Enhancing Pneumonia Detection: A Comparative Study of CNN, VGG16, and ResNet50 Models

COM S-573 Project Final Report

Roop Sagar Mangineni, Ankit Jyothish, Lakshmi Narayana Aditya Akilesh Mantha
Department of Computer Science, Electrical and Computer Engineering, Iowa State University, Ames, IA, United States, 50010

Abstract

Pneumonia remains a significant global health concern, necessitating advanced diagnostic tools for timely and accurate detection. However, several pneumonia detection methodologies use Machine Learning and Deep Learning technologies. This study focuses on pneumonia detection conducting a comparative study of Convolutional Neural Network (CNN), VGG16, and ResNet50 models. The implementation involves training these deep learning models on a comprehensive dataset of chest X-ray images to evaluate their respective performance in pneumonia detection. Further, a comparative analysis between CNN, VGG16, and ResNet50 was performed to identify the most effective model for detecting Pneumonia. The purpose is to achieve higher accuracy and reliability in pneumonia detection, advancing the field of medical image analysis. By synergizing effective architecture design, hyperparameter optimization, and interpretability, this research aims to contribute to the development of advanced deep-learning models for vital healthcare applications.

1 Introduction

Pneumonia, a prevalent respiratory infection, poses significant public health challenges worldwide. It is considered the greatest cause of child fatalities all over the world. Approximately 1.4 million children die of Pneumonia yearly, 18% of children under five years old. Globally, two billion people are suffering from Pneumonia every year. However, early and accurate detection of Pneumonia is crucial for timely and effective treatment, improving patient outcomes, and reducing mortality rates.

In recent years, advancements in artificial intelligence (AI) and deep learning, specifically Convolutional Neural Networks (CNNs), have shown promising potential in medical image analysis, including diagnosing Pneumonia. CNNs are deep learning models designed to automatically and adaptively learn hierarchical features from input data, making them particularly effective for image-related tasks. The intersection of healthcare and machine learning offers a promising avenue to enhance diagnostic capabilities. This project delves into pneumonia detection using Convolutional Neural Networks (CNN), specifically focusing on our proposed customized CNN, VGG16, and ResNet50 architectures. We also review existing studies and research that have utilized CNNs for pneumonia detection, highlighting their methodologies, performance, and potential areas for improvement. By understanding the advancements in this field, we aim to contribute to the ongoing research efforts to enhance pneumonia diagnosis and ultimately improve patient care and outcomes.

The dataset employed for this task is obtained from Kaggle and contains many chest X-ray images annotated as pneumonia-positive or pneumonia-negative. The dataset is already structured for supervised learning, making it an ideal resource for training and testing machine learning models. The machine learning task at hand is binary classification: distinguishing between images of chest X-rays that depict normal cases and those that reveal Pneumonia. This task holds substantial clinical relevance and demonstrates the potential of deep learning in medical image analysis. This project compares the various model architectures of CNNs used to detect Pneumonia from chest X-ray images. We discuss the architecture and workings of CNNs, emphasizing their ability to extract features and patterns from medical images. Furthermore, we delve into training a CNN model, including data preprocessing, model selection, and optimization techniques, to achieve optimal performance in pneumonia detection.

The primary machine learning models under examination are the VGG16 architecture, ResNet50V2 architecture, and our customized CNN architecture. Both VGG16 and ResNet50V2 are recognized for their

proficiency in image classification tasks. We aim to accurately distinguish Pneumonia from normal cases by deploying these models. Furthermore, the project also investigates model interpretability, aiming to uncover key features and patterns that drive the model's decision-making process.
The expected outcomes of this project are twofold: first, to develop a robust pneumonia detection model with high accuracy and generalization, and second, to gain insights into the features and characteristics that influence the model's predictions. These insights could be valuable in improving the diagnosis of Pneumonia and advancing our understanding of medical image analysis using deep learning techniques. This research's potential benefits span healthcare, where early and accurate diagnosis can significantly impact patient outcomes, and the broader field of machine learning, where interpretable models are growing interest.

2 Related Work

Pneumonia detection using chest X-rays has been an open research problem for many years. During COVID-19, the need for research in Pneumonia spiked to help with diagnosis. There are many earlier works where the authors have tried to perform pneumonia detection using state-of-the-art machine-learning techniques. Work has been done in Pneumonia detection using deep learning techniques like CNN using ReLU activation function with 8M parameters generating an accuracy of 88.9% [1]. A technique was proposed for classifying pulmonary tuberculosis using two different DCNNs, AlexNet and GoogleNet, with a very AUC of 0.99 [2]. Using multiple architectures and ensembling has also shown a significant increase in accuracy to 98.91% [3]. An ensemble-learning approach was made to detect Pneumonia in chest X-rays in which pretraining of the model was done on ImageNet to help ease the detection, and the models of three base CNN models, GoogLeNet, ResNet-18, and DenseNet-121, were used for ensembling. The ensemble achieved an accuracy rate of 98.81%. [4] Four different pre-trained deep Convolutional Neural Networks (CNN): AlexNet, ResNet18, DenseNet201, and SqueezeNet for transfer learning for detecting pneumonia images, among which DenseNet201 produced the highest accuracy for both training and testing, achieving the accuracy of 98% and precision of 97% [5]. Marco La Salvia and Gianmarco introduced a new evaluation technique using Lung Ultrasound Scores for detection. ReLU activation was used for resNet18(11M parameters) and resNet50(23M parameters) based residual CNNs, which had already undergone optimization based on the ImageNet dataset. The restNet-based architectures could achieve nearly 98.5% accuracy on test data, which utilized a lot of domain-specific insights about Pneumonia from the field of medicine to help understand the nature of the images [6]. Detection of Pneumonia using convolutional neural networks and deep learning was done by adding a dropout layer for efficient calculation, leading to an accuracy of 97% in 122ms [7]. A Hybrid Deep Convolutional Neural Network Based on Two Parallel Visual Geometry Group Architectures and Machine Learning Classifiers was introduced [8]. A deep learning model for the classification of COVID and Pneumonia using DenseNet-201 was created, considering HIPAA compliance for maintaining patient information privacy, which attained a training accuracy of 98% but a test accuracy of 80%; this paper had a strong focus on the compliance requirements for HIPAA [7]. A Deep Learning-based model for detecting Pneumonia from Chest X-ray images using VGG-16 and Neural Networks was created [9]. Pediatric pneumonia detection was done using a machine learning approach on chest X-rays, and it used the quadratic SVM model to yield an accuracy of 97.58%; the possibility of using SVM showed huge promise since it helps identify the site points which help the medical professional provide concrete diagnosis [10]. Using transfer learning on top of enhanced CNN helped retrieve a high accuracy of 92.4% [11]. A multimodal algorithm called CheXMed was developed for Pneumonia Detection in the Elderly, opening up huge possibilities and achieving an AUC of 0.768 and accuracy of 64.5% [12]. Pneumonia was done using several variants of CNN using Inception V3, VGG-16, ResNet, and DenseNet, attaining a high accuracy of Inception V3 of 90.26% [13]. Most techniques involved convolutional neural networks and their variants, indicating high potential for these models in detecting Pneumonia.

3 Methodology

In this section, we discussed the methodology employed for pneumonia detection using deep learning models, emphasizing various aspects such as dataset collection, data pre-processing, feature selection, hyper-parameter choices, and evaluation metrics, as shown in Fig. 1.

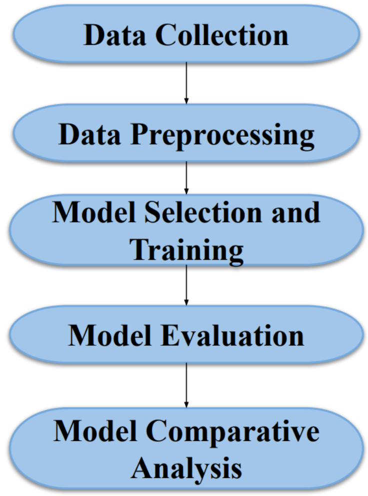

Figure 1: Methodology

3.1 Dataset Collection

The dataset used in the study was from Kaggle and contains 5,916 chest X-ray images. These images were annotated and labeled by expert radiologists to indicate the presence or absence of Pneumonia, as shown in Fig 2 and Fig 3. The dataset is diverse, encompassing a variety of patient demographics, X-ray machine settings, and clinical conditions.

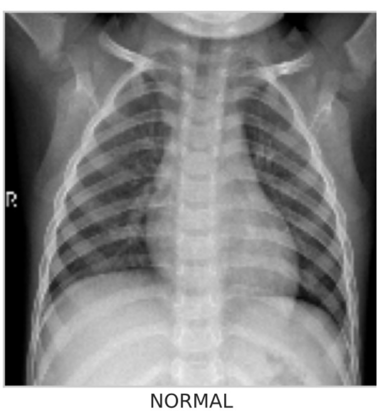

Figure 2: Pneumonia Negative

Figure 3: Pneumonia Positive

3.2 Data Preprocessing

Before training the deep learning models, preprocessing steps were applied to the chest X-ray images. This section outlines the preprocessing techniques employed, such as Gray Filtering, Image Normalization, Image Resizing, and Image Augmentation. These preprocessing steps are crucial for enhancing the model's ability to learn relevant features and patterns from the dataset.

3.2.1 Gray Scaling

The RGB chest X-ray images were converted into grayscale, reducing the input dimensionality and computational complexity while retaining essential information. This process involves a weighted combination of the red, green, and blue channels to create a single-channel representation.

3.2.2 Image Normalization

Normalization was applied to standardize pixel values across all images. This step is crucial for enhancing model convergence and performance by reducing the impact of variations in pixel intensity. Normalization typically involves scaling pixel values to a standardized range of 0 to 1.

3.2.3 Image Resizing

All images were resized to a predefined resolution to ensure consistency in model input dimensions and facilitate efficient computation. Resizing mitigates variations in image dimensions and assists the model in learning relevant features irrespective of the original input size.

3.2.4 Image Augmentation

Augmentation techniques, such as rotation, flipping, and zooming, were employed to artificially increase the diversity of the training dataset. It helps prevent overfitting and enhances the model's generalization ability to unseen data. Augmentation introduces variability in the training set, simulating real-world conditions and improving the model's robustness.

3.3 Model Selection and Architecture

This section outlines the selection and training process of three distinct deep learning architectures – CNN, VGG16, and ResNet50 for pneumonia detection. The chosen models represent a spectrum of complexities, ranging from a basic CNN to more sophisticated architectures like VGG16 and ResNet50. The training process involves optimizing the models using the preprocessed dataset to achieve high accuracy and reliability in pneumonia classification.

3.3.1 CNN

CNNs are foundational deep-learning models for image classification tasks. The CNN architecture chosen for this study comprises convolutional layers for feature extraction and pooling layers for down-sampling. Fully connected layers at the end of the network facilitate the final classification. Fig. 4 illustrates the architecture of the CNN model.

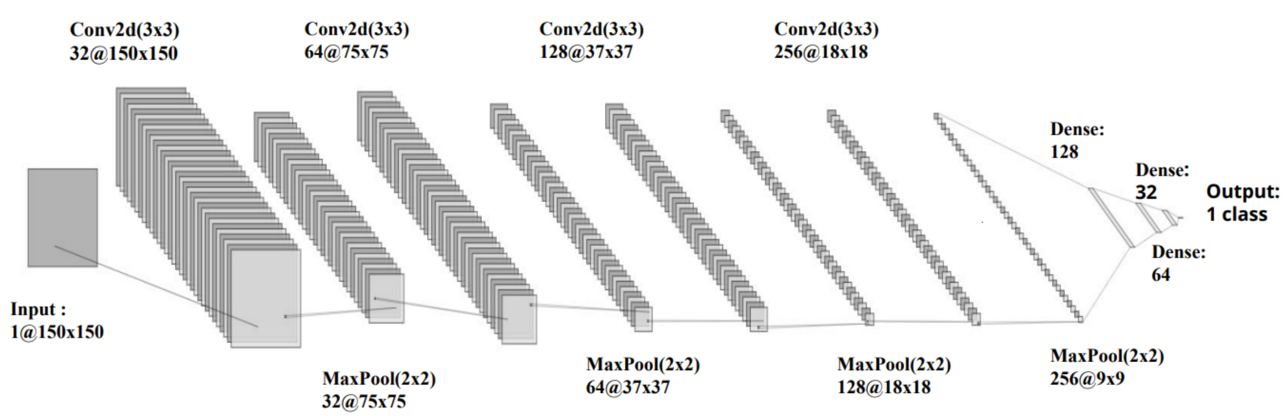

Figure 4: CNN Architecture

- **Input Shape**: The input images have a shape of (150, 150, 3).
- **Convolutional Layer:** The first convolutional layer *'conv2d 23'* with 32 filters and a kernel size of (3, 3) detects basic features in the input image. The second convolutional layer *'conv2d 24'* with 64 filters and a kernel size of (3, 3). As you move deeper into the network, these layers can detect more complex patterns and features. The third convolutional layer *'conv2d 25'* with 128 filters and a kernel size of (3,3). The fourth convolutional layer *'conv2d 26 '* with 256 filters and a kernel size of (3, 3).
- **Pooling Layers:** *'max pooling2d23'*, *'max pooling2d24'*, *'max pooling2d25'*, and *'max pooling2d26'* are max-pooling layers. Max-pooling layers reduce the spatial dimensions of the feature maps and retain the most important information. The pool size is (2, 2), which means the feature maps are downsampled by a factor of 2 in each dimension.
- **Dropout Layers:** It is a regularization technique that helps prevent overfitting by randomly setting a fraction of input units to 0 at each update during training. 'dropout 16' and 'dropout 17' are dropout layers.
- **Flatten Layers:** *'flatten 5'* is a layer that transforms the 2D feature maps into a 1D vector, preparing the data for the fully connected layers.
- **Fully Connected (Dense) Layers:** *'dense 14'* is a fully connected layer with 128 neurons. This layer is responsible for learning complex patterns in the data. *'dense 15'*, *'dense 16'*, and *'dense 17'* are subsequent fully connected layers with 64, 32, and 1 neuron(s), respectively. The final layer typically has a single neuron for binary classification tasks.

3.3.2 VGG16

VGG16 is known for its deep architecture with 16 weight layers, making it suitable for complex image recognition tasks. The VGG16 model in this study leverages multiple convolutional layers with small filter sizes, aiming to capture intricate patterns in chest X-ray images. Fig. 5 illustrates the architecture of the CNN model. We have configured the model with CNN parameters for comparative analysis.

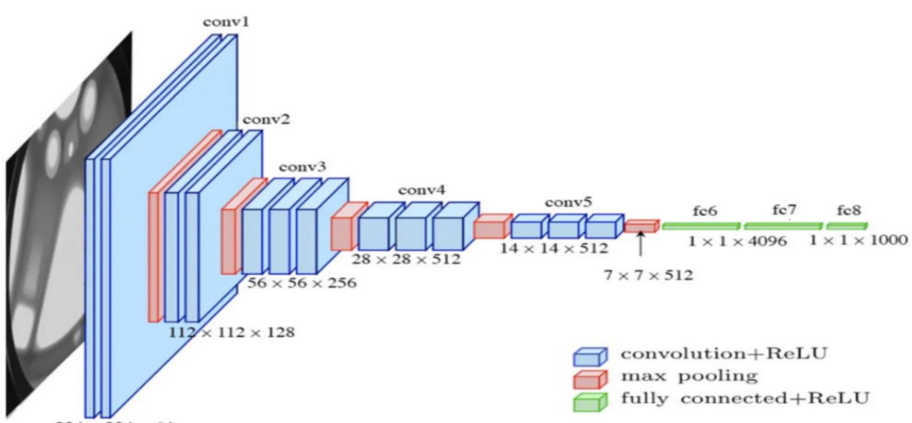

Figure 5: VGG16 Architecture

3.3.3 ResNet50

ResNet50, a variant of the ResNet architecture, is characterized by its residual learning framework, which facilitates the training of extremely deep networks. This design mitigates the vanishing gradient problem, allowing for the effective training of deep neural networks. Fig. 6 illustrates the architecture of the CNN model. We have configured the model with CNN parameters for comparative analysis.

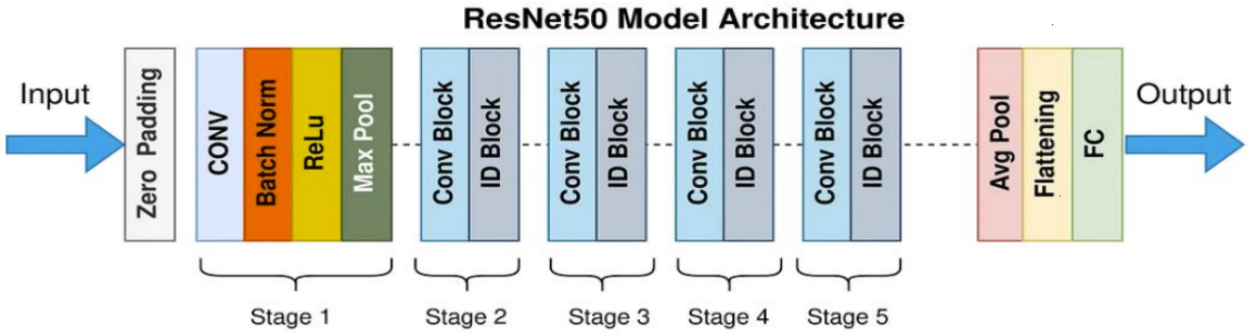

Figure 6: ResNet50 Architecture

3.4 Training and Hyperparamters

The training process involved fine-tuning the selected deep learning models – CNN, VGG16, and ResNet50 for pneumonia detection, utilizing a comprehensive dataset of preprocessed chest X-ray images. The Adam optimizer, a popular optimization algorithm, minimizes the models' loss functions during backpropagation. This adaptive learning rate optimizer adjusts the learning rates for each parameter individually, enhancing convergence speed and overall training efficiency. An Early Stopping function was implemented to prevent overfitting and ensure generalization capability. Early Stopping regularly monitored the model's performance on a separate validation set during training. If the validation performance plateaued or started degrading, the training process was halted, preventing the model from memorizing the training set and improving its ability to generalize to unseen data. The dataset was divided into training, validation, and test sets during training. The models were initialized with pre-trained weights from ImageNet to leverage knowledge gained from diverse image datasets. Fine-tuning involved updating model parameters based on the gradients computed during backpropagation on the pneumonia dataset. Gradient Descent (GD) with the Adam optimizer iteratively adjusted the model's weights to minimize the discrepancy between predicted and actual pneumonia labels. The training process iterated through multiple epochs, each representing a complete pass through the training dataset. Table. 1 shows the configuration of hyperparameters of each model.

Model	Total Parameters	Trainable Parameters	Learning Rate	Number of Epochs	Batch Size
CNN	3,053,121	3,053,121	0.0001	10	16
VGG16	138 million	134.7 million	0.0001	10	32
ResNet50	25,636,712	25,583,592	0.0001	10	32

Model	Loss Function	Optimizer	Number of Hidden Layers	Activation Function	Dropout Rate
CNN	Binary Cross Entropy	Adam	11	ReLU, Sigmoid	0.2
VGG16	Binary Cross Entropy	Adam	16	ReLU, Sigmoid	0.2
ResNet50	Binary Cross Entropy	Adam	50	ReLU, Sigmoid	0.2

Table 1: Hyperparameters Description

3.5 Evaluation Metrics

The performance of the trained CNN, VGG16, and ResNet50 models in pneumonia detection was assessed using a suite of key evaluation metrics, providing a comprehensive understanding of their effectiveness in classification tasks.

3.5.1 Accuracy

Accuracy measures the overall correctness of the model's predictions and is the ratio of correctly predicted instances to the total number of instances. High accuracy indicates the model's proficiency in correctly classifying pneumonia-positive and pneumonia-negative cases.

3.5.2 Precision

Precision, also known as positive predictive value, quantifies the accuracy of positive predictions made by the model. It is calculated as the ratio of true positive predictions to the sum of true positives and false positives. Precision is particularly relevant in medical diagnostics, reflecting the ability to avoid false positive pneumonia identifications.

3.5.3 Recall (Sensitivity)

Recall, or sensitivity, gauges the model's ability to correctly identify all positive instances in the dataset. It is computed as the ratio of true positive predictions to the sum of true positives and false negatives. High recall is crucial in medicine, ensuring that pneumonia cases are not overlooked.

3.5.4 F1 Score

The F1 score represents the harmonic mean of precision and recall, providing a balanced measure of a model's performance. It is calculated as two times the product of precision and recall divided by the sum of precision and recall. The F1 score is particularly useful when there is an imbalance between positive and negative instances, offering a combined assessment of precision and recall.

4 Experimental Results

This section presents the experimental results of the pneumonia detection models, including CNN, VGG16, and ResNet50. The evaluation metrics discussed in the previous section—accuracy, precision, recall, and F1 score reported. Furthermore, visualizations depicting model loss and accuracy across epochs showcase the training dynamics and convergence behavior.

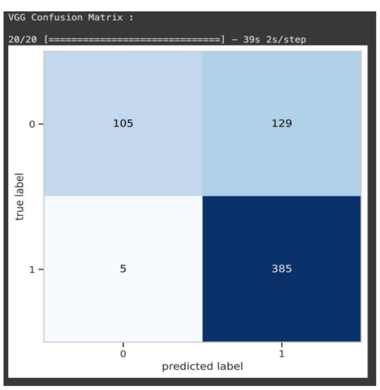

Figure 7: VGG16 Confusion Matrix

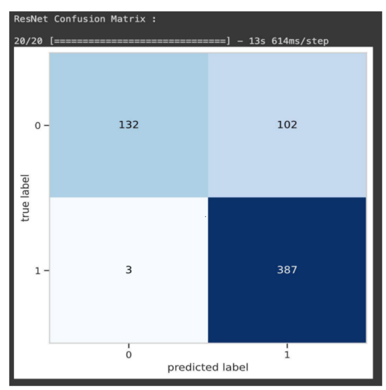

Figure 8: ResNet50 Confusion Matrix

4.1 Confusion Matrix

The confusion matrices provide a detailed breakdown of the model's predictions, categorizing them into true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN). These matrices visually represent the model's performance in pneumonia detection, highlighting areas of strength and potential improvement. The matrices were constructed based on the test set, enabling a granular analysis of the CNN, VGG16, and ResNet50 models' classification outcomes. Figs 7-8 shows confusion matrixes of VGG16 and ResNet50 models respectively. Fig 9 illustrates the comparative analysis of all models using evaluation metrics.

Model	Accuracy	Precision	Recall	F1-Score
CNN	**64.750%**	**63.070%**	**93.785%**	**75.110%**
VGG16	**78.525%**	**74.902%**	**98.717%**	**85.176%**
ResNet50	**83.173%**	**79.141%**	**99.230%**	**88.054%**

Figure 9: Evaluation Metrics

4.2 Model Loss vs. Epochs

Visualization of the training and validation loss over epochs provides insights into the models' convergence and potential overfitting. The plot illustrates the progression of the loss function during training, indicating whether the models effectively learn the underlying patterns in the pneumonia dataset. Figs 10-11 illustrates the model loss with respective to epochs of VGG16 and ResNet50. Fig 12 shows the model loss of CNN, VGG16, and ResNet50 with respective epochs.

Figure 10: VGG16 Model Loss vs Epochs

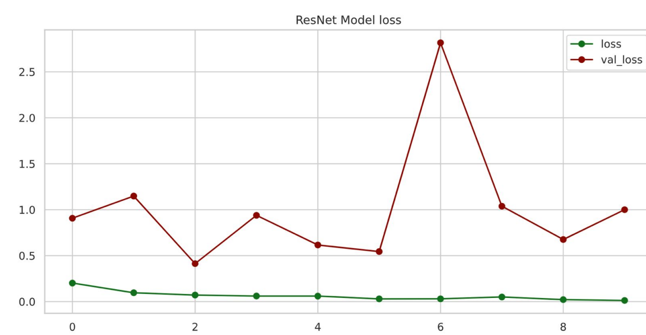

Figure 11: ResNet50 Model Loss vs Epochs

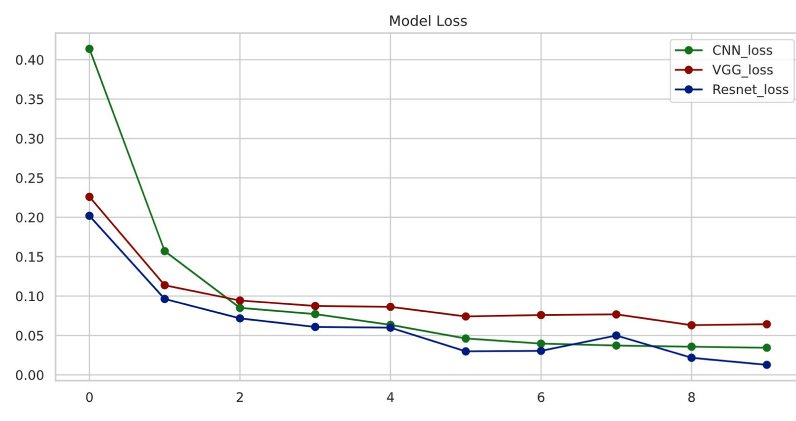

Figure 12: Model Loss vs Epochs for all Models

4.3 Model Accuracy vs. Epochs

Similar to the loss plots, the accuracy vs epochs visualizations depict the models' learning trajectories. Examining how accuracy evolves over training epochs helps assess each model's convergence and generalization capabilities, shedding light on their stability and robustness. Figs 13-14 illustrates the model accuracy with respective to epochs of VGG16 and ResNet50. Fig 15 shows the model accuracy of CNN, VGG16, and ResNet50 with respective epochs.

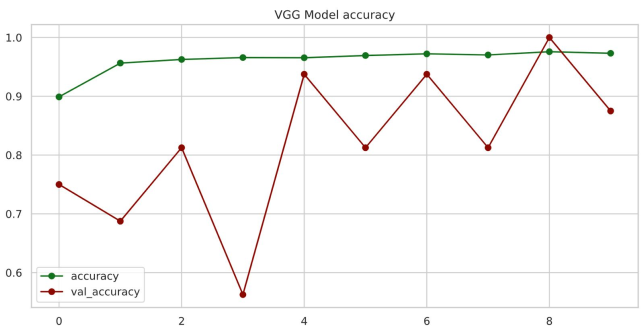

Figure 13: VGG16 Model Accuracy vs Epochs

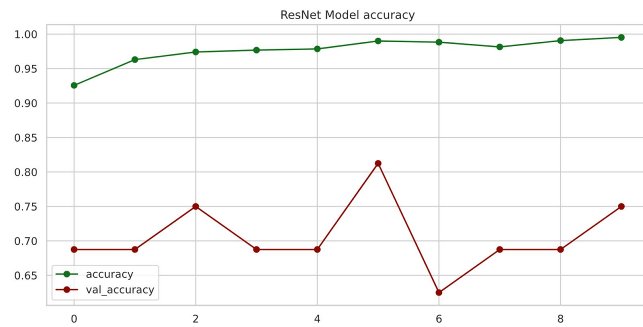

Figure 14: ResNet50 Model Accuracy vs Epochs

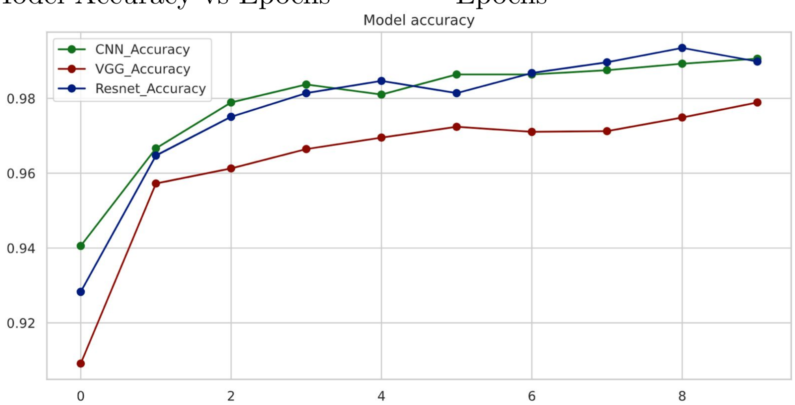

Figure 15: Model Accuracy vs Epochs

4.4 Discussion of Results

The discussion section interprets the experimental results, delving into the implications of the confusion matrices, summarizing key metrics, and providing insights gleaned from the loss and accuracy plots. Patterns, trends, and potential areas of improvement are explored, allowing for a nuanced understanding of the CNN, VGG16, and ResNet50 models' performance in pneumonia detection.

- From the above table and results, we can see that ResNet50 outperforms CNN and VGG16.
- We have observed less training time per epoch for ResNet50 than VGG16 and CNN.
- During testing, the model loss of ResNet50 converges faster than VGG16 and CNN.
- Overall, ResNet50 performs better than other models in detecting Pneumonia.

5 Conclusion and Future Work

In conclusion, this study conducted a comprehensive comparative analysis of three deep learning architectures, CNN, VGG16, and ResNet, to enhance pneumonia detection from chest X-ray images. Incorporating advanced image preprocessing techniques, including normalization, resizing, and image augmentation, contributed to the robustness of the models. Our findings reveal that ResNet50 exhibits the highest overall accuracy. While each architecture displayed strengths in different aspects, the choice of the optimal model may depend on specific clinical priorities, such as minimizing false negatives or positives. As we move forward, this research provides valuable insights into the potential applications of deep learning in medical imaging. Further, we can use GAN instead of data augmentation and the possibility of using autoencoders and transfer learning.

References

[1] L. Račić, T. Popović, S. čakić, and S. Šandi, "Pneumonia detection using deep learning based on convolutional neural network," in *2021 25th International Conference on Information Technology (IT)*, 2021, pp. 1–4.

[2] P. Lakhani and B. Sundaram, "Deep learning at chest radiography: Automated classification of pulmonary tuberculosis by using convolutional neural networks," *Radiology*, vol. 284, no. 2, pp. 574–582, 2017, pMID: 28436741. [Online]. Available: https://doi.org/10.1148/radiol.2017162326

[3] 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. 1–29, 09 2021. [Online]. Available: https://doi.org/10.1371/journal.pone.0256630

[4] T. Rahman, M. E. H. Chowdhury, A. Khandakar, K. R. Islam, K. F. Islam, Z. B. Mahbub, M. A. Kadir, and S. Kashem, "Transfer learning with deep convolutional neural network (cnn) for pneumonia detection using chest x-ray," *Applied Sciences*, vol. 10, no. 9, 2020. [Online]. Available: https://www.mdpi.com/2076-3417/10/9/3233

[5] M. La Salvia, G. Secco, E. Torti, G. Florimbi, L. Guido, P. Lago, F. Salinaro, S. Perlini, and F. Leporati, "Deep learning and lung ultrasound for covid-19 pneumonia detection and severity classification," *Computers in Biology and Medicine*, vol. 136, p. 104742, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0010482521005369

[6] P. Szepesi and L. Szilágyi, "Detection of pneumonia using convolutional neural networks and deep learning," *Biocybernetics and Biomedical Engineering*, vol. 42, no. 3, pp. 1012–1022, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0208521622000742

[7] H. A. Sanghvi, R. H. Patel, A. Agarwal, S. Gupta, V. Sawhney, and A. S. Pandya, "A deep learning approach for classification of covid and pneumonia using densenet-201," *Int J Imaging Syst Technol*, 2022. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9537800/

[8] M. Yaseliani, A. Z. Hamadani, A. I. Maghsoodi, and A. Mosavi, "Pneumonia detection proposing a hybrid deep convolutional neural network based on two parallel visual geometry group architectures and machine learning classifiers," *IEEE Access*, vol. 10, pp. 62 110–62 128, 2022.

[9] S. Sharma and K. Guleria, "A deep learning based model for the detection of pneumonia from chest x-ray images using vgg-16 and neural networks," *Procedia Computer Science*, vol. 218, pp. 357–366, 2023, international Conference on Machine Learning and Data Engineering. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1877050923000182

[10] N. Barakat, M. Awad, and B. A. Abu-Nabah, "A machine learning approach on chest x-rays for pediatric pneumonia detection," *Digital Health*, vol. 9, p. 20552076231180008, 2023.

[11] H. Bhatt and M. Shah, "A convolutional neural network ensemble model for pneumonia detection using chest x-ray images," *Healthcare Analytics*, vol. 3, p. 100176, 2023.

[12] H. Ren, F. Jing, Z. Chen, S. He, J. Zhou, L. Liu, R. Jing, W. Lian, J. Tian, Q. Zhang *et al.*, "Chexmed: a multimodal learning algorithm for pneumonia detection in the elderly," *Information Sciences*, p. 119854, 2023.

[13] D. Pippiri, R. Kode, and K. Ravikiran, "Pneumonia detection using cnn variants," in *AIP Conference Proceedings*, vol. 2754, no. 1. AIP Publishing, 2023.