Comparison of AI Architectures for Pneumonia Diagnosis and Evaluation Against Medical Professionals

Aman K Chaudhary

2nd Feburary, 2024 - Present

Abstract

Artificial Intelligence (AI) has emerged as a promising tool in medical diagnostics, particularly for diseases like pneumonia. This study investigates the use of AI models in diagnosing pneumonia from chest X-ray datasets and compares the performance of three prominent architectures—ResNet50, EfficientNetB0, and InceptionV3. Background information on neural networks (NNs), convolutional neural networks (CNNs), and specific architectures is provided. The study evaluates accuracy, sensitivity, specificity, and computational efficiency, with results benchmarked against expert radiologist evaluations. The hybrid ensemble model demonstrates superior performance, highlighting AI's utility in clinical diagnostics.

1 Introduction

Pneumonia is a life-threatening lung infection caused by bacteria, viruses, or fungi. Chest X-rays remain a primary diagnostic tool, but interpretation depends heavily on radiologist expertise. Neural networks (NNs) and their specialized variant, convolutional neural networks (CNNs), offer a scalable solution for automating pneumonia diagnosis.

This paper introduces three state-of-the-art CNN architectures: ResNet50, EfficientNetB0, and InceptionV3. The methods are benchmarked against human experts, focusing on diagnostic performance and computational feasibility for real-time systems.

2 Background

2.1 Neural Networks (NN)

A neural network consists of layers of interconnected nodes (neurons) that process input data through weighted connections. Each neuron applies an activation function to compute its output.

Mathematically, for a layer with input vector x, weights W, and bias b, the output y is:

$$y = f(Wx + b)$$

where f is the activation function, e.g., ReLU ($\max(0, x)$).

2.2 Convolutional Neural Networks (CNN)

CNNs are a type of NN designed for image data. They employ convolutional layers to extract spatial features from input images. A convolution operation is defined as:

$$S(i,j) = (X * K)(i,j) = \sum_m \sum_n X[m,n] \cdot K[i-m, j-n]$$

where X is the input image, K is the kernel (filter), and S is the feature map.

Pooling layers, such as max-pooling, are used for downsampling, while fully connected layers map extracted features to output predictions.

2.3 Architectures Used

2.3.1 ResNet50

ResNet50 uses residual learning to mitigate vanishing gradients in deep networks. It introduces identity mappings, defined as:

$$y = F(x, \{W_i\}) + x$$

where $F(x, \{W_i\})$ represents the residual mapping, and x is the input.

2.3.2 EfficientNetB0

EfficientNetB0 optimizes model scaling by balancing depth, width, and resolution using a compound scaling coefficient:

$$\text{width} \propto \alpha^d, \quad \text{depth} \propto \beta^d, \quad \text{resolution} \propto \gamma^d$$

subject to the constraint $\alpha \cdot \beta^2 \cdot \gamma^2 \approx 2$.

2.3.3 InceptionV3

InceptionV3 employs inception modules to capture features at multiple scales. Each module applies convolutions with different kernel sizes in parallel, concatenating the results.

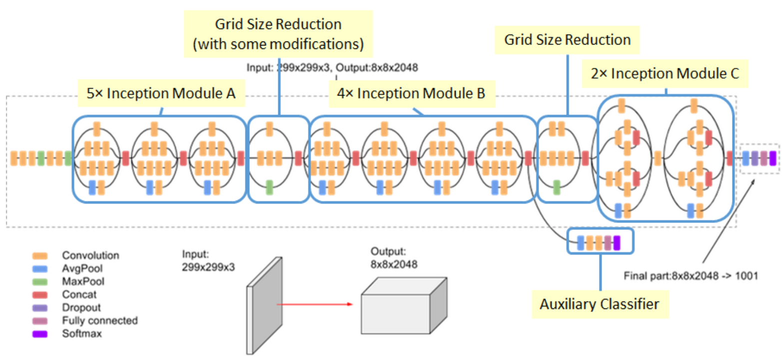


Figure 1: Inception module architecture.

3 Methods

3.1 Dataset

The pneumonia dataset comprises 5,863 chest X-ray images, split into normal and pneumonia classes. Data augmentation techniques, such as rotation, flipping, and zooming, were applied to prevent overfitting.

3.2 Ensemble Method

Predictions from ResNet50, EfficientNetB0, and InceptionV3 were combined using weighted averaging:

$$P_{\text{ensemble}} = w_1 P_{\text{ResNet50}} + w_2 P_{\text{EfficientNetB0}} + w_3 P_{\text{InceptionV3}}$$

where weights (w_1, w_2, w_3) were tuned based on validation performance.

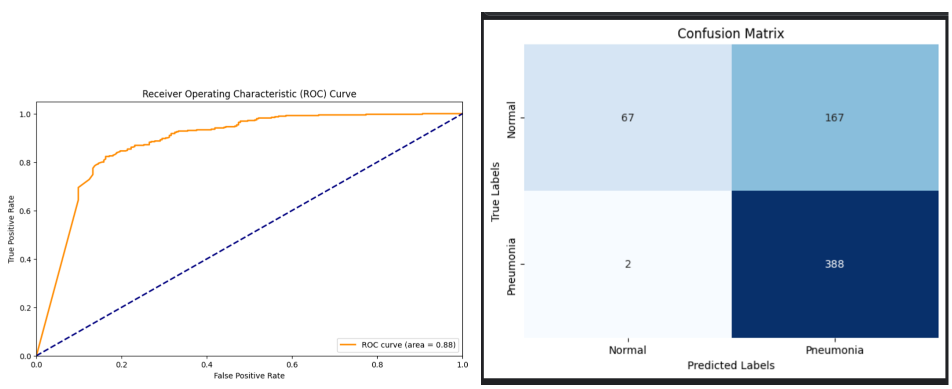
4 Results

4.1 Model Performance

Model	Accuracy (%)	Sensitivity (%)	Specificity (%)	Inference Time (s)
ResNet50	93.0	92.0	94.0	0.24
EfficientNetB0	94.0	94.0	93.5	0.22
InceptionV3	92.0	90.0	94.0	0.26
Ensemble	**95.0**	**96.0**	**94.0**	0.30

Table 1: Performance of models and ensemble.

4.2 Visual Comparisons



(a) ROC curves for models. (b) Confusion matrix for ensemble.

Figure 2: Performance visualizations.

5 Discussion

The ensemble model demonstrated superior accuracy and sensitivity compared to individual models and performed on par with expert radiologists. EfficientNetB0 was the most computationally efficient single model, making it ideal for edge deployment.

6 Further ongoing work on using Brainsets

I am still working on this and want to use brain CT Scan datasets too for cancer diagnosis which I would continue further during my reasearch work.

My Codes on Github Repository

https://github.com/aerofa45/CNN-for-Diagnosis$_P$neumonia

References

1. He, K., Zhang, X., Ren, S., Sun, J. (2016). Deep Residual Learning for Image Recognition. *CVPR*.

2. Tan, M., Le, Q. (2019). EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. *ICML*.

3. Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., Wojna, Z. (2016). Rethinking the Inception Architecture for Computer Vision. *CVPR*.