MULTI-SCALE RESIDUAL NETWORK FOR COVID-19 DIAGNOSIS USING CT-SCANS

Pratyush Garg
Department of Electrical Engineering
Indian Institute of Technology, Delhi
pratyush.garg1999@gmail.com

Rishabh Ranjan
Department of Electrical Engineering
Indian Institute of Technology, Delhi
rishabhranjan456@gmail.com

Kamini Upadhyay
Centre for Applied Research in Electronics
Indian Institute of Technology, Delhi
kaminiup23@gmail.com

Monika Agrawal
Centre for Applied Research in Electronics
Indian Institute of Technology, Delhi
maggarwal@care.iitd.ernet.in

Desh Deepak
Department of Respiratory Medicine
Dr. Ram Manohar Lohia Hospital, Delhi
drdeepak.rml@gmail.com

Abstract—To mitigate the outbreak of highly contagious COVID-19, we need a sensitive, robust automated diagnostic tool. This paper proposes a three-level approach to separate the cases of COVID-19, pneumonia from normal patients using chest CT scans. At the first level, we fine tune a multi-scale ResNet50 model for feature extraction from all the slices of CT scan for each patient. By using multi-scale residual network, we can learn different sizes of infection, thereby making the detection possible at early stages too. These extracted features are used to train a patient-level classifier, at the second level. Four different classifiers are trained at this stage. Finally, predictions of patient level classifiers are combined by training an ensemble classifier. We test the proposed method on three sets of data released by ICASSP, COVID-19 Signal Processing Grand Challenge (SPGC). The proposed method has been successful in classifying the three classes with a validation accuracy of 94.9% and testing accuracy of 88.89%.

Index Terms—COVID-19, ResNet, BiLSTM, Computed Tomography Scans, Deep Learning

I. INTRODUCTION

Coronavirus Disease-2019 or COVID-19 is a highly contagious respiratory illness, which started spreading from Wuhan, China in December 2019 [8]. In March 2020, World Health Organisation declared this disease a pandemic due to its exponential rate of rising cases all over the world [7]. The novelty of disease makes its diagnosis a big challenge before the medical community. Presently, Real-time Reverse Transcription–Polymerase Chain Reaction (RT–PCR) is the most common COVID-19 diagnostic method [6]. However, this test is expensive, time-consuming and sometimes less sensitive $(71-98\%)$. Also, specialised laboratories are required for these tests [5]. Other than RT-PCR, lung imaging modalities like chest X-rays (CXR) and computerised tomography (CT) scans are also helping the medical experts to visualize the infection and thereby using them for diagnosis. Various opacities like bilateral ground-glass opacities (GGOs), bilateral multilobar GGO and consolidations, mostly in peripheral regions of lungs have been reported as observable visual patterns in COVID-19 positive cases [1]- [4]. For COVID-19 diagnosis, in comparison to CXR images, CT scans are reported to perform better in terms of sensitivity (Sen) and specificity (Spe) measures [1].

Several research groups are analysing and experimenting with CT-scans to develop an automatic COVID-19 diagnostic tool to assist the medical experts. Amyar et al. [19] proposed a multitasking deep model to segment the COVID-19 infection lesions and classify the CT scans of COVID-19. Inf-Net [20] model is proposed to locate the infected regions of CT-slices and thus detect COVID-positive cases. Hu et al. [17] developed a mutli-scale classifier to detect COVID-19, community acquired pneumonia and non-pneumonia cases. Despite all these algorithms, there is still a large scope of improvement in COVID-19 diagnosis. We need to minimize the false negatives to alleviate the risk. Moreover, with rapidly increasing variety of coronavirus strains, we need a robust and sensitive algorithm which can detect the slightest possibility of infection. In the proposed work, we use chest CT-scans to train a robust three level classifier. At the first level, slice based feature extraction is done for each patient. These features are then used to train patient level classifiers. We train four different patient level classifiers. These classifiers are then combined by training an ensemble classifier. Main contribution of the proposed algorithm lie in multi-scale feature extraction which leads into better detection of different dimensions of infection. Moreover, three-level classification brings robustness to the algorithm. The proposed method achieves an accuracy of 94.9% and a sensitivity of 100%, 89.47% and 94.55% for Normal, CAP and COVID-19 patients, respectively, on the validation set. We also visualize the heat maps from different convolution layers using Grad-CAM which reveals that the model is looking at the correct infectious areas.

II. DATASET

In this section, we explain the different datasets used for our study. The first dataset, COVIDx-CT [9] is used for pre-training the model. The second dataset, COVID-CT-MD [11] is then used to fine tune the model.

A. COVIDx-CT

This dataset [22] comprises of $194,922$ chest CT slices from $3,745$ patients. It consists of COVID-19 positive patients with confirmed diagnoses (i.e., RT-PCR, radiologist-confirmed, etc.). It has chest CT-scans of patients with slice level labels across three different classes, i.e., COVID-19, Community Acquired Pneumonia (CAP) and Normal. It contains $94,548$ slices for COVID-19, $40,291$ slices for CAP and $60,083$ slices for Normal patients.

The dataset is provided with a $60\% - 20\% - 20\%$ split for training, validation and testing, respectively. We combine the training and validation dataset to train our model. It has images of size 512×512 pixels in png format.

B. COVID-CT-MD

This dataset [11] contains the chest CT scans of 307 patients out of which 171 are COVID-19 positives, 60 are patients with community acquired pneumonia (CAP), and 76 are normal patients. Along with this, 55 COVID-19 and 25 CAP patients have slice level labels. The slice level dataset consists of $4,993$ slices with infection and $18,501$ slices without infection. The dataset is provided in DICOM format with images of size 512×512 pixels. The dataset is provided with a $70\% - 30\%$ split for training and validation.

C. ICASSP-SPGC Test Set

The test dataset consists of three sets with each set having 30 patients. All the three sets have different exposure dose, slice thickness etc. One of the test sets also contain CT scans of patients with manifestations related to non-infectious disease.

III. PROPOSED METHOD

In order to develop a sensitive and robust COVID-19 diagnosis using chest CT-scans, we adopt a three-level classification approach. First level is extraction of features corresponding to normal, COVID-19 and CAP cases. The second level uses the extracted features to train four different classifiers at patient level. The last level combines the predictions from the multiple classifiers of previous level in an optimized manner.

A. Preprocessing

Before applying the core algorithm, we need some pre-processing steps to clean the raw images. Here, we extract region of interest (RoI), normalise the intensity range, resize the images etc. This step facilitates the model to focus on the relevant details. As we use two different datasets to train our model, we need some different sets of pre-processing steps in each case. Fig 1 shows the raw and pre-processed images.

1) Preprocessing of CT slices from COVIDx-CT dataset: The raw CT slices have projections of pulmonary region along-with patient's body and some other external objects. So, as the first step of pre-processing, we remove these extra projections and extract our RoI i.e. lungs along with patient's body from each of the CT slices using image contours. The image contours are computed using the binary image of the given CT slices. We compute the binary image by applying a
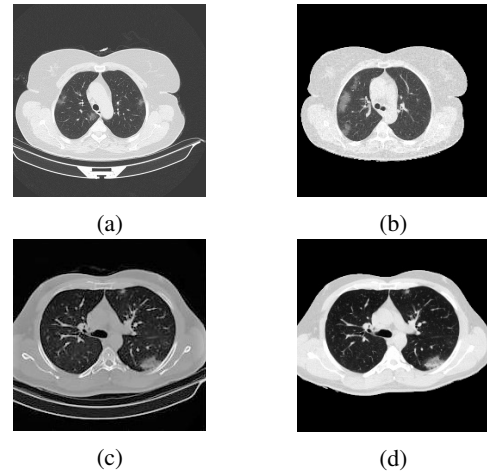

Fig. 1: (a), (b): Sample slices from COVIDx-CT dataset before and after preprocessing; (c), (d): Sample slices from COVID-CT-MD dataset before and after preprocesssing

thresholding step after performing Gaussian smoothing. The contour with the maximum area is taken as the body of the patient [9]. The obtained contour is then used to create a binary mask for the body, which is then multiplied with the original image to remove the background and obtain the RoI.

The images are resized to 224×224 pixels, the size for which the ResNet50 architecture was defined. The standard pre-processing functions used by the pre-trained ResNet50 model, are then applied before giving it as the input [15].

2) Preprocessing of CT slices from COVID-CT-MD dataset: At first, we normalise the images to an intensity range of 0-255. Then, we convert the data from float to unsigned integer values in order to reduce the memory requirements during training. Again, we segment out the RoI i.e. the body of the patient as done with COVIDx-CT dataset. Also, we increase the brightness of the images by a factor of 2 to match the brightness of the images with that of COVIDx-CT dataset.

Since the CT scans from different sources could have different number slices, we interpolate the given dataset using spline interpolation to increase the number of slices. The interpolated images are only used to train the slice level classifier. In order to keep the slice level labels intact we didn't interpolate the infected slices with non infected slices.

The images are then resized to a size of 224×224 pixels, the size for which the ResNet50 architecture was defined. The standard pre-processing functions used by the pre-trained ResNet50 model is then applied before giving it as the input [15].

B. Data Augmentation

To avoid overfitting, we apply the following data augmentation techniques to both the datasets during training of the slice level classifier.

- Rotation: -25° to 25°.
- Translation: -20 pixels to 20 pixels
- Zoom: 0.8 to 1.2
- Shear: -0.2 to 0.2

- Brightness: 0.8 to 1.2 times the original
- Random Horizontal flips

C. *Model*

The proposed model consists of three levels. At the first level, we train a slice-level classifier. This classifier extracts the features from each slice of a patient. These extracted features from all the slices of a patient are then used to train a patient-level classifier. We train four different classifiers for this purpose. At the final level, we train an ensemble classifier, which combines the predicted scores of each of the four patient-level classifiers. Fig. 2 shows the overall architecture of the proposed model. In this work, all the implementations are done using Keras with Tensorflow backend.
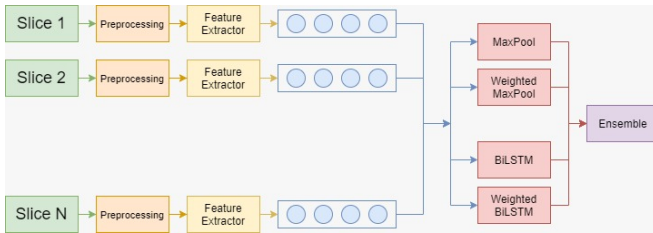

Fig. 2: Overall architecture of the proposed three-level model.

1) Slice-Level Classifier: The slice-level classifier i.e. the first level classifier, is the backbone of the proposed model. We train it to extract features from each CT-scan slice of the patient.

Architecture. We use ResNet50 [15] model pre-trained on ImageNet database. We remove the last layer of the model and replace it with a dropout layer and two fully connected layers of 1024 and 3 neurons. As the infection in different slices can be of different sizes, depending upon its severity, we consider adapting the multi-scale features. Hence, apart from the features of the $Stage5$ of ResNet50 (i.e, the last stage) we also use the features from the $Stage3$ and $Stage4$ [17]. We concatenate the features from each of these stages before passing them to the fully connected layer as shown in Fig. 3. ReLU activation is used in all the layers except for the last layer where we use softmax activation function. As the loss function, we use categorical cross-entropy with class weights to handle the class imbalance.

Training. The model is first trained on COVIDx-CT dataset. We train the complete model using Adam optimizer with a batch size of 64 and a learning rate of 10^{-3} for 10 epochs. Next we train the model on COVID-CT-MD dataset. First, we train only the last two layers with a learning rate of 10^{-4} for 10 epochs. Once the last layers of the model are trained we unfreeze all the layers except for the Batch normalization layers which is kept frozen throughout because if we update the running mean and variance the weights of the subsequent layers are rendered useless. Then we retrain the model with a lower learning rate of 10^{-5} for 10 epochs. We do not train the model on the normal slices of infected patients since these slices might have traces of infection hidden from human eyes.

Finally we use the data which has only the patient level labels to fine-tune the weights. We note the fact that the model performs reasonably well in classifying normal slices. Hence we use the model to get infected slices from the unlabeled dataset and then retrain the model for 25 epochs. We use Adam optimizer with a batch size of 64 in all the three training steps of COVID-CT-MD dataset. At each training step, we save the best model performing in terms of maximum accuracy on validation set.
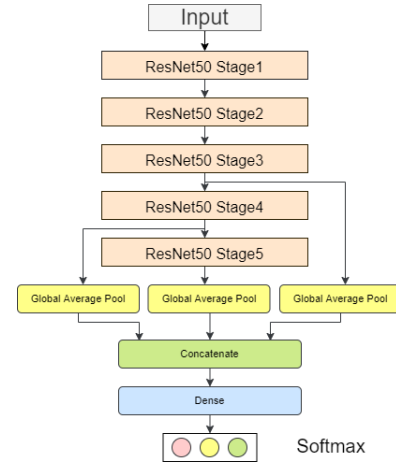

Fig. 3: Architecture of slice-level classifier

2) Patient-Level Classifier: After extracting the features from each slice corresponding to all the patients, we train four different classifiers at this level.

Feature Extraction. We extract two different types of features from each slice. First is a vector of dimension 1024 from the penultimate layer of slice-level classifier. Second is a vector of dimension 3072 created by concatenating the features after multiplying the probabilities of the three classes predicted by the slice level classifier. This is done so as to make it easier for the model to distinguish between the infected and normal slices.

Architecture. Next we train four patient level classifier:
- Maxpool: In this model we take the dimension-wise maximum of each slice features to get a single vector representative of the whole patient [23]. Then we train a simple two layer neural network with 512 and 3 neurons.
- Maxpool-Weighted: This model is same as Maxpool with feature vector replaced with the weighted features of dimension 3072.
- BiLSTM: In this model we train a Bidirectional Long Short Term Memory network [13] with 512 hidden units. Then we take the dimension-wise average of the output of each time step, over which we train a two layer neural network with 1024 and 3 neurons, as shown in Fig. 4.
- BiLSTM-Weighted: This model is same as BiLSTM with feature vector replaced with the weighted features of dimension 3072.

ReLU activation function is used in all the layers except for the last layer where we use softmax activation function. We

use categorical crossentropy as the loss function with class weights to handle class imbalance. L1 and L2 regularization with regularizing coefficient 10^{-5} and 10^{-4} is used.

Training. All the four models are trained using Adam optimizer with a batch size of 64 and a learning rate of 10^{-3} for 25 epochs. For all four models, we save the model with maximum accuracy on validation set.
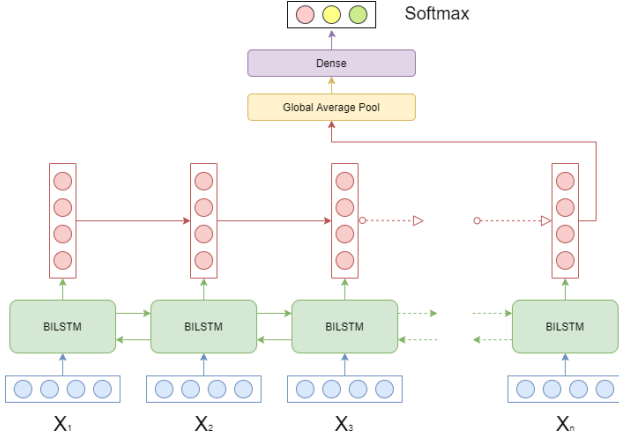

Fig. 4: Architecture of BiLSTM Network

3) Ensemble Learning: Predictions of four patient-level classifier need to be combined in an optimized manner to obtain the final prediction. Here, we implement ensemble learning for this purpose.

Architecture. We concatenate the scores of all the four patient-level classifiers to get a 12 dimensional vector corresponding to each patient. Then, we train a two layer neural network with 10 and 3 neurons.

We use categorical crossentropy as the loss function with class weights to handle class imbalance. L1 and L2 regularization with regularizing coefficient 10^{-5} and 10^{-4} is used.

Training. The model is trained using Adam optimizer with a batch size of 64 and a learning rate of 10^{-3} for 200 epochs. Again, the model with best validation accuracy is saved.

IV. Results and Discussion

The pre-trained model is validated on COVIDx-CT dataset with 6018 slices for COVID-19, 7395 slices for CAP and 12245 normal slices. The model achieves an accuracy of 93.06%. The slice-level model is validated on COVID-CT-MD dataset with 1333 slices for COVID-19, 436 slices for CAP and 3659 normal slices. The model achieves an accuracy of 95.76%. The patient-level model is validated on COVID-CT-MD dataset with 55 COVID-19 cases, 19 CAP cases and 24 normal cases. The model achieves an accuracy of 94.9%.

Finally, the model achieves an accuracy of 88.89% on the test dataset.

A. Grad-CAM

We visualise the convolution feature maps at three different stages using Grad-CAM [14] as shown in Fig. 5. It shows an infected slice of a COVID-positive patient from the validation
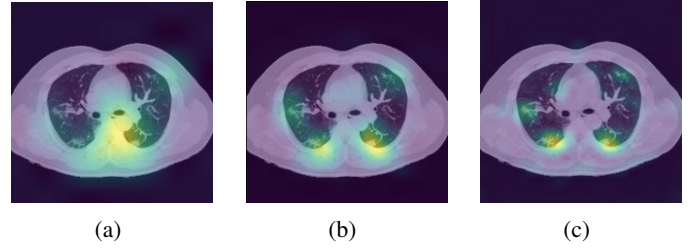

Fig. 5: Heat maps for outputs corresponding to (a) $Stage5$ (b) $Stage4$ (c) $Stage3$.

set. Clearly the model is able to locate the infected region in the lung slice and is able to make the correct prediction.

B. t-SNE Plots

We visualize the patient-level feature vector learned by the model for the validation and the test dataset using t-SNE as shown in Fig. 6. Clearly, we can see the formation of clusters of Normal, COVID-19 and CAP patients. It is important to note that the test dataset overlaps with the validation set indicating there is no significant distribution drift in the features extracted, which otherwise would have formed a separate cluster.
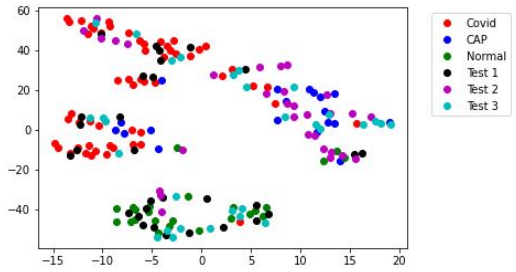

Fig. 6: Visualizing patient level features using t-SNE

TABLE I: Results on Validation Dataset

Model	Accuracy (%)	Sensitivity(%)		
		Normal	CAP	COVID
Pretrained	93.06	95.58	98.44	81.32
Slice Level	95.76	98.17	81.65	93.77
Patient Level	94.90	100.00	89.47	94.55

TABLE II: Results on Test Dataset

Accuracy (%)	Sensitivity(%)		
	Normal	CAP	COVID
88.89	88.57	90.00	88.57

V. Conclusion

The proposed work implements a three-level classifier to detect the cases of normal, COVID-19 and CAP using chest CT-scans. The model learns the features at slice-level, patient-level and then applies ensemble learning to combine the patient-level scores. Such multi-level training enhances the robustness of the model. Moreover, it uses residual network at multiple scales which covers different dimensions of infection and thereby helps in detecting the early stages of infection. The algorithm has performed quite well in terms of validation accuracy and sensitivity.

REFERENCES

[1] Ng, M.Y., Lee, E.Y., Yang, J., et al. (2020). Imaging profile of the covid-19 infection: radiologic findings and literature review. *Radiology: Cardiothoracic Imaging*, 2(1), 200034.

[2] Rodriguez-Morales, A., Cardona-Ospina, J., Gutierrez-Ocampo, E., et al. (2020). Clinical, laboratory and imaging features of covid-19: a systematic review and meta-analysis. *Travel Medicine and Infectious Disease*, 34, 101623.

[3] Salehi, S., Abedi, A., Balakrishnan, S., et al. (2020). Coronavirus disease 2019 (COVID-19): a systematic review of imaging findings in 919 patients. *American Journal of Roentgenology*, 215, 87–93.

[4] Zhao, W., Zhong, Z., Xie, X., et al. (2020). Relation between chest ct findings and clinical conditions of coronavirus disease (covid-19) pneumonia: a multicenter study. *American Journal of Roentgenology*, 214, 1072–1077.

[5] Watson, J., Whiting, P.F., Brush, J.E. (2020). Interpreting a covid-19 test result. *BMJ*, 369.

[6] Jin, Y.H., Cai, L., Cheng, Z.S., et al. (2020). A rapid advice guideline for the diagnosis and treatment of 2019 novel coronavirus (2019-ncov) infected pneumonia (standard version). *Military Medical Research*, 7, 4.

[7] WHO. (2020). Coronavirus disease 2019 (COVID-19): situation report.

[8] Wu, F., Zhao, S., Yu, B., et al. (2020). A new coronavirus associated with human respiratory disease in china. *Nature*, 579(7798), 265–269.

[9] Gunraj, H., Wang, L., Wong, A. (2020). COVIDNet-CT: A Tailored Deep Convolutional Neural Network Design for Detection of COVID-19 Cases From Chest CT Images. *Frontiers in Medicine*, 7, 1025.

[10] Heidarian, S., Afshar, P., Enshaei, N., et al.(2020). COVID-FACT: A Fully-Automated Capsule Network-based Framework for Identification of COVID-19 Cases from Chest CT scans.

[11] Afshar, P., Heidarian, S., Enshaei, N., et al.(2020). COVID-CT-MD: COVID-19 Computed Tomography (CT) Scan Dataset Applicable in Machine Learning and Deep Learning.

[12] Hofmanninger, J., Prayer, F., Pan, J. et al. (2020). Automatic lung segmentation in routine imaging is primarily a data diversity problem, not a methodology problem. *European Radiology* Exp 4, 50.

[13] Shoeibi, A., Khodatars, M., Alizadehsani, R., et al. (2020). Automated Detection and Forecasting of COVID-19 using Deep Learning Techniques: A Review.

[14] Selvaraju, R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., & Batra, D. (2019). Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization. *International Journal of Computer Vision*, 128(2), 336–359.

[15] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep Residual Learning for Image Recognition. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 770-778.

[16] Gao, K., Su, J., Jiang, Z., Zeng, L.L., Feng, Z., Shen, H., Rong, P., Xu, X., Qin, J., Yang, Y. & Wang, W. (2021). Dual-branch combination network (DCN): Towards accurate diagnosis and lesion segmentation of COVID-19 using CT images. *Medical image analysis*, 67, 101836.

[17] Hu, S., Gao, Y., Niu, Z., Jiang, Y., Li, L., Xiao, X., Wang, M., Fang, E.F., Menpes-Smith, W., Xia, J. & Ye, H. (2020). Weakly supervised deep learning for covid-19 infection detection and classification from ct images. *IEEE Access*, 8, 118869-118883.

[18] Wang, J., Bao, Y., Wen, Y., Lu, H., Luo, H., Xiang, Y., Li, X., Liu, C. & Qian, D. (2020). Prior-attention residual learning for more discriminative COVID-19 screening in CT images. *IEEE Transactions on Medical Imaging*, 39(8), 2572-2583.

[19] Amyar, A., Modzelewski, R., Li, H. & Ruan, S. (2020). Multi-task deep learning based CT imaging analysis for COVID-19 pneumonia: Classification and segmentation. *Computers in Biology and Medicine*, 126, 104037.

[20] Fan, D.P., Zhou, T., Ji, G.P., Zhou, Y., Chen, G., Fu, H., Shen, J. & Shao, L. (2020). Inf-net: Automatic covid-19 lung infection segmentation from ct images. *IEEE Transactions on Medical Imaging*, 39(8), 2626-2637.

[21] He, K., Zhao, W., Xie, X., Ji, W., Liu, M., Tang, Z., Shi, Y., Shi, F., Gao, Y., Liu, J. & Zhang, J. (2021). Synergistic learning of lung lobe segmentation and hierarchical multi-instance classification for automated severity assessment of COVID-19 in CT images. *Pattern recognition*, 107828.

[22] Gunraj, H. COVIDx CT: A large-scale chest CT dataset for COVID-19 detection. `https://www.kaggle.com/hgunraj/covidxct`

[23] Li, L., Qin, L., Xu, Z., Yin, Y., Wang, X., Kong, B., et al. (2020). Using Artificial Intelligence to Detect COVID-19 and Community-acquired Pneumonia Based on Pulmonary CT: Evaluation of the Diagnostic Accuracy. *Radiology* 296, E65–E71. doi:10.1148/radiol.2020200905