2020 11th International Conference on Electrical and Computer Engineering (ICECE)

A Comparative Study of CNN Transfer Learning Classification Algorithms with Segmentation for COVID-19 Detection from CT Scan Images

Ashek Seum, Amir Hossain Raj, Shadman Sakib, Tonmoy Hossain
Department of Computer Science and Engineering
Ahsanullah University of Science and Technology
Dhaka, Bangladesh
Email: asheksekseum86@gmail.com, raj.amir@outlook.com, shadmsakib@gmail.com, tonmoy.cse@aust.edu

***Abstract*—After it's inception, COVID-19 has spread rapidly all across the globe. Considering this outbreak, by far, it is the most decisive task to detect early and isolate the patients quickly to contain the spread of this virus. In such cases, artificial intelligence and machine learning or deep learning methods can come to aid. For that purpose, we have conducted a qualitative investigation to inspect 12 off-the-shelf Convolution Neural Network (CNN) architectures in classifying COVID-19 from CT scan images. Furthermore, a segmentation algorithm for biomedical images – U-Net, is analyzed to evaluate the performance of the CNN models. A publicly available dataset (SARS-COV-2 CT-Scan) containing a total of 2481 CT scan images is employed for the performance evaluation. In terms of feature extraction by excluding the segmentation technique, a performance of 88.60% as the F1 Score and 89.31% as accuracy is achieved by training DenseNet169 architecture. Adopting the U-Net segmentation method, we accomplished the most optimal accuracy and F1 Scores as 89.92% and 89.67% respectively on DenseNet201 model. Furthermore, evaluating the performances, we can affirm that a combination of a Transfer Learning architecture with a segmentation technique (U-Net) enhances the performance of the classification model.**

***Index Terms*—COVID-19, Transfer learning, CNN, CT scan, DenseNet, U-Net**

I. Introduction

The Coronavirus Disease 2019 (COVID-19) is an ongoing pandemic. Thousands of people are being affected by this highly contagious disease every day. To defend from the impacts of this virus, early detection is very crucial and a paramount task for saving the lives of millions. Up until now, the Reverse Transcription-Polymerase Chain Reaction (RT-PCR) test is the reliable gold standard test that is being used to diagnose COVID-19 disease [1]. However, the RT-PCR test is both costly and time-consuming. Hence, there is an indisputable need for diagnosing this disease in a much more efficient and economical way. The detection of COVID-19 from CT scan images is a challenging task. One of the aspects that makes the task difficult is that there is similarity in the affected region with other diseases such as pneumonia, pulmonary infections, etc. Also, bacterial and viral pneumonia have a similar effect on the lungs as COVID-19. Therefore, it becomes rather difficult to distinguish between these diseases and COVID-19 from CT scans. For this reason, deep learning is very useful in the classification of COVID and non-COVID from the other lung centered diseases. Another factor that makes pre-trained deep learning models useful in this task is the scarcity of a large dataset of COVID positive CT scans on which a CNN model could be trained from scratch. Furthermore, to efficiently work on deep learning models, a considerable amount of images are needed to train the model. For that purpose, Transfer Learning architectures come in handy as they can train a deep network with little data and produce good results [2].

In this study, we have conducted two types of feature extraction techniques based on the presence of segmentation techniques. It is necessary to exclude the redundant areas or pixels of a CT scan image as it can degrade the performance of the model. To work on this issue, the U-Net architecture is adopted which works accurately in the segmentation tasks of biomedical images [3]. It is by far the most commonly used technique for the segmentation of the lung regions in COVID detection studies [4].

This study represents a comparative analysis between the pre-trained models which can be beneficial to determine the future roadmap to build a COVID-19 classification model. Also, it will be helpful for the medical officials to exploit the outcome from these automated architectures before the RT-PCR test in the decision-making process. The rest of this paper is organized as follows. Section II provides the literature review on various deep learning systems developed for COVID-19 detection. This is followed by presenting the proposed approach of COVID-19 classification for the comparative analysis. Finally, section IV provides a discussion of the experiment results.

II. Related Work

Since the beginning of this pandemic, researchers in the field of Computer Vision have utilized different tools focusing on early detection of this virus. Das et al. [5] proposed a truncated inception model for the detection of COVID-19 from Chest X-ray (CXR) images. They applied a transfer learning technique using a modified InceptionNetV3 with pre-trained ImageNet weights. Then, a multiclass classification approach is carried out in this work by detecting COVID-19 positive, pneumonia positive, and tuberculosis positive. A deep learning approach based on a Generative Adversarial Network (GAN) is introduced by Loey et al. [6] for the detection of COVID-19 from Chest X-rays. Three CNN-based models — AlexNet, GoogleNet, and ResNet-18, are used for this purpose. Experimentation is done on several scenarios by using four, three, or two classes for classification. In another work, Miranda et al. [7], developed an approach for COVID-19 detection from Chest X-Ray using hierarchical and multiclass views. A new dataset called RYDLS-20 is introduced, which contains CXR images from different pathogen-induced pneumonia and CXR images of healthy lungs.

Xuehai et al. [8] proposed a Self-Trans approach that integrates self-supervised learning with transfer learning subsiding the risk of overfitting to address the problem of small datasets of COVID-19 CT images. Chen et al. [3] proposed a novel algorithm for the segmentation of COVID-19 infected areas using Aggregated Residual Transformations for measuring the infection area in CT images caused by COVID-19. An encoder-decoder architecture with a multi-layer perceptron is proposed by Amyar et al. [9] for jointly identifying COVID-19 patients by segmenting infected areas from CT images. A simple convolutional neural network and modified pre-trained AlexNet are put on test on both X-ray and CT scan images by Maghdid et al. [10].

In the existing works, several different approaches have been explored for detecting COVID-19 disease. Moreover, a comparison of two or three different models has been carried out in some of these studies. However, there are various significant models that have been left unexplored. Some of these models may perform better at this automatic detection task than the prevalent ones. Hence, in this work, several such models have been explored trying to determine the best fit for the task of COVID-19 classification from CT scan images. A thorough comparison between the models is performed for both excluding and including a segmentation approach to comprehensively understand the possible roadmap of automatic COVID-19 classification.

III. COVID-19 Classification Architecture

The primary objective of this study is to evaluate different pre-trained deep learning architectures available against a COVID-19 CT scan dataset and analyze which one performs better in classification task. The steps can be divided as: Image Pre-processing and Segmentation, Feature Extraction and Classification, and Performance Evaluation. In Fig. 1, a basic architecture of COVID-19 classification is depicted.

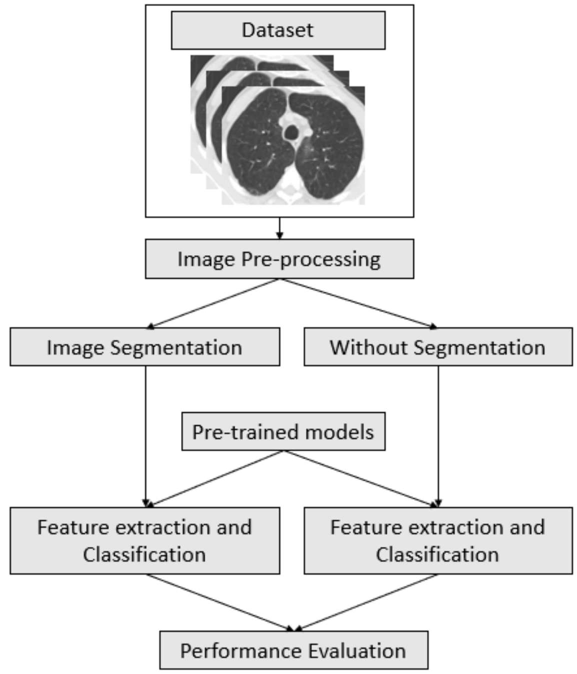

Fig. 1: A basic architecture of COVID-19 classification

A. Image Pre-processing and Segmentation

Before using these models without adopting the segmentation technique, pre-processing was carried out on the images. The images were resized to $224 \times 224 \times 3$ dimensions for all the models except Inception_v3. For Inception_v3, they were resized to $299 \times 299 \times 3$ dimension as per the requirement of the model. The images were then normalized, which helps the models to learn faster and better. In the case of segmentation, the images were first resized to $256 \times 256 \times 3$ dimensions. After resizing, normalization was applied to them. After pre-processing the images, they were fed into the U-Net architecture with ImageNet weights to get the segmented images.

B. Feature Extraction and Classification

In this study, 12 different pre-trained deep learning models are employed to evaluate the efficiency of the architectures to work on the classification of COVID-19 disease. These models are: AlexNet, VGG16, VGG19, ResNet18, ResNet50, ResNet101, ResNet152, DenseNet121, DenseNet169, DenseNet201, Inception_v3 and GoogleNet. In the AlexNet model, data augmentation and dropout was employed to reduce the overfitting of the model [11]. VGG is one of the most used pre-trained architectures for different computer vision tasks with different variations. All the variants end with three fully connected layers [12]. Residual Networks (ResNet) is one of the complex architectures, which truly defines how convoluted a deep learning architecture can be. ResNet consists of multiple subsequent residual modules, which are the basic building block of ResNet architecture [13]. Inception network or GoogleNet is a CNN architecture designed by the researchers from Google. They made a novel approach called the inception module, which introduces a drastic change from the sequential architectures which we observed previously [14]. In DenseNet architecture, a dense block where all the CNN layers are connected with all descendant layers. The major advantages of this architecture are strong gradient flow, parameter, and computational efficiency, diversified, and low complexity features [15]. Fundamentally, the U-Net architecture is a convolutional neural network architecture that is mainly developed for the segmentation of biomedical images, which uses upsampling and downsampling to learn features [16].

The models mentioned above are all capable of classifying up to 1000 different classes. So, for our experiment, we replaced the classification layer or the last Fully Connected (FC) layer with a distinct layer for the purpose of our experiment. The number of outputs of this new layer is equal to the binary classes of the CT scan dataset on which the model is trained. The number of input features in these newly constructed last FC layers remained the same as before.

C. Performance Evaluation

During the training phase, we used Stochastic Gradient Descent (SGD) optimizer with a learning rate of 0.001 and momentum 0.9. Furthermore, we used 25 epochs for training each pre-trained model on the dataset. A learning rate scheduler with the decay of the learning rate by a factor of 0.1 in every seven epochs was operated. Then, a validation set was employed during the training process. The best performing model on the validation set was selected for final testing. To evaluate the performances of the models, the metrics used were – sensitivity, specificity, precision, F1 Score, and accuracy.

IV. Performance Analysis and Discussions

In this section, the performance of the above mentioned models on the CT scan dataset along with their performance comparison is discussed.

A. Dataset Preperation

SARS-COV-2 CT-Scan dataset from Kaggle is used for experiments in this study [17]. It is a publicly available dataset that contains data collected from patients from Sau Paulo, Brazil. There are a total of 2481 CT scan images in this dataset, out of which 1252 CT scans are from COVID positive patients and 1229 CT scans are from COVID negative patients. Fig. 2 represents samples from each category (COVID positive and negative). The dataset was split in train, validation, and test set at a ratio of 70:10:20. This means that 70% data is used for training and 10% is used for validation, and the rest 20% for testing purpose. The distribution is shown in TABLE I.

TABLE I: A distribution on train, validation, and test sets of SARS-COV-2 dataset.

	Number of CT scan images		
Sets	**COVID**	**Non-COVID**	**Total**
Train	876	860	1736
Validation	126	123	249
Test	250	246	496

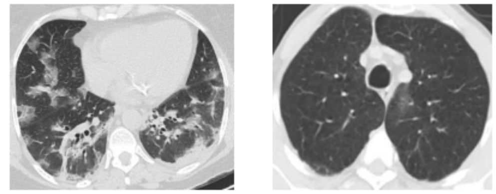

Fig. 2: COVID positive (left) and COVID negative (right) samples

B. Evaluation metrics

Classification performance of the pre-trained models are evaluated on the testing data based on the following parameters where TP, TN, FP, and FN defined as the True Positive, True Negative, False Positive, and False Negative respectively.

1) Sensitivity: It indicates the proportion of positives that are correctly classified. That is the correct detection COVID-19 CT scans.

$$Sensitivity = \frac{TP}{TP + FN}$$

2) Specificity: It indicates what proportion of the negatives are correctly classified. That is the correct detection Non-COVID CT scans.

$$Specificity = \frac{TN}{TN + FP}$$

3) Precision: It defines the probability that the COVID positive classification made by the classifier is indeed COVID patients.

$$Precision = \frac{TP}{TP + FP}$$

4) F1 Score: It is the weighted average of precision and recall. It takes into account both precision and recall when evaluating performance.

$$F1\ Score = \frac{2 \times TP}{2 \times TP + FP + FN}$$

5) Accuracy: It is the ratio of correct predictions to total predictions. It measures how often the classifier predicts the correct result.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

C. Experimental Results

Classification performance of transfer learning models without segmentation based on aforementioned parameters are stated in TABLE II.

From the experiment, it can be seen that DenseNet121 scores the highest precision – 98.31%. It means that the DenseNet121 model detected the most COVID-19 scans out of all the positives. ResNet18 and DenseNet121 get the most optimal scores in terms of sensitivity and specificity – 90.80% and 98.78%, respectively. But we find that, based on overall accuracy, DenseNet169 outperforms all other models with an F1 Score of 88.60% and accuracy of 89.31%.

TABLE II: Performance of transfer learning models without segmentation on test data.

Model	Sensitivity	Precision	Specificity	F1	Accuracy
AlexNet	82.0%	80.08%	79.27%	81.03%	80.65%
VGG 16	68.40%	95.53%	96.75%	79.72%	82.46%
VGG 19	71.60%	89.95%	91.87%	79.73%	81.65%
ResNet18	**90.80%**	75.67%	70.33%	82.55%	80.65%
ResNet50	68.40%	96.61%	97.56%	80.09%	82.86%
ResNet101	81.20%	95.31%	95.93%	87.69%	88.51%
ResNet152	76.80%	92.31%	93.50%	83.84%	85.08%
DenseNet121	70.00%	**98.31%**	**98.78%**	81.78%	84.27%
DenseNet169	82.40%	95.81%	96.34%	**88.60%**	**89.31%**
DenseNet201	74.40%	97.89%	98.37%	84.55%	86.29%
Inception_v3	75.60%	95.94%	96.75%	84.56%	86.09%
GoogleNet	82.80%	81.18%	80.49%	81.98%	81.65%

TABLE III: Performance of transfer learning models with segmentation on testing data.

Model Name	Sensitivity	Precision	Specificity	F1 Score	Accuracy
VGG 16	83.20%	83.20%	82.93%	83.20%	83.06%
VGG 19	84.80%	77.09%	74.39%	80.76%	79.64%
ResNet18	80.40%	**99.50%**	**99.59%**	88.94%	**89.92%**
ResNet50	80.80%	92.66%	93.50%	86.32%	87.10%
ResNet101	82.80%	95.83%	96.34%	88.84%	89.52%
ResNet152	77.20%	91.90%	93.09%	83.91%	85.08 %
DenseNet121	79.60%	98.03%	98.37%	87.86%	88.91%
DenseNet169	79.20%	98.51%	98.78%	87.80%	88.91%
DenseNet201	**86.80%**	92.74%	93.09%	**89.67%**	**89.92%**

Classification performance of the models with segmentation is stated in TABLE III. From this experiment of including the segmentation approach we see that, ResNet18 here scores the best precision, specificity, and accuracy – 99.50%, 99.59%, and 89.92% respectively whereas DenseNet201 gets best scores in F1 Score, sensitivity, and accuracy – 86.80%, 89.67% and 89.92% respectively.

TABLE IV: Comparison between best performing models before and after segmentation.

Subjects	Model Name	F1 Score	Accuracy
Without segmentation	DenseNet169	88.60%	89.31%
With Segmentation	ResNet18	88.94%	89.92%
	DenseNet201	89.67%	89.92%

Without segmentation, the ResNet18 scored 82.55% and 80.65% respectively in F1 Score and accuracy. These increase to 88.94% and 89.92%, respectively, after segmentation was applied. For DenseNet201, F1 Score and accuracy were 84.55% and 86.29% respectively before segmentation. These enhanced to 89.67% and 89.92% after segmentation, respectively. Both of the optimal performing models after segmentation scored better than the best performing model without adopting segmentation. The performance accuracies before and after segmentation are depicted in TABLE IV.

D. Comparative Analysis

The results in TABLE II and III suggest that accuracy and F1 Score increases for most of the models after segmentation was applied on the dataset. A comparison can be seen in TABLE IV between the best performing models based on the adoption of the segmentation technique.

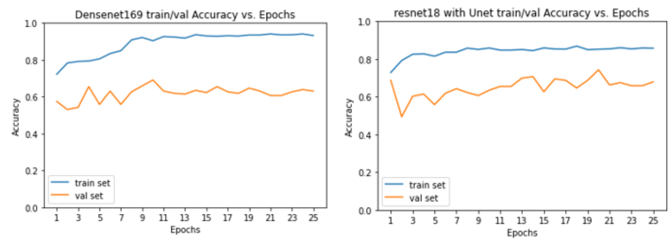

Fig. 3: Train and validation accuracy vs epochs in the best performing models without (left) and with (right) segmentation

All of these results were based on the test dataset. During the training phase, the model that performed best on the validation dataset was chosen for the evaluation of the test set. Fig. 3 shows the behavior of the best performing models during the training and validation phase.

V. CONCLUSIONS

In this study, a comparative analysis is carried out based on some of the pre-trained convolutional neural network architectures to classify between COVID and Non-COVID classes. The comparison is done in two phases based on the adoption of segmentation. In the first phase, the pre-trained models are first trained on the training dataset without segmentation and then the trained model is used on the test dataset to classify the images. Here, DenseNet169 outperformed all the existing models in F1 Score and accuracy with scores of 88.60% and 89.31% respectively. In the second phase, the experiments continued with segmentation using the U-Net algorithm. Here, ResNet18 and DenseNet201 both scored the optimal accuracy as 89.92%. From the results, it can be seen that using segmentation before classification improves the overall performance of the models in most of the cases.

REFERENCES

[1] T. Ai, Z. Yang, H. Hou, C. Zhan, C. Chen, W. Lv, Q. Tao, Z. Sun, and L. Xia, "Correlation of chest ct and rt-pcr testing for coronavirus disease 2019 (covid-19) in china: A report of 1014 cases," *Radiology*, vol. 296, p. 200642, 02 2020.

[2] S. Ahuja, B. K. Panigrahi, N. Dey, V. Rajinikanth, and T. K. Gandhi, "Deep transfer learning-based automated detection of covid-19 from lung ct scan slices," 2020.

[3] X. Chen, L. Yao, and Y. Zhang, "Residual attention u-net for automated multi-class segmentation of covid-19 chest ct images," *arXiv preprint arXiv:2004.05645*, 2020.

[4] F. Shi, J. Wang, J. Shi, Z. Wu, Q. Wang, Z. Tang, K. He, Y. Shi, and D. Shen, "Review of artificial intelligence techniques in imaging data acquisition, segmentation and diagnosis for covid-19," *IEEE Reviews in Biomedical Engineering*, pp. 1–1, 2020.

[5] D. Das, K. Santosh, and U. Pal, "Truncated inception net: Covid-19 outbreak screening using chest x-rays," *Physical and Engineering Sciences in Medicine*, vol. 43, 06 2020.

[6] M. Loey, F. Smarandache, and N. E. Khalifa, "Within the lack of chest covid-19 x-ray dataset: A novel detection model based on gan and deep transfer learning," *Symmetry*, vol. 12, p. 651, 04 2020.

[7] R. Miranda Pereira, D. Bertolini, L. Teixeira, C. Silla, and Y. Costa, "Covid-19 identification in chest x-ray images on flat and hierarchical classification scenarios," *Computer Methods and Programs in Biomedicine*, vol. 194, p. 105532, 05 2020.

[8] X. He, X. Yang, S. Zhang, J. Zhao, Y. Zhang, E. Xing, and P. Xie, "Sample-efficient deep learning for covid-19 diagnosis based on ct scans," *medRxiv*, 2020.

[9] A. Amyar, R. Modzelewski, and S. Ruan, “Multi-task deep learning based ct imaging analysis for covid-19: Classification and segmentation,” *medRxiv*, 2020.
[10] H. S. Maghdid, A. T. Asaad, K. Ghafoor, A. S. Sadiq, and M. Khan, “Diagnosing covid-19 pneumonia from x-ray and ct images using deep learning and transfer learning algorithms,” *ArXiv*, vol. abs/2004.00038, 2020.
[11] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “Imagenet classification with deep convolutional neural networks,” in *Advances in neural information processing systems*, pp. 1097–1105, 2012.
[12] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” *arXiv preprint arXiv:1409.1556*, 2014.
[13] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, 2016.
[14] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, “Going deeper with convolutions,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1–9, 2015.
[15] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, “Densely connected convolutional networks,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 4700–4708, 2017.
[16] O. Ronneberger, P. Fischer, and T. Brox, “U-net: Convolutional networks for biomedical image segmentation,” in *International Conference on Medical image computing and computer-assisted intervention*, pp. 234–241, Springer, 2015.
[17] PlamenEduardo, “Sars-cov-2 ct-scan dataset,” May 2020. Accessed: 2020-09-28.