Valid	270 (17.1%)	777 (18.2%)	1047 (17.9%)
Test	234 (14.8%)	390 (9.1%)	624 (10.6%)
Total	1583 (100%)	4273 (100%)	5856 (100%)

Backbone Network: EfficientNet-B1

EfficientNet-B1 is based on a compound scaling principle that uniformly scales network depth d, width w, and input resolution r using a set of predefined scaling coefficients. Instead of arbitrarily increasing one dimension, EfficientNet applies a balanced scaling strategy defined as:

$$d = \alpha^{\phi}, w = \beta^{\phi}, r = \gamma^{\phi} \qquad (1)$$

subject to the constraint:

$$\alpha \cdot \beta^2 \cdot \gamma^2 \approx 2 \qquad (2)$$

where ϕ is a user-defined scaling coefficient controlling model size, and α, β, γ are constants determined through grid search. This formulation ensures that depth, width, and resolution are scaled proportionally, leading to improved accuracy–efficiency trade-off compared to conventional single-dimension scaling strategies.

The fundamental building unit of EfficientNet-B1 is the Mobile Inverted Bottleneck Convolution (MBConv) block. Given an input tensor $\boldsymbol{X} \in \mathbb{R}^{H \times W \times C}$, the MBConv block first expands the channel dimension using a pointwise convolution:

$$X_e = \sigma(\mathrm{BN}(W_e * X)) \qquad (3)$$

where W_e denotes the expansion weights, $*$ represents convolution, BN denotes batch normalization, and $\sigma(\cdot)$ is the Swish activation function. This is followed by a depth wise convolution and a squeeze-and-excitation (SE) operation that adaptively recalibrates channel-wise responses. When the input and output dimensions match, a residual connection is applied:

$$Y = X + \mathcal{F}(X) \qquad (4)$$

where $\mathcal{F}(\cdot)$ denotes the MBConv transformation. This design improves parameter efficiency while maintaining strong representational capacity.

The detailed architectural configuration of the EfficientNet-B1 backbone is presented in Table 2. As shown, the network begins with a standard convolutional stem layer followed by a sequence of MBConv blocks with varying expansion ratios, kernel sizes, and channel depths. The spatial resolution is progressively reduced while the number of feature channels increases, enabling hierarchical feature extraction from low-level representations to high-level semantic features. Each MBConv block incorporates depth wise separable convolution, SE modules, Swish activation, and batch normalization. The backbone concludes with a 1×1 convolutional head and a global average pooling layer, producing a compact and information-rich feature representation for subsequent classification.

Preprocessing techniques

To analyze the effectiveness of preprocessing techniques in feature extraction, we applied a series of preprocessing methods.

Histogram equalization

Histogram equalization is a technique for adjusting image intensities to enhance contrast. The method transforms the intensity values so that the histogram of the output image approximately matches a uniform distribution. This can be mathematically represented as:

$$CDF(i) = \frac{\sum_{j=0}^{i} h(j)}{N} \qquad (5)$$

where $CDF(i)$ is the cumulative distribution function of intensity level i, $h(j)$ is the histogram value at intensity j, and N is the total number of pixels in the image. The new intensity value for each pixel is then given by:

$$I' = \mathrm{round}\big(\mathrm{CDF}(I) \times (L-1)\big) \qquad (6)$$

Table 2. Architecture of EfficientNet-B1 Backbone

Stage	Operator	Kernel	Output Resolution	#Channels	#Layers	Activation	SE	BatchNorm
Stem	Conv2D	3×3	240×240	32	1	Swish	No	Yes
1	MBConv1	3×3	240×240	16	1	Swish	Yes	Yes
2	MBConv6	3×3	120×120	24	2	Swish	Yes	Yes
3	MBConv6	5×5	60×60	40	2	Swish	Yes	Yes
4	MBConv6	3×3	30×30	80	3	Swish	Yes	Yes
5	MBConv6	5×5	15×15	112	3	Swish	Yes	Yes
6	MBConv6	5×5	15×15	192	4	Swish	Yes	Yes
7	MBConv6	3×3	15×15	320	1	Swish	Yes	Yes
Head	Conv2D	1×1	15×15	1280	1	Swish	No	Yes
Pooling	Global Avg Pool	–	1×1	1280	–	–	–	–

where I is the original intensity, I' is the new intensity, and L is the number of possible intensity levels.

Bilateral filtering

Bilateral filtering is a non-linear, edge-preserving, and noise reduction technique replaces the intensity of each pixel with a weighted average of intensity values from nearby pixels, considering both the spatial distance and the intensity difference. The filter's mathematical formulation is:

$$I'(x) = \frac{1}{W_p}\sum_{x_i \in \Omega} I(x_i) \cdot exp\left(-\frac{|x - x_i|^2}{2\sigma_d^2}\right) \cdot exp\left(-\frac{|I(x) - I(x_i)|^2}{2\sigma_r^2}\right) \tag{7}$$

where $I'(x)$ is the filtered image, $I(x_i)$ is the intensity of the pixel x_i in the neighborhood Ω of x, σ_d controls the spatial distance, σ_r controls the intensity difference, and W_p is the normalization factor defined as:

$$W_p = \sum_{x_i \in \Omega} exp\left(-\frac{|x - x_i|^2}{2\sigma_d^2}\right) \cdot exp\left(-\frac{|I(x) - I(x_i)|^2}{2\sigma_r^2}\right) \tag{8}$$

CLAHE (Contrast Limited Adaptive Histogram Equalization)

CLAHE is a variation of adaptive histogram equalization (AHE) that prevents over-amplification of noise by limiting the contrast. The image is divided into small contextual regions, and histogram equalization is applied to each. The method can be described as:

1. Divide the image into non-overlapping contextual regions (tiles).
2. Apply histogram equalization to each tile.
3. Clip the histogram at a predefined limit to avoid over-amplification.
4. Interpolate the intensity values in the boundaries of the tiles to avoid artifacts.

Mathematically, the clipping limit can be set as a threshold T, and the redistributed histogram h_clip is given by:

$$h_\mathrm{clip}(i) = min(h(i), T) + \frac{1}{n}\sum_{j=0}^{L-1} max(0, h(j) - T) \tag{9}$$

where $h(i)$ is the histogram value of intensity i, L is the number of intensity levels, and n is the number of pixels in the contextual region.

Gaussian filtering

Gaussian filtering is a linear smoothing filter used to reduce noise and detail in the image. The filter uses a Gaussian function to assign weights to the neighboring pixels. The mathematical representation of a Gaussian filter is:

$$G(x, y) = \frac{1}{2\pi\sigma^2} exp\left(-\frac{x^2 + y^2}{2\sigma^2}\right) \tag{10}$$

where $G(x, y)$ is the Gaussian function, and σ is the standard deviation of the Gaussian distribution. The filtered image I′ is obtained by convolving the original image I with the Gaussian function:

$$I'(x, y) = \sum_{u=-k}^{k} \sum_{v=-k}^{v} I(x - u, y - v) \cdot G(u, v) \tag{11}$$

where k is the kernel size, determined by the standard deviation σ.

Weight initialization approaches

This section describes the various model training approaches for training models on the CXR dataset. The approaches include random initialization, transfer learning with fine-tuning of only the top layers, and transfer learning with complete fine-tuning.

Random initialization

Training a model from scratch involves initializing the network with random weights and training it on the entire dataset. The process starts by defining the architecture of the neural network, which includes selecting the number of layers, types of layers (e.g., convolutional, fully connected), and activation functions. The weights are typically initialized randomly or using specific initialization methods such as Xavier [37] or He [38] initialization.

Training proceeds with a higher learning rate and continues through an optimization process, adjusting the weights based on the computed loss until predefined stopping criteria are met or the number of epochs is exhausted. This approach requires substantial amounts of data and computational resources to achieve good performance especially on large datasets. The training process can be mathematically described using the loss function L and the optimization algorithm, such as stochastic gradient descent (SGD), as follows:

$$\theta_{t+1} = \theta_t - \eta \nabla_\theta L(\theta_t) \tag{12}$$

where θ represents the model parameters, η is the learning rate, and $\nabla_\theta L(\theta_t)$ is the gradient of the loss function with respect to the parameters.

Transfer learning with fine-tuning only top layers

In this approach, a pre-trained model is used where only the top (dense) layers are fine-tuned. The process begins by selecting a pre-trained model that has been trained on a large dataset (e.g., ImageNet) and using it as a feature extractor. The pre-trained layers remain fixed, preserving the learned features from the original dataset. A new dense layer is added to adapt the model to the CXR dataset.

Fine-tuning involves training only the newly added top layers while keeping the weights of the pre-trained layers frozen. This method enables the model to adapt to specific features of the CXR images without altering the pre-trained features. This approach is particularly useful when dealing with limited data, as it leverages the pre-trained model's existing knowledge.

Transfer learning with complete fine-tuning

This method involves comprehensive fine-tuning of the pre-trained model, including both the top (dense) layers and the underlying layers. Initially, the network architecture is initialized using weights from a pre-trained model (e.g.,

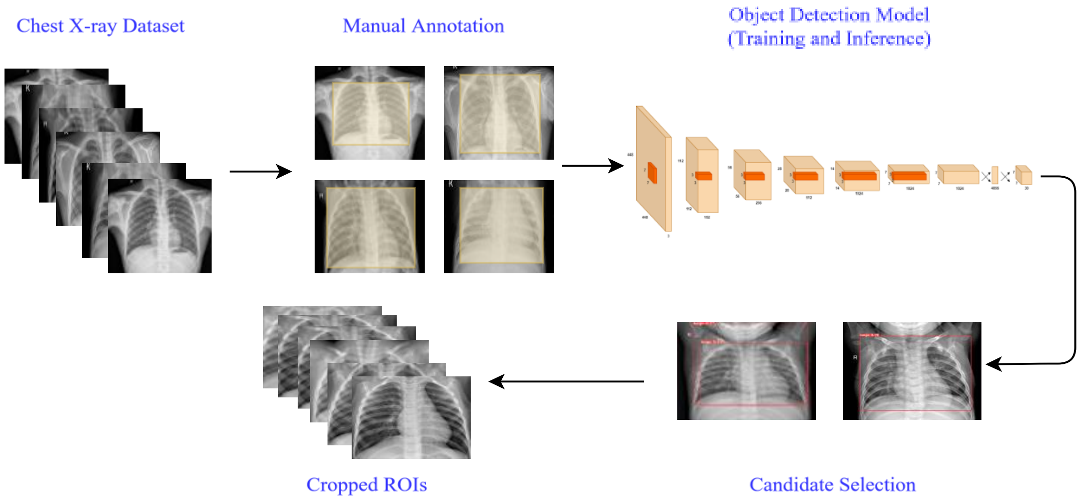

Figure 3. The workflow of the ROI estimation process

trained on ImageNet). The model is then further trained with a smaller learning rate to fine-tune all layers, adapting the weights to better suit the specific characteristics of the CXR dataset.

Complete fine-tuning allows the model to adjust both low-level and high-level features to better fit the new dataset. This approach typically involves adjusting hyperparameters such as the learning rate to balance between learning new features and preserving previously learned representations. The comprehensive adjustment of all layers often results in improved performance but requires more computational resources compared to using only the top layers.

Evaluation metrics

To evaluate model performance, we employed standard classification metrics, including accuracy A, precision P, recall R, and F1-score $F1$. Accuracy is defined as:

$$A = \frac{TP + TN}{TP + TN + FP + FN} \tag{13}$$

Precision, recall, and F1-score are computed as:

$$P = \frac{TP}{TP + FP} \tag{14}$$

$$R = \frac{TP}{TP + FN} \tag{15}$$

$$F1 = \frac{2 \cdot P \cdot R}{P + R} \tag{16}$$

where TP, TN, FP, and FN denote true positives, true negatives, false positives, and false negatives, respectively.

Experiments

The first stage of the experiments examined the effect of ROI selection on CXR image analysis. ROI selection was introduced to reduce the influence of irrelevant background regions and to guide the model's attention toward anatomically meaningful areas. In chest X-ray images, large portions of the image may contain non-diagnostic regions such as background artifacts, image borders, or acquisition-related markers. These regions may introduce unnecessary variability and distract the learning process. By focusing on the rib cage area, which contains the primary diagnostic structures relevant to pneumonia detection, ROI-based cropping aims to enhance feature learning by reducing noise and emphasizing clinically informative regions. Furthermore, restricting the input to a consistent anatomical region helps improve spatial alignment across samples, potentially leading to more stable and discriminative representations.

To identify the ROIs, an object detection model (YOLOv9) was employed [39]. For training the detection model, 100 images were manually annotated, including 50 pneumonia and 50 normal samples. The bounding boxes were defined to encompass the rib cage region in order to obtain anatomically consistent crops. Since the objective of the detection stage was coarse rib cage localization rather than fine-grained pathological detection, a limited yet representative annotation set was considered sufficient due to the relatively stable anatomical structure and spatial consistency of the rib cage in CXR images.

The performance of the trained YOLOv9 model was evaluated using standard object detection metrics, including precision, recall, and mAP@0.5. The detection model successfully localized the rib cage region in the majority of images. Rib cage detection failed in 35 images due to contrast variations. To avoid data exclusion bias and preserve dataset integrity, these images were retained in their original form without cropping.

After the detection stage, the dataset was processed using the predicted bounding boxes to obtain rib cage-centered crops. All images were subsequently resized to 256×256 pixels to standardize the input size for the classification network. The ROI estimation workflow, including annotation, detection model training, and bounding box selection, is illustrated in Figure 3. At this stage, both full-size and cropped images were used to train separate EfficientNet-B1 networks for comparative analysis.

In the second stage, we examined the impact of weight initialization strategies on the analysis of CXR images using

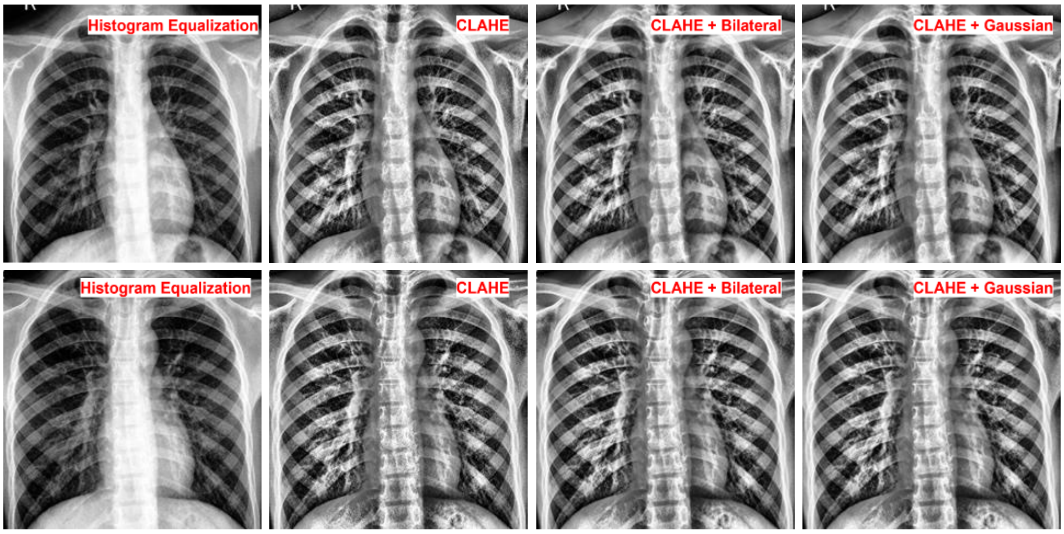

Figure 4. Example outputs for the preprocessing algorithms.

cropped inputs obtained from the ROI detection stage. To ensure a controlled comparison, the network architecture, optimizer settings, and training protocol were kept identical across configurations, while only the initialization strategy was varied. In the first configuration, the EfficientNet-B1 model was trained from scratch using random weight initialization. In the second configuration, the same architecture was initialized with ImageNet-pretrained weights and subsequently trained using a complete fine-tuning scheme. In this setting, all layers were updated during training, allowing the model to adapt both low-level and high-level feature representations to the target CXR dataset.

In the third stage, we evaluated the impact of different preprocessing techniques on the analysis of CXR images. The preprocessing methods investigated included histogram equalization, CLAHE, CLAHE combined with bilateral filtering, and CLAHE combined with Gaussian filtering. For the bilateral filter, the window size (Ω) was set to 9, with spatial standard deviation σ_d=75 and intensity standard deviation σ_r=75. For Gaussian filter, a kernel size of 5×5 and a standard deviation σ=1.5 were used. In the CLAHE configuration, the clipping limit (threshold) was set to 2.0 and the grid size to 8×8. Example outputs of the applied preprocessing algorithms are provided in Figure 4.

We conducted experiments under three settings. To standardize all experimental work, we used the EfficientNet B1 network with a batch size of 32 and input size of 256×256. The training was performed using the SGD optimizer with a cosine learning rate schedule with warm restarts [40]. For random initialization, the learning rate was set to 1×10^{-2} a warm restart period of 10 epochs. For fine-tuning, a lower learning rate of 1×10^{-4} was selected. The network was trained for 40 epochs, and no data augmentation techniques were used during training or testing, as our aim was to assess the differences between various preprocessing techniques. The hyperparameter settings and training configurations used across different experimental setups are summarized in Table 3. All experiments were conducted using the EfficientNet-B1 backbone with standardized input resolution and batch size. Differences between configurations are primarily related to initialization strategy and learning rate selection.

All experiments were conducted on a workstation equipped with an NVIDIA RTX 4090 GPU and a 32-core CPU. The experiments were implemented in Python 3.10 using the PyTorch framework.

Table 3. Hyperparameter Configuration Used in the Experiments

Parameter	Random Initialization	Fine-Tuning (Top Layers)	Fine-Tuning (Full Model)
Backbone	EfficientNet-B1	EfficientNet-B1	EfficientNet-B1
Input Size	256 × 256	256 × 256	256 × 256
Batch Size	32	32	32
Optimizer	SGD	SGD	SGD
Learning Rate	1×10^{-2}	1×10^{-4}	1×10^{-4}
LR Scheduler	Cosine Annealing + Warm Restarts	Cosine Annealing	Cosine Annealing
Warm Restart Period	10 epochs	–	–
Number of Epochs	40	40	40
Data Augmentation	None	None	None
Weight Initialization	Random	ImageNet Pretrained	ImageNet Pretrained

Results

Effect of ROI Determination

All experimental results are summarized in Table 4, which reports training, validation, and test accuracy alongside macro, weighted, and binary (Pneumonia-class) precision, recall, and F1-score for all configurations. The full-size dataset (Exp 1a) achieved a test accuracy of 0.857, with macro precision, recall, and F1-score of 0.884, 0.818, and 0.836, respectively, and a binary F1-score of 0.895. The cropped dataset (Exp 1b) achieved a test accuracy of 0.881, with macro precision, recall, and F1-score of 0.903, 0.849, and 0.865, and a binary F1-score of 0.912.

Effect of Weight Initialization

The model trained with random initialization (Exp 2a) achieved a test accuracy of 0.912, with macro precision, recall, and F1-score of 0.924, 0.889, and 0.902, and a binary F1-score of 0.933. The model trained with complete fine-tuning using ImageNet pre-trained weights (Exp 2b) achieved a test accuracy of 0.925, with macro precision, recall, and F1-score of 0.944, 0.900, and 0.916, and a binary F1-score of 0.943.

Effect of Preprocessing Techniques

Histogram equalization (Exp 3a) achieved a test accuracy of 0.934, with macro precision, recall, and F1-score of 0.945, 0.917, and 0.928, and a binary F1-score of 0.949. CLAHE (Exp 3b) achieved a test accuracy of 0.939, with macro precision, recall, and F1-score of 0.950, 0.922, and 0.933, and a binary F1-score of 0.953. CLAHE combined with bilateral filtering (Exp 3c) achieved a test accuracy of 0.946, with macro precision, recall, and F1-score of 0.952, 0.932, and 0.941, weighted F1-score of 0.945, and a binary F1-score of 0.958. CLAHE combined with Gaussian filtering (Exp 3d) achieved a test accuracy of 0.921, with macro precision, recall, and F1-score of 0.939, 0.898, and 0.913, and a binary F1-score of 0.940.

Learning Curves and Performance Metrics

The training and validation accuracy and loss curves for all experimental configurations are presented in Figure 5. In addition to the aggregated performance metrics reported in Table 4, the confusion matrices for all configurations are illustrated in Figure 6, providing a detailed class-wise breakdown of correct and incorrect predictions on the test set.

Discussion

Interpretation of Results

The results demonstrate that ROI determination, weight initialization strategies, and preprocessing techniques each have a significant impact on CXR classification performance.

To provide a comprehensive evaluation of model performance, three averaging strategies were employed for precision, recall, and F1-score: macro averaging, which assigns equal weight to each class regardless of support; weighted averaging, which accounts for class imbalance by weighting each class proportionally to its number of samples in the test set; and binary averaging, which evaluates performance exclusively on the Pneumonia class. Given the clinical objective of the task, detecting Pneumonia in chest X-rays, binary metrics are particularly relevant, as they directly reflect the model's sensitivity and precision with respect to the positive class. Reporting all three strategies together allows for a more transparent assessment of how class imbalance affects the reported scores and ensures that the results are interpretable from both a statistical and a clinical perspective.

Focusing on the ROI consistently enhanced test accuracy by directing the model's attention to anatomically relevant regions. Cropped images centered on the rib cage (Exp 1b) achieved a test accuracy of 0.881, compared to 0.857 for the full-size dataset (Exp 1a), indicating that removing

Table 4. Classification performance of all experimental configurations evaluated on the test set. Precision, Recall, and F1-Score are reported under three averaging strategies: macro (equal weight per class), weighted (proportional to class support), and binary (Pneumonia class only). The highest F1-Score in each averaging group is highlighted.

Exp #	Dataset & Weight Init	Preprocessing	Accuracy			Macro Avg			Weighted Avg			Binary (Pneumonia)		
			Train	Val	Test	P	R	F1	P	R	F1	P	R	F1
1a	Original + ImageNet (top layer finetune)	None	0.942	0.941	0.857	0.884	0.818	0.836	0.870	0.857	0.851	0.828	0.974	0.895
1b	Cropped + ImageNet (top layer finetune)	None	0.926	0.928	0.881	0.903	0.849	0.865	0.891	0.881	0.877	0.853	0.979	0.912
2a	Cropped + Random init (complete finetune)	None	0.975	0.943	0.912	0.924	0.889	0.902	0.916	0.912	0.910	0.890	0.979	0.933
2b	Cropped + ImageNet (complete finetune)	None	0.998	0.962	0.925	0.944	0.900	0.916	0.932	0.925	0.923	0.894	0.997	0.943
3a	Cropped + ImageNet (complete finetune)	Histogram equalization	0.999	0.959	0.934	0.945	0.917	0.928	0.937	0.934	0.933	0.914	0.987	0.949
3b	Cropped + ImageNet (complete finetune)	CLAHE	0.999	0.949	0.939	0.950	0.922	0.933	0.942	0.939	0.938	0.919	0.990	0.953
3c	Cropped + ImageNet (complete finetune)	CLAHE + Bilateral	0.999	0.958	0.946	0.952	0.932	**0.941**	0.947	0.946	**0.945**	0.932	0.985	**0.958**
3d	Cropped + ImageNet (complete finetune)	CLAHE + Gaussian	1.000	0.971	0.921	0.939	0.898	0.913	0.928	0.921	0.920	0.894	0.992	0.940

* P: Precision, R: Recall, F1: F1-Score

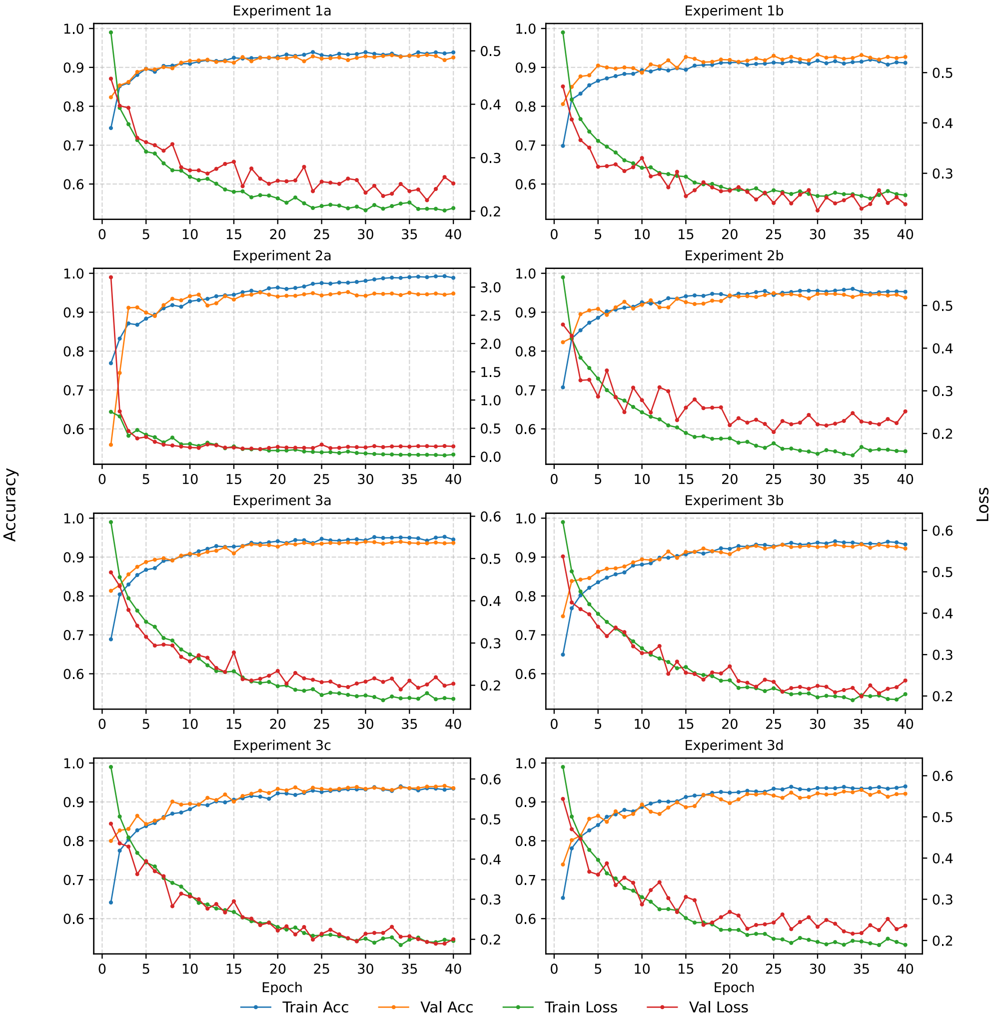

Figure 5. Training and validation performance curves for all experimental configurations. Each subplot shows the training and validation accuracy (left y-axis) and loss (right y-axis) over epochs for a specific experiment (1a–3d).

irrelevant background information allows the network to learn more discriminative features effectively, improving both learning efficiency and classification performance.

Weight initialization also had a substantial effect on performance. The model trained from scratch using random initialization (Exp 2a) achieved a test accuracy of 0.912, with macro precision, recall, and F1-score of 0.924, 0.889, and 0.902, whereas the same model with ImageNet-pretrained weights and complete fine-tuning (Exp 2b) reached a test accuracy of 0.925, with macro precision, recall, and F1-score of 0.944, 0.900, and 0.916. These results highlight the benefit of transfer learning, particularly for small to medium-sized datasets, where pre-trained features provide a strong foundation and reduce the computational resources required for training from scratch.

The impact of preprocessing techniques was similarly notable. Histogram equalization (Exp 3a) achieved a test accuracy of 0.934 with a macro F1-score of 0.928, demonstrating moderate improvement with stable performance across runs. CLAHE alone (Exp 3b) yielded a test accuracy of 0.939 and a macro F1-score of 0.933. The combination of CLAHE with bilateral filtering (Exp 3c) achieved the highest overall performance, with a test accuracy of 0.946, macro precision, recall, and F1-score of 0.952, 0.932, and 0.941, and a binary F1-score of 0.958, exceeding all other configurations across all averaging strategies. CLAHE combined with Gaussian filtering (Exp 3d) produced a lower test accuracy of 0.921 and a macro F1-score of 0.913, indicating that while contrast enhancement and noise smoothing can improve feature extraction, certain filter combinations may not generalize optimally.

To assess generalization, best validation accuracy was compared against the corresponding training accuracy. Experiments 1a and 1b, which employed top-layer fine-tuning, exhibited the smallest train–validation gap (at most 0.010), indicating that restricting trainable parameters effectively limits overfitting. When complete fine-tuning was applied, training accuracy rose substantially, reaching 0.998 to 1.000 in Exp 2b and Group 3, while best validation accuracy remained between 0.925 and 0.951, yielding gaps

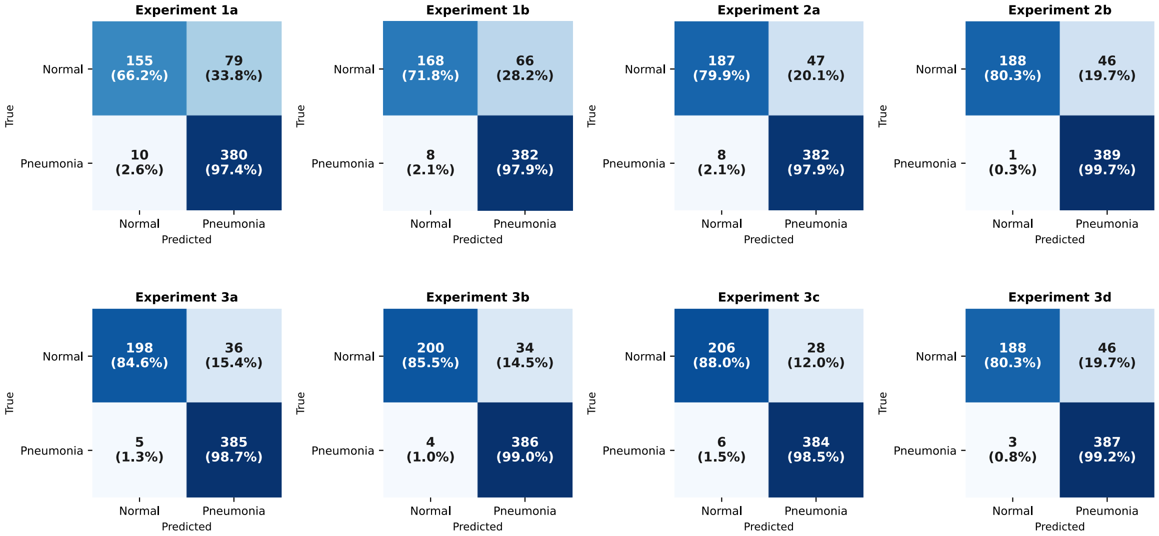

Figure 6. Confusion matrices for all experimental configurations evaluated on the test set

of 0.049 to 0.075. Among the fully fine-tuned configurations, Exp 2a achieved a comparatively lower training accuracy (0.975) alongside competitive validation accuracy (0.951), implying that random initialization acts as an implicit regularizer by slowing convergence. Within Group 3, preprocessing did not produce a consistent improvement in validation accuracy; Exp 3c and Exp 3a tied for the highest validation accuracy (0.939), while Exp 3d yielded the lowest (0.925) despite perfect training accuracy.

Across all three averaging strategies, Exp 3c (CLAHE + Bilateral filtering with complete fine-tuning using ImageNet initialization) consistently achieved the highest F1-score, with values of 0.941 (macro), 0.945 (weighted), and 0.958 (binary). This consistency across different evaluation criteria indicates that Exp 3c does not merely perform well under a single metric but generalizes effectively across both classes while maintaining strong sensitivity to Pneumonia cases specifically, making it the most robust configuration among all those evaluated.

Overall, these findings emphasize that performance gains in CXR classification can be obtained not only through architectural modifications but also via careful design of preprocessing pipelines, ROI-based cropping, and informed weight initialization strategies. Each component independently contributes to enhancing model generalizability, stability, and efficiency, highlighting the importance of a systematic evaluation of methodological factors beyond network architecture alone.

Comparison with existing studies

We compare the performance and methodology of our proposed approach with existing studies in the field of CXR analysis. The studies, their methodological approaches, datasets used, and their accuracy values are provided in Table 5.

Sharma et al. [41] proposed a VGG-Net-like architecture for detecting lung diseases from CXR images. They trained four different models with varying combinations of data augmentation and dropout techniques. The combination of data augmentation and dropout achieved an accuracy of 90.68%, with an input size of 64 x 64 pixels. Reshan et al. [42] utilized a MobileNet model for detecting pneumonia from CXR images. They trained and evaluated a total of eight pre-trained models on NIH CXR [43] and pediatric CXR [35] datasets. To address class imbalance, they applied data augmentation to equalize the number of pneumonia and non-pneumonia samples. Among these models, MobileNet achieved the highest accuracy of 94.23%. Chouhan et al. [44] proposed an ensemble learning approach for the computer-aided classification of pneumonia in CXR images. Their method leverages DenseNet169, MobileNetV2, and ViT, which were fine-tuned on pediatric CXR dataset [35]. By combining the features extracted from these models and utilizing random augmentation techniques, their approach achieved an accuracy of 93.91%. Gazda et al. [45] introduced a self-supervised deep neural network for CXR classification, utilizing contrastive learning to pretrain the model on an unlabeled dataset. Their method achieved an accuracy of 91.50% on the pediatric CXR dataset [35] without requiring large amounts of labeled training data.

Our proposed methodology leveraged ROI selection to focus the model on anatomically relevant regions, thereby enhancing classification performance. Subsequently, we applied complete fine-tuning using ImageNet-pretrained weights, enabling the network to adapt both low- and high-level features to the target CXR dataset. Finally, incorporating CLAHE combined with Gaussian filtering for image preprocessing yielded a peak classification accuracy of 95.80%. We anticipate that further improvements could be achieved through the adoption of alternative network architectures, higher-resolution inputs, and fine-tuning

Table 5. Comparison of Methodological Approaches, Datasets, and Accuracy for Various Studies.

Study	Approach	Dataset	Accuracy
Sharma et al. [41]	VGG-Net-like architecture with data augmentation and dropout	Pediatric CXR [35]	90.68%
Gazda et al. [45]	Self-supervised deep neural network with contrastive learning	Pediatric CXR [35]	91.50%
Chouhan et al. [44]	Ensemble learning with three networks; random augmentation	Pediatric CXR [35]	93.91%
Reshan et al. [42]	MobileNet model; data augmentation to address class imbalance	Pediatric CXR [35]	94.23%
Proposed	**ROI; complete finetune; clahe + gaussian preprocessing**	**Pediatric CXR** [35]	**94.60%**

strategies with finer granularity and shorter learning schedules.

Limitations

The dataset used in this study, although sufficient for initial experimentation, is relatively small and specific to pediatric patients from a single medical center. This limits the generalizability of the findings to broader populations and different age groups. Also, the dataset's demographic homogeneity may not capture the variability in CXR images that could arise from different ethnicities, geographical locations, or adult populations.

While several preprocessing techniques were applied, we did not explore all possible methods. There may be other techniques that could further enhance image quality and model performance. Additionally, some preprocessing methods might not be applicable to CXR images as they could alter the inherent nature of the data. For example, techniques that adjust color distributions could inadvertently obscure or modify critical diagnostic features, such as white shadows that are indicative of certain pathologies in CXRs. We evaluated the impact of preprocessing techniques without incorporating data augmentation, which is commonly used in practice to enhance model robustness.

The study focused on EfficientNet B1 and did not extensively compare other state-of-the-art architectures. While EfficientNet is known for its performance, other models might offer different advantages. The decision to limit the comparison was influenced by processing load considerations and computational constraints, which restricted the scope of the study.

The ROI selection process relied on an object detection model trained on a small subset of labeled data. This might introduce bias or errors if the object detection model is not generalizable to all types of CXR images. Also, images where the rib cage was not detected were included in the dataset without modifications, potentially affecting the consistency of the dataset.

Challenges

We faced complexities in determining the optimal weight initialization strategy and fine-tuning approach, as balancing between learning new features and preserving pre-trained knowledge was challenging, particularly with a smaller dataset.

The inherent stochastic nature of deep learning training processes posed difficulties in consistently evaluating model performance. Variability in results necessitated the introduction of additional metrics to provide a more robust measure of performance. We ensured reproducibility and stability of the results across different runs by carefully managing random seeds and training conditions.

We encountered significant computational demands when training deep learning models, especially with fine-tuning and multiple preprocessing techniques, and limited access to high-performance computing facilities slowed down the experimentation process.

Conclusion

In this study, we assessed the influence of various preprocessing techniques and weight initialization strategies on deep neural networks for CXR image analysis. Our experiments revealed that ROI selection significantly enhances model performance by focusing on relevant image areas . We observed that using pre-trained ImageNet weights with complete fine-tuning outperformed random initialization, highlighting the benefits of transfer learning, especially with limited data. Among preprocessing techniques, combining CLAHE with Gaussian filtering achieved the highest validation accuracy (94.60%), indicating that advanced preprocessing can substantially improve model performance. Other methods showed varying results, emphasizing the importance of selecting appropriate preprocessing strategies for specific applications.

The findings of this study have significant implications for clinical practice and medical imaging analysis. Effective ROI selection and advanced preprocessing techniques can enhance diagnostic accuracy and efficiency in analyzing CXRs. Implementing these methods in clinical settings could lead to more accurate and timely diagnoses, potentially improving patient outcomes. Additionally, leveraging transfer learning with pre-trained models can be particularly advantageous when working with smaller datasets, offering a practical approach for developing robust diagnostic tools in resource-constrained environments.

Future research should explore additional preprocessing techniques and their impact on model performance, as well as evaluate other state-of-the-art architectures beyond EfficientNet B1. Investigating how data augmentation methods could further enhance model robustness and accuracy is also crucial. Additionally, expanding the dataset to include a wider variety of cases and incorporating other imaging modalities could provide a more comprehensive understanding of model performance and applicability. Further studies could also examine the integration of real-time image analysis systems in clinical practice to assess their practical benefits and limitations in diverse healthcare settings.

References

[1] WHO, "The top 10 causes of death." Accessed: Aug. 03, 2024. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death

[2] WHO, "Chronic obstructive pulmonary disease (COPD)." Accessed: Aug. 03, 2024. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/chronic-obstructive-pulmonary-disease-(copd)

[3] WHO, "Lung cancer." Accessed: Aug. 03, 2024. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/lung-cancer

[4] C. M. Jones *et al.*, "Chest radiographs and machine learning – Past, present and future," *J. Med. Imaging Radiat. Oncol.*, vol. 65, no. 5, pp. 538–544, 2021, doi: 10.1111/1754-9485.13274.

[5] M. A. Thanoon, M. A. Zulkifley, M. A. A. Mohd Zainuri, and S. R. Abdani, "A Review of Deep Learning Techniques for Lung Cancer Screening and Diagnosis Based on CT Images," *Diagnostics*, vol. 13, no. 16, 2023, doi: 10.3390/diagnostics13162617.

[6] C. Deniz, "Modern Computer Tomography with Artificial Intelligence and Deep Learning Applications," *Artif. Intell. Theory Appl.*, vol. 3, no. 2, pp. 123–136, 2023.

[7] R. Booij, R. P. J. Budde, M. L. Dijkshoorn, and M. van Straten, "Technological developments of X-ray computed tomography over half a century: User's influence on protocol optimization," *Eur. J. Radiol.*, vol. 131, p. 109261, Oct. 2020, doi: 10.1016/j.ejrad.2020.109261.

[8] J. M. Jiang, L. Miao, X. Liang, Z. H. Liu, L. Zhang, and M. Li, "The Value of Deep Learning Image Reconstruction in Improving the Quality of Low-Dose Chest CT Images," *Diagnostics*, vol. 12, no. 10, p. 2560, Oct. 2022, doi: 10.3390/diagnostics12102560.

[9] P. Celard, E. L. Iglesias, J. M. Sorribes-Fdez, R. Romero, A. S. Vieira, and L. Borrajo, "A survey on deep learning applied to medical images: from simple artificial neural networks to generative models," *Neural Comput. Appl.*, vol. 35, no. 3, pp. 2291–2323, 2023, doi: 10.1007/s00521-022-07953-4.

[10] X. Xie, J. Niu, X. Liu, Z. Chen, S. Tang, and S. Yu, "A survey on incorporating domain knowledge into deep learning for medical image analysis," *Med. Image Anal.*, vol. 69, 2021, doi: 10.1016/j.media.2021.101985.

[11] E. Çallı, E. Sogancioglu, B. van Ginneken, K. G. van Leeuwen, and K. Murphy, "Deep learning for chest X-ray analysis: A survey," *Med. Image Anal.*, vol. 72, 2021, doi: 10.1016/j.media.2021.102125.

[12] D. Meedeniya, H. Kumarasinghe, S. Kolonne, C. Fernando, I. D. la T. Díez, and G. Marques, "Chest X-ray analysis empowered with deep learning: A systematic review," *Appl. Soft Comput.*, vol. 126, 2022, doi: 10.1016/j.asoc.2022.109319.

[13] F. Shamshad *et al.*, "Transformers in medical imaging: A survey," in *Medical Image Analysis*, 2023. doi: 10.1016/j.media.2023.102802.

[14] A. Dosovitskiy *et al.*, "an Image Is Worth 16X16 Words: Transformers for Image Recognition At Scale," in *ICLR 2021 - 9th International Conference on Learning Representations*, 2021.

[15] X. Wu *et al.*, "CheXNet: Combing Transformer and CNN for Thorax Disease Diagnosis from Chest X-ray Images," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2024, pp. 73–84. doi: 10.1007/978-981-99-8558-6_7.

[16] C. C. Ukwuoma *et al.*, "A hybrid explainable ensemble transformer encoder for pneumonia identification from chest X-ray images," *J. Adv. Res.*, vol. 48, pp. 191–211, 2023, doi: 10.1016/j.jare.2022.08.021.

[17] O. F. Soylemez and B. Ergen, "Facial Landmark Based Region of Interest Localization for Deep Facial Expression Recognition," *Teh. Vjesn.*, vol. 29, no. 1, pp. 38–44, 2022, doi: 10.17559/TV-20200423145443.

[18] R. G, "A Study to Find Facts Behind Preprocessing on Deep Learning Algorithms," *J. Innov. Image Process.*, vol. 3, no. 1, pp. 66–74, 2021, doi: 10.36548/jiip.2021.1.006.

[19] M. V. Narkhede, P. P. Bartakke, and M. S. Sutaone, "A review on weight initialization strategies for neural networks," *Artif. Intell. Rev.*, vol. 55, no. 1, pp. 291–322, 2022, doi: 10.1007/s10462-021-10033-z.

[20] U. Avni, H. Greenspan, E. Konen, M. Sharon, and J. Goldberger, "X-ray categorization and retrieval on the organ and pathology level, using patch-based visual words," *IEEE Trans. Med. Imaging*, vol. 30, no. 3, pp. 733–746, 2011, doi: 10.1109/TMI.2010.2095026.

[21] M. R. Zare, W. C. Seng, and A. Mueen, "Automatic classification of medical X-ray

images," *Malaysian J. Comput. Sci.*, vol. 26, no. 1, pp. 9–22, 2013, doi: 10.22452/mjcs.vol26no1.2.

[22] N. R. S. Parveen and M. M. Sathik, "Detection of Pneumonia in chest X-ray images," *J. Xray. Sci. Technol.*, vol. 19, no. 4, pp. 423–428, 2011, doi: 10.3233/XST-2011-0304.

[23] L. Houam, A. Hafiane, A. Boukrouche, E. Lespessailles, and R. Jennane, "Texture characterization using local binary pattern and wavelets. Application to bone radiographs," *2012 3rd Int. Conf. Image Process. Theory, Tools Appl. IPTA 2012*, pp. 371–376, 2012, doi: 10.1109/IPTA.2012.6469546.

[24] V. Kumar and C. Science, "Neural Network Based Approach for Detection of Abnormal Regions of Lung Cancer in X-Ray Image," vol. 1, no. 5, pp. 1–7, 2012.

[25] S. Jaeger *et al.*, "Automatic tuberculosis screening using chest radiographs," *IEEE Trans. Med. Imaging*, vol. 33, no. 2, pp. 233–245, Feb. 2014, doi: 10.1109/TMI.2013.2284099.

[26] H. Arimura, S. Katsuragawa, Q. Li, T. Ishida, and K. Doi, "Development of a computerized method for identifying the posteroanterior and lateral views of chest radiographs by use of a template matching technique," *Med. Phys.*, vol. 29, no. 7, pp. 1556–1561, 2002, doi: 10.1118/1.1487426.

[27] K. C. Santosh *et al.*, "Rotation detection in chest radiographs based on generalized line histogram of rib-orientations," *Proc. - IEEE Symp. Comput. Med. Syst.*, pp. 138–142, 2014, doi: 10.1109/CBMS.2014.56.

[28] Ö. F. Söylemez, "Forward selection-based ensemble of deep neural networks for melanoma classification in dermoscopy images," *Int. J. Imaging Syst. Technol.*, vol. 33, no. 6, pp. 1929–1943, 2023, doi: 10.1002/ima.22912.

[29] Q. Wang *et al.*, "Automated segmentation and diagnosis of pneumothorax on chest X-rays with fully convolutional multi-scale ScSE-DenseNet: a retrospective study," *BMC Med. Inform. Decis. Mak.*, vol. 20, 2020, doi: 10.1186/s12911-020-01325-5.

[30] M. Owais, M. Arsalan, T. Mahmood, Y. H. Kim, and K. R. Park, "Comprehensive computer-aided decision support framework to diagnose tuberculosis from chest X-Ray images: Data mining study," *JMIR Med. Informatics*, vol. 8, no. 12, 2020, doi: 10.2196/21790.

[31] S. Rajaraman, S. Sornapudi, P. O. Alderson, L. R. Folio, and S. K. Antani, "Analyzing inter-reader variability affecting deep ensemble learning for COVID-19 detection in chest radiographs.," *PLoS One*, vol. 15, no. 11, p. e0242301, 2020, [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/33180877/

[32] X. Wang *et al.*, "Potential of deep learning in assessing pneumoconiosis depicted on digital chest radiography," *Occup. Environ. Med.*, vol. 77, no. 9, pp. 597–602, 2020, doi: 10.1136/oemed-2019-106386.

[33] G. I. Okolo, S. Katsigiannis, and N. Ramzan, "IEViT: An enhanced vision transformer architecture for chest X-ray image classification," *Comput. Methods Programs Biomed.*, vol. 226, 2022, doi: 10.1016/j.cmpb.2022.107141.

[34] T. Chen *et al.*, "A vision transformer machine learning model for COVID-19 diagnosis using chest X-ray images," *Healthc. Anal.*, vol. 5, 2024, doi: 10.1016/j.health.2024.100332.

[35] D. Kermany, K. Zhang, M. Goldbaum, and others, "Labeled optical coherence tomography (oct) and chest x-ray images for classification," *Mendeley data*, vol. 2, no. 2, p. 651, 2018.

[36] D. S. Kermany *et al.*, "Identifying Medical Diagnoses and Treatable Diseases by Image-Based Deep Learning," *Cell*, vol. 172, no. 5, pp. 1122-1131.e9, 2018, doi: 10.1016/j.cell.2018.02.010.

[37] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," *J. Mach. Learn. Res.*, vol. 9, pp. 249–256, 2010.

[38] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on imagenet classification," in *Proceedings of the IEEE International Conference on Computer Vision*, 2015, pp. 1026–1034. doi: 10.1109/ICCV.2015.123.

[39] C.-Y. Wang, I.-H. Yeh, and H.-Y. M. Liao, "YOLOv9: Learning What You Want to Learn Using Programmable Gradient Information," 2024, [Online]. Available: http://arxiv.org/abs/2402.13616

[40] I. Loshchilov and F. Hutter, "SGDR: Stochastic gradient descent with warm restarts," in *5th International Conference on Learning Representations, ICLR 2017*, 2017. doi: 10.48550/arXiv.1608.03983.

[41] H. Sharma, J. S. Jain, P. Bansal, and S. Gupta, "Feature extraction and classification of chest X-ray images using CNN to detect pneumonia," in *Proceedings of the Confluence 2020 - 10th International Conference on Cloud Computing, Data Science and Engineering*, 2020, pp. 227–231. doi: 10.1109/Confluence47617.2020.9057809.

[42] M. S. Al Reshan *et al.*, "Detection of Pneumonia from Chest X-ray Images Utilizing MobileNet Model," *Healthcare*, vol. 11, no. 11, p. 1561, May 2023, doi: 10.3390/healthcare11111561.

[43] X. Wang, Y. Peng, L. Lu, Z. Lu, M. Bagheri, and R. M. Summers, “ChestX-ray8: Hospital-scale chest X-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases,” *Proc. - 30th IEEE Conf. Comput. Vis. Pattern Recognition, CVPR 2017*, vol. 2017-Janua, pp. 3462–3471, 2017, doi: 10.1109/CVPR.2017.369.

[44] V. Chouhan *et al.*, “A novel transfer learning based approach for pneumonia detection in chest X-ray images,” *Appl. Sci.*, vol. 10, no. 2, 2020, doi: 10.3390/app10020559.

[45] M. Gazda, J. Plavka, J. Gazda, and P. Drotar, “Self-Supervised Deep Convolutional Neural Network for Chest X-Ray Classification,” *IEEE Access*, vol. 9, pp. 151972–151982, 2021, doi: 10.1109/ACCESS.2021.3125324.