Deep Feature Based COVID Detection from CT Scan Images Using Support Vector Machine

S.Lokesh Sai Phani Babu, U.Sri Ranganath, P.Bharath Anuj, C.Divyanth,Sikha O K*
Amrita School of Engineering, Coimbatore
Amrita Vishwa Vidyapeetham-India
*ok_sikha@cb.amrita.edu

Abstract Corona Virus (COVID-19) is an air-borne disease that has affected the lifestyle of people all around the world. Tracing patients infected with coronavirus has become a difficult process because of the limitation of tests based on Reverse transcription-polymerase Chain Reaction (RT-PCR). Recently methodologies based on imaging have been proposed by various researchers especially using deep learning-based models for the detection of COVID infection. This paper analyses the effectiveness of deep features for COVID detection from CT scan images. Deep features were extracted from the final layers of deep learning models which are then fed into machine learning frameworks for classification. Transfer learned features obtained from Resnet50, Inception V3 and Efficientb7 were employed for the study. A combination of InceptionV3 and SVM gave the best accuracy of 86.12 and precision and recall with 83.11 and 80.44 respectively. These results are comparable to recent transfer learning approaches and architecture that is about to be discussed is having an advantage of minimized time when compared to traditional deep learning approaches.

1 Introduction

COVID-19(coronavirus disease) is a highly contagious disease which is caused by SARS-COV-2(severe acute respiratory syndrome coronavirus) [1]. This virus is highly communicable and can be transferred to other people by direct contact or indirect contact. According to WHO statistics, currently, 226,844,344 people have been affected by COVID-19 and there were 4,666,334 deaths across the globe. [1]. In 2021, many new contagious and deadly variants of SARS-COV-19 like Alpha (B.1.1.7), Delta(AY.1) have been reported in countries like Brazil, UK, and India [2]. These new variants have increased mortality and infection rates in many countries when compared to the previous variants. According to WHO the most common symptoms are fever, cough, loss of taste and smell [3].

One of the causes for the rapid development of COVID in densely populated nations such as India is a lack of efficient tools capable of reliably diagnosing COVID in a short period of time. Out of many COVID diagnosing techniques, scientists have declared RT-PCR (reverse transcription polymerase chain reaction) as the standard procedure for diagnosing COVID-19 as this is one of the most successful identification tests till date [4]. Although it requires a huge number of certified personnel, large equipped labs and most importantly it requires a minimum of 5 hours of time [5]. Considering the drawbacks of this diagnosis method, scientists started to work on different techniques to detect COVID-19 infection. Apart from laboratory methods such as Antigen deduction, Antibody detection, point-of-care test, there are studies that explore various imaging techniques such as CT (computed tomography) scan images and X-ray images.

CT scans have been used extensively in the diagnosis of COVID-19 infections. When compared to other major diagnostic procedures such as RT-PCR, it has a high sensitivity rate. In rare cases, persons have been diagnosed with COVID-19 disease, yet CT imaging shows that they are not infected [4],[6]. The primary disadvantage of CT scan-based diagnosis is that it requires skilled radiologists to work and takes time. As a result, we cannot rely on this in an emergency. This underlines the requirement for an AI model capable of automating and detecting COVID in a short amount of time while providing superior performance. In addition to accurate COVID identification, the model's time complexity is critical. In comparison to existing deep convolutional neural networks, the proposed model is less complex in terms of learnable parameters and provides faster and accurate prediction.

The proposed model uses advanced deep learning and machine learning models for the accurate detection of COVID from CT scan images[7], which have been proved to be effective in the field of medical imaging. Various transfer learned models ResNet50V2, InceptionV3, EfficientNetB7 are explored to extract features from CT scan images which are then fed into using Machine learning models which are faster than many deep learning models for final classification. The prime contributions of this paper are summarized as follows:

- Deep features were extracted from the deep convolutional models such as InceptionnetV3, Resnet50V2 and EfficientnetB7.
- Supervised machine learning models: SVM, Decision tree and Random Forest were learned on the extracted deep features for Covid-19 detection.
- Effect of preprocessing on deep feature-based classification is analyzed

2 Related Works

Medical image analysis flourished buoyed by the advancement of deep learning models [8], such as Residual Network (Res Net), Dense net [9], etc. Yang *et al.*[8] proposed two deep learning models to assist us in the process of detecting COVID-19, a Dense-Net based model, and a Res net based model. In [8] Yang et al. created a dataset by collecting 760 preprints of COVID-19 patients from medRxiv and bioRxiv. With the help of MuPDF, they have extracted minute information and preserved the quality of preprints. Finally, a collection of 349 CT images labeled as positive training samples is generated, 397 non-COVID CT scan samples were also added to make a complete classification dataset. The DenseNet-

169 based model was then trained on the dataset which attained an accuracy of 69.8, 79.5, and 57.8% on the COVID-seg, COVID-CT-349, and COVID-CT-118 Datasets [8] respectively. Similarly, ResNet-50 has attained 66.3, 77.4, and 60.4% on the three Datasets as mentioned above. The authors created a Lung mask and lesion mask for each CT image and combined COVID-seg and COVID-CT-349 Datasets to attain an accuracy of 87.1%. In [10], the authors proposed a light CNN without bypass configuration for the detection of COVID-19. Instead of bypass configuration, they have added a transpose convolutional layer to the final Custom Fire Module. This Design has proved to achieve better accuracy compared to the Squeeze net model. Silva et al.[11] used Efficient net for Detecting COVID-19 Diagnosis in CT images with a voting-based approach and cross-data analysis. In this approach, the images from a particular patient are classified into groups based on voting. In [12], M. J. Horry *et al*. have taken three types of images for detecting COVID-19 i.e, X-Rays, Computerized Tomography (CT images), and Ultrasound Images. The author has tested the Dataset on the Optimized VGG-16 Model. The Experimental Results indicated that Ultrasound images provide better Accuracy than X-Rays and Computerized Tomography images. The model has attained a precision of 86%, 100%, and 84% respectively on X-Rays, Ultrasound, and CT images. Emtiaz Hussian *et al.* in [13], has proposed a 22- Layer CNN architecture called Coro-Det. The proposed model is developed to evaluate accurate diagnostics for 2 class Classification(COVID-19, Normal), 3 class Classification(COVID-19, Normal and Non-COVID Pneumonia) and 4 class Classification(COVID-19, Normal, Non-COVID viral pneumonia and Non-COVID Bacterial pneumonia). The model has achieved an accuracy of 99.1%, 94.2%, and 91.2% on 2 class classification, 3 class classification, and 4 class classification respectively. Diaz-Escobar J *et al.* in [14] used the publicly available dataset POCUS encompassing 3360 ultrasound images of Normal, COVID-19, and pneumonia patients. The author has used pre-trained models including Resnet50, Inception v3, VGG-19, and Xception and evaluated the dataset considering Three classes (COVID-19, Normal and Pneumonia) and two classes (COVID-19 and Normal). InceptionV3 has outperformed remaining state of art models by achieving the best average accuracy of 89.1, best balanced accuracy of 89.3, and area under the receiver operating curve (AUC) of 97.1%. In our work, we have used the COVID-19 CT Dataset that comprises 746 CT Scans of COVID and Non-COVID. We have extracted deep learning features from pre-trained Convolutional neural networks like Resnet50V2, Inception V3, and Efficient net b7. We have used machine learning models to evaluate the features extracted from the deep learning models. The primary aim of this work is to measure the efficiency of Machine Learning models on Deep Learning features.

3 Methodology

3.1 Dataset

COVID-19 CT Dataset proposed by Yang et al.[8] is employed for the proposed study. The Dataset comprises 746 CT images of COVID-19 patients (from 216 patients) from medRxiv and biorxiv. To Preserve the quality of images the authors have used PyMuPdf to extract Low-Level structure information of the pdf files of preprints and locate all the embedded figures. The final dataset includes 349 CT images labeled as positive samples and 397 CT images as Negative samples i.e, Non-COVID CT images. The maximum, average, and minimum height of the images are 1853, 491, and 153 whereas the minimum, average and maximum width of the images are 124, 383, and 1485. The details of the dataset is tabulated in Table.1

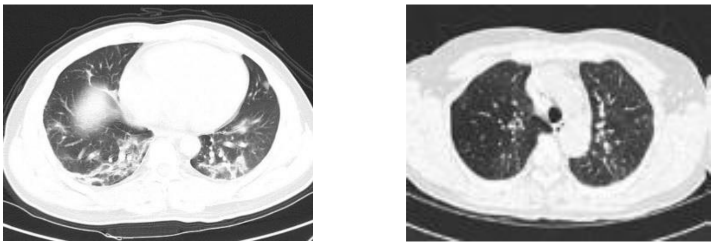

Figure 1: The Left Image represents a CT scan of a person affected with COVID-19 and the right image represents a CT Scan not affected with COVID.

Table 1: Details of COVID-19 CT-Data set [8]

Category	COVID-19 Data	Non-COVID-19 Data	Total Data
Train set	253	292	545
Test set	96	105	201

3.2 Data Augmentation and Pre-Processing

As mentioned in the previous section, the dataset consists CT scan images of different sizes, all the images were reshaped to (224,224) to make it uniform. Various data augmentation techniques such as Random Vertical Flips, Horizontal Flips, Rotation of images from 0 to 10 degrees, height shift range, width shift range, and Normalization of images between 0.0 and 1.0 degrees were applied on the dataset to avoid Overfitting. To further enhance the quality of the images pseudo coloring [15], [16] is employed, which gave a better classification accuracy also. Figure.2 shows the original image and the pseudo-colored image.

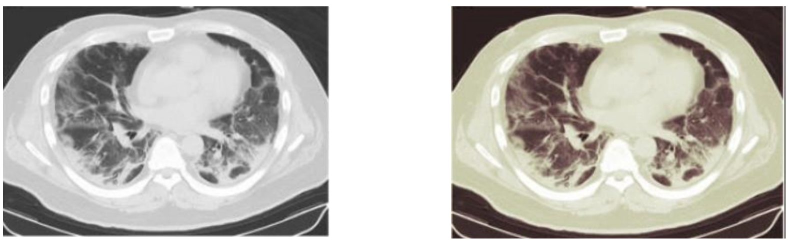

Figure 2: The Left Image shows the original image whereas the right image is after applying pseudo coloring and dimension normalization to the original image

3 Proposed Architecture

A Convolutional Neural Network (Conv Net) is a special category of neural network architectures that takes an image as input, assigns importance to distinct aspects in the image, and can distinguish between them [9], [5]. They have made significant progress toward improved image classification performance. The proposed model exploited various state-of-the-art CNN models for classification such as InceptionV3, Resnet50V2, and Efficient Net b7 for extracting features. The high-dimensional nature of the transfer learned deep features produced from multiple deep convolutional models could have a negative impact on the classification model. This unveils the limitation of this study. High-level features obtained from the final FC layer of the aforementioned deep learning models are then fed into machine learning models to analyze how well the ML models are performing on the deep learning features. The workflow of the proposed model is shown in Figure.3.

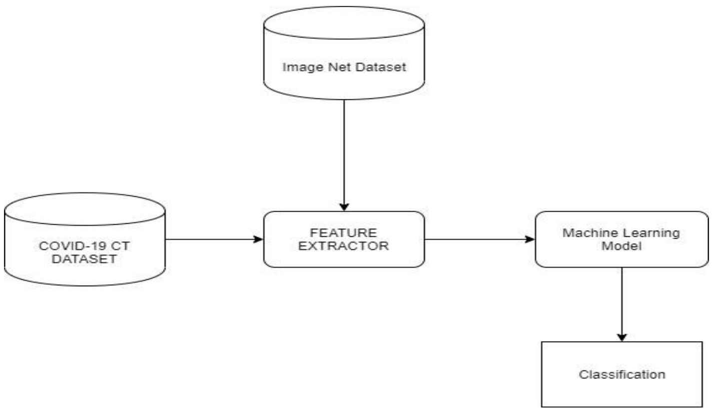

Figure 3: Workflow of the proposed Architecture

4 Results and Discussions

This section compares the effectiveness of transfer learned features from state-of-the-art models such as Efficient B7, InceptionNetV3, Resnet50V2 for classification. The extracted feature vectors were fed into advanced machine learning models like Support Vector Machine (SVM), Decision Tree, Random Forest for prediction. The experiment was carried out on the Google Colab Pro which provides NVidia Tesla T4 or P100 GPU. Colab Pro is chosen for experimentation as it provides a longer GPU when compared to Colab and has a high computational power that is required for the experiments. As the dataset is balanced, we have used evaluation metrics such as Accuracy (Equation.1), Precision (Equation.2), and Recall (Equation.3) that are calculated using the values extracted from the confusion matrix for the experimental analysis.

$$\text{Accuracy} = \frac{TP+TN}{TP+FP+TN+FN} \quad \text{.....(1)}$$

$$\text{Precision} = \frac{TP}{TP+FP} \quad \text{.....(2)}$$

$$\text{Recall} = \frac{TP}{TP+FN} \quad \text{.....(3)}$$

We got the best results using a combination of InceptionV3 and SVM out of all the Transfer Learning Models we tried. Tables 2 and 3 show the results obtained with and without image pre-processing as explained in Section 3.2. On the SVM model, we achieved an accuracy of 86.12 percent, 85.24 percent, and 78.19 percent with the Incep-tionV3, Resnet 50V2, and Efficient B7, respectively. According to Tables 2 and 3, there is an increase in accuracy for all three models after applying to pre-process.

Table 2: Performance Comparison of various combinations of deep learning and machine learning models on classification without Image pre-processing

	InceptionV3			Resnet50V2			EfficientnetB7		
	Accuracy	Precision	Recall	Accuracy	Precision	Recall	Accuracy	Precision	Recall
SVM	85.24	87.34	90.70	83.80	88.14	85.65	77.81	83.90	79.19
Decision Tree	75.12	79.002	76.76	74.96	81.50	76.56	68.70	74.06	76.16
Random Forest	78.56	85.49	78.58	79.30	85.17	79.6	78.31	78.60	88.48

Table 3: Performance Comparison of various combinations of deep learning and machine learning models on classification with Image pre-processing

	InceptionV3			Resnet50V2			EfficientnetB7		
	Accuracy	Precision	Recall	Accuracy	Precision	Recall	Accuracy	Precision	Recall
SVM	86.12	87.92	89.96	85.24	85.25	87.27	78.19	83.43	81.14
Decision Tree	76.70	82.62	82.62	75.34	75.55	75.55	73.72	77.57	77.57
Random Forest	80.42	81.61	78.93	78.93	78.98	78.98	72.24	87. 87	87. 87

The proposed model was compared against the state-of-the-art classification models for the COVID classification task. The obtained results are tabulated in table.4. The proposed model produces equivalent results (in terms of accuracy) to the benchmark models in the literature, as shown in table 4.

Table 4: Performance Comparison of the proposed architecture with the existing approaches

	Accuracy	Precision	Recall
Jangam et al [17]	84.73	0.791	0.928
Polsinelli et al[10]	83.00	0.817	0.850
Silva et al [11]	**87.68**	0.939	0.795
Proposed Architecture	**86.12**	83.11	80.44

5 Conclusion

There are possibilities that patients affected with COVID might end up damaging their lungs permanently, which can lead to death after a long time. In this paper, we employ image features extracted from various deep learning models for classification. The extracted features were fed into machine learning models for classification, as machine learning models are faster than deep learning models but lack the ability to extract features. With the COVID19 dataset, the combination of InceptionV3 and SVM performed the best, with an accuracy of 86.12%.

6 References

1. Cucinotta D, Vanelli M. WHO Declares COVID-19 a Pandemic. *Acta Biomed*. 2020;91(1):157-160. Published 2020 Mar 19. doi:10.23750/abm.v91i1.9397
2. Janik, E.; Niemcewicz, M.; Podogrocki, M.; Majsterek, I.; Bijak, M. The Emerging Concern and Interest SARS-CoV-2 Variants. *Pathogens* **2021**, *10*, 633. https://doi.org/10.3390/pathogens10060633
3. Christopher K C Lai, Wilson Lam,Laboratory testing for the diagnosis of COVID-19,Biochemical and Biophysical Research Communications,Volume 538,2021,Pages 226-230,ISSN 0006-291X,https://doi.org/10.1016/j.bbrc.2020.10.069.
4. Alireza Tahamtan & Abdollah Ardebili (2020) Real-time RT-PCR in COVID-19 detection: issues affecting the results, Expert Review of Molecular Diagnostics, 20:5, 453-454, DOI: 10.1080/14737159.2020.1757437
5. S. Tamuly, Jyotsna, C., and J, A., “Deep Learning Model for Image classification”, in 3rd International Conference on Computational Vision and Bio Inspired Computing, (ICCVBIC 2019), , RVS Technical Campus, Coimbatore, 2019.

6. U. Subbiah, R. V. Kumar, S. A. Panicker, R. A. Bhalaje and P. S, "An Enhanced Deep Learning Architecture for the Classification of Cancerous Lymph Node Images," *2020 Second International Conference on Inventive Research in Computing Applications (ICIRCA)*, Coimbatore, India, 2020, pp. 381-386, doi: 10.1109/ICIRCA48905.2020.9183250.
7. R. Ramachandran, Rajeev, D. C., Krishnan, S. G., and Subathra P., “Deep learning – An overview”, International Journal of Applied Engineering Research, vol. 10, pp. 25433-25448, 2015
8. Yang, Xingyi, et al. "COVID-CT-Dataset: a CT scan dataset about COVID-19." (2020).
9. Vinayakumar, R., Soman, K.P., Poornachandran, P.:” Applying convolutional neural network for network intrusion detection.” *In: 2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI), pp. 1222–1228. IEEE (2017).*
10. Polsinelli, Matteo, Luigi Cinque, and Giuseppe Placidi. "A light CNN for detecting COVID-19 from CT scans of the chest." *Pattern Recognition Letters* 140 (2020): 95-100.
11. Silva, Pedro, *et al*. "COVID-19 detection in CT images with deep learning: A voting-based scheme and cross-datasets analysis." *Informatics in Medicine Unlocked* 20 (2020): 100427
12. M. J. Horry et al., "COVID-19 Detection Through Transfer Learning Using Multimodal Imaging Data," in IEEE Access, vol. 8, pp. 149808-149824, 2020, doi: 10.1109/ACCESS.2020.3016780.
13. Emtiaz Hussain, Mahmudul Hasan, Md Anisur Rahman, Ickjai Lee, Tasmi Tamanna, Mohammad Zavid Parvez,CoroDet: A deep learning based classification for COVID-19 detection using chest X-ray images,Chaos, Solitons & Fractals,Volume 142,2021,110495,ISSN 0960-0779,https://doi.org/10.1016/j.chaos.2020.110495.
14. Diaz-Escobar J, Ordóñez-Guillén NE, Villarreal-Reyes S, Galaviz-Mosqueda A, Kober V, Rivera-Rodriguez R, et al. (2021) Deep-learning based detection of COVID-19 using lung ultrasound imagery. PLoS ONE 16(8): e0255886. https://doi.org/10.1371/journal.pone.0255886
15. Liang, S., Liu, H., Gu, Y. *et al.* Fast automated detection of COVID-19 from medical images using convolutional neural networks. *Commun Biol* 4, 35 (2021). https://doi.org/10.1038/s42003-020-01535-7

16. Saj, TK Sachin, *et al*. "Facial Emotion Recognition Using Shallow CNN." *Symposium on Machine Learning and Metaheuristics Algorithms, and Applications*. Springer, Singapore, 2019

17. Jangam, E., Barreto, A.A.D. & Annavarapu, C.S.R. Automatic detection of COVID-19 from chest CT scan and chest X-Rays images using deep learning, transfer learning and stacking. *Appl Intell* (2021). https://doi.org/10.1007/s10489-021-02393-4