InRFNet: Involution Receptive Field Network for COVID-19 Diagnosis

Dhruv M\textsuperscript{1a)}, Sai Chandra Teja R\textsuperscript{1b), Sri Devi R\textsuperscript{2a), Nagesh Kumar S\textsuperscript{2b)}

\textsuperscript{1}CKM VIGIL Pvt Ltd, Hyderabad, Telangana, India, 
\textsuperscript{2}Department of Anaesthesiology, \textsuperscript{2}Department of Surgical Oncology, 
Sri Venkateswara Institute of Medical Sciences SVIMS Tirupati, Chittoor District, Andhra Pradesh, India.

Email: \textsuperscript{1a) dnakwana503@gmail.com, \textsuperscript{1b) saichandrateja@alumni.iith.ac.in, 
\textsuperscript{2a) srive.radhapuram@gmail.com, \textsuperscript{2b) nageshkumarsingaram@gmail.com

Abstract. COVID-19 is an emerging infectious disease that has been rampant worldwide since its onset causing Lung irregularity and severe respiratory failure due to pneumonia. The Community-Acquired Pneumonia (CAP), Normal, and COVID-19 Computed Tomography (CT) scan images are classified using Involution Receptive Field Network from Large COVID-19 CT scan slice dataset. The proposed lightweight Involution Receptive Field Network (InRFNet) is spatial specific and channel-agnostic with Receptive Field structure to enhance the feature map extraction. The InRFNet model evaluation results show high training (99\%) and validation (96\%) accuracy. The performance metrics of the InRFNet model are Sensitivity (94.48\%), Specificity (97.87\%), Recall (96.34\%), F1-score (96.33\%), kappa score (94.10\%), ROC-AUC (99.41\%), mean square error (0.04), and the total number of parameters (33100).

1. Introduction
Around the world, the new coronavirus infection (COVID-19) has spread rapidly. In late August 2021, the cumulative number of infected people worldwide was about 222 million, and the number of deaths was about 4.59 million [1]. Accurate diagnostic methods are needed to treat COVID-19 patients appropriately and prevent infection to others. The COVID-19 testing is done through RT-PCR (Reverse Transcription-Polymerase Chain Reaction) whose Sensitivity is as low as 42\% to 71\% [2]. On the other hand, to support doctors’ diagnosis and treatment in clinical settings, computer-aided diagnosis/detection (CAD) is being considered. CAD imaging results with quantitative evaluation help doctors make comprehensive judgments.

CAD is often developed using deep learning (DL) technology. In the application of DL to infectious diseases like COVID-19 spreading rampantly, there have been many efforts to develop a CAD model that estimates the disease at an early stage so that the infected patients can be isolated which may protect health care workers and other patients. Some of the DL models are diagnostic imaging models, non-imaging diagnostic models, prognosis prediction models, application to treatment development, medical staff, patient support model, and others. There is a global shortage of COVID-19 laboratories and test kits. There are many research cases as to whether computed tomography (CT) scanning of the lungs can be used as the first screening/alternative test for real-time RT-PCR. The Sensitivity of the COVID-19 test using CT images is as high as 97\% [3].
DL-type CAD is data-driven, and the outcome that determines performance depends on the image data collection. It is critical not only to have a big volume of data but also to have high-quality data. For training, preferably high-end hardware like GPU and annotated or labeled dataset is required. In DL, a system (also called a model) is built based on Convolution Neural Network (CNN) layers and is trained on a large amount of data. The medical images database has a small amount of data compared to general natural images. The Data Augmentation method is used to enrich the data through (image rotation, flipping, noise addition, etc.). In addition to this, there is a technology called Generative Adversarial Networks (GAN) which is employed to generate new images. By using the GAN technique termed CovidGAN [4] to create a chest X-ray image and learning it alongside the existing image, the detection accuracy was improved by 10%.

A DL model network configuration is roughly composed of two parts as shown in Fig.1. The first half part consists of a structure in which a convolution layer and a pooling layer are overlapped, and feature extraction is performed from the image. The latter half is composed of a dense layer and an output layer for image classification. In image processing, the convolution layer matches the filtering and highlights the picture's local characteristics. In addition, the pooling layer can absorb the image fluctuations due to translation and rotation making it robust against the position invariance. For a COVID-19 positive CT scan image used as an input in Fig.1, the output is classified as COVID-19 positive/normal (discrimination). In addition, in the output, the region of interest is surrounded by a rectangular frame (detection), or the site is determined on a pixel-by-pixel basis (segmentation).

Many DL models for the automated screening of COVID-19 from CT scans have been suggested in recent research examples. There is a case where researchers in the United States (Northwestern University), have developed an AI platform [5]. The developed AI algorithm called DeepCOVID-XR takes the approach of outputting COVID-19 negative or positive from the output of 6 types of DL models, and over 4000 positive images in the dataset. The study in [6] applies DL to AI-CAD that estimates COVID-19 infectious diseases at an early stage. A study by researchers at Mount Sinai School of Medicine in the United States reported diagnostic support by AI (conventional machine learning and CNN) based on CT image differentiation of lungs [7]. Since CT images are also 3D images, there are cases of AI-CAD development that display which area of the lung the infection has spread by utilizing the segmentation technology of the lung structure [8].

In addition, a method called Federated Learning [9] is introduced recently, wherein the learning data is decentralized without possessing it, on devices like smartphones, etc. There is an advantage of ensuring data privacy and security for DL models. Our paper focuses on building a custom end-to-end light-weight Involution Receptive Field Network (InRFNet) that can distinguish COVID-19 from healthy subjects in a large lung CT scan dataset.

The rest of the paper is arranged as follows. Section 2 presents the Involution Receptive Field Network (InRFNet). The experiment results and performance of the InRFNet are shown in Section 3, and the conclusion is drawn in Section 4.
2. Involution Receptive Field Network (InRFNet)

A novel Involution Neural Network based deep learning model with Receptive Field (RF) block is introduced.

CNNs are space invariant and depend on the channel characteristics. Recent research has proposed Involution [11], a new operation that is space-dependent and channel-independent, as opposed to convolution. CNNs are limited by inter-channel redundancies whereas in involution redundancy is less severe because of kernel sharing along channel dimensions.

The involution (Inv) kernel, unlike convolution (Conv) kernels, is reliant on the feature map but has an extendable and switchable spatial relationship. The Inv kernel is based on a single pixel rather than its interaction with its neighbor's pixels to form an improved feature map.

A Custom architecture named InNet (Fig.2) based on Involution Neural Network is built for image classification and better visual identification. To construct improved feature representation combining low-level and high-level information, it requires high computation deep neural networks. The feature map calculated with the same kernel size, uniform resolution, and fixed sampling grid makes it less robust and not emphasized (Fig.3).

Receptive Field (RF) [12] enhances feature representation through varying kernels with different sizes and multi-branches. Feature representation channels can be decreased by using various kernels of different size and deeper non-linear layers. The proposed novel InRFNet (Fig.4) is a lightweight Involution Receptive Field network with enhanced feature representation and less parameters. To minimize the number of parameters the original CxC Inv-layer is replaced by 1xC Inv-layer plus Cx1 Inv-layer (Fig.5).

Figure 2. Architecture of InNet

Figure 3. Involution Network.
3. Experiments
The performance of the InNet and InRFNet is evaluated on the large lung CT scan dataset [10]. There are three classes in the dataset: Community-Acquired Pneumonia (CAP), Normal, and COVID-19. Total images: 17104, with CAP images: 2618, normal images: 6893, and COVID-19 images: 7593 as distribution among them. Images are augmented by horizontal and vertical flip, rotation, width and height shifting, and zooming. The experiments are executed on Windows system Intel® Core™ i7-9750H CPU@ 2.60GHz with Tesla RTX 2060 Graphics card, using TensorFlow-GPU 2.6, and CUDA 11.1. Different parameter settings of the proposed InNet and InRFNet are shown in Table 1.

| Table 1. Experimental Parameter Settings |
|-----------------------------------------|
| Description                              | Value                        |
| Learning rate                            | 0.001                        |
| Image size                               | 128x128                      |
| Batch size                               | 64                           |
| Optimizer                                | Adam                         |
| Number of Epochs                         | 30                           |
| Validation-split                         | 20                           |
| Loss                                     | categorical-cross entropy    |
Table 2 Performance of InRFNet and InNet

| Model   | Class      | Precision | Recall | F1-Score |
|---------|------------|-----------|--------|----------|
| InRFNet | Normal     | 97%       | 94%    | 96%      |
|         | COVID-19   | 95%       | 98%    | 96%      |
|         | CAP        | 98%       | 99%    | 98%      |
| InNet   | Normal     | 96%       | 93%    | 94%      |
|         | COVID-19   | 94%       | 96%    | 95%      |
|         | CAP        | 98%       | 99%    | 99%      |

Figure 6. InRFNet accuracy graph

Figure 7. InNet accuracy graph.

Figure 8. InRFNet loss graph

Figure 9. InNet loss graph.
Table 3 Performance comparison of the proposed InNet and InRFNet with other models

| Model                  | Recall | Precision | F1 Score | Accuracy | ROC AUC |
|------------------------|--------|-----------|----------|----------|---------|
| DenseNet-121           | 87.55  | 94.42     | 90.85    | 92.57    | 91.37   |
| Residual Attention-92  | 90.99  | 90.47     | 90.73    | 91.76    | 94.39   |
| Ensemble with FC       | 89.84  | 98.32     | 93.89    | 95.07    | 96.72   |
| Ensemble with FC + SVM | 90.80  | 97.93     | 94.23    | 95.31    | 98.06   |
| InNet                  | 95.20  | 95.20     | 95.19    | 95       | 98.96   |
| InRFNet                | 96.34  | 96.34     | 96.33    | 96       | 99.41   |

In the training phase proposed InNet architecture provides sensitivity, specificity, kappa score, the total number of parameters, and mean square error of 92.75%, 96.67%, 92.25%, 65867, and 0.05. The training accuracy and validation accuracy of InNet are 98% and 95% respectively with a recall score of 95%. The proposed InRFNet model achieved 99% training accuracy and 96% validation accuracy with a high recall score of 96%. Furthermore, the InRFNet model also obtained a 94.10% kappa score, 94.48% Sensitivity, 97.87% Specificity, 0.04 mean square error and 33100 total number of parameters. Table 2 shows Class-wise classification scores.

Fig.6 and Fig.7 shows InRFNet and InNet accuracy graphs. The InRFNet model has reached 95% training accuracy and 93% validation accuracy by 10 epochs. InRFNet and InNet loss graphs are shown in Fig.8 and Fig.9. InRFNet and InNet training loss is as low as 0.01 and 0.04 after 30 epochs respectively.

Performance comparison of the proposed InNet and InRFNet with existing models is shown in Table 3, it is observed that DenseNet-121 [14] has better precision and accuracy than Residual Attention-92 [13], the latter outperforms the former in terms of recall and ROC AUC measures. InRFNet outperforms Ensemble with FC + SVM. However, the average recall of InRFNet is 96.34% for the validation dataset which is 5.54% higher than Ensemble with FC + SVM which demonstrates the effectiveness of InRFNet for Federated Learning.

4. Conclusion

COVID-19 infection is still widespread worldwide even now that vaccination has begun. The deep learning (DL) based computer aided diagnosis (CAD) model which helps in screening the infections to prevent the transmission would help reduce the pressure on medical institution resources. With high accuracy and recall scores, the proposed InRFNet: Involution Receptive Field Network has demonstrated efficient classification.

InRFNet has obtained better performance metrics with less parameters making it suitable for low computation devices like smartphones in federated learning. InRFNet's discriminating capability is illustrated by its high sensitivity and maximum ROC-AUC values, making it a reliable tool for supporting health care systems.

References

[1] Coronavirus Update(accessed on September 9 2021)
https://www.worldometers.info/coronavirus/

[2] Simpson S, Kay FU, Abbara S, et al. (2020) Radiological Society of North America Expert Consensus Statement on Reporting Chest CT Findings Related to COVID-19. Endorsed by the Society of Thoracic Radiology, the American College of Radiology, and RSNA Radiology Cardiothoracic Imaging 2(2)

[3] Ai T, Yang Z, Hou H, et al. (2020) Correlation of Chest CT and RT-PCR Testing for Coronavirus Disease 2019 (COVID-19) in China: A Report of 1014 Cases. Radiology 296(2)

[4] A. Waheed, M. Goyl, D. Gupta. et al.(2020) CovidGAN: Data augmentation using auxiliary classifier GAN for improved covid-19 detection, IEEE Access, Vol 8, No 8, pp 91916-91923
[5] RM. Wehbe, J. Sheng, S. Dutta, et al. (2020) DeepCOVID-XR: An artificial intelligence algorithm to detect COVID-19 on chest radiographs trained and tested on a large US clinical dataset, *Radiology*

[6] K. Hammoudi, H. Benhabiles, M. Melkemi, et al. (2020) Deep learning on chest X-ray images to detect and evaluate pneumonia cases at the era of COVID-19, *arXiv:2004.03399*

[7] X. Mei, H-C. Lee, K. Diao, et al. (2020) Artificial intelligence–enabled rapid diagnosis of patients with COVID-19, *Nature Medicine*, Vol 26, pp 1224-1228

[8] F. Shan, Y. Gao, J. Wang, et al. (2020) Abnormal lung quantification in chest CT images of COVID-19 patients with deep learning and its application to severity prediction. *Medical Physics*

[9] Jakub Konečný and H. Brendan McMahan and Felix X. Yu and Peter Richtárik and Ananda Theertha Suresh and Dave Bacon. (2017) Federated Learning: Strategies for Improving Communication Efficiency, *arXiv:1610.05492*

[10] Maftouni, M., Law, A.C, Shen, B., Zhou, Y., Yazdi, N., and Kong, Z.J. (2021) A Robust Ensemble-Deep Learning Model for COVID-19 Diagnosis based on an Integrated CT Scan Images Database, in Proceedings of the 2021 Industrial and Systems Engineering Conference

[11] Duo Li and Jie Hu and Changhu Wang and Xiangtai Li and Qi She and Lei Zhu and Tong Zhang and Qifeng Chen. (2021) Involution: Inverting the Inherence of Convolution for Visual Recognition, in Proceedings of the IEEE conference on computer vision and pattern recognition, *arXiv:2103.06255*

[12] Songtao Liu and Di Huang and Yunhong Wang. (2017) Receptive Field Block Net for Accurate and Fast Object Detection, *arXiv:1711.07767*

[13] F. Wang et al., (2017) Residual attention network for image classification, in Proceedings of the IEEE conference on computer vision and pattern recognition, *arXiv:1704.06904* pp 3156-3164

[14] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, (2017) Densely connected convolutional networks, in Proceedings of the IEEE conference on computer vision and pattern recognition, *arXiv:1608.06993* pp 4700-4708