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/ICASSP39728.2021.9413707

DIAGNOSING COVID-19 FROM CT IMAGES BASED ON AN ENSEMBLE LEARNING FRAMEWORK

Bingyang Li Qi Zhang Yinan Song Zhicheng Zhao Zhu Meng Fei Su

Beijing University of Posts and Telecommunications, China

ABSTRACT

Research on automated diagnosis of Coronavirus Disease 2019 (COVID-19) has increased in recent months. SPGC COVID19 aims at classifying the grouped images of the same patient into COVID, Community Acquired Pneumonia(CAP) or normal. In this paper, we propose a novel ensemble learning framework to solve this problem. Moreover, adaptive boosting and dataset clustering algorithms are introduced to improve the classification performance. In our experiments, we demonstrate that our framework is superior to existing networks in terms of both accuracy and sensitivity.

Index Terms— COVID-19, Deep Learning, Computed Tomography(CT) Scan Image, CAP, Ensemble Learning

1. INTRODUCTION

COVID-19 is an ongoing pandemic and has spread to more than 200 countries and territories worldwide. The World Health Organization declared COVID-19 as a public health emergency when involved 120,383,919 confirmed cases of COVID-19, including 2,664,386 deaths (as of 17 March, 2021).

The COVID-19 pandemic is caused by Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) which can be confirmed by using the Reverse Transcription Polymerase Chain Reaction (RT-PCR) test[1]. Though considered as the gold standard for SARS-CoV-2 diagnosis, the RT-PCR test is actually time-consuming with a low sensitivity and a high false-negative rate, especially in the early stage which has posed giant challenges to prevent the spread of the pandemic[2]. The clinical experience has implied that CT and Chest Radiographs are more efficient when identifying COVID-19 infection or diagnosing pneumonia with a high sensitivity[3]. Unfortunately, CT scans may share similar features between COVID-19 and other types of pneumonia, which may be considered as CAP, leading to a high false-negative rate. Hence, it is challenging to distinguish COVID-19 from other viral pneumonia.

A few previous works based on artificial intelligence have made significant advancement for automated screening of COVID-19 from chest CT scans due to the outstanding performance in feature representation[4]. For example, Wang et al.[5] modified the Inception model to establish the supervised model. However, the supervised learning algorithm requires a considerable scale of manual annotations of CT images, thus it is unrealistic. Some successors focused on 3D convolutional neural networks. Han et al.[6] proposed an attention-based deep 3D multiple instance learning (AD3D-MIL) approach in which labels were assigned to CT scans as a bag of instances. AD3D-MIL was a weakly-supervised learning framework in which attention mechanism and deep multiple instance learning were combined. However, 3D convolution is obviously more complicated than 2D convolution.

In this paper, we propose a hierarchical classifer framework which mainly includes two aspects. 1) Based on semi-supervised learning, a deep network is trained to extract features on slice-level and generate contextual features for each patient (including patients without slice-level label). 2) Taking the extracted features as the input, a sequence classifier based on supervised learning is constructed to classify COVID-19 cases from CAP and normal cases.

The main contributions of this paper can be summarized as follows:

- A novel ensemble learning framework via integrating the semi-supervising and supervised learning methods is proposed to distinguish COVID-19 from CAP and normal cases.
- To deal with different data distributions, an adaptive boosting and dataset clustering method is introduced to improve the generalization capability of the model.
- The experimental results demonstrate the effectiveness of the proposed algorithm.

This work is supported by The National Key Research and Development Program of China (2020YFB2104604), BUPT innovation and entrepreneurship support program, and Chinese National Natural Science Foundation (U1931202, 62076033)

The remainder of this paper is organized as follows: Section 2 introduces the proposed methods. Section 3 analyzes the results of experiments on the SPGC COVID19[7] dataset. Section 4 summaries the conclusions.

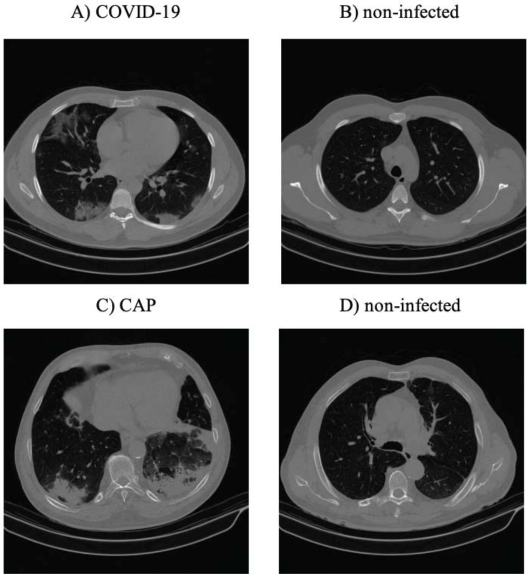

Fig. 1. A, B: Sample slices with and without infection in a COVID-19 case; C, D: Sample slices with and without infection in a non-COVID-19 Pneumonia case.

2. THE ALGORITHM

The classification method proposed in this paper is based on an ensemble learning scheme, which is formed by two essential components, as shown in Fig.2. First, deep neural networks based on semi-supervised learning are constructed for slice-level classification task. Second, according to outputs of slice classifier, another sequence classifier based on the supervised learning is learned to make final decisions.

2.1. Slice classifier

Our model applies a hierarchical classification strategy to perform patient-level diagnosis. Inspired by [8] which combines consistency regularization and pseudo-labeling, a slice-level classifier is built to extract multi-level features. A two-step procedure is adopted:

- Strong and weak data augmentation are applied to unlabeled samples, and the labels of two augmentations are predicted respectively.
- The weak classifier is trained by fixing the label on the weak augmentation, and then used to teach the strong augmentation.

Considering that only a small amount of labeled data exist in the SPGC-COVID19[7] dataset, semi-supervised learning method is proposed to train the slice classifier for the classification of chest CT scan images into COVID, CAP and normal categories. In our implementation, two types of data augmentation named weak augmentation and strong augmentation are used. Specifically, weak augmentation is a standard flip-and-shift augmentation strategy, while strong augmentation uses RandAugment[9] to find an augmentation strategy comprising transformations. During the course of training, the model will try to minimize the deviation between the labels on the two augmentations. Note that the errors are back-propagated only through strong augmentation.

Our model takes a CT scan image as input, and computes a supervised loss ℓ_s and an unsupervised loss ℓ_u respectively. The training loss function $\mathcal{L}_{\mathrm{cls}}$ for classification is thus defined as below:

$$\mathcal{L}_{cls} = \ell_s + \lambda_u \ell_u \tag{1}$$

$$\ell_s = \frac{1}{B} \sum_{b=1}^{B} \mathrm{H}\left(p_b, p_{\mathrm{m}}\left(y \mid \alpha\left(x_b\right)\right)\right) \tag{2}$$

$$\ell_u = \frac{1}{\mu B} \sum_{b=1}^{\mu B} \nVdash \left(\max\left(q_b\right) \geq \tau\right) \mathrm{H}\left(\hat{q}_b, p_{\mathrm{m}}\left(y \mid \mathcal{A}\left(u_b\right)\right)\right) \tag{3}$$

where ℓ_s is the cross-entropy loss, τ is a scalar hyperparameter denoting the threshold above which we retain a pseudo-label. $\hat{q}_b = \arg\max\left(q_b\right)$ would be used as a pseudo-label when $\left(\max\left(q_b\right) \geq \tau\right)$ satisfied. λ_u is a fixed hyperparameter denoting the relative weight of the unlabeled loss.

2.2. Sequence classifier

Since each patient corresponds to a sequence of CT scan images, a supervised sequence classifier is employed for classification of CT image sequences.

In this paper, we introduce AdaBoost[10] for CT image sequence classification. The features extracted from the slice-level classifier are used as the input of the sequence classifier. By minimizing the empirical risk, the weights of the weak classifiers are learned. Finally the linearly combined strong classifier makes the final decisions based on the votes.

2.3. Dataset clustering

In task 2 of SPGC-COVID19[7], the test set has lower radiation doses. Therefore, this task is more challenging than task 1 and requires some additional preprocessing.

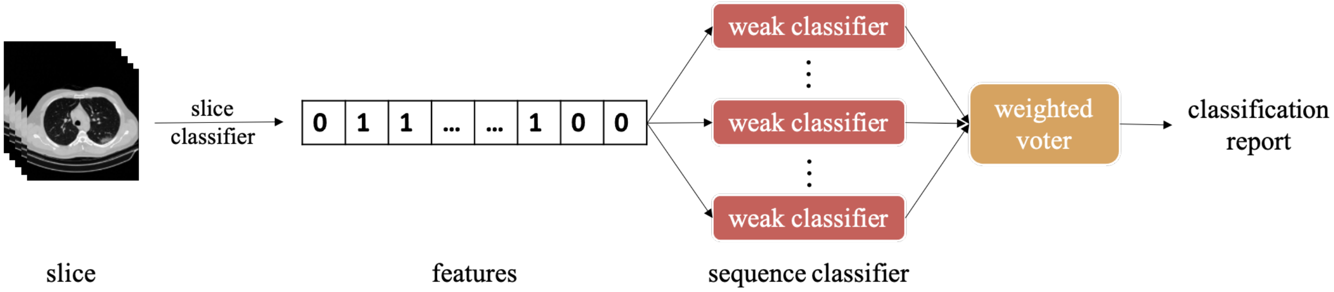

Fig. 2. The architecture of our method.

Inspired by the similarity of grouped images of the same lung lobe, we divided the CT scan images into three sub-data sets to make the neural network focus on lobe-specific disease patterns.

The clustering starts from the encoding step. In this step, the EfficientNet[11] pre-trained on the ImageNet dataset[12] is used to extract digital feature descriptors of each image. The output from the penultimate layer can be used as a feature representation of the image. These features are used to perform subsequent clustering steps, with the purpose of grouping images with similar distance definitions.

Here, we use K-means[13] algorithm to cluster the data. Consider that the right lung segments is biologically divided into three parts: superior lobe, inferior lobe, and middle one, k = 3 was used as the number of clusters. Then, single classifier is employed in the three subsets respectively. The output of each classifier could be different because different training data are used in the single classifier itself. The final prediction is obtained by a majority voting strategy, which votes for each label of all classifiers and uses the label with the most votes.

3. EXPERIMENTS

In this section, we first describe the dataset, then, we introduce our implementation details. Thirdly, ablation experiments are performed to validate the effectiveness of each modules of the proposed method. Finally, we compare the proposed method to the existing models.

3.1. Dataset

We use the SPGC-COVID[7] dataset (Fig. 1) provided by the 2021 IEEE ICASSP Signal Processing Grand Challenge. The dataset contains volumetric chest CT scans of COVID-19, CAP, and normal cases acquired with various imaging settings from different medical centers. The training dataset includes volumetric CT scans with all slices of 116 patient positive for COVID-19, 41 CAP, and 52 normal cases, while the validation dataset includes 55 COVID-19, 19 CAP, and 24 normal cases. Besides the patient-level labels, a subset of 55 COVID-19, and 25 CAP cases in the training and validation dataset have slice-level labels. All CT scans in the trainning and validation dataset are obtained by a SIEMENS, SOMATOM Scope scanner with the normal radiation dose and the slice thickness of 2mm. However, besides this situation, in the test dataset, Low-Dose CT scans (LDCT) and patients with a history of hearth disease or operation in which an abnormal manifestation related to a non-infection disease is demonstrated in some cases.

Table 1. Classification results on the validation set with three classes: COVID-19, CAP and Normal.

model	Accuracy	Sensitivity		
		COVID-19	CAP	Normal
ours*†	82.65%	78.18%	73.68%	100.0%
ours*	86.73%	100.0%	42.11%	91.67%
ours†	82.65%	100.0%	26.32%	70.83%
ours	86.73%	87.27%	68.42%	100.0%

*: w/o FixMatch †: w/o sequence classifier

3.2. Implementation Details

The proposed network and all experiments are implemented through Pytorch[14] on a Tesla T4 GPU. For the labeled dataset, the infectious slices of COVID/CAP CT scan are labeled as COVID/CAP, and the non-infectious slices of COVID/CAP CT scan and all slices of normal CT scan are labeled as normal. For the trainning of sclice-level classifier, the learning rate is initially set to 1e-4 and decreased by the cosine learning decay. The batch size is set to 16. The Stochastic Gradient Descent ($weight\ decay = 0.0001, momentum = 0.9$) is adopted to optimize the models. The models of the 50th epoch are stored for evaluation. We use cross entropy as the loss function($\lambda_u = 1$). For the trainning of sequence classifier, the Adaboost is chosen as the base estimator,

Table 2. Classification results on the validation set with two classes: COVID-19 and Normal.

model	Accuracy	Sensitivity	
		COVID-19	Normal
Single classifier†	92.81%	93.29%	92.41%
Ensemble	93.23%	93.38%	93.03%

†: w/o clustering

Table 3. Classification results on the validation set with sllice-level three-way classification: COVID-19, CAP and Normal.

model	Accuracy	Sensitivity		
		COVID-19	CAP	Normal
ResNet18[16]	86.78%	92.22%	80.05%	89.03%
ResNet34[16]	86.67%	93.85%	82.11%	85.08%
SEResNext50[17]	**91.51%**	92.02%	**88.99%**	**93.94%**
InceptionV4[18]	87.35%	**94.39%**	80.96%	87.82%
EfficientNet-B0	88.31%	92.34%	81.42%	91.95%
EfficientNet-B1	90.23%	93.79%	85.09%	92.31%

and it is performed with the learning rate of 1 and maximum of 50 SAMME.R[15] estimators. The accuracy and sensitivity are used as the evaluation criteria.

3.3. Ablation Experiments

Table 1 shows the effectiveness of the introduced FixMatch and sequence classifier on the patient-level classification. Note that *w/o sequence classifier* means a CT scan is COVID-19 if there exists at least one slice which classified as COVID-19. Compared with the backbone, the FixMatch can increase the sensitivity of CAP by 26.31% for the three-way classification. We attribute this improvement to the introduction of a large number of unlabeled slices as training data. Based on this, the sequence classifier can further improve the performance of the model in both accuracy and sensitivity. This is because it can integrate the spatial information of all the slices of a CT scan and make a comprehensive judgment without being interfered by the misjudgment of some slices.

Table 2 compares the performance of the best single and ensemble classifiers. Though the number of cases for each classifier are reduced, the integrated classifier has better performance and finally achieves higher accuracy. This result is not surprising because the clustering method increases the similarity of the images. Note that the positive samples of CAP are rare (about 1,000 labels), and clustering will exacerbate this problem, thus the single classifier is adopted in task1 and task3.

Table 4. Classification results on the validation set with patient-level three-way classification: COVID-19, CAP and Normal.

model	Accuracy	Sensitivity		
		COVID-19	CAP	Normal
EfficientNet3D-B1	58.16%	58.18%	52.63%	62.50%
AD3D-MIL[6]	63.26%	63.63%	57.89%	66.67%
ours	86.73%	87.27%	68.42%	100.0%

Table 5. Classification results on the test set and validation set.

model	Accuracy	Sensitivity		
		COVID-19	CAP	Normal
validation	86.73%	87.27%	68.42%	100.0%
test	80.00%	88.57%	35.00%	97.14%

3.4. Comparison

Table 3 compares different backbones on the slice-level three-way classification. It is observed that SEResNext50 performs best, but EfficientNet-B1 has better sensitivity on COVID-19, thus we choose it as the backbone of our model. To show the superiority of our model on the patient-level classification task, we compare it with two 3D CNNs as shown in Table 4. Note that the EfficientNet3D-B1 is a 3D version of EfficientNet-B1. These two 3D CNNs take all slices of a CT scan as input and output the classification result end-to-end. Because 3D data has a lot more information than 2D data, the training of 3D CNNs is more complex and challenging especially when the training dataset is small. As a result, these two 3D CNNs perform worse than the proposed method. Finally, results of the proposed method on the test set are shown in Table 5.

4. CONCLUSION

In this paper, a novel ensemble learning framework for distinguishing COVID-19 from CAP and normal cases is proposed. First, FixMatch is introduced for the semi-supervised learning of slice-level classifier. Second, AdaBoost algorithm is used as sequence classifier to make the patient-level judgement. Finally, dataset clustering is performed to achieve better robustness. The experimental results demonstrate that our framework is superior to existing networks in both accuracy and sensitivity.

5. REFERENCES

[1] Xingzhi Xie, Zheng Zhong, Wei Zhao, et al., “Chest ct for typical coronavirus disease 2019 (covid-19) pneumonia: relationship to negative rt-pcr testing,” Radiology, vol. 296, no. 2, pp. E41–E45, 2020.

[2] Tao Ai, Zhenlu Yang, Hongyan Hou, et al., “Correlation of chest ct and rt-pcr testing for coronavirus disease 2019 (covid-19) in china: a report of 1014 cases,” Radiology, vol. 296, no. 2, pp. E32–E40, 2020.

[3] Yicheng Fang, Huangqi Zhang, Jicheng Xie, et al., “Sensitivity of chest ct for covid-19: comparison to rt-pcr,” Radiology, vol. 296, no. 2, pp. E115–E117, 2020.

[4] Lin Li, Lixin Qin, Zeguo Xu, et al., “Artificial intelligence distinguishes covid-19 from community acquired pneumonia on chest ct,” Radiology, 2020.

[5] Shuai Wang, Bo Kang, Jinlu Ma, et al., “A deep learning algorithm using ct images to screen for corona virus disease (covid-19). medrxiv,” Preprint at https://www. medrxiv. org/content/10.1101/2020.02, vol. 14, pp. v5, 2020.

[6] Zhongyi Han, Benzheng Wei, Yanfei Hong, et al., “Accurate screening of covid-19 using attention-based deep 3d multiple instance learning,” IEEE transactions on medical imaging, vol. 39, no. 8, pp. 2584–2594, 2020.

[7] Parnian Afshar, Shahin Heidarian, Nastaran Enshaei, et al., “COVID-CT-MD: COVID-19 Computed Tomography (CT) Scan Dataset Applicable in Machine Learning and Deep Learning,” arXiv preprint arXiv:2009.14623, 2020.

[8] Kihyuk Sohn, David Berthelot, Chun-Liang Li, et al., “Fixmatch: Simplifying semi-supervised learning with consistency and confidence,” arXiv preprint arXiv:2001.07685, 2020.

[9] Ekin D Cubuk, Barret Zoph, Jonathon Shlens, et al., “Randaugment: Practical automated data augmentation with a reduced search space,” in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, 2020, pp. 702–703.

[10] Yoav Freund and Robert E Schapire, “A decision-theoretic generalization of on-line learning and an application to boosting,” Journal of computer and system sciences, vol. 55, no. 1, pp. 119–139, 1997.

[11] Mingxing Tan and Quoc Le, “Efficientnet: Rethinking model scaling for convolutional neural networks,” in International Conference on Machine Learning. PMLR, 2019, pp. 6105–6114.

[12] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei, “Imagenet: A large-scale hierarchical image database,” in 2009 IEEE conference on computer vision and pattern recognition. Ieee, 2009, pp. 248–255.

[13] James MacQueen et al., “Some methods for classification and analysis of multivariate observations,” in Proceedings of the fifth Berkeley symposium on mathematical statistics and probability. Oakland, CA, USA, 1967, vol. 1, pp. 281–297.

[14] Adam Paszke, Sam Gross, Francisco Massa, et al., “Pytorch: An imperative style, high-performance deep learning library,” arXiv preprint arXiv:1912.01703, 2019.

[15] Trevor Hastie, Saharon Rosset, Ji Zhu, and Hui Zou, “Multi-class adaboost,” Statistics and its Interface, vol. 2, no. 3, pp. 349–360, 2009.

[16] Kaiming He, Xiangyu Zhang, Shaoqing Ren, et al., “Deep residual learning for image recognition,” in Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 770–778.

[17] Jie Hu, Li Shen, and Gang Sun, “Squeeze-and-excitation networks,” in Proceedings of the IEEE conference on computer vision and pattern recognition, 2018, pp. 7132–7141.

[18] Christian Szegedy, Sergey Ioffe, Vincent Vanhoucke, et al., “Inception-v4, inception-resnet and the impact of residual connections on learning,” in Proceedings of the AAAI Conference on Artificial Intelligence, 2017, vol. 31.