Transfer Learning by Feature Extraction and Fine-Tuning Approaches with DenseNet121 for Multi-Class Classification of COVID-19, Pneumonia, and Healthy Chest X-ray.

Fatima Rabia Yapicioglu[†]

***Abstract*—The Coronavirus Disease 2019 (COVID-19) pandemic is still wreaking havoc on the global population's health and well-being. Due to the misleading results and slow process of RT-PCR test kits during the detection of COVID-19, its detection employing computer aided medical imaging techniques plays a crucial role. Early research discovered that patients with COVID-19 infection have anomalies in chest radiography imaging.**

Also, there is another common infection which threats lung condition called Pneumonia. It is a bacterial or viral infection similar to the COVID-19 cases in terms of respiratory symptoms. Therefore, it is also critical to diagnose and discriminate between Pneumonia and COVID-19 in order to provide correct curative treatment and boost survival chances.

This study proposes a computer-aided diagnosis system for automatic disease detection using Chest X-ray images of three classes COVID-19, Pneumonia, and Healthy. Two ways of customizing Transfer Learning, Feature Extraction and Fine-Tuning approach with the pre-trained network called DenseNet121 have been utilized for classification and performance comparison purpose. The scores of five evaluation metrics can be listed as Accuracy, Loss, Precision, Recall, F1-score and Confusion Matrix. The approach utilized was evaluated on COVID19, Pneumonia and Healthy Chest X-ray PA Dataset from Khulna University of Engineering and Technology[1]. The model that we created achieved Accuracy rate of 94.58%, and on average Precision rate of 89% , Recall rate of 87% and F1-score of 87%. The results were quite efficient even with the the less number of iterations and epochs, this can be further trained but we would like to consider the case of over-fitting.

***Index Terms*—Deep Learning, Transfer Learning, Fine-tuning, COVID-19 Detection, Chest X-ray, Feature Extraction.**

I. Introduction

The ongoing COVID-19 phenomenon has been causing major lung damage and breathing issues since December 2019. Antibiotics are rendered ineffective in the treatment of COVID-19 patients, worsening the illness if their immune system is compromised. Various pharmaceutical laboratories and research institutions have recently created vaccinations, and several countries have begun immunization; however, the possibility of finally defeating this pandemic has been adversely affected by the appearance of several mutants with higher highly infectious levels. RT-PCR includes obtaining samples from the nose or throat and analyzing them experimentally to identify and classify all discovered infectious genetic variants of the disease. The RT-PCR testing has a high false-negative rate, as well as a long time to obtain results and kits that are inconvenient in some areas.

This study will provide a comparison between two ways of customizing a pre-trained model using transfer learning approach by fine-tuning and feature extraction, thus will show an effective way of using pre-trained models for classifying between COVID-19, Pneumonia, and Healthy images. Due to the high false-negative rates of RT-PCR test kits and fast spreading of genetic variants, the early detection and treatment should be taken into consideration in terms of helping to prevent the disease from progressing to a deadly stage. According to the hemotology fellow Atilla Uslu from Ankara University School of medicine, Uslu, Atilla(@ativitta) "We started to have cases where the COVID-19 test was negative twice, but lung tomography seemed compatible with coronavirus. How the samples are taken from the mouth and nose affects the results. When not properly taken from the mouth and nose affect the results. When not properly taken these test can produce incorrect results."

As a result, early diagnosis of lung disorders and finding the quick and correct way of building neural networks for early detection are more critical than ever. Following their construction training and generalization performance along with the high rates of Precision, Recall and F1-Score is important. For this, when especially the community of researchers are in need of high-resolution and huge number of samples, transfer learning approaches as proposed in this paper will be boon to those who spend effort in this field. This study can be further improved with different pre-trained model architectures and more data in order to get better accuracy results. Thus, effective screening of infected people will help in the fight against COVID-19 so that those who are sick can receive timely care and treatment, and also be isolated to avoid the spread of the virus.

[†]Department of Information Engineering, University of Padova, email: fatimarabia.yapicioglu@studenti.unipd.it

Special thanks to Michele Rossi, ICT for Internet and Multimedia Department, University of Padova.

II. Related Work

Since the beginning of pandemics researchers spending great effort on the development and evolution of systems which are helpful in forecasting, detection, organizing of pandemics. Computer Vision and Artificial Intelligence emerged in the medical industry in order to boost problems that are hard to solve with existing tools. In recent times, several developments are done utilizing Convolutional Neural Network to solve problems which the human eye is not sufficient enough. COVID-19 and Pneumonia-Affected Chest X-ray Image Classification Using a Transfer Learning-Based Approach with Deep CNN research by Chakraborty et al. [2] demonstrated a successful way of using pre-trained VGG-19 architecture for transfer learning usage for classification purpose. For diagnosing COVID-19, Rahimzadeh and Attar et al. [3] employed transfer learning methodology, which entails fine-tuning a concatenated Xception and Resnet50V2 structure. Ardakani et al. [4] described the 10 CNNs that were used to differentiate COVID-19 infection from non-COVID-19 infection, thus provides an extensive comparison. Ucar and Korkmaz et al. [5] proposed an Artificial Intelligence framework based on a pre-trained Squeezenet network with Bayesian optimization. With its small model size, the suggested deep Bayes-SqueezeNet is simple to implement in hardware deployments.To detect COVID-19, Abraham and Nair et al. [6] built models based on an ensemble of CNNs. The results of this study's studies showed that pre-trained multi-CNNs outperformed single CNNs in detecting COVID-19. Rousan, L.A. et al.[7] proposes a study to describe the chest x-ray findings as well as the temporal radiographic changes in COVID-19 patients. As a result, they found nearly half of the COVID-19 patients showed abnormal chest x-ray findings, with GGO in a peripheral distribution and lower lobe preference being the most common. Another study by Phankokkruad M. [8] the models will be created from the three pre-trained models include Xception, VGG16, and Inception-Resnet-V2 model in order to make performance comparison but COVID-19 X-ray dataset has a small number of images containing 323 images.Another important paper by Gu et al. is the Temporal association between serial RT-PCR results and serial chest CT imaging, as well as serial CT alterations in coronavirus 2019 (COVID-19) pneumonia: a descriptive research of 155 cases in China [9] discovered that Chest CT diagnosed COVID-19 pneumonia sooner than RT-PCR results and that it can be utilized to track the progression of the disease. The early diagnosis of COVID-19 pneumonia could be aided by combining imaging findings with epidemiological history and clinical information. Having explained the importance of using Chest X-ray images for the detection of COVID-19 and Pneumonia, finding the correct model and adjusting hyper-parameters in an optimum way plays a significant role. As provided above research has been made generally about comparing different pre-trained models on the same method or approach, but this paper will provide a comparison between two techniques.

III. Processing Pipeline

The proposed method's structure is represented in Figure 1. After examining and understanding the data we built a input pipeline by using Keras ImageDataGenerator.Then the pre-processed and rescaled images between [0,1] in size of (224,224) are fed into the base pre-trained model then the classification head, which consists of the GlobalAveragePooling2D layer, which generates predictions from the block of features, averaged over the spatial 7x7 spatial locations and converts the features to a single 1024-element vector per image according to the structure we built. The composition of our model therefore structured as loading the pre-trained base model(and pre-trained weights), then in the next step stacking the classification layers on top. In order to adjust a decision mechanism we also add a dense layer which consist of three nodes at the very end of the pre-trained network with the softmax activation function. During the training process we would like to emphasize the difference between two ways of customizing a pre-trained model: Feature Extraction and Fine-Tuning. In the Feature Extraction approach we freeze the convolutional base and prevent it to be trained, therefore we only train the additional classifier that we added on top of the base model. In the second approach, on the other hand, we unfreeze a few of the top layers of a frozen model base and train both the new classifier layers and the base model's final layers at the same time. This allows us to "fine-tune" the underlying model's higher-order feature representations to make them more relevant for the task at hand. Finally, we evaluated the metrics of Accuracy, Loss, Precision, Recall, F1-Score rates and computed the confusion matrix for the Fine-Tuning approach.

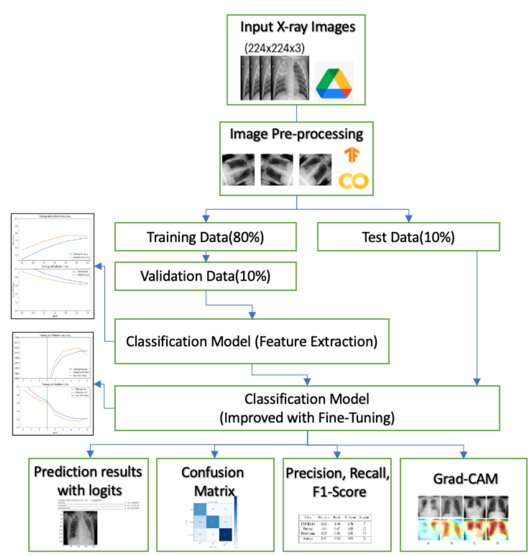

Fig. 1: Overall Experiment from Input X-ray Images to the Evaluation Metrics
This figures shows the experiment tasks and evaluation metrics in the order of their exposure.

IV. DATASET DESCRIPTION

Many diseases can affect the human lungs, including pneumonia, lung cancer, and, more recently, COVID-19. Chest CT or X-ray pictures are required for the diagnosis of various disorders, as they serve a key and vital role. In this sense, ”COVID19, Pneumonia and normal Chest X-ray PA Dataset” which is organized into 3 different folders (covid-19, pneumonia, normal) has been used. The provided X-ray samples of COVID-19 has been retrieved from various sources which can be investigated in detail in the COVID19, Pneumonia and Normal Chest X-ray PA Dataset Paper published by Amanullah Asraf, Zabirul Islam[1]. The dataset consist of 1525 chest x-ray images per each class which sums up to 4575 in total. The dataset satisfies the great expectation of neural networks from the perspective of keeping balance in the dataset among different classes. Then, we applied data augmentation method in order to artificially introduce sample diversity by applying random transformations such as random flip and random rotation. For the feature extraction, we utilized transfer learning which is the process of applying features acquired on one problem to a new, related problem.Along with Fine-Tuning approach by unfreezing the few top layers of the pre-trained base model and retraining it on the new data, we get more specific features belong to our data-set during the feature extraction. Finally, we split the data by reserving 80% for the training, 10% for the validation data and 10% for the test data via a custom function that we provided.

V. LEARNING FRAMEWORK

A. Proposed Method

Transfer learning is the process of employing models that have been trained on one problem as a starting point for a new challenge. It’s adaptable, allowing you to employ pre-trained models directly and quickly adapting them to a different situation. This paper proposes a comparison between two types of Transfer Learning approaches, Feature extraction and Fine-Tuning.

B. Feature Extraction

In the first approach we experimented Transfer Learning by Feature Extraction as we indicated earlier in this paper, the convolutional base has been frozen and additionally we added a classifier top on top it and train the top-level classifier. By freezing the convolutional base before we compile will prevent weights in a network from being updated during training. The complete model does not need to be (re)trained. The fundamental convolutional network already has features that can be used to identify images in general. The final, classification element of the pre-trained model, on the other hand, is particular to the original classification task and, as a result, to the collection of classes on which it was trained.

C. Fine-Tuning

Along with the training of the classifier you added, one method to improve performance even more is to train (or ”fine-tune”) the weights of the upper layers of the pre-trained model. Transfer learning by Fine-Tuning is our second approach where the weights will be forced to be modified from general feature maps to features particular to the dataset during the training process. The first few layers learn very basic and generic aspects that apply to practically all image formats. The features get more particular to the dataset on which the model was trained as you move higher up. Rather of overwriting the generic learning, Fine-Tuning aims to adapt these specialized features to operate with the new dataset. Therefore we fine tune till the layer number 100. It’s as simple as unfreezing the base model and making the lowest layers untrainable.

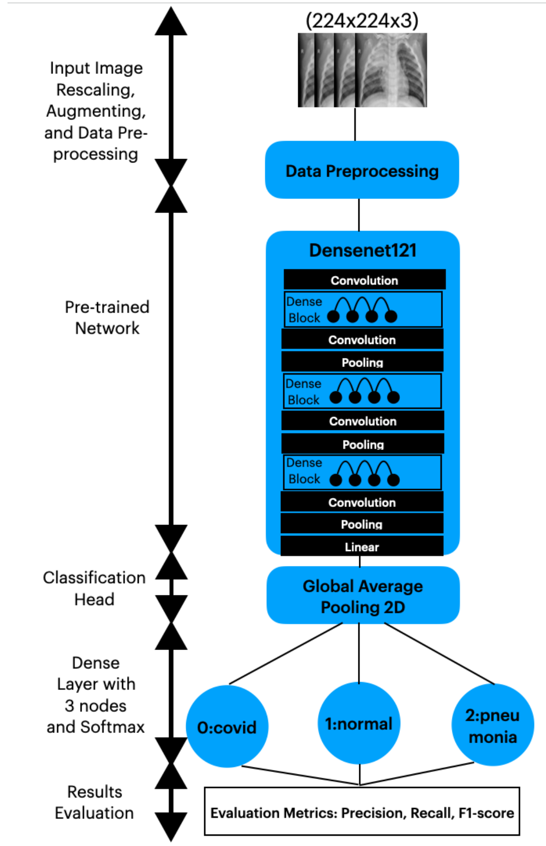

Fig. 2: Model Structure

D. Classification

In total, 4575 images belong to 3 classes which has been set to have image size as 224 x 224 and batch size as 32 have been used during training. Adam optimizer with the base learning rate 0.0001 and the Sparse Categorical Cross Entropy as loss function have been used to compile the model. With the first approach, feature extraction, the model has been trained with 4 epochs, following this the model has been trained with 4 more epochs with the adjustments which the second approach, transfer learning by Fine-Tuning, requires. As a background Tensorflow background has been utilized along with Keras libraries for optimization and loss functions. Also, Dropout and Early Stopping techniques have been applied for regularization purposes.

E. Model Evaluation

Precision (1), Recall (2), and F1-score (3) are used to evaluate the classification model's performance. A plot line which demonstrate the accuracy and loss performances of the two methods feature extraction and Fine-Tuning is provided. Also a confusion matrix which consist of three classes is plotted.

$$Precision = \frac{TP}{TP + FP}. \tag{1}$$

$$Recall = \frac{TP}{TP + FN}. \tag{2}$$

$$F1 - Score = 2x\frac{Precision + Recall}{PrecisionxRecall}. \tag{3}$$

F. Experimental Organization

The experiments have been carried out on Google Colab platform and the dataset imported from drive via Colab's 'mount' method. Runtime had take the advantage of GPU utilization provided by Google's servers. Tensorflow background has been utilized with diverse libraries such as Keras, Matplotlib, Numpy, Pandas, OpenCV, and GraphViz.

VI. Results

During the implementation of Feature Extraction after 4 epochs, the model achieved 77.23% test accuracy. Then additional 4 epochs has been applied to train the model after fine-tuning approach adjustments, the model validation accuracy reached up to 94.42%. The evaluation metric of test accuracy is 94.58%. As an optimizer we used Adam with the initial learning rate of 0.0001 and as a loss function we used Sparse Categorical Cross Entropy. Because the classes are presented in a mutually exclusive format, the model used sparse categorical crossentropy. (For example, when each sample belongs to a single class). Training accuracy and validation accuracy values before and after the fine tuning point can be seen as in the Figure 2.

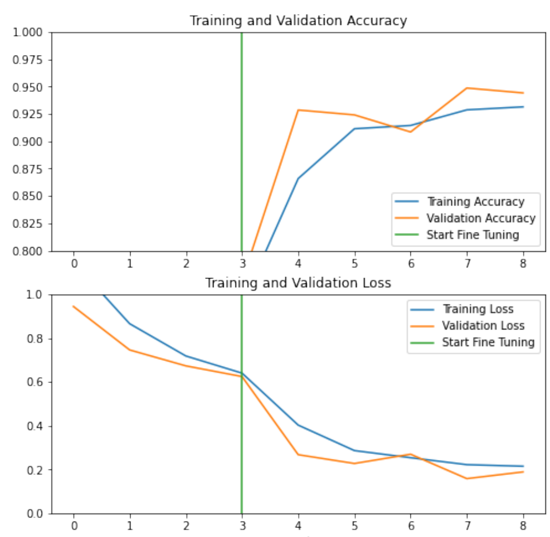

Fig. 3: Training and Validation Accuracy Over Fine-Tuning

Also the Precision, Recall, and F1-score values which belong to Fine-Tuning approach can be found in Table 1. As we can see from Figure 2, Fine-Tuning has an excellent effect on the rising value of the training and the validation accuracy.

The green line on the plot represent where the transition from feature extraction to fine-tuning. On the right hand side the model performed better with the fine-tuning approach.

Class	Precision	Recall	F1-Score	Support
COVID-19	0.81	0.93	0.87	14
Normal	0.86	0.67	0.75	9
Pneumonia	1.00	1.00	1.00	9
Average	0.89	0.87	0.87	32

TABLE 1: Class, Precision, Recall, F1-Score, and Support

Then we also visualized a couple of images from a batch of test dataset. Since we utilized the softmax activation function at the final dense layer, it is possible to get logits, and produce the probabilities belong to each class during the decision mechanism of our model.

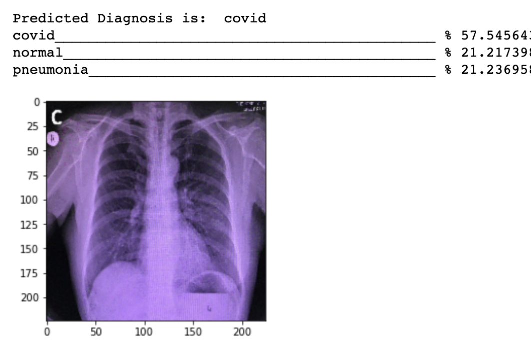

Fig. 4: Demo of the model performance

Finally, the Grad-CAM technique has been used to visualize the regions of input that are "essential" for predictions from these models or visual explanations using COVID-19 and Pneumonia chest X-ray.

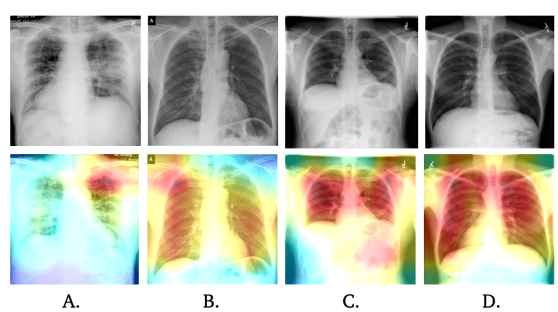

Fig. 5: Grad-CAM Visualization of COVID-19 and Penumonia

A confusion matrix for the test dataset which has been prepared by taking the 10% of dataset is generated .The confusion matrix that is computed by our model can be seen as in the Figure 3. The model is good at detecting COVID-19 cases and Pneumonia but on the other hand detecting or differentiating normal cases from other classes is harder according to this confusion matrix.

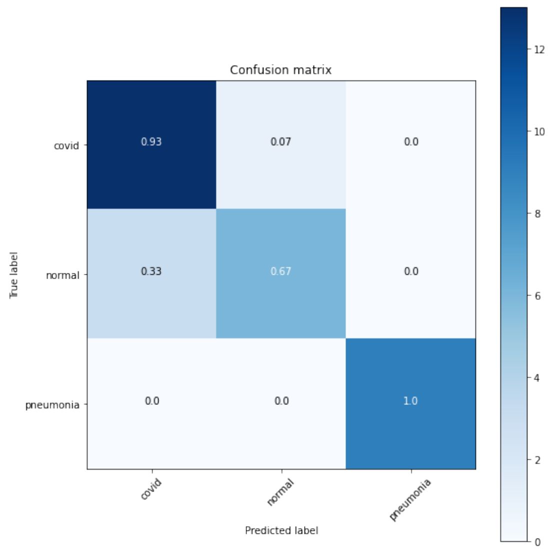

Fig. 6: Confusion Matrix

VII. Concluding Remarks

The goal of this work was to build a deep learning-based approach for classifying COVID-19, Pneumonia-affected and Healthy chest X-ray pictures utilizing two methods of adapting pre-trained models. These two ways were the feature extracting and fine-tuning, and the final message of this paper states that the fine-tuning dramatically increase the performance of the model. In total, 4575 images which are balanced in terms of number of pictures belong to each class have been fed into the feed-forward neural network. With fine-tuning approach unfreezing a few of the top layers of a frozen model base and jointly training both the newly added classifier layers and the base model's last levels allows us to "fine-tune" the base model's higher-order feature representations to make them more relevant for the current job. These findings show that deep learning can support human-level effort in image classification tasks, making the diagnostic procedure more pleasant.

References

1. Asraf, Amanullah; Islam, Zabirul (2021), â€œCOVID19, Pneumonia and Normal Chest X-ray PA Datasetâ€, Mendeley Data, V1, doi: 10.17632/jctsfj2sfn.1

2. Chakraborty S, Paul S, Hasan KMA. A Transfer Learning-Based Approach with Deep CNN for COVID-19- and Pneumonia-Affected Chest X-ray Image Classification. SN Comput Sci. 2022;3(1):17. doi: 10.1007/s42979-021-00881-5. Epub 2021 Oct 26. PMID: 34723208; PMCID: PMC8547126.

3. Mohammad Rahimzadeh, Abolfazl Attar, A modified deep convolutional neural network for detecting COVID-19 and pneumonia from chest X-ray images based on the concatenation of Xception and ResNet50V2, Informatics in Medicine Unlocked, Volume 19,2020, 100360, ISSN 2352-9148.https://doi.org/10.1016/j.imu.2020.100360.

4. Ali Abbasian Ardakani, Alireza Rajabzadeh Kanafi, U. Rajendra Acharya, Nazanin Khadem, Afshin Mohammadi, Application of deep learning technique to manage COVID-19 in routine clinical practice using CT images: Results of 10 convolutional neural networks, Computers in Biology and Medicine, Volume 121, 2020, 103795, ISSN 0010-4825.

5. Ferhat Ucar, Deniz Korkmaz, COVIDiagnosis-Net: Deep Bayes-SqueezeNet based diagnosis of the coronavirus disease 2019 (COVID-19) from X-ray images, Medical Hypotheses, Volume 140, 2020, 109761, ISSN 0306-9877, https://doi.org/10.1016/j.mehy.2020.109761.

6. Bejoy Abraham, Madhu S. Nair, Computer-aided detection of COVID-19 from X-ray images using multi-CNN and Bayesnet classifier, Biocybernetics and Biomedical Engineering, Volume 40, Issue 4, 2020, Pages 1436-1445, ISSN 0208-5216, https://doi.org/10.1016/j.bbe.2020.08.005.

7. Liqa A. Rousan, Eyhab Elobeid, Musaab Karrar, Yousef Khader, Chest X-ray findings and temporal lung changes in patients with covid-19 pneumonia, BMC Pulm. Med. 20 (1) (2020) 1â€“9. https://bmcpulmmed.biomedcentral.com/articles/10.1186/s12890-020-01286-5.

8. Manop Phankokkruad. 2020. COVID-19 Pneumonia Detection in Chest X-ray Images Using Transfer Learning of Convolutional Neural Networks. In Proceedings of the 3rd International Conference on Data Science and Information Technology (DSIT 2020). Association for Computing Machinery, New York, NY, USA, 147â€“152. DOI:https://doi.org/10.1145/3414274.3414496.

9. Gu J, Yang L, Li T, Liu Y, Zhang J, Ning K, Su D. Temporal relationship between serial RT-PCR results and serial chest CT imaging, and serial CT changes in coronavirus 2019 (COVID-19) pneumonia: a descriptive study of 155 cases in China. Eur Radiol. 2021 Mar;31(3):1175-1184. doi: 10.1007/s00330-020-07268-9. Epub 2020 Sep 15. PMID: 32930834; PMCID: PMC7491042.