 Detecting COVID-19 and Other Respiratory Conditions Using Deep Learning Image Classification on Chest X-Ray Scans

RONI SHTERNBERG (034466862), ADIR LEIBOVICH (308247790)

[Link to notebook](https://colab.research.google.com/drive/1ySr53pCxvgGk4MvKppDTIRwgBTO2H6A0?usp=sharing)

Abstract

This study leverages deep learning models to classify Chest X-ray (CXR) scans into four categories - COVID-19 positive, Normal, Lung Opacity (Non-COVID lung infection), and Viral Pneumonia. The ResNet-18, ResNet-50, VGG19 and DenseNet-121 models, pre-trained on the ImageNet dataset (with millions of images), are employed for this task. A systematic methodology involving image preprocessing, data augmentation, model training, hyperparameter tuning, and model evaluation is used. The study uses the "COVID-19 Radiography Database" from Kaggle, containing a total of 21,165 cases. Results indicate that the ResNet-18 and DenseNet-121 models outperform the other two models with an average F1 score of 0.76 and 0.74, respectively, despite some challenges in accurately classifying 'COVID' and 'Normal' images. Future work may involve data augmentation, hyperparameter tuning, use of ensemble methods, gathering additional data, adjustments to the model architecture, or trying different architectures such as ResNet-101 and EfficientNet to improve performance.

1. Introduction:

Deep learning has shown great promise in various domains, including the field of medical imaging. With the onset of the COVID-19 pandemic, the need for rapid and accurate diagnostic tools has become paramount. In this regard, we propose the application of deep learning models for the classification of Chest X-Ray (CXR) scans into four categories - COVID-19 positive, Lung Opacity (Non-COVID lung infection), Viral Pneumonia, and Normal. For this project, we leverage the power of the ResNet-18, ResNet-50, VGG19 and DenseNet-121 models.

1.1 ResNet-18

ResNet, or Residual Network, is a deep learning model that introduced the concept of skip connections or shortcuts to prevent the problem of vanishing gradient during training. ResNet-18 is a version of this network that contains 18 layers, making it one of the smaller and computationally efficient variants of ResNet. This is particularly useful when dealing with large datasets, as in the case of our project, where we process thousands of images. Despite its smaller size, ResNet-18 can still capture complex patterns in the data and perform well on various tasks, which makes it a suitable choice for initial exploratory analysis in this project.

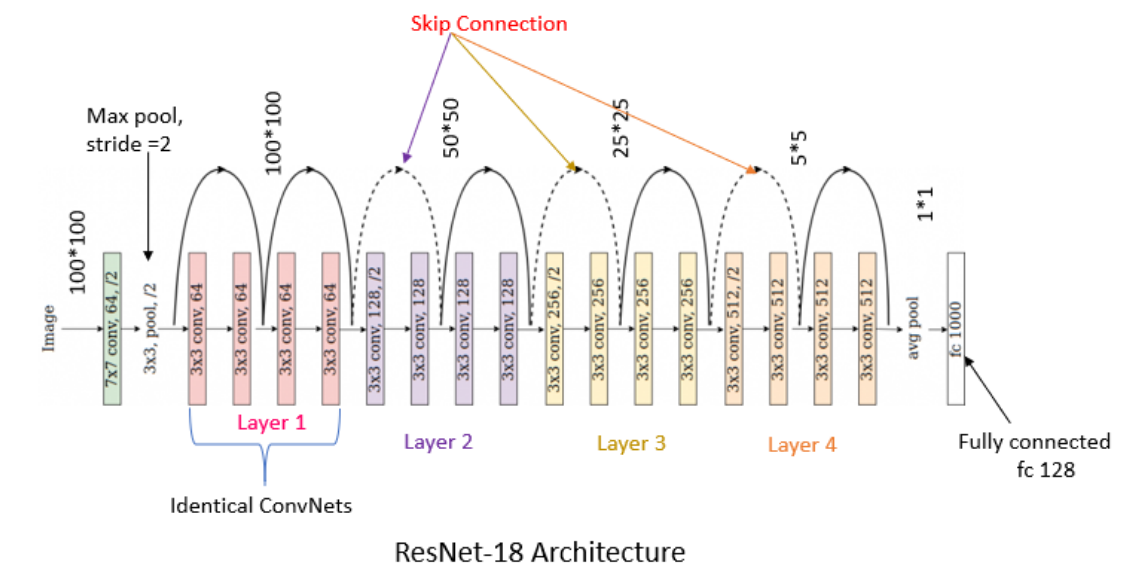


Figure 1 ResNet-18 Architecture [pluralsight 2020]

1.2 ResNet-50

ResNet-50 is a larger variant of ResNet, with 50 layers, offering a more complex model that can capture even more intricate patterns in the data. While it requires more computational power than ResNet-18, it also tends to yield better results, especially in cases where the data is complex and varied. In the context of our project, this model can be beneficial in capturing the subtleties in CXR scans, which can be crucial in distinguishing between different conditions.

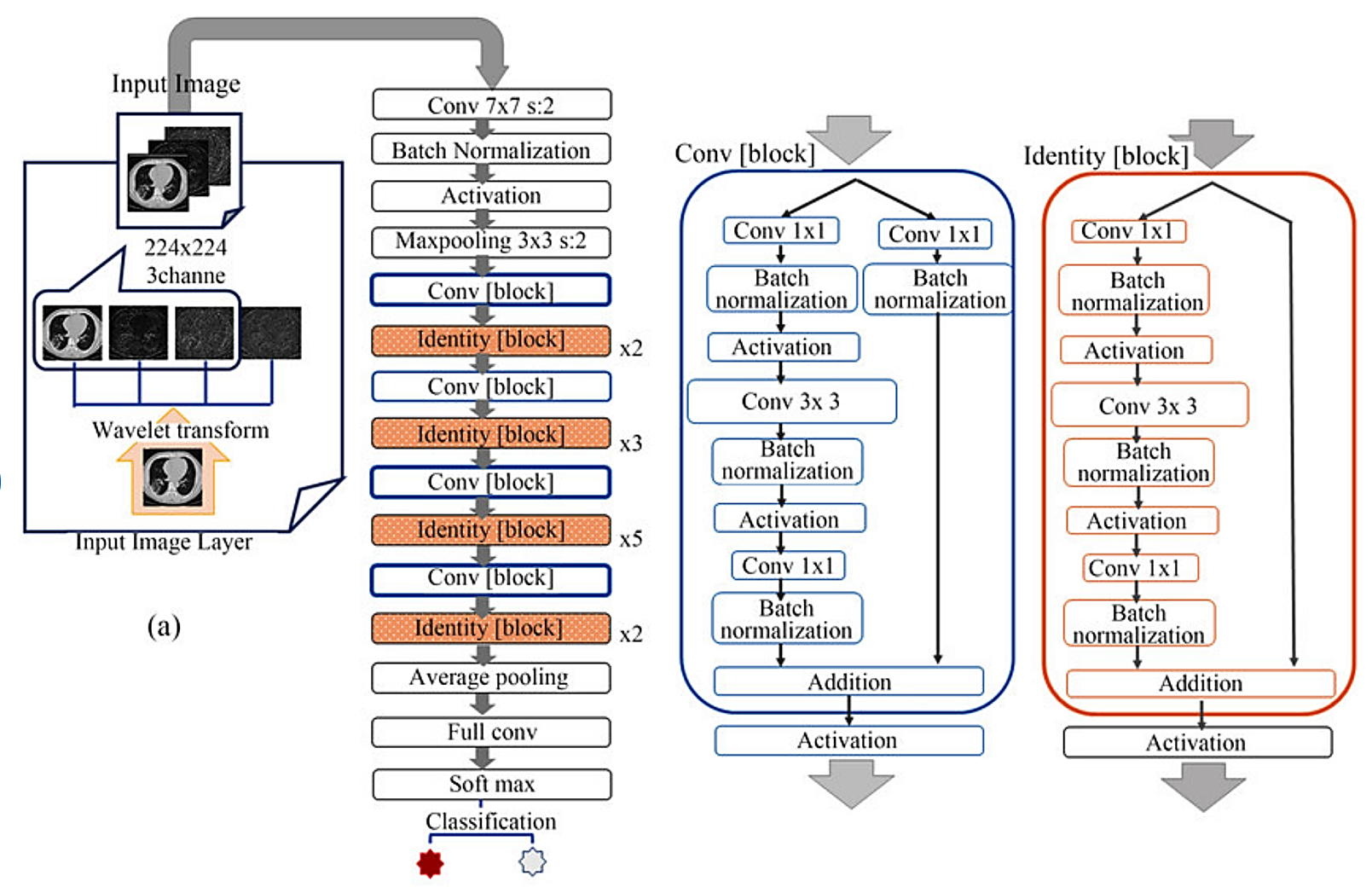


Figure ResNet-50 Architecture [ResearchGate]

1.3 VGG19

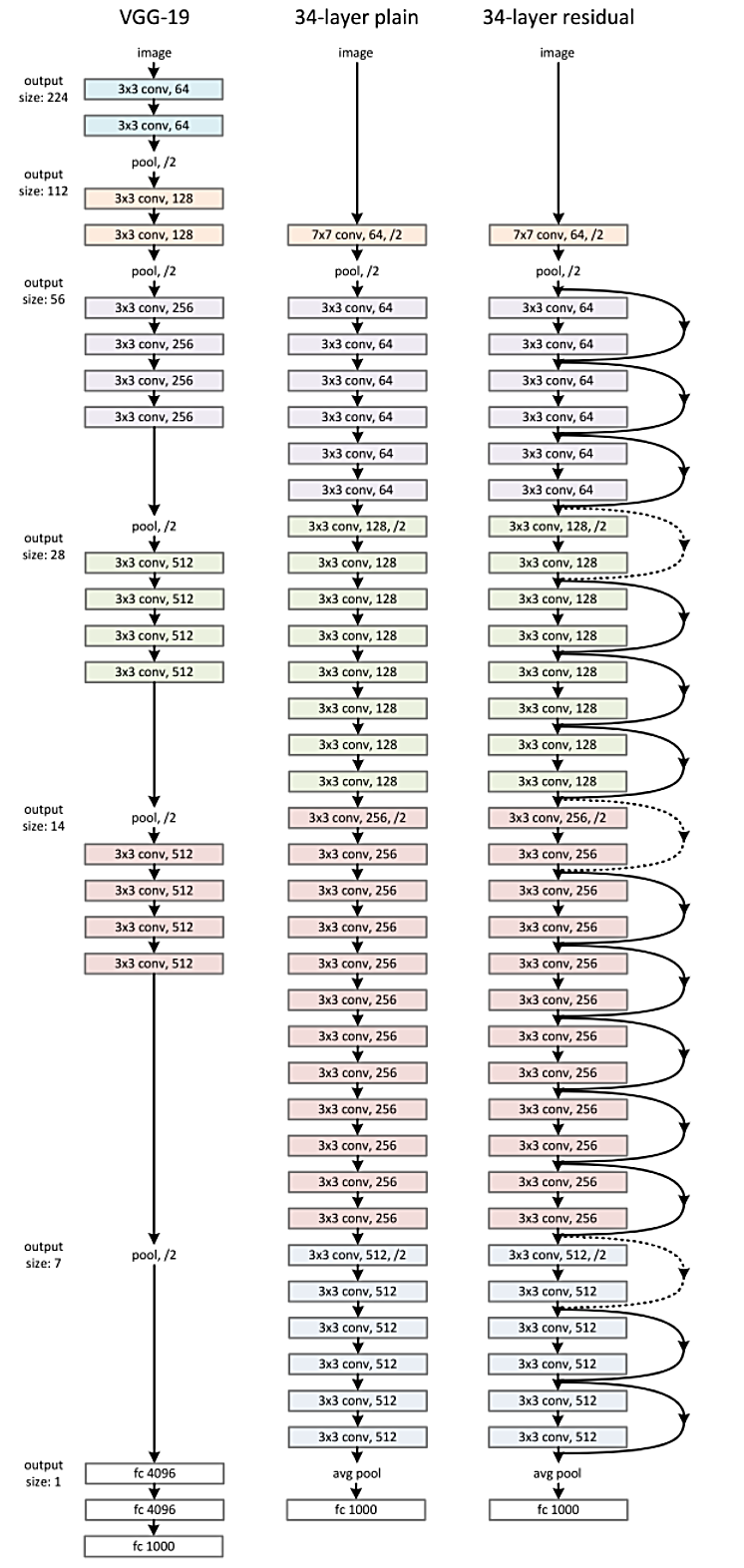


Figure VGG19 compared to 34-layer-plain and ResNet-34 [Kaggle]

VGG19 is a deep learning model that is known for its simplicity and effectiveness. It comprises 19 layers and is characterized by its use of small (3x3) convolution filters, which allows it to capture local features in the image more effectively. It has been widely used in the field of image classification and has demonstrated excellent performance. Despite its increased computational demand compared to the ResNet models, it was included in this project to provide a comparative study between architectures that adopt different design principles.

By employing these three models, we aim to explore their respective strengths in dealing with the classification task at hand and potentially draw conclusions on which architectural principles could be more effective in diagnosing conditions from CXR scans. It is important to note that these models serve as tools for research and analysis and should not be relied upon for clinical diagnosis without further validation and regulatory approval.

1.4 DenseNet-121

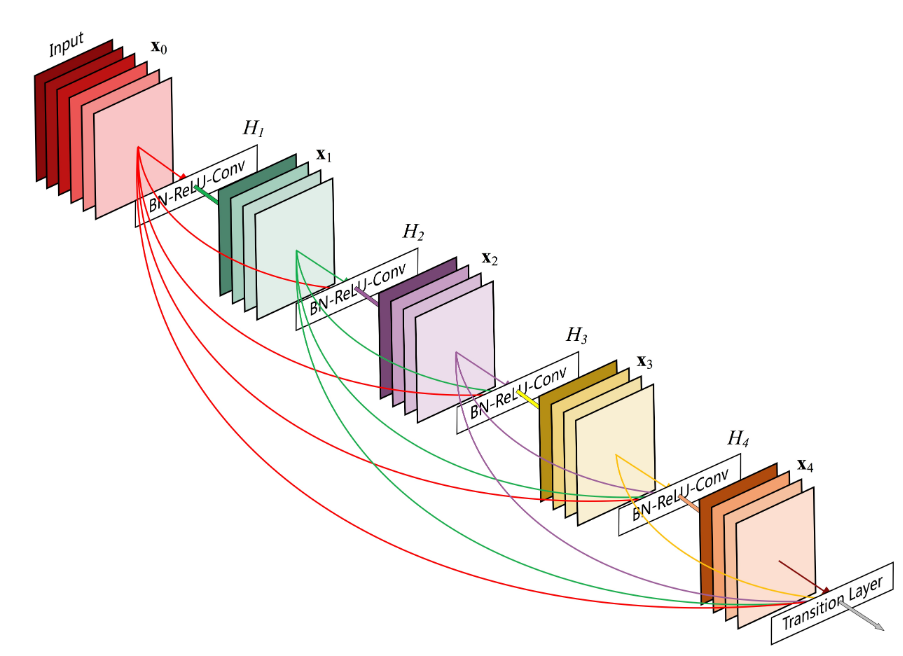


Figure : A Dense block with 5 layers and growth rate 4 [Denesly Connected Convolutional Networks, Arxiv, 2018]

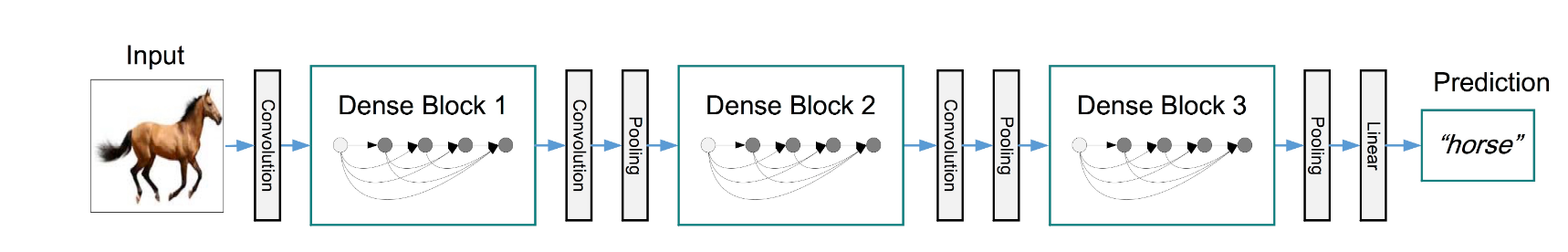


Figure : A deep DenseNet with three dense blocks [Denesly Connected Convolutional Networks, Arxiv, 2018]

Dense Convolutional Network (DenseNet) connects each layer to every other layer in a feed-forward fashion. Whereas traditional convolutional networks with L layers have L connections - one between each layer and its subsequent layer - our network has L(L+1)/2 direct connections. For each layer, the feature-maps of all preceding layers are used as inputs, and its own feature-maps are used as inputs into all subsequent layers. DenseNets have several compelling advantages: they alleviate the vanishing-gradient problem, strengthen feature propagation, encourage feature reuse, and substantially reduce the number of parameters. This connectivity pattern yields state-of-the-art accuracies on CIFAR10/100 (with or without data augmentation) and SVHN (Street View House Numbers dataset). On the large scale ILSVRC 2012 (ImageNet) dataset, DenseNet achieves a similar accuracy as ResNet, but using less than half the number of parameters and roughly half the number of FLOPs.

1.5 Data Source

The dataset: ["COVID-19 Radiography Database” taken from Kaggle](https://www.kaggle.com/datasets/tawsifurrahman/covid19-radiography-database)

We have 21,165 total cases, which are categorized in four classes:

1. 3616 COVID-19 positive cases

2. 10,192 Normal cases

3. 6012 Lung Opacity (Non-COVID lung infection)

4. 1345 Viral Pneumonia cases

Image Formats: All the images are in Portable Network Graphics (PNG) file of each image is 256\*256 pixels.format and the resolution

Citations using during the project:

Article (1): ["Can AI Help in Screening Viral and COVID-19 Pneumonia?"](https://arxiv.org/abs/2003.13145)

Article (2): "[Exploring the effect of image enhancement techniques on COVID-19 detection using chest X-ray images](https://pubmed.ncbi.nlm.nih.gov/33799220/)"

1. METHODOLGY

The methodology adopted in this project follows a systematic approach involving several stages such as image preprocessing, data augmentation, model selection and training, hyperparameter tuning, and model evaluation. Here is a detailed breakdown of the steps:

Figure 6 Steps in BOX Chart

2.1 Data Collection

The data used in this study is the "COVID-19 Radiography Database" available on Kaggle. The database includes a total of 21,165 X-ray images of the anterior-posterior view. These images are categorized into four classes: COVID-19 cases, Lung Opacity cases, Viral Pneumonia cases, and Normal cases. The images are in PNG format and have a resolution of 256\*256 pixels.

2.2 Data Preprocessing

In the preprocessing step, the images are first normalized to have pixel values ranging from 0 to 1. This is a common practice for image classification tasks as it scales down the input and hence, makes the optimization problem more manageable for the model. Next, the images are resized to a standard size of 224x224 pixels. The rationale behind this size is that many pre-trained models used in transfer learning were originally trained on images of this size.

2.3 Data Augmentation

Data augmentation is used to artificially expand the training dataset with new, altered versions of the images. In this study, random horizontal flipping is applied to the training images. This technique not only helps to increase the amount of training data but also allows the model to generalize better by learning to handle different orientations of the images.

2.4 Model Selection and Training

Three deep learning architectures are used in this project: ResNet-18, ResNet-50, and VGG19. These models are pre-trained on the ImageNet dataset, which allows us to leverage the hierarchical feature representations learned from this large-scale dataset.

The training process involves feeding the images to the model and adjusting the model parameters to minimize a loss function. The Adam optimizer is used with a learning rate set adaptively based on a learning rate schedule. This method is often more effective than a fixed learning rate as it can automatically adjust the learning rate during training based on the progress of the model.

2.5 Hyperparameter Tuning (manual)

Hyperparameter tuning is carried out to find the optimal hyperparameters for the model. This includes adjusting the learning rate, batch size, and the number of epochs for training (this should be done automatically using Grid Search or Optuna in future work).

2.6 Model Evaluation

After the model is trained, it is evaluated on the testing dataset. The testing dataset is a set of images that the model has not seen during training, which allows us to assess how well the model generalizes to unseen data.

The performance of the model is evaluated using several metrics including accuracy, precision, recall, and F1 score. Additionally, a confusion matrix is generated to visualize the performance of the model in terms of true positive, true negative, false positive, and false negative predictions.

1. Results

3.1 ResNet-18:

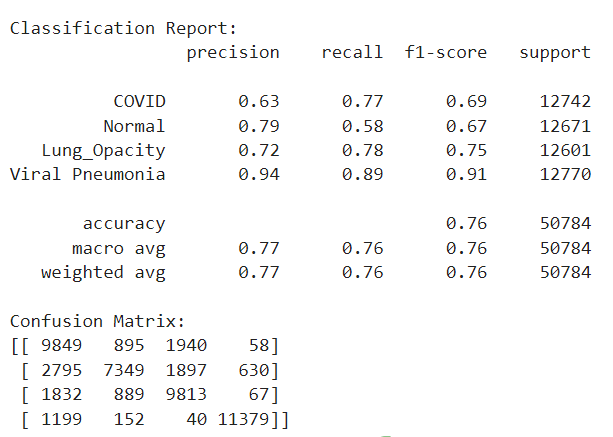


Figure 7 ResNet-18 Classification Metrics

Our project aimed at classifying different types of lung conditions from chest X-ray images using the ResNet-18 model. The model was tasked to differentiate between four classes: COVID-19, Normal (healthy), Lung Opacity, and Viral Pneumonia.

Our results show that the model achieved an overall accuracy of 76% across the four classes. It suggests that the model could identify the different types of lung conditions from the X-ray images with a moderate degree of success.

In the breakdown of the classification report, we see a more detailed view of the model's performance. The f1-score, which combines precision and recall in a single metric, is 69% for COVID-19, 67% for Normal, 75% for Lung Opacity, and 91% for Viral Pneumonia.

The model performed exceptionally well in classifying Viral Pneumonia, with high precision (94%) and recall (89%). However, for COVID-19, the model showed a slightly lower f1-score, with a precision of 63% and recall of 77%. This might be due to the overlapping symptoms of COVID-19 with other lung conditions, making it harder for the model to distinguish. Similarly, for Normal (healthy) lungs, the model showed a precision of 79% and recall of 58%, indicating a degree of misclassification for healthy images.

In the confusion matrix, the model showed some confusion, especially between COVID-19 and Lung Opacity, and Normal with both COVID-19 and Lung Opacity. This could be attributed to the fact that these conditions may share similar patterns on the X-ray images, making them harder to distinguish.

3.2 ResNet-50

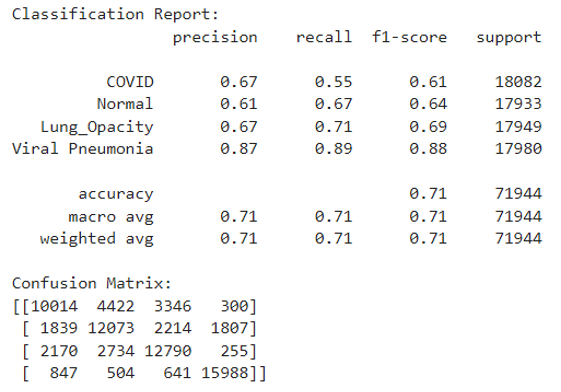


Figure ResNet-50 Classification Metrics

ResNet-50 demonstrated a solid performance in our task, with an overall accuracy of 0.71. In terms of individual class performance, the F1-score for COVID was 0.61, indicating moderate precision and recall balance for COVID detection. ResNet-50 identified 'Normal' conditions with an F1-score of 0.64, while 'Lung\_Opacity' and 'Viral Pneumonia' were recognized with higher F1-scores of 0.69 and 0.88, respectively.

The high F1-score for 'Viral Pneumonia' at 0.88 shows the model was particularly effective in distinguishing this class from the others. This could be due to distinguishing features in the images of 'Viral Pneumonia' cases, which might be more apparent or unique compared to those of other categories.

The confusion matrix further breaks down the model's performance. There was some misclassification across classes, especially between 'COVID', 'Normal', and 'Lung\_Opacity', which could suggest the presence of similar features in these classes that caused confusion for the model.

3.3 VGG19

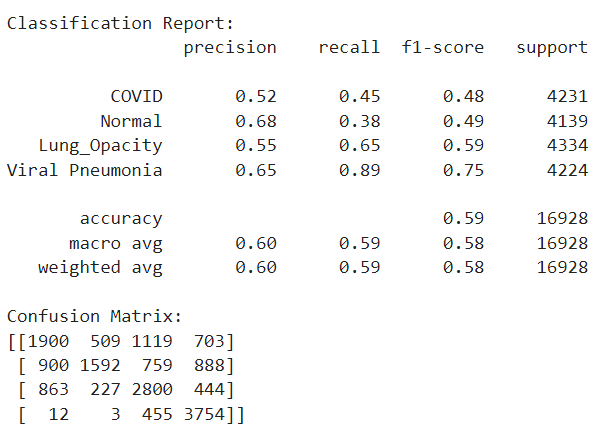


Figure VGG19 Classification Metrics

During the evaluation at step 60, the VGG19 model demonstrated a validation loss of 0.6312 and an overall accuracy of 0.7297. These metrics, while decent, suggest that the model might have room for further refinement to improve its performance.

Upon examining the individual class performance, the F1-scores for 'COVID', 'Normal', and 'Lung\_Opacity' were relatively low, with values of 0.48, 0.49, and 0.59 respectively. This suggests that the model had challenges accurately predicting these classes. 'Viral Pneumonia', on the other hand, had a notably higher F1-score of 0.75, indicating a well-balanced precision and recall for this class.

The confusion matrix provides further insights into the model's performance. Significant confusion can be observed between the 'COVID', 'Normal', and 'Lung\_Opacity' classes. However, 'Viral Pneumonia' had a significantly higher true positive rate with lesser misclassifications, which is consistent with the high F1-score for this class.

3.4 DenseNet-121

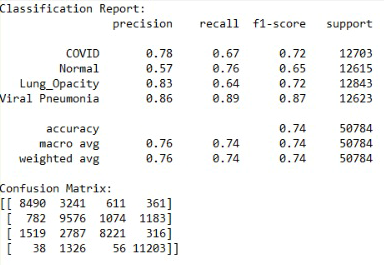


Figure 10 DenseNet-121 Classification Metrics

DenseNet-121 was slightly inferior to the Resnet-18 except in the class COVID-19 where it slightly outperformed it (f1 score of 0.72 vs 0.69) as can be seen from the confusion matrix a lot of images were classified as Normal where they shouldn’t have (False Negative).

1. Conclusions and further steps

ResNet-50 achieved an overall f1 score of 0.69 and showed relatively balanced performance across all classes. However, it demonstrated some challenges in distinguishing between 'COVID', 'Normal', and 'Lung\_Opacity' classes.

VGG19 had a lower overall f1 score of 0.57. Its performance was particularly weak for the 'COVID' and 'Normal' classes, as indicated by the lower F1-scores of 0.48 and 0.49, respectively. The confusion matrix supports this, revealing a higher rate of misclassification for these two classes.

Densnet-121 showed slightly inferior results compared to the Resnet 18 with an overall f1 score of 0.74.

The ResNet-18 model outperformed the other three models with an overall f1 score of 0.76. While it demonstrated improved performance across all classes compared to the other models, it still showed some difficulty accurately classifying 'COVID' and 'Normal' images, as demonstrated by the relatively lower F1-scores and the confusion matrix.

All models struggled with False Negatives, where they classified many images as Normal where they shouldn’t have but as can be seen from the confusion matrices in ResNet-18 it classified only 630 images of Viral Pneumonia as Normal while DenseNet-121 classified 1183 as such which is a considerable difference.

Based on these results, one possible conclusion is that the ResNet-18 model is the best performer among the four for this specific task. This could be due to the ability of ResNet-18 to capture important image features more effectively than the other models. The specific architecture of ResNet-18, which includes shortcut connections to avoid the problem of vanishing gradients, might be better suited for this image classification task.

To improve the models' performance in future work, there are several potential strategies:

1. Data Augmentation: Augmenting the dataset by introducing variations of the existing images (such as rotated or zoomed images) can help make the model more robust to different image representations.
2. Hyperparameter Tuning (using Grid Search or Optuna): Tuning hyperparameters like learning rate, batch size, and number of epochs can often lead to better performance.
3. Ensemble Methods: Combining the predictions from different models can often lead to better performance than any single model.
4. Additional Data: If possible, gathering additional labeled data for the 'COVID' and 'Normal' classes could help improve performance.
5. Model Architecture Adjustments: It might be beneficial to experiment with different layers or activation functions to see if they can better capture the relevant patterns in the data.
6. Trying different architectures: such as ResNet-101 and EfficientNet.