# Deep Learning Models for COVID-19, Pneumonia, and Normal Chest X-ray Classification

Davide Christian Mancosu Bustos[†], Karim Eugenio Hamdar[‡]

*Abstract*—**The COVID-19 pandemic has put immense pressure on healthcare systems globally, leading to an urgent need for reliable diagnostic tools. Chest X-ray imaging, as a non-invasive and readily available diagnostic method, plays a critical role in identifying respiratory conditions like COVID-19 and pneumonia. Accurate classification of chest X-ray images into COVID-19, pneumonia, and normal categories is essential for effective patient treatment and resource management. This paper evaluates and compares the performance of three deep learning models -ResNet, DenseNet, and a custom CNN- on classifying chest X-rays. Preprocessing techniques such as image resizing, padding, and grayscale conversion were applied, and data augmentation was utilized to improve model generalization. The models were trained using the Kaggle COVID-19 Radiography Database, which contains 4,575 chest X-ray images evenly distributed across the three categories. The results show that ResNet outperformed DenseNet and the custom CNN in terms of accuracy, precision, and recall, achieving an overall accuracy of 93.1%. DenseNet also demonstrated strong performance, while the custom CNN, although computationally efficient, produced lower classification accuracy. Our findings highlight the importance of advanced deep learning architectures and proper preprocessing techniques for improving the accuracy and reliability of medical image classification systems. Further research on larger datasets and more diverse patient populations is needed to generalize these results.**

*Index Terms*—**COVID-19 classification, chest X-ray, deep learning, ResNet, DenseNet, convolutional neural networks (CNNs).**

## I. Introduction

The COVID-19 pandemic has put immense pressure on global healthcare systems, highlighting the need for rapid, accurate diagnostic tools. Chest X-ray imaging, being widely accessible and non-invasive, plays a crucial role in diagnosing respiratory conditions such as COVID-19 and pneumonia. Accurately classifying chest X-rays into COVID-19, pneumonia, and normal categories is essential for effective patient management and resource allocation, especially in resource-constrained settings.

The challenge lies in differentiating between conditions like COVID-19 and pneumonia, which often present similar radiographic features. Several studies, such as those by Cohen et al. (2020) and Wang et al. (2020), have applied deep learning models like ResNet and DenseNet to chest X-ray classification with promising results [1]. However, these studies focused on individual architectures or datasets, without extensively evaluating the effects of image preprocessing techniques like resizing, padding, or data augmentation, leaving gaps in understanding their impact on model performance.

This paper aims to address these gaps by systematically comparing three deep learning models -ResNet, DenseNet, and a custom CNN- for classifying chest X-rays into COVID-19, pneumonia, and normal categories. We analyze the influence of preprocessing techniques (resizing, padding) and explore the role of data augmentation in enhancing model robustness. By leveraging a balanced dataset from the Kaggle COVID-19 Radiography Database, we evaluate these models in terms of accuracy, precision, recall, and computational efficiency, aiming to provide insights for real-world clinical deployment.

The main contributions of this work are:

1) A comprehensive comparison of ResNet, DenseNet, and CNN architectures for chest X-ray classification.
2) An in-depth analysis of preprocessing techniques (resizing, padding) and their impact on model performance.
3) Exploration of data augmentation's effects on model robustness and generalization.
4) Evaluation of model performance with a focus on computational efficiency for practical deployment.

This paper is structured as follows: Section II reviews related work in deep learning for chest X-ray classification. Section III details the dataset and preprocessing methods. Section IV describes the model architectures. Section V presents the experimental setup and results. Section VI concludes the paper.

## II. Related Work

The application of deep learning models to medical imaging, particularly in the classification of chest X-rays for detecting respiratory conditions like COVID-19 and pneumonia, has gained significant attention during the pandemic. Numerous studies have demonstrated the effectiveness of various deep learning architectures, including convolutional neural networks (CNNs) like ResNet, DenseNet, and -designed models.

Cohen et al. (2020) made a substantial contribution by developing a publicly available dataset of COVID-19 chest X-ray and CT images [2], which became foundational for many studies. In their work, they implemented DenseNet with transfer learning to predict the severity of COVID-19 pneumonia, achieving an accuracy of 80.6%. While this study showed the potential of deep learning for COVID-19 classification, it primarily focused on the DenseNet architecture and did

[†]Department of Mathematics, University of Padua, email: {davidechristian.mancosubustos}@studenti.unipd.it

[‡]Department of Mathematics, University of Padua, email: {karimeugenio.hamdar}@studenti.unipd.it

not explore other architectures or preprocessing techniques, leaving room for comparison and further optimization.

Another influential work by Wang et al. (2020) introduced COVID-Net, a deep learning model specifically designed for detecting COVID-19 from chest X-rays [3]. The model was trained on a large dataset of 13,975 images and achieved an accuracy of 93.3%, outperforming standard architectures like ResNet-50 and VGG-19. Wang et al. also focused on making COVID-Net computationally efficient for real-time clinical deployment. However, their research primarily evaluated a single architecture and did not compare the effectiveness of alternative models like DenseNet or ResNet under varying preprocessing conditions.

He et al. (2016) developed the widely-used ResNet architecture, which introduced residual learning techniques to solve the vanishing gradient problem in deep networks [4]. This architecture has since been adapted in numerous medical imaging studies due to its strong performance in feature extraction and classification tasks. Studies like Khalifa et al. (2020) applied ResNet-50 for COVID-19 detection, showing excellent results in terms of accuracy and recall. Despite these promising findings, a comprehensive comparison of ResNet with other state-of-the-art architectures in chest X-ray classification remains scarce.

A major limitation in the existing literature is the lack of systematic evaluation of preprocessing techniques, such as resizing and padding, and how these techniques affect model performance across different architectures. For instance, Loey et al. (2020) explored image augmentation methods like flipping and rotation in a COVID-19 classification task using VGGNet, ResNet, and InceptionV3 [5]. Their results showed that data augmentation could enhance generalization, but the study lacked a detailed exploration of resizing or padding techniques and their impact on different architectures.

In addition, CNNs have been employed in several studies due to their computational simplicity and adaptability to specific tasks. For instance, Apostolopoulos and Besiana (2020) built a CNN for detecting COVID-19 from X-ray images, achieving notable accuracy [6]. However, CNNs often underperform compared to deeper models like ResNet and DenseNet when applied to large, diverse datasets.

While prior studies have shown the effectiveness of individual models and certain preprocessing methods, a comprehensive comparison of multiple architectures, preprocessing techniques, and the effects of data augmentation is still lacking in the context of chest X-ray classification. Our study aims to fill this gap by systematically comparing ResNet, DenseNet, and CNN architectures, evaluating the effects of image resizing, padding, and data augmentation, which have not been thoroughly explored in the literature.

## III. Processing Pipeline

The processing pipeline (Fig. 1) in this study is designed to build an efficient deep learning-based system for classifying chest X-ray images into three categories: COVID-19, pneumonia, and normal. It consists of several key stages that ensure data consistency, model robustness, and accurate classification. First, the dataset is collected from the publicly available

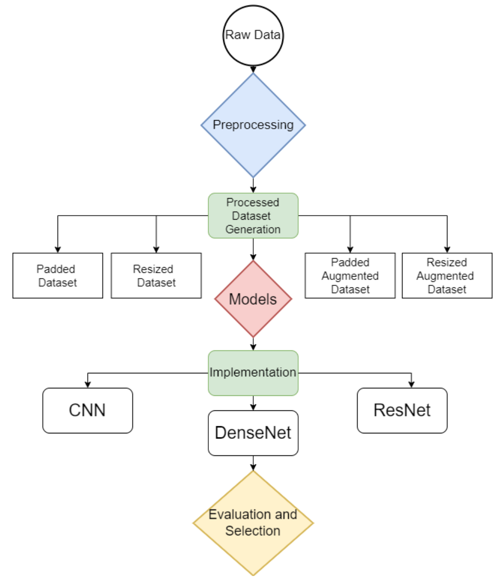



Fig. 1: Processing Pipeline.

Kaggle COVID-19 Radiography Database, containing 4,575 chest X-ray images. These images are evenly distributed across the three categories to ensure balanced learning (Fig. 2).

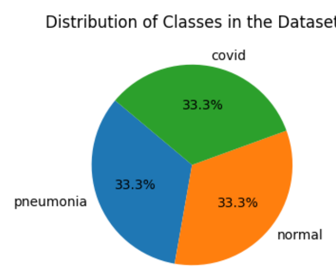



Fig. 2: Classes distribution.

In the preprocessing stage, the images undergo resizing to 224x224 pixels, symmetrical padding for non-square images to preserve the aspect ratio, grayscale conversion for consistency, and normalization of pixel values between 0 and 1 to improve the convergence rate of the models during training.

Three deep learning architectures -ResNet, DenseNet, and a Convolutional Neural Network (CNN)- are implemented and trained from scratch. Each model is assessed under identical preprocessing and augmentation conditions to ensure a fair comparison of their classification performance. Data augmentation techniques, including random rotations, flips, and scaling, are applied to the training data to introduce variability and enhance model generalization. The dataset is splitted into training, validation, and test subsets, with stratified sampling ensuring class balance. The final evaluation measures the accuracy, precision, recall, F1-score, and computational efficiency of each model.

## IV. Signals and Features

In this study, the primary data source consists of 4,575 chest X-ray images, each representing a patient diagnosis of COVID-19, pneumonia, or normal (Fig. 3).

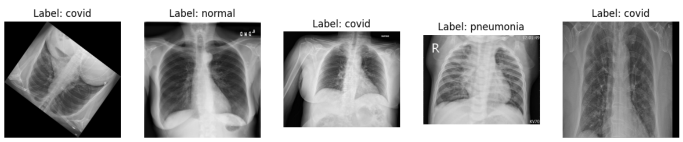



Fig. 3: X-ray scans of Covid, Pneuomonia and Normal healthy lungs.

The images, acquired from the Kaggle COVID-19 Radiography Database, are preprocessed to ensure consistency across the dataset. Preprocessing involves converting images to grayscale, resizing them to 224x224 pixels (Fig. 4), and applying padding (Fig. 5) to retain the aspect ratio. These transformations reduce noise, simplify computational complexity, and ensure uniform input dimensions across all models.

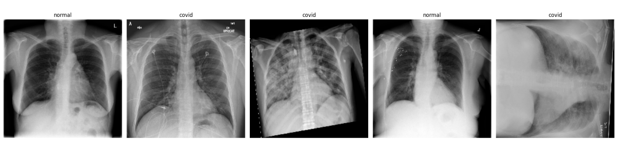


Fig. 4: Resized images.

Fig. 5: Padded images.

The features extracted from the preprocessed images are essential for classification. Since chest X-rays provide high-level structural information, the deep learning models are tasked with learning critical patterns like lung opacity, texture variations, and other visual markers indicative of COVID-19 or pneumonia. The models leverage pixel intensity values (after normalization) as the primary feature set. Additionally, random augmentations such as rotations, flips, and scaling are applied during training to expose the models to a wider variety of image conditions, enhancing their ability to generalize to unseen cases.

The dataset is splitted into training (80%) and test (20%) subsets, with further division of the training data into validation sets. Stratified sampling ensures that the proportions of each class (COVID-19, pneumonia, normal) are preserved throughout the splits, maintaining data integrity and balance in each set.

## V. Learning Framework

In this section, we describe the learning strategy employed to solve the problem of classifying chest X-rays into three categories: COVID-19, pneumonia, and normal. The framework encompasses the design and optimization of the deep learning models, including their architectures, parameters, and the techniques used to improve performance. The key models compared in this study are ResNet, DenseNet, and a Convolutional Neural Network (CNN). The learning process follows a well-defined pipeline, including model training, optimization, and evaluation.

### *A. Model Architectures*

Three deep learning architectures were selected for this task:

- **ResNet**: The ResNet (Residual Network) architecture employs residual connections to mitigate the vanishing gradient problem, which often occurs in deep networks. ResNet-50 (Fig. 6) was chosen due to its balance between depth and computational efficiency. Residual connections allow the gradient to flow easily across the network layers, making it well-suited for handling complex patterns in medical images.

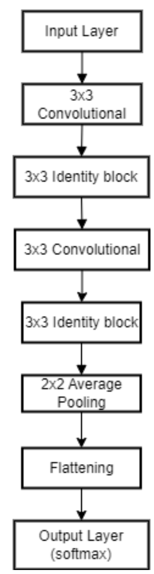



Fig. 6: ResNet.

- **DenseNet**: DenseNet (Densely Connected Convolutional Networks) (Fig. 7) introduces dense connections between layers, where each layer receives input from all preceding layers. This connectivity pattern encourages feature reuse, improving the model's ability to learn fine-grained details in the X-ray images. DenseNet-121 is used, offering a good balance between model complexity and performance.

- **CNN**: A Convolutional Neural Network (CNN) (Fig. 8) is included as a baseline. This simpler model is designed with multiple convolutional layers, followed by pooling and fully connected layers. The CNN serves as a benchmark to assess the performance improvements

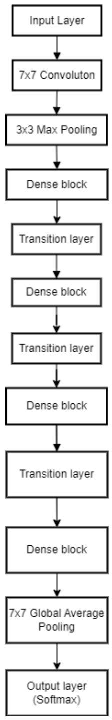



Fig. 7: DenseNet.

offered by more advanced architectures like ResNet and DenseNet.

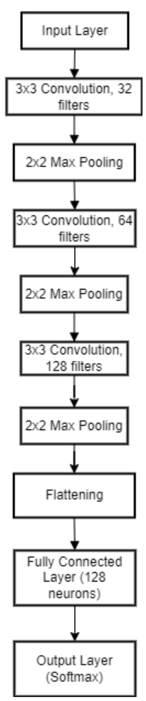



Fig. 8: CNN.

### B. Training Strategy

The models are trained using the **Adam optimizer**, which is known for its adaptive learning rate and efficient handling of sparse gradients. The learning rate is initially set to **1e-5**. The **categorical cross-entropy** loss function is employed, as it is well-suited for multi-class classification problems. Each model is trained for up to **100 epochs**, with **early stopping** based on validation loss to prevent overfitting. This ensures that training halts when the model performance stops improving on the validation set, reducing the risk of overfitting to the training data.

### C. Data Augmentation and Regularization

To increase model robustness and generalization, **data augmentation** is applied during the training phase. Augmentation techniques include:

- **Random Rotations** (up to 30 degrees) to simulate different orientations.
- **Horizontal and Vertical Flips** to reflect possible image variability.
- **Scaling** (between 0.8x and 1.2x) to introduce size variations and further diversify the training data.

Regularization techniques like **dropout** are also employed to mitigate overfitting. Dropout prevents co-adaptation of neurons by randomly setting a fraction of the input units to zero during training, effectively improving the model's ability to generalize to unseen data.

### D. Model Optimization

The models are optimized over a given parameter set, including the learning rate, batch size, and dropout rate. Hyperparameter tuning is carried out to determine the optimal configuration for each model. The **batch size** is set to 32, ensuring efficient use of memory while maintaining good convergence. The **learning rate** is fine-tuned through experimentation, and **dropout rates** between 0.3 and 0.5 are tested to strike a balance between model performance and regularization.

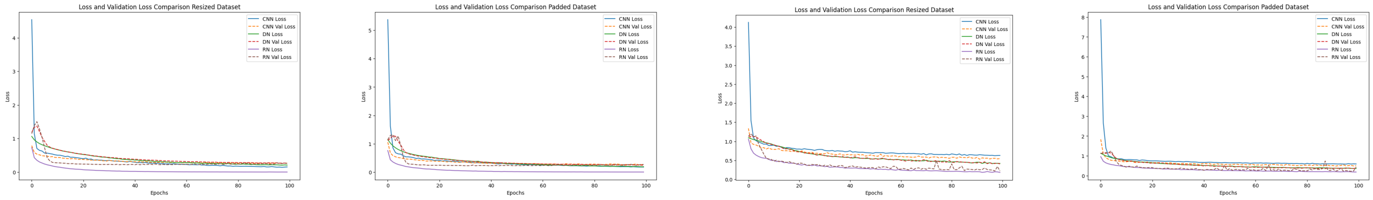
### E. Evaluation Metrics

Once the models are trained, their performance is evaluated using multiple metrics:

- **Accuracy**: The proportion of correctly classified images.
- **Precision**: The ratio of true positive predictions to the total predicted positives.
- **Recall**: The ratio of true positives to the actual positives.
- **F1-Score**: The harmonic mean of precision and recall, providing a balance between the two.
- **AUC (Area Under the Curve)**: This measures the ability of the model to distinguish between the three classes, summarizing the performance across all possible thresholds.

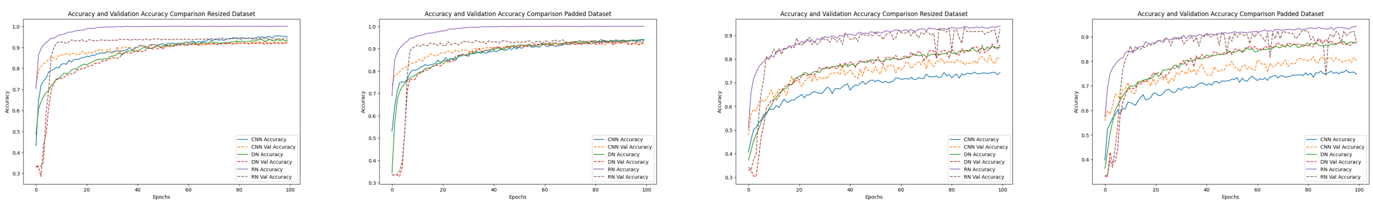
## VI. Results

This section presents a comparative analysis of the performance of three deep learning models -ResNet, DenseNet, and a Convolutional Neural Network (CNN)- for classifying chest X-ray images into three categories: COVID-19, pneumonia,




(a) (b) (c) (d)

Fig. 9: Loss comparison between resized(a), padded(b) dataset without augmentation and resized(c), padded(d) dataset with augmentation




(a) (b) (c) (d)

Fig. 10: Accuracy comparison between resized(a), padded(b) dataset without augmentation and resized(c), padded(d) dataset with augmentation

and normal. The models were evaluated using multiple preprocessing and augmentation conditions. The performance was measured based on precision, recall, accuracy, area under the curve (AUC), and loss.

### A. Performance Metrics

The models' performance metrics are presented in Tables 1 and 2. These tables show the precision, recall, accuracy, AUC, and loss for each model, trained and tested with different preprocessing techniques (resizing, padding) and data augmentation (with and without). The results indicate that data augmentation did not consistently improve accuracy, with some models showing reduced accuracy when augmentation was applied.

| **Model** | **Preprocessing** | **Precision** | **Recall** | **Accuracy** | **AUC** | **Loss** |
|---|---|---|---|---|---|---|
| CNN | Resized | 0.9239 | 0.9191 | 0.9208 | 0.9799 | 0.2709 |
| CNN | Padded | 0.8431 | 0.8168 | 0.8282 | 0.9380 | 0.4951 |
| DenseNet | Resized | 0.9121 | 0.9013 | 0.9047 | 0.9770 | 0.2814 |
| DenseNet | Padded | 0.9229 | 0.9153 | 0.9162 | 0.9805 | 0.2592 |
| ResNet | Resized | 0.9311 | 0.9278 | 0.9310 | 0.9838 | 0.2741 |
| ResNet | Padded | 0.9272 | 0.9272 | 0.9272 | 0.9822 | 0.2831 |

TABLE 1: Performance metrics of models without data augmentation.

| **Model** | **Preprocessing** | **Precision** | **Recall** | **Accuracy** | **AUC** | **Loss** |
|---|---|---|---|---|---|---|
| CNN | Resized | 0.8453 | 0.7649 | 0.8217 | 0.9387 | 0.4912 |
| CNN | Padded | 0.6939 | 0.4826 | 0.5256 | 0.7594 | 0.9002 |
| DenseNet | Resized | 0.8592 | 0.8237 | 0.8411 | 0.9524 | 0.4201 |
| DenseNet | Padded | 0.8970 | 0.8734 | 0.8884 | 0.9623 | 0.3718 |
| ResNet | Resized | 0.8853 | 0.8527 | 0.8727 | 0.9731 | 0.3148 |
| ResNet | Padded | 0.9067 | 0.8951 | 0.8975 | 0.9772 | 0.2978 |

TABLE 2: Performance metrics of models with data augmentation.

### B. Impact of Preprocessing Techniques

The choice of preprocessing techniques significantly influenced model performance. Resizing consistently enhanced model accuracy across all architectures, with ResNet achieving the highest accuracy (93.11%) when resizing was applied without data augmentation. Conversely, padding led to reduced accuracy, particularly in the case of the CNN, where the accuracy dropped to 52.56% when both padding and augmentation were applied.

### C. Effect of Data Augmentation

Data augmentation did not consistently lead to improvements in model performance. Both DenseNet and ResNet demonstrated a decrease in accuracy when augmentation was applied, with DenseNet's accuracy dropping from 91.62% to 88.84% in padded images, and ResNet showing similar patterns. The CNN performed better without augmentation, reaching its highest accuracy of 92.08% when images were resized without any augmentation. These results indicate that data augmentation, while beneficial for increasing the robustness of some models, may introduce variability in performance for others.

### D. Plot Comparison

The results illustrated in Figures 9 and 10 highlight the distinct patterns observed when comparing the models across different preprocessing techniques and augmentation strategies. The loss curves (Fig. 9) show that resizing the dataset tends to lead to faster convergence and lower overall losses compared to padding, regardless of the presence of data augmentation. This is especially evident in models like ResNet and DenseNet, where resized datasets demonstrate smoother learning curves with lower final validation losses. In contrast, CNN shows more variability in its loss values, particularly when padding is applied with augmentation.

For accuracy (Fig. 10), we observe that the accuracy improves more steadily and reaches higher values with resized datasets, particularly when no augmentation is applied. In augmented cases, padding tends to introduce more instability, as seen in the validation accuracy fluctuations, especially for CNN. The clear distinction between resizing and padding effects reinforces the observation that resizing, without augmentation, provides more reliable and consistent performance across all models.

### *E. Computational Efficiency*

The computational efficiency of each model, measured in terms of training time and memory usage, is presented in Tables 3 and 4. While ResNet demonstrated the shortest training time and lowest memory usage in both datasets, DenseNet showed the highest memory usage and longer training times. Interestingly, the CNN, although having a larger model size compared to both ResNet and DenseNet, exhibited competitive training times in the resized dataset but required more time for training in the padded dataset. This trade-off between computational efficiency and model performance is essential for considering real-world deployment scenarios. ResNet, with its balance of speed and memory efficiency, might be ideal for accuracy-critical tasks in resource-constrained environments, while the CNN could be better suited for applications where model size is not a primary concern but computational speed is important.

| **Model** | **Training Time** | **Size** |
|---|---|---|
| ResNet | 8m 38s | 7.52Mb |
| DenseNet | 17m 40s | 82.4Mb |
| CNN | 8m 40s | 127Mb |

TABLE 3: Training time and memory usage with resized dataset.

| **Model** | **Training Time** | **Size** |
|---|---|---|
| ResNet | 7m 45s | 7.52Mb |
| DenseNet | 10m 16s | 82.4Mb |
| CNN | 9m 33s | 127Mb |

TABLE 4: Training time and memory usage with padded dataset.

### *F. Discussion*

The results indicate that ResNet is the most effective model for classifying chest X-ray images, achieving the highest accuracy across both datasets (with and without data augmentation) and preprocessing techniques. DenseNet also performed well but required significantly more computational resources, especially in terms of memory usage. Preprocessing, particularly resizing, consistently led to better model performance across all models, with ResNet achieving the highest accuracy (93.11%) when images were resized without augmentation. Padding generally resulted in lower accuracy, with the CNN being particularly affected, dropping to 52.56% accuracy with both padding and augmentation applied.

Data augmentation did not consistently improve the performance of the models. Both DenseNet and ResNet experienced a decrease in accuracy when augmentation was applied, particularly for padded images. The CNN, although computationally efficient, demonstrated better performance without data augmentation, achieving its highest accuracy (92.08%) with resized images and no augmentation.

These findings suggest that while ResNet is well-suited for applications requiring high accuracy, DenseNet, despite its memory demands, can be effective when computational resources allow. The CNN, although more lightweight, is a viable option for resource-constrained environments, though it comes at the cost of reduced accuracy.

## VII. Concluding Remarks

This study evaluated three deep learning models -ResNet, DenseNet, and a Convolutional Neural Network (CNN)- for the classification of chest X-ray images into COVID-19, pneumonia, and normal categories. The models were trained on the Kaggle COVID-19 Radiography Database, with preprocessing techniques such as resizing, padding, grayscale conversion, and normalization applied to standardize the dataset. In addition, data augmentation was used to improve model robustness and generalization. The results demonstrated that ResNet consistently outperformed the other models, achieving the highest accuracy and F1-score, while DenseNet also delivered strong results. The CNN, while computationally efficient, exhibited lower accuracy.

The findings of this study have several practical implications. First, the superior performance of ResNet and DenseNet suggests that more advanced architectures are well-suited for complex medical imaging tasks, where high accuracy is critical. These models could be deployed in clinical settings to assist radiologists in diagnosing COVID-19 and other respiratory diseases, especially in resource-constrained environments where rapid and accurate diagnostics are crucial. However, the increased computational demand of these models must be considered when deploying them in real-world applications. The CNN, while less accurate, may be a viable option for faster, lightweight applications where computational resources are limited.

Despite the promising results, this study has certain limitations. The dataset used is relatively small, and further experiments on larger and more diverse datasets are needed to improve the generalizability of the models. Additionally, while the models performed well on the given dataset, real-world deployment would require rigorous testing on a broader range of image sources to account for variations in X-ray equipment, patient demographics, and other factors. Future work should also explore more advanced architectures, such as EfficientNet or Transformer-based models, to further improve classification accuracy and computational efficiency.

Finally, one of the key lessons learned from this study is the significant impact of preprocessing techniques on model performance. Simple adjustments, such as resizing images, led to substantial improvements in accuracy, underscoring the importance of data preparation in medical imaging tasks. Additionally, the mixed impact of data augmentation highlights the need for careful tuning of augmentation techniques, as they may improve performance for some models while negatively affecting others. Addressing these challenges required extensive experimentation and parameter tuning, but the insights gained will be valuable for future research in this field.

## References


[1] J. P. Cohen, L. Dao, K. Roth, and et al., "Predicting covid-19 pneumonia severity on chest x-ray with deep learning," *Cureus*, vol. 12, p. e9448, July 2020.

[2] J. P. Cohen, P. Morrison, and L. Dao, "Covid-19 image data collection: Prospective predictions are the future," *arXiv preprint arXiv:2006.11988*, 2020.

[3] L. Wang, Z. Q. Lin, and A. Wong, "Covid-net: A tailored deep convolutional neural network design for detection of covid-19 cases from chest x-ray images," *Scientific Reports*, vol. 10, p. 19549, 2020.

[4] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778, 2016.

[5] M. Loey, F. Smarandache, and N. E. M. Khalifa, "Within the lack of chest covid-19 x-ray dataset: A novel detection model based on gan and deep transfer learning," *Symmetry*, vol. 12, no. 4, p. 651, 2020.

[6] S. Apostolopoulos and T. Bessiana, "Covid-19: Automatic detection from x-ray images utilizing transfer learning with convolutional neural networks," *Physica A: Statistical Mechanics and its Applications*, vol. 564, pp. 124–134, 2020.