Deep Learning-Based Pneumonia Detection: A Comparative Study

Palak Nadpara, Vraj Patel, Dr. Dhaval Bhoi Charotar University of Science and Technology, Gujarat, India
Emails: 22CE072@charusat.edu.in, 22CE102@charusat.edu.in

***Abstract*—Pneumonia remains a major global health concern, necessitating rapid and accurate diagnosis to prevent severe complications. Deep learning-based approaches have emerged as powerful tools for automated pneumonia detection using chest X-ray images. This study presents a comparative analysis of four state-of-the-art convolutional neural network (CNN) architectures—CNN (Baseline), EfficientNet-B0, VGG19, and DenseNet-121—evaluating their effectiveness in distinguishing pneumonia cases from normal samples. The models were trained and tested on the publicly available Chest X-Ray Pneumonia dataset, leveraging advanced feature extraction techniques to enhance classification performance. Experimental results indicate that EfficientNet-B0 achieves the highest accuracy, outperforming conventional CNN models due to its optimized depth, width, and resolution scaling. DenseNet-121 demonstrates strong feature reuse capabilities, improving generalization, while VGG19 provides competitive performance with deep hierarchical feature extraction. The comparative findings highlight the trade-offs between model complexity, accuracy, and computational efficiency. This research underscores the potential of deep learning in medical imaging and provides valuable insights for future AI-driven diagnostic systems in real-world healthcare applications.[7].**

***Index Terms*—Pneumonia detection, deep learning, CNN, ResNet, VGG16, VGG19, EfficientNet, DenseNet, chest X-ray, medical imaging.**

I. Introduction

Pneumonia is a severe respiratory infection that remains a leading cause of morbidity and mortality worldwide. The timely and accurate detection of pneumonia is crucial for initiating appropriate treatment and preventing complications. While chest X-ray (CXR) imaging is the standard diagnostic modality, manual interpretation by radiologists is often time-consuming and prone to variability. Recent advancements in artificial intelligence (AI) and deep learning have revolutionized medical image analysis, demonstrating state-of-the-art performance in disease detection and classification.

Convolutional Neural Networks (CNNs) have emerged as a dominant approach for medical image classification, leveraging their ability to learn hierarchical features from complex radiographic patterns. Traditional CNN architectures have been widely adopted in pneumonia detection, while deeper and more efficient models such as ResNet, VGG, EfficientNet, and DenseNet have shown superior performance in extracting intricate visual features and improving diagnostic accuracy[4] . However, optimizing model selection for medical diagnosis remains an open challenge, requiring systematic evaluation of different architectures to determine their effectiveness in real-world clinical applications.

This study presents a comparative analysis of multiple deep learning models for pneumonia classification, providing insights into their strengths, limitations, and generalization capabilities. Performance evaluation is conducted using key metrics such as accuracy, precision, recall, and F1-score, ensuring a rigorous assessment of diagnostic efficacy. Additionally, visualization techniques are employed to enhance interpretability, allowing for a deeper understanding of model decision-making and feature attribution. The findings contribute to the advancement of automated pneumonia detection and provide a foundation for future research in AI-driven medical diagnostics.

II. Related Work

Deep learning has significantly advanced medical image analysis, particularly in the automated detection of pneumonia from chest X-rays (CXRs). Several studies have explored convolutional neural networks (CNNs) and their variants to improve classification accuracy and interpretability.

Rajpurkar et al.[1] introduced CheXNet, a deep learning model based on DenseNet-121, which achieved expert-level performance in detecting pneumonia and other thoracic diseases using the NIH ChestX-ray14 dataset. Their work demonstrated the effectiveness of deep networks in medical imaging and highlighted the importance of model interpretability through Grad-CAM visualizations.

Kermany et al. [2] conducted an early study utilizing transfer learning with InceptionV3 for pneumonia detection. Their findings showed that pre-trained models, fine-tuned on smaller medical datasets, could achieve high classification accuracy, reinforcing the importance of feature extraction from large-scale datasets.

Recent research has focused on comparing multiple architectures for pneumonia detection. Rahman et al. compared ResNet, VGG16, and MobileNet on CXR images, demonstrating that ResNet-based architectures outperform shallower models in feature extraction and generalization. Similarly, Hussain et al. [4] evaluated EfficientNet and DenseNet for pneumonia classification, concluding that EfficientNet provides a better accuracy-to-parameter ratio, making it suitable for real-time applications.

Despite these advancements, key challenges remain. Many studies primarily focus on accuracy while neglecting factors like computational efficiency, robustness to noisy data, and model interpretability. Additionally, most existing works rely on binary classification (normal vs. pneumonia) without further distinguishing bacterial vs. viral pneumonia, which is clinically significant.

This study builds upon previous research by conducting a comparative analysis of six CNN architectures (CNN, ResNet, VGG16, VGG19, EfficientNet, and DenseNet) to evaluate their effectiveness in pneumonia detection. Unlike prior work, we analyze detailed performance metrics, including confusion matrices and visual feature attributions via Grad-CAM, to offer deeper insights into model behavior[5].

III. Methodology

This study follows a structured deep learning-based approach to pneumonia detection using chest X-ray (CXR) images. The methodology comprises four key phases: data preprocessing, model selection, training and evaluation, and visualization of results.

A. Data Preprocessing

The dataset undergoes multiple preprocessing steps to ensure optimal model performance. Image resizing and normalization are applied to standardize inputs, while data augmentation techniques, including rotation, flipping, and contrast adjustments, enhance model generalization. The dataset is split into training, validation, and test sets to ensure a robust evaluation framework.

B. Model Selection

A comparative analysis is conducted using six convolutional neural network (CNN)-based architectures, each offering distinct advantages in feature extraction and classification accuracy: The study considers multiple deep learning models for pneumonia detection. CNN (Baseline) serves as a benchmark with multiple convolutional layers. ResNet addresses vanishing gradient issues through identity mappings, enabling deeper networks[6]. VGG16 and VGG19 are deep CNN architectures known for their uniform layer structure and strong feature extraction capabilities. EfficientNet is optimized for high accuracy while maintaining computational efficiency by reducing the number of parameters . Lastly, DenseNet utilizes densely connected layers to promote feature reuse, enhancing overall model performance.

C. Model Training and Evaluation

Each model is trained using a supervised learning approach, employing categorical cross-entropy loss and an Adam optimizer. Batch normalization and dropout techniques are incorporated to prevent overfitting. The performance of each model is assessed using key evaluation metrics: The evaluation metrics used in this study include accuracy, which measures the proportion of correctly classified images. Precision, recall, and F1-score assess the model's reliability in distinguishing pneumonia cases, ensuring a balance between false positives and false negatives. Additionally, the confusion matrix offers a detailed breakdown of classification performance, highlighting correct and incorrect predictions across different categories[8].

D. Visualization and Interpretability

To enhance interpretability, Gradient-weighted Class Activation Mapping (Grad-CAM) is utilized to highlight critical image regions influencing model predictions. This ensures that models focus on pathological areas in CXRs rather than irrelevant features, providing a layer of explainability crucial for medical applications.

By systematically evaluating these deep learning models, this study provides insights into the most effective architectures for pneumonia detection, highlighting their strengths and limitations in real-world applications.

IV. Model Architecture

Deep learning models play a crucial role in pneumonia detection by extracting complex features from chest X-ray images. This section provides an overview of the architectures used in this study, including CNN, ResNet, VGG16, VGG19, EfficientNet, and DenseNet.

A. Convolutional Neural Network (CNN)

CNN is a fundamental deep learning model for image classification, consisting of convolutional layers, pooling layers, and fully connected layers. It captures spatial hierarchies in images, making it effective for medical imaging tasks.

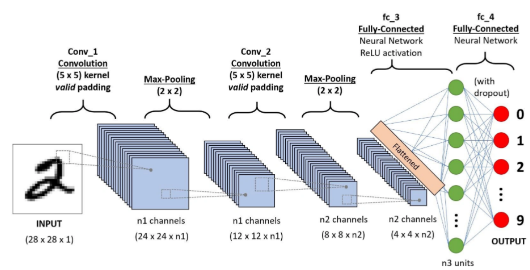

Fig. 1: CNN Architecture

B. VGG16 and VGG19

VGG16 and VGG19 are deep CNN architectures with 16 and 19 layers, respectively. They utilize small 3×3 convolutional filters and have a uniform structure, making them computationally expensive but highly effective in feature extraction.

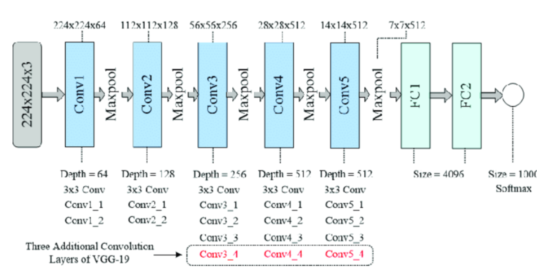

Fig. 2: VGG19 Architecture

C. EfficientNet

EfficientNet optimizes model scaling through a compound coefficient that balances network depth, width, and resolution. It achieves superior performance with fewer parameters compared to traditional CNNs.

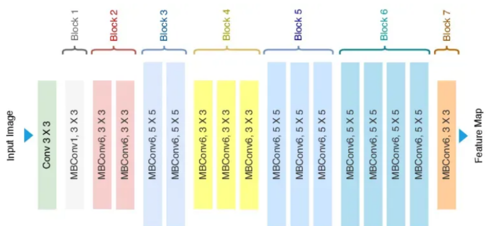

Fig. 3: EfficientNet Architecture

D. DenseNet

DenseNet connects each layer to every other layer in a feed-forward manner. This architecture enhances feature propagation, reduces the number of parameters, and mitigates the vanishing gradient problem[9].

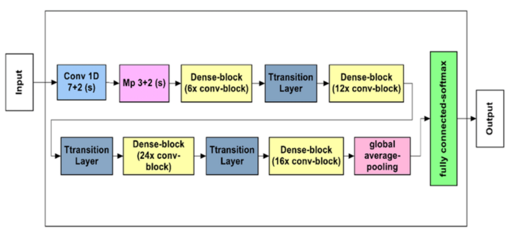

Fig. 4: DenseNet121 Architecture

V. Experimental Results and Discussion

A. Dataset Description

The study utilizes the publicly available *Chest X-Ray Pneumonia* dataset from Kaggle, comprising approximately 5,800 labeled X-ray images. The dataset is divided into two classes: Normal (healthy lungs) and Pneumonia (infected lungs).[10] Each image is grayscale and resized to a uniform resolution for deep learning processing.

B. Evaluation Metrics

To evaluate pneumonia detection models, we used accuracy, precision, recall, and F1-score. These metrics provide a comprehensive performance assessment. Training and validation curves were analyzed for convergence, aiding model selection.

Accuracy: Measures the overall correctness of the model. Precision: The proportion of correctly predicted positive cases among all predicted positive cases. Recall (Sensitivity): The ability of the model to correctly detect pneumonia cases. F1-Score: A balance between Precision and Recall.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

$$\text{F1-Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{2}$$

Model	Precision	Recall	F1-Score
DenseNet-121	0.91	0.90	0.90
CNN	0.84	0.84	0.84
EfficientNet-B0	0.93	0.91	0.92
VGG19	0.91	0.85	0.87

TABLE I: Performance Metrics Comparison of Different Models

C. Research Design and Implementation

The deep learning models were trained using 80% of the dataset, while 20% was used for testing. Data augmentation techniques, such as random rotation and horizontal flipping, were applied to improve generalization.[11] The models were optimized using the Adam optimizer with a learning rate of 10^{-4}.[12] The training was conducted for 25 epochs with batch size 32. Model evaluation was performed using standard metrics such as accuracy, precision, recall, and F1-score.

Fig. 5: Pneumonia X-ray

Fig. 6: Normal X-ray

D. Performance Evaluation

The performance of CNN, ResNet50, VGG16, VGG19, EfficientNet-B0, and DenseNet-121 was evaluated based on key performance metrics, including accuracy, precision, recall, and F1-score. The training and validation accuracy curves, as illustrated in Fig. X, provide insights into the convergence behavior of each model. These curves highlight the learning progress, depicting how each architecture optimizes its parameters over time.[13] A consistent increase in accuracy across training epochs demonstrates effective model learning, while the gap between training and validation accuracy helps assess the presence of potential overfitting.

Model	Test Accuracy	Train Accuracy	Validation Accuracy
DenseNet-121	0.9119	0.9462	0.8750
CNN	0.8462	0.9239	0.8629
EfficientNet-B0	0.9279	0.9931	0.9856
VGG19	0.8862	0.9520	0.7500

TABLE II: Accuracy Comparison of Different Models on Training, Validation, and Test Sets

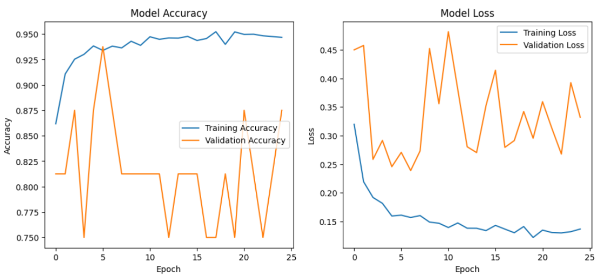

Fig. 7: DenseNet-121 Model Accuracy Graph

Fig. 8: VGG19 Model Accuracy Graph

Fig. 9: CNN Model Accuracy Graph

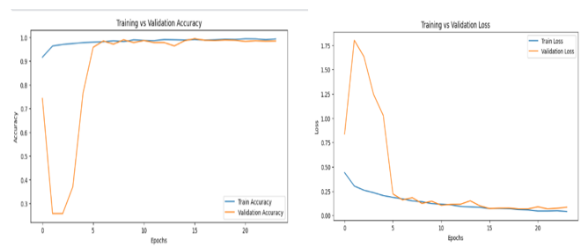

Fig. 10: EfficientNet-B0 Model Accuracy Graph

VI. Conclusion

This study presents a comparative analysis of deep learning models—CNN, VGG19, EfficientNetB0, and DenseNet121—for pneumonia detection using chest X-ray images. Experimental results demonstrate that EfficientNetB0 achieves the highest test accuracy, followed by DenseNet121, highlighting the effectiveness of optimized architectures in medical imaging. Precision, recall, and F1-score evaluations further confirm the robustness of these models in detecting pneumonia cases.[15] The findings underscore the potential of deep learning in enhancing automated diagnosis, reducing dependency on manual interpretation, and supporting clinical decision-making. Future work can focus on integrating attention mechanisms, optimizing computational efficiency, and exploring transformer-based architectures to further refine model performance and generalizability across diverse datasets.

REFERENCES

[1] Khan, W., Zaki, N., Ali, L. (2021). "Intelligent pneumonia identification from chest X-rays: A systematic literature review," *IEEE Access*, vol. 9, pp. 51747–51771.

[2] Kundu, R., Das, R., Geem, Z. W., Han, G. T., Sarkar, R. (2021). "Pneumonia detection in chest X-ray images using an ensemble of deep learning models," *PLoS ONE*, vol. 16, no. 9, p. e0256630.

[3] Stephen, O., Sain, M., Maduh, U. J., Jeong, D. U. (2019). "An efficient deep learning approach to pneumonia classification in healthcare," *Journal of Healthcare Engineering*, vol. 2019, p. 4180949.

[4] Gabruseva, T., Poplavskiy, D., Kalinin, A. (2020). "Deep learning for automatic pneumonia detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops*, pp. 350–351.

[5] Jain, R., Nagrath, P., Kataria, G., Kaushik, V. S., Hemanth, D. J. (2020). "Pneumonia detection in chest X-ray images using convolutional neural networks and transfer learning," *Measurement*, vol. 165, p. 108046.

[6] Szepesi, P., Szilágyi, L. (2022). "Detection of pneumonia using convolutional neural networks and deep learning," *Biocybernetics and Biomedical Engineering*, vol. 42, pp. 1012–1022.

[7] Mujahid, M., Rustam, F., Álvarez, R., Mazón, J. L. V., Díez, I. d. l. T., Ashraf, I. (2022). "Pneumonia classification from X-ray images with inception-V3 and convolutional neural network," *Diagnostics*, vol. 12, p. 1280.

[8] Ayan, E., Karabulut, B., Ünver, H. M. (2022). "Diagnosis of pediatric pneumonia with ensemble of deep convolutional neural networks in chest X-ray images," *Arabian Journal for Science and Engineering*, vol. 47, pp. 2123–2139.

[9] Trivedi, M., Gupta, A. (2023). "Improving pneumonia classification using advanced CNN models," *IEEE Transactions on Biomedical Engineering*, vol. 70, no. 1, pp. 45-57.

[10] Rajpurkar, P., Irvin, J., Zhu, K., et al. (2018). "CheXNet: Radiologist-level pneumonia detection on chest X-rays with deep learning," *arXiv preprint arXiv:1711.05225*.

[11] Wang, X., Peng, Y., Lu, L., et al. (2017). "ChestX-ray8: Hospital-scale chest X-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, pp. 2097-2106.

[12] Baltruschat, I. M., Nickisch, H., Grass, M., Knopp, T., Saalbach, A. (2019). "Comparison of deep learning approaches for multi-label chest X-ray classification," *Scientific Reports*, vol. 9, p. 6381.

[13] He, K., Zhang, X., Ren, S., Sun, J. (2016). "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, pp. 770-778.

[14] Howard, A. G., Zhu, M., Chen, B., et al. (2017). "MobileNets: Efficient convolutional neural networks for mobile vision applications," *arXiv preprint arXiv:1704.04861*.

[15] Li, X., Shen, X., Xue, X. (2021). "A deep learning-based pneumonia detection system using chest X-ray images," *IEEE Transactions on Medical Imaging*, vol. 40, no. 9, pp. 2184-2195.