MEKATRONIKA
e-ISSN: 2637-0883
VOL. 3, ISSUE 2, 19 – 24
DOI: https://doi.org/10.15282/mekatronika.v3i2.7161

ORIGINAL ARTICLE

The Diagnosis of COVID-19 through X-ray Images via Transfer Learning and Fine-Tuned Dense Layer on Pipeline

Amiir Haamzah Mohamed Ismail[1], Mohd Azraai Mohd Razman[1], Ismail Mohd Khairuddin[1], Muhammad Amirul Abdullah[1], Rabiu Muazu Musa[2] and Anwar P.P. Abdul Majeed[1,3,4,5*]

[1]Innovative Manufacturing, Mechatronics and Sports Laboratory, Faculty of Manufacturing and Mechatronics Engineering Technology, Universiti Malaysia Pahang, 26600 Pahang, Malaysia.
[2]Centre for Fundamental and Continuing Education, Department of Credited Co-curriculum, Universiti Malaysia Terengganu, Terengganu. Malaysia
[3]Centre for Software Development & Integrated Computing, Universiti Malaysia Pahang, 26600 Pekan, Pahang, Malaysia
[4]Faculty of Engineering, Technology and Built Environment, UCSI University. Kuala Lumpur Campus. 56000 Cheras. Kuala Lumpur. Malaysia
[5]EUREKA Robotics Centre, Cardiff School of Technologies, Cardiff Metropolitan University, CF5 2YB. Cardiff. UK

ABSTRACT – X-ray is used in medical treatment as a method to diagnose the human body internally from diseases. Nevertheless, the development in machine learning technologies for pattern recognition have allowed machine learning of diagnosing diseases from chest X-ray images. One such diseases that are able to be detected by using X-ray is the COVID-19 coronavirus. This research investigates the diagnosis of COVID-19 through X-ray images by using transfer learning and fine-tuning of the fully connected layer. Next, hyperparameters such as dropout, p, number of neurons, and activation functions are investigated on which combinations of these hyperparameters will yield the highest classification accuracy model. InceptionV3 which is one of the common neural network is used for feature extraction from chest X-ray images. Subsequently, the loss and accuracy graphs are used to find the pipeline which performs the best in classification task. The findings in this research will open new possibilities in screening method for COVID-19.

ARTICLE HISTORY
Received: 14th May 2021
Revised: 25th June 2021
Accepted: 29th July 2021

KEYWORDS
InceptionV3
Transfer learning
Hyperparameter
Dropout
OVAT

INTRODUCTION

As a consequence of the ever-rising cases of COVID-19 [1], it appears proper testing method are needed in order to curb the spreading of this virus. The most common test kit for COVID-19 is the polymerase chain reaction (RT-PCR) kits which is designed to detect SARS-CoV-2 genetically. However, this type of test kits was reported giving false positive and false negative detections. Further COVID-19 diagnostic is emerging such as CRISPR, LAMP, RT-LAMP to improve the current testing method but this method will take time to produce the results and will suffer from shortages and high price. Successively, in Hubei, patients were scanned using computed tomography (CT) as COVID-19 patients exhibits pneumonia like infections as stated by Udugama et al. [2]. Interestingly, the use of deep learning is possible to distinguish between healthy patients and COVID-19 patients by analysing the CT scan or lung X-ray images. This will introduce endless possibility for future detection of Covid-19 in terms of convenience, cost, accuracy, and reliability.

Machine learning may further expedite the diagnosis of COVID-19 from X-ray images and in turn bringing instant results, help with understaff hospitals and be able to apply at any health facilities because the wide availability of X-ray machines. This work aims to investigate the diagnosis of COVID-19 through X-ray images by using transfer learning and fine-tuning the fully connected layers so that the performance is comparable or better with pre-existing models. The specific objectives are formulated in order to achieve the overall aim of the study are to evaluate the performance of the different Transfer Learning (TL) models extracting features from the images and to identify the suitable hyperparameters of the fully connected layers in the classification of COVID-19 cases with non-COVID-19 cases from the Transfer Learning (TL) based extracted features.

It is worth to note that the achievements of the aforesaid objectives are not mutually exclusive as the overall framework shall determine which combination of the models that will yield the best classification for the disease.

RELATED WORK

An effort in improving the performance of CNN to predict the likelihood of COVID-19 have been undertaken by Heidari et al. [3] using chest X-ray images with pre-processing algorithms. Unlike other researchers, the authors utilized a vast amount of dataset; 8474 chest X-rays which consisted of 415 Covid-19 images, 5179 non Covid-19 pneumonia, and 2880 normal chest X-ray images. Subsequently, a CNN based model was used, VGG 16, which correlates for its performance winning the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) using 14 million image

*CORRESPONDING AUTHOR | Anwar P.P. Abdul Majeed | amajeed@ump.edu.my

datasets. Adam optimizer was used in training the model and two improve the accuracy, the authors added two image pre-processing steps and generates a pseudo colour image. A strong relationship between image pre-processing and classification accuracy has been reported in the literature with 94.5% overall accuracy and 88% accuracy without using the two pre-processing steps.

In another major study, Panwar et al. [4] have identified the application of deep learning for fast detection of COVID-19 in X-Rays using the developed NCOVnet which refers to the CNN type model of VGG16. The datasets used in this research are open source; 142 images of COVID-19 positive patients, for normal patients the authors used Kaggle's Chest X-Ray Images (pneumonia) consisting of 5863 images. These images are divided into 'Normal', 'Bacteria Pneumonia', and 'Viral Pneumonia'. 70% of the data were used for training purpose while 30% of data for testing purpose. The CNN based VGG16 model was used to extract the features from the datasets. Subsequently, hyperparameter optimization applied using Adam optimizer, CNN learning rate, known epochs, CNN maximum number of iterations, and number of images covered in one iteration. The results obtained were rated for their performance by using confusion matrix, Area Under Curve (AUC) of Receiver Operating Characteristics (ROC), and accuracy. Overall, the proposed model was able to detect a COVID-19 positive patient with an overall accuracy of 88.10% with reference to the confusion matrix.

Vaid et al. [5] draws our attention to deep learning COVID-19 detection bias often observed in artificial intelligence accuracy in detecting COVID-19 in chest X-ray images. The authors utilized datasets from public data with an overall healthy to COVID-19 ratio of 1:2 to be used for the modified VGG 19 model. Accordingly, the modified model was trained previously on ImageNet. Subsequently, CNN was used in extraction of features from the training images and optimized by using early stopping and learning rate reduction. The results of this study show that 96.3% accuracy was achieved although the limitations dealing with new strain of virus are present.

PROPOSED ARCHITECTURE FOR TRANSFER LEARNING BASED CLASSIFICATION

Dataset

The X-ray images for COVID-19 patients and non-COVID-19 patients were retrieved from the Extensive COVID-19 X-Ray and CT Chest Images Dataset, curated by Walid El-Shafai, and Fathi Abd El-Samie [6]. The aforementioned dataset consists of 17099 X-ray and CT images of patients with and without COVID-19 infection. The linked dataset is extended with various augmentation techniques to increase the variety for the transfer learning (TL) model. For the purpose of this research, the directory for X-ray images was used. Subsequently, the folder contains folders separated into Covid, 4044 X-ray images: and Non-Covid, 5500 X-ray images. It is crucial for the transfer learning (TL) to be unbiased in order to produce valid classification accuracy (CA). Therefore, the dataset is further managed into 4000 X-ray images, split into 2000 Covid and 2000 non-Covid. The pre-processed images are shown in Figure 1(a) and (b), which shows the distinction between the two categories of dataset.

(a) (b)

Figure 1. (a) COVID-19 infected lung (b) Normal condition lung.

InceptionV3

InceptionV3 attained more than 78% accuracy on the ImageNet dataset which makes it widely used as an image recognition model. The model is made up of several symmetric and asymmetric building blocks as shown in Figure 2.

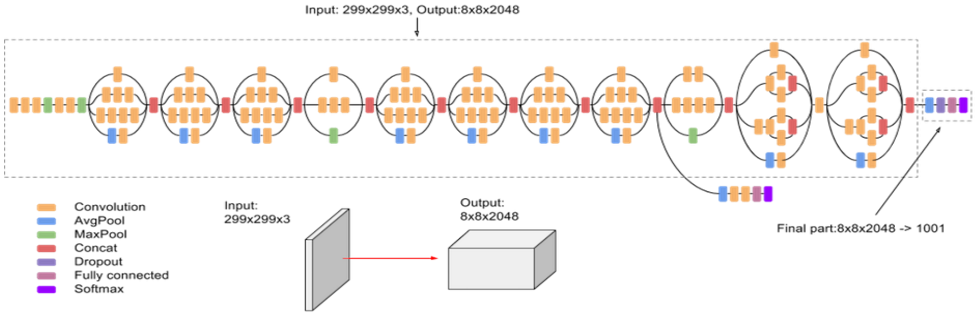

Figure 2. InceptionV3 architecture. Source: [7]

Performance metric

Since the dataset used for this research is balance, classification accuracy (CA) is best suited to show the performance of the proposed model because a balance dataset should avoid training bias. CA shows what percentage of predictions were correct and can be calculated as shown in Equation 1.

$$CA = \frac{(TP + TN)}{(TP + TN + FP + FN)} \tag{1}$$

One-value-at-a-time (OVAT) research workflow

This research is generally divided into three phases. Therefore, the proposed architecture will be synthesized after the best performing classification model is acquire by following the one-value-at-a-time (OVAT) strategy. The proposed architecture is developed by using Spyder version 5.0.0 an integrated development environment software which is initialized by using Anaconda Navigator version 2.0.3. An environment created in Anaconda with Python version 3.6.13. Next, GPU-accelerated deep learning is exploited for this research. Thus, modules installed in Spyder are keras-gpu version 2.3.1, tensorflow-gpu version 2.1.0, cudnn version 7.6.5, and cudatoolkit version 10.1.243. The computation specifications used in this research are Windows 10 Home Single Language, AMD Ryzen 5 3600 6-Core Processor @ 3.60GHz, NVIDIA GeForce RTX 2060, 16 GB DDR4 installed RAM, and Team Group GX2 2.5" 512GB SATA III Internal Solid-State Drive.

In this present study, Phase 1 is defined in the terms of the base model as the feature extraction from X-ray images. Initially, all the required libraries were imported which mainly consists of the Keras and Scikit modules. Then, the datasets path was specified and loaded into the model. Subsequently, labels were assigned to each class with label 0 for non_COVID class and label 1 for COVID class. Moreover, data splitting is important as it affects the outcome of the research. In this study, the dataset was split into a 70:30 ratio which consist of 70% for training set while 30% was further split into 15% validation set and 15% test set. Moreover, a single numpy array was created for the loaded images and the labels were also converted to numpy array. Then, the numpy array was normalize into 0 and 1 range only to match the classes defined prior. Consequently, the chosen learning model, either VGG16, VGG19, or InceptionV3, was loaded with Imagenet weights and the dense layer removed as this model is used for feature extraction only. In order to be productive, the array of extracted features and the matching array of expected values were saved into a file and will be loaded for Phase 2.

In Phase 2, this is where the hyperparameter tuning is executed which saves the author's time as the base model in Phase 1 would have been loaded only once because the features from the X-ray images were already extracted. The layers added in the dense layer are based on the sequential model with hyperparameters based on the required values. Then, the model is compiled, and Adam optimizer was used with model trained for 50 epochs and batch size of 10. Subsequently, the model is evaluated in terms of train accuracy and validation accuracy. Next, the weights and the model were saved for future references for that pipeline. The classification report and loss and accuracy graphs were generated at the end of the simulation for Phase 3.

Finally, Phase 3 is where all the data is sorted and analysed. It is noted that the variables manipulated for this research is the dropout rate, *p* ranging from 0.0 until 1.0, the number of neurons in a logarithmic sequence of 10, 100, 1000, and the activation functions of sigmoid, tanh, ReLu. Best performing pipelines are evaluated based on the loss and accuracy graphs for each pipeline.

EXPERIMENTAL RESULTS

Dropout, *p*

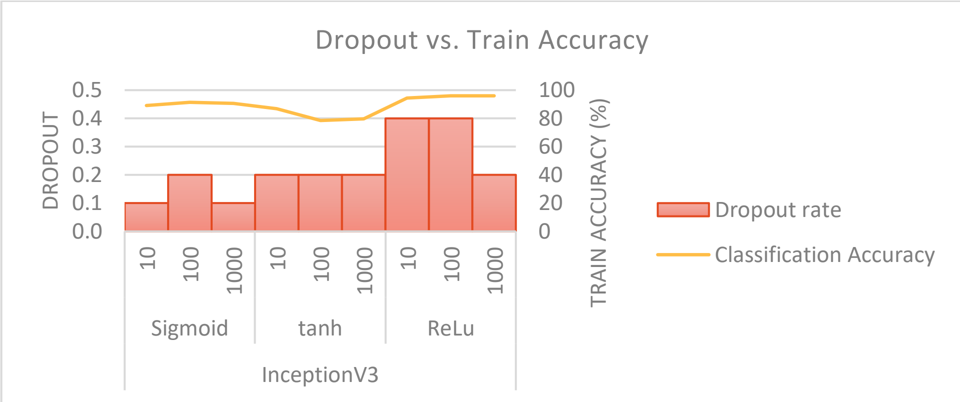

Figure 3. Dropout, *p* evaluated.

The loss and accuracy graphs of the various dropout-based pipeline are evaluated for selecting the suitable pipeline which exhibits low underfitting and overfitting properties. Consequently, the selected pipelines are shown in Figure 3 for the best performing. The data shows a trend of dropout, p = 0.5 and lower are more suitable in the dense layer. These findings are in line with theories that large dropout, p would not prevent overfitting because not enough dropout, p is able to be deliver which agrees with the ideas from Srivastava et al. [8]

Number of neurons

The data gained from the experiments confirms the effects when tuning the number of neurons in the dense layer. However, there are no definitive number of neurons to be defined for particular dense layer as stated by Arifin et al. [9]. Hence, a logarithmic increment approach is implemented with 10, 100, and 1000 number of neurons to estimate the suitable number of neurons. The experiment results shown in Figure 4 indicates that lower number of neurons are preferable. The data from the loss and accuracy graphs generated proves the view by Sheela and Deepa [10] that asserted that excessive hidden neurons caused overfitting.

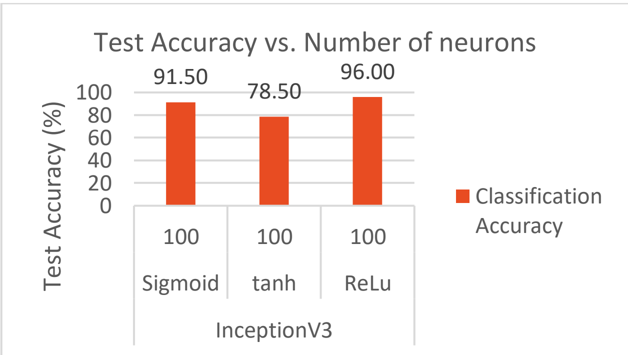

Figure 4. Number of neurons effects on test accuracy.

Activation function

Three activation functions were evaluated based on their pipelines as illustrated in Figure 4. The sigmoid-100-0.2 pipeline has a training accuracy of 93.68% , validation accuracy of 89.70%, and test accuracy of 91.50%. Next, the tanh-100-0.2 pipeline has a training accuracy of 80.54%, validation accuracy of 78.20%, and test accuracy of 78.50%. Subsequently, the ReLu-100-0.4 pipeline has a training accuracy of 99.10%, validation accuracy of 94.50%, and test accuracy of 96%. Thus, the ReLu-100-0.4 pipeline is chosen as the best performing pipeline.

Best performing pipeline

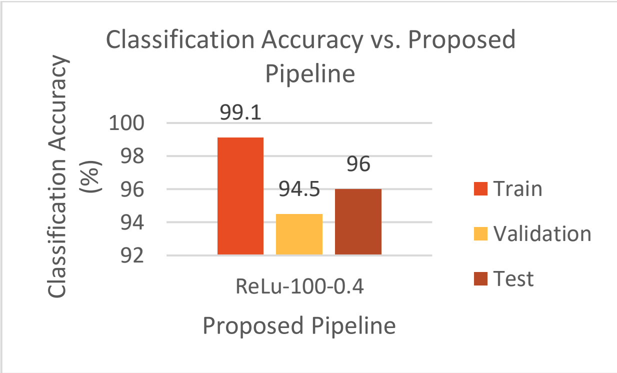

Figure 5. Best pipeline: InceptionV3-ReLu-100-0.4.

To recap, this study determines to find the best performing model by tuning the hyperparameters and evaluating the classification abilities. Overall, the prior findings presented in this paper shows that by using InceptionV3 to extract features from chest X-ray images, ReLu as activation function, 100 neurons in the dense layer and additionally p = 0.4 in the dropout layer will produce the highest performance in classification task. The chart in Figure 5 shows the proposed model has 99.1% testing accuracy, 94.5% validation accuracy, and 96% testing accuracy. The proposed model ability to make prediction is further demonstrated by the confusion matrix in Figure 7 in classifying non-COVID-19 (label 0) X-ray images and COVID-19 (label 1) X-ray images. Figure 6 shows that 285 images correctly classified as non-COVID-19 and 15 images were misclassified. Consequently, the COVID-19 class has 9 images which were misclassified but 291 images still correctly been classified as COVID-19. In short, the proposed model shows competitive performance compare to other models in detection ability of COVID-19 in chest X-ray images, shown in Table 1. The model by Vaid et al. [5] is marginally better than the proposed model. However, future works can help to increase the performance of the present model.

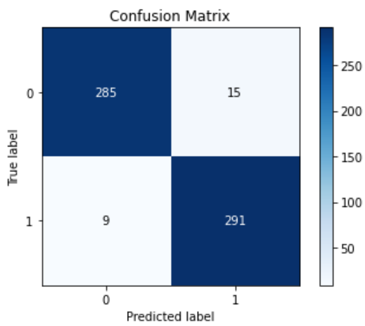

Figure 6. Confusion matrix for testing dataset.

Table 1. Comparison between present and past work.

Work	Learning models	Classification Accuracy, CA
Present study	InceptionV3	96%
Heidari et al. 2020	VGG16	94.5%
Panwar et al. 2020	VGG16	88.10%
Vaid et al. 2020	VGG19	96.3%

CONCLUSION

The present study effectively explained the effect of hyperparameter tuning in the fully connected layer in finding the best performing model. This study successfully shows that InceptionV3-ReLu-100-0.4 is the best pipeline for this research configuration. Future works may implement new hyperparameters to be tuned such as the learning rate in the fully connected layer in increasing the classification accuracy. The results and data of the present study are not bias to any party or institution, and solely based on the actual findings observed in the present research.

ACKNOWLEDGEMENT

The author would like to thank the lecturers of UMP for supervising this work and sharing the expertise in improvements for this research.

REFERENCES

[1] World Health Organization, "WHO Coronavirus Disease (COVID-19) Dashboard With Vaccination Data | WHO Coronavirus (COVID-19) Dashboard With Vaccination Data," *World Health Organization*, 2021. https://covid19.who.int/ (accessed Jun. 08, 2021).

[2] B. Udugama *et al.*, "Diagnosing COVID-19: The Disease and Tools for Detection," *ACS Nano*, vol. 14, no. 4, pp. 3822–3835, 2020, doi: 10.1021/acsnano.0c02624.

[3] M. Heidari, S. Mirniaharikandehei, A. Z. Khuzani, G. Danala, Y. Qiu, and B. Zheng, "Improving the performance of CNN to predict the likelihood of COVID-19 using chest X-ray images with preprocessing algorithms," *Int. J. Med. Inform.*, vol. 144, no. September, p. 104284, 2020, doi: 10.1016/j.ijmedinf.2020.104284.

[4] H. Panwar, P. K. Gupta, M. K. Siddiqui, R. Morales-Menendez, and V. Singh, "Application of deep learning for fast detection of COVID-19 in X-Rays using nCOVnet," *Chaos, Solitons and Fractals*, vol. 138, p. 109944, 2020, doi: 10.1016/j.chaos.2020.109944.

[5] S. Vaid, R. Kalantar, and M. Bhandari, "Deep learning COVID-19 detection bias: accuracy through artificial intelligence," *Int. Orthop.*, vol. 44, no. 8, pp. 1539–1542, 2020, doi: 10.1007/s00264-020-04609-7.

[6] F. El-Shafai, Walid; Abd El-Samie, "Extensive COVID-19 X-Ray and CT Chest Images Dataset", Mendeley Data, V3," 2020. https://data.mendeley.com/datasets/8h65ywd2jr/3 (accessed Dec. 29, 2020).

[7] Google Cloud TPU, "Advanced Guide to Inception v3 on Cloud TPU | Google Cloud," *Google*, 2016. https://cloud.google.com/tpu/docs/inception-v3-advanced#learning_rate_adaptation (accessed Jan. 09, 2021).

[8] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A Simple Way to Prevent Neural Networks from Overfitting," *J. Mach. Learn. Res.*, vol. 15, pp. 1929–1958, Jun. 2014.

[9] F. Arifin, H. Robbani, T. Annisa, and N. N. M. I. Ma'Arof, "Variations in the Number of Layers and the Number of Neurons in Artificial Neural Networks: Case Study of Pattern Recognition," *J. Phys. Conf. Ser.*, vol. 1413, no. 1, 2019, doi: 10.1088/1742-6596/1413/1/012016.

[10] K. G. Sheela and S. N. Deepa, "Selection of number of hidden neurons in neural networks in renewable energy systems," *J. Sci. Ind. Res. (India).*, vol. 73, no. 10, pp. 686–688, 2014.