# AugCXR Dataset: An Augmented Chest X-Ray Image Dataset for Robust Deep Learning Pneumonia Diagnosis

**Waqar Ahmad**
Department of Computer Science, University of Hertfordshire, United Kingdom
wa23aat@herts.ac.uk

**Deepak Panday**
Department of Computer Science, University of Hertfordshire, United Kingdom
d.panday@herts.ac.uk

**Muhammad Ibrahim**
Institute of Biotechnology and Genetic Engineering, The University of Agriculture, Pakistan
ibrahimfaqir@aup.edu.pk

**Tahir Mehmood**
School of Information Technology, UNITAR International University, Malaysia
tahir.mehmood@unitar.my (corresponding author)

**Muhammad Yaqoob**
Department of Computer Science, University of Hertfordshire, United Kingdom
m.yaqoob3@herts.ac.uk

*Received: 14 November 2025 | Revised: 6 January 2026, 16 February 2026, and 26 February 2026 | Accepted: 28 February 2026*

 *DOI: https://doi.org/10.48084/etasr.16271*

## ABSTRACT

**Chest X-ray (CXR) imaging is commonly used to detect pneumonia, but reliance on expert radiologists may delay diagnosis and increase costs. The shortage of radiologists and the risk of diagnostic errors highlight the need for automated solutions. Deep learning models using Convolutional Neural Networks (CNNs) have shown potential in computer-aided diagnosis of pneumonia from CXR images. Most studies use the publicly available dataset from Guangzhou Women and Children's Hospital, which includes CXR images of children aged 1 to 5. However, the dataset is imbalanced, with more pneumonia cases than normal. This imbalance may affect model performance and generalizability. This study proposes a geometric data augmentation method using five transformations: rotation, width-shift, height-shift, zoom, and brightness to balance the dataset and improve model accuracy. The proposed Augmented Chest X-ray (AugCXR) dataset was validated using three widely adopted architectures: Improved Visual Geometry Group-13 (IVGG13), MobileNetV2, and EfficientNetV2L. The results demonstrate that the proposed augmentation method enhances classification performance across all three pretrained deep learning models.**

***Keywords-pneumonia; chest X-ray; data augmentation; deep learning***

## I. INTRODUCTION

X-ray scans are used to diagnose many diseases, including pneumonia and a wide range of other medical conditions, such as bone fractures. A key advantage of using X-ray images is that they are cost-efficient and provide quick results in emergencies. X-ray imaging plays a crucial role in diagnosing pneumonia in children since it is the leading cause of death in children under five. That is, around 14% of the total deaths in children happen due to pneumonia [1]. The latter is a life-threatening respiratory infection that requires early and accurate diagnosis for timely treatment. Advancements in medical imaging technologies using X-ray images may speed up the diagnostic process and allow early detection of pneumonia. Consequently, lowering infection risk and saving countless young lives. Traditionally, pneumonia is diagnosed

by expert radiologists using Chest X-Ray (CXR) images. This can be a time-consuming and expensive process depending on the availability and expertise of radiologists. Moreover, pneumonia diagnosis carried out by technicians rather than a qualified radiologist in the developing world may be prone to errors [2]. Even in developed countries, pneumonia diagnosis based on clinical coding can be unreliable, with up to 30% of coded cases lacking radiographic evidence [3]. As a result, automated image analysis using artificial intelligence has gained significant attention. Deep learning techniques for pneumonia diagnosis using CXR images have been explored [4]. Although deep learning models have shown promising results, they require a huge amount of data to be trained [5]. However, obtaining large and well-balanced datasets for training is a great challenge. Data augmentation techniques have been explored to balance the dataset and thus reduce bias and overfitting in deep learning models [12-14].

The performance of machine learning models generally depends on the ability of the model to learn from complex and relevant features from the input data [6], and the quality and variety of the data used for training. A model with complex feature extraction capabilities can detect important patterns. High-quality data help the model learn from a variety of patterns and features, enabling the model to generalize better to new or unseen test cases. Studies on deep models focus on using transfer learning with pre-trained Convolutional Neural Network (CNN) models, fine-tuned for classification tasks. Visual Geometry Group (VGG) models are widely employed in medical image classification tasks due to their relatively simple yet robust architecture, which enables effective feature extraction with high accuracy. In particular, VGG16 [7] and VGG19 are among the top-performing models for image classification tasks, including CXR images. These pre-trained VGG models are resource-efficient, and they achieve fast convergence during training through knowledge transfer. Authors in [8] optimized pre-trained CNNs (AlexNet, ResNet18, DenseNet201, and SqueezeNet) for pneumonia detection and demonstrated high recognition rates with the optimized architectures. Authors in [9] confirmed the robustness of the VGG16 model in classifying pathological brain images. Moreover, authors in [10] compared several CNN architectures, including VGG16, ResNet50, and EfficientNetV2L, for pneumonia detection using CXR images. They concluded that EfficientNetV2L achieved the highest accuracy, whereas VGG16 also performed competitively. Authors in [11] proposed Improved VGG-13 (IVGG13) as an improved version of VGG16 for pneumonia CXR classification. Their model used geometric data augmentation and demonstrated better performance for pneumonia classification using CXRs.

Augmentation techniques, including geometric, histogram-based enhancement, Generative Adversarial Networks (GANs), and Synthetic Minority Over-sampling Technique (SMOTE), can improve the classification accuracy. Authors in [12] studied the impact of geometric data augmentation techniques, such as rotation, translation, scaling, shearing, and flipping, on the detection of COVID-19 using 17 different deep learning models trained on the CXR datasets. It was found that removing the augmentation improved performance, as the models without augmentation achieved higher accuracy and Matthews Correlation Coefficient (MCC) scores, emphasizing the selection of appropriate augmentation parameters for the optimal performance of the model. Authors in [13] investigated traditional data augmentation methods consisting of horizontal flip, vertical flip, rotation, and brightness through object detection on CXR security inspection images using Faster R-CNN and ResNet50. It was established that data augmentation methods reflecting real-world changes, specifically geometric ones, can improve model generalizability. However, inappropriate or excessive augmentations, such as extreme rotation or incorrect brightness adjustment, can reduce model accuracy, which shows the importance of selecting the right augmentation parameters for a given dataset. Authors in [14] presented CovidXrayNet using EfficientNet-B0 with specific data augmentation methods, such as resizing along with flipping, rotation, zooming, lighting, warping, normalization, and tuning of CNN hyperparameters to identify COVID-19 in CXRs. Their model reached an accuracy of 95.82% for both balanced and unbalanced datasets.

GAN-based data augmentation techniques have been evaluated. Authors in [15] implemented DCGAN to augment their imbalanced CXR dataset, which led to a 95.5% CNN accuracy score combined with a Fréchet Inception Distance (FID) score of 1.289, indicating a high quality of synthetic images produced by GANs. Authors in [16] demonstrated that GAN-based augmented images generated for cardiovascular abnormality datasets reached an accuracy level of 84.19%, which surpassed traditional methods as well as non-augmented techniques. Authors in [17] developed Inception-Augmentation GAN (IAGAN) that outperformed traditional augmentation and DCGAN by producing more authentic and diverse images and resulting in improved detection of pneumonia and COVID-19 by using Area Under the Curve (AUC), sensitivity, and specificity measurements.

The impact of the SMOTE on model performance has also been studied. Authors in [18] investigated pneumonia detection using SMOTE with VGG16 and several CNN architectures, including VGG16, VGG19, Xception, Inception-ResNet v2, and DenseNet201. They reported an accuracy of 93.75% with SMOTE enabling a 4% improvement in overall performance. To address class imbalance in early COVID-19 diagnosis, authors in [19] conducted research on tuberculosis classification by applying SMOTE oversampling techniques to InceptionV3, which delivered 99.94% accuracy, 100% recall, and 99.94% F1-score. Authors in [20] applied SMOTE and Weighted Categorical Loss (WCL) on three datasets, achieving 98.87% accuracy and 99.15% AUC using CheXNet with WCL.

The role of image enhancement techniques in improving CXR image quality and model performance has been extensively studied. Authors in [21] combined log-normalization and Contrast Limited Adaptive Histogram Equalization (CLAHE) in a two-stage Normalized CLAHE (N-CLAHE) and reported improved diagnostic image quality compared to standard techniques. Authors in [22] applied combined techniques of CLAHE-HE to CNN-based medical image segmentation, which delivered outstanding performance metrics with DSC at 0.92 and SSIM at 0.97. The combination

of CLAHE and DWT produced improved diagnostic results when authors in [23] applied these methods to CXR images according to MSE, PSNR, and AMBE metric measurements.

Research has explored pneumonia detection from CXR images using feature-engineering and ensemble-based machine learning approaches. Authors in [24] proposed PneuX-Net, an ensemble-based feature extraction and transformation framework for pneumonia detection from CXR images, and reported classification accuracy. In contrast, the present work focuses on dataset construction and augmentation-driven generalization for improved reproducibility, leakage prevention, and external validation rather than feature-level optimization. The present study proposes an Augmented Chest X-ray (AugCXR) dataset, a structured dataset construction and augmentation framework in which preprocessing and data splitting are performed before augmentation, and synthetic samples are generated exclusively for the training set. This design ensures balanced learning while maintaining unbiased validation and testing, as summarized in Figure 1.

In this work, a geometric augmentation method was proposed to balance the CXR image dataset. In addition, the study introduces an optimal method to augment the CXR images using five different transformations (rotation, width-shift, height-shift, zoom, and brightness) aiming to address the class imbalance problem and increase the data size. The results indicate that, regardless of the pre-trained architecture, the proposed augmentation outperforms all the current state-of-the-art models across all performance measures.

## II. THE PROPOSED DATA AUGMENTATION METHOD

The dataset used in this research consists of CXR images of child patients (ages 1 to 5) from the Guangzhou Women and Children's Hospital [25]. The dataset suffers from class imbalance, limited intra-class variability, and inconsistent data partitioning, which can lead to biased performance estimates and reduced generalizability. It contains 5856 images with 4273 pneumonia and 1583 normal patients, and is further divided into training, test, and validation sets. To address the class imbalance and improve the robustness of the model, the present study proposes a geometric data augmentation technique. The proposed augmentation synthetically expands the dataset by generating new images through controlled transformations of the existing ones. The proposed data augmentation and dataset restructuring for pneumonia detection using CXR images are illustrated in Figure 1. The training set and validation set images are merged into the train set to maximize the data availability for augmentation. To ensure a balanced and sufficiently large test set, 500 images are randomly sampled from the training set for each class (normal and pneumonia). The test set contains only original, un-augmented images for an unbiased evaluation of the trained model. Augmentation is then applied exclusively to the remaining training data. After augmentation, a portion of the augmented training set is reassigned as the new validation set to support model tuning and performance monitoring.

A grid search is conducted to identify the optimal augmentation parameters. The transformations and their respective search ranges are: rotation angles from 1° to 45° with a step size of 5°, width, height, and shear shifts ranging from 0 to 0.2 (in steps of 0.01), and zoom factors from 0 to 0.5 (step size 0.05). Horizontal flipping is included as a binary option (enabled or disabled). For filling pixels in newly created image regions, three modes are explored: 'nearest', 'wrap', and 'constant'. Image brightness for newly created images is adjusted within a constant range of 0.5-1.5. For each combination of augmentation parameters, the model is trained, and its classification accuracy is evaluated on a validation set. The training set comprised a total of 7,200 images evenly distributed between two classes: 3,600 Pneumonia and 3,600 Normal. Similarly, the validation set contained 1,800 images, equally divided between the two classes. The exhaustive grid search was used to identify the optimal augmentation parameter: a rotation angle of 10°, width and height shifts of 0.09, a shear factor of 0 (shear is not applied), a zoom factor of 0.15, horizontal flipping enabled, a brightness range of 0.8-1.0, and the 'nearest' pixel filling for newly introduced pixels. These parameters are utilized to generate new images from the original dataset. This data augmentation is employed to address the limited diversity and inherent class imbalance commonly observed in CXR datasets. By introducing controlled geometric variations while preserving clinical realism, the proposed augmentation strategy aims to enhance model generalization and reduce overfitting.

## III. DEEP LEARNING MODELS

To evaluate the performance of the proposed balanced AugCXR dataset, three state-of-the-art CNN models (from lightweight to high-performing), commonly used for pneumonia classification, are selected: Improved VGG13, MobileNetV2, and EfficientNetV2L. IVGG13 [11] is a modified version of the original VGG13 architecture, designed to improve classification performance on small datasets through structural adjustments and data augmentation techniques. It is relatively shallow, with 13 hidden layers and roughly 2.5 million parameters. Its ability to extract complex features makes it suitable for medical imaging tasks where data availability is limited. In this study, the IVGG13 model is implemented, and then it is trained from scratch on the datasets. MobileNetV2 [26] is a lightweight CNN architecture optimized for deployment on mobile and embedded devices. The architecture incorporates two key innovations, inverted residual blocks and linear bottlenecks, which reduce model size while maintaining classification performance. In addition, MobileNetV2 employs depth-wise separable convolutions, which significantly reduce the number of parameters and computational cost compared to standard convolutions. The standard MobileNetV2 architecture in its Keras implementation comprises 155 layers, including convolutional, normalization, and activation layers, and approximately 3.4 million trainable parameters. These characteristics make MobileNetV2 feasible for real-time clinical applications. In this study, MobileNetV2 is deployed with pre-trained ImageNet weights. It applies transfer learning by freezing the base layers and training only the top classification layers on the datasets.

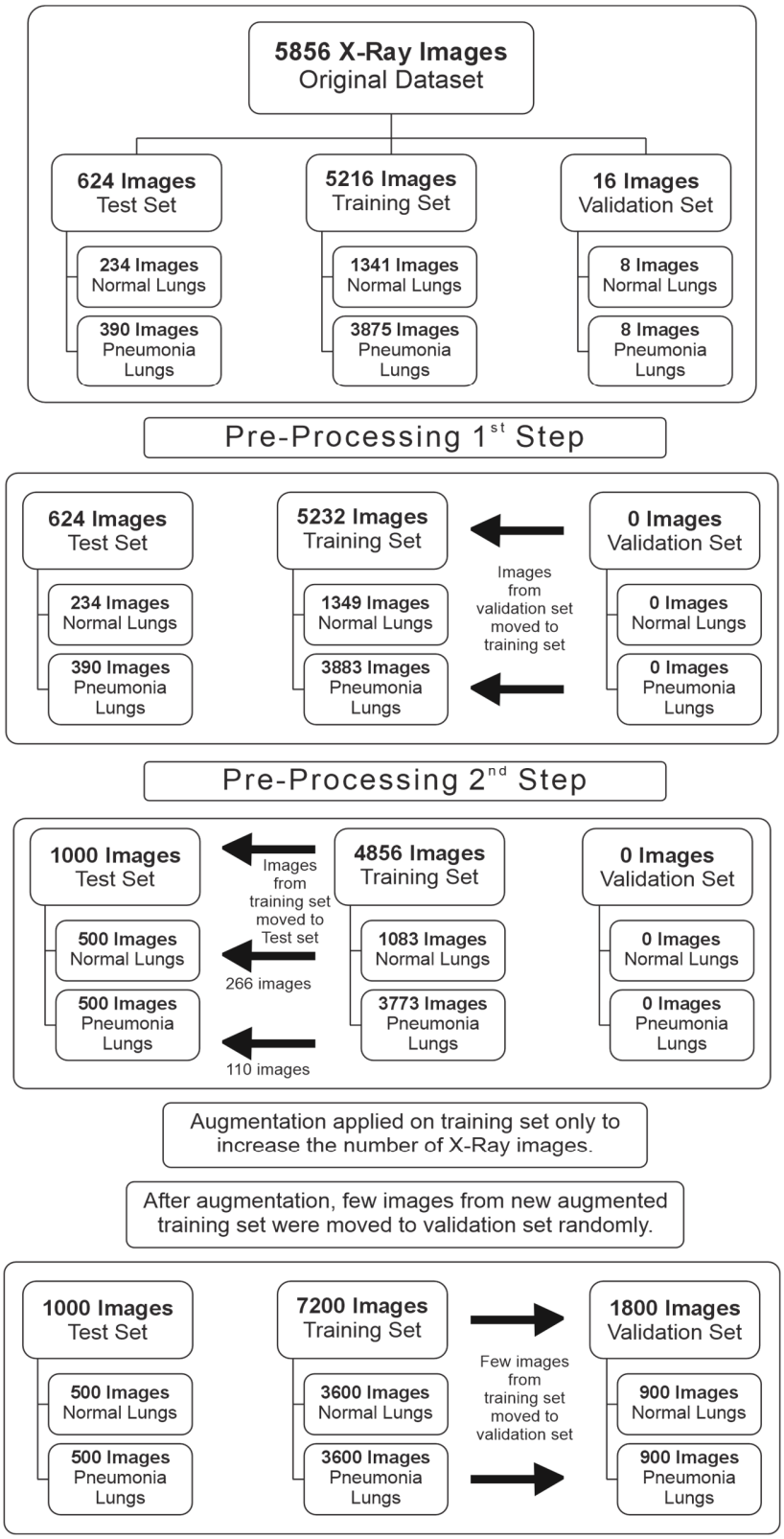


Fig. 1. Proposed data augmentation and restructuring for the CXR image dataset.

EfficientNetV2L [27] is a high-performance CNN that extends the EfficientNet family by improving training speed, accuracy, and model scaling techniques. It employs compound scaling, which uniformly scales network depth, width, and resolution, along with fused-MBConv blocks to improve computational efficiency. EfficientNetV2L consists of 479 layers and approximately 119 million parameters, making it one of the largest variants in the EfficientNetV2 series. Despite its size, it achieves state-of-the-art accuracy on image classification tasks while maintaining relatively efficient training. These properties make EfficientNetV2L highly suitable for pneumonia detection in CXR images, where high precision is significant. In this work, the pre-trained version is used, and only the top layers are fine-tuned to adapt it for pneumonia detection in CXR images.

IVGG13, MobileNetV2, and EfficientNetV2L are trained under three experimental settings: (i) without augmentation of the original dataset, (ii) using SMOTE-based augmentation on the original dataset, and (iii) using the proposed AugCXR dataset. The performance of each model is then evaluated using four standard metrics: precision, recall, accuracy, and F1-score.

## IV. AUGCXR DATASET VALIDATION

To validate the AugCXR dataset, three models, IVGG13, MobileNetV2, and EfficientNetV2L, were initially trained on the original dataset without any augmentation, and were tested on the original test set. Their performance is reported in Table I. After that, the SMOTE-based augmentation was employed to address the class imbalance in the training dataset. The selected three models were then trained, and their performance was evaluated. Finally, the three selected models were trained on the proposed AugCXR dataset for pneumonia detection. The performance of the trained models was subsequently evaluated on the test set included within the AugCXR dataset.

TABLE I. PERFORMANCE COMPARISON OF MODELS ON THE ORIGINAL DATASET, SMOTE, AND THE PROPOSED AUGMENTATION METHOD

| Description | Without augmentation | SMOTE-based augmentation | Proposed augmentation |
|---|---|---|---|
| MobileNetV2 model | | | |
| Accuracy | 90.38% | 79.00% | 93.20% |
| Precision | 91.00% | 75.29% | 91.00% |
| Recall | 94.00% | 99.23% | 96.00% |
| F1-score | 92.00% | 85.61% | 93.00% |
| EfficientNetV2L | | | |
| Accuracy | 78.85% | 78.00% | 84.60% |
| Precision | 80.00% | 76.98% | 82.00% |
| Recall | 88.00% | 91.79% | 88.00% |
| F1-score | 84.00% | 83.74% | 85.00% |
| Model IVGG13 model | | | |
| Accuracy | 77.00% | 78.00% | 95.00% |
| Precision | 73.16% | 74.61% | 93.30% |
| Recall | 99.23% | 98.71% | 97.40% |
| F1-score | 84.22% | 84.98% | 95.30% |

As shown in Table I, the proposed AugCXR dataset provides more robust performance for pneumonia diagnosis compared to no augmentation and SMOTE-based augmentation. Augmentation introduces image variations that reflect real-world conditions, i.e., changes in brightness, orientation, and small distortions that occur in CXR capture. These variations allow the models to learn features that are more general and less sensitive to noise. In contrast, SMOTE only works on tabular feature space by creating synthetic samples from existing data points, which is less suitable for image data because it does not create variation that reflects the real world. The MobileNetV2, which is designed for lightweight deployment, showed improved performance (93.20% accuracy, 96.00% recall) on the AugCXR dataset. The high recall is crucial in medical screening because it means fewer missed pneumonia cases. Moreover, the false positives are also controlled with a precision of 91.00%. The performance of EfficientNetV2L improved modestly with AugCXR but still showed a gain in accuracy and F1-score, suggesting that even high-capacity, well-regularized models could learn more patterns when trained on the proposed AugCXR. The IVGG13 benefits the most from AugCXR, achieving an accuracy of 95% and an F1-score of 95.30% on the test set. This indicates that the data augmentation enabled this model to learn more robust textures and patterns in CXRs, which are essential for detecting pneumonia-related opacities.

The proposed model was compared with the IVGG13 model reported in [11], which achieved an accuracy of 89.10% for pneumonia classification, as displayed in Table II. It can be seen that when the proposed augmentation method is applied to the original dataset, it improves IVGG13 performance, resulting in an accuracy of 93.20%. This confirms an overall improvement in classification robustness.

TABLE II. PERFORMANCE COMPARISON OF THE PROPOSED MODEL

| Methods | Accuracy | Precision | Recall | F1-score |
|---|---|---|---|---|
| IVGG13 [11] | 89.10% | 83.30% | 97.80% | 90.00% |
| IVGG13 on Proposed AugCXR | 93.20% | 91.00% | 96.00% | 93.00% |

Receiver Operating Characteristic (ROC) curves, as depicted in Figure 2, and the AUC scores were used to check the performance of the three models on the proposed AugCXR dataset. The IVGG13 model achieved the highest AUC score of 0.95, indicating that it can accurately differentiate between pneumonia and normal CXR images. MobileNetV2 achieved a slightly lower AUC score of 0.93, indicating accurate classification ability with fewer false positives at different thresholds. In contrast, EfficientNetV2L had a lower AUC score of 0.85 with lower sensitivity and specificity compared to the other two models. The ROC curves for IVGG13 and MobileNetV2 rise steeply toward the top-left corner, demonstrating that these models accurately detect the pneumonia cases and make fewer false positive errors. These results support that the proposed AugCXR dataset is suitable for reliable pneumonia diagnosis regardless of the underlying deep learning architecture.

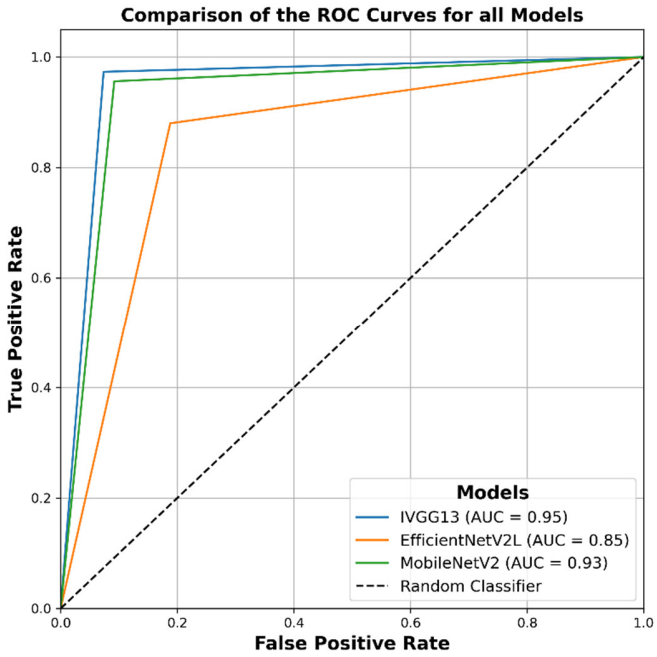


Fig. 2. ROC curves for IVGG13, MobileNetV2, and EfficientNetV2L.

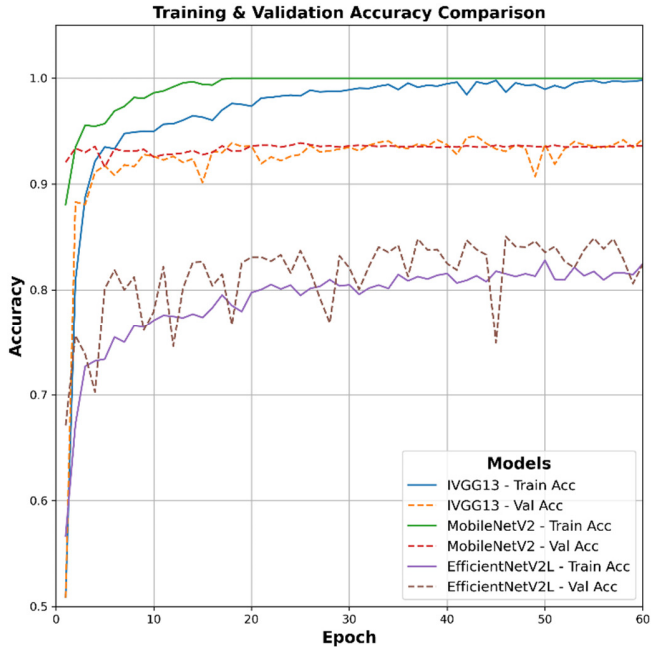


Fig. 3. Training and validation accuracies of the models across epochs.

Figures 4-6 present the confusion matrices for all three models. These confusion matrices illustrate the number of samples that were correctly classified and misclassified for each class. To further evaluate model performance, a comparative analysis was conducted based on the number of trainable parameters and the training time required for each model. Table II summarizes the parameter information for training IVGG13, MobileNetV2, and EfficientNetV2L on the proposed AugCXR dataset. Both the number of trainable parameters and the total training time for each model are included. The IVGG13 model, trained from scratch, required approximately 41 min to train 2.5 million parameters. MobileNetV2, using transfer learning, took around 25 min to train 1.3 million parameters. Due to the larger size of EfficientNetV2L, it required approximately 4 h and 41 min to train 8.8 million parameters. This comparison highlights the trade-off between model complexity, runtime, and classification performance.

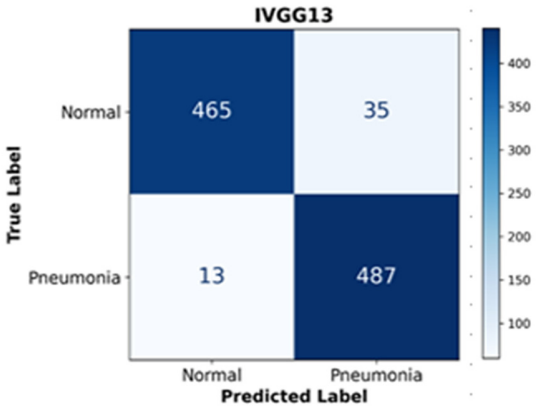


Fig. 4. Confusion matrix for VGG13 on the test set.

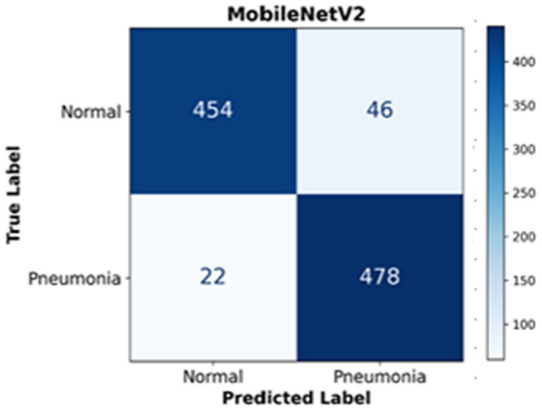


Fig. 5. Confusion matrix for MobileNetV2 on the test set.

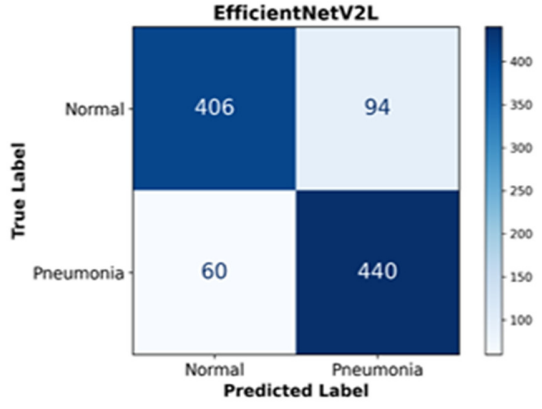


Fig. 6. Confusion matrix for EfficientNetV2L on the test set.

TABLE III. COMPARISON OF TRAINABLE PARAMETERS AND TRAINING TIME

| | **IVGG13** | **MobileNetV2** | **EfficientNetV2L** |
|---|---|---|---|
| Total parameters | 2.5 millions | 3.5 millions | 119.58 millions |
| Trainable parameters | 2.5 millions | 1.3 millions | 8.8 millions |
| Training time | 41 min approx. | 25 min approx. | 281 min approx. |

## V. CONCLUSION

The study presented an Augmented Chest X-ray (AugCXR) dataset, demonstrating that the proposed geometric data augmentation method improves both the discriminative ability and the generalization of pre-trained models for pneumonia detection. Improved Visual Geometry Group-13 (IVGG13) and MobileNetV2 achieved high accuracy, precision, recall, and F1-scores, with high Area Under the Curve (AUC) values, indicating that the model can accurately separate pneumonia from normal cases. Even EfficientNetV2L showed improvement when trained with AugCXR. The augmentation method created realistic variations in the images to reduce the risk of overfitting and enabled the models to learn patterns that performed well on unseen data. Thus, the proposed augmentation strategy is suitable for building AI-based diagnosis systems, with reliable performance in real-time clinical settings.

A future study will employ fine-tune Stable Diffusion (SD) to generate high-quality synthetic CXR images to further improve data diversity and model performance. In addition, this work will be extended to other imbalanced medical imaging datasets to test the adaptability of the proposed augmentation approach across different medical imaging datasets.

## DECLARATION OF COMPETING INTERESTS

The authors declare no competing interests.

## ACKNOWLEDGMENT

The authors appreciate UNITAR International University (Malaysia) for their financial support for this research.

## DATA AVAILABILITY

The dataset [25] used in this study is publicly available at: https://data.mendeley.com/datasets/rscbjbr9sj/2. Code is available at https://github.com/waqarahmad2311/AugCXR-Dataset.

## REFERENCES

[1] "Pneumonia in Children," World Health Organization, Nov. 2022. https://www.who.int/news-room/fact-sheets/detail/pneumonia.

[2] S. Andronikou *et al.*, "Guidelines for the Use of Chest Radiographs in Community-Acquired Pneumonia in Children and Adolescents," *Pediatric Radiology*, vol. 47, no. 11, pp. 1405–1411, Oct. 2017, https://doi.org/10.1007/s00247-017-3944-4.

[3] K. Pink, I. Mitchell, and H. Davies, "P17 The Accuracy of a Diagnosis of Pneumonia in a UK Teaching Hospital," *Thorax*, vol. 67, no. Suppl 2, Dec. 2012, Art. no. A71, https://doi.org/10.1136/thoraxjnl-2012-202678.158.

[4] H. N. T. Al-Azzawi *et al.*, "Utilization of a Deep Convolutional Neural Network for the Binary Classification of Chest X-Ray Pneumonia," *Engineering, Technology & Applied Science Research*, vol. 15, no. 1, pp. 20471–20483, Feb. 2025, https://doi.org/10.48084/etasr.9788.

[5] T. Mehmood, A. E. Gerevini, A. Lavelli, M. Olivato, and I. Serina, "Distilling Knowledge with a Teacher's Multitask Model for Biomedical Named Entity Recognition," *Information*, vol. 14, no. 5, Apr. 2023, Art. no. 255, https://doi.org/10.3390/info14050255.

[6] A. M. Nababan *et al.*, "Extreme Learning Machine Approach on Heart Abnormalities Identification in ECG Images," *International Journal of Electronics and Telecommunications*, pp. 473–480, Jun. 2024, https://doi.org/10.24425/ijet.2024.149568.

[7] N. C. Kundur, B. C. Anil, P. M. Dhulavvagol, R. Ganiger, and B. Ramadoss, "Pneumonia Detection in Chest X-Rays Using Transfer Learning and TPUs," *Engineering, Technology & Applied Science Research*, vol. 13, no. 5, pp. 11878–11883, Oct. 2023, https://doi.org/10.48084/etasr.6335.

[8] T. Rahman *et al.*, "Transfer Learning with Deep Convolutional Neural Network (CNN) for Pneumonia Detection Using Chest X-ray," *Applied Sciences*, vol. 10, no. 9, May 2020, Art. no. 3233, https://doi.org/10.3390/app10093233.

[9] T. Kaur and T. K. Gandhi, "Automated Brain Image Classification Based on VGG-16 and Transfer Learning," in *International Conference on Information Technology*, Bhubaneswar, India, Dec. 2019, pp. 94–98, https://doi.org/10.1109/ICIT48102.2019.00023.

[10] M. Ali *et al.*, "Pneumonia Detection Using Chest Radiographs with Novel EfficientNetV2L Model," *IEEE Access*, vol. 12, pp. 34691–34707, 2024, https://doi.org/10.1109/ACCESS.2024.3372588.

[11] Z.-P. Jiang, Y.-Y. Liu, Z.-E. Shao, and K.-W. Huang, "An Improved VGG16 Model for Pneumonia Image Classification," *Applied Sciences*, vol. 11, no. 23, Nov. 2021, Art. no. 11185, https://doi.org/10.3390/app112311185.

[12] M. Elgendi *et al.*, "The Effectiveness of Image Augmentation in Deep Learning Networks for Detecting COVID-19: A Geometric Transformation Perspective," *Frontiers in Medicine*, vol. 8, Mar. 2021, Art. no. 629134, https://doi.org/10.3389/fmed.2021.629134.

[13] B. Jing and Y. Du, "DTDG-Net: A Few-shot Data Augmentation Method for X-ray Security Inspection Images," in *International Conference on Image Processing, Computer Vision and Machine Learning*, Shenzhen, China, Nov. 2024, pp. 493–502, https://doi.org/10.1109/ICICML63543.2024.10957903.

[14] M. M. A. Monshi, J. Poon, V. Chung, and F. M. Monshi, "CovidXrayNet: Optimizing Data Augmentation and CNN Hyperparameters for Improved COVID-19 Detection from CXR," *Computers in Biology and Medicine*, vol. 133, Jun. 2021, Art. no. 104375, https://doi.org/10.1016/j.compbiomed.2021.104375.

[15] S. Kora Venu and S. Ravula, "Evaluation of Deep Convolutional Generative Adversarial Networks for Data Augmentation of Chest X-ray Images," *Future Internet*, vol. 13, no. 1, Dec. 2020, Art. no. 8, https://doi.org/10.3390/fi13010008.

[16] M. Moradi, A. Madani, A. Karargyris, and T. F. Syeda-Mahmood, "Chest X-Ray Generation and Data Augmentation for Cardiovascular Abnormality Classification," in *Medical Imaging 2018: Image Processing*, Houston, TX, United States, Mar. 2018, Art. no. 57, https://doi.org/10.1117/12.2293971.

[17] S. Motamed, P. Rogalla, and F. Khalvati, "Data Augmentation Using Generative Adversarial Networks (GANs) for GAN-Based Detection of Pneumonia and COVID-19 in Chest X-Ray Images," *Informatics in Medicine Unlocked*, vol. 27, 2021, Art. no. 100779, https://doi.org/10.1016/j.imu.2021.100779.

[18] Y. Pamungkas, M. R. N. Ramadani, and E. N. Njoto, "Effectiveness of CNN Architectures and SMOTE to Overcome Imbalanced X-Ray Data in Childhood Pneumonia Detection," *Journal of Robotics and Control*, vol. 5, no. 3, pp. 775–785, Apr. 2024.

[19] A. Alqahtani, Q. Abu Al-Haija, A. A. Alsulami, B. Alturki, N. Alqahtani, and R. Alsini, "Optimizing Chest Tuberculosis Image Classification with Oversampling and Transfer Learning," *IET Image Processing*, vol. 18, no. 5, pp. 1109–1118, Apr. 2024, https://doi.org/10.1049/ipr2.13010.

[20] E. Chamseddine, N. Mansouri, M. Soui, and M. Abed, "Handling Class Imbalance in COVID-19 Chest X-Ray Images Classification: Using SMOTE and Weighted Loss," *Applied Soft Computing*, vol. 129, Nov. 2022, Art. no. 109588, https://doi.org/10.1016/j.asoc.2022.109588.

[21] K. Koonsanit, S. Thongvigitmanee, N. Pongnapang, and P. Thajchayapong, "Image Enhancement on Digital X-Ray Images Using N-CLAHE," in *2017 10th Biomedical Engineering International Conference (BMEiCON)*, Hokkaido, Japan, Aug. 2017, pp. 1–4, https://doi.org/10.1109/BMEiCON.2017.8229130.

[22] S. Saifullah and R. Drežewski, "Modified Histogram Equalization for Improved CNN Medical Image Segmentation," *Procedia Computer Science*, vol. 225, pp. 3021–3030, 2023, https://doi.org/10.1016/j.procs.2023.10.295.

[23] M. Sharma and D. Kumar, "Comparative Analysis of Image Enhancement Techniques for Chest X-ray Images," in *International Conference on Computational Intelligence and Sustainable Engineering Solutions*, Greater Noida, India, May 2022, pp. 130–135, https://doi.org/10.1109/CISES54857.2022.9844364.

[24] K. Munir, M. Usama Tanveer, H. J. Alyamani, A. Bermak, and A. Ur Rehman, "PneuX-Net: An Enhanced Feature Extraction and Transformation Approach for Pneumonia Detection in X-Ray Images," *IEEE Access*, vol. 13, pp. 84024–84037, 2025, https://doi.org/10.1109/ACCESS.2025.3568885.

[25] D. Kermany, "Labeled Optical Coherence Tomography (OCT) and Chest X-Ray Images for Classification." Mendeley, Jan. 06, 2018, [Online]. Available: https://data.mendeley.com/datasets/rscbjbr9sj/2.

[26] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "MobileNetV2: Inverted Residuals and Linear Bottlenecks," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, UT, USA, Jun. 2018, pp. 4510–4520, https://doi.org/10.1109/CVPR.2018.00474.

[27] M. Tan and Q. Le, "EfficientNetV2: Smaller Models and Faster Training," in *Proceedings of the 38th International Conference on Machine Learning*, Virtual, Jul. 2021, Art. no. 139.