Volume XI Issue 1 Year 2026 | Page 215-225 ISSN: 2527-9866
Received: ***20-12-2025*** *|* *Revised:* ***12-01-2026*** *|* *Accepted:* ***27-01-2026***

Comparison of MobileNetV3-Small and EfficientNetV2-Small for Low-Resolution X-ray Image Classification

Andhika Rizky Cahya Putra[1], Siska Devella[2]
[1,2] Multi Data Palembang University, Palembang, South Sumatra, Indonesia, 30113
e-mail: andhikarizkycahyaputra_2226250071@mhs.mdp.ac.id[1], siskadevella@mdp.ac.id[2]
*Correspondence: andhikarizkycahyaputra_2226250071@mhs.mdp.ac.id

Abstract: *Lung diseases are a global health concern that require accurate and efficient automated diagnostic systems, particularly in healthcare facilities with limited resources. This study evaluates the performance and computational efficiency of two lightweight convolutional neural network architectures, namely MobileNetV3-Small and EfficientNetV2-Small, on the multi-label classification task of low-resolution ChestMNIST chest X-ray images. Experiments were conducted across eight testing scenarios with and without light spatial data augmentation. The evaluation encompassed predictive performance using accuracy and Area Under the Curve (AUC-ROC) metrics, as well as computational efficiency based on the number of parameters, FLOPs, model size, training time, and inference time. Results indicated that although both models achieved high accuracy (0.93–0.95), MobileNetV3-Small consistently produced higher and more stable AUC-ROC values compared to EfficientNetV2-Small, while being significantly more computationally efficient. Moreover, the application of light spatial data augmentation on low-resolution datasets such as ChestMNIST did not provide consistent performance improvements and instead increased training costs, indicating the limited effectiveness of simple geometric variations when spatial information in the images is highly constrained. These findings provide insight that, in low-resolution medical image multi-label classification, the suitability of an efficient CNN architecture design has a greater impact on overall performance than increasing model complexity or applying light spatial augmentation.*

Keywords: *MobileNetV3-Small, EfficientNetV2-Small, ChestMNIST, Low-Resolution X-ray Images*

1. Introduction

Lung diseases remain one of the major global health challenges with significant impact. According to the World Health Organization (WHO) [1], in 2022, an estimated 10.6 million tuberculosis (TB) cases were reported worldwide, resulting in approximately 1.30 million deaths from the disease. In addition, the global burden of chronic respiratory diseases (CRDs) is substantial. According to the Global Burden of Disease Study 2019, approximately 454.6 million people worldwide suffered from chronic lung conditions, including asthma and chronic obstructive pulmonary disease (COPD) [2]. Among these, about 212.3 million were estimated to have COPD, while 262.4 million were affected by asthma [2]. COPD accounts for millions of deaths annually, with the majority of premature fatalities occurring in low- and middle-income countries [3]. Therefore, early identification of lung diseases plays a crucial role in reducing mortality and controlling transmission, highlighting the need for diagnostic systems that are rapid, accurate, and applicable in resource-constrained healthcare settings.

Chest radiography (chest X-ray) is a widely accessible and commonly used imaging modality for initial assessment of lung diseases, though more advanced techniques like CT scans may be needed for detailed diagnosis [4]. Chest X-ray (CXR) images are used to diagnose conditions such as pneumonia, tuberculosis, and COVID-19 [5]. However, Manual interpretation of chest X-ray

images can be challenging, and recent research has explored the use of machine learning and deep learning-based AI to support and potentially improve diagnostic performance from radiographs [6]. Advances in artificial intelligence have significantly impacted medical image analysis, and deep learning (a subset of machine learning) is able to automatically extract complex features from image data, which has been widely explored in recent research [7]. Within this domain, Convolutional Neural Networks (CNNs) have become the most widely used algorithms because they have shown high classification performance compared with traditional approaches [8]. Numerous chest radiography studies employ various CNN architectures such as VGG, ResNet, DenseNet, Inception, and EfficientNet for lung disease classification [5].

Several large models achieve near-perfect accuracy. For instance, Shazia et al. [9] reported accuracies of 99.32% for ResNet50 and 99.48% for DenseNet121 in binary classification of COVID-19 versus pneumonia. Sadoon et al. [10] demonstrated 95.49% accuracy in nine-class lung disease classification using DenseNet201. Despite their high performance, large architectures such as VGG, ResNet, and DenseNet require substantial computational resources. ResNet demands high processing power, while DenseNet is prone to memory bottlenecks due to its feature reuse mechanism [11], making them less suitable for deployment on resource-constrained devices such as edge devices [12]. In addition to architectural factors, image resolution also affects computational load. Reducing resolution can significantly decrease FLOPs, the number of parameters, training time, and inference time, albeit with a slight reduction in accuracy [13]. This creates a trade-off between accuracy and efficiency. To address efficiency issues on resource-constrained devices, lightweight CNN architectures such as MobileNet and EfficientNet have been increasingly adopted. These architectures are designed to be efficient without significantly compromising accuracy [14]. In the medical domain, both architectures have demonstrated excellent performance. Zhao et al. [15] employed EfficientNetV2-S for liver fibrosis classification, achieving over 92% accuracy for distinguishing specific fibrosis stages, while Sriwiboon and Phimphisan [16] reported that MobileNetV3-Small reached 99.5% accuracy on COVID-19 X-ray images with a model size of only 2.5 MB. These architectures are also effective on low-resolution images. MobileNetV3 achieved 88.93% accuracy on the CIFAR-10 dataset with 32×32 pixel images [17], whereas EfficientNetV2-S reached 98.7% accuracy at the same resolution [18]. This demonstrates the capability of both architectures to efficiently handle low-resolution images.

Although numerous studies have applied CNNs for X-ray image classification, most have focused on improving accuracy without considering computational efficiency. In real-world applications, particularly in healthcare facilities with limited hardware, factors such as model size, memory requirements, and inference speed are equally important as accuracy. Previous research has also shown that using low-resolution images can reduce computational load without drastically compromising accuracy, making the combination of low-resolution images and lightweight CNN architectures highly promising for clinical scenarios. Therefore, this study focuses on evaluating MobileNetV3-Small and EfficientNetV2-S in the multi-label classification of low-resolution ChestMNIST images, encompassing accuracy, Area Under the Curve (AUC), computational efficiency, and performance stability with augmentation, to provide a realistic assessment of lightweight CNN usage in X-ray-based diagnostic systems.

2.Methods

This study was conducted through experiments on two lightweight deep learning models, namely MobileNetV3-Small and EfficientNetV2-Small, focusing on the classification of low-resolution X-ray images from the multi-label ChestMNIST dataset for lung disease classification. The experiments were divided into eight testing scenarios, comprising four scenarios for the EfficientNetV2-Small model and four scenarios for the MobileNetV3-Small model. For each model, testing was performed with and without the application of data augmentation on the training data, employing various augmentation methods to evaluate their effects on model performance.

The details of each experimental scenario are presented in Table 1, providing a systematic overview of the model configurations used.

Table 1. Experimental Skenario

	Scenarios	Architecture	Augmentation
1.	Scenario 1	MobileNetV3-Small	None
2.	Scenario 2	MobileNetV3-Small	Rotation
3.	Scenario 3	MobileNetV3-Small	Gaussian Blur
4.	Scenario 4	MobileNetV3-Small	Translation
5.	Scenario 5	EfficientNetV2-Small	None
6.	Scenario 6	EfficientNetV2-Small	Rotation
7.	Scenario 7	EfficientNetV2-Small	Gaussian Blur
8.	Scenario 8	EfficientNetV2-Small	Translation

To ensure a fair comparison, all experimental scenarios followed the same testing workflow and configuration. The systematic steps of the general testing scenario in this study are illustrated in Figure 1.

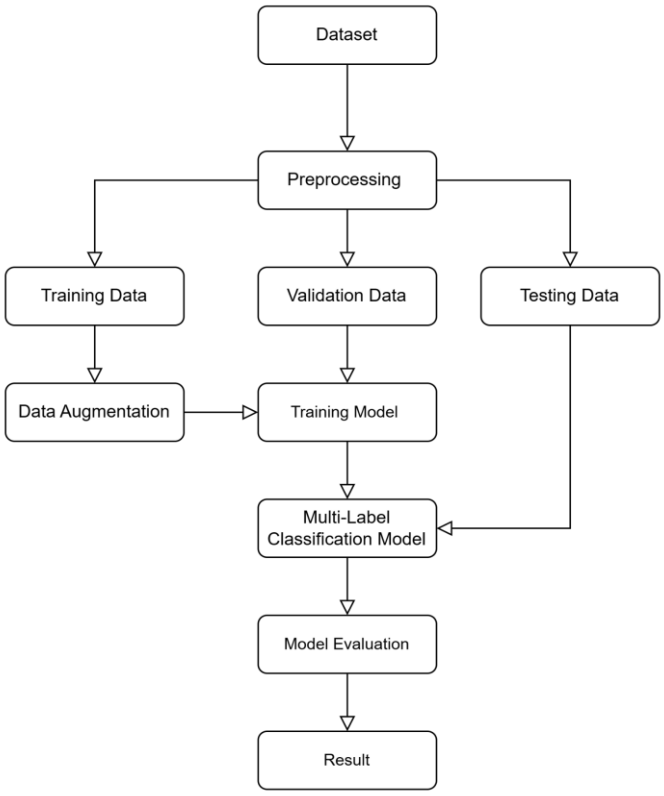

Figure 1. General Experimental Scenarios

A. Dataset Collection

In the initial phase of this study, dataset collection was conducted. The data used consisted of chest X-ray images obtained from the public ChestMNIST dataset, which is part of the MedMNIST v2 medical dataset collection [19]. This dataset is used for chest X-ray-based lung disease classification and has been widely utilized in deep learning research within the medical domain. All images in the ChestMNIST dataset have a resolution of 28 × 28 pixels and are in grayscale format.

The ChestMNIST dataset comprises 14 lung disease labels, where each image may have more than one label, classifying the task as multi-label classification. The dataset provides predefined splits for training, validation, and testing sets, eliminating the need for manual data partitioning. Additionally, the distribution of samples across labels is imbalanced, which poses a challenge in model training and classification performance evaluation [19]. Sample images from the dataset are shown in Figure 2, and the independent label distribution in the ChestMNIST dataset is presented in Table 2

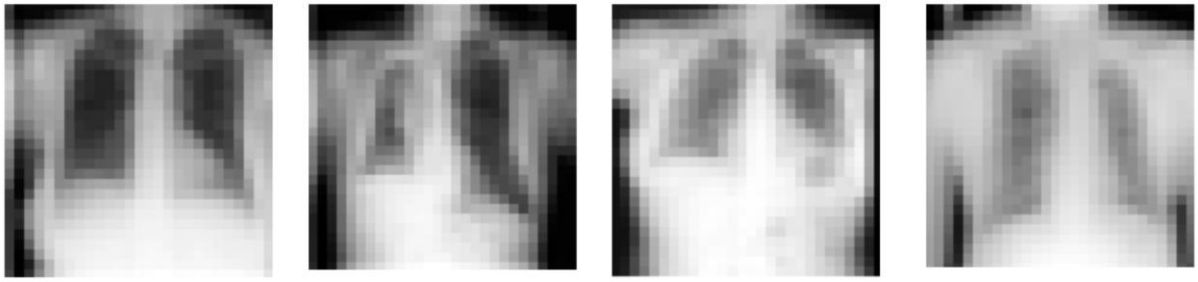

Figure 2. Sample Dataset ChestMNIST

Table 2. Independent Label Distribution in the ChestMNIST Dataset

	Label (Disease)	Train	Val	Test	Total Occurrences
1	atelectasis	7,996	1,119	2,420	11,535
2	cardiomegaly	1,950	240	582	2,772
3	effusion	9,261	1,292	2,754	13,307
4	**infiltration**	**13,914**	**2,018**	**3,938**	**19,870**
5	mass	3,988	625	1,133	5,746
6	nodule	4,375	613	1,335	6,323
7	pneumonia	978	133	242	1,353
8	pneumothorax	3,705	504	1,089	5,298
9	consolidation	3,263	447	957	4,667
10	edema	1,690	200	413	2,303
11	emphysema	1,799	208	509	2,516
12	fibrosis	1,158	166	362	1,686
13	pleural	2,279	372	734	3,385
14	**hernia**	**144**	**41**	**42**	**227**

B. Preprocessing and Augmentation

The preprocessing stage was carried out to prepare the dataset for optimal use and compatibility with the CNN models. The ChestMNIST dataset, which had already been split into training, validation, and test sets, was processed through several preprocessing steps. All X-ray images were converted to three channels to ensure compatibility with the input of MobileNetV3-Small and EfficientNetV2-Small models using pretrained ImageNet weights. Subsequently, the images were converted into PyTorch tensors with pixel values scaled to [0,1] using ToTensor(), followed by normalization using a mean of [0.485, 0.456, 0.406] and a standard deviation of [0.229, 0.224, 0.225], in accordance with ImageNet normalization standards.

Data augmentation was also applied to the training data based on prior studies addressing augmentation methods for X-ray and MRI images [20]. The augmentations used included light rotation up to 15 degrees, slight translation within a range of 10%, and Gaussian blur. Each type of augmentation was applied separately in an ablation study scheme to analyze the effect of each augmentation on model performance. No augmentation was applied to the validation and test data to maintain the objectivity of model evaluation. With these preprocessing steps, the ChestMNIST X-ray dataset was prepared for model training across the various experimental scenarios established in this study.

C. MobileNetV3-Small

MobileNetV3-Small is a lightweight CNN architecture that integrates layers from MobileNetV1 and MobileNetV2 to form a more efficient model [21]. The Swish activation function is modified to Hard-Swish, and the sigmoid function is replaced with Hard-Sigmoid to enhance computational efficiency. The Squeeze-and-Excite (SE) module is applied to residual layers, and activation functions vary across layers [21]. In the final layer of MobileNetV3-Small, a redesign places a 1×1 convolution layer after average pooling, allowing the final features to be computed at a 1×1 resolution. This design reduces latency by 11% and decreases operational complexity by 30 million MAdds, while the number of initial filters is reduced from 32 to 16 for inference efficiency without compromising accuracy [21]. The efficient last-stage design enables the removal of three computationally heavy final layers without sacrificing performance.

D. EfficientNetV2-Small

EfficientNetV2-Small is designed for parameter efficiency and high training speed using training-aware neural architecture search [18]. This architecture introduces Fused-MBConv blocks that replace some of the previous MBConv blocks, resulting in a smaller and faster model without significantly compromising accuracy. The small variant has 22 million parameters, 8.8 billion FLOPs, 83.9% accuracy, and an inference time of 24 ms per batch [18]. Its architectural structure consists of an initial 3×3 convolution, followed by several stages of Fused-MBConv (expansion ratios 1 and 4), MBConv stages (expansion ratios 4 and 6) with Squeeze-and-Excitation, and a final stage of 1×1 convolution + pooling + fully connected layer producing 1,280 output channels. The last stage also removes stride-1 from EfficientNetV1 to reduce parameters and memory usage [18].

E. Model Training

Model training was conducted on MobileNetV3-Small and EfficientNetV2-Small using pretrained ImageNet weights, with the final classifier layer adjusted to 14 outputs corresponding to the number of ChestMNIST labels. The loss function employed was Binary Cross Entropy with Logits (BCEWithLogitsLoss), the optimizer was Adam with a learning rate of 0.001, the batch size was set to 128, and training was performed for 100 epochs. To accelerate convergence, a MultiStepLR learning rate scheduler was utilized.

All model parameters were fine-tuned end-to-end without freezing the backbone layers. Model predictions were converted to binary labels using a threshold of 0.5 prior to performance evaluation. This parameter configuration follows the baseline from the MEDMNIST v2 study [19], enabling a fair comparison between lightweight CNN architectures and conventional, heavier CNN models. All experiments were conducted on the Kaggle platform using an NVIDIA Tesla P100 GPU (16GB VRAM). The models were implemented using Python 3.10.12 with PyTorch 2.1.0, Torchvision 0.16.0, and the MEDMNIST library for dataset management.

F. Model Evaluation

During the evaluation stage, the trained models were tested using the test data to assess their performance. Performance assessment was conducted using Accuracy and Area Under the Curve (AUC) metrics. Since the ChestMNIST dataset is multi-label, AUC values were calculated using the macro-average, ensuring that the performance of each label was equally considered [22].

Average Accuracy reflects the overall performance in multi-label classification. The equation for average accuracy is presented in Equation 1.

$$Average\ Accuracy = \frac{1}{L}\sum_{j=1}^{L} \frac{TP_j + TN_j}{TP_j + TN_j + FP_j + FN_j} \quad (1)$$

Notation: TP = True Positive, TN = True Negative, FP = False Positive, FN = False Negative.

Macro-AUC is calculated by taking the AUC of each label from the ROC curve and then averaging across all classes. The ROC curve plots the True Positive Rate (TPR) against the False Positive Rate (FPR), as shown in Equations 2 and 3.

$$TPR = \frac{TP}{TP + FN} \quad (2)$$

$$FPR = \frac{TP}{FP + TN} \quad (3)$$

The AUC for each class is calculated as the area under the ROC curve, as shown in Equation 4.

$$AUC = \int_0^1 TPR(FPR)\, d(FPR) \quad (4)$$

Meanwhile, the Macro-AUC across all classes is given in Equation 5.

$$Macro\ AUC = \frac{1}{C}\sum_{i=1}^{C} AUC_i \quad (5)$$

In addition to classification performance evaluation, this study also assessed computational efficiency to determine the feasibility of the models in resource-constrained environments. Computational efficiency was measured using several indicators, namely the number of parameters, FLOPs (Floating Point Operations), training time, inference time, and model size. The number of parameters and FLOPs represent the model's complexity, while training and inference times indicate practical computational requirements. Model size reflects storage requirements and the potential for deploying the model on devices with limited memory capacity.

3. Results and Discussion

Based on the eight testing scenarios conducted in this study, the model performance results are presented in Tables 3, 4, and 5.

Table 3. Accuracy Performance per Label Across Each Scenario

	Label	S1	S2	S3	S4	S5	S6	S7	S8
1	atelectasis	0.89	0.89	0.89	0.89	0.85	0.89	0.88	0.89
2	cardiomegaly	0.97	0.97	0.97	0.97	0.96	0.97	0.97	0.97
3	Effusion	0.87	0.87	0.87	0.87	0.86	0.87	0.87	0.87
4	infiltration	0.82	0.82	0.82	0.82	0.78	0.82	0.82	0.82
5	mass	0.94	0.94	0.94	0.94	0.93	0.94	0.94	0.94
6	nodule	0.94	0.98	0.94	0.94	0.92	0.94	0.93	0.94
7	pneumonia	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.98

	Label	S1	S2	S3	S4	S5	S6	S7	S8
1	atelectasis	0.89	0.89	0.89	0.89	0.85	0.89	0.88	0.89
2	cardiomegaly	0.97	0.97	0.97	0.97	0.96	0.97	0.97	0.97
3	Effusion	0.87	0.87	0.87	0.87	0.86	0.87	0.87	0.87
8	pneumothora	0.95	0.95	0.95	0.95	0.94	0.95	0.94	0.95
9	consolidation	0.95	0.95	0.95	0.95	0.93	0.95	0.95	0.95
10	Edema	0.98	0.98	0.98	0.98	0.97	0.98	0.97	0.98
11	emphysema	0.97	0.97	0.97	0.97	0.97	0.97	0.97	0.97
12	Fibrosis	0.98	0.98	0.98	0.98	0.97	0.98	0.98	0.98
13	Pleura	0.96	0.96	0.96	0.96	0.95	0.96	0.96	0.96
14	hernia	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99

Table 3 indicates that MobileNetV3-Small (Scenarios 1–4) produced more stable and consistent accuracy performance compared to EfficientNetV2-Small (Scenarios 5–8) across almost all disease labels. For labels with a large number of samples, such as atelectasis, effusion, and infiltration, MobileNetV3-Small was able to maintain relatively consistent accuracy across different augmentation scenarios, whereas EfficientNetV2-Small tended to experience accuracy drops in scenarios without augmentation. The impact of data augmentation on per-label accuracy improvement was very limited. For most labels, the accuracy values across scenarios differed only slightly for both MobileNetV3-Small and EfficientNetV2-Small. This indicates that light augmentations, such as rotations up to 15 degrees, 10% translation, and Gaussian blur, did not provide significant improvements in the accuracy of ChestMNIST models.

Moreover, for labels with imbalanced data distribution, such as hernia and pneumonia, all scenarios produced very high accuracy values. However, these accuracy values should be interpreted with caution, as they are likely influenced by the dominance of negative classes in multi-label classification. This condition can cause accuracy to not fully reflect the model's ability to recognize positive cases, which are relatively few. Therefore, the per-label accuracy results in Table 3 should be understood in the context of class imbalance in the ChestMNIST dataset. Accordingly, model performance evaluation is also presented using Area Under the Curve (AUC) values in Table 4 to provide a more representative measure of performance.

Table 4.AUC-ROC Performance per Label Across Each Scenario

	Label	S1	S2	S3	S4	S5	S6	S7	S8
1	atelectasis	**0.73**	**0.73**	0.72	**0.73**	0.63	0.62	0.68	0.66
2	cardiomegaly	0.82	**0.84**	0.82	0.80	0.70	0.63	0.65	0.67
3	Efussion	0.80	**0.81**	0.79	0.80	0.69	0.68	0.73	0.74
4	infiltration	0.65	**0.66**	0.65	**0.66**	0.57	0.61	0.63	0.63
5	mass	**0.68**	0.67	0.67	0.66	0.58	0.54	0.59	0.58
6	nodule	0.61	**0.62**	0.60	0.61	0.54	0.55	0.57	0.53
7	pneumonia	**0.68**	**0.68**	0.67	**0.68**	0.55	0.64	0.65	0.66
8	pneumothorax	0.76	**0.77**	0.73	**0.77**	0.62	0.59	0.66	0.62
9	consolidation	0.76	**0.77**	0.75	0.76	0.56	0.68	0.73	0.74
10	Edema	**0.85**	**0.85**	0.84	**0.85**	0.67	0.75	0.79	0.79
11	emphysema	**0.76**	**0.76**	0.74	**0.76**	0.62	0.54	0.64	0.60
12	Fibrosis	0.70	**0.72**	0.70	0.70	0.57	0.62	0.61	0.64
13	Pleura	**0.69**	**0.69**	0.67	0.68	0.54	0.56	0.58	0.58
14	hernia	**0.84**	0.83	0.77	0.80	0.58	0.61	0.61	0.68

Based on Table 4, MobileNetV3-Small in Scenarios 1–4 consistently demonstrated higher and more stable AUC-ROC values compared to EfficientNetV2-Small in Scenarios 5–8 across almost all disease labels. These findings indicate that MobileNetV3-Small is more effective in distinguishing between positive and negative classes in multi-label classification tasks. The performance difference was most pronounced for the labels infiltration, mass, and nodule. For these labels, MobileNetV3-Small exhibited higher and relatively stable AUC values compared to EfficientNetV2-Small, which tended to produce lower AUC values across all testing scenarios.

The performance differences presented in Table 4 indicate that MobileNetV3-Small achieves higher AUC-ROC values compared to EfficientNetV2-Small on low-resolution X-ray images. Furthermore, the variation in AUC values across augmentation scenarios for each model was relatively small and did not exhibit a consistent improvement pattern, indicating that the application of light spatial augmentations, such as rotation, translation, and Gaussian blur, did not result in significant performance changes on the ChestMNIST dataset. For labels with highly imbalanced class distributions, such as hernia and pneumonia, AUC values were in the mid-range despite high accuracy, suggesting that the AUC-ROC metric provides a more appropriate representation of performance than accuracy in the context of multi-label classification with class imbalance. Overall, the study's findings indicate that the choice of model has a greater impact on classification performance than the application of light data augmentation variations.

In addition to assessing per-label accuracy and AUC performance, this study also conducted a comprehensive evaluation focusing on two main aspects: computational efficiency and overall performance in terms of accuracy and AUC. The computational efficiency metrics evaluated in this study include the number of model parameters, FLOPs, resulting model size, training time, and inference time. The computational efficiency results of the models are presented in Table 5.

Table 5. Overall Model Performance in Terms of Computational Efficiency

No	Scenarios	Params	FLOPs	Model Size	infer-time	Train-time (minutes)
1	Scenario 1	**1.53 M**	**0.06 B**	**6.02 MB**	6.17 ms	**33.72**
2	Scenario 2	**1.53 M**	**0.06 B**	**6.02 MB**	6.15 ms	50.53
3	Scenario 3	**1.53 M**	**0.06 B**	**6.02 MB**	**6.05 ms**	52.86
4	Scenario 4	**1.53 M**	**0.06 B**	**6.02 MB**	5.99 ms	51.82
5	Scenario 5	20.20 M	2.90 B	78.1 MB	21.07 ms	117.38
6	Scenario 6	20.20 M	2.90 B	78.1 MB	21.78 ms	118.29
7	Scenario 7	20.20 M	2.90 B	78.1 MB	21.36 ms	117.83
8	Scenario 8	20.20 M	2.90 B	78.1 MB	20.88 ms	11894

Based on Table 5, in terms of model complexity, MobileNetV3-Small has approximately 1.53 million parameters, whereas EfficientNetV2-Small has about 20.20 million parameters, meaning that EfficientNetV2-Small has roughly 13 times more parameters. The computational requirements also show a more pronounced difference, with EfficientNetV2-Small requiring around 2.90 GFLOPs, while MobileNetV3-Small requires only about 0.06 GFLOPs, approximately 48 times lower. Correspondingly, the model size of EfficientNetV2-Small reaches 78.1 MB, about 13 times larger than MobileNetV3-Small at only 6.02 MB. Regarding inference time, EfficientNetV2-Small takes approximately 20.88–21.78 ms, whereas MobileNetV3-Small ranges from 5.99–6.17 ms, indicating that EfficientNetV2-Small is about 3.4–3.6 times slower. Additionally, EfficientNetV2-Small's training time ranges from 117–118 minutes, while MobileNetV3-Small requires approximately 34–53 minutes, or about 2–3 times less. Meanwhile,

variations in augmentation scenarios for each model did not show significant differences in the number of parameters, FLOPs, or model size, and only produced minor variations in training and inference times.

In addition to evaluating computational efficiency, this study also conducted a comprehensive assessment of prediction performance using Accuracy and AUC-ROC metrics to evaluate the overall model performance across the eight testing scenarios. The overall Accuracy and AUC-ROC evaluation results for all testing scenarios are presented in Table 6.

Table 6. Overall Model Performance in Terms of Accuracy and AUC

No	Scenarios	Akurasi	AUC
1	Scenario 1	0.947	0.742
2	Scenario 2	0.947	0.748
3	Scenario 3	0.947	0.726
4	Scenario 4	0.947	0.737
5	Scenario 5	0.932	0.607
6	Scenario 6	0.947	0.620
7	Scenario 7	0.944	0.655
8	Scenario 8	0.947	0.656

Based on Table 6, MobileNetV3-Small (Scenarios 1–4) demonstrated highly stable overall accuracy performance, with consistent accuracy values of 0.947 across all scenarios. Meanwhile, the AUC-ROC values for MobileNetV3-Small ranged from 0.726 to 0.748, showing relatively small variation across augmentation scenarios. This indicates that the application of data augmentation did not result in significant changes in overall prediction performance, in terms of both accuracy and AUC-ROC, for the MobileNetV3-Small model.

For EfficientNetV2-Small (Scenarios 5–8), overall accuracy ranged from 0.932 to 0.947, generally slightly lower than that of MobileNetV3-Small. The performance difference was more pronounced in AUC-ROC values, with EfficientNetV2-Small producing AUCs in the range of 0.607–0.656, consistently lower than those of MobileNetV3-Small. Furthermore, although applying augmentation to EfficientNetV2-Small resulted in a slight increase in AUC compared to scenarios without augmentation, this improvement was limited and did not match the AUC performance of MobileNetV3-Small. Overall, the experimental results presented in Tables 3 through 6 indicate that MobileNetV3-Small provides a more balanced performance between prediction accuracy and computational efficiency compared to EfficientNetV2-Small for multi-label, low-resolution ChestMNIST X-ray image classification. In terms of prediction performance, MobileNetV3-Small exhibited stable per-label accuracy (Table 3) as well as higher and more consistent AUC-ROC values (Tables 4 and 6), particularly for labels with imbalanced distributions.

From the perspective of computational efficiency (Table 5), MobileNetV3-Small exhibits significantly lower values in terms of the number of parameters, computational requirements (FLOPs), model size, training time, and inference time compared to EfficientNetV2-Small, with differences reaching several orders of magnitude for some metrics. Meanwhile, the application of light data augmentation had only a limited impact on accuracy and AUC performance, primarily contributing to increased training time without affecting inference efficiency or the inherent characteristics of the model architecture. Consequently, the results of this study indicate that the differences in model architecture are a more dominant factor than variations in light augmentation strategies in determining the overall model performance, both in terms of discriminative capability and computational efficiency, for multi-label ChestMNIST X-ray image classification tasks.

4. Conclusions

The findings of this study indicate that for multi-label, low-resolution chest X-ray image classification tasks, the advantages of lightweight CNN architectures are not solely determined by larger model capacity, but rather by the suitability of the architecture design to the data characteristics. MobileNetV3-Small consistently produced higher and more stable AUC-ROC values compared to EfficientNetV2-Small. These results suggest that a more compact architecture with efficient activation functions and blocks, such as MobileNetV3-Small, is capable of extracting more relevant feature representations from low-resolution images, compared to deeper and more complex models that may suffer from feature redundancy or inefficient representations at limited spatial scales.

From the perspective of computational efficiency, the stark performance differences between the two models underscore the trade-off between complexity and effectiveness. EfficientNetV2-Small requires computational resources up to several orders of magnitude higher, yet this is not accompanied by commensurate improvements in predictive performance, in terms of both accuracy and AUC-ROC. Furthermore, the impact of light spatial data augmentation on model performance was shown to be marginal, indicating that simple geometric variations are insufficient to enrich feature representations in the extremely low-resolution and highly imbalanced ChestMNIST dataset. Overall, these findings emphasize that for low-resolution, multi-label medical image classification scenarios, selecting a CNN architecture that is efficient and well-suited to the data characteristics is more crucial than increasing model complexity or applying light augmentation. They also open avenues for further research on semantic-based augmentation or more adaptive strategies for handling class imbalance.

References

[1] World Health Organization, “Global Tuberculosis Report 2023: Factsheet.” Accessed: Dec. 19, 2025. [Online]. Available: https://cdn.who.int/media/docs/default-source/hq-tuberculosis/global-tuberculosis-report-2023/global-tb-report-2023-factsheet.pdf

[2] J. B. Soriano et al., “Global, regional, and national prevalence and risk factors of chronic respiratory diseases from 1990 to 2019,” The Lancet, 2023. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/37229504/

[3] Institute for Health Metrics and Evaluation, “Chronic respiratory disease is third leading cause of death globally,” IHME News Release, 2024. [Online]. Available: https://www.healthdata.org/news-events/newsroom/news-releases/chronic-respiratory-disease-third-leading-cause-death-globally.

[4] E. Callı, E. Sogancioglu, B. van Ginneken, K. G. van Leeuwen, and K. Murphy, "Deep learning for chest X-ray analysis: A survey," Medical Image Analysis, vol. 72, p. 102105, 2021. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/34171622/

[5] A. Ait Nasser and M. A. Akhloufi, "A review of recent advances in deep learning models for chest disease detection using radiography," Diagnostics, vol. 13, no. 1, p. 159, 2023. DOI: https://doi.org/10.3390/diagnostics13010159

[6] H. Iqbal, A. Khan, N. Nepal, F. Khan, and Y.-K. Moon, "Deep Learning Approaches for Chest Radiograph Interpretation: A Systematic Review," Electronics, vol. 13, no. 23, p. 4688, 2024. DOI: https://doi.org/10.3390/electronics13234688

[7] Tedyyana, A., Ghazali, O., & W. Purbo, O. (2024). Enhancing intrusion detection system using rectified linear unit function in pigeon inspired optimization algorithm. *IAES International Journal of Artificial Intelligence (IJ-AI), 13*(2), 1526-1534. doi:http://doi.org/10.11591/ijai.v13.i2.pp1526-1534

[8] R. Egala and M. V. S. Sairam, "A review on medical image analysis using deep learning," Engineering Proceedings, vol. 66, no. 1, p. 7, 2024. DOI: https://doi.org/10.3390/engproc2024066007

[9] A. Shazia, Z. X. Tan, J. H. Chuah, J. Usman, P. Qian, and K. W. Lai, "A comparative study of multiple neural network for detection of COVID-19 on chest X-ray," EURASIP Journal on

Advances in Signal Processing, 2021, Art. no. 50. DOI: https://doi.org/10.1186/s13634-021-00755-1

[10] R. N. Sadoon and A. M. Chaid, "Deep learning technique for the classification of lung diseases from X-ray images," Journal of Al-Qadisiyah for Computer Science and Mathematics, vol. 16, no. 2, pp. 70–83, 2024. DOI: https://doi.org/10.29304/jqcsm.2024.16.21544

[11] A. M. Wahid, T. Hariguna, and G. Karyono, "Optimization of recommender systems for image-based website themes using transfer learning," Journal of Applied Data Sciences, vol. 6, no. 2, pp. 952–968, 2025. DOI: https://doi.org/10.47738/jads.v6i2.671

[12] H. Li, Z. Wang, X. Yue, W. Wang, H. Tomiyama, and L. Meng, "An architecture-level analysis on deep learning models for low-impact computations," Artificial Intelligence Review, vol. 56, no. 3, pp. 1971–2010, 2022. DOI: https://doi.org/10.1007/s10462-022-10221-5

[13] X. Du, Y. Zhang, H. Li, and J. Wang, "Impact of input image resolution on deep learning performance for side-scan sonar classification: An accuracy–efficiency analysis," Remote Sensing, vol. 17, no. 14, p. 2431, 2025. DOI: https://doi.org/10.3390/rs17142431

[14] M. A. Abu Talib, S. Setumin, S. J. Abu Bakar, A. I. Che Ani, and D. E. Cahyani, "Systematic review of a lightweight convolutional neural network architectures on edge devices," International Journal of Reconfigurable and Embedded Systems (IJRES), vol. 14, no. 2, pp. 339–352, 2024. DOI: https://doi.org/10.11591/ijres.v14.i2.pp339-352

[15] H. Zhao, X. Zhang, Y. Gao, L. Wang, L. Xiao, S. Liu, B. Huang, and Z. Li, "Diagnostic performance of EfficientNetV2-S method for staging liver fibrosis based on multiparametric MRI," Heliyon, vol. 10, no. 1, Art. no. e11334657, 2024. [Online] Available: https://pubmed.ncbi.nlm.nih.gov/39165928/

[16] N. Sriwiboon and S. Phimphisan, "Efficient COVID-19 detection using optimized MobileNetV3-Small with SRGAN for web application," Engineering, Technology & Applied Science Research, vol. 15, no. 2, pp. 9977–9985, 2025. [Online] Available:https://etasr.com/index.php/ETASR/article/view/9977

[17] S. P. K. Prasad and M. El-Sharkawy, "Deployment of Compressed MobileNet V3 on i.MX RT1060," in 2021 IEEE International IoT, Electronics and Mechatronics Conference (IEMTRONICS), Oct. 2021, pp. 1–4. DOI: https://doi.org/10.1109/IEMTRONICS52119.2021.9422512

[18] M. Tan and Q. V. Le, "EfficientNetV2: Smaller models and faster training," in *Proc. 38th International Conference on Machine Learning (ICML)*, P. F. Gehler et al., Eds., 2021, pp. 10096–10106. [Online] Available: https://proceedings.mlr.press/v139/tan21a/tan21a.pdf

[19] J. Yang, R. Shi, D. Wei, Z. Liu, L. Zhao, B. Ke, H. Pfister, and B. Ni, "MedMNIST v2: A large-scale lightweight benchmark for 2D and 3D biomedical image classification," Scientific Data, vol. 10, Art. no. 67, 2023. [Online] Available: https://www.nature.com/articles/s41597-022-01721-8

[20] O. Rainio and R. Klén, "Comparison of simple augmentation transformations for a convolutional neural network classifying medical images," Signal, Image and Video Processing, vol. 18, pp. 3353–3360, 2024. DOI: https://doi.org/10.1007/s11760-024-02998-5

[21] Howard, A., Sandler, M., Chu, G., et al. Searching for MobileNetV3. In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), pp. 1314–1324, 2019. DOI: https://doi.org/10.1109/ICCV.2019.00140

[22] G. Wu, C. Li, and Y. Yin, "Towards understanding generalization of Macro-AUC in multi-label learning," in Proc. 40th International Conference on Machine Learning (ICML 2023), vol. 202 PMLR, 2023. DOI: https://dl.acm.org/doi/10.5555/3618408.3619971