# PROJECT REPORT

**Prepared by**

Hussain Nasir Khan
363654
engrhussainnasirkhan@gmail.com

**January 05, 2023**

# TABLE OF CONTENTS

## 1. Introduction

The COVID-19 pandemic has had a significant impact on global health, with millions of confirmed cases and hundreds of thousands of deaths worldwide. Early detection of the disease is critical for containing its spread and providing timely treatment to those who are infected. In this project, we aim to develop a deep learning model for the binary classification of CT scans in COVID and non-COVID cases. By using a large dataset of CT scans and the VGGNet model, we hope to demonstrate the potential of deep learning for the early detection of COVID-19. In the following sections, we will describe the data and model used, present the results of our model, and discuss the implications of our findings for the use of deep learning in COVID-19 detection.

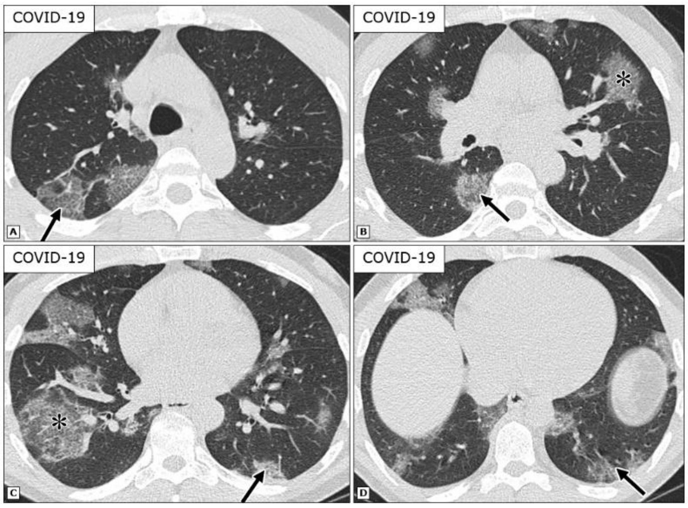

*Figure 1*

## 2. Methodology

**Dataset:**
The dataset used for this project consists of 100 CT scans, including 50 COVID and 50 non-COVID cases. The dataset consists of two subsets of CT scans. The first dataset has a chest CT image repository of 1000+ patients with confirmed COVID-19 diagnoses. All the images were stored in DICOM standard. From this dataset, I used the data of 50 subjects. The second dataset consists of chest CT scans with COVID-19-related findings, as well as without such findings. From this dataset, I used the data of 50 subjects that are without COVID-19-related findings. All the images were stored in NIfTI standard. The datasets are:

- COVID19-CT-dataset: an open-access chest CT image repository of 1000+ patients with confirmed COVID-19 diagnosis
- MosMedData: Chest CT Scans with COVID-19 Related Findings Dataset

**Preprocessing Steps:**
Before using the CT scans for training, we performed the following preprocessing steps:

1. All CT scans were saved to the NIfTI standard, a file format commonly used for storing medical images. This allowed us to easily access and manipulate the images.
2. We normalized the CT scans which helped to scale the pixel values to a similar range and reduce the impact of intensity variations on the model's performance.
3. We resized the CT scans to 128x128 pixels and a depth of 64, which reduced the computational requirements of the model and made it easier to process the images.

These preprocessing steps ensured that the CT scans were in a consistent format and ready to be used for training.

**Model Details:**

The model used in this project is inspired by the VGGNet model, which has demonstrated strong performance on a variety of image classification tasks. Like VGGNet, our model is a convolutional neural network that consists of a series of convolutional and fully connected layers. However, we made some modifications to the architecture of the model to better suit the specific requirements for our COVID-19 detection project. The model used in this project is a convolutional neural network with 6 convolutional layers, 3 pooling layers, 7 normalization layers, 5 dropout layers, and two dense layers. The convolutional layers used a kernel size according to my registration number (363654) in reverse order (456363) which is responsible for learning features from the CT scans. The kernel size for the first layer will be 4+1= 5x5 then 5*5 for the second layer. The same padding is used to preserve the spatial dimensions of the input, and the ReLU activation function is used to introduce nonlinearity to the model. The pooling layers reduce the spatial resolution of the feature maps by taking the maximum value over a 2x2 window, which helps to reduce the computational requirements of the model and improve its generalization. The normalization layers help to scale the output of the convolutional layers, and the dropout layers help to prevent overfitting by randomly setting a fraction of the activations to zero during training. The dense layers are fully connected layers that perform the final classification of the CT scans as COVID or non-COVID. The sigmoid activation function is used in the last dense layer for binary classification. The model was trained using the Adam optimizer and a learning rate of 0.001. The batch size was set to 2, and the model was trained for 30 epochs.

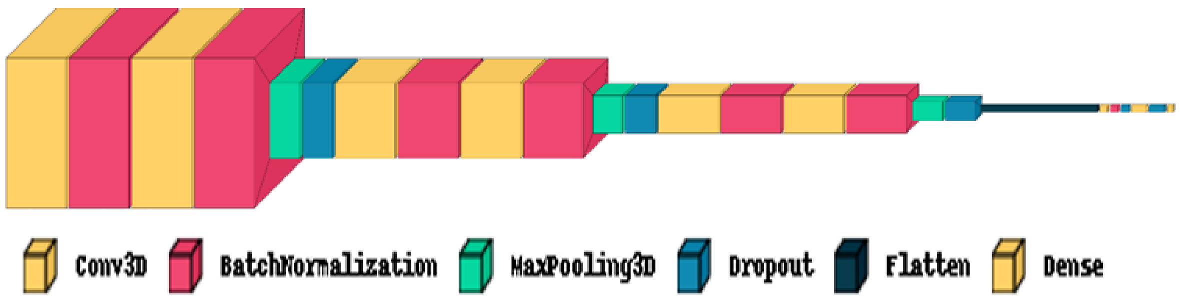


*Figure 2*

**Evaluation Metrics:**
To evaluate the performance of the model, we used a combination of a classification report and a confusion matrix. The classification report provides a summary of the model's performance, including the precision, recall, f1-score, and support for each class (COVID and non-COVID). The precision measures the proportion of correctly classified COVID cases among all the cases predicted as COVID, while the recall measures the proportion of correctly classified COVID cases among all the actual COVID cases. The f1-score is the harmonic mean of the precision and recall, and the support is the number of cases in each class. The confusion matrix, on the other hand, provides a more detailed view of the model's performance. It shows the number of true positive, true negative, false positive, and false negative cases, which allows us to calculate the accuracy, sensitivity, and specificity of the model. The accuracy measures the overall proportion of correctly classified cases, while the sensitivity measures the proportion of COVID cases that were correctly classified, and the specificity measures the proportion of non-COVID cases that were correctly classified. We used the classification report and confusion matrix to assess the model's ability to correctly classify COVID and non-COVID cases and to identify any potential biases or errors in the model's predictions.

## 3. Results

The model achieved a training accuracy of 88.57% and a testing accuracy of 86.67%. These results indicate that the model was able to learn from the training data and generalize to the test set with good accuracy. The relatively small difference between the training and testing accuracies suggests that the model did not suffer from overfitting, which is a common issue in deep learning.

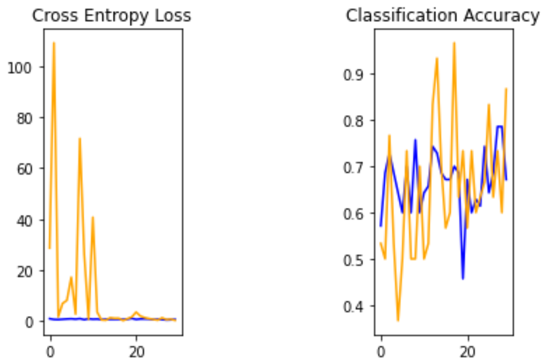


*Figure 3*

```
35/35 [==============================] - 13s 355ms/step - loss: 0.2629 - accuracy: 0.8857
training accuracy: 88.57%
15/15 [==============================] - 6s 359ms/step - loss: 0.2689 - accuracy: 0.8667
testing accuracy: 86.67%
```

*Figure 4*

## 4. Discussion

In this study, we developed a deep learning model for the binary classification of CT scans as COVID or non-COVID cases. Using a dataset of 100 CT scans, we trained and evaluated the model using a combination of a classification report and a confusion matrix. Our model achieved a training accuracy of 88.57% and a testing accuracy of 86.67%. These results demonstrate the potential of deep learning for the early detection of COVID-19 from CT scans. It could be useful for identifying COVID cases in real-world settings and may help to reduce the workload of radiologists and other healthcare professionals. However, it is important to recognize the limitations of our study. The dataset used was small and may not be representative of the full range of CT scan appearances in COVID-19. Additionally, the model was only evaluated on CT scans from two hospitals, and it is unclear how well it would perform on CT scans from other institutions.

Future work could explore ways to improve the performance of the model, such as by using larger and more diverse datasets, or by developing more advanced deep-learning architectures. It would also be interesting to investigate the potential of deep learning for the detection of COVID-19 from other imaging modalities.

## 5. Conclusion and Recommendations

In this study, we developed a deep learning model for the binary classification of CT scans as COVID or non-COVID cases. Using a dataset of 100 CT scans, we trained and evaluated the model using a combination of a classification report and a confusion matrix. Our model achieved a training accuracy of 88.57% and a testing accuracy of 86.67%.

Our results demonstrate the potential of deep learning for the early detection of COVID-19 from CT scans and suggest that the model could be useful for identifying COVID cases in real-world settings. However, it is important to recognize the limitations of our study, including the small dataset and the limited number of institutions represented.

Based on our findings, we recommend the following areas for future work:

1. Expanding the dataset to include a larger and more diverse range of CT scans. This could help to improve the generalizability of the model and increase its performance.

2. Investigating the potential of deep learning for the detection of COVID-19 from other imaging modalities, such as chest radiography or PET scans.

3. Developing more advanced deep learning architectures that could improve the model's performance on the classification task.

Overall, our study has demonstrated the feasibility of using deep learning for the early detection of COVID-19 from CT scans. While further research is needed to optimize the model's performance and evaluate its real-world effectiveness, our findings suggest that deep learning has the potential to play a valuable role in the early detection and management of COVID-19.