Employing Artificial Intelligence to Assist Doctors in Diagnosing COVID-19

Mohiniben Prajapati
Registration Number: 2009326
School of Computer Science and Electronics Engineering
University of Essex
Colchester, England, UK
mp20469@essex.ac.uk

***Abstract*—COVID-19 has changed the way we live, work, and travel around the world. Even after almost a year into this pandemic we still have places where there is a rise of COVID-19 cases. It is vital that the presence of the virus is detected early. This can be accomplished with the help of artificial intelligence and manual input from humans. To achieve this goal, deep learning models were employed to get a better accuracy, recall, precision and other metrics compared to the baseline metrics. DenseNet and NASNetLarge pre-trained models were tried for investigation. Transfer learning was used to classify the images. Feature extraction and fine tuning was done on the models. With the use of DenseNet better metrics were achieved.**

***Index Terms*—classification, COVID-19, CT scans, DenseNet, NASNetLarge, pandemic, SARS-CoV-2**

I. Introduction

Machine learning and data science are the emerging and most important fields of this and the previous decade. Artificial intelligence has been integrated and incorporated in various fields and applications. The field of medicine, however, has been reluctant on relying on deep learning models because of their black box nature. This however does not mean that it cannot help practitioners diagnose certain diseases. Practitioners' own experience along with the help of artificial intelligence can significantly improve the efficiency and decrease false positives and false negatives in diagnosis.

SARS-CoV-2, also known colloquially as COVID-19, C-OVID or simply as Coronavirus has changed the course of human history. The virus has a high death toll and has devastated so many lives. The psychological impact of this is also staggering. In these hard times it is important that patients with COVID-19 are diagnosed precisely and quickly.

Chest CT scans can be used to find the severity of the SARS-CoV-2 [1]. For most of the COVID-19 cases lesions were in lower lobes or in more than two lobes [2]. Bilateral lesions were also common; lesions in a single lobe were uncommon. This is a vital information which can be used to do feature extraction and improve the classification accuracy.

COVID-19 is a type of pneumonia [3] and it is different influenza and other types of pneumonia. Hence, it is critical that we employ models which can correctly classify non-COVID-19 cases from COVID-19 cases.

Accuracy alone is not enough for the problem at hand. False positive identification means that the non-COVID patient was identified as a COVID patient; false negative identification means that the COVID patient was identified as a non-COVID patient. For medical diagnosis false negative rate is much more important than false positive rate. This is even more apparent in a global pandemic where even a few false negative cases can severely impact the patient's health, can have negative impacts on the hospital, and can increase the spread of the virus. The metric precision considers false positives and the metric recall considers false negatives. Hence, for our investigation recall and accuracy are the measures that should be considered to evaluate the performance of the models. Receiver Operating Characteristic Curve (ROC Curve) and Area Under the ROC Curve (AUC) are the two metrics that measure the separability of the classes. These two metrics are also employed in our investigation.

II. Literature Review

Hyper-tuning deep transfer learning model DenseNet201 brings improvements over VGG-16 [4]. The improvement is only 1%; but this can save many lives. DenseNet is more condensed [5] which can explain the improvements over VGG-16. CoroDet CNN model consists of 22-layers such as convolutional layer, dense layer, max pooling, and activation layers such as ReLU, Leaky ReLU, and Sigmoid [6]. The given model has more than 2 million parameters and achieves a reasonable classification accuracy. The given model also assists clinicians in finding clear demarcations; unsurprisingly, model performance is low on low resolution images [6].

The authors in [7] have used ResNet50V2 as a base model and modified feature pyramid network (FPN) to extract important features, dropout was also used to prevent overfitting. This model achieves better metrics than Xception and ResNet50V2 model. On RYDLS-20 dataset [8] have used various combinations to classify COVID cases; the proposed schema has various phases, in the first phase the images are inputted in various learning models. In the next phase early fusion (combination of extracted features is done before training and test) is done. In the last two phases resampling and learning is done; in the last phase multi-class learning and hierarchical learning was done, late fusion (features are trained individually and the predictions are combined after the training) was also done in this phase [8]. The proposed schema achieves upwards of 90% F1-score in hierarchical learning. By combining both classification and segmentation task the authors in [9] have achieved better metrics than U-NET and image classification model CNN. Adding image reconstruction task improves the accuracy even more.

Using Q-Deformed entropy feature extraction and deep features on CNN model achieves respectable classification accuracy [10]. The authors have extracted spectral and spatial deep features to better classify the images. Multi-objective differential evolution based CNN model applied on chest CT scans achieves better accuracy than CNN, ANN, and ANFIS models [11]. In [12] targeted self supervision was employed (which starts with a downstream task instead of providing a pretext task for different downstream tasks). This improves performance over transfer learning from other models from ImageNet.

III. Methodology

Data set[1] from GitHub was cloned and unzipped. The zip files for both COVID and non-COVID images were 40 megabytes each. The image format is png.

A. Exploratory Data Analysis

There are 348 images of scans which are COVID positive; and 396 images of scans which are COVID negative.

B. Preprocessing

Data augmentation was done on the images. The images were resized to 224x224 resolution. This standardizes the image resolution. Then the images were normalized; alpha and beta values were zero and one. Min-max normalization was performed. OpenCV[2] library was utilized for both resizing and normalizing the images.

C. Modeling

Keras[3] library was utilized for modeling. Transfer learning was used to for modeling. First, the model is loaded from the library. Then both feature extraction and fine tuning is performed for each model.

The code for the models was adapted from a GitHub Repository[4].

a) Feature Extraction: Top layer was not included when loading the models. The pre-trained models' output were flattened and then a dense layer with a single neuron and sigmoid activation was created. Binary cross-entropy loss and adam optimizer were used. Learning rate was reduced on a plateau. Batch size and epochs used were 10 and 100.

b) Fine Tuning: Top layer was not included when loading the model. The pre-trained models' first half was frozen. Then the models' output was flattened and a dense layer with a single neuron and sigmoid activation was created. Binary cross-entropy loss and adam optimizer were used. Batch size and epochs used were 20 and 100.

[1]https://github.com/UCSD-AI4H/COVID-CT
[2]https://opencv.org/
[3]https://keras.io/
[4]https://github.com/sagihaider/TransferLearning_COVID19

D. Evaluation

As mentioned in the introduction section, accuracy, recall, precision, F1 Score, AUC, and ROC metrics were used to evaluate the models' performance.

IV. Results

TABLE I
DenseNet121 Results

DenseNet121 Metrics	Feature Extraction	Fine Tuning
Train Accuracy	1.0	0.996
Validation Accuracy	0.87	0.84
Test Accuracy	0.885	0.843
Precision	0.849	0.851
Recall	0.938	0.83
F1 Score	0.891	0.842
AUC	0.885	0.843

As it can be seen from the tables, better metrics were achieved with DenseNet169 pre-trained model.

Figure 1 shows the accuracy plot for the DenseNet121 model with feature extraction. The training and validation accuracy stagnate after five epochs. Figure 2 shows the loss plot for the DenseNet121 model with feature extraction. The training and validation loss stagnate after five epochs. Figure 3 shows the confusion matrix for the DenseNet121 model with feature extraction. There are less false positives and false negatives.

Figure 4 shows the accuracy plot for the DenseNet121 model with fine tuning. Figure 5 shows the loss plot for the DenseNet121 model with fine tuning. Figure 6 shows the confusion matrix for the DenseNet121 model with fine tuning.

Figure 7 shows the accuracy plot for the DenseNet169 model with feature extraction. The training and validation accuracy stagnate after five epochs. Figure 8 shows the loss plot for the DenseNet169 model with feature extraction. The training and validation loss stagnate after five epochs. Figure 9

TABLE II
DenseNet169 Results

DenseNet121 Metrics	Feature Extraction	Fine Tuning
Train Accuracy	1.0	0.998
Validation Accuracy	0.85	0.91
Test Accuracy	0.927	0.927
Precision	0.956	0.902
Recall	0.896	0.958
F1 Score	0.924	0.929
AUC	0.927	0.927

TABLE III
NASNetLarge Results

NASNetLarge Metrics	Feature Extraction
Train Accuracy	0.99
Validation Accuracy	0.76
Test Accuracy	0.781
Precision	0.745
Recall	0.854
F1 Score	0.796
AUC	0.781

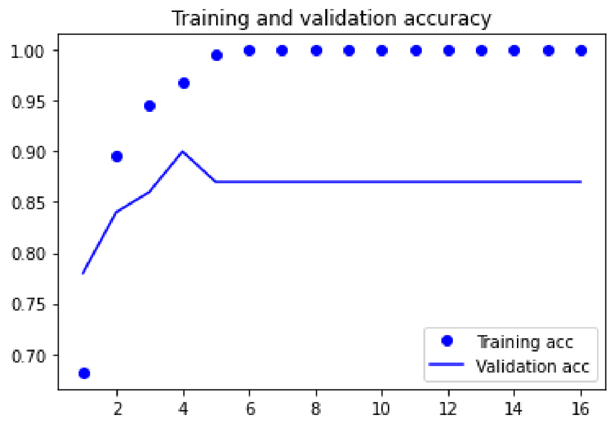

Fig. 1. DenseNet121 Feature Extraction — Accuracy Plot

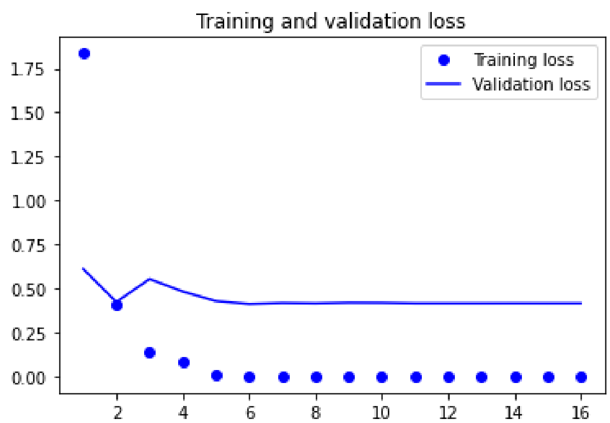

Fig. 2. DenseNet121 Feature Extraction — Loss Plot

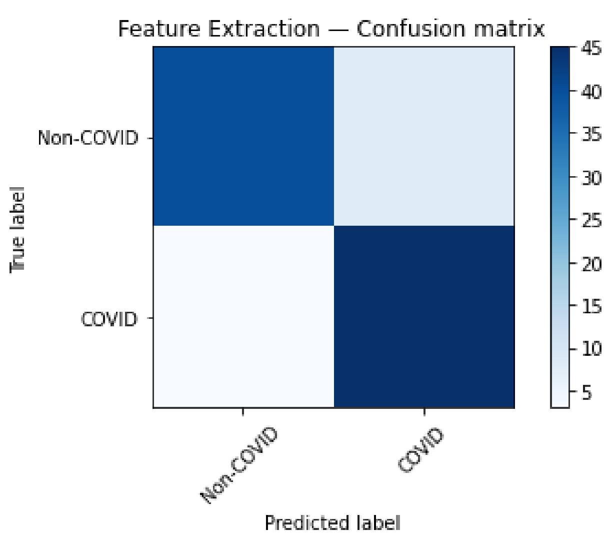

Fig. 3. DenseNet121 Feature Extraction — Confusion Matrix

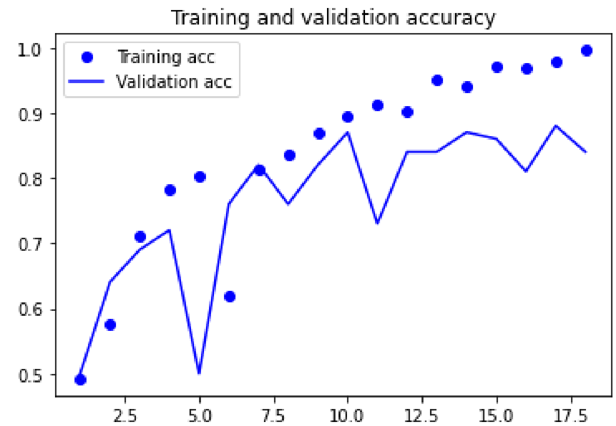

Fig. 4. DenseNet121 Fine Tuning — Accuracy Plot

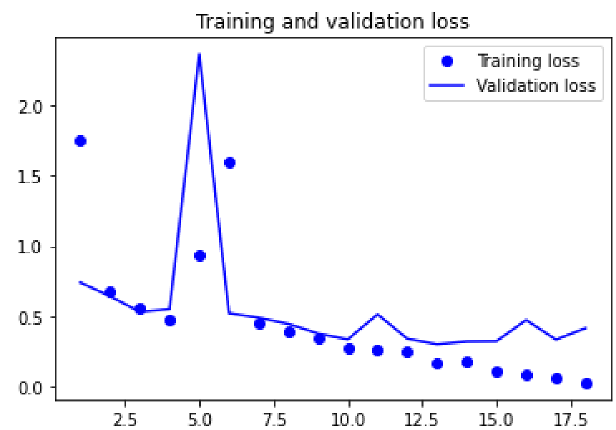

Fig. 5. DenseNet121 Fine Tuning — Loss Plot

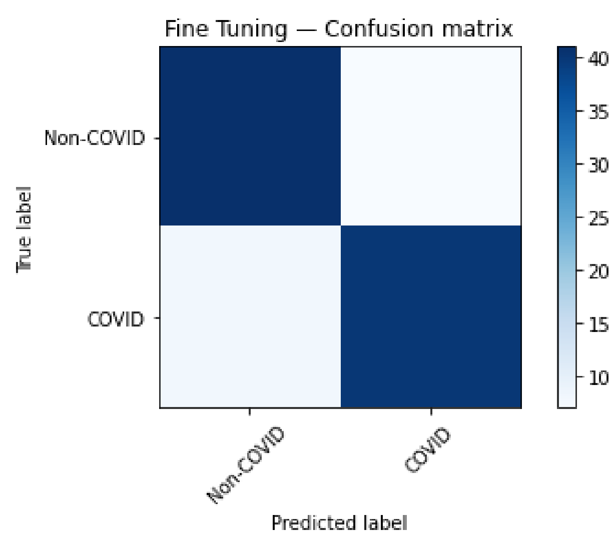

Fig. 6. DenseNet121 Fine Tuning — Confusion Matrix

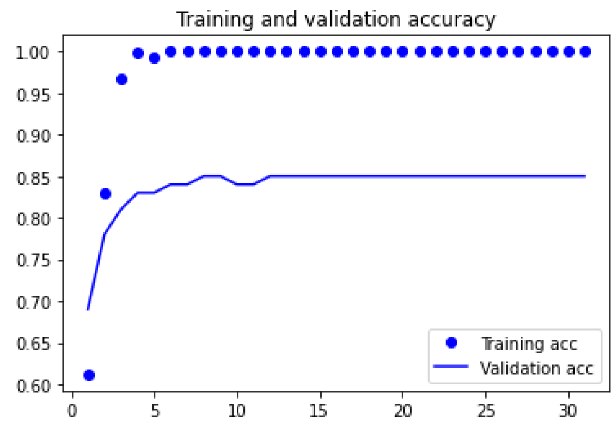

Fig. 7. DenseNet169 Feature Extraction — Accuracy Plot

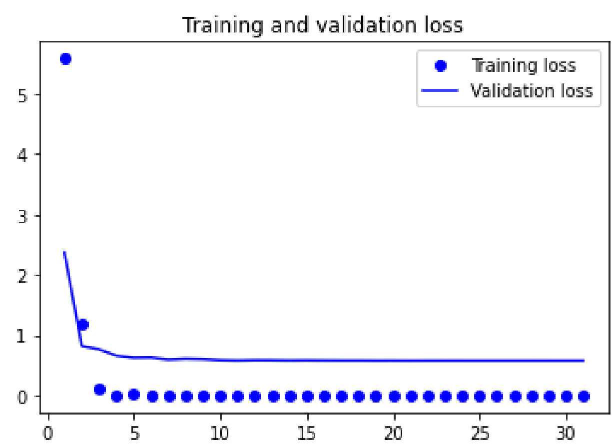

Fig. 8. DenseNet169 Feature Extraction — Loss Plot

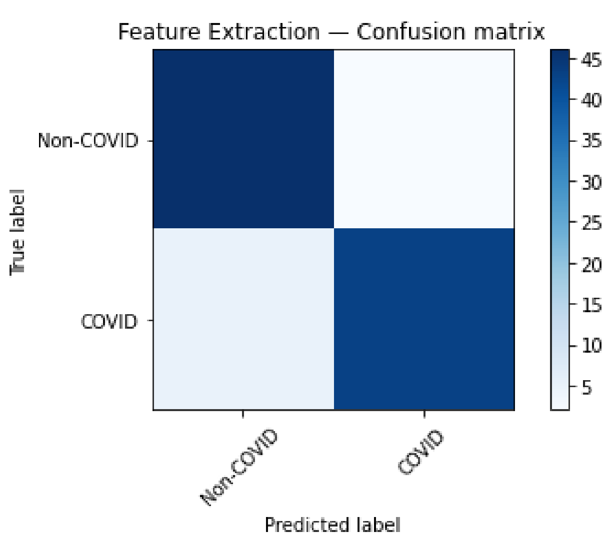

Fig. 9. DenseNet169 Feature Extraction — Confusion Matrix

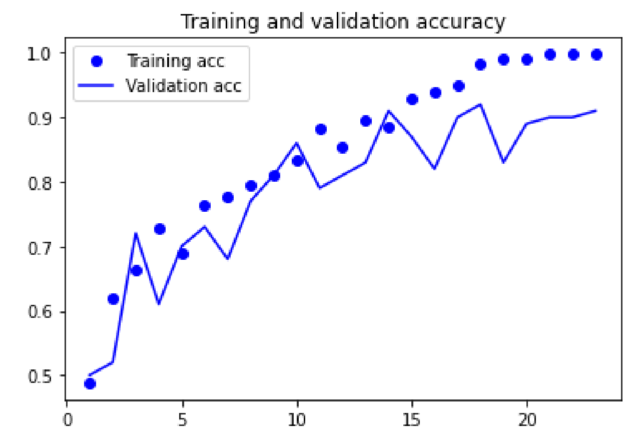

Fig. 10. DenseNet169 Fine Tuning — Accuracy Plot

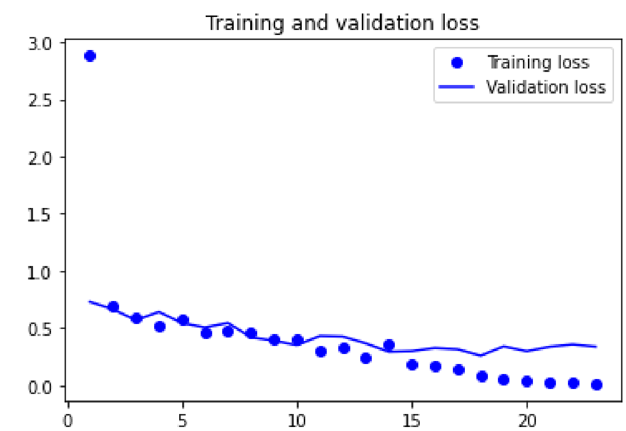

Fig. 11. DenseNet169 Fine Tuning — Loss Plot

shows the confusion matrix for the DenseNet169 model with feature extraction.

Figure 10 shows the accuracy plot for the DenseNet169 model with fine tuning. The training and validation accuracy stagnate after five epochs. Figure 11 shows the loss plot for the DenseNet169 model with fine tuning. The training and validation loss stagnate after five epochs. Figure 12 shows the confusion matrix for the DenseNet169 model with fine tuning.

Figure 13 shows the accuracy plot for the NASNetLarge model with feature extraction. The training and validation accuracy stagnate after five epochs. Figure 14 shows the loss plot for the NASNetLarge model with feature extraction. The training and validation loss stagnate after five epochs. Figure 15 shows the confusion matrix for the NASNetLarge model with feature extraction.

V. Discussion

Presence or absence of COVID-19 can be detected with lungs' CT scans. Using machine learning the scanned images can be classified as a COVID-19 or non-COVID-19 CT scan image. Pre-trained models were trained for the purpose of classification. Along with Xception, Inception and its variants

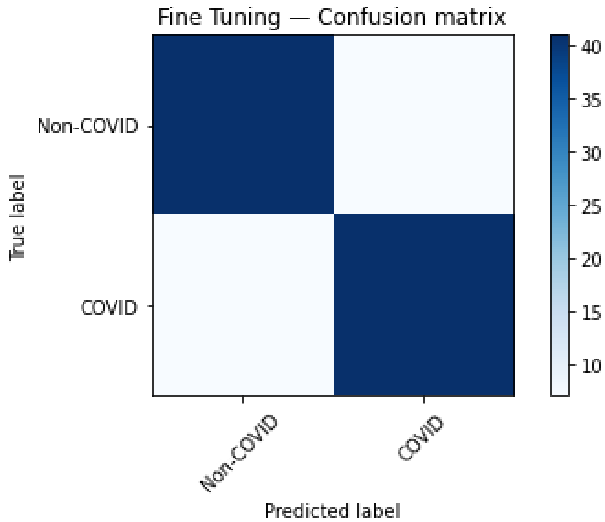

Fig. 12. DenseNet169 Fine Tuning — Confusion Matrix

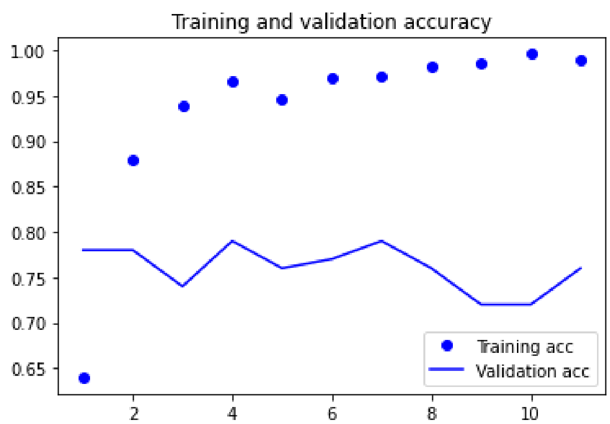

Fig. 13. NASNetLarge Feature Extraction — Accuracy Plot

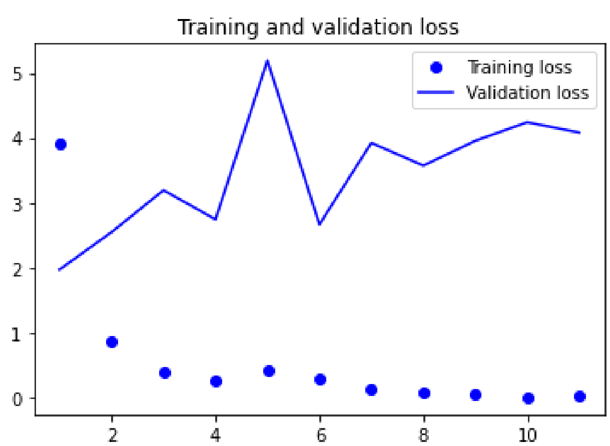

Fig. 14. NASNetLarge Feature Extraction — Loss Plot

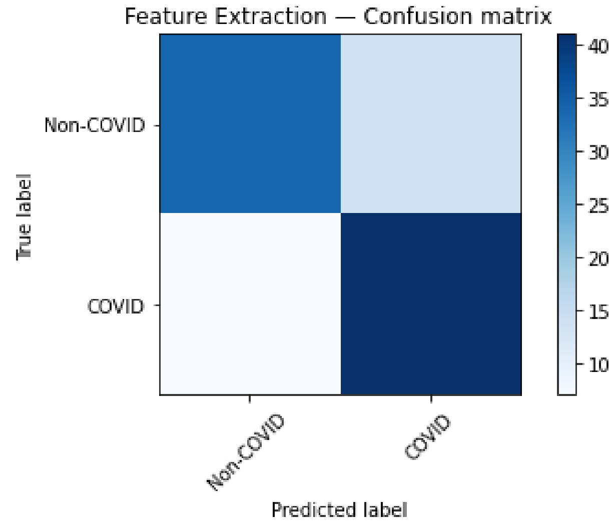

Fig. 15. NASNetLarge Feature Extraction — Confusion Matrix

were also tried. But the performance was significantly less than the baseline metrics. Out of the tried models, NASNetLarge's results are shown in the results section. DenseNet models gave the best results during the investigation.

Surprisingly, DenseNet121's feature extraction gives better metrics than fine tuning. In DenseNet169, we clearly see a significant improvement with fine tuning.

VI. Conclusion

Significant improvements were made compared to the baseline results[5].

Various pre-trained models were trained and DenseNet169 performed the best. This can have a major impact on the efficiency of COVID-19 diagnosis.

There is still a room for improvement. If more data is available than there is a potential for better results. Also, various methods and models from literature review can significantly reduce the validation loss. Various models can be trained and a voting scheme can be developed; the image would be classified as a COVID or non-COVID based on the number of votes.

Models that can extract precise edges of the medical images can be developed which can have a major impact on the medical field and artificial intelligence as a whole.

References

[1] E. Poggiali, A. Dacrema, D. Bastoni, V. Tinelli, E. Demichele, P. Mateo Ramos, T. Marcianò, M. Silva, A. Vercelli, and A. Magnacavallo, "Can lung us help critical care clinicians in the early diagnosis of novel coronavirus (covid-19) pneumonia?," *Radiology*, vol. 295, no. 3, pp. E6–E6, 2020.

[2] K. Li, Y. Fang, W. Li, C. Pan, P. Qin, Y. Zhong, X. Liu, M. Huang, Y. Liao, and S. Li, "Ct image visual quantitative evaluation and clinical classification of coronavirus disease (covid-19)," *European radiology*, vol. 30, no. 8, pp. 4407–4416, 2020.

[3] M. Liu, W. Zeng, Y. Wen, Y. Zheng, F. Lv, and K. Xiao, "Covid-19 pneumonia: Ct findings of 122 patients and differentiation from influenza pneumonia," *European radiology*, vol. 30, no. 10, pp. 5463–5469, 2020.

[5] https://github.com/sagihaider/TransferLearning_COVID19

[4] A. Jaiswal, N. Gianchandani, D. Singh, V. Kumar, and M. Kaur, "Classification of the covid-19 infected patients using densenet201 based deep transfer learning," *Journal of Biomolecular Structure and Dynamics*, pp. 1–8, 2020.

[5] 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.

[6] E. Hussain, M. Hasan, M. A. Rahman, I. Lee, T. Tamanna, and M. Z. Parvez, "Corodet: A deep learning based classification for covid-19 detection using chest x-ray images," *Chaos, Solitons & Fractals*, vol. 142, p. 110495, 2021.

[7] M. Rahimzadeh, A. Attar, and S. M. Sakhaei, "A fully automated deep learning-based network for detecting covid-19 from a new and large lung ct scan dataset," *Biomedical Signal Processing and Control*, vol. 68, p. 102588, 2021.

[8] R. M. Pereira, D. Bertolini, L. O. Teixeira, C. N. Silla Jr, and Y. M. 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, 2020.

[9] A. Amyar, R. Modzelewski, H. Li, and S. Ruan, "Multi-task deep learning based ct imaging analysis for covid-19 pneumonia: Classification and segmentation," *Computers in Biology and Medicine*, vol. 126, p. 104037, 2020.

[10] A. M. Hasan, M. M. Al-Jawad, H. A. Jalab, H. Shaiba, R. W. Ibrahim, and A. R. AL-Shamasneh, "Classification of covid-19 coronavirus, pneumonia and healthy lungs in ct scans using q-deformed entropy and deep learning features," *Entropy*, vol. 22, no. 5, p. 517, 2020.

[11] D. Singh, V. Kumar, M. Kaur, *et al.*, "Classification of covid-19 patients from chest ct images using multi-objective differential evolution–based convolutional neural networks," *European Journal of Clinical Microbiology & Infectious Diseases*, vol. 39, no. 7, pp. 1379–1389, 2020.

[12] N. Ewen and N. Khan, "Targeted self supervision for classification on a small covid-19 ct scan dataset," in *2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI)*, pp. 1481–1485, IEEE, 2021.