Analysis of the Performance of Deep Neural Networks in Diagnosing Pneumonia from Chest X-Ray Images

Reham Faisal Alsubhi-Nehal Hamed Alzhrani-Lamar Waleed Fattah
Data Science Department
S444003014@uqu.edu.sa- S444001073@uqu.edu.sa- S444006719@uqu.edu.sa

Abstract- Deep Learning is revolutionizing medical diagnostics, enabling faster and more accurate disease detection. This study utilizes chest X-ray images and advanced models such as DenseNet201, ResNet-18, EfficientNetB0, and CNN for pneumonia classification, achieving remarkable accuracy through data augmentation techniques. The results highlight the power of artificial intelligence in detecting complex patterns and its potential to enhance diagnostic workflows. Despite challenges like noisy data and class imbalance, this study underscores the significant promise of AI-driven tools in transforming healthcare and paving the way for broader diagnostic applications.

■ INTRODUCTION

In recent years, the healthcare sector has experienced a significant transformation driven by rapid advancements in Artificial Intelligence (AI) technologies. These developments have particularly impacted medical image analysis, where deep learning techniques have emerged as powerful tools to enhance diagnostic accuracy and efficiency. Among these techniques, Convolutional Neural Networks (CNNs) have proven to be highly effective in processing and analyzing complex visual data, such as medical images. These networks excel at recognizing intricate patterns in images, making them particularly suitable for applications such as classifying chest X-ray (CXR) images, which are widely used in diagnosing various diseases, including pneumonia.

Pneumonia is one of the most prevalent and serious respiratory diseases, contributing to high mortality rates, particularly among children under the age of five and the elderly. This condition arises from bacterial, viral, or fungal infections, leading to the alveoli filling with fluid and impairing respiratory functions. According to global statistics, pneumonia remains a major public health challenge, accounting for significant hospital admissions and deaths each year. Consequently, early and accurate diagnosis is critical to ensuring timely treatment, reducing complications, and alleviating the burden on healthcare systems.

Traditionally, the diagnosis of pneumonia relies on the analysis of CXR images by specialized radiologists. However, this process is not without challenges, as it is prone to human errors and variability in expertise, especially in regions with limited access to qualified medical professionals. These challenges highlight the urgent need for additional technological tools to enhance diagnostic accuracy and efficiency. In this context, AI technologies have demonstrated their ability to analyze medical images with greater speed and precision.

Recent studies have highlighted the significant potential of deep learning models, particularly CNNs, in improving pneumonia diagnosis. Advanced architectures such as DenseNet201, ResNet, and EfficientNet have achieved outstanding performance in classifying CXR images and differentiating between types of pneumonia, such as bacterial and viral infections. Additionally, techniques like transfer learning have been employed to enhance the performance of these models when dealing with limited datasets, thereby accelerating the training process and improving accuracy.

These AI-driven advancements offer tremendous opportunities for clinical applications. Not only do they enhance diagnostic accuracy, but they also provide scalable solutions to address the shortage of medical professionals in underserved areas. Moreover, these models can reduce the cognitive workload on radiologists, allowing them to focus on more complex cases.

This study aims to evaluate the performance of deep neural networks in diagnosing pneumonia using CXR images. By analyzing the capabilities of modern models, this research seeks to identify their strengths and limitations and provide insights into how these technologies can be effectively integrated into clinical workflows to support medical decision-making and improve patient outcomes.

Literature Review

Recent research has demonstrated the promise of deep neural networks in automated medical diagnostics by highlighting notable developments in its application for the diagnosis of pneumonia using chest X-ray (CXR) pictures. In their evaluation of both pre-trained and custom CNN models, Jain et al. (2020) highlighted the significance of high recall rates in lowering false negatives and achieved up to 92.31% accuracy with custom designs [2]. Similarly, Rahman et al. (2020) demonstrated the usefulness of transfer learning in overcoming data limitations by applying transfer learning with pre-trained CNNs, such as DenseNet201, on a dataset of 5,247 pictures, obtaining 98% accuracy for differentiating between normal and pneumonia cases [4]. Gabruseva et al. (2020) used sophisticated models such as SE-ResNext101 in the RetinaNet framework the same year. They utilized a dataset of more than 26,000 X-rays with bounding box annotations to efficiently identify lung opacities using single-shot detector (SSD) approaches. Their strategy, which combined multi-task learning, dropout regularization, and personalized augmentations, achieved high mean average accuracy (mAP) across a range of IoU thresholds, improving performance and reducing

In their study of simpler CNN designs, Rautaray et al. (2021) discovered that these models were computationally effective while retaining competitive accuracy, making them appropriate for contexts with limited resources [7]. Within the same year, Alsharif et al. (2021) presented PneumoniaNet, a 50-layer CNN designed for pediatric pneumonia. Using a dataset of 5,852 pictures, it achieved 99.7% accuracy with strong augmentation approaches and interpretability through Class Activation Mapping (CAM), demonstrating its usefulness in underserved areas [5]. With 98.1% accuracy, Nessipkhanov et al. (2023) addressed scalability and ethical issues while highlighting Deep CNNs' promise in medical imaging [6]. Lastly, Varshni et al. (2023) used DenseNet-169 in conjunction with SVM, demonstrating the usefulness of automated systems in remote locations with an AUC of 0.8002 [3]. All of these studies highlight how quickly deep learning techniques are developing, increasing their precision, effectiveness, and interpretability while opening the door for real-world applications in pneumonia detection.

Data Collection and Structure

The dataset used for this study has been taken from Kaggle and includes a total of 5863 chest X-ray images of children, aged between 1 and 5, collected from the Guangzhou Women and Children Medical Center. These diagnoses were verified by two specialists and further reviewed to make sure each diagnosis was correct. The given data is divided into three main folders: Train, Test, and Validation. Each of the mentioned folders includes two categories: Normal (0 label) and Pneumonia (1 label). Preprocessing included grayscale conversion to reduce computational complexity, normalization of pixel values in the range [0, 1], and reshaping into a single channel for compatibility. Data augmentation was done by randomly rotating, adjusting for zoom, and shifting; no flipping was done since anatomical directionality needs to be maintained in chest X-rays. Then the dataset was split with balanced classes into training, testing, and validation sets, and images were loaded into machine learning-compatible arrays that could easily fit into the model.

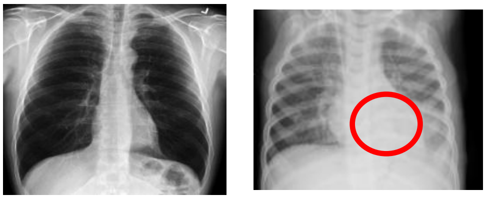

Fig. 1. Samples of normal and pneumonia chest x-rays

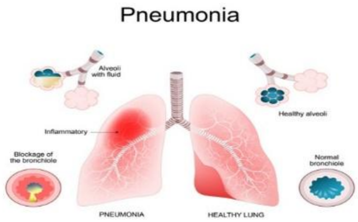

Fig. 2. Chest pneumonia explanation.

METHODOLOGY

this study utilizes four deep learning models—DenseNet201, EfficientNetB0, a custom CNN model, and ResNet-18—to classify chest X-ray images into two categories: NORMAL and PNEUMONIA. Across all models, the preprocessing stage involved resizing images to either 196×196 or 224×224 to standardize input dimensions and normalizing pixel values to the range [0, 1]. Data augmentation techniques such as rotation, horizontal flipping, zooming, and shifting were applied to address class imbalance and enhance data diversity. In some models, images were converted to grayscale to reduce complexity, and the datasets were split into training (80%), validation (10–20%), and testing (10–20%) sets.

The DenseNet201 model was chosen for its efficient feature extraction and dense connectivity, which improves feature reusability and reduces the number of parameters compared to other architectures like ResNet and VGG. DenseNet201 was initialized with pre-trained weights from the ImageNet dataset, and its top layers were customized to include a GlobalAveragePooling2D layer, a dense layer with 256 neurons and ReLU activation, and a Softmax output layer. The Adam optimizer (learning rate = 0.001) and the Categorical Crossentropy loss function were used,

along with Early Stopping to prevent overfitting.

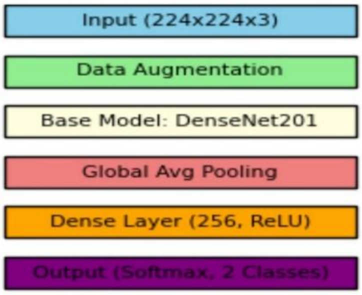

FIG.3.DENSENET201 ARCHITECTURE

EfficientNetB0 was employed for its scalability and efficiency in feature extraction. Batch normalization and dropout layers were used to stabilize training and reduce overfitting. The feature extraction progressed from detecting simple edges to complex patterns. The extracted features were passed through a dense layer with 256 units and ReLU activation, followed by a Softmax output layer for classification. The model used the Adamax optimizer and Categorical Crossentropy loss function to ensure effective weight adjustments and high accuracy.

The custom CNN model utilized convolutional layers with filters of various sizes (7×7, 5×5,3×3) to extract features. MaxPooling layers reduced dimensionality and enhanced computational efficiency, while dropout layers mitigated overfitting. The number of filters was incrementally increased from 8 to 128, enabling the model to capture both simple and complex patterns. Flatten and dense layers followed, with the final Softmax output layer providing class probabilities. The Adam optimizer with a low learning rate ensured stable training and effective differentiation between the two classes. The ResNet-18 model employed convolutional layers organized within ResNet blocks, each containing two layers with ReLU activation. The filters ranged from 8 to 128, with MaxPooling and Flatten layers extracting and reshaping features. A Dropout layer (20%) was included to prevent overfitting, while the final Softmax output layer performed binary classification. The model was trained using the Adam optimizer (learning rate = 0.001) and the Categorical Crossentropy loss function. Data augmentation and class weighting addressed the imbalance between NORMAL and PNEUMONIA cases. By integrating preprocessing, feature extraction, and classification, these models leveraged data augmentation and optimization techniques to improve performance and generalization. The use of pre-trained weights in DenseNet201 and EfficientNetB0 enhanced feature learning, while the custom CNN and ResNet-18 models demonstrated flexibility in capturing disease-specific patterns, ensuring accurate classification of chest X-ray images.

Fig.4.Convolutional Neural Network (CNN) architecture

RESULTS

DenseNet201 was used in this study to identify pneumonia in chest X-rays. As can be seen from the confusion matrix (Figure 5: "Confusion matrix highlighting recall challenges for healthy cases"), the model initially performed 82% accurately with simple preprocessing but also had trouble with class imbalance, When class weights were included, the accuracy fell to 56%, indicating that it was not very effective at improving generalization.

```
[[121 113]
 [  2 394]]
```

	precision	recall	f1-score	support
0	0.98	0.52	0.68	234
1	0.78	0.99	0.87	396
accuracy			0.82	630
macro avg	0.88	0.76	0.78	630
weighted avg	0.85	0.82	0.80	630

The accuracy increased by 88% as a result of data augmentation (rotation, shifts, zoom, shearing, flipping) and fine-tuning (extending training epochs, adding custom layers) (Figure 6: "Confusion matrix showing balanced recall and precision after enhancements").
Along with highlighting the importance of fine-tuning and augmentation in addressing class imbalance, the study also emphasizes DenseNet201's benefits in pattern detection and feature reuse. even if the precision was high.

```
[[171  63]
 [ 11 385]]
```

	precision	recall	f1-score	support
0	0.94	0.73	0.82	234
1	0.86	0.97	0.91	396
accuracy			0.88	630
macro avg	0.90	0.85	0.87	630
weighted avg	0.89	0.88	0.88	630

ResNet-18 was used to classify pneumonia patients, and the model's overall accuracy was 0.50, meaning that only 50% of the cases were successfully classified. Due mostly to the subpar performance in recognizing the PNEUMONIA class, the macro average (Macro Avg) analysis revealed both precision and recall to be noticeably low. The weighted average (Weighted Avg), which takes into consideration the support of each class, likewise produced poor results because the model was completely unable to recognize pictures that represented PNEUMONIA instances. This failure can be explained by the model's effectiveness in properly detecting normal images, which are by contrast less noisy, but failing to categorize noisy photos.

On the test dataset, the **EfficientNetB0** model performed exceptionally well, achieving 95.94% accuracy as well as high precision and recall, especially for the "PNEUMONIA" class (Recall: 0.99). With few misclassifications, the confusion matrix displays 346 accurate "PNEUMONIA" classifications and 103 accurate "NORMAL" classifications. Training was successful, reducing loss steadily, and epoch 5 produced the best results. to raise the recall of the "NORMAL" class (0.87). Pneumonia may be detected with great reliability using the EfficientNetB0 model.

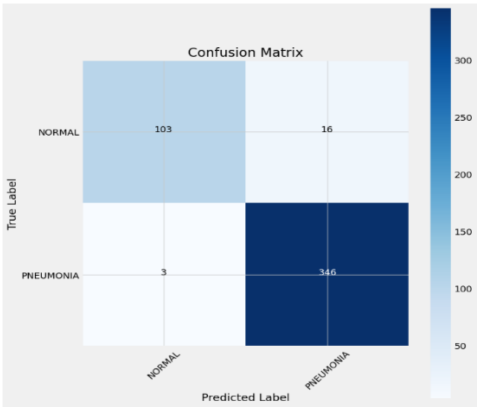

Fig.7. EfficientNetB0 confusion matrix

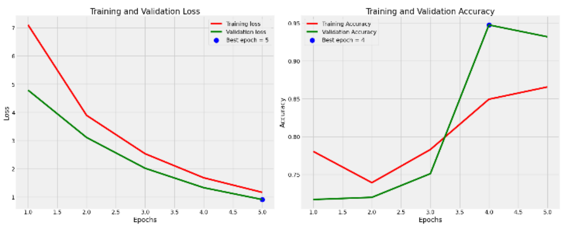

Fig.8. EfficientNetB0 validation Loss and accuracy

The accuracy of the **CNN** model for detecting pneumonia from chest X-rays was 91%, with 92% precision, 93% recall, and 93% F1-score for pneumonia cases. Along with 364 pneumonia cases and 201 normal cases, it accurately identified 33 false positives and 26 false negatives.

The reliability of the model is demonstrated by its weighted average F1-score of 91%. Its generality was improved and overfitting was avoided by early stopping. These findings show how useful the model is and how it could help radiologists diagnose pneumonia more accurately and quickly.

DISCUSSION

The effectiveness of several deep learning models in identifying pneumonia from chest X-ray pictures is demonstrated in this work. When class weights were applied, the DenseNet201 model's accuracy dropped to 56% because to its struggles with class imbalance, which had previously reached 82%. Its accuracy was raised to 88% by implementing data augmentation techniques (like rotation, zooming, and shifting) and fine-tuning tactics (like adding custom layers and increasing training epochs), which successfully addressed class imbalance and demonstrated its prowess in pattern detection.In contrast, the ResNet-18 model's total accuracy of 50% indicated subpar performance. It had poor precision and recall for this class as a result of its severe difficulties with noisy pneumonia images. However, because normal photos had lower noise levels, it did better with them.shown remarkable dependability, successfully classifying the majority of cases and attaining an overall accuracy of 95.94% with a recall of 0.99 for the pneumonia class. Nevertheless, there is still opportunity for improvement given the recall for the usual class (0.87). EfficientNetB0's training approach proved effective, consistently reducing loss and producing the best results at epoch 5.Moreover, the CNN model demonstrated strong performance, attaining a weighted F1-score of 91%, 92% precision, 93% recall, and 91% accuracy. Through early halting, it prevented overfitting and successfully classified the majority of pneumonia and normal patients.

These results highlight how crucial data pretreatment, augmentation, and fine-tuning are to enhance model performance.

efficientnetb0 was the most dependable framework, even if densenet201 and efficientnetb0 demonstrated impressive capabilities. This study shows how deep learning models can help radiologists diagnose pneumonia more quickly and accurately, which will eventually improve patient outcomes.

to increase the models' generalizability, future research should concentrate on improving them even more by tackling important issues like growing the dataset to include more varied samples from a range of demographics and imaging circumstances. Furthermore, more sophisticated denoising methods or training approaches might be created to better handle low-quality and noisy x-ray images, particularly for models such as resnet-18. More research into optimizing pretrained models using domain-specific datasets may improve their capacity to extract features. Additionally, it would be beneficial to use explainability strategies like grad-cam,

to enhance usability and trust by giving radiologists more in-depth knowledge of model predictions. finally, implementing these models in real-time clinical processes and evaluating their performance in larger healthcare environments may show their usefulness and open the door for their application in the diagnosis of further lung conditions as covid-19 or tuberculosis.

CONCLUSION

Deep learning techniques have been highly instrumental in the development of a range of diagnostic tools, with the rapid development of technologies and changes in the face of medicine. Deep learning model studies on the classification of pneumonia based on chest X-ray images make a great contribution to this context. The high performance, allowed by the strategic incorporation of the DenseNet201 architecture with data augmentation techniques of rotation, zooming, and adding custom layers, underlines deep learning algorithms for the finding of complex patterns in medical images.

These findings extend beyond proving the efficiency of the models used but also provide a wide-angle view on the impact of the employment of such computational tools in healthcare. The models' clinical values were reaffirmed with their high accuracy combined with advanced levels of recall and precision. This would mean that AI can, if applied appropriately, turn out to be an effective partner to human experience for hastening diagnosis and increasing the precision of therapeutic interventions. But like any other computational system, it does need cautious interpretation. Though the performance of the model in this study turned out to be efficient within the scope of the dataset used for training, further diversity might show up in a real-world application. There is a dire need therefore to continually upgrade and adapt these models to various clinical environments and situations. In other words, this work contributes to the ever-growing body of evidence that deep learning tools have a place in medical diagnostics. These promising results in pneumonia classification represent a significant step toward future applications in other diagnostic domains. Yet, the road is endless, as success in this domain is sustainable only with continuous efforts toward model enhancement and increased usage to meet ever-evolving healthcare demands.

REFERENCES

[1] R. Kundu, R. Das, Z. W. Geem, G.-T. Han, and R. Sarkar, "Pneumonia detection in chest X-ray images using an ensemble of deep learning models," *PLoS ONE*, vol. 16, no. 9, pp. e0256630, Sep. 2021. doi: 10.1371/journal.pone.0256630.

[2] [2] R. Jain, P. Nagrath, G. Kataria, V. S. Kaushik, and D. J. Hemanth, "Pneumonia detection in chest X-ray images using convolutional neural networks and transfer learning," *Measurement*, vol. 173, p. 108046, 2020, doi: 10.1016/j.measurement.2020.108046.

[3] D. Varshni, R. Nijhawan, K. Thakral, A. Mittal, and L. Agarwal, "Pneumonia detection using CNN-based feature extraction," in *2019 International Conference on Signal Processing and Communication (ICSC)*, 2019, pp. 1–6, doi: 10.1109/ICSC45622.2019.8938373.

[4] 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, pp. 3233, May 2020, doi: 10.3390/app10093233.

[5] R. Alsharif, Y. Al-Issa, A. M. Alqudah, I. Abu Qasmieh, W. A. Mustafa, and H. Alquran, "PneumoniaNet: Automated detection and classification of pediatric pneumonia using chest X-ray images and CNN approach," *Electronics*, vol. 10, no. 11, pp. 2949, Nov. 2021. doi: 10.3390/electronics10232949

[6] D. Nessipkhanov, V. Davletova, N. Kurmanbekkyzy, and B. Omarov, "Deep CNN for the identification of pneumonia respiratory disease in chest X-ray imagery," *International Journal of Advanced Computer Science and Applications (IJACSA)*, vol. 14, no. 10, pp. 652–661, Oct. 2023, doi: 10.14569/IJACSA.2023.0141065.

[7] H. G. M., M. K. Gourisaria, S. S. Rautaray, and M. Pandey, "Pneumonia detection using CNN through chest X-ray," *Journal of Engineering Science and Technology*, vol. 16, no. 1, pp. 861–876, Feb. 2021.