Mathematical Biosciences and Engineering

http://www.aimspress.com/journal/mbe

MBE, 20(4): 6838–6852.
DOI: 10.3934/mbe.2023294
Received: 29 November 2022
Revised: 16 January 2023
Accepted: 29 January 2023
Published: 06 February 2023

Research article

Data augmentation based semi-supervised method to improve COVID-19 CT classification

Xiangtao Chen[1,*], Yuting Bai[1], Peng Wang[2] and Jiawei Luo[1]

[1] College of Computer Science and Electronic Engineering, Hunan University, Changsha 410082, Hunan, China

[2] College of Computer Science and Engineering, Hunan Institute of Technology, Hengyang 421002, China

* **Correspondence:** Email: xtchen2009@sina.cn.

Abstract: The Coronavirus (COVID-19) outbreak of December 2019 has become a serious threat to people around the world, creating a health crisis that infected millions of lives, as well as destroying the global economy. Early detection and diagnosis are essential to prevent further transmission. The detection of COVID-19 computed tomography images is one of the important approaches to rapid diagnosis. Many different branches of deep learning methods have played an important role in this area, including transfer learning, contrastive learning, ensemble strategy, etc. However, these works require a large number of samples of expensive manual labels, so in order to save costs, scholars adopted semi-supervised learning that applies only a few labels to classify COVID-19 CT images. Nevertheless, the existing semi-supervised methods focus primarily on class imbalance and pseudo-label filtering rather than on pseudo-label generation. Accordingly, in this paper, we organized a semi-supervised classification framework based on data augmentation to classify the CT images of COVID-19. We revised the classic teacher-student framework and introduced the popular data augmentation method Mixup, which widened the distribution of high confidence to improve the accuracy of selected pseudo-labels and ultimately obtain a model with better performance. For the COVID-CT dataset, our method makes precision, F1 score, accuracy and specificity 21.04%, 12.95%, 17.13% and 38.29% higher than average values for other methods respectively, For the SARS-COV-2 dataset, these increases were 8.40%, 7.59%, 9.35% and 12.80% respectively. For the Harvard Dataverse dataset, growth was 17.64%, 18.89%, 19.81% and 20.20% respectively. The codes are available at https://github.com/YutingBai99/COVID-19-SSL.

Keywords: COVID-19; semi-supervised; pseudo-labels; Mixup; teacher-student framework

1. Introduction

Since 2019, COVID-19 has been rampant all over the world, which has greatly damaged people's health and social economy [1, 2]. How to quickly judge whether a patient is suffering from COVID-19 has become an urgent problem, which has attracted extensive attention from scholars in related fields [3]. Previously, deep learning technology has been proven to be helpful to assist doctors in rapid diagnosis [4]. The deep learning method related to COVID-19 is mainly to classify and segment the X-ray or CT images to distinguish patients [5]. This paper mainly discusses the classification of COVID-19 CT images.

At present, some scholars have opened diagnostic datasets of COVID-19, and researchers have proposed classification methods for COVID-19 CT images on this basis [6]. Due to the extensive application of deep learning in the field of medical images and the lack of early data sets of COVID-19 [7, 8], transfer learning, contrastive learning, ensemble strategy, semi-supervised learning and even the combination of the above, have been favored by researchers [9–11]. Additionally, there are some classifications based on human units. Next, we will introduce these methods in more detail.

Transfer Learning is a common machine learning method by reusing the model developed for Task A in the process of developing the model for Task B [12]. In the classification problem of the medical image field, it is often based on the model pretrained on ImageNet [13] dataset for further training, most of the COVID-19 classification methods proposed by researchers adopt the same way. Panwar et al. [14] proposed a new deep transfer learning algorithm based on VGG-19 and performed Grad-CAM analysis on the test results. Similarly, Jaiswal et al. [15] applied DenseNet201 based deep transfer learning to classify the patients as COVID-infected or not. Further, Alshazly et al. [16] discussed several deep network architectures and employed custom-sized input tailored for each deep architecture to achieve the best performance. Even the methods based on other technologies also used the pretrained model [17].

Ensemble strategy is also commonly used in COVID-19 classification problems in order to alleviate the problems of overfitting and low accuracy caused by less data [18]. That can be embodied in integrating multiple datasets or classifiers, which are called data integration [19, 20] or ensemble learning [21] respectively. In terms of data integration, Wang et al. [22] proposed a joint learning framework for COVID-19 diagnosis by learning from heterogeneous datasets, in which they conducted separate feature normalization to tackle the inter-site data discrepancy. For integrated learning, Kundu et al. [23] used three standard CNN models and assigned fuzzy ranks to the constituent classifiers employing a re-parameterized Gompertz function. Further, Kundu et al. [24] designed a Sugeno fuzzy integral ensemble of four pretrained deep learning models for the classification of chest CT-scan images. In addition, Shaik et al. [25] designed a composite neural network architecture and trained it in an end-to-end manner based on eight different pre-trained models. Jangam et al. [26] adopted advanced ensemble technology, namely stack ensemble, to integrate three models, so as to obtain high recall and accuracy. The difference between these technologies lies not only in the number of models but also in the way of integration.

Contrastive learning is a machine learning technique used to learn the general features of unlabeled or less labeled datasets by teaching the model which data points are similar or different [27]. Researchers often combine contrastive learning and supervised learning to classify COVID-19 datasets. Similarity regularization (SR) derived from contrastive learning is proposed to enable CNNs to learn

more parameter-efficient representations by Xu et al. [28]. Ewen et al. [29] organized targeted self-supervision around the downstream task, proving that gradually increasing targeted self-supervision results in a significant increase in performance. He et al. [30] developed Self-Trans approach integrating contrastive self-supervised learning and transfer learning as the baseline of the COVID-CT dataset. Han et al. [31] constructed a new loss function, which is the addition of contrastive learning's loss and traditional cross entropy. Although these methods adopt the idea of contrastive learning, they still use more or fewer labels, so they are not unsupervised learning.

The methods mentioned above are to judge whether the image comes from COVID-19 patients, while some approaches also judge on the basis of patients. Silva et al. [32] considered all images of an individual, they employed a voting scheme to achieve a diagnosis per individual instead of by instance or image. Wu et al. [33] developed a novel COVID-19 diagnosis system to perform explainable Joint Classification and accurate lesion Segmentation (JCS). If the number of infected CT images is above a threshold, the patient is diagnosed as COVID-19 positive.

In a word, these methods all use a lot of labeled data. However, COVID-19 is still rampant, experienced medical personnel is extremely scarce, and the cost of manual labels is expensive, so some scholars used semi-supervised learning to classify COVID-19 CT images [34]. At the early stage of the COVID-19 outbreak, because there were too few positive labels, researchers mainly solved the problem of unbalanced classification [35, 36]. Calderon-Ramirez et al. [37] re-weighted each observation in the loss function, giving a higher weight to the observations corresponding to the under-represented class. Calderon-Ramirez et al. [38] utilized the feature space to construct an OOD scoring approach, they discarded harmful observations that might create wrong low-density regions to increase distribution mismatch robustness in Semi-supervised learning. There are also some scholars who made improvements in pseudo-label screening, improving classification accuracy. Alizadehsani et al. [39] and Calderon-Ramirez et al. [40] introduced uncertainty to determine whether the pseudo-labels are reliable. Nevertheless, these semi-supervised methods are not very concerned about the quality of pseudo-label generation [41].

In order to save human resources and use as few labeled data as possible, we designed a semi-supervised learning framework, which is different from the semi-supervised method dedicated to alleviating class imbalance. Our framework focuses on obtaining better pseudo-labels, that is, generating pseudo-labels with less noise. So we upgraded the training process to make the traditional pseudo-label filtering more effective. Contributions of this paper are summarized as follows:

We organized a semi-supervised classification framework based on data augmentation to classify the CT images of COVID-19.

We discussed and verified the advantageous performance of Mixup [42] on pseudo-label quality.

We have designed comparative experiments on three datasets to prove the effectiveness of our method.

2. Methods

2.1. Notation

Set $D_L = (x^i, y^i)_{i=1}^{N_L}$ as a labeled dataset with N_L samples, where x is the input and $y = [y_0, y_1]$ is the corresponding label. The label can only be 0 or 1, indicating whether the input image is from a patient with COVID-19, which means $y_c \in \{0, 1\}, c = [0, 1]$. Set $D_U = (x^i)_{i=1}^{N_U}$ as an unlabeled dataset with

N_U samples, which does not contain labels corresponding to its input samples, pseudo-labels $\tilde{y}$ will be generated for their samples. Set $D = (x^i, \tilde{y}^i)_{i=1}^{N_L+N_U}$ as merged dataset from D_L and D_U. For labeled samples, $\tilde{y}_i = y_i$. Set D_T as the test dataset. The neural network teacher t_θ and student s_θ are trained on dataset D_L and D respectively.

2.2. *Research framework*

An overview of our approach is shown in Figure 1, which can be divided into 7 steps. We divide the training dataset into labeled dataset D_L and unlabeled dataset D_U, and take the iteration number I together as the input of the method. For labeled data, we use different colors to represent samples belonging to different categories, while the square without colors but only a black outline represents unlabeled data. We use the classic teacher-student [43] based semi-supervised learning framework to construct the entire method, moreover, the teacher and student share a common architecture. After we obtain the partitioned dataset, we can start to perform the steps in Figure 1. Step 1: Mixup on dataset D_L, that is, mixing the samples and labels to get D_{LM}. Step 2: Train teacher t_θ on D_{LM}. Step 3: Use the teacher model to generate pseudo-labels for D_U, the blocks of D_{UPSE} are dyed correspondingly. Step 4: Fuse D_{UPSE} and D_L to get D. Shuffle is implied here. Step 5: Mixup on D to get D_M. Step 6: Train student s_θ on D_M. Step 7: Judge whether the number of iterations i that has been run is less than or equal to the specified total number of iterations. If yes, assign the parameters of the student to the teacher, return to step 1 and start the next iteration. Otherwise, output the student as the final result model. Next, we will detail the vital parts of the framework.

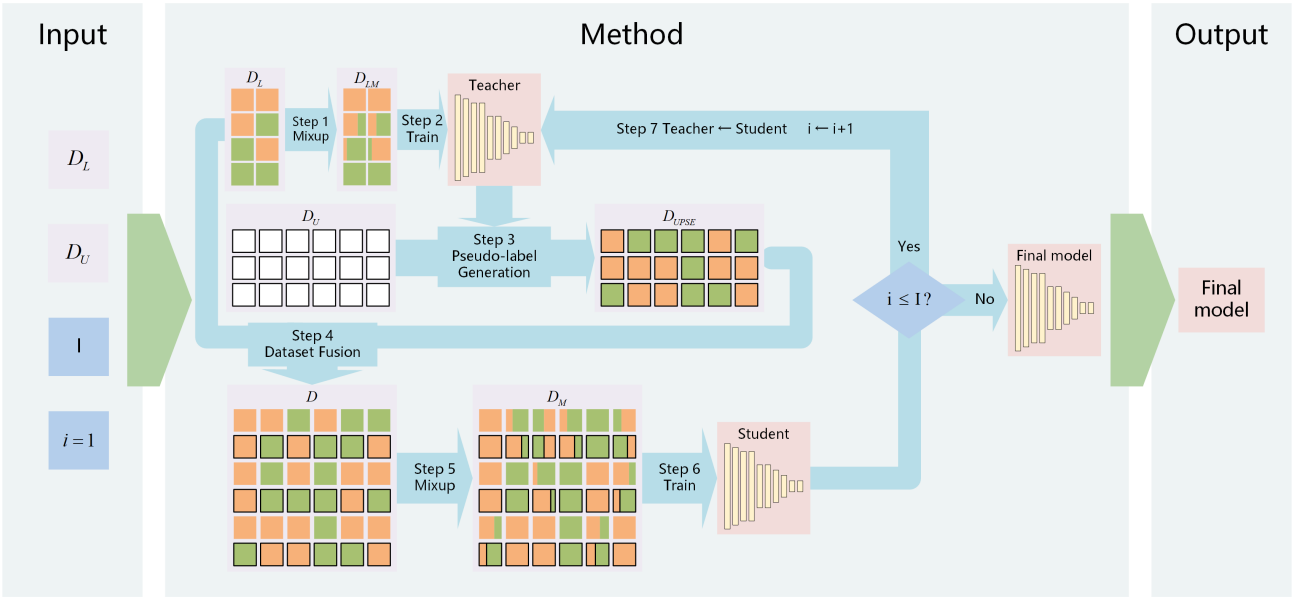


Figure 1. Overview.

2.3. *Pseudo-label generation*

There are several approaches to creating the pseudo-labels, we adopt a simple approach where hard pseudo-labels are obtained directly from network predictions [44]. The generation process is as follows: firstly, get the probability outputs p_i of a trained network on the input image i; then, select

the maximum probability in p_i as the confidence that sample i belongs to the corresponding class, like Eqs 2.1 and 2.2; finally, filter the samples whose confidence is greater than the threshold, and set the class with high confidence as their pseudo-labels, the process can be defined as Eq 2.3.

$$\text{class}_\text{i} = \begin{cases} 0 & p_{i0} \geq p_{i1} \\ 1 & p_{i0} < p_{i1} \end{cases} \tag{2.1}$$

$$\text{confidence}_\text{i} = \begin{cases} p_{i0} & p_{i0} \geq p_{i1} \\ p_{i0} & p_{i0} < p_{i1} \end{cases} \tag{2.2}$$

$$\tilde{y} = \begin{cases} \text{class} & \text{confidence} \geq \text{th} \\ \text{Null} & \text{confidence} < \text{th} \end{cases} \tag{2.3}$$

where $\text{th} \in (0, 1)$ is a threshold used to produce hard labels. This threshold is a hyperparameter, and its influence experiment is proposed in Section 3.5.

2.4. Mixup

In semi-supervised learning, the accuracy of pseudo-labels is crucial, and the generation accuracy of pseudo-labels actually reflects the generalization ability of the model. The stronger the generalization ability of the model is, the wider the range of knowledge owned by the model can be used, and the better the accuracy of pseudo-labels will be. We use Mixup [42] to enhance the generalization performance of the teacher model, so as to improve the accuracy of pseudo-labels. The usage of Mixup is shown in Eq 2.4.

$$\begin{cases} x' = \lambda x_i + (1 - \lambda) x_j \\ y' = \lambda y_i + (1 - \lambda) y_j \end{cases} \tag{2.4}$$

In addition, Mixup can also expand the range of high confidence, which is conducive to the threshold screening of pseudo-labels. This is specifically discussed in Section 4.2.

2.5. Pseudocode

The pseudocode of this framework is as Algorithm 1.

After several iterations, we get the final model. The number of iterations is related to the scale of the dataset. The larger the dataset, the more iterations are required to obtain the optimal model.

3. Results

3.1. Datasets

To evaluate the performance of the proposed framework, we have used three publicly available datasets with different sizes, namely the COVID-CT dataset by Zhao et al. [45], the SARS-COV-2 dataset by Soares et al. [46] and the Harvard Dataverse chest CT dataset [47]. Their distribution is shown in Table 1. For all datasets, we integrate the original training set, verification set and test set, and then re-divide them into labeled training set, unlabeled training set and test set. Among them,

Algorithm 1 Research framework

Input: Labeled datasets D_L; Unlabeled datasets D_U; Test datasets D_T; Iterations I;
Output: Final model f_θ;

Initialize teacher model t_θ
$D_{LM} \leftarrow Mixup(D_L)$
t_θ are trained on D_{LM}
$i = 1$
repeat
    Use t_θ to generate confidence on D_U
    Pseudo-label Generation
    $D \leftarrow D_L + D_U$
    Initialize student model s_θ
    $D_M \leftarrow Mixup(D)$
    s_θ are trained on D_M
    $M_T \leftarrow M_S$
    $i \leftarrow i + 1$
until i > I
return $f_\theta \leftarrow s_\theta$

20% of the total data is divided into test sets according to the distribution of the dataset, after that, the remaining parts are divided into labeled training datasets and unlabeled training datasets on a scale of 1:4.

Table 1. Class distribution, total number and division of COVID-CT, SARS-COV-2 and Harvard Dataverse chest CT dataset.

	COVID	NonCOVID	Total Number	Labeled	Unlableled	Test
COVID-CT	349	397	746	118	478	150
SARS-COV-2	1252	1229	2481	396	1588	497
Harvard Dataverse	2167	2004	4171	666	2670	835

Since many images in the three datasets were from the same patients, the datasets were not randomly divided after shuffling but sorted according to the patients of CT images to avoid data leakage.

3.2. Experimental settings

For all experiments, we start training with a model pre-trained on ImageNet, we set the batch size to 10, the learning rate to 0.0001, the confidence threshold to 0.995, and 100 epochs are trained for each iteration. Without explicitly declaring the model, we use the EfficientNet-b0 [48] as the architecture in the framework. To avoid the shock caused by the randomness of model training, all experiments were carried out three times, and the results were averaged. About Iterations, for COVID-CT and SARS-COV-2, we get the best model in three iterations, while for the Harvard Dataverse chest CT dataset, we get the best model in the fourth iteration.

3.3. Main results

Table 2. Performance analysis of the proposed semi-supervised framework, Kundu et al. [23], Self-Trans [30], Ewen et al. [29], DNN-GFE [49], CNN-SVM + Sobel [50] and EfficientNet-b0 on the COVID-CT dataset.

Method	Precision	F1	Accuracy	Specificity
Ours	**0.8459**	**0.7947**	**0.7956**	**0.8429**
Kundu et al.	0.5344	0.6635	0.5267	0.1286
Self-Trans	0.7391	0.7907	0.7600	0.5143
Ewen et al.	0.6591	0.4677	0.5600	0.7857
DNN-GFE	0.5780	0.6667	0.5880	0.3429
CNN-SVM + Sobel	0.6667	0.7374	0.6867	0.5286
EfficientNet-b0	0.7888	0.7879	0.7733	0.7572

Table 3. Performance analysis of the proposed semi-supervised framework, Kundu et al., Self-Trans, Ewen et al., DNN-GFE, CNN-SVM + Sobel and EfficientNet-b0 on the SARS-COV-2 dataset.

Method	Precision	F1	Accuracy	Specificity
Ours	**0.7534**	**0.8353**	**0.8162**	**0.6959**
Kundu et al.	0.6562	0.7726	0.7264	0.5179
Self-Trans	0.7160	0.8140	0.7867	0.6335
Ewen et al.	0.6442	0.7747	0.7203	0.4741
DNN-GFE	0.5958	0.6748	0.6258	0.5378
CNN-SVM + Sobel	0.7348	0.7608	0.7545	0.6764
EfficientNet-b0	0.7246	0.8345	0.8068	0.6335

Table 4. Performance analysis of the proposed semi-supervised framework, Kundu et al., Self-Trans , Ewen et al., DNN-GFE, CNN-SVM + Sobel and EfficientNet-b0 on the Harvard Dataverse dataset.

Method	Precision	F1	Accuracy	Specificity
Ours	**0.8704**	**0.9255**	**0.9234**	0.8633
Kundu et al.	0.6573	0.6795	0.6814	0.6613
Self-Trans	0.7184	0.7851	0.7725	0.6866
Ewen et al.	0.8427	0.7926	0.8120	**0.8710**
DNN-GFE	0.5889	0.6667	0.6311	0.5046
CNN-SVM + Sobel	0.6629	0.7591	0.7293	0.5829
EfficientNet-b0	0.8751	0.9322	0.9301	0.8679

We take 20% of the three datasets mentioned above as the labeled train dataset and compare our framework with three methods for COVID-19 CT image classification using the same parameters and models. Since the method proposed by Kundu et al. [23] integrates multiple models, we also used VGG-19 [51] and Wide-ResNet50-2 [52] based on EfficientNet-b0 to reproduce this approach. Besides that, the method proposed by Ewen et al. [29] has its regulations on the number of epochs, so instead of using the unified 100 epochs, we used a total of 150 epochs in five stages of training. Moreover, methods DNN-GFE [49] and CNN-SVM + Sobel [50] have constructed a new network structure to solve this problem, and we have reproduced these two methods according to their papers. At the same time, we also used all train datasets to calculate the results of full supervision on EfficientNet-b0. There are four evaluation indicators including precision, F1 score, accuracy and specificity used to evaluate the above methods. As shown in Tables 2–4, the best result under this indicator is shown in bold. Our method achieves the best performance in all indicators on datasets COVID-CT and SARS-COV-2, and outperforms the full-supervised method. For the Harvard Dataverse dataset, only the specificity is slightly lower than the optimal indicator, and all indicators are slightly lower than the full supervision method. This is because we discard samples that may lead to training errors, and high-quality samples contribute to the better performance of neural networks.

3.4. Ablation experiments

We present an ablation study to measure the contribution of the method's different components. We run experiments on SARS-COV-2 with 396 labeled samples, their results are reported in Table 5. In the ablation study, only labeled data were used in the full-supervised experiment. Teacher-student framework comprehensively improves upon the full-supervised baseline due to pseudo-labels expanding the available training samples. The addition of Mixup improves the accuracy of the pseudo-label, thereby improving the performance of the model. Except for the 0.01% decline in F1 score, other indicators have been greatly improved. Experiments have proved that the main components of our method are effective.

Table 5. Ablation Study on SARS-COV-2 dataset.

Method	Precision	F1	Accuracy	Specificity
Full-Supervised	0.7141	0.8192	0.7894	0.6202
T-S Framework	0.7281	**0.8354**	0.8089	0.6414
Ours	**0.7534**	0.8353	**0.8162**	**0.6959**

3.5. Confidence threshold experiments

We organized experiments on Harvard Dataverse with different confidence thresholds, as seen in Table 6. Except that the first row represents the iteration number, each row of the table represents an experiment. The first column records the confidence threshold used in the experiment, and the remaining columns represent the accuracy of the iteration. We can observe that with the increase of the confidence threshold, the number of iterations for the optimal result increases. This is because the lower confidence threshold passes more pseudo-labels at the beginning, while the higher confidence threshold can make full use of the unlabeled dataset after several iterations.

Table 6. Accuracy of different confidence thresholds on Harvard Dataverse dataset.

Th	1	2	3	4	5
0.980	0.7557	**0.8251**	0.7569	0.7844	0.7569
0.985	0.7884	0.8595	**0.8663**	0.8638	0.8478
0.990	0.7341	0.7928	0.7509	**0.7940**	0.7138
0.995	0.8739	0.8655	0.8890	**0.9234**	0.9018

3.6. Results based on different models

We also put other models into the framework proposed in this paper, proving that the framework is also useful for other architectures, the results are shown in Table 7. Take 20% of COVID-CT's training set as labeled dataset, set the confidence threshold as 0.985, conduct full- supervised and our method based on DenseNet-169 [20] and ResNet-50 [53], take the result of the fifth iteration, our method is better in all indicators, with average improvements of 8.67% and 2.79% respectively.

Table 7. Results based on DenseNet-169 and ResNet-50.

Model	Method	Precision	F1	Accuracy	Specificity
DenseNet-169	Full-Supervised	0.7234	0.7816	0.7467	0.6286
	Ours	**0.8148**	**0.8199**	**0.8067**	**0.7857**
ResNet-50	Full-Supervised	0.7727	0.6986	0.7067	0.7857
	Ours	**0.7971**	**0.7383**	**0.7400**	**0.8000**

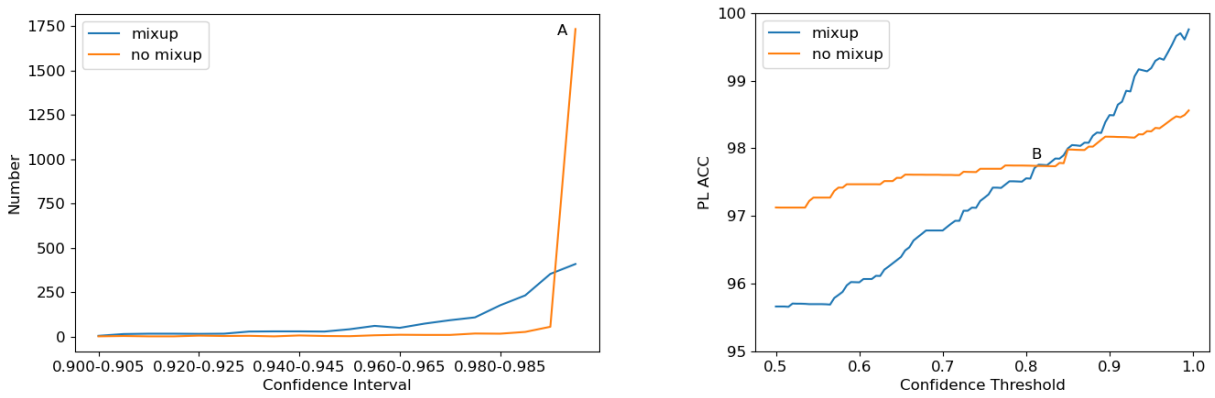
4. Discussion

4.1. Influence of confidence threshold

The confidence threshold affects both the precision and the number of pseudo-labels. When the confidence threshold is higher, the precision of pseudo-labels is higher, but the number of pseudo-labels is less; The lower the confidence threshold, the lower the precision of the pseudo-labels, but the more the number of pseudo-labels. Meanwhile, both the precision and the number of pseudo-labels will affect the performance of the next iteration training model. Therefore, the setting of the confidence threshold needs to make a balance between them. The framework proposed in this paper uses a relatively strict confidence threshold. Although the quantity of pseudo-labels is small at the beginning, it will gradually become large with the increase of iterations, which ensures the high quality of pseudo-labels. In addition, a higher confidence threshold will lead to an increase in the number of iterations.

4.2. The role of Mixup

Take the first generation of pseudo-label data on the EfficientNet-b0 by SARS-COV-2 as an example to illustrate the effect of Mixup on high confidence, as shown in Figure 2(a),(b). Figure 2(a) shows the counting distribution of confidence intervals greater than 0.9. The x-axis represents the confidence interval with a step size of 0.005, and the y-axis represents the number of confidence within this interval.



(a) (b)

Figure 2. (a): Confidence interval counting chart. (b): The Relationship between confidence threshold and pseudo-label accuracy.

Since more than 90% of confidence is within the range of 0.9 to 1, we only make statistics on this interval. Figure 2(b) records the accuracy of pseudo-labels filtered under different confidence thresholds *th*. When the confidence is less than 0.5, the filter is almost meaningless. We only focus on the case where *th* is greater than 0.5. The two lines in Figure 2(a),(b) represent two situations where Mixup is used or not. In Figure 2(a), when Mixup is not used, point A in the figure shows that the confidence is highly concentrated in the range of 0.995 to 1; otherwise, the slope of the broken line is significantly slower, at this time, using threshold filtering can achieve better results. In Figure 2(b), although the accuracy of pseudo-labels directly obtained without filtering without using Mixup is higher than that after using Mixup, after the *th* rise exceeds the intersection point B of two lines, the line using Mixup achieves better filter performance.

5. Conclusions

In this work, we propose a semi-supervised classification framework based on Mixup to classify the CT images of COVID-19. It is easy to implement and has practical significance. Because the noise in the pseudo-label will affect the subsequent training, we introduced Mixup to reduce the prediction mistake during the pseudo-label threshold screening based on the classic teacher-student semi-supervised learning framework. To prove the effectiveness of our method, we organized comparative experiments on three datasets and adopted four evaluation indicators, namely, precision, accuracy, F1 score, and specificity. By averaging the results of the four indicators of the three comparison methods, our method is 22.35%, 9.53% and 19.14% higher than the average on the three datasets respectively. In addition, based on DenseNet-169 and ResNet-50, our method also achieved 8.67% and 2.79% average improvement respectively.

Compared with other methods for CT image classification of COVID-19, our method requires significantly less labeled data, but the performance is not weaker than the method using all labeled data training. Moreover, our method can be effective for different datasets based on different neural net-

works and is easy to popularize.

However, our method also has defects. Under different conditions, the optimal confidence threshold may have deviations. We will study this problem in the follow-up work.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China(Grant No. 62006074, No. 61873089); and in part by the Postdoctoral Science Foundation of China (Grant No. 2020M672487); and in part by Hunan Provincial Natural Science Foundation of China (No. 2021JJ30134, No. 2021JJ40120).

Conflict of interest

The authors declare there is no conflict of interest.

References

1. D. Zavras, Healthcare access as an important element for the EU's socioeconomic development: Greece's residents' opinions during the COVID-19 pandemic, *Natl. Account. Rev.*, **4** (2022), 362–377. https://doi.org/10.3934/NAR.2022020
2. D. Panarello, G. Tassinari, The consequences of COVID-19 on older adults: Evidence from the share corona survey, *Natl. Account. Rev.*, **4** (2022), 56–73. https://doi.org/10.3934/NAR.2022004
3. M. Islam, F. Karray, R. Alhajj, J. Zeng, A review on deep learning techniques for the diagnosis of novel coronavirus (COVID-19), *IEEE Access*, **9** (2021), 30551–30572. https://doi.org/10.1109/ACCESS.2021.3058537
4. J. Xu, J. Xu, Y. Meng, C. Lu, L. Cai, X. Zeng, et al., Graph embedding and gaussian mixture variational autoencoder network for end-to-end analysis of single-cell rna sequencing data, *Cell Rep. Methods*, **2023** (2023), 100382. https://doi.org/10.1016/j.crmeth.2022.100382
5. A. Shoeibi, M. Khodatars, R. Alizadehsani, N. Ghassemi, M. Jafari, P. Moridian, et al., Automated detection and forecasting of COVID-19 using deep learning techniques: A review, preprint, arXiv:2007.10785.
6. T. Ai, Z. Yang, H. Hou, C. Zhan, C. Chen, W. Lv, et al., orrelation of chest CT and RT-PCR testing in coronavirus disease 2019 (COVID-19) in china: a report of 1014 cases, *Radiology*, **296** (2020), E32–E40. https://doi.org/10.1148/radiol.2020200642
7. N. Ayoobi, D. Sharifrazi, R. Alizadehsani, A. Shoeibi, J. M. Gorriz, H. Moosaei, et al., Time series forecasting of new cases and new deaths rate for COVID-19 using deep learning methods, *Results Phys.*, **27** (2021), 104495. https://doi.org/10.1016/j.rinp.2021.104495
8. F. Khozeimeh, D. Sharifrazi, N. H. Izadi, J. H. Joloudari, A. Shoeibi, R. Alizadehsani, et al., Combining a convolutional neural network with autoencoders to predict the survival chance of COVID-19 patients, *Sci. Rep.*, **11** (2021), 1–18. https://doi.org/10.1038/s41598-021-93543-8

9. A. Khan, S. Khan, M. Saif, A. Batool, A. Sohail, M. Khan, A survey of deep learning techniques for the analysis of COVID-19 and their usability for detecting omicron, preprint, arXiv:2202.06372.
10. A. Parvaiz, M. Khalid, R. Zafar, H. Ameer, M. Ali, M. Fraz, Vision transformers in medical computer vision–a contemplative retrospection, preprint, arXiv:2203.15269.
11. X. Yang, X. He, Y. Liang, Y. Yang, S. Zhang, P. Xie, Transfer learning or self-supervised learning? a tale of two pretraining paradigms, preprint, arXiv:2007.04234.
12. S. Pan, Q. Yang, A survey on transfer learning, *IEEE Trans. Knowl. Data Eng.*, **22** (2009), 1345–1359. https://doi.org/10.1109/TKDE.2009.191
13. J. Deng, W. Dong, R. Socher, L. J. Li, K. Li, F. Li, ImageNet: A large-scale hierarchical image database, in *2009 IEEE Conference on Computer Vision and Pattern Recognition*, (2009), 248–255. https://doi.org/10.1109/CVPR.2009.5206848
14. H. Panwar, P. Gupta, M. Siddiqui, R. Morales-Menendez, P. Bhardwaj, V. Singh, A deep learning and grad-cam based color visualization approach for fast detection of COVID-19 cases using chest X-ray and CT-scan images, *Chaos, Solitons Fractals*, **140** (2020), 110190. https://doi.org/10.1016/j.chaos.2020.110190
15. A. Jaiswal, N. Gianchandani, D. Singh, V. Kumar, M. Kaur, Classification of the COVID-19 infected patients using densenet201 based deep transfer learning, *J. Biomol. Struct. Dyn.*, **39** (2021), 5682–5689. https://doi.org/10.1080/07391102.2020.1788642
16. H. Alshazly, C. Linse, E. Barth, T. Martinetz, Explainable COVID-19 detection using chest CT scans and deep learning, *Sensors*, **21** (2021), 455. https://doi.org/10.3390/s21020455
17. T. Pham, Classification of COVID-19 chest X-rays with deep learning: New models or fine tuning, *Health Inf. Sci. Syst.*, **9** (2021), 1–11. https://doi.org/10.1007/s13755-020-00135-3
18. Y. Cao, T. Geddes, J. Yang, P. Yang, Ensemble deep learning in bioinformatics, *Nat. Mach. Intell.*, **2** (2020), 500–508. https://doi.org/10.1038/s42256-020-0217-y
19. M. Lenzerini, Data integration: A theoretical perspective, in *Proceedings of the the 21st ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems (PODS02)*, (2002), 233–246. https://doi.org/10.1145/543613.543644
20. Z. Wang, Q. Liu, Q. Dou, Contrastive cross-site learning with redesigned net for COVID-19 CT classification, *IEEE J. Biomed. Health. Inf.*, **24** (2020), 2806–2813. https://doi.org/10.1109/JBHI.2020.3023246
21. O. Sagi, L. Rokach, Ensemble learning: A survey, *Wiley Interdiscip. Rev.: Data Min. Knowl. Discovery*, **8** (2018), e1249. https://doi.org/10.1002/widm.1249
22. Z. Wang, Q. Liu, Q. Dou, Contrastive cross-site learning with redesigned net for COVID-19 CT classification, *IEEE J. Biomed. Health. Inf.*, **24** (2020), 2806–2813. https://doi.org/10.1109/JBHI.2020.3023246
23. R. Kundu, H. Basak, P. Singh, A. Ahmadian, M. Ferrara, R. Sarkar, Fuzzy rank-based fusion of cnn models using gompertz function for screening COVID-19 CT-scans, *Sci. Rep.*, **11** (2021), 1–12. https://doi.org/10.1038/s41598-021-93658-y

24. R. Kundu, P. Singh, S. Mirjalili, R. Sarkar, COVID-19 detection from lung ct-scans using a fuzzy integral-based cnn ensemble, *Comput. Biol. Med.*, **138** (2021), 104895. https://doi.org/10.1016/j.compbiomed.2021.104895
25. N. Shaik, T. Cherukuri, Transfer learning based novel ensemble classifier for COVID-19 detection from chest CT-scans, *Comput. Biol. Med.*, **141** (2022), 105127. https://doi.org/10.1016/j.compbiomed.2021.105127
26. E. Jangam, C. S. Annavarapu, A stacked ensemble for the detection of COVID-19 with high recall and accuracy, *Comput. Biol. Med.*, **135** (2021), 104608. https://doi.org/10.1016/j.compbiomed.2021.104608
27. A. Jaiswal, A. R. Babu, M. Z. Zadeh, D. Banerjee, F. Makedon, A survey on contrastive self-supervised learning, *Technologies*, **9** (2020), 2. https://doi.org/10.3390/technologies9010002
28. Y. Xu, H. Lam, G. Jia, J. Jiang, J. Liao, X. Bao, Improving COVID-19 CT classification of CNNS by learning parameter-efficient representation, preprint, arXiv:2208.04718.
29. N. Ewen, 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)*, (2021), 1481–1485. https://doi.org/10.1109/ISBI48211.2021.9434047
30. X. He, X. Yang, S. Zhang, J. Zhao, Y. Zhang, E. Xing, et al., Sample-efficient deep learning for COVID-19 diagnosis based on CT scans, medRxiv 2020.04.13.20063941, 2020. https://doi.org/10.1101/2020.04.13.20063941
31. C. Han, M. Kim, J. Kwak, Semi-supervised learning for an improved diagnosis of COVID-19 in CT images, *PLoS One*, **16** (2021), e0249450. https://doi.org/10.1371/journal.pone.0249450
32. P. Silva, E. Luz, G. Silva, G. Moreira, R. Silva, D. Lucio, et al., COVID-19 detection in CT images with deep learning: A voting-based scheme and cross-datasets analysis, *Inf. Med. Unlocked*, **20** (2020), 100427. https://doi.org/10.1016/j.imu.2020.100427
33. Y. Wu, S. Gao, J. Mei, J. Xu, D. Fan, R. Zhang, et al., JCS: An explainable COVID-19 diagnosis system by joint classification and segmentation, *IEEE Trans. Image Process.*, **30** (2021), 3113–3126. https://doi.org/10.1109/TIP.2021.3058783
34. G. Kostopoulos, S. Karlos, S. Kotsiantis, O. Ragos, Semi-supervised regression: A recent review, *J. Intell. Fuzzy Syst.*, **35** (2018), 1483–1500. https://doi.org/10.3233/JIFS-169689
35. J. Zhou, B. Jing, Z. Wang, H. Xin, H. Tong, Soda: Detecting COVID-19 in chest X-rays with semi-supervised open set domain adaptation, *IEEE/ACM Trans. Comput. Biol. Bioinf.*, **2021** (2021). https://doi.org/10.1109/TCBB.2021.3066331
36. A. More, Survey of resampling techniques for improving classification performance in unbalanced datasets, preprint, arXiv:1608.06048.
37. S. Calderon-Ramirez, S. Yang, A. Moemeni, D. Elizondo, S. Colreavy-Donnelly, L. Chavarría-Estrada, et al., Correcting data imbalance for semi-supervised COVID-19 detection using X-ray chest images, *Appl. Soft Comput.*, **111** (2021), 107692. https://doi.org/10.1016/j.asoc.2021.107692

38. S. Calderon-Ramirez, S. Yang, D. Elizondo, A. Moemeni, Dealing with distribution mismatch in semi-supervised deep learning for COVID-19 detection using chest X-ray images: A novel approach using feature densities, *Appl. Soft Comput.*, **123** (2022), 108983. https://doi.org/10.1016/j.asoc.2022.108983
39. R. Alizadehsani, D. Sharifrazi, N. Izadi, J. Joloudari, A. Shoeibi, J. Gorriz, et al., Uncertainty-aware semi-supervised method using large unlabeled and limited labeled COVID-19 data, *ACM Trans. Multimedia Comput. Commun. Appl.*, **17** (2021), 1–24. https://doi.org/10.1145/3462635
40. S. Calderon-Ramirez, S. Yang, A. Moemeni, S. Colreavy-Donnelly, D. Elizondo, L. Oala, et al., Improving uncertainty estimation with semi-supervised deep learning for COVID-19 detection using chest X-ray images, *IEEE Access*, **9** (2021), 85442–85454. https://doi.org/10.1109/ACCESS.2021.3085418
41. H. Asgharnezhad, A. Shamsi, R. Alizadehsani, A. Khosravi, S. Nahavandi, Z. A. Sani, et al., Objective evaluation of deep uncertainty predictions for COVID-19 detection, *Sci. Rep.*, **12** (2022), 1–11. https://doi.org/10.1038/s41598-022-05052-x
42. H. Zhang, M. Cisse, Y. Dauphin, D. Lopez-Paz, mixup: Beyond empirical risk minimization, preprint, arXiv:1710.09412.
43. Q. Xie, M. Luong, E. Hovy, Q. Le, Self-training with noisy student improves imagenet classification, in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, (2020), 10687–10698. https://doi.org/10.1109/CVPR42600.2020.01070
44. M. Rizve, K. Duarte, Y. Rawat, M. Shah, In defense of pseudo-labeling: An uncertainty-aware pseudo-label selection framework for semi-supervised learning, preprint, arXiv:2101.06329.
45. J. Zhao, Y. Zhang, X. He, P. Xie, COVID-CT-dataset: a CT scan dataset about COVID-19, preprint, arXiv:2003.13865.
46. E. Soares, P. Angelov, S. Biaso, M. Froes, D. Abe, SARS-COV-2 CT-scan dataset: A large dataset of real patients CT scans for SARS-COV-2 identification, medrxiv, 2020.
47. E. Soares, P. Angelov, A large dataset of real patients CT scans for COVID-19 identification, *Harvard Dataverse*, **1** (2020). https://doi.org/10.7910/DVN/SZDUQX
48. M. Tan, Q. Le, Efficientnet: Rethinking model scaling for convolutional neural networks, in *Proceedings of the 36th International Conference on Machine Learning*, **97** (2019), 6105–6114. https://doi.org/10.48550/arXiv.1905.11946
49. D. Sharifrazi, R. Alizadehsani, M. Roshanzamir, J. H. Joloudari, A. Shoeibi, M. Jafari, et al., Fusion of convolution neural network, support vector machine and sobel filter for accurate detection of COVID-19 patients using X-ray images, *Biomed. Signal Process. Control*, **68** (2021), 102622. https://doi.org/10.1016/j.bspc.2021.102622
50. J. H. Joloudari, F. Azizi, I. Nodehi, M. A. Nematollahi, F. Kamrannejhad, A. Mosavi, et al., DNN-GFE: A deep neural network model combined with global feature extractor for COVID-19 diagnosis based on CT scan images, *Tech. Rep.*, **2021** (2021).
51. K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, preprint, arXiv:1409.1556.
52. S. Zagoruyko, N. Komodakis, Wide residual networks, preprint, arXiv:1605.07146.

53. K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, (2016), 770–778.

AIMS Press