Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Computers in Biology and Medicine 146 (2022) 105531

Contents lists available at ScienceDirect

Computers in Biology and Medicine

journal homepage: www.elsevier.com/locate/compbiomed

TSRNet: Diagnosis of COVID-19 based on self-supervised learning and hybrid ensemble model

Junding Sun [a,*], Pengpeng Pi [a], Chaosheng Tang [a], Shui-Hua Wang [a,b], Yu-Dong Zhang [a,b,**]

[a] School of Computer Science and Technology, Henan Polytechnic University, Jiaozuo, Henan, 454000, PR China
[b] School of Computing and Mathematical Sciences, University of Leicester, Leicester, LE1 7RH, UK

ARTICLE INFO

Keywords:
Machine learning
COVID-19
CT
Ensemble
Deep learning
Transfer learning
Self-supervised learning
Attention

ABSTRACT

Background: As of Feb 27, 2022, coronavirus (COVID-19) has caused 434,888,591 infections and 5,958,849 deaths worldwide, dealing a severe blow to the economies and cultures of most countries around the world. As the virus has mutated, its infectious capacity has further increased. Effective diagnosis of suspected cases is an important tool to stop the spread of the pandemic. Therefore, we intended to develop a computer-aided diagnosis system for the diagnosis of suspected cases.

Methods: To address the shortcomings of commonly used pre-training methods and exploit the information in unlabeled images, we proposed a new pre-training method based on transfer learning with self-supervised learning (TS). After that, a new convolutional neural network based on attention mechanism and deep residual network (RANet) was proposed to extract features. Based on this, a hybrid ensemble model (TSRNet) was proposed for classifying lung CT images of suspected patients as COVID-19 and normal.

Results: Compared with the existing five models in terms of accuracy (DarkCOVIDNet: 98.08%; Deep-COVID: 97.58%; NAGNN: 97.86%; COVID-ResNet: 97.78%; Patch-based CNN: 88.90%), TSRNet has the highest accuracy of 99.80%. In addition, the recall, f1-score, and AUC of the model reached 99.59%, 99.78%, and 1, respectively.

Conclusion: TSRNet can effectively diagnose suspected COVID-19 cases with the help of the information in unlabeled and labeled images, thus helping physicians to adopt early treatment plans for confirmed cases.

1. Introduction

1.1. Background

Coronavirus disease (COVID-19) is caused by a virus called SARS-CoV-2, which spreads widely and rapidly. People infected with the SARS-CoV-2 virus usually have symptoms such as cough, fever, and loss of smell or taste. Some older infected people develop life-threatening conditions [1]. In December 2019, the world's first new case of coronavirus pneumonia infection was identified. A few weeks later, the pandemic swept through most countries and regions of the world. As of Dec 15, 2021, there were 22,061,730 confirmed cases of COVID-19 and 5,334,236 deaths worldwide, according to the latest outbreak data released by the World Health Organization (WHO) [2]. This pandemic has devastated economies and cultures worldwide and brought global health care systems to the brink of collapse. Although some economically developed countries and regions have developed vaccines and completed vaccination at this stage, there are still two major problems: First, the virus is an mRNA virus, and it is very prone to mutation, such as the Delta variant discovered in October 2020 [3]. This variant is more infectious, with higher viral loads within patients after infection, and there have been cases of vaccination but still infection. Secondly, most countries and regions do not have enough financial resources to develop and purchase vaccines. Therefore, efficient diagnosis and proper isolation of suspected cases are still the main pandemic prevention measures at this stage. Real-time reverse transcription-polymerase chain reaction (RT-PCR) is considered the "gold standard" for COVID-19 diagnosis [4]. This test converts the RNA of the virus into DNA by reverse transcription. It then uses PCR to amplify the DNA for appropriate analysis to detect the infected subject. However, RT-PCR tests suffer from long

* Corresponding author.
** Corresponding author. School of Computer Science and Technology, Henan Polytechnic University, Jiaozuo, Henan, 454000, PR China.
E-mail addresses: sunjd@hpu.edu.cn (J. Sun), pipengpeng@home.hpu.edu.cn (P. Pi), tcs@hpu.edu.cn (C. Tang), shuihuawang@ieee.org (S.-H. Wang), yudongzhang@ieee.org (Y.-D. Zhang).

https://doi.org/10.1016/j.compbiomed.2022.105531
Received 30 December 2021; Received in revised form 12 March 2022; Accepted 13 April 2022
Available online 16 April 2022

acceptance times, high false-negative rates, and lack of test kits.

The researchers found that CT scan images or X-ray images of patients infected with COVID-19 had typical features such as cloudiness, ground-glass nodules or faint black dots. This opens up the possibility for physicians to diagnose subjects from their CT scan images or X-ray images. In addition, a related study showed that the sensitivity of RT-PCR detection was 71%, and that of CT scan image detection was 98% at the time of the patient's initial visit [5]. At the same time, it is faster and easier for physicians to diagnose by CT scan images or X-ray images of the subject than by RT-PCR [6]. Therefore, CT scans or X-ray images can be used as a complementary tool to RT-PCR to diagnose COVID-19.

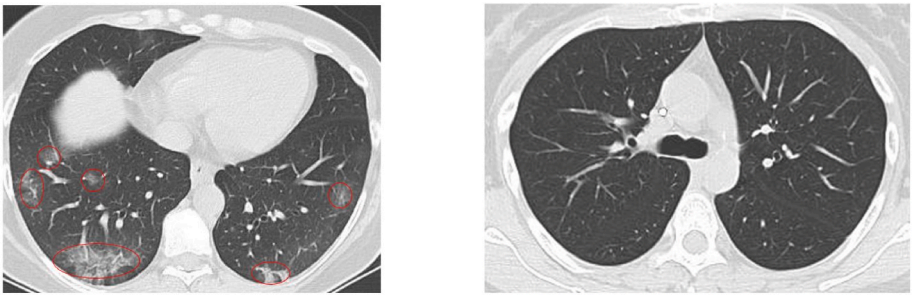
1.2. Related work

However, traditional manual diagnostic methods struggle to cope with the crisis with a widespread epidemic. Deep learning opens up a new path for healthcare systems. A binary classification model for automatic COVID-19 detection was proposed by Qzturk et al. [7]. The accuracy of the model was 98.08%, and there is still room for further improvement of the model. Nour and Cömert [8] designed a new CNN network structure. They used the deep features extracted by this model in machine learning algorithms such as k-nearest neighbors, support vector machine (SVM), and decision tree. The highest accuracy of this model was 98.97%. Xu et al. [9] modified the initial transfer learning model accordingly and built the corresponding algorithm, which was later validated by overall internal and external validation with an accuracy of 89.5%. Ashkan et al. [10] proposed an auxiliary diagnostic system for COVID-19 that uses a modified AlexNet network for feature extraction and a majority voting system for final classification and diagnosis. The system's accuracy was 93.20% on the CT dataset. Danial Sharifrazi et al. [11] proposed a hybrid model that fuses convolutional neural networks with support vector machines and Sobel filters, and the method achieved 99.02% accuracy in the automatic detection of COVID-19. Matteo et al. [12] improved the proposed lightweight network structure based on SqueezeNet and obtained an accuracy of 85.03% on the new crown CT dataset. Suat Toraman et al. [13] proposed a new artificial neural network, CapsNet, which has an accuracy of 97.24% in COVID-19 dichotomous detection and 84.22% correct rate in multiclassification detection. Gonçalo et al. [14] performed COVID-19 detection based on the EfficientNet architecture and showed an average accuracy of 99.63% on dichotomous classification and 96.69% on multiclassification. However, this paper uses a single network model for diagnosis. The accuracy still needs to be improved. D. Apostolopoulos et al. [15] used transfer learning for COVID-19 detection on X-ray images. The results showed that an accuracy of 96.78% was achieved when performing multiclassification detection. Yu-Dong Zhang et al. [16] introduced random pooling instead of global pooling and maximum pooling based on the traditional deep convolution. The model achieved an accuracy of 93.64% and performed better for lung CT detection of COVID-19 patients. H. Benbrahim et al. [17] performed classification of COVID-19 X-ray images based on the Apache Spark framework and using the InceptionV3 and ResNet50 models, which achieved a maximum accuracy of 99.01%. Xiang Yu et al. [18] developed a deep learning framework. The CNN network first extracts the framework, then reconstructed based on the image, and finally classified by the classifier. The accuracy of this framework structure on the COVID-19 CT dataset is 99%. Mujeeb Ur Rehman et al. [19] developed a supervised learning method. Instead of relying on a single salient symptom for diagnosis, this method used multiple symptoms of the subject as features for diagnosis and achieved an accuracy of 97% for COVID-19 diagnosis.

The above work focuses on the diagnosis of COVID-19 but has the disadvantage of not allowing for an analysis of the extent of the patient's condition. Segmentation of chest CT images of patients with COVID-19 can help physicians analyze the extent of the patient's disease and thus adopt the best treatment plan for the patient. Adel Oulefki et al. [20] developed a method for automatic segmentation and measuring chest CT images of COVID-19 patients. The method achieves 0.98 and 0.99 in accuracy and specificity, respectively, which is more accurate compared to methods such as medical image segmentation (MIS). To solve the problem of relying more on real image label information in previous deep learning segmentation methods, Xiaoming Liu et al. [21] proposed a weakly supervised segmentation method for COVID-19 patient chest CT images. Juanjuan He et al. [22] developed an evolvable adversarial framework for COVID-19 patients. The method used three different mutation-evolving generator networks and incorporated gradient penalties into the model to achieve excellent performance in segmenting chest CT images of patients with COVID-19. Nan Mu et al. [23] proposed a network for segmenting chest CT images of patients with pneumonia caused by the SARS-CoV-2 virus. The network fused multiple feature information to identify the boundaries of the patient's lung infection accurately and thus outperformed other segmentation models in terms of performance.

With the rapid development of artificial intelligence, deep learning has demonstrated its unique capabilities in speech recognition, automatic machine translation, and autonomous driving. Deep learning relies on using large amounts of data for training, but insufficient data have been an unavoidable problem in some specific areas. For example, the dataset samples in medical imaging are usually patients suffering from painful diseases, so the dataset may be insufficient to protect patient privacy. Therefore, to solve this problem, transfer learning becomes an essential part. In transfer learning, as pointed out by Ref. [24], the training and test data are not required to be independent and identically distributed (i.i.d.). Also, we do not have to train the target domain model from scratch, significantly reducing the training time and the need for data from the target domain. It is now common practice to pre-train on the natural-image (e.g., ImageNet [25]) datasets or directly use the weights of already trained network models (e.g., ResNet [26], DenseNet [27]) and fine-tune them accordingly in the target medical imaging domain. However, it is pointed out in Ref. [28] that networks pre-trained on large natural image datasets are often over-fitted on the target dataset due to the large differences between medical image datasets and large natural image datasets in terms of quantity and variety. Therefore, in this paper, we will explore a more suitable pre-training scheme in the medical image domain.

The main goal of self-supervised learning (SSL) is not to rely on manual annotation but to learn meaningful representations of input data with the help of pretext. Self-supervised learning methods are mainly divided into three categories. One is to construct corresponding auxiliary tasks based on the context information of the data itself. For example, Word2sec, the most important algorithm in NLP, uses contextual information to make corresponding predictions. In the image field [29], seeks the self-supervised context by predicting the rotation of the input picture from different angles. The research shows that data enhancement brings more training data and improves the robustness of the model, which is very beneficial to self-supervised learning. The second is to construct corresponding constraints based on timing. For example, in Ref. [30], the feature of the adjacent frames in the video is constructed as positive samples, and the far apart frames are constructed as negative samples for self-supervising constraints. The third is a self-supervised learning method based on contrast constraints. This method builds representations by learning to encode the similarity or dissimilarity of two things, which is the current mainstream method. The core idea of this method is to realize self-supervised learning by measuring the distance between positive and negative samples, and the distance between the sample and the positive sample is much larger than the distance between the sample and the negative sample. DIM [31] uses the local features of the image as positive samples and other images' local features as negative samples to achieve contrast constraints. CMC [32] proposes to take multiple modalities of one sample as positive samples and multiple modalities of other samples as negative samples. Related researchers put forward the related concepts of memory bank



(a) Covid-19 (b) Normal

Fig. 1. Sample dataset example.

[33] to efficiently access and calculate the loss. MoCo [34] uses momentum to update encoder parameters based on the idea of a memory bank, which solves the problem of inconsistent encoding of new and old candidate samples. SimCLR [35] makes the model training effect close to the supervised model by adopting methods such as increasing nonlinear mapping, larger batch size, and data enhancement. MoCo v2 [36] improves the data enhancement method based on MoCo, and adds the same nonlinear layer to the representation of the encoder during training. The model is better than SimCLR under a smaller batch size.

Compared with the annotation of natural images, the annotation of medical images needs to be performed by experts with relevant specialties. In addition, the annotation of medical images requires very specialized medical knowledge, and a lot of subjective ideas are added in the annotation process. Therefore, there exists a large amount of information in unlabeled medical images that cannot be utilized. In this paper, a new pre-training method based on migration learning and self-supervised learning (TS) is proposed. The method effectively alleviates the problem of excessive differences between the source and target domains in traditional migration learning methods while enabling the information in unlabeled images to be effectively utilized.

1.3. Our work

The above methods have achieved good results in diagnosing COVID-19, but there are still some problems. 1) In the field of natural images, most images are manually annotated by ordinary people. However, due to the particularity of the medical image field, the label of the dataset must be annotated by experts in the relevant field. Therefore, there is a large amount of information in unlabeled images that cannot be used. 2) In medical imaging, dataset samples are relatively scarce.

The general solution is to migrate the model to downstream target tasks for fine-tuning after ImageNet pre-training to improve the model's generalization ability and speed up the training speed. However, natural and medical images are quite different in type and quantity, so this may not be the best pre-training method. 3) The accuracy of the above method needs to be further improved. We propose a new CAD method for diagnosing suspected COVID-19 patients to solve the above problems. The main contributions of this article are as follows:

1) To improve the accuracy of the diagnosis of suspected CVOID-19 cases, we proposed a new hybrid ensemble model (TSRNet) to diagnose suspected cases.
2) To solve the problems in the traditional pre-training method and use the information in the unlabeled image, we proposed a new pre-training method (TS).
3) To enable the model to better focus on the lesion area, we proposed a new convolutional neural network based on attention mechanism and deep residual network to extract features.
4) Our model has the highest accuracy of 99.80% compared to the five existing models, which suggests that our proposed model can help radiologists diagnose COVID-19 suspected patients more accurately.

Table 1
Dataset used in this study.

Class	Train	Val	Test	Total
Non-COVID	737	246	246	1229
COVID-19	751	251	250	1252

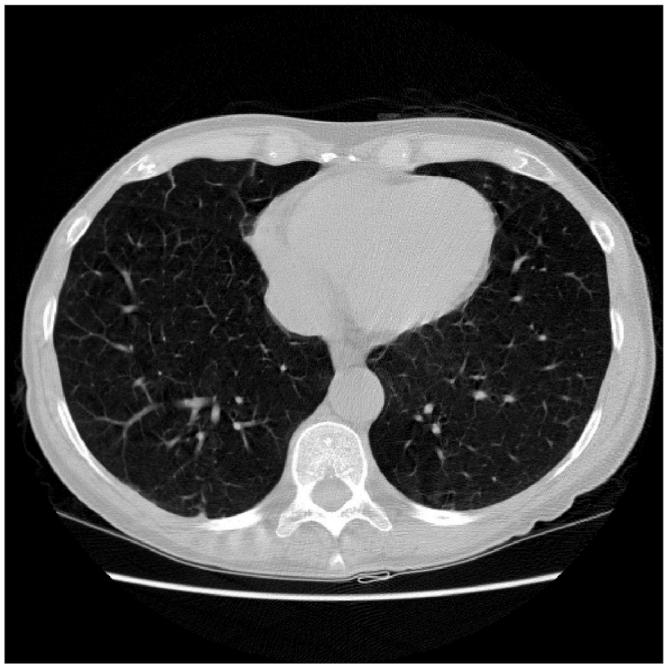


Fig. 2. A sample of LUNA dataset.

The other subsections of the paper are structured as follows: Section 2 introduces the dataset required for the experiments, Section 3 presents the principles of the methods used and the general structure of the model, Section 4 presents the experimental part, and Section 5 concludes with a summary.

2. Dataset

A total of four datasets were used in this experiment. The first is the ImageNet dataset, which was established to facilitate the development of computer image recognition technology. The dataset is huge and contains most of the image categories we will see in our lives. Therefore, this dataset is usually regarded as the source domain of transfer learning. The second one is the COVID dataset from Ref. [37]. This dataset contains 1252 COVID-19 patients and 1229 lung CT images of healthy people. Fig. 1 shows a sample example of this dataset. This dataset is

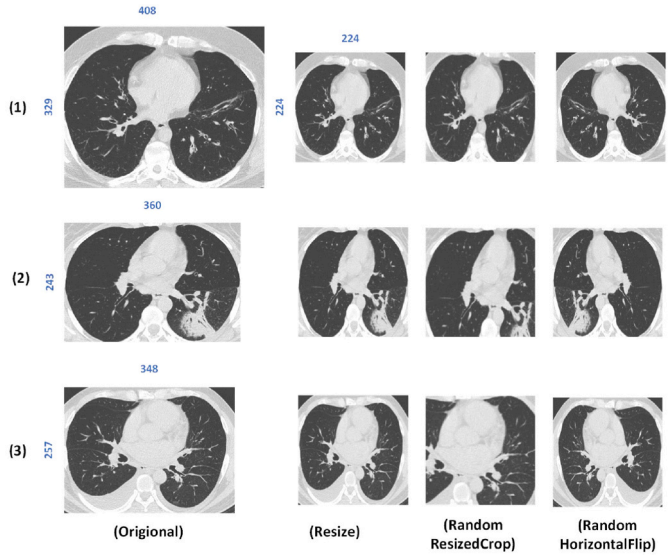


Fig. 3. Data augmentation.

divided into a training set, a validation set, and a test set in the ratio of 0.6:0.2:0.2. Table 1 lists the relevant detailed data. In addition, the Lung Nodule Analysis (LUNA) [38] was used as the source of the unlabeled CT data. This dataset has a total of 1000 images and will be used for unlabeled images for self-supervised learning but will not be used as negative examples in CVOID-19. Fig. 2 shows a sample example of the LUNA dataset. To meet the input requirements of the model, we converted all CT images to $224 \times 224 \times 3$ before input. Finally, to verify the impact of the difference between the source domain and the target domain on transfer learning, the data in Ref. [39] was used as a secondary dataset for validation.

In order to improve the generalization ability of the network model, different data enhancement operations, such as random cropping and horizontal flipping, are performed on the COVID dataset to improve the robustness of the network model. The specific operations are shown in Fig. 3.

3. Methodology

3.1. Transfer learning with self-supervised learning (TS)

Although transfer learning is top-rated in medical imaging, according to Ref. [40] studies, there are two main problems with transfer learning in medical imaging at this stage. One is that pre-trained networks on ImageNet are usually over-parameterized in downstream target tasks. Second, there is a big difference between the dataset image in the original task and the target task dataset image. For example, the images in the ImageNet dataset are natural images of flowers, birds, fish, insects, etc., while the experimental dataset is the chest CT images of COVID-19 patients. Secondly, there are many dataset images in the natural image dataset. There are 1000 image categories in the ImageNet dataset, but fewer dataset images are in the medical imaging field.

Furthermore, compared with the 1 million images in ImageNet, the medical imaging dataset may range from a few thousand to a few hundred thousand. Therefore, the characterization information learned by the network model on ImageNet may not be well adapted to CT images. We doubt the feasibility of transplanting the network model from natural images to chest CT images of COVID-19 patients.

There are problems in the medical image domain, such as fewer datasets and more expensive annotation costs. Meanwhile, the commonly used transfer learning methods suffer from the problem of large bias between the source and target domain datasets. Therefore, a new pre-training method (TS) is proposed to solve the problems in the traditional pre-training methods. TS incorporates self-supervised learning into transfer learning, enabling the model to learn efficient and unbiased representational information on unlabeled datasets. The pseudocode of TS is shown in Table 2.

We adopted a self-supervised learning task based on contrast loss in this work. We first constructed a queue dictionary to store relevant negative samples. After that, the model is used to determine whether the two samples generated by data enhancement are from the same source image so that the model can be trained.

In the comparative self-supervised learning task, image enhancement is first performed on the original data to obtain two pictures Z_q and Z_k. Z_q is called a query, and Z_k is called a key. Use the query encoder Q and the key encoder K initialized with the weight ε_q and the weight ε_k to obtain the potential representation of the query and the keywords $Q = F_q(Z_q; \varepsilon_q)$ and $K = F_k(Z_k; \varepsilon_k)$. We mark query and key from the same

Table 2
TS pseudocode.

Algorithm 1 Algorithm of TS

Input: batch size N_S, temperature τ, LUNA dataset D_L, model F pretrained on ImageNet. The queue S serves as a storage container for the key K. a is the same as b for the data enhancement operation.

Initialize encoder networks for query F_q and keys F_k: $F_q = F_k = F$.

for minibatch $\{x_l\}_{l=1}^{N_S}$ where $x_l \in D_L$ **do**

$x_{lq} = a(x_l)$, $x_{lk} = b(x_l)$,

$Q = F_q(x_{lq})$, $K = F_k(x_{lk})$,

$L_{possitive} = \{Q_i^T K_i\}_{i=1}^{N}$, $L_{negative} = \{Q_i^T S_i\}_{i=1}^{N}$

$L = concate([L_{possitive}, L_{negative}], dim = 1)$,

$L^{\wedge} = zero(length(x))$,

$loss = CrossEntropyLoss(L/\tau, L^{\wedge})$,

Update F_q by equation (6), F_k by equation (7), and S

end for

return F_q

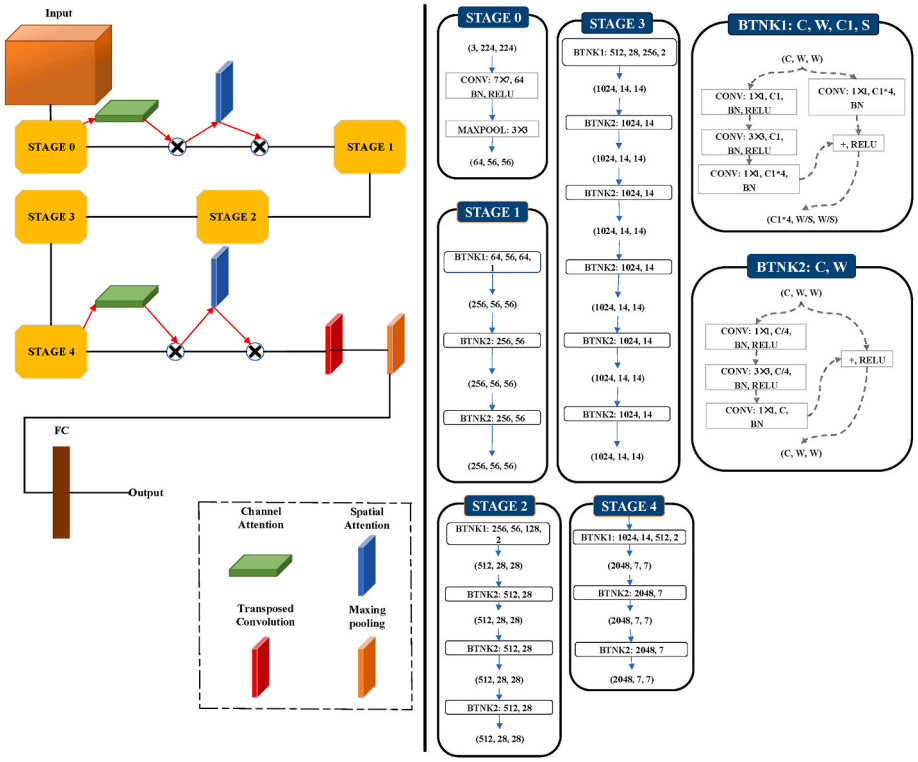


Fig. 4. The structure of the proposed RANet.

image as a positive pair and query and key from different images as a negative pair. We use the queue as a dictionary to store a set of keywords $\{K_j\}$ from different images. According to the first-in-first-out feature of the queue, the oldest batch of keys will be used as negative keys and replaced by new queries. This mechanism prevents irregular sampling of negative sampling. We are given the pair (Q_i, K_i) obtained from the new image. The contrast loss can be expressed as [41]:

$$Loss = -\log \frac{\exp^{\left(\frac{Q_i \cdot K_i}{\tau}\right)}}{\exp^{\left(\frac{Q_i \cdot K_i}{\tau}\right)} + \sum_{j=1}^{N-1} \exp^{\left(\frac{Q_i \cdot K_j}{\tau}\right)}} \tag{1}$$

where $Q_i.K_i$ represents the representation vector corresponding to the sample and the positive example, and $Q_i.K_j$ represents the representation vector corresponding to the sample and the negative example. The numerator part of this function will close the distance between the sample and the positive example, while the denominator part, which encourages the vector similarity between the sample and the negative example to be as low as possible, pushes away the distance between the sample and the negative example.

We set the queue size to 512 and add a multilayer perceptron with 2048 hidden units to the model structure. In addition, in specific experiments, the optimizer uses the stochastic random gradient descent method.

$$h = encoder(Z) \tag{2}$$

$$Q = F(Z) = W^{(2)}\sigma(W^{(1)}h) \tag{3}$$

Here *encoder* is the neural network that extracts the representation information from the image sample and Z is the image sample. Equation (3) shows that the representation information h extracted by the *encoder* is projected by a multilayer perceptron to obtain Q, where σ is the ReLU nonlinear activation function.

$$\sigma = \max(0, x) \tag{4}$$

$$Q_i \cdot K_i = \frac{Q_i^T \cdot K_i}{\|Q_i\| \|K_i\|} \tag{5}$$

In this study, the parameter ε_q of the query encoder F_q is updated by back propagation, while the parameter ε_k of the key encoder F_k is updated by momentum. ε_q and ε_k are updated using the following rules:

$$\varepsilon_q = \varepsilon_q - \alpha \frac{\delta_{Loss}}{\delta_{\varepsilon_q}} \tag{6}$$

$$\varepsilon_k = m\varepsilon_k + (1-m)\varepsilon_q \tag{7}$$

where $m = 0.999$ is the momentum coefficient and α is the learning rate of the query encoder. The specific reasons are as follows. In the initial end-to-end self-supervised learning approach, the parameters of the query encoder F_q and the key encoder F_k are updated at each step. Since the input of the key encoder F_k is a negative sample of batch (i.e., N-1), the number of inputs cannot be too large. Therefore, the dictionary cannot be too large, and the batch size cannot be too large.

Now, the key encoder F_k is updated using the momentum method, which does not involve backpropagation, so the number of negative samples of the input can be large. Specifically, the queue size can be larger than the mini-batch, which is a hyperparameter. The queue is gradually updated in each iteration, and the current mini-batch samples are listed, as are the oldest mini-batch samples in the queue, i.e., the more negative samples, the better, of course. The form of momentum update allows the dictionary to contain more negative samples. In addition, the update of the key encoder F_k is extremely slow ($m = 0.999$ very close to 1), so the update of the key encoder F_k is equivalent to looking at many negative samples.

In addition, in the original end-to-end self-supervised learning approach, the representations of all samples are stored in the dictionary, and the K obtained from the latest Q sampling may be the K obtained from the encoders encoded several steps ago, thus losing the consistency problem. However, in this paper, the encoder F_k is updated by momentum m for each step. Although the update is slow, the query encoder

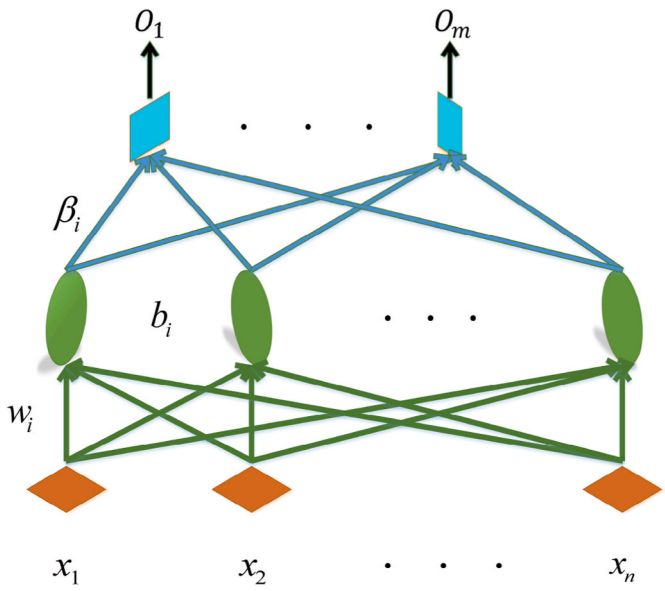


Fig. 5. Structure of ELM

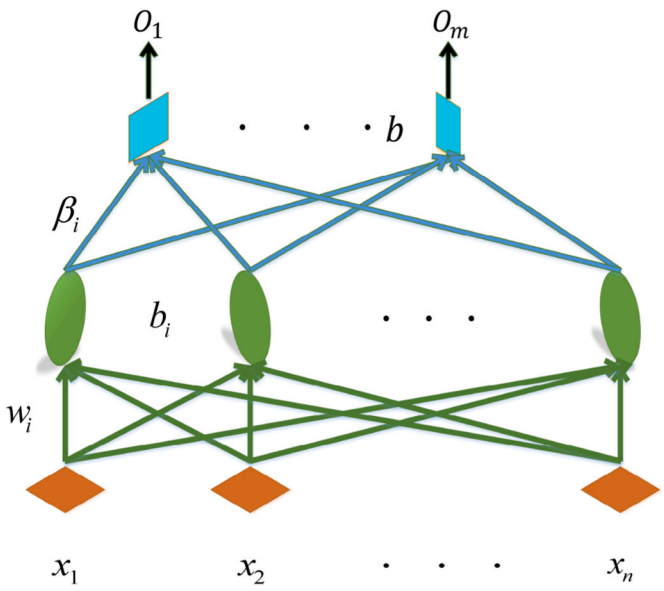


Fig. 6. Structure of SNN.

F_q and key encoder F_k are updated for each step, so the consistency problem is solved.

3.2. RANet

The attention mechanism is a kind of data processing in machine learning, widely used in natural language processing, image recognition, and other fields. The attention mechanism operates through a neural network to generate a mask that evaluates the rating of the current point to be attended. Attention mechanisms are divided into three categories: (1) Channel attention mechanisms generate a mask for the channel and score it. (2) Spatial attention mechanism: the mask is generated for the space and scored. (3) Hybrid attention mechanism: scoring both channel and spatial attention.

The convolutional block attention module (CBAM) [42] was proposed by Sanghyun Woo et al., in 2018. Compared with SENet [43], which only focuses on the channel attention mechanism, CBAM combines both spatial and channel attention mechanisms to achieve better results than SENet.

In this paper, a novel convolutional neural network based on attention mechanism and deep residual network (RANet) is designed. The specific structure and hyperparameters of the network are shown in Fig. 4. We label the similar parts of the network as five stages. Afterward, we insert the CBAM between stage0 and stage1, insert another CBAM after stage4, and add the transposed convolution layer and the max-pooling layer before the final fully-connected layer. In subsequent related experiments, we will conduct comparative experiments with other classic convolutional neural networks to prove the effectiveness of the proposed RANet.

3.3. Classifier

Extreme learning machine (ELM) is a single-hidden-layer feedforward neural network algorithm proposed by Prof. Guangbin Huang in 2004. Like the traditional single-hidden neural network structure, ELM has only a three-layer structure, containing an input layer, an intermediate hidden layer, and a final output layer. BPNN is the most famous classical feedforward neural network. Still, it often stays at local extremes because gradient descent is a greedy algorithm that cannot jump out of locally optimal solutions.

ELM uses a new training algorithm that randomly generates input weights and deviations during training while deriving output weights

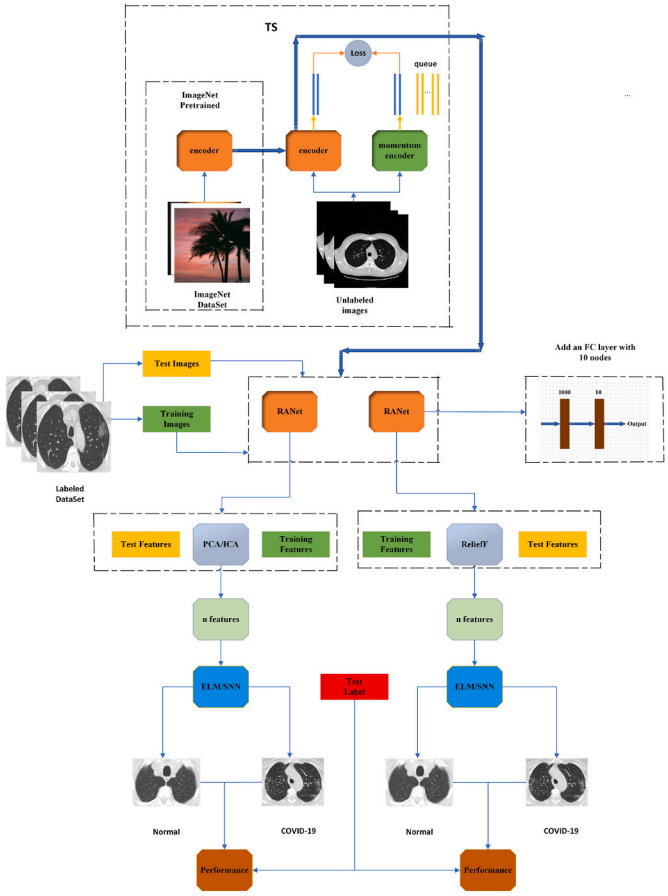


Fig. 7. TSRNet overall architecture diagram.

based on the generalized inverse matrix principle. As a result, the training time of ELM is significantly reduced compared to traditional neural networks based on gradient descent algorithms.

The network structure of the ELM is shown in Fig. 5. The structure of the Schmidt neural network (SNN) is similar to that of ELM. The output bias of the ELM is always 0, but the output bias of the SNN may not be 0.

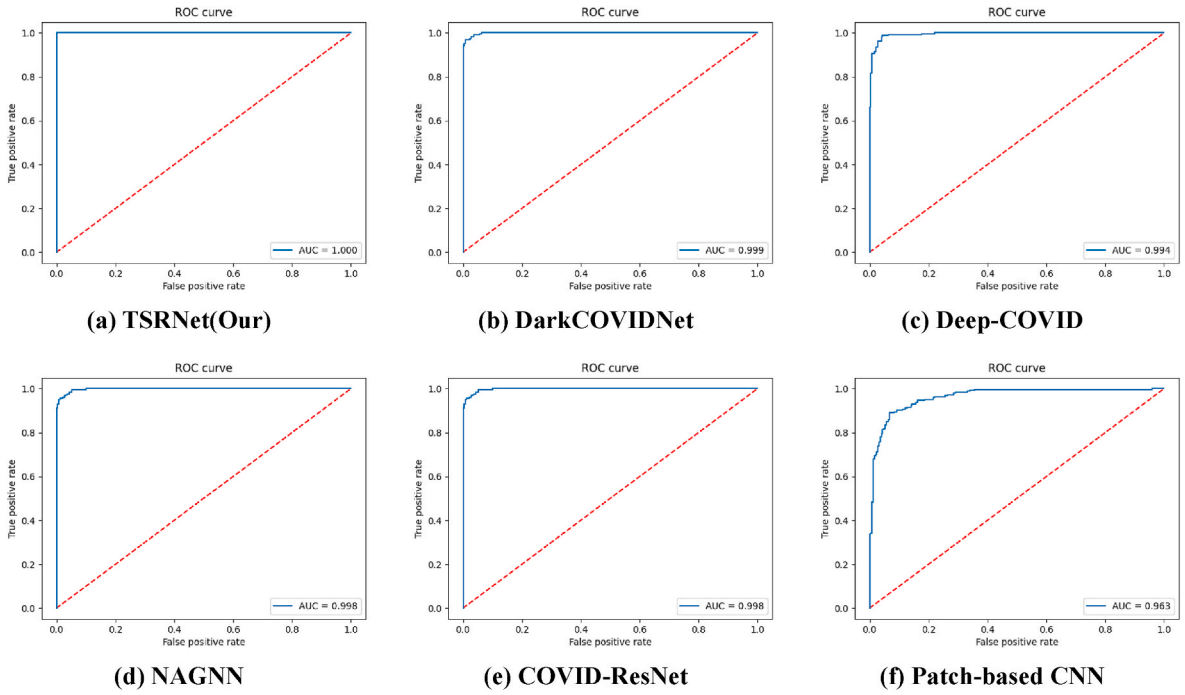


Fig. 8. ROC curves.

The structure of the SNN is shown schematically in Fig. 6. Since the feature variables extracted by RANet are highly correlated, the classifier may not be satisfactory if all feature variables are used to fit the classifier. Therefore, we optimize the data using data downscaling and feature selection.

3.4. TSRNet

We propose a new CAD system (TSRNet) for COVID-19 diagnosis to solve the problems of traditional pre-training methods in the field of medical images and improve the accuracy of diagnosis of suspected COVID-19 patients. The overall structure of the TSRNet is shown in Fig. 7. We first pre-trained the RANet by TS method and then fine-tuned the pre-trained RANet on the target dataset for feature extraction. Due to the high similarity of the features extracted by RANet, we use two different optimization methods for the extracted features, data downscaling and feature selection, to achieve better results for the classifier. For data dimensionality reduction, we use two algorithms, principal component analysis (PCA) and independent component analysis (ICA). The solution steps of PCA are generally to obtain the covariance matrix of the data first, then calculate the eigenvalues and eigenvariables of the covariance matrix, and finally select the final principal component based on the contribution of the eigenvalues. PCA extracts the features unrelated to each other, while ICA extracts the features independent of each other. We use the ReliefF algorithm for feature selection, a feature weighting algorithm. First, each feature is given a different weight according to its relevance to the category, and features with weights less than a certain threshold will be removed. Since the feature selection algorithm generally selects in a small number of feature spaces, we will add a 10-node fully connected layer to the output layer of RANet. After that, we perform the final classification by Extreme Learning Machine (ELM) and Schmidt Neural Network (SNN). Finally, we find the best model solution by comparing and analyzing the results.

4. Experiments

4.1. Performance measures

This study used five evaluation metrics to evaluate the proposed method: accuracy, recall, F1-score, accuracy, and AUC. They are defined as follows:

$$\text{Pression} = \frac{TP}{TP + FP} \times 100\% \tag{8}$$

$$\text{Recall} = \frac{TP}{TP + FN} \times 100\% \tag{9}$$

$$F1 - score = \frac{2TP}{2TP + FP + FN} \times 100\% \tag{10}$$

$$Accuracy = \frac{TP + TN}{TP + FN + TN + FP} \times 100\% \tag{11}$$

AUC (Area under Curve): indicates the area under the Roc (receiver operating characteristic) curve, and its value is between 0.5 and 1. The larger the value of AUC, the greater the probability that the current model ranks positive cases over negative cases, and the better the classification effect. Among them, TP (true positive) represents that the predicted attribute is COVID-19, and the true attribute is CVOID-19. TN (true negative) represents that the predicted attribute is normal, and the true attribute is normal. FP (false positive) indicates a predicted attribute of COVID-19 and a true attribute of normal. FN (false negative) indicates that the true attribute is COVID-19, but the predicted attribute is normal.

As seen from the ROC curves in Fig. 8, our model again shows strong performance. Compared with other models, the AUC value of our proposed model is 1.0, which indicates that our model performs better in performing image classification. In summary, the comprehensive model proposed in this paper can perform the diagnostic task of COVID-19 with high accuracy, which can assist the radiology department in diagnosing suspicious individuals and facilitate subsequent treatment and corresponding isolation.

In addition, the confusion matrix was used in this experiment to

Table 3
Classification result.

ImageNet Pretrained	Precision	Recall	F1	Accuracy
ResNet50	98.74%	95.93%	97.32%	97.38%
RANet	99.58%	96.75%	98.14%	98.19%
RANet + ELM	100%	98.37%	99.18%	99.19%
RANet + PCA + ELM	**100%**	**99.18%**	**99.59%**	**99.59%**

evaluate the model's performance. In the confusion matrix, the columns of the matrix represent the predicted attributes of the sample, and the rows of the matrix represent the true attributes of the sample.

4.2. Experimental settings

Thanks to your suggestion, we have supplemented the experimental setup in subsection 4.2 and marked it in yellow. All network models in this study were pre-trained on a server with 64 GB RAM, CPU Intel Xeon Silver 4214, and GPU Quadro RTX 8000. All subsequent experiments were performed on RTX 2060 GPUs. In addition, all networks used for feature extraction were generated based on the PyTorch [44] framework. We use the scikit-learn [45] for feature dimensionality reduction. Finally, we get the classification results through the classifier implemented in Python.

4.3. Evaluation of RANet

To verify the effectiveness of our proposed RANet, we conducted a comparison experiment with ResNet50. The network in this experiment was pre-trained on the ImageNet dataset only. The experimental results are shown in Table 3. The error rate of RANet in image recognition is reduced by 30.92% compared to the ResNet network, proving that the network will also outperform the traditional ResNet50 network as a feature extractor. Fig. 9 shows the visualization of RANet and ResNet50

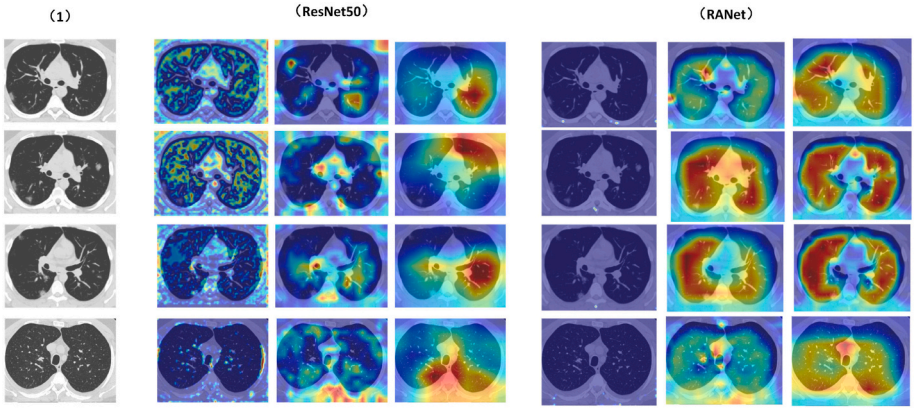
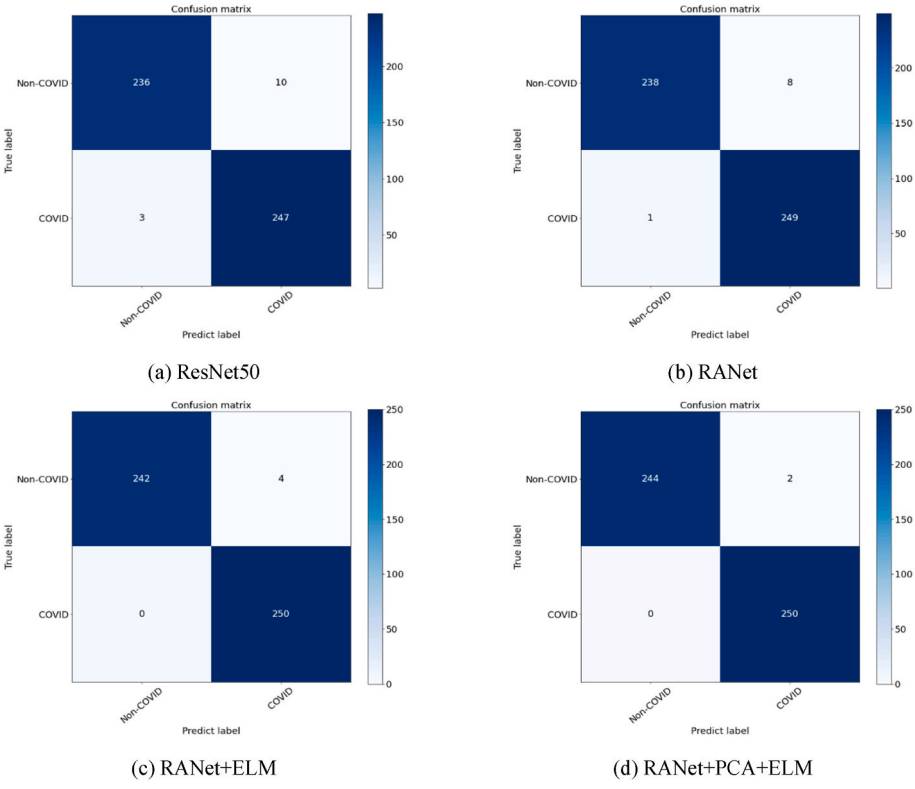


Fig. 9. Grad-CAM visualization of ResNet50 with RANet.



(a) ResNet50 (b) RANet

(c) RANet+ELM (d) RANet+PCA+ELM

Fig. 10. Confusion matrix.

Table 4
The classification effect of RANet under different transfer learning methods.

Transfer learning method	Precision	Recall	Accuracy
Random initialization	99.57%	93.50%	96.57%
Pre-training on ImageNet	99.58%	96.75%	98.19%
Pre-training on CT	99.80%	97.32%	98.39%
Pre-training on ImageNet and CT	**99.91%**	**97.56%**	**98.54%**

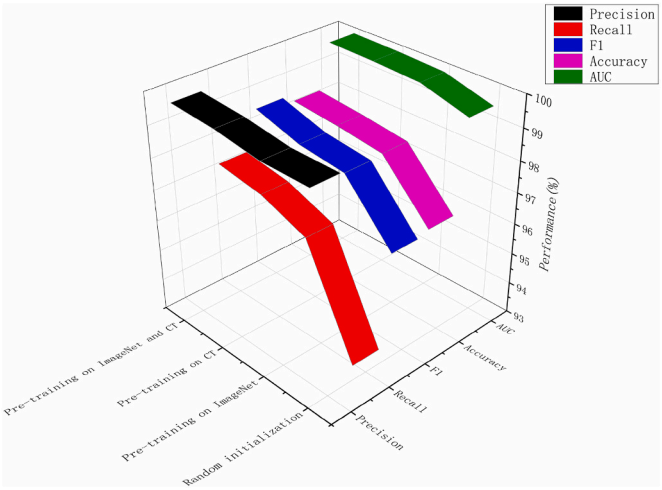


Fig. 11. Transfer learning performance of RANet.

on the same layer of Grad-CAM. By comparison, we find that ResNet50 always focuses on regions unrelated to the lesion area.

In contrast, RANet can accurately focus on the lesion region, suggesting that RANet can detect COVID-19 more accurately than ResNet50. After that, we conducted more in-depth experiments to verify the effect of the optimized classifier. The confusion matrix plot for the experiments is shown in Fig. 10. All the data on this plot are the results of the test set. We found that the model's accuracy in performing correct classification improved by 1% after adding the classifier. After the PCA data downscaling, the model's accuracy reached 99.59%. The results show that both data downscaling and classifiers are beneficial to improving the model structure's accuracy, but adding classifiers to the model structure is more effective than data downscaling.

4.4. Evaluation of transfer learning

To verify the effectiveness of transfer learning and to validate the effect of the difference between source and target domain data on transfer learning, we selected pre-trained RANet networks under different datasets for the corresponding tests. (a) We randomly initialize the parameters of the RANet network. (b) We pre-trained RANet on the ImageNet dataset and then fine-tuned it accordingly on the target dataset. (c) We pre-trained RANet on the [39] dataset and then fine-tuned it accordingly in the downstream task. (d) We first pre-trained the RANet network on the ImageNet dataset, then transferred it to the [39] dataset for training, and finally fine-tuned the network on the target dataset. The experimental results are shown in Table 4. We found that transfer learning can greatly improve the network's performance on the target dataset compared to the random initialization of the network in Fig. 11. In addition, since the ImageNet dataset is more different from the target dataset in terms of number and type, the [39] dataset is closer to the target dataset.

Therefore, transfer learning performance is better when the source domain data is closer to the target domain data. Finally, the [39] dataset has a single data type and small data volume. Therefore, after pre-training the network model on the large dataset, transferring it to a smaller dataset closer to the target domain data, and pre-training it again, the network will learn more unbiased information about the image representation. The performance will be improved again on the target dataset.

Table 5
Classification results under different pre-training methods.

TS					
Experiment 1	Precision	Recall	F1	Accuracy	AUC
TSRNet	**100%**	**99.59%**	**99.79%**	**99.80%**	**1**
DenseNet169	100%	99.19%	99.59%	99.60%	1
Only ImageNet Pretrained					
Experiment 2	Precision	Recall	F1	Accuracy	AUC
TSRNet	100%	99.18%	99.59%	99.59%	0.9996
DenseNet169	99.59%	98.78%	99.18%	99.19%	0.9992
Random Initialization					
Experiment 3	Precision	Recall	F1	Accuracy	AUC
TSRNet	100%	97.96%	98.97%	98.99%	0.9998
DenseNet169	97.98%	98.37%	98.17%	98.19%	0.9999

4.5. Evaluation of TS

As seen in Section 4.4, pre-training the network model on the ImageNet dataset and then transferring it to a dataset similar to the target dataset for pre-training again will facilitate the network model to learn unbiased image representation information, which is more beneficial for downstream tasks. However, these pre-training processes are performed under supervised learning. Therefore, to enable the network model to exploit the information of unlabeled images, we consider introducing self-supervised learning in the pre-training of the model.

This section tests the role of self-supervised learning in transfer learning. We use TSRNet and DenseNet169 to test together under the same experimental conditions to avoid experimental chance. The experiments are divided into three groups. The first set of experiments was conducted under the TS pre-training method. The second set of experiments was just pre-training of the network on the ImageNet dataset. The third set of experiments was just a random initialization of the network. After that, the pre-trained networks from the three sets of experiments are fine-tuned as feature extractors on the target dataset. Finally, the final classification results are obtained by downscaling and classifier classification.

As shown by Table 5, the TSRNet accuracy of the scheme of Experiment 2 reaches 99.59%, which is 0.6% higher compared with the model accuracy of the scheme of Experiment 3. In contrast, the model accuracy in the scheme of Experiment 1 reached 99.80%, which is 0.21% higher compared to the model accuracy of Experiment 2. In performing self-supervised learning, we learn with the help of data from the target task without using any labels. In contrast, transfer learning under supervised learning is performed with the help of data labels, which makes the parameters in the network model more biased towards the source data and labels. Thus, by incorporating self-supervised learning into transfer learning, the model structure learns more unbiased image representations than transfer learning under supervision, which is more beneficial for fine-tuning the target data.

4.6. Comparison with state-of-the-art approaches

To verify the overall effectiveness of TSRNet, we compared it with five models (DarkCOVIDNet [7], Deep-COVID [46], NAGNN [47], COVID-ResNet [48], Patch-based CNN [49]). The results are shown in Fig. 12. In terms of accuracy, compared with the existing models (DarkCOVIDNet: 98.80%; Deep-COVID: 97.58%; NAGNN: 97.86%; COVID-ResNet: 97.78%; Patch-based CNN: 88.90%), TSRNet has the highest accuracy of 99.80%. In terms of other indicators, TSRNet still

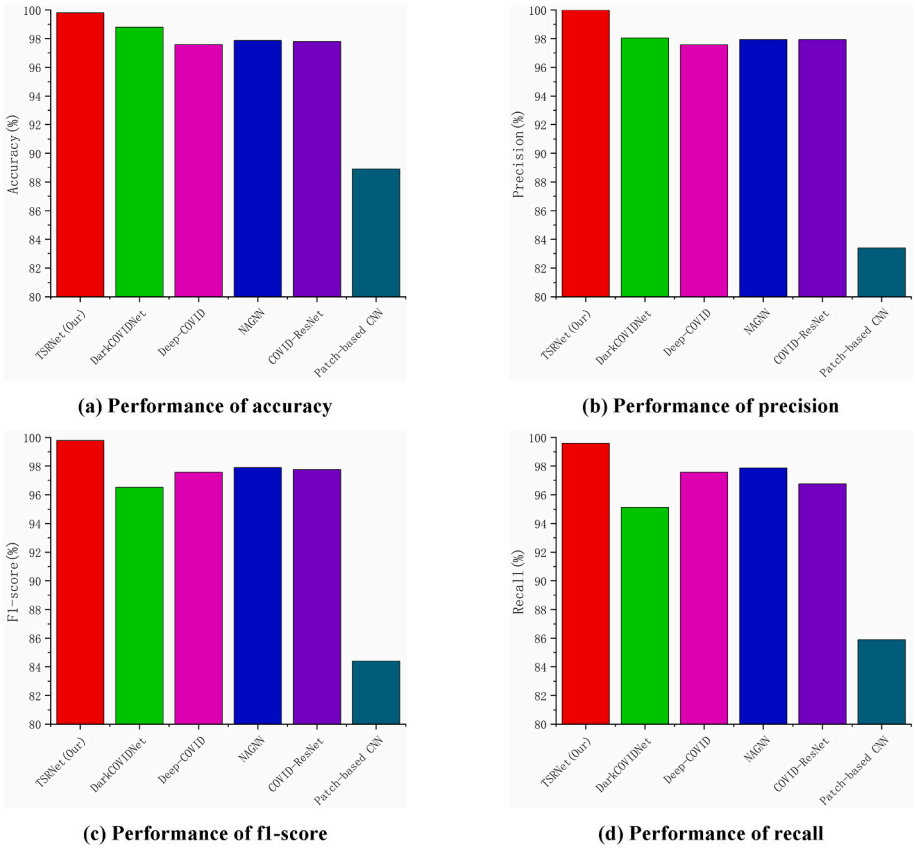


Fig. 12. Comparison of experimental results.

has a great advantage. This indicates that our model has the outstanding ability both in performing discrimination as a whole and still excels in performing specificity discrimination, which suggests that our method shows a stronger ability to diagnose patients with suspected COVID-19.

4.7. Ablation experiment

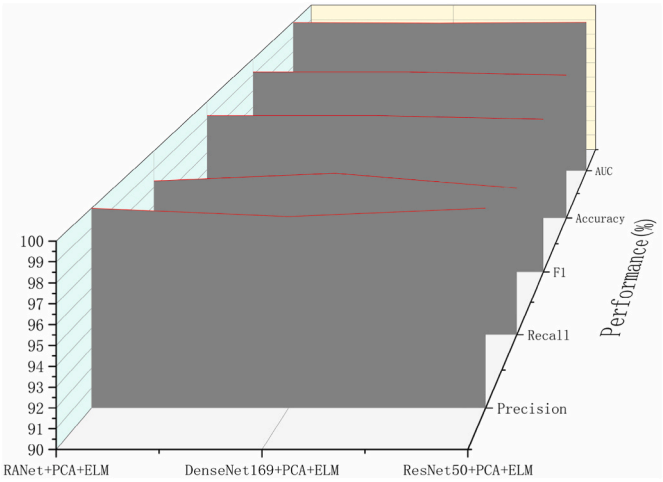
In the previous subsections, we validate the effectiveness of our proposed model. Also, we demonstrated that incorporating self-supervised learning into transfer learning as a model pre-training method can significantly improve the model's accuracy. In addition, our model consists of three components after pretraining: (feature extraction, feature dimension reduction, and classifier). These three components can be replaced with each other using different components. To verify the effectiveness of the combination between different components, we designed three sets of experiments: (a) keeping the feature downscaling and classifier unchanged and replacing the feature extractor; (b) fixing the feature extractor and classifier and choosing a different feature downscaling algorithm (since ReliefF is a feature selection algorithm that acts on small samples, we replace the output of RANet during the experiments with dimensionality to 10); (c) we fix the feature extractor with the classifier accordingly and choose a different classifier for validation.

The network components in this experiment were pre-trained on the ImageNet dataset to maintain experimental rigor. The experimental results are shown in Fig. 13. Our proposed model (TSRNet) has the highest accuracy for COVID-19 diagnosis, which proves the effectiveness of our proposed method. 1) The model using different components still achieves high performance. 2) The comprehensive performance of our proposed model has reached the best. 3) Changing different classifier components has a relatively large impact on the classification results. As shown in Fig. 13, our proposed TSRNet has an accuracy of 99.59% when using ELM as the classifier, but 99.08% when the classifier is replaced with SNN. Therefore, we have to choose the best classifier when designing the model. In addition, the experiment is still not perfect. We uniformly set the number of features after feature dimensionality reduction to 4, which may not be the best value. We will further investigate this value in the follow-up work.

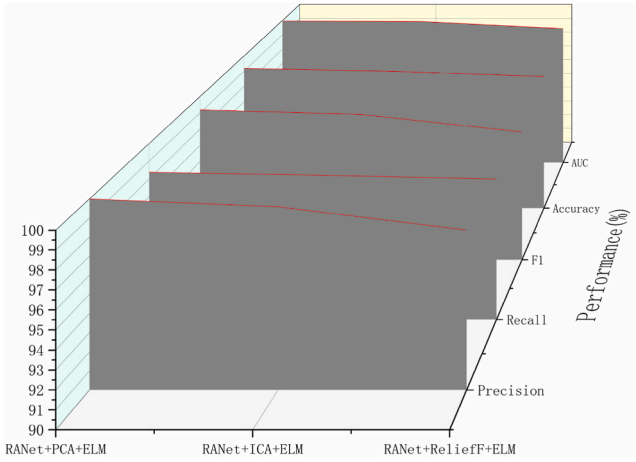
5. Conclusion

Since image labeling in medical images needs to be labeled accordingly by experts, this task is very costly and labor-intensive. To enable the network model to exploit the information in unlabeled data and improve the classification model's accuracy, we propose a hybrid model (TSRNet). The model first incorporates self-supervised learning into transfer learning to replace the traditional pre-training process under supervised learning. The pre-trained model is used as a feature extractor to fine-tune the feature extraction on the target dataset. Finally, data dimensionality reduction and classifier classification obtain the final classification results. Compared with the existing (DarkCOVIDNet: 98.08%; Patch-based CNN: 88.90%; NAGNN: 97.86%; Deep-COVID: 97.58%; COVID-ResNet: 97.78%) models, our model achieves an accuracy of 99.80%. This indicates that our model can complete the detection of new coronary pneumonia with high accuracy, thus helping doctors diagnose suspected patients more accurately and thus preventing the spread of the epidemic.

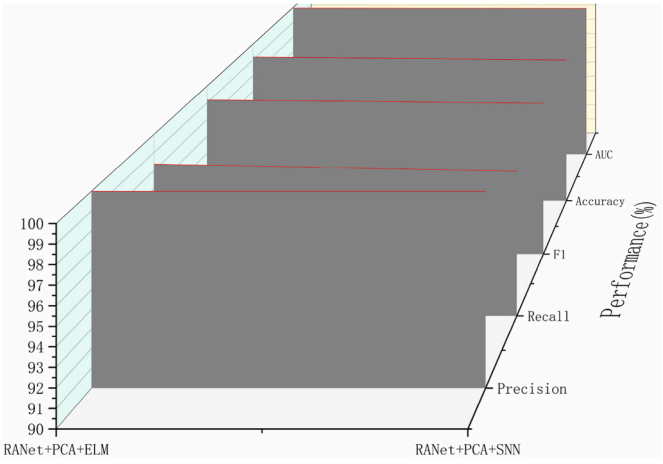
However, our proposed approach still has some drawbacks. The first is the interpretability of the CNN model, although Grad-CAM was able to show that the model's predictions for chest CT images were derived from the relevant lesion regions. Secondly, our model can only predict whether a patient is sick or not; however, segmentation of chest CT images of COVID-19 patients can help physicians analyze the extent of the patient's disease and thus adopt the best treatment plan for the



(a) Using different backbone networks



(b) Using different data processing methods



(c) Using different classifiers

Fig. 13. Classification performance.

patient. Therefore, we will try this aspect in our future work. Finally, the pre-training scheme in this study is more resource-consuming. Therefore, we will optimize it to reduce the demand on hardware resources in future work.

Author contributions

Junding Sun: Conceptualization, Methodology, Formal analysis, Investigation, Resources, Writing - Review & Editing, Project administration, Funding acquisition; Pengpeng Pi: Conceptualization, Methodology, Software, Data Curation, Writing - Original Draft, Visualization; Chaosheng Tang: Software, Validation, Resources, Visualization, Supervision, Project administration; Shui-Hua Wang: Methodology, Validation, Formal analysis, Resources, Visualization, Supervision, Project administration; Yu-Dong Zhang: Conceptualization, Validation, Formal analysis, Investigation, Writing - Original Draft, Writing - Review & Editing, Supervision, Funding acquisition.

Funding

Key Science and Technology Program of Henan Province, China (212102310084); Key Scientific Research Projects of Colleges and Universities in Henan Province(22A520027).

Institutional review board statement

Not applicable.

Informed consent statement

Not applicable.

Data availability statement

Not applicable.

Declaration of competing interest

Not applicable.

References

[1] T. Struyf, et al., Signs and symptoms to determine if a patient presenting in primary care or hospital outpatient settings has COVID-19, Cochrane Database Syst. Rev. 2 (Feb 23 2021) CD013665.
[2] W. H. Organization, Coronavirus disease (COVID-19) pandemic [Online]. Available: https://www.who.int/emergencies/diseases/novel-coronavirus-2019.
[3] A. Tahamtan, A. Ardebili, Real-time RT-PCR in COVID-19 detection: issues affecting the results, Expert Rev. Mol. Diagn 20 (5) (May 2020) 453–454.
[4] V.M. Corman, et al., Detection of 2019 novel coronavirus (2019-nCoV) by real-time RT-PCR, Euro Surveill. 25 (3) (Jan 2020).
[5] Y. Fang, et al., Sensitivity of chest CT for COVID-19: comparison to RT-PCR, Radiology 296 (2) (Aug 2020) E115–E117.
[6] U. Teichgraber, et al., Ruling out COVID-19 by chest CT at emergency admission when prevalence is low: the prospective, observational SCOUT study, Respir. Res. 22 (1) (Jan 12 2021) 13.
[7] T. Ozturk, M. Talo, E.A. Yildirim, U.B. Baloglu, O. Yildirim, U. Rajendra Acharya, Automated detection of COVID-19 cases using deep neural networks with X-ray images, Comput. Biol. Med. 121 (Jun 2020), 103792.
[8] M. Nour, Z. Comert, K. Polat, A novel medical diagnosis model for COVID-19 infection detection based on deep features and bayesian optimization, Appl. Soft Comput. 97 (Dec 2020), 106580.
[9] S. Wang, et al., A deep learning algorithm using CT images to screen for Corona virus disease (COVID-19), Eur. Radiol. 31 (8) (Aug 2021) 6096–6104.
[10] A. Shakarami, M.B. Menhaj, H. Tarrah, Diagnosing COVID-19 disease using an efficient CAD system, Optik 241 (Sep 2021), 167199.
[11] D. Sharifrazi, 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 (Jul 2021), 102622.
[12] M. Polsinelli, L. Cinque, G. Placidi, A light CNN for detecting COVID-19 from CT scans of the chest, Pattern Recogn. Lett. 140 (Dec 2020) 95–100.
[13] S. Toraman, T.B. Alakus, I. Turkoglu, Convolutional capsnet: a novel artificial neural network approach to detect COVID-19 disease from X-ray images using capsule networks, Chaos, Solit. Fractals 140 (Nov 2020), 110122.
[14] D. Agarwal, Automated medical diagnosis of COVID-19 through EfficientNet convolutional neural network, Appl. Soft Comput. 96 (2020).
[15] I.D. Apostolopoulos, T.A. Mpesiana, Covid-19: automatic detection from X-ray images utilizing transfer learning with convolutional neural networks, Phys. Eng. Sci. Med. 43 (2) (Jun 2020) 635–640.
[16] Y.-D. Zhang, S.C. Satapathy, L.-Y. Zhu, J.M. Gorriz, S.-H. Wang, A seven-layer convolutional neural network for chest CT based COVID-19 diagnosis using stochastic pooling, IEEE Sensor. J. (2021), 1-1.

[17] H. Benbrahim, H. Hachimi, A. Amine, Deep transfer learning with Apache spark to detect covid-19 in chest x-ray images, Rom. J. Inf. Sci. Technol. 23 (2020) S117–S129. S, SI.
[18] X. Yu, S.H. Wang, Y.D. Zhang, CGNet: a graph-knowledge embedded convolutional neural network for detection of pneumonia, Inf. Process. Manag. 58 (1) (Jan 2021), 102411.
[19] M.U. Rehman, A. Shafique, S. Khalid, M. Driss, S. Rubaiee, Future forecasting of COVID-19: a supervised learning approach, Sensors 21 (10) (May 11 2021).
[20] A. Oulefki, S. Agaian, T. Trongtirakul, A. Kassah Laouar, Automatic COVID-19 lung infected region segmentation and measurement using CT-scans images, Pattern Recogn. 114 (Jun 2021), 107747.
[21] X. Liu, et al., Weakly supervised segmentation of COVID19 infection with scribble annotation on CT images, Pattern Recogn. 122 (Feb 2022), 108341.
[22] J. He, Q. Zhu, K. Zhang, P. Yu, J. Tang, An evolvable adversarial network with gradient penalty for COVID-19 infection segmentation, Appl. Soft Comput. 113 (Dec 2021), 107947.
[23] N. Mu, H. Wang, Y. Zhang, J. Jiang, J. Tang, Progressive global perception and local polishing network for lung infection segmentation of COVID-19 CT images, Pattern Recogn. 120 (Dec 2021), 108168.
[24] C. Tan, F. Sun, T. Kong, W. Zhang, C. Yang, C. Liu, A Survey on Deep Transfer Learning, arXiv, 2018.
[25] J. Deng, ImageNet : a large-scale hierarchical image database, Proc. CVPR (2009), 2009.
[26] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: Presented at the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016.
[27] G. Huang, Z. Liu, V. Laurens, K.Q. Weinberger, Densely Connected Convolutional Networks, IEEE Computer Society, 2016.
[28] M. Raghu, C. Zhang, J. Kleinberg, S. Bengio, Transfusion: understanding transfer learning for medical imaging, Adv. Neural Inf. Process. Syst. 32 (2019).
[29] S. Gidaris, P. Singh, N. Komodakis, Unsupervised Representation Learning by Predicting Image Rotations, 2018.
[30] P. Sermanet, et al., Time-Contrastive Networks, Self-Supervised Learning from Video, 2017.
[31] R.D. Hjelm, et al., Learning Deep Representations by Mutual Information Estimation and Maximization, 2018.
[32] Y. Tian, D. Krishnan, P. Isola, Contrastive Multiview Coding, 2019.
[33] Z. Wu, Y. Xiong, S.X. Yu, D. Lin, Unsupervised Feature Learning via Non-parametric Instance Discrimination, IEEE, 2018.
[34] K. He, H. Fan, Y. Wu, S. Xie, R. Girshick, Momentum contrast for unsupervised visual representation learning, in: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020.
[35] T. Chen, S. Kornblith, M. Norouzi, and G. Hinton, "A Simple Framework for Contrastive Learning of Visual Representations," 2020.
[36] X. Chen, H. Fan, R. Girshick, K. He, Improved Baselines with Momentum Contrastive Learning, 2020.
[37] Soares, et al., SARS-CoV-2 CT-scan dataset: a large dataset of real patients CT scans for SARS-CoV-2 identification" [Online], Available: https://www.kaggle.com/plameneduardo/sarscov2-ctscan-dataset, 2020.
[38] Lung Nodule Analysis, 2016 [Online]. Available: https://luna16.grand-challenge.org/data/.
[39] Maftouni, et al., Robust ensemble-deep learning model for COVID-19 diagnosis based on an integrated CT scan images database [Online]. Available: https://www.kaggle.com/maedemaftouni/large-covid19-ct-slice-dataset, 2021.
[40] M. Raghu, C. Zhang, J. Kleinberg, S. Bengio, Transfusion: Understanding Transfer Learning for Medical Imaging, 2019.
[41] A. Oord, Y. Li, O. Vinyals, Representation Learning with Contrastive Predictive Coding, 2018.
[42] S. Woo, J. Park, J.Y. Lee, I.S. Kweon, CBAM: convolutional block Attention module, in: European Conference on Computer Vision, 2018.
[43] H. Jie, S. Li, S. Gang, S. Albanie, Squeeze-and-Excitation Networks, vol. 99, IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017.
[44] A. Paszke, et al., Pytorch: an imperative style, high-performance deep learning library, Adv. Neural Inf. Process. Syst. 32 (2019).
[45] F. Pedregosa, et al., Scikit-learn: Machine Learning in Python, 2012.
[46] S. Minaee, R. Kafieh, M. Sonka, S. Yazdani, G. Jamalipour Soufi, Deep-COVID: predicting COVID-19 from chest X-ray images using deep transfer learning, Med. Image Anal. 65 (Oct 2020), 101794.
[47] S. Lu, Z. Zhu, J.M. Gorriz, S.H. Wang, Y.D. Zhang, NAGNN: classification of COVID-19 based on neighboring aware representation from deep graph neural network, Int. J. Intell. Syst. (2021).
[48] M. Farooq, A. Hafeez, COVID-ResNet: A Deep Learning Framework for Screening of COVID19 from Radiographs, 2020.
[49] Y. Oh, S. Park, J.C. Ye, Deep learning COVID-19 features on CXR using limited training data sets, IEEE Trans. Med. Imag. 39 (8) (Aug 2020) 2688–2700.