COVID-19 Classification from Chest X-ray Images using ResNet50, DenseNet, and EfficientNetB0

TzuHung Cheng
tcheng11@depaul.edu

November 20, 2024

Abstract

Coronavirus Disease 2019 (COVID-19) is a strongly contagious virus spread around the world and depletes the healthcare sources. Chest X-ray (CXR) imaging offers a quick way to detect lung abnormalities caused by COVID-19, but its manual interpretation is time-consuming and prone to error. In this study, we utilized three CNN architectures—ResNet, DenseNet, and EfficientNet to enhance the accuracy of COVID-19 classification from CXR.

1 Introduction

The rapid spread of COVID-19 has overwhelmed the global healthcare system. CXR images represent a potential solution due to their ability to highlight lung abnormalities associated with COVID-19. However, accurately interpreting images is time-consuming and requires expertise, making it difficult to scale for a large population. This research investigates the use of deep learning techniques to classify COVID-19 cases with CXR images.

The primary challenges in this research include the high variability of the infections. The lung abnormalities in CXR evolve over different stages. The classifier might struggle to recognize all possible variations of COVID-19, as it needs to generalize across a diverse range of images. This study aims to address these challenges by utilizing deep learning models to create reliable diagnoses.

2 Related Work

Machine learning methods have been applied in the field of medical imaging for disease detection, with a particular focus on deep learning models such as Convolutional Neural Networks (CNNs) [7]. Several studies have explored the use of CNNs to detect COVID-19 from CXR images [6, 3, 9, 2]. For example, the study [6] utilizes a deep learning approach based on a pre-trained AlexNet model. This model is specifically chosen for its efficiency in image classification tasks, particularly in medical imaging, though their model struggled with generalization on smaller datasets.

One major challenge in COVID-19 detection lies in data variability and class imbalance, where most datasets contain far fewer COVID-19 images than normal or pneumonia cases. Studies [1, 2] state that individual CNNs can overfit or misclassify due to this variability. Recent studies have explored the use of deep learning for medical image classification. Wang et al. (2020) introduced the COVID-Net [11], a custom CNN for detecting COVID-19 in chest X-rays. However, these approaches often lacked systematic comparisons across diverse architectures. In this project, we adopt a comparative approach, focusing on pre-trained models such as ResNet50, DenseNet, and EfficientNetB0.

3 Background

3.1 Convolution Neural Networks (CNNs)

CNNs are widely used in image processing tasks due to their ability to automatically extract hierarchical features from raw pixel data. Layers such as convolution, pooling, and fully connected layers enable CNNs to learn spatial patterns.

3.2 Transfer Learning

Training a model from scratch requires large datasets and computational resources. Transfer learning uses pre-trained models on large datasets (e.g., ImageNet) and fine-tunes them for a specific task, significantly reducing training time and improving accuracy on smaller datasets.

3.3 Evaluation Metrics

In this study, we will use a dataset of Chest X-ray Image(COVID19, PNEUMONIA, and NORMAL) from Kaggle, which includes samples labeled as COVID-19, pneumonia, and normal cases. The dataset will be split into training, validation, and test sets to ensure a fair evaluation of the model's generalization ability. To assess the performance of each model, we will employ a confusion matrix, which provides detailed insights into the precision and recall. This will help us evaluate how well the model distinguishes between COVID-19 and other classes. The results will be presented with the following metrics derived from the confusion matrix:

- **Accuracy:** The percentage of correctly classified samples across all categories.
- **Precision:** Indicates how many of the predicted positive cases are actually positive.
- **Recall:** Reflects how many of the actual positive cases were correctly identified by the model.

3.4 Challenges

X-ray image classification presents several challenges, including low contrast, which makes it difficult to distinguish subtle features, and noise or artifacts caused by variations in image quality, medical devices, or differences in equipment. Additionally, class imbalance, with COVID-19 images often underrepresented compared to other categories, further complicates model performance and necessitates balancing strategies to ensure fair learning across all classes.

4 The Methodology

4.1 Model Selection

In this research, we will utilize three deep learning architectures: ResNet, DenseNet, and EfficientNet. These models have demonstrated state-of-the-art performance in image classification tasks, and stacking them allows us to leverage their complementary strengths.

4.1.1 Resnet

ResNet [4] model architecture allows the training error to be reduced with a deeper network through connection skip. Thanks to the feature of residual connections, the model enables to learn deeper representations without degradation of performance.

4.1.2 DenseNet

DenseNet [5] connections promote feature reuse, making the model highly efficient in extracting intricate patterns. It is ideal for detecting subtle differences in medical imaging data, such as identifying early-stage COVID-19 features.

4.1.3 EfficientNet

EfficientNet [10] is a convolutional neural network (CNN) architecture known for its impressive balance of accuracy and efficiency. It uses a compound coefficient to uniformly scale network width, depth, and resolution in a principled way. This method enables the creation of models that are both highly accurate and computationally efficient. This makes it a compelling choice for image classification tasks with limited resources, such as chest X-ray analysis

4.2 Model Architecture

Three state-of-the-art architectures, ResNet50, DenseNet, and EfficientNetB0, were chosen for their proven performance in image classification tasks. These pre-trained models were loaded with weights initialized from the ImageNet dataset to leverage prior feature extraction capabilities.

To adapt these architectures for our task, we appended custom layers after each base model:

- GlobalAveragePooling2D Layer: This layer condenses feature maps generated by the base model into a compact representation, reducing spatial dimensions while retaining essential information.
- Dense Layer with ReLU Activation: Adds a fully connected layer to perform high-level feature learning, activated by the ReLU function to introduce non-linearity.
- Dropout Layer: This layer was employed to reduce overfitting by randomly deactivating neurons during training.
- Output Dense Layer with Softmax Activation: A final fully connected layer with three output neurons (corresponding to the COVID-19, pneumonia, and healthy classes) and softmax activation was added to produce class probabilities.

4.3 Training Process

Each pre-trained model (ResNet50, DenseNet, EfficientNetB0) was loaded with ImageNet weights. The base model was frozen during initial training, allowing only the custom layers to be optimized. After training, evaluation metrics such as accuracy, precision, recall, F1-score, and inference time were computed on the test set to compare models.

4.3.1 Pre-Trained Model

A pre-trained base model is a neural network that has been previously trained on a large, diverse dataset, such as ImageNet, to learn general features that can be applied to a wide range of tasks. These models, including ResNet, DenseNet, and EfficientNet, act as feature extractors, capturing low- to high-level patterns such as edges, textures, and object structures.

Using a pre-trained model as a starting point provides several advantages. It reduces the need for extensive labeled data, as the model already possesses a strong foundation of knowledge. This approach also accelerates training and improves performance, especially in cases where datasets are small or imbalanced.

Transfer learning adapts these models to new tasks by adding custom layers for task-specific outputs and fine-tuning the weights of certain layers. Pre-trained models are particularly effective in fields like medical imaging and natural language processing, where acquiring large, annotated datasets can be challenging.

4.3.2 Optimizer

The Adam optimizer (Adaptive Moment Estimation) [8] is a popular algorithm for training deep learning models due to its efficiency and versatility. It combines the strengths of two other optimization methods: Momentum and RMSProp. Adam computes individual adaptive learning rates for each parameter by maintaining an exponentially decaying average of past gradients (momentum) and their squared values.

This dual adaptation allows Adam to accelerate convergence on sparse gradients while mitigating oscillations in noisy settings. Unlike traditional gradient descent, Adam automatically adjusts the learning rate during training, reducing the need for manual tuning. Its computational efficiency and ability to handle large datasets or models with high-dimensional data make it suitable for a wide range of applications.

4.3.3 Callbacks

The EarlyStopping [12] callback is a widely used technique in deep learning to prevent overfitting and reduce unnecessary computation during model training. It monitors a specified performance metric, such as validation loss or accuracy, and halts training if the metric stops improving for a predefined number of epochs (patience).

By stopping training early, this callback ensures that the model does not continue to learn patterns from noise in the data, which can degrade its generalization performance on unseen datasets. EarlyStopping also saves computational resources by terminating training once optimal performance is achieved.

5 Numerical Experiments

5.1 Experimental Setup

All models were trained on a dataset divided into training (80%) and validation (20%)subsets. Each model was trained for a maximum of 30 epochs using the Adam optimizer with an initial learning rate of 0.001. EarlyStopping and ReduceLROnPlateau callbacks were applied to prevent overfitting. The models were fine-tuned with ImageNet-pretrained weights, and additional custom layers were appended for the classification task.

5.2 Training Performance

The training accuracy of ResNet50, DenseNet, and EfficientNetB0 over epochs is visualized in Figure 1, Figure 2, and Figure 3. Early stopping triggered at different epochs for each model reflects their learning dynamics.

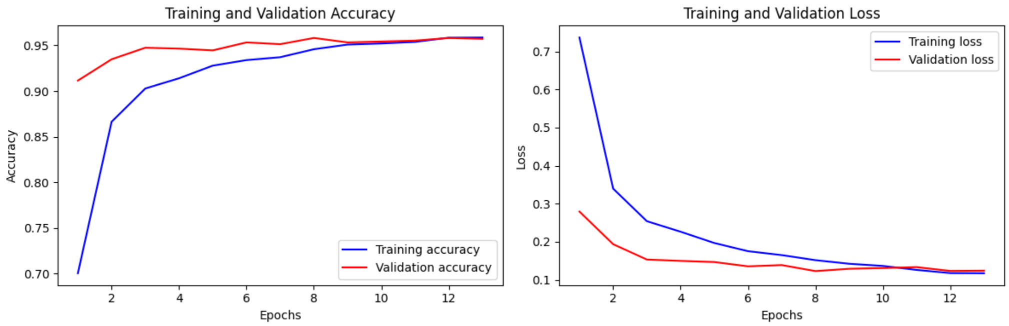

Figure 1: **ResNet**'s Training Accuracy and Validation Accuracy Over Epochs

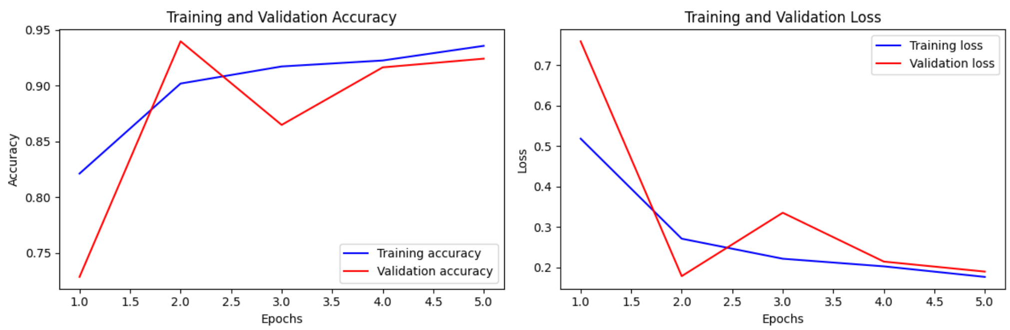

Figure 2: **DenseNet**'s Training Accuracy and Validation Accuracy Over Epochs

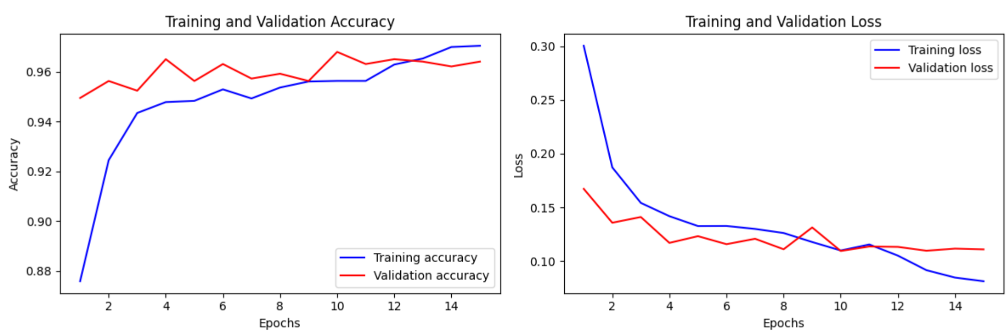

Figure 3: **EfficientNet**'s Training Accuracy and Validation Accuracy Over Epochs

5.3 Testing Performance

Each model's performance was evaluated on the test set using precision, recall, and F1-score metrics. The classification reports are detailed in Table 1.

5.4 Observations

In the training process, ResNet50 and EfficientNetB0 demonstrate steady improvement, achieving high accuracy before early stopping (epochs 13 and 15, respectively). DenseNet, however, stops at epoch 5, reflecting a faster but potentially suboptimal convergence. For the classification report, ResNet50 and EfficientNetB0 outperform DenseNet

Class	ResNet50		DenseNet		EfficientNetB0	
	Precision	Recall	Precision	Recall	Precision	Recall
COVID19	1.00	0.93	0.99	0.67	0.98	0.97
NORMAL	0.91	0.93	0.52	0.97	0.94	0.90
PNEUMONIA	0.97	0.97	0.98	0.71	0.96	0.98
Overall	**0.96**	**0.96**	**0.87**	**0.77**	**0.96**	**0.96**
Accuracy	**96%**		**77%**		**96%**	

Table 1: Classification Report for ResNet50, DenseNet, and EfficientNetB0 on Test Data

with an overall accuracy of 96%. Both models exhibit balanced precision and recall across all classes. DenseNet struggles with the NORMAL class (precision 0.52) and COVID-19 (recall 0.67), leading to lower overall performance. Imbalanced class representation in the dataset (fewer COVID-19 samples) might have negatively impacted DenseNet's ability to generalize. In conclusion, EfficientNetB0 demonstrates the most consistent balance across classes, while DenseNet's early stopping may have hindered its ability to learn robust features.

6 Future Work

First, experimenting with additional deep learning architectures can provide insights into the strengths and weaknesses of various models for this specific task. Models such as InceptionV3, MobileNet, or Vision Transformers (ViTs) could be evaluated for their ability to learn from the complex features of X-ray images. These models have demonstrated promising results in other medical imaging domains and may offer competitive performance in this task as well.

Second, building an ensemble model by combining the outputs of multiple base models is a promising approach. Ensembles leverage the diversity of predictions from different architectures to improve overall accuracy and robustness. For example, predictions from ResNet50, DenseNet, and EfficientNetB0 can be combined using techniques such as averaging probabilities, majority voting, or a meta-classifier trained on the predictions of the individual models. This approach could reduce model biases and improve generalization to new datasets.

Third, a comparative study of data preprocessing techniques would provide a deeper understanding of their impact on model performance. Preprocessing methods such as normalization, augmentation, and contrast enhancement play a critical role in medical imaging tasks. Comparing the performance of models trained with preprocessing against those trained on raw images could highlight the significance of these techniques in mitigating noise, handling class imbalances, and improving feature extraction. The analysis could also explore whether certain preprocessing steps are more beneficial for specific architectures or datasets.

In addition, analyzing misclassified cases could reveal the limitations of the current models and preprocessing techniques. For example, identifying patterns in images where the model struggles, such as low-contrast regions or images with artifacts, could guide future improvements in both the dataset and the model design.

Lastly, expanding the dataset to include more samples and diverse sources can help improve the model's generalizability. This could involve collecting X-ray images from different hospitals, regions, or imaging equipment to ensure the model performs well across varied real-world conditions.

7 Conclusion

This study demonstrates the efficacy of multiple CNNs for COVID-19 detection from chest X-ray images. EfficientNetB0 emerged as the most promising model due to its high accuracy and computational efficiency. The insights from these experiments would not only enhance the current system but also contribute valuable knowledge to the broader field of medical imaging and artificial intelligence. By exploring deeper in the field of computer vision, our future work aims to develop a more accurate, robust, and generalizable system for COVID-19 detection and potentially other medical classification tasks.

References

[1] Parnian Afshar, Shahin Heidarian, Farnoosh Naderkhani, Anastasia Oikonomou, Konstantinos N Plataniotis, and Arash Mohammadi. Covid-caps: A capsule network-based framework for identification of covid-19 cases from x-ray images. *Pattern Recognition Letters*, 138:638–643, 2020.

[2] Priya Aggarwal, Narendra Kumar Mishra, Binish Fatimah, Pushpendra Singh, Anubha Gupta, and Shiv Dutt Joshi. Covid-19 image classification using deep learning: Advances, challenges and opportunities. *Computers in Biology and Medicine*, 144:105350, 2022.

[3] Goram Mufarah M Alshmrani, Qiang Ni, Richard Jiang, Haris Pervaiz, and Nada M Elshennawy. A deep learning architecture for multi-class lung diseases classification using chest x-ray (cxr) images. *Alexandria Engineering Journal*, 64:923–935, 2023.

[4] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[5] Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q Weinberger. Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4700–4708, 2017.

[6] Abdullahi Umar Ibrahim, Mehmet Ozsoz, Sertan Serte, Fadi Al-Turjman, and Polycarp Shizawaliyi Yakoi. Pneumonia classification using deep learning from chest x-ray images during covid-19. *Cognitive computation*, 16(4):1589–1601, 2024.

[7] Huiyan Jiang, Zhaoshuo Diao, Tianyu Shi, Yang Zhou, Feiyu Wang, Wenrui Hu, Xiaolin Zhu, Shijie Luo, Guoyu Tong, and Yu-Dong Yao. A review of deep learning-based multiple-lesion recognition from medical images: classification, detection and segmentation. *Computers in Biology and Medicine*, 157:106726, 2023.

[8] Diederik P Kingma. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[9] FM Javed Mehedi Shamrat, Sami Azam, Asif Karim, Kawsar Ahmed, Francis M Bui, and Friso De Boer. High-precision multiclass classification of lung disease through customized mobilenetv2 from chest x-ray images. *Computers in Biology and Medicine*, 155:106646, 2023.

[10] Mingxing Tan and Quoc Le. Efficientnet: Rethinking model scaling for convolutional neural networks. In *International conference on machine learning*, pages 6105–6114. PMLR, 2019.

[11] Linda Wang, Zhong Qiu Lin, and Alexander Wong. Covid-net: A tailored deep convolutional neural network design for detection of covid-19 cases from chest x-ray images. *Scientific reports*, 10(1):19549, 2020.

[12] Yuan Yao, Lorenzo Rosasco, and Andrea Caponnetto. On early stopping in gradient descent learning. *Constructive Approximation*, 26(2):289–315, 2007.

A Appendix

Acronyms

CNNs Convolutional Neural Networks

COVID-19 Coronavirus Disease 2019

CXR Chest X-ray