ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) | 978-1-7281-7605-5/20/$31.00 ©2021 IEEE | DOI: 10.1109/ICASSP39728.2021.9414745

A MULTI-STAGE PROGRESSIVE LEARNING STRATEGY FOR COVID-19 DIAGNOSIS USING CHEST COMPUTED TOMOGRAPHY WITH IMBALANCED DATA

Zaifeng Yang[1], Yubo Hou[2*], Zhenghua Chen[2], Le Zhang[2] and Jie Chen[3]*

[1] Institute of High Performance Computing, A*STAR, Singapore
[2] Institute for Infocomm Research, A*STAR, Singapore
[3] Hong Kong Baptist University, Hong Kong, China

ABSTRACT

In this paper, a multi-stage progressive learning strategy is investigated to train classifiers for COVID-19 Diagnosis using imbalanced Chest Computed Tomography Data acquired from patients infected with COVID-19 Pneumonia, Community Acquired Pneumonia (CAP) and from normal healthy subjects. In the first learning stage, pre-processed volumetric CT data together with the segmented lung masks are fed into a 3D ResNet module, and an initial classification result can be obtained. However, due to categorical data imbalance, we observe large differences in sensitivity between COVID-19 and CAP cases. In the second stage, five learning models are independently trained over data with only COVID-19 and CAP cases, and are then ensembled to further discriminate the two classes. The final classification results are obtained by combining the predictions from both stages. Based on the validation dataset, we have evaluated our method and compared it with up-to-date methods in terms of overall accuracy and sensitivity for each class. The validation results validate the accuracy of the proposed multi-stage learning strategy. The overall accuracy of the validation dataset is 88.8%, and the sensitivities are 0.873, 0.789 and 1 for COVID-19, CAP and normal cases, respectively.

***Index Terms*—** COVID-19, computed tomography, multi-stage learning, imbalanced data classification, 3D ResNet

1. INTRODUCTION

Coronavirus disease 2019 (COVID-19) has been spread to almost every corner of the world since December 2019. The reverse transcription-polymerase chain reaction (RT-PCR) test is accepted as the most effective tool to examine whether a person is carrying COVID. However, the accuracy may be as low as 60%-70%. Thus the patients could be infected by COVID-19 with lung abnormalities but obtain an initially negative RT-PCR results [1]. At the beginning of the outbreak in Wuhan, China, due to the lack of the RT-PCR test, chest computed tomography (CT) entered the fight against this unknown virus [2]. Subsequently, more and more countries started to use CT for COVID diagnosis as a complementary measure in addition to RT-PCR [3]. In the CT diagnosis for COVID-19, features such as bilateral patchy shadows or ground glass opacity in lung can be scanned and identified by professional radiologists [4]. However, radiologists need to analyze each slice of the 3D reconstructed lung for a large number of patients, and it will inevitably result in lower accuracy and efficiency.

With the rapid development of deep learning-based artificial intelligence in medical diagnoses through the past few years, AI-assisted CT diagnosis for lung diseases could improve the diagnostic accuracy and reduce the radiologists' overwhelming workload [5, 6]. In the ongoing fight against the COVID-19 pandemic, various deep learning-based methods have been developed for the chest CT data analysis and classification [7–9]. In [7], ResNet50 is used as the backbone. The CNN features from each slice of the CT series are combined by a max-pooling operation and the resulting feature map is fed into a fully connected layer to generate a probability score for each class. In addition to supervised learning, a weakly supervised deep learning framework was developed using 3D CT volumes for COVID-19 classification and lesion localization [10], but it only applies for COVID-19 and non-COVID classes. Similarly, The proposed method in [11] can minimize the requirements of manual labeling of CT images but still cannot obtain accurate infection detection and distinguish COVID-19 from non-COVID-19 cases such as community acquired pneumonia and non-pneumonia scans. In addition to COVID-19 diagnosis, many other attempts based on AI have also been made to fight COVID-19 [12]. ResNet is a widely used deep neural network for classification problems. However, for a problem with 3D data such as CT, MRI and videos, the capacity of the 3D CNN based ResNet is highly constrained due to the expensive computational cost and memory demand. Thus, the full 3D ResNet with deep depth is difficult to be leveraged.

This paper was submitted for the Signal Processing Grand Challenges (SPGC) Programme in 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP '21).
Z. Yang[1*] and Y. Hou[2*] share equal contributions to this work, and Y. Hou is the corresponding author.

This paper presents a multi-stage progressive learning strategy to train classifiers for COVID-19 Diagnosis using imbalanced Chest Computed Tomography Data acquired from patients infected with COVID-19 Pneumonia, CAP and from normal healthy subjects. In Section II, the data analysis and data preprocessings, including lung segmentation masks, window settings, cropping and resizing, will be introduced. In the next Section, the pre-processed volumetric CT data together with the segmented lung masks are fed into a 3D ResNet module, and an initial classification result can be obtained. The detailed architecture of the Stem, 3D ResNet101 module, the classifier and the data augmentation will be described. After the first learning stage, the rest learning stages for improving the sensitivities will be introduced accordingly in Section IV. Five learning models are independently trained over data with only COVID-19 and CAP cases, and are then ensembled to further discriminate the two classes. The final classification results are obtained by combining the predictions from both stages. In the last Section, we have evaluated our method and compared it with up-to-date methods in terms of overall accuracy and sensitivity for each class based on the validation dataset. The experimental results validate the accuracy of the proposed multi-stage learning strategy.

2. DATA PREPROCESSINGS

2.1. The Acquired CT Dataset

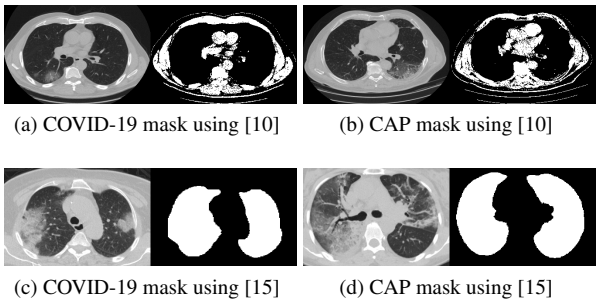
The dataset released by the challenge committee includes volumetric CT scans of 171 patient positive for COVID-19, 60 Community Acquired Pneumonia (CAP), and 76 normal cases [13]. The thickness of all the CT slices are 2 mm, and the dimension of each slice is 512×512. 30% of the data were randomly selected as the validation set by the challenge committee. In the validation dataset, the number of the COVID-19, cap and normal cases are 55, 19 and 24, respectively. It should be noted that the numbers of each class in both training and validation dataset are not close to each other, so the accuracy of the validation dataset should not be the only indicator to assess the quality of the trained model. The sensitivity is also significant to evaluate the accuracy of each class.

2.2. Data Preprocessings

The body and the lung shapes of each patient differ from each other. In addition, the CT settings such as dose and reconstruction regions by the radiologist could also be different. Thus, it is indispensable to process the acquired CT data for a stable input of the subsequent training.

2.2.1. Window of the Hounsfield Scale

From the provided CT files in DICOM format, the volumetric CT data in Hounsfield unit can be extracted by stacking the slices sorted by the z-position. The window of the Hounsfield scale is set to [-1200, 400] HU [14]. Using this window, most features within the lung area such as lung tissues and infections can be well persevered, while the other body parts including bones, fat and liver can be remained as constant.



(a) COVID-19 mask using [10] (b) CAP mask using [10]

(c) COVID-19 mask using [15] (d) CAP mask using [15]

Fig. 1: Comparison of the lung segmentation for COVID-19 and CAP cases using [10] and [15].

2.2.2. Lung Segmentation

In [15], it is concluded that accurate lung segmentation does not require complex methodology and a proven deep learning-based segmentation architecture yields state-of-the-art results once diverse (but not necessarily larger) training data are available. The U-net(R231CovidWeb) model [16] is used for the lung segmentation in our preprocessing. It is obvious that the lung segmentation from [10] is noisy and the infectious area in the lung cannot be well segmented, as shown in Fig. 1. Moreover, the masks in Fig. 1(a) and (b) are noisy even in the chest fat and bone region, which will definitely cause errors in the subsequent training. The generated mask is concatenated with the CT data as the input of the proposed deep learning model.

2.2.3. Crop and Resize

To further reduce the computational demand for a more efficient training, the dimensions for each slice are cropped to [224,336] based on the generated lung mask. The 3D CT data together with the lung segmentation data can be both reduced by cropping according to the boundaries of lung mask. To get a constant dimension along slice dimension for each CT data, both the CT and segmentation data are resized to [224,336] based on spline interpolation. Note that there is no need to implement isotropic processing based on physical dimension because multiple interpolations will cause accuracy degradation and the infectious lung features could not be well recognized by the trained model. In addition, the distance between each slice could be up to 5 mm, and the interpolated isotropic CT data could be totally meaningless.

Table 1: The Structure of Classifier.

Stage	Layers	Output size
Classifier	AdaptiveMaxPool3d(8,7,11)	[512, 8,7,11]
	Conv3d(512,128, kernel_size= [3,3,3])	[128,8,7,11]
	ReLU+AdaptiveMaxPool3d(4,4,6)	[128,4,4,6]
	Dropout3d(p = 0.5)	[128,4,4,6]
	Conv3d(128,32, kernel_size= [3,3,3])	[32,4,4,6]
	ReLU+AdaptiveMaxPool3d(1,1,1)	[32,1,1,1]
	Linear(32,32)+ ReLu	32
	Linear(32,3) + Relu	3

3. THE PROPOSED TRAINING METHOD

To feed the train model with segmented lung volume, the data will be passed into the stem module and pseudo-3D ResNet by concatenating the segmented lung mask with the pre-processed CT data. The structure of the deep learning based model and data augmentation will be introduced in the following subsections.

3.1. The Architecture of the Training Model

In the Stem module, the input data size after preprocessing is (2, T, 224, 336), where T is the number of the CT slices for each patient. Borrowing the ideas from [10, 17, 18] , the proposed pseudo 3D ResNet101 comprises of 2D spatial convolutions ($1\times3\times3$) along slice plane and ($3\times1\times1$) along z-direction (normal to slice plane). The alternating dimensional convolutions are used to achieve comparative performance of 3D convolution kernel. Taking advantages of the alternating 3D (two 2D spatial convolutions) kernels, deeper neural network can be used for higher accuracy. 3D ResNet101 (3, 4, 23, 3) is used as the backbone of the proposed method. The architecture of the proposed method including the stem module and 3D ResNet101 is illustrated in Fig. 2. One layer of the 3D ResNet101 named bottleneck transform layer is shown in this figure. 3D Instance normalization is adopted throughout the model. After four-time expanse of the feature channels, the output feature channel is expanded from 16 to 512.

The slice number of each CT is different because the acquisition setting for each patient varies. Before the output results are fed into the classifier, adapativeMaxPool is applied to ensure the output size, especially for the number of slices, should remain to be equal, as illustrated in Table 1. Subsequently, a progressive classifier is leveraged to reduce the feature channels from 512 to 3 step by step. A dropout3d layer is introduced here to avoid overfitting.

3.2. Data Augmentation

In practical, the CT equipment settings differ from various hospitals and radiologists. Additionally, the different dosage and reconstruction methods also cause variations on the acquired CT data. Data augmentation [19] are adopted to avoid overfitting during training, and take these variations into considerations. The following affine transformation parameters are utilized: brightness (50%) and contrast adjustment(30%), rotation (10 degrees), horizontal and vertical translations (10%), scaling (factor is from 1 to 1.2) and shearing (0.1) in the width dimension.

4. THE PROPOSED MULTI-STAGE PROGRESSIVE STRATEGY

In the first learning stage, as discussed in the previous section, the preprocessed volumetric CT data together with the segmented lung mask are fed into a 3D ResNet module. Thus, an initial classified results can be obtained. As long as the output results from the first stage belong to COVID-19 cases, further classification in the second stage will be executed. In this section, the subsequent stages including ensembled trained models to improve the sensitivity of imbalanced training data will be introduced.

4.1. Ensembled Binary Classification Model

Due to the imbalanced data over the three classes (the number of the COVID-19 cases are three times more than CAP cases) and similarity of CAP and COVID-19, most of CAP cases would be classified as COVID-19 by mistake in the first stage. Ensembling based on different combination of the training data is able to improve the performance when the training data is imbalanced [20, 21]. At the second stage, we prepared five training datasets by sampling from the COVID-19 cases. For the first three groups, the numbers of the training data for COVID-19 cases are all 39, which is identical to the number of CAP cases. For the rest two groups, the numbers of the COVID-19 are 60, with slightly larger than the number of CAP cases. These five deep learning models are independently trained over data with only COVID-19 and CAP cases, and are then ensembled to further discriminate the two classes. The structures of the five deep learning models are similar to the one shown in Table 1 except the output size is 2. Note that cross-entropy loss is used at the first stage for multi-class problem, and the binary cross-entropy loss is applied for the five models at the second stage. Ensembling via voting from each model will give the final decision for COVID-19 and CAP cases.

4.2. Final Combining Stage

The final classification results are obtained by combining the predictions from both stages. The normal cases are determined directly by the results from the first stage, and the COVID-19 cases are from the second stage via ensembling. The CAP cases from the first stage are merged with the ensembled results at the second stage.

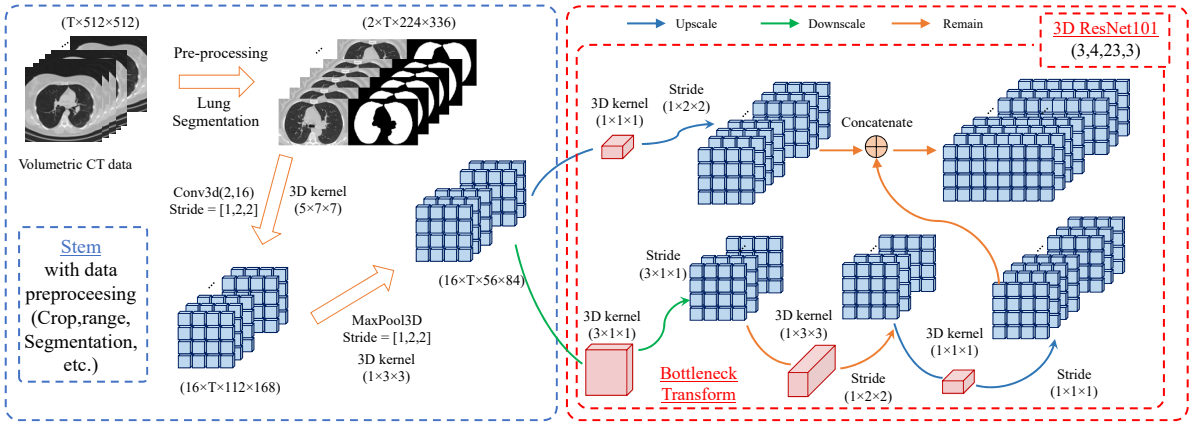


Fig. 2: Architecture of the Stem and 3D ResNet101 (bottleneck transform) at the first learning stage.

5. EXPERIMENTAL RESULTS AND ANALYSIS

We have trained our proposed multi-stage learning method using a PC with dual Intel Xeon Gold 6240 (36 cores) CPUs and two NVIDIA Geforce RTX 2080 Ti GPUs under Ubuntu operating system. The initial learning rate is 1e-4, and it will be decayed after 120 epochs by 0.995 for each following epoch. Cross-entropy loss is used for the first stage for multi-class problem, and binary cross-entropy loss is leveraged for the five models at the second stage.

For the validation dataset chosen by the challenge committee, the number of the COVID-19, cap and normal cases are 55, 19 and 24, respectively. It is obvious the validation dataset is also imbalanced among the three classes, so the accuracy on the validation dataset should not be the only indicator to evaluate the performance of the trained model. The sensitivities of each class is also important, especially for the CAP cases. The accuracy and sensitivities at different stages are given in Table 2. At the second stage where only COVID-19 and CAP cases are involved in the training, the sensitivity varies due to the different COVID-19 training data in these five models. By voting from five different trained models, the final decision on the COVID-19 and CAP cases can be determined. At the final stage, it can be seen that the sensitivity of the CAP cases is increased from 0.632 to 0.789, while the overall accuracy is only dropped a little from 0.897 to 0.888, compared with the result from the first stage.

Moreover, comparing with a recent published method [10] (altered to multi-class classification and used the data from the challenge), the proposed method outperforms in both accuracy and sensitivities of each classes. The overall accuracy is increased by 11.2%, and the sensitivities are increased by 0.037, 0.368, and 0.083 for COVID-19, CAP and normal cases, respectively.

Table 2: Overall accuracy and sensitivities of each classes at different learning stage.

Model		Accuracy	Sensitivity		
			COVID-19	CAP	Normal
Baseline [10]		0.776	0.836	0.421	0.917
Stage1		0.897	0.945	0.632	1
Stage2	1st	0.824	0.8	0.895	-
	2nd	0.703	0.655	0.842	-
	3rd	0.838	0.855	0.789	-
	4th	0.797	0.855	0.632	-
	5th	0.784	0.782	0.789	-
Final		0.888	0.873	0.789	1

6. CONCLUSION

In this paper, a multi-stage progressive learning strategy has been proposed to train classifiers for COVID-19 Diagnosis using imbalanced chest CT data acquired from patients infected with COVID-19, CAP and from normal cases. In the first learning stage, pre-processed volumetric CT data together with the segmented lung masks are fed into a 3D ResNet module, and an initial classification result can be obtained. In the second stage, five learning models are independently trained over data with only COVID-19 and CAP cases, and are then ensembled to further discriminate the two classes. The final classification results are obtained by combining the predictions from both stages. We have evaluated our method using validation dataset and compared it with up-to-date methods in terms of overall accuracy and sensitivity for each class. It is promising that the proposed method can be further applied in practical low-dose CT, CT with large slice thickness, and for those patients with the other lung diseases such as lung tumor.

7. REFERENCES

[1] J. P. Kanne, B. P. Little, J. H. Chung, B. M. Elicker, and L. H. Ketai, "Essentials for radiologists on covid-19: an update—radiology scientific expert panel," 2020.

[2] C. Huang, Y. Wang, X. Li, L. Ren, J. Zhao, Y. Hu, L. Zhang, G. Fan, J. Xu, X. Gu *et al.*, "Clinical features of patients infected with 2019 novel coronavirus in wuhan, china," *The lancet*, vol. 395, no. 10223, pp. 497–506, 2020.

[3] F. Pan, T. Ye, P. Sun, S. Gui, B. Liang, L. Li, D. Zheng, J. Wang, R. L. Hesketh, L. Yang *et al.*, "Time course of lung changes on chest ct during recovery from 2019 novel coronavirus (covid-19) pneumonia," *Radiology*, 2020.

[4] T. C. Kwee and R. M. Kwee, "Chest ct in covid-19: what the radiologist needs to know," *RadioGraphics*, vol. 40, no. 7, pp. 1848–1865, 2020.

[5] M. Anthimopoulos, S. Christodoulidis, L. Ebner, A. Christe, and S. Mougiakakou, "Lung pattern classification for interstitial lung diseases using a deep convolutional neural network," *IEEE transactions on medical imaging*, vol. 35, no. 5, pp. 1207–1216, 2016.

[6] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. Van Der Laak, B. Van Ginneken, and C. I. Sánchez, "A survey on deep learning in medical image analysis," *Medical image analysis*, vol. 42, pp. 60–88, 2017.

[7] L. Li, L. Qin, Z. Xu, Y. Yin, X. Wang, B. Kong, J. Bai, Y. Lu, Z. Fang, Q. Song *et al.*, "Artificial intelligence distinguishes covid-19 from community acquired pneumonia on chest ct," *Radiology*, 2020.

[8] F. Shan, Y. Gao, J. Wang, W. Shi, N. Shi, M. Han, Z. Xue, and Y. Shi, "Lung infection quantification of covid-19 in ct images with deep learning," *arXiv preprint arXiv:2003.04655*, 2020.

[9] Y. Song, S. Zheng, L. Li, X. Zhang, X. Zhang, Z. Huang, J. Chen, H. Zhao, Y. Jie, R. Wang *et al.*, "Deep learning enables accurate diagnosis of novel coronavirus (covid-19) with ct images," *MedRxiv*, 2020.

[10] X. Wang, X. Deng, Q. Fu, Q. Zhou, J. Feng, H. Ma, W. Liu, and C. Zheng, "A weakly-supervised framework for covid-19 classification and lesion localization from chest ct," *IEEE Transactions on Medical Imaging*, vol. 39, no. 8, pp. 2615–2625, 2020.

[11] S. Hu, Y. Gao, Z. Niu, Y. Jiang, L. Li, X. Xiao, M. Wang, E. F. Fang, W. Menpes-Smith, J. Xia *et al.*, "Weakly supervised deep learning for covid-19 infection detection and classification from ct images," *IEEE Access*, pp. 118 869–118 883, 2020.

[12] J. Bullock, A. Luccioni, K. H. Pham, C. S. N. Lam, and M. Luengo-Oroz, "Mapping the landscape of artificial intelligence applications against covid-19," *Journal of Artificial Intelligence Research*, vol. 69, pp. 807–845, 2020.

[13] P. Afshar, S. Heidarian, N. Enshaei, F. Naderkhani, M. J. Rafiee, A. Oikonomou, F. B. Fard, K. Samimi, K. N. Plataniotis, and A. Mohammadi, "Covid-ct-md: Covid-19 computed tomography (ct) scan dataset applicable in machine learning and deep learning," *arXiv preprint arXiv:2009.14623*, 2020.

[14] E. A. Kazerooni and B. H. Gross, *Cardiopulmonary imaging*. Lippincott Williams & Wilkins, 2004, vol. 4.

[15] J. Hofmanninger, F. Prayer, J. Pan, S. Röhrich, H. Prosch, and G. Langs, "Automatic lung segmentation in routine imaging is primarily a data diversity problem, not a methodology problem," *European Radiology Experimental*, vol. 4, no. 1, pp. 1–13, 2020.

[16] Automated lung segmentation in ct under presence of severe pathologies. [Online]. Available: https://github.com/JoHof/lungmask

[17] Z. Qiu, T. Yao, and T. Mei, "Learning spatio-temporal representation with pseudo-3d residual networks," in *proceedings of the IEEE International Conference on Computer Vision*, 2017, pp. 5533–5541.

[18] J. Carreira and A. Zisserman, "Quo vadis, action recognition? a new model and the kinetics dataset," in *proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 6299–6308.

[19] Y. Xu, R. Jia, L. Mou, G. Li, Y. Chen, Y. Lu, and Z. Jin, "Improved relation classification by deep recurrent neural networks with data augmentation," *arXiv preprint arXiv:1601.03651*, 2016.

[20] T.-Y. Liu, "Easyensemble and feature selection for imbalance data sets," in *2009 international joint conference on bioinformatics, systems biology and intelligent computing*. IEEE, 2009, pp. 517–520.

[21] M. Galar, A. Fernandez, E. Barrenechea, H. Bustince, and F. Herrera, "A review on ensembles for the class imbalance problem: bagging-, boosting-, and hybrid-based approaches," *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 42, no. 4, pp. 463–484, 2011.