

Automated Chest X-Ray Pneumonia Classification

Mustafa Melik Ayanoğlu, Emir Çil, Emir Ersel Bilgiç, Batu Arıbakır
Department of Electrical and Electronics Engineering,
Izmir Katip Celebi University, Izmir, Turkey

Abstract

Chest radiographs are among the most widely used imaging techniques in the evaluation of pneumonia conditions. In recent years, the interpretation of these images has increasingly involved the use of computer-aided diagnosis (CAD) systems, which aim to improve diagnostic accuracy and minimize inter-observer variability. This study proposes a deep learning-based approach for the automatic pneumonia classification of chest X-ray. Therefore, various CNN architectures have been trained to detect anatomical differences and categorize the images into frontal views. Among the fine-tuned models using transfer learning, ShuffleNet_V2_x1_0 and EfficientNet-B0 yielded the best results on the Chest X-Ray Images (Pneumonia) dataset. The training process was conducted using the Adam optimization algorithm and the cross-entropy loss function. Evaluation metrics such as accuracy, sensitivity, specificity, and F1-score demonstrated the effectiveness of the proposed model in pneumonia classification tasks.

Keywords: Pneumonia, Image Classification, Chest X-ray, Convolutional Neural Networks.

Özet

Akciğer radyografileri, pnömoni olgularının değerlendirilmesinde en yaygın kullanılan görüntüleme yöntemleri arasındadır. Son yıllarda, bu görüntülerin yorumlanmasında tanısız doğruluğu artırmayı ve gözlemciler arası varyasyonu en aza indirmeyi amaçlayan bilgisayar destekli tanı (BDT) sistemlerinin kullanımı giderek yaygınlaşmaktadır. Bu çalışma, akciğer radyografilerinde pnömoninin otomatik sınıflandırılmasına yönelik derin öğrenme tabanlı bir yaklaşım önermektedir. Bu amaçla, anatomik farklılıkları algılayarak görüntüleri frontal projeksiyonlara göre kategorize edebilen çeşitli Evrişimli Sinir Ağı (ESA) mimarileri eğitilmiştir. Transfer öğrenimiyle ince ayar yapılan modeller arasında ShuffleNet_V2_x1_0 ve EfficientNet-B0, Chest X-Ray Images (Pneumonia) veri kümesinde en yüksek performansı göstermiştir. Eğitim süreci Adam optimizasyon algoritması ve çapraz entropi kayıp fonksiyonu kullanılarak yürütülmüştür. Doğruluk, duyarlılık, özgüllük ve F1-skora gibi değerlendirme metrikleri, önerilen modelin pnömoni sınıflandırma görevlerindeki etkinliğini ortaya koymuştur.

Anahtar Kelimeler: Pnömoni, Görüntü Sınıflandırma, Göğüs Röntgeni, Evrişimsel Sinir Ağları.

I. INTRODUCTION

Contemporary medical imaging technologies play a significant role in early diagnosis and disease detection processes [1], [2]. In particular, chest X-ray images are widely used for the assessment of lung diseases, cardiac abnormalities, and breast tissue pathologies [3]–[9]. However, the interpretation of these images depends largely on the experience of radiologists, and variations in expertise among healthcare professionals can lead to time-consuming evaluation processes. In this context, this study aims to develop a deep learning-based artificial intelligence model for the automatic classification of chest X-ray images.

Frontal X-ray images provide a general view of the heart, lung fields, ribs, and diaphragm domes such as the posterior border of the heart, retrosternal, and retrocardiac spaces. Therefore, correct classification of the pneumonia directly affects the clinical interpretability of the images.

The model developed in this study integrates the lightweight and efficient ShuffleNet_V2_x1_0 and EfficientNet-B0 architecture with transformer-based Vision Transformer (ViT) components to learn anatomical differences and perform projection classification tasks. ShuffleNet_V2_x1_0 and EfficientNet-B0, known for its depthwise separable convolutions and inverted residuals, achieved the highest performance scores among the evaluated architectures, making it the core of the proposed method. The model was trained on the Chest X-Ray Images (Pneumonia) dataset, utilizing pre-processing techniques such as contrast enhancement to emphasize anatomical details like organs and bone structures. This approach facilitates the identification of biological differences.

II. METHOD

In this section, the proposed CNN-based X-ray pneumonia classification approach is presented as shown in 1 and 2. The methodology involves pre-processing the Chest X-Ray Images (Pneumonia) dataset, correcting enlarging the corpus through extensive image-augmentation techniques that generate diverse synthetic samples, and subsequently training multiple deep-learning models for automatic classification.

To enhance contrast, the Contrast Limited Adaptive Histogram Equalization method was applied. This technique improves local contrast, making anatomical structures more distinguishable and allowing the model to extract more effective features. The

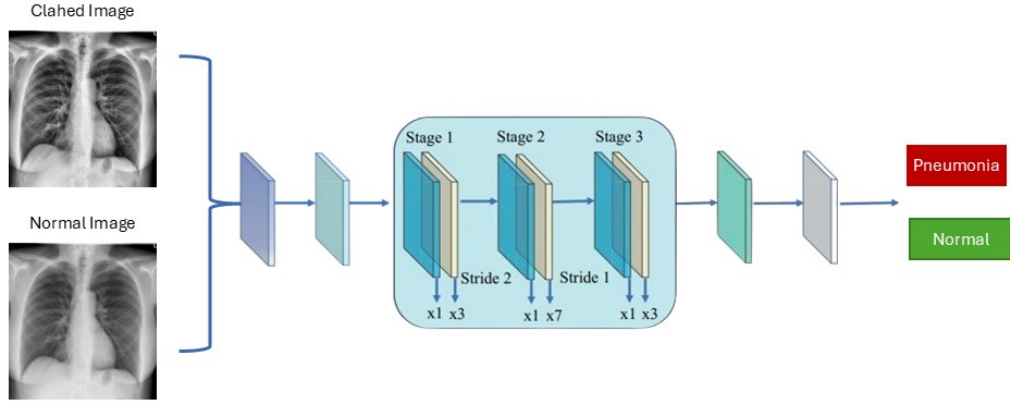


Figure 1: The Proposed ShuffleNetV2 Approach

images processed with CLAHE were directly fed into CNN models, ensuring that the learning process was based on sharpened visual representations.

Low-level features, such as edges, textures, and fundamental geometric structures learned in the early convolutional layers, remain valid for medical imaging, enabling these models to achieve higher accuracy with less data. However, training deep learning models from scratch requires a large amount of labeled data, significantly increasing computational costs and the risk of overfitting [10], [11]. Therefore, a transfer learning approach was adopted, utilizing models pre-trained on large-scale datasets to enable the network to learn general visual representations beforehand. Transfer learning facilitates an optimal initialization in the parameter space, thereby accelerating and stabilizing the gradient descent process [12], [13]. Various pretrained CNN models, including ShuffleNet_V2_x1_0 and EfficientNet-B0 were employed to pneumonia classify. The performance of the trained models was evaluated using metrics such as accuracy, sensitivity, specificity, and F1-score, providing a comprehensive analysis of classification effectiveness. The developed model aims to automatically classify X-ray pneumonia and reduce the workload of radiologists by accelerating the evaluation process. By comparing various CNN architectures, the most effective model for pneumonia classification was determined, ensuring a more reliable classification in medical imaging workflows.

III. EXPERIMENTAL EVALUATIONS

In this section, we evaluate the performance of our proposed approach for each radiology report, using the Chest X-Ray Images (Pneumonia) dataset for our experimental evaluations.

A. Dataset and Performance Metrics

Pneumonia detection in medical imaging aims to accurately classify pneumonia in chest X-ray images. The dataset used in this study was obtained from the Chest X-Ray Images (Pneumonia) dataset, which contains chest X-ray images labeled with pneumonia information. In addition to applying CLAHE preprocessing was also performed to enhance contrast and improve the model's feature extraction capability [14].

The performance of the proposed model was evaluated using various assessment metrics. Accuracy represents the proportion of correctly classified instances among all samples. Sensitivity (recall) measures the model's ability to correctly identify a specific projection type, while specificity indicates the extent to which incorrect classifications are avoided. The F1-score is calculated as the harmonic mean of precision and recall, providing a balanced evaluation of the model's performance [15].

Additionally, the area under the receiver operating characteristic curve (AUC-ROC) was computed to evaluate the model's ability to distinguish between different pneumonia [16]. The final classification performance of the model was determined by

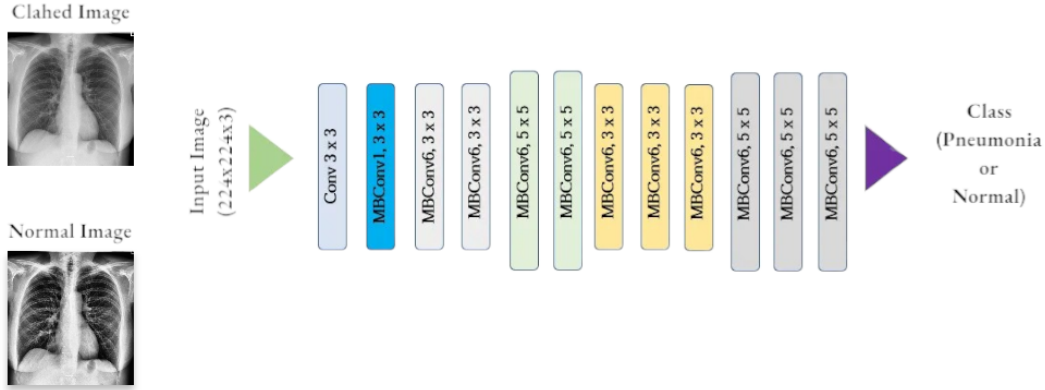


Figure 2: The Proposed Efficientnet-B0 Approach

averaging all evaluation metrics. Through this comprehensive evaluation approach, the reliability of the CNN-based pneumonia detection model was systematically validated.

Model	Train Loss	Train Accuracy	Validation Loss	Validation Accuracy	Precision	Recall	F1-Score	ROC-AUC	IoU
EfficientNet-B0	0.00918	0.99674	0.50752	0.875	0.800	1.000	0.889	1.000	0.800
ShuffleNetV2_x1_0	0.01257	0.99617	0.14781	0.875	0.800	1.000	0.889	1.000	0.800

Table I: Chest X-ray Pneumonia Classification Performance Metrics Results (Normal)

Model	Train Loss	Train Accuracy	Validation Loss	Validation Accuracy	Precision	Recall	F1-Score	ROC-AUC	IoU
EfficientNet-B0	0.00842	0.99636	0.16219	0.9375	0.8889	1.0000	0.9412	1.000	0.8889
ShuffleNetV2_x1_0	0.00341	0.99847	1.85789	0.6875	0.6154	1.0000	0.7619	1.000	0.6154

Table II: Chest X-ray Pneumonia Classification Performance Metrics Results (CLAHE)

Model	Train Loss	Train Accuracy	Validation Loss	Validation Accuracy	Precision	Recall	F1-Score	ROC-AUC	IoU
EfficientNet-B0	0.01696	0.99382	0.07094	0.96875	0.9412	1.0000	0.9697	1.000	0.9412
ShuffleNetV2_x1_0	0.02127	0.99262	0.12081	0.93750	0.8889	1.0000	0.9412	1.000	0.8889

Table III: Chest X-ray Pneumonia Classification Performance Metrics Results (Augmented)

B. Results and Discussion

This section presents a comparative analysis of the classification performance of various Convolutional Neural Network (CNN)-based architectures in identifying chest X-ray projection views. The models were evaluated under two preprocessing conditions: original images and contrast-enhanced images using CLAHE. The performance metrics considered include accuracy, sensitivity, specificity, precision, F1-score, and loss, and are summarized in Tables I.

The results demonstrate that CNN-based models consistently in all preprocessing scenarios. For original images, modern architectures such as ShuffleNetV2 and EfficientNetB0 achieved accuracies above 0.99 and maintained a strong balance between sensitivity and specificity, as shown in Table I.

As shown in Table I, nearly all CNN-based models achieved near-perfect or perfect scores, with ShuffleNetV2 and EfficientNetB0 reaching 1.000 in accuracy, sensitivity, specificity, precision, and F1-score. This indicates that CNNs are highly effective in utilizing the improved contrast. However, the ViT model struggled under the same conditions, with substantially lower scores across all metrics. This outcome supports existing findings that transformer-based models often require larger and more balanced datasets to generalize effectively in medical imaging tasks.

Overall, the findings highlight the robustness and adaptability of CNN-based models, particularly lightweight and optimized architectures, across different preprocessing techniques.

Model	Train Loss	Train Accuracy	Validation Loss	Validation Accuracy	Precision Macro	Recall Macro	F1-Score Macro
EfficientNet-B0	0.03532	0.98715	0.66981	0.7500	0.6667	0.5000	0.5556
ShuffleNetV2_x1_0	0.07041	0.97508	2.44627	0.5625	0.6667	0.3750	0.4444

Table IV: Chest X-ray Pneumonia Classification Performance Metrics Results (Multiclass)

IV. CONCLUSION

In this study, we evaluated the performance of various CNN architectures and a Vision Transformer model for classifying pneumonia views of chest X-ray images using the Chest X-Ray Images (Pneumonia) dataset. Each model was trained and tested on two image variations: the original images and contrast-enhanced images using CLAHE. The results demonstrate that preprocessing algorithms have a decisive impact on medical image analysis and that convolutional and transformer-based architectures offer effective approaches in radiological classification tasks.

REFERENCES

- [1] Y. Alaca, "Machine learning via darts-optimized mobilevit models for pancreatic cancer diagnosis with graph-based deep learning," *BMC Medical Informatics and Decision Making*, vol. 25, no. 1, pp. 1–21, 2025.
- [2] S. Bharati, P. Podder, and M. R. H. Mondal, "Hybrid deep learning for detecting lung diseases from x-ray images," *Informatics in Medicine Unlocked*, vol. 20, p. 100391, 2020.
- [3] N. R. S. Parveen and M. M. Sathik, "Detection of pneumonia in chest x-ray images," *Journal of X-ray Science and Technology*, vol. 19, no. 4, pp. 423–428, 2011.
- [4] J.-I. Toriwaki, Y. Suenaga, T. Negoro, and T. Fukumura, "Pattern recognition of chest x-ray images," *Computer Graphics and Image Processing*, vol. 2, no. 3–4, pp. 252–271, 1973.
- [5] D. A. Moses, "Deep learning applied to automatic disease detection using chest x-rays," *Journal of Medical Imaging and Radiation Oncology*, vol. 65, no. 5, pp. 498–517, 2021.
- [6] V. Chang, V. R. Bhavani, A. Q. Xu, and M. Hossain, "An artificial intelligence model for heart disease detection using machine learning algorithms," *Healthcare Analytics*, vol. 2, p. 100016, 2022.
- [7] N. Ghaffar Nia, E. Kaplanoglu, and A. Nasab, "Evaluation of artificial intelligence techniques in disease diagnosis and prediction," *Discover Artificial Intelligence*, vol. 3, no. 1, p. 5, 2023.
- [8] A. Alqahtani, S. Alsubai, M. Sha, L. Vilcekova, and T. Javed, "Cardiovascular disease detection using ensemble learning," *Computational Intelligence and Neuroscience*, vol. 2022, no. 1, p. 5267498, 2022.
- [9] F. Yasmin, S. M. I. Shah, A. Naeem, S. M. Shujaiddin, A. Jabeen, S. Kazmi, S. A. Siddiqui, P. Kumar, S. Salman, S. A. Hassan *et al.*, "Artificial intelligence in the diagnosis and detection of heart failure: the past, present, and future," *Reviews in cardiovascular medicine*, vol. 22, no. 4, pp. 1095–1113, 2021.
- [10] K. Weiss, T. M. Khoshgoftaar, and D. Wang, "A survey of transfer learning," *Journal of Big data*, vol. 3, pp. 1–40, 2016.
- [11] S. Panigrahi, A. Nanda, and T. Swarnkar, "A survey on transfer learning," in *Intelligent and Cloud Computing: Proceedings of ICICC 2019, Volume 1*. Springer, 2021, pp. 781–789.
- [12] I. Goodfellow, Y. Bengio, A. Courville, and Y. Bengio, *Deep learning*. MIT press Cambridge, 2016, vol. 1, no. 2.
- [13] Y. Bengio, A. Courville, and P. Vincent, "Representation learning: A review and new perspectives," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 8, pp. 1798–1828, 2013.
- [14] S. M. Pizer, M. Amburn, J. Austin *et al.*, "Adaptive histogram equalization and its variations," *Computer Vision, Graphics, and Image Processing*, vol. 39, no. 3, pp. 355–368, 1987.
- [15] C. J. Van Rijsbergen, "Precision and recall," *Journal of Information Retrieval*, vol. 7, pp. 313–323, 2013.
- [16] T. Fawcett, "An introduction to roc analysis," *Pattern Recognition Letters*, vol. 27, no. 8, pp. 861–874, 2006.