

Detection of COVID-19 in Computed Tomography (CT) Scan Images using Deep Learning



Abdulrazak Yahya Saleh¹, Letchmikanthan Ilango¹

¹ Faculty of Cognitive Sciences & Human Development, Universiti Malaysia Sarawak,
94300 Kota Samarahan, Sarawak, Malaysia
ysahabdulrazak@unimas.my

ABSTRACT

Coronavirus disease (COVID-19) is a fresh genus found in 2019 that was not previously known in humans. On the other hand, Deep learning is one of the most important fields of medical imaging science at present. In this article, a model of deep learning is being trained for the COVID-19 detection in CT Scan images. This study is implemented using Python programming language. To build and train the Convolution Neural Network (CNN) model, Python Deep Learning libraries such as Keras and TensorFlow 2.0 have been utilized. As for the dataset, the open source dataset of COVID-19 chest computed tomography (CT) images were used. These image where been confirmed by the senior radiologist who performed Diagnosis of and treatment of patients with COVID-19. There were total of 745 images belonging to two classes were sampled. 348 positive (+) COVID-19 images and 397 negative (-) COVID-19 images. Based on the training process, the model was able to detect 79 per cent accuracy on the test set. The performance of the model, Convolution Neural Network were evaluated by comparing with Logistic Regression model. Findings from the research proves that Convolution Neural Network are reliable by producing higher accuracy rate of 79% while Logistic Regression produce a rate of 54%. However, in the future more reliable and quality image datasets should be used along with the metadata of the patients to train the model.

Key words: Convolution Neural Network (CNN), COVID-19, CT Scan Images, Deep learning.

1. INTRODUCTION

Coronavirus (CoV) is a vast group of viruses which provoke cold-related diseases including Severe Acute Respiratory Syndrome (SARS-CoV) and Respiratory Syndrome in the Middle East (MERS-CoV). Coronavirus

disease (COVID-19) is the latest genus found in 2019 that was not previously known in humans. These viruses are zoonotic transmission by wildlife to humans [1]. International Committee on Taxonomy of Viruses (ICTV) identified the virus as SARS-CoV-2 [2]; and the infectious diseases contaminated with this coronavirus is referred to as Coronavirus Virus 2019 (COVID-19) by the World Health Organization (WHO). The recent coronavirus is being identified as human-to-human transmission. The further airborne transmission of the disease from one human to another has triggered widespread infection [3]. Symptoms of the disease comprise cough, fever and respiratory symptoms. The infection can lead to septic shock, pneumonia, serious acute respiratory syndrome, multi-organ failure, and death in much more extreme cases. [4]. It had since been reported that women are less affected than men and there is no infant mortality between 0-9 years old [5]. COVID-19, viral pneumonia caused by a unique coronavirus, has spread worldwide exponentially within a few months, threatening the health of billions of human beings [6]. Although COVID-19 triggers milder symptoms in about 82 percent of cases, the others are serious or severe [5]. These viruses invade the lungs, forcing them to enlarge and obstruct the supply of oxygen that may be life-threatening. With the threat of SARS-CoV-2 further transmission, the WHO has therefore proclaimed a Public Health Emergency of International Concern (PHEIC) by 30 January 2020 [2]. Moreover, disease outbreak is growing all over the world at a pace that has never been seen before with any infectious disease. One of the powerful strategies suggested by WHO to monitor the transmission of viral infection is social isolation and contact mapping. Farooq & Hafeez in [7] mentioned that a crucial step in this path is efficient and reliable monitoring of COVID-19 patients not just for rapid diagnosis but also for isolation from the public to mitigate the transmission of infectious disease. The restricted supply of test kits makes it difficult to diagnose a large population of infected individuals by the epidemic. In the current public health crisis, poor RT-PCR sensitivity

indicates that one cannot easily detect too many COVID-19 cases and will not seek adequate care. In regards, given the increasingly infectious nature of the virus, there is a chance of infecting a greater population [8]. Besides, these experiments take from a few hours to a day or two to generate the result, which is far too energy-consuming, time-consuming, and much of the time error-prone under the present state of emergency. A faster and more efficient screening procedure, which could be further validated by the PCR method, is therefore critically required [7]. Recent findings indicate that chest radiography can be used in diagnostic categories for initial screening of COVID-19 [8]. The scanning of CT scan images may also be used as an alternative to the PCR process, which in some cases shows greater sensitivity [9]. However, the central bottleneck that radiologists face in the study of CT images is the conceptual screening with detailed observations. It requires the use of advanced techniques that can continuously derive valuable information from CT scan images that are the features of COVID-19 [7]. In brief, precise and efficient diagnosis of COVID-19 suspected cases at a very initial phase plays a key role in fast quarantine and medical care, which is therefore of considerable significance for the prognosis of patients, the containment of this outbreak and the safety of public health. As a substitute for the patients waiting to get their test results, detection could now include whoever shows a prominent pneumonia pattern of chest scan COVID-19. With this approach, the officials should be able to identify and treat patients more rapidly [10]. Chest computed tomography (CT) is a quicker and simpler form of clinical diagnosis of COVID-19 as well by comparing the clinical symptoms and signs of the patient with their previous contact tracing, travel records, and laboratory tests, which makes it possible to identify them as easily as possible in clinical practice. This is also effective to identify affected patients on a timely basis and to suppress the outbreak. CT scan is a central component of the treatment process for suspected patients and the CT representations have been demonstrated in many recent studies [11]. In this paper, therefore, Convolution Neural Network (CNN) is used as a computer tool to diagnose images of COVID-19 CT scan images. Considering that there are fewer studies, that used CT scanning images of COVID-19. CNN is considered as one of the deep learning algorithms which currently utilized a lot in medical imaging research. Deep learning algorithms learn by permitting the multi-level representation of information in a neural network across multiple hidden layers. Studies have shown that profound training approaches can be more accurate than traditional methods [12]- [19]. In the world of medical imaging, deep learning has brought countless developments. Deep learning also represents a growing field of machine

learning research that has recently been used for classification. Deep learning algorithms learn from the multi-stage representation of information in a neural network utilizing several hidden layers. There are very few published papers concerning research have been working with deep learning to detect COVID 19. The paper is divided into several sections: Section 1 addresses analysis of the literature, presents the specifics of the proposed algorithm. Section 2 describes the steps of the proposed algorithm; the experimental study shows. Section 3 displays findings and discussion. In Section 4, the paper is eventually concluded.

2. RELATED WORK

This section consists of the literature review, Preliminaries and Convolution Neural Network.

2.1 Literature Review

This sub-section addresses the findings of the literature survey carried out. Past work is reviewed in-depth and the techniques used to identify COVID-19 are analysed. The hypothesis behind COVID-19 is also explored to support this research in history. The limitation of the earlier research is analysed to achieve features that need to be changed in the design that would be developed in this research. A research performed by Prof. Zhong 's team, analyzing 1099 laboratory-confirmed cases, showed that typical clinical manifestations included headaches (13.6 percent), sputum development (33.4 percent), cough (67.8 percent), fever (88.7 percent), sore throat (13.9 percent), fatigue (38.1 percent) and difficulty breathing (18.6 percent) [20]. In fact, some patients have bowel problems of vomiting (5.0%) and diarrhea (3.8%). Experimental findings remained compatible with prior analyzes of 41, 99 and 138 patients in the Hubei area [21]. Cough and fever had been the primary signs, while gastrointestinal issues and upper respiratory signs existed uncommon, reflecting variations in viral tropism relative to SARS-CoV, MERSCoV and influenza [22]- [24]. Senior citizens and those with medical conditions (i.e. asthma, chronic respiratory diseases, cardiovascular disease, diabetes) rapidly developed into septic shock, acute respiratory distress syndrome, metabolic acidosis difficult to remedy, and coagulation disorder, leading to passing away [11].

Of laboratory study tests, Majority of the patients have elevated or reduced levels of white blood cells and lymphocytopenia [20]. Nonetheless, blood urea, neutrophil concentration, D-dimer and creatinine levels remained slightly greater in serious patients and lymphocyte numbers proceeded to decline [11]. Medical

diagnosis of COVID-19 is primarily dependent on clinical symptoms, epidemiological data and certain supplementary studies. However, clinical manifestations of patients diagnosed with SARS-CoV-2 are particularly different, including dyspnea, cough, respiratory issues, nausea and viral pneumonia. As an epidemiological analysis, consequently, Alternative tests are crucial for COVID-19 diagnosis [25].

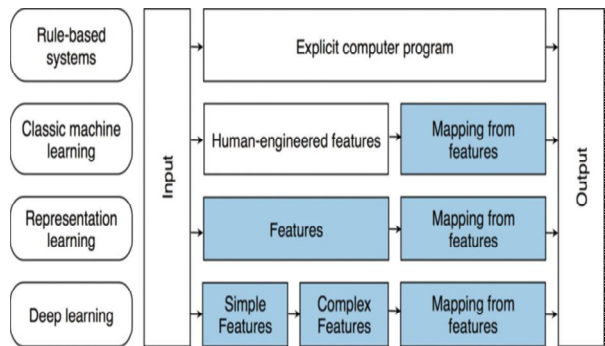
Many such doctors proposed that CT scans could be one of the required alternative diagnostic approaches since they are more versatile. High resolution CT (HRCT) for the chest is critical in patients with SARS-CoV-2 for the early diagnosis and assessment of disease seriousness [26]. Many such studies have studied the CT images of people with SARS-CoV-2 diagnoses [27]. Typical CT scans demonstrate bilateral pulmonary parenchymal glazing and combined pulmonary opacities, often with peripheral lung distribution and rounded morphology. Lung activity with peripheral prevalence has been shown in patients with MERS-CoV infections and SARS-CoV, and chest CT has shown progression of the disorder with ground-glass opacity and accumulation, similar to SARS-CoV-2 [28]. Therefore, CT scans have strong clinical and prognostic significance for COVID-19 according to these results, particularly in the high prevalence region of SARS-CoV-2 infection. However, there are also some deficiencies in CT scans, as such where, it is undistinguishable from other viral pneumonia and hysteresis of unusual CT imaging. In consideration of the shortcomings in the identification of commonly used nucleic acid and CT scans for the diagnosis of COVID-19, certain immunological testing kits containing viral antigens or antibodies could be used by clinical laboratories as quickly as possible. It is therefore important and urgent to establish other critical and precise auxiliary strategies for the detection of COVID-19.

2.2 Preliminaries

Artificial intelligence (AI) is a derivative of computer science for constructing systems that actually require human intelligence to perform tasks. Machine learning, where computers are equipped to perform tasks without complex programming, is the field of AI. Traditionally, human engineering features allow a machine to learn how to discern patterns of data. A type of machine learning that does not use functional engineering is representation learning. The computer instead learns what the data is classified. Deep learning is a form of representation learning using compositional or hierarchical characteristics learned. It is clear to see the flowcharts

relating to each other of the various sections of an AI system as in Figure1.

Classically designed machine learning relies on features that require human expertise and complex tasks. Deep learning circumvents science by using large amounts of data and dynamic hierarchical structures. Deep learning recently brought significant improvements in performance in various fields such as classification of images, voice recognition, processing of natural



languages and playgrounds. The blue boxes are elements learned by adapting a model to data examples; deep learning allows learning from input to output and end-to-end mapping.

Figure 1: Flowcharts relating to each other of the various sections of an AI system.

In the last fifty years, various approaches to chest imaging research have been established with computer aid. Ginneken in [13] reviews chest imaging computer analysis and explain that three kinds of methods which are deep learning, machine learning and regulatory image processing. It is found that has the new prevalent approach is deep learning with extremely promising results. Deep learning algorithms learn through multiple hidden layers in a neural network that allows the multilevel representation of data [29].

Deep learning techniques, the most astounding part of the field of machine learning, are an efficient way to build a complete model which can calculate final classification labels using the medical images raw pixels. Deep models in the field of analysing medical images need to be used in other areas of imaging as large data sets are required to obtain excellent features in deep architectures.

Medical images are commonly challenging to attain, however, and therefore traditionally medical data sets are rather small. Thus, if the deep model is directly used to address a small data set, the approach is apt to override the model. The interpretability of the model has proved quite poor except for these problems, and training in a profound model usually requires a great deal of

computation. A different model is now being proposed, a deeper model incorporating traditional elements to address these issues about traditional methods contrasted with deep models; Not only can this model take full advantage of existing medical knowledge but use deep architectures also to automatically extract high-level features to classify medical objects in fields focused on imaging data, including radiology and pathology, dermatology and ophthalmology, deep learning showed great promise [30]- [32]. Deep learning automatically detects complex data patterns and therefore makes statistical assessments. Deep learning networks allow automatic quantification and selection of the most solid features, contrasting to feature engineering approaches, and require little to no human input. In several activities, deep learning methods have outperformed their expected counterparts, including mammogram detection, mortality prediction and multimodal picture recording [33]- [34].

Deep learning is a type of machine learning that employs artificial neural network architectures comparable to human cognitive function design. It is a way in which the learning algorithm composes features that represent a hierarchy of structures in the data [35]. CNN and recurring Networks (RNNs) are different types of profound learning methods that use artificial neural networks (ANN).

CNNs are deep learning algorithms that, most applicable to radiology, can be used efficiently for segmentation and classification of images. The computational process of convolution by using convolutional and pools can be modelled by CNNs rather than by using more than billions in weights to create total connectivity. Spatial invariance can also be provided by CNN's. Many kernels can be educated and then learn other position invariant features for different convolutional layers. Since key features can be learned automatically, image extraction is not necessary prior to the learning process. CNN can, therefore, be applied in clinical practice relatively easily [36]- [38].

2.3 Convolution Neural Network

The CNNs are neural networks that have proven to be excellent in the processing of picture data. CNN has a three-dimensional structure that analyses the red, green and blue layers of a coloured picture across three advanced neural networks. CNN scans a picture one area at a time, recognizes, extracts, and uses essential features for the classification of the item. For computer viewing tasks such as image classification, CNNs are mostly used.

In three phases, a CNN operates. The first is a convolution that "scans" the image a few pixels at a time, and creates a feature map that will probably have the

necessary class (in a simple classification example). The second phase is the pooling of (also called sampling), which reduces each feature's dimensionality while retaining its most important information.

The stage of pooling provides a "summary" of the key features of the image. Some CNNs use max pooling, which takes the maximum amount of each CNN scanned pixel region as seen below. A CNN is then able to accomplish multiple rounds of convolution. Finally, if the properties are granular, CNN enters the third level, a completely connecting neural network evaluating the ultimate probabilities and deciding the class to which the image belongs. Figure 2 presents CNN Model Architecture.

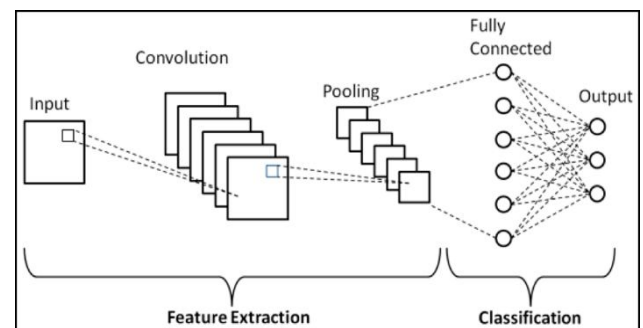


Figure 2: CNN Model Architecture. Adapted from [39]

3. METHODOLOGY

This section describes the research framework of the research. Explanations on research data are also covered. In addition, this section provides adequate data on the steps of the analysis that will be conducted to achieve the objectives of the study. The goals of the research can be accomplished by applying specific methods. There are three phases which are research design, experimental design and finally the performance evaluations phase. In phase 1 (research design), research framework is discussed and used as a guideline to conduct this research. Figure 3 shows the steps involved in research framework.

Moreover, in phase2 (Experimental Design), Deep learning is very computationally intensive because it needs high computing power. However, in this study a moderate level of laptop hardware was used. The following is the system information of HP Pavilion Notebook used in this project.

- Processor Type: Intel(R) Core (TM) i5-5200U CPU @ 2.20GHz
- Current Processor Speed: 2500 MHz
- Memory Size: 8192 MB RAM

- Storage Capacity: 931.51 GB
- Graphics Card: NVIDIA GeForce 940M
- Graphics Memory: 4188 MB

The method was implemented using data collection, pre-processing the data and validation. In the analysis, the programming language used is Python. Python is an open source programming language commonly used in deep learning. The main reason is its abundance of libraries and frameworks that are available on the Python ecosystem like Keras and TensorFlow 2.0. The simple syntax of Python also means that it is faster in development than many programming languages and allows the developer to test algorithms quickly without having to implement them. Thus, it is the best choice to use Python to train the deep learning model.

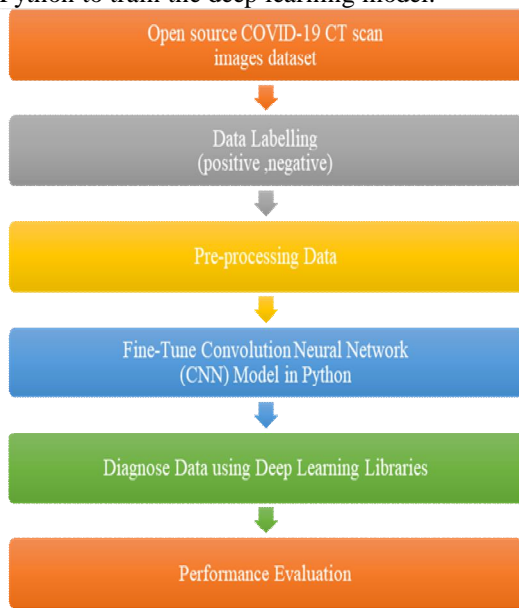


Figure3: The research framework of study. Adapted from [40].

It is possible to break the data preparation process into two portions. The first portion is the data acquisition and the second portion are the data pre-processing. For data acquisition, the type of data that was obtained in this research is the open-source dataset of COVID-19 chest computed tomography (CT) images. This image where been confirmed by the senior radiologist who performed the diagnosis and treatment of COVID-19 patients. There were total of 745 images belonging to two classes were sampled. The two classes consist of 348 COVID-19 positive (+) patients’ images and the remaining 397 COVID-19 negative (-) patients’ images. All copyrights of data belong to the authors and publishers of these papers (cite).

Figure 4 shows some of the chest CT scan images that were used in the research. On the left is positive (infected

patients) COVID-19 CT scan images whereas on the right is negative (healthy patients) COVID-19 CT scan images. In the second stage (data preprocessing), the collected data has been pre-processed to enhance the Computed Tomography (CT) scan images of chest quality where several techniques are used for that. This was done by using OpenCV library in Python. The pre-processed steps begin with reading the CT scan images in the dataset. The dataset of chest CT scan images are placed in the dataset directory folder where the data two classes are divided into covid and normal folders. Then for each image path, the class label of covid or normal were extracted. Next, the images are converted into RGB channel ordering. The last step is to resize all the varying image. Some images of the dataset vary in size; therefore, we should establish a base size for all images fed into the model. Thus, the images are resized to 224 pixels by 224 pixels and fed back into the model.

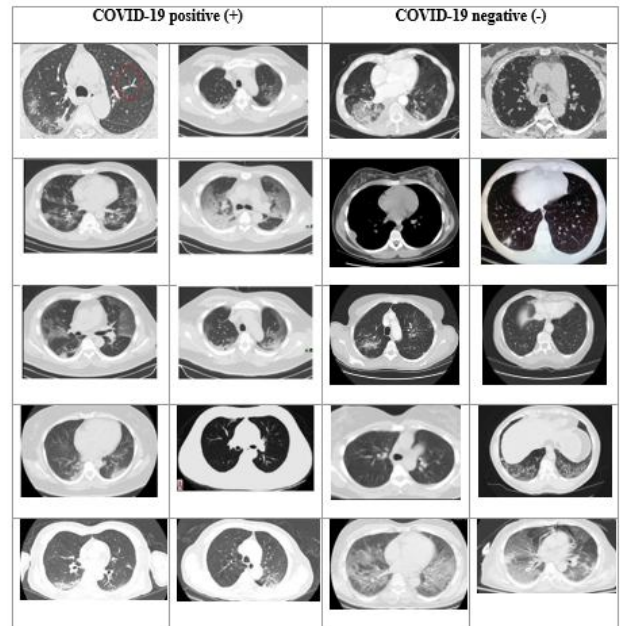


Figure 4: Sample images of CT scan

In this study, Convolutional Neural Network (CNN) method was applied. This technique is generally employed for image processing applications from pixel images with minimal pre-processing. CNN consists of neurons arranged in a 1, 2, or 3-dimensional way (width, height, depth) which transform input values into neuron output class values activations via a series of hidden layers. The popular layers on CNN are CONV, POOL and FC with the input and output layers [41]. CNN models of the neural network may take long time for data set training. One way to shorten this method is to reuse the weights of pre-trained models improved for typical

computer vision datasets such as the image recognition tasks in ImageNet. Best models for your computer vision issues can be downloaded and used or incorporated directly into a new one. Transfer learning means utilizing models that are trained on one problem to continue with a similar problem. Transfer training is versatile and allows for the direct use of pre-trained models as a preprocessing function and is incorporated into completely different models. Keras offers easy access to many highest performing models on ImageNet image recognition tasks, such as VGGNet.

Researchers from Oxford Visual Geometry Community (VGG) won the image recognition tasks in 2014 for the Convolutional Neural Network (CNN) models. Following the competition, the participants wrote their findings in the paper and made their models available online. VGGNet comprises 16 convolutional layers and its very uniform architecture makes it very appealing. This network is characterized by simplicity by a growing depth of only 3 x 3 convolution layers. Volume size reduction is achieved by max pooling. A softmax classifier will follow up two fully connected layers, each with 4,096 nodes. Currently, it is the community's most popular choice for image extraction. In many other applications, the VGGNet architecture model has since been used as a basis for function extractors in many other challenges. For this analysis, the research was carried out using the Convolution Neural Network with the VGGNet architecture. The following steps showed the proposed algorithm:

- Step1: obtaine the dataset from the open-source dataset of COVID-19 chest computed tomography (CT) images*
- Step2: pre-processe the dataset using OpenCV library*
- Step3: perform initialising of batch size hyperparameters, number of training epochs and initial learning rate*
- Step4: split data 80% for training and 20% for testing.*
- Step 5: Perform generalisation data augmentation by setting the random picture rotation setting to 15 degrees clockwise or counterclockwise.*
- Step6: Initialize and fine-tune the VGGNet model*
- Step7: Compile and train the new fully connected layers built and mounted to top of VGG16 (COVID-19 deep learning model)*
- Step8: Evaluate the predictions performance of the model using the accuracy metrics on the testing set.*

As for the evaluation of the model, the most popular assessment methods were used which is the accuracy. Accuracy is the number of correct cases predicted in all cases. Formally, accuracy has the following definition:

$$Accuracy = \frac{No. of correct predictions}{Total number of prediction.} \quad (1)$$

However, it can also be formulated as follows in terms of positive and negative for binary classification.

$$Accuracy = \frac{tp + tn}{tp + tn + fp + fn} \quad (2)$$

Where fn = False Negatives, tp = True Positives, fp = False Positives, and tn = True Negatives.

4. RESULTS AND DISCUSSION

This section describes the model implementation in this research of detecting COVID-19 in CT Scan images. The results obtained from the research were analysed and discussed in this section. Along with that, a comparative study between Convolution Neural Network and Logistic Regression was performed to compare the models accuracy. The dataset for this study is obtained from the open-source dataset of COVID-19 chest computed tomography (CT) images. These images where been confirmed by the senior radiologist who performed the diagnosis and treatment of COVID-19 patients. A total of 745 images were sampled belonging to two groups of 348 COVID-19 positive (+) images, and 397 COVID-19 negative (-) images. The chest CT scan image dataset is stored in the separate dataset directory folder where both of our data groups are COVID and regular. Then the dataset was pre-processed using OpenCV library. The steps involved are reading the CT scan images in the dataset and converting it to RGB channel ordering and resize all the varying image to 224 pixels by 224 pixels and fed back into the model. In this study, Python programming language were utilized to develop the deep learning model. In deep learning, Python is widely used because it is an open source programming language that has abundance of libraries and frameworks in its ecosystem such as Keras and TensorFlow 2.0. This deep learning model was tested and run on laptop with processor type of Intel(R) Core (TM) i5-5200U CPU @ 2.20GHz with 8192 MB RAM and Graphics Card of NVIDIA GeForce 940M. Firstly, all the libraries are imported using deep learning libraries such as Keras and TensorFlow 2.0. Moreover, OpenCV for loading and pre-processing images in the dataset and scikit-learn, matplotlib for plotting were used. Next, the path to the input dataset of chest CT scan images are defined. Along with that, initialising of batch size hyperparameters, number of training epochs and initial learning rate are performed. A model hyperparameter is a model-external configuration the value of which cannot be determined from data. It is often tuned for a given predictive modelling problem. Table 1 shows the hyperparameter used to train the model.

Table 1: Data Hyperparameter of the Model

No.	Hyperparameter	Values
1.	Batch Size	18
2.	Epochs	9
3.	Learning rate	10 ⁻³
4.	Steps per epochs	63

Afterwards, the dataset of chest CT scan images from the dataset folder were loaded. Then the class label of covid or normal were extracted. Following by the pre-processing of data was carried out by converting the images to RGB channels and resize the varying image to 224 pixels by 224 pixels. Thereafter, one hot encoding is performed on the label of the dataset and create the testing and training dataset. Each encoded label contains a two-features array with one of the features being “hot” (1) versus “not” (0). The data divided by 80% for training and 20% for testing. For model generalizes, data augmentation is performed by establishing the random image rotation setting to 15 degrees clockwise or counterclockwise. After that, Initializing and fine-tuning the VGGNet model has been performed. From there, the new fully connected layers built and mounted to top of VGG16. Subsequently, the COVID-19 deep learning model are compiled and trained. To compile the network with the Adam optimizer and learning rate decay. Since this is a binary problem, "binary_crossentropy" loss were used. To evaluate the model, predictions are done by using the testing set. The evaluation of the model performance of is done by using the accuracy metrics on the testing set. The results of both testing and training are presented on Figure 5.

```

Train for 63 steps, validate on 144 samples
Epoch 1/14
63/63 [=====] - 227s 4s/step - loss: 0.7202 - accuracy: 0.5311 - val_loss: 0.6440 - val_accuracy: 0.6736
Epoch 2/14
63/63 [=====] - 223s 4s/step - loss: 0.6477 - accuracy: 0.6199 - val_loss: 0.6251 - val_accuracy: 0.7153
Epoch 3/14
63/63 [=====] - 221s 4s/step - loss: 0.6297 - accuracy: 0.6590 - val_loss: 0.6151 - val_accuracy: 0.6111
Epoch 4/14
63/63 [=====] - 224s 4s/step - loss: 0.5929 - accuracy: 0.6909 - val_loss: 0.5712 - val_accuracy: 0.7222
Epoch 5/14
63/63 [=====] - 221s 4s/step - loss: 0.5975 - accuracy: 0.6963 - val_loss: 0.5604 - val_accuracy: 0.7222
Epoch 6/14
63/63 [=====] - 220s 3s/step - loss: 0.5562 - accuracy: 0.7194 - val_loss: 0.5657 - val_accuracy: 0.6528
Epoch 7/14
63/63 [=====] - 220s 3s/step - loss: 0.5525 - accuracy: 0.7194 - val_loss: 0.5305 - val_accuracy: 0.7361
Epoch 8/14
63/63 [=====] - 226s 4s/step - loss: 0.5498 - accuracy: 0.7185 - val_loss: 0.5272 - val_accuracy: 0.7508
Epoch 9/14
63/63 [=====] - 221s 4s/step - loss: 0.5472 - accuracy: 0.7389 - val_loss: 0.5506 - val_accuracy: 0.7431
Epoch 10/14
63/63 [=====] - 220s 3s/step - loss: 0.5854 - accuracy: 0.7726 - val_loss: 0.5149 - val_accuracy: 0.7222
Epoch 11/14
63/63 [=====] - 220s 3s/step - loss: 0.4999 - accuracy: 0.7602 - val_loss: 0.5100 - val_accuracy: 0.7431
Epoch 12/14
63/63 [=====] - 220s 3s/step - loss: 0.5932 - accuracy: 0.7691 - val_loss: 0.5214 - val_accuracy: 0.7153
Epoch 13/14
63/63 [=====] - 221s 4s/step - loss: 0.5863 - accuracy: 0.7549 - val_loss: 0.5108 - val_accuracy: 0.7639
Epoch 14/14
63/63 [=====] - 222s 4s/step - loss: 0.4936 - accuracy: 0.7780 - val_loss: 0.4954 - val_accuracy: 0.7917
    
```

Figure 5: Training and testing dataset of the model

The plots are provided below in Figure 6 were generated from Matplot library. Visualizing data is one of

the best ways to humanize data to make it easy to understand and get the relevant trends from it. From the plot it shows that the model is not overfitting despite having less training data

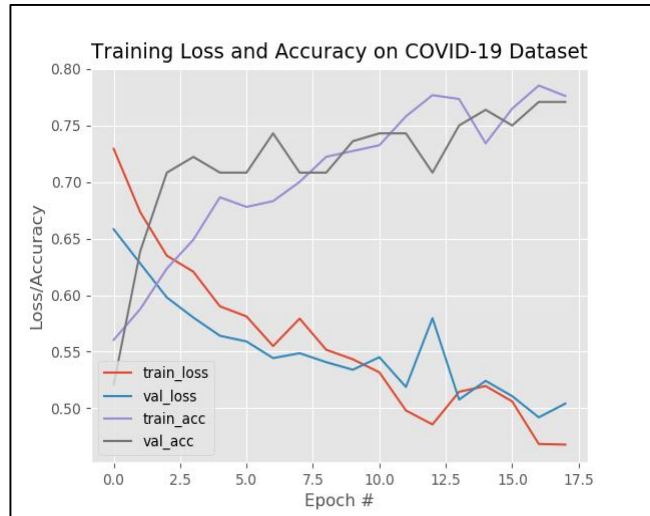


Figure 6: Training history plot of accuracy and loss curves.

A comparative study was done between Convolution Neural Network (CNN) and Logistic Regression (LogReg) to compare the accuracy achieved by both models. The same dataset was used in this study. Logistic regression is a simple form of a neural network that classifies data categorically. The “classic” application of logistic regression model is binary classification. Logistic regression takes an input, passes it through a function called sigmoid function then returns an output of probability between 0 and 1. This sigmoid function is responsible for classifying the input. The wrong classification by the sigmoid function is also known as weight (w) or loss. The comparative study was performed, and the accuracy of both studies were compared in Figure7.

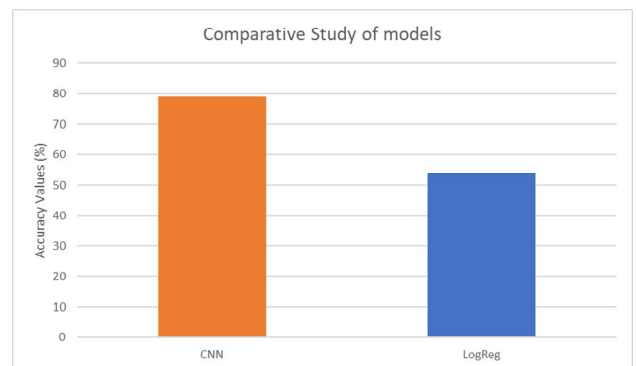


Figure 7: Comparative study of CNN and LogReg model

Both algorithms used the same dataset, the accuracy produced by the both algorithms were comparatively different from each other. The Convolution Neural Network algorithm has the higher accuracy of 79% compared to the Logistic Regression which only obtained 54% accuracy. The comparison was taken to find out the effectiveness of this CNN model compared with different algorithms like Logistic Regression. By achieving a higher accuracy, it proves the effectiveness of this model in classification of images.

5. CONCLUSION

In this study, a deep learning model was trained to detect COVID-19 in CT Scan images. This study was implemented on Python programming language. To build and train the Convolution Neural Network (CNN) model, Python Deep Learning libraries such as TensorFlow 2.0 and Keras had been utilized. As for the dataset, the open-source dataset of COVID-19 chest computed tomography (CT) images were used. This image where been confirmed by the senior radiologist who performed the diagnosis and treatment of COVID-19 patients. There were a total of 745 sampled images that belonged to two classes. 348 COVID-19 images positive and 397 COVID-19 images negative. Based on the training process, the model was able to detect with accuracy of 79% on the testing set. All the objectives of the study have been achieved by completing the study despite having very limited training data. The performance of the model, Convolution Neural Network were evaluated by comparing with Logistic Regression model. Findings from the research proves that Convolution Neural Network are reliable by producing higher accuracy rate of 79% while Logistic Regression produce a rate of 54%. However, in the future more reliable and quality image datasets should be used along with the metadata of the patients to train the model. By doing so, this can enhance the results of the model. In the future, larger local dataset image datasets should be used. Since COVID-19 is still new and given that the patients' rights and confidentiality it is harder to get local dataset with high quality images. But in coming few months, we can get the local datasets.

In addition, the results that are being detected by the model should be validated and evaluated by trained medical professional. As prediction models is very crucial process, high accuracy and reliability is important because missed diagnosis can cost lives.

Next, the metadata of the patients should be used as well to train the model. This is because this image data by itself is not sufficient to conclude the results. However, this model can be a baseline for anyone interested in continuing this topic of the 2019 new coronavirus

(COVID-19) to enhance the results to be more accurate and reliable. More experiments will be carried to do more comparisons with other standard algorithms to evaluate the performance of CNN algorithm.

ACKNOWLEDGMENT

This research is supported and funded by RIEC of University Malaysia Sarawak (UNIMAS), under the Cross Disciplinary Research (F04/CDRG/1839/2019).

REFERENCES

- [1] World Health Organization. (2020a). **Coronavirus**. Retrieved July 20, 2020, from World Health Organization website: <https://www.who.int/health-topics/coronavirus>
- [2] ICTV. (2019). **Naming the 2019 Coronavirus**. Retrieved July 20, 2020, from New Taxonomy Release website: <https://talk.ictvonline.org/>
- [3] Li, Q., Guan, X., Wu, P., Wang, X., Zhou, L., Tong, Y., ... Feng, Z. (2020). **Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus-Infected Pneumonia**. *New England Journal of Medicine*, 382(13), 1199–1207. <https://doi.org/10.1056/NEJMoa2001316>
- [4] Mahase, E. (2020). **Coronavirus: covid-19 has killed more people than SARS and MERS combined, despite lower case fatality rate**. *BMJ*, 368. <https://doi.org/10.1136/bmj.m641>
- [5] McKeever, A. (2020). **Here's what coronavirus does to the body**. Retrieved July 20, 2020, from National Geographic Society website: <https://www.nationalgeographic.com/science/2020/02/here-is-what-coronavirus-does-to-the-body>
- [6] Bai, Y., Yao, L., Wei, T., Tian, F., Jin, D.-Y., Chen, L., & Wang, M. (2020). **Presumed Asymptomatic Carrier Transmission of COVID-19**. *JAMA*, 323(14), 1406–1407. <https://doi.org/10.1001/jama.2020.2565>
- [7] Farooq, M., & Hafeez, A. (2020). **COVID-ResNet: A Deep Learning Framework for Screening of COVID19 from Radiographs**. Retrieved from <http://arxiv.org/abs/2003.14395>
- [8] Ai, T., Yang, Z., Hou, H., Zhan, C., Chen, C., Lv, W., ... Xia, L. (2020). **Correlation of Chest CT and RT-PCR Testing for Coronavirus Disease 2019 (COVID-19) in China: A Report of 1014 Cases**. *Radiology*, 296(2), E32–E40. <https://doi.org/10.1148/radiol.2020200642>

- [9] Fang, Y., Zhang, H., Xie, J., Lin, M., Ying, L., Pang, P., & Ji, W. (2020). **Sensitivity of Chest CT for COVID-19: Comparison to RT-PCR.** *Radiology*, 296(2), E115–E117. <https://doi.org/10.1148/radiol.2020200432>
- [10] Narin, A., Kaya, C., & Pamuk, Z. (2020). **Automatic Detection of Coronavirus Disease (COVID-19) Using X-ray Images and Deep Convolutional Neural Networks.** arXiv preprint arXiv:2003.10849.
- [11] Huang, C., Wang, Y., Li, X., Ren, L., Zhao, J., Hu, Y., ... Cao, B. (2020). **Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China.** *The Lancet*, 395(10223), 497–506. [https://doi.org/10.1016/S0140-6736\(20\)30183-5](https://doi.org/10.1016/S0140-6736(20)30183-5)
- [12] Evangeline, P. P., & Batri, K. (2019). **Detection of Lung Cancer by Machine Learning.** *International Journal of Engineering Research & Technology (IJERT)*, 8(09), 783–784.
- [13] Ginneken, B. (2017). **Fifty years of computer analysis in chest imaging: rule-based, machine learning, deep learning.** *Radiological Physics and Technology*, 10(1), 23–32. <https://doi.org/10.1007/s12194-017-0394-5>
- [14] Golan, R., Jacob, C., & Denzinger, J. (2016). **Lung nodule detection in CT images using deep convolutional neural networks.** *Proceedings of the International Joint Conference on Neural Networks*, 2016-October (1), 243–250. <https://doi.org/10.1109/IJCNN.2016.7727205>
- [15] Hua, K. L., Hsu, C. H., Hidayati, S. C., Cheng, W. H., & Chen, Y. J. (2015). **Computer-aided classification of lung nodules on computed tomography images via deep learning technique.** *OncoTargets and Therapy*, 8, 2015–2022. <https://doi.org/10.2147/OTT.S80733>
- [16] Kumar, D., Wong, A., & Clausi, D. A. (2015). **Lung Nodule Classification Using Deep Features in CT Images.** *Proceedings -2015 12th Conference on Computer and Robot Vision, CRV 2015*, 133–138. <https://doi.org/10.1109/CRV.2015.25>
- [17] Kooi, T., Litjens, G., van Ginneken, B., Gubern-Mérida, A., Sánchez, C. I., Mann, R., ... Karssemeijer, N. (2017). **Large scale deep learning for computer aided detection of mammographic lesions.** *Medical Image Analysis*, 35, 303–312. <https://doi.org/https://doi.org/10.1016/j.media.2016.07.007>
- [18] Madan, B., Panchal, A., & Chavan, D. (2019). **Lung Cancer Detection Using Deep Learning.** *SSRN Electronic Journal*, 266–271. <https://doi.org/10.2139/ssrn.3370783>
- [19] Sharif, S. M. N. and M. (2017). **Recent Developments in Computer Aided Diagnosis for Lung Nodule Detection from CT images: A Review.** *Current Medical Imaging*, Vol. 13, pp. 3–19. <https://doi.org/http://dx.doi.org/10.2174/1573405612666160610093453>
- [20] Guan, W., Ni, Z., Hu, Y., Liang, W., Ou, C., He, J., ... Zhong, N. (2020). **Clinical Characteristics of Coronavirus Disease 2019 in China.** *New England Journal of Medicine*, 382(18), 1708–1720. <https://doi.org/10.1056/NEJMoa2002032>
- [21] Wang, D., Hu, B., Hu, C., Zhu, F., Liu, X., Zhang, J., ... Peng, Z. (2020). **Clinical Characteristics of 138 Hospitalized Patients with 2019 Novel Coronavirus-Infected Pneumonia in Wuhan, China.** *JAMA*, 323(11), 1061–1069. <https://doi.org/10.1001/jama.2020.1585>
- [22] Lee, N., Hui, D., Wu, A., Chan, P., Cameron, P., Joynt, G. M., ... Sung, J. J. Y. (2003). **A major outbreak of severe acute respiratory syndrome in Hong Kong.** *The New England Journal of Medicine*, 348(20), 1986–1994. <https://doi.org/10.1056/NEJMoa030685>
- [23] Assiri, A., Al-Tawfiq, J. A., Al-Rabeeh, A. A., Al-Rabiah, F. A., Al-Hajjar, S., Al-Barrak, A., ... Memish, Z. A. (2013). **Epidemiological, demographic, and clinical characteristics of 47 cases of Middle East respiratory syndrome coronavirus disease from Saudi Arabia: a descriptive study.** *The Lancet. Infectious Diseases*, 13(9), 752–761. [https://doi.org/10.1016/S1473-3099\(13\)70204-4](https://doi.org/10.1016/S1473-3099(13)70204-4)
- [24] Wang, H., Xiao, X., Lu, J., Chen, Z., Li, K., Liu, H., ... Yang, Z. (2016). **Factors associated with clinical outcome in 25 patients with avian influenza A (H7N9) infection in Guangzhou, China.** *BMC Infectious Diseases*, 16(1), 534. <https://doi.org/10.1186/s12879-016-1840-4>
- [25] Li, X., Geng, M., Peng, Y., Meng, L., & Lu, S. (2020). **Molecular immune pathogenesis and diagnosis of COVID-19.** *Journal of Pharmaceutical Analysis*, 10(2), 102–108. <https://doi.org/10.1016/j.jpha.2020.03.001>
- [26] Pan, Y., Guan, H., Zhou, S., Wang, Y., Li, Q., Zhu, T., ... Xia, L. (2020). **Initial CT findings and temporal changes in patients with the novel coronavirus pneumonia (2019-nCoV): a study of 63 patients in Wuhan, China.** *European Radiology*, 30(6), 3306–3309. <https://doi.org/10.1007/s00330-020-06731-x>
- [27] Chung, M., Bernheim, A., Mei, X., Zhang, N., Huang, M., Zeng, X., ... Shan, H. (2020). **CT Imaging Features of 2019 Novel Coronavirus (2019-**

- nCoV**). *Radiology*, 295(1), 202–207. <https://doi.org/10.1148/radiol.2020200230>
- [28] Ooi, G. C., Khong, P. L., Müller, N. L., Yiu, W. C., Zhou, L. J., Ho, J. C. M., ... Tsang, K. W. T. (2004). **Severe Acute Respiratory Syndrome: Temporal Lung Changes at Thin-Section CT in 30 Patients**. *Radiology*, 230(3), 836–844. <https://doi.org/10.1148/radiol.2303030853>
- [29] Paul, R., Hawkins, S. H., Schabath, M. B., Gillies, R. J., Hall, L. O., & Goldgof, D. B. (2018). **Predicting malignant nodules by fusing deep features with classical radiomics features**. *Journal of Medical Imaging*, 5(01), 1. <https://doi.org/10.1117/1.jmi.5.1.011021>
- [30] Cruz-Roa, A., Gilmore, H., Basavanthally, A., Feldman, M., Ganesan, S., Shih, N. N. C., ... Madabhushi, A. (2017). **Accurate and reproducible invasive breast cancer detection in whole-slide images: A Deep Learning approach for quantifying tumor extent**. *Scientific Reports*, 7(1), 46450. <https://doi.org/10.1038/srep46450>
- [31] Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., & Thrun, S. (2017). **Dermatologist-level classification of skin cancer with deep neural networks**. *Nature*, 542(7639), 115–118. <https://doi.org/10.1038/nature21056>
- [32] Gulshan, V., Peng, L., Coram, M., Stumpe, M. C., Wu, D., Narayanaswamy, A., ... Webster, D. R. (2016). **Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs**. *JAMA*, 316(22), 2402–2410. <https://doi.org/10.1001/jama.2016.17216>
- [33] Carneiro, G., Oakden-Rayner, L., Bradley, A. P., Nascimento, J., & Palmer, L. (2017). **Automated 5-year mortality prediction using deep learning and radiomics features from chest computed tomography**. *Proceedings - International Symposium on Biomedical Imaging*, 130–134. <https://doi.org/10.1109/ISBI.2017.7950485>
- [34] Yang, X., Kwitt, R., Styner, M., & Niethammer, M. (2017). **Quicksilver: Fast predictive image registration – A deep learning approach**. *NeuroImage*, 158, 378–396. <https://doi.org/https://doi.org/10.1016/j.neuroimage.2017.07.008>
- [35] Chartrand, G., Cheng, P. M., Vorontsov, E., Drozdal, M., Turcotte, S., Pal, C. J., ... Tang, A. (2017). **Deep learning: A primer for radiologists**. *Radiographics*, 37(7), 2113–2131. <https://doi.org/10.1148/rg.2017170077>
- [36] Zaharchuk, G., Gong, E., Wintermark, M., Rubin, D., & Langlotz, C. P. (2018). **Deep learning in neuroradiology**. *American Journal of Neuroradiology*, 39(10), 1776–1784. <https://doi.org/10.3174/ajnr.A5543>
- [37] Samhitha, B. K., Mana, S. C., Jose, J., Vignesh, R., & Deepa, D. (2020). **Prediction of Lung Cancer Using Convolutional Neural Network (CNN)**. *International Journal Advanced Trends in Computer Science and Engineering*, 9(3), 3361–3365.
- [38] Reddy, A. V., Vege, H. K., & Prakash, K. B. (2020). **Efficient and accurate hybrid deep learning model for multimodal disease risk prediction**. *International Journal of Advanced Trends in Computer Science and Engineering*, 9(2), 1262–1267.
- [39] Phung, V. H., & Rhee, E. J. (2019). **A High-accuracy model average ensemble of convolutional neural networks for classification of cloud image patches on small datasets**. *Applied Sciences (Switzerland)*, 9(21). <https://doi.org/10.3390/app9214500>
- [40] K.Mohanambal, Y.Nirosha, E.Oliviya Roshini, S.Punitha, M. S. (2019). **Lung Cancer Detection Using Machine Learning Techniques**. *International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering*, 8(2), 6. <https://doi.org/DOI:10.15662/IJAREEIE.2019.0802011>
- [41] Alkhaleefah, M., & Wu, C. C. (2019). **A Hybrid CNN and RBF-Based SVM Approach for Breast Cancer Classification in Mammograms**. *Proceedings - 2018 IEEE International Conference on Systems, Man, and Cybernetics, SMC 2018*, 894–899. <https://doi.org/10.1109/SMC.2018.00159>