Medical Image Triage and Classification System——Based on COVID-19 CT and X_ray Scan Dataset

Chenglong Duan
SUSTECH
Email: 11912604@mail.sustech.edu.cn

Haoxuan Feng
SUSTECH
Email: 11912205@mail.sustech.edu.cn

Feirui Li
SUSTECH
Email: 11910308@mail.sustech.edu.cn

ABSTRACT

Nowadays, COVID-19 is raging. To cope with the epidemic and improve the accuracy and efficiency of COVID-19 detection, we propose a medical image triage and classification system based on COVID-19 CT and X-ray scan datasets. In our system, before classifying a specific CT dataset for the presence or absence of COVID-19, a two-step ood detection was first completed to filter the specific dataset from the bid datasets which contains other kinds of CT and X_ray images, finally we proposed a classification model based on transfer learning and self-supervised learning for whether a CT image is suffering from COVID-19. The dataset we selected for the final classification is an open sourced dataset——COVID-CT, which contains 349 COVID-19 CT images from 216 patients and 463 NonCOVID-19 CTs. Among our system, The accuracy of two-step OOD detection was 100% for CT and X-ray images, and 81% for different types of CT images. Our final classification model classifies CT images as COVID or NonCOVID based on DenseNet169 with an accuracy of 94.4%. Therefore, our system is suitable for reality, which can help hospitals complete triage, classification and improve the accuracy of detection.

Code:https://github.com/feary-li/Medical-Image-Triage-and-Classification-System-Based-on-COVID-19-CT-and-X-ray-Scan-Dataset.git

Index Terms—— COVID-19, CT, X_ray, ood detection, GLCM, DenseNet169, transfer learning, supervised learning, OOD detection

1. Introduction

In December 2019, a large-scale infection of the COVID-19 occurred in Wuhan, China. Because of the frequent population movement between countries and the negligence of gov- ernments around the world except China, the new coronavirus has spread globally since January 2020.The World Health Organization (WHO) named the infectious disease infected by the coronavirus as Coronavirus Disease 2019 (COVID-19) on January 30, 2020.

When COVID-19 epidemic just broke out, RT-PCR was the only technology to detect and diagnose SARS-CoV-2 from breath samples. However, due to certain technical defects and limitations of this technology, it is not the best detection method. In contrast, chest computed tomography (CT) is a faster and easier clinical diagnosis method for COVID-19, but a large number of CT images have brought a serious burden to professional medical staff, and it is difficult to complete everything by manpower alone.

Therefore, it is necessary to develop a computer-assisted method to automatically detect the new coronary pneumonia virus from lung images. Deep learning technology, as a novel core technology of artificial intelligence in recent years, has a wide range of applications in the automatic analysis of radiological images. Therefore, this artificial intelligence method that can automatically detect COVID-19, when it reaches a certain degree of accuracy, can bring certain help to the work of radiologists and reduce their burden.

In hospitals, there is a mixture of medical images, and those that can be used directly to detect COVID-19 are CT and X_ray, so a system was built that can be used to triage these two types of medical images and to classify CT images. Specifically the classification of CT images is done by first classifying the images taken by different machines, and later by classifying the CT images of a specific category to predict whether the patient is have COVID-19 or not.

This project will use the relevant knowledge of machine learning to build a triage and classification system for medical image of COVID-19, provide help for relevant diagnoses in the medical community, and provide other researchers with references in this regard. The objectives and content of this research are summarized as follows:

- Use SVM method to classify lung CT images and X_ray images to complete the first step of OOD;
- Use Mahalanobis distance and GLCM to construct a comprehensive evaluation system to classify the various types of CT images taken by different machines to complete the second step of OOD;
- Use transfer learning and self-supervision learning to construct a classification model based on DenseNet169 to classify COVID CT images or NonCOVID CT images;
- The medical image triage and classification system is built by combining the above three parts, which can meet the needs of hospitals for automatic recognition of CT images and help hospitals to carry out diagnostic work more efficiently.

2. Related Works

Many research works have been intensively and rapidly conducted on developing some deep learning methods in responding to COVID-19 global pandemic.We here briefly review deep learning approaches for the task of image-level classification for diagnosis which are closely relevant to our article.

2.1. Classification of CT and X_ray

Due to the differences between CT images and X_ray images, there are basically no deep learning methods to classify these two kinds of images. Therefore, we have looked for specific classification methods for other types of images, and we can refer to these methods for this kind of classification of CT and X_ray. Abien Fred et al.[1] proposed an image classifier combining CNN and SVM methods, the test accuracy on mnist dataset is 99.23%. Also Yiqing Guo et al.[2] proposed a SVM-based sequential classifier training (SCT-SVM) approach for multitemporal remote sensing image classification, the overall classification accuracy of the incoming image was improved from 76.18% to 94.02% with the proposed SCT-SVM, compared with those obtained without the assistance from previous images. Yijie Dang et al.[3] proposed a method based on quantum K-Nearest-Neighbor algorithm, the classification accuracy is 83.1% on Graz-01 dataset and 78% on Caltech-101 dataset, this method has a good classification performance while greatly improving the efficiency.

2.2. Out-of-distribution(OOD) Detection

There are some methods that show good OOD detection performance in classification models. Lee et al. [15] propose that estimated Gaussian distribution from in-distribution(ID) samples and calculated uncertainty using the Mahalanobis distance. Lee et al.[14] focus on KL divergence of softmaxed output from the uniform distribution. Other method[4] use a pre-trained nnU-Net models to evaluate the suitability to new data, this method exploits the Mahalanobis distance in the feature space.

2.3. Usage of the COVID-19 CT Datesets

More people are using this open source dataset when developing deep learning medical image processing. Yang et al.[5] were the first to propose and built this open source dataset, and it was endorsed by radiologists. They obtained a prediction accuracy of 85% with the Densenet169 network model based on this dataset after multi-task learning and self-supervised learning. Nicolas Ewen et al.[6] used all the data in this dataset and added some additional data to obtained an accuracy rate of 86.21% through self-supervised learning combined with Densenet network structure model. Wang et al.[7] designed a COVID-19 CT classification network with an accuracy of 76.6% through comparative cross-site learning.

2.4. Data Augmentation

Before performing the actual classification, data augmentation operation are required on the dataset due to the limitation of the number of medical images. Singh et al.[18] improves Progressive Growing of GANs with a structural similarity loss function (PGGAN-SSIM) to solve image blurriness problems and model collapse, successfully generates 256x256 realistic brain tumor MR images which fill the real image distribution uncovered by the original dataset. Arjun Krishna et al.[19] proposed a deep reinforcement learning (DRL) approach that is integrated with a style-transfer (ST) methodology to generate high resolution CT images to complete data augmentation.

2.5. Self Supervised Learning

We examine supervised learning methods for image processing. He et al.[8] proposed a method of deep sample-based coloring through self supervised learning. Su et al.[9] proposed a method to perceive image coloring based on instances through self supervised learning. Aaron van den Oord et al.[10] proposed a definition of a comparative self supervised learning loss function.

3. Methods

3.1. Classification of CT and X_ray

Based on the differences in the features of the two images, some basic classification methods can be used to classify them, so the methods available in sklearn for image classification are used, and the following is a brief description of the specific methods:

- Logistic Regression
  Despite the name 'regression', logistic regression is actually a classification method, mainly used in two classification problems (as the output has only two kinds, representing two categories). The essence of logistic regression is that the data is assumed to obey this distribution, and then the parameters are estimated using the maximum likelihood estimation.
- K-Near-Neighbor
  The main idea of the algorithm is to determine the category to which one belongs based on the neighboring categories that are close to one another. The specific calculation steps are divided into three steps: calculate the distance between the test object and all the objects in the training set; find the nearest K objects in the distance calculated in the previous step as the test object's neighbors; find the object with the highest frequency among the K objects, and the category to which the test object belongs is the category to which the test object belongs.
- Support Vector Machine
  In layman's terms, SVM is a two-class classification model whose basic model is defined as a linear classifier with maximum interval on the feature space, and whose learning strategy is interval maximization, which can eventually be translated into

the solution of a convex quadratic programming problem.

3.2. Detection of Specific CT images

We propose a method for OOD detection that is a comprehensive evaluation system. All CT images we use are gray scale images, and the gray-level co-occurrence matrix(GLCM) can extract the texture features of gray scale images. GLCM reflects the brightness characteristics of pixels, it is also a second-order statistical feature about the image brightness variation. Hence, the input CT samples can be preliminary screening detected based on the statistics of the GLCM. In this step, the system will exclude the CT input images that do not belong to ID and flow the remaining images to the next detector. Inspired by Camlia et al.[4] propose a strategy to detect the CT samples which from the first step. This strategy will passes the images through a pre-trained network first, VGGNet16 is chosen here. Then estimates the Gaussian distribution of the obtained features, when predict the class of the CT samples, the KL divergence is used to obtain the classification.

The whole comprehensive evaluation system needs to be trained first, but the VGGNet network in the second step is a pre-trained model. The details are shown as follows.

The feature of GLCM

In the first step of the GLCM, we used its six statistics as our features used for classification, which are contrast, dissimilarity, homogeneity, angular second moment, energy and correlation. Contrast reflects the sharpness of the image and the depth of the texture grooves.

$$Con = \sum_i \sum_j (i-j)^2 P(i,j)$$

Angular second moment is a measure of the uniformity of the image gray distribution and the thickness of the texture.

$$ASM = \sum_i \sum_j P(i,j)^2$$

The remaining concentration statistics represent different features such as the consistency of the image texture, respectively.

Image similarity comparison

In training, training images belonging to the same class are grouped into the same multivariate Gaussian distribution, assuming that the data coming from the pre-trained VGGNet16 is z_i, then we can then estimate its Gaussian distribution as:

$$\mu = \frac{1}{N}\sum_{i=1}^{N} \hat{z}_i \quad \Sigma = \frac{1}{N}\sum_{i=1}^{N}(\hat{z}_i-\mu)(\hat{z}_i-\mu)^T$$

During inference, the Mahalanobis distance can be calculated from the covariance and mean of the obtained Gaussian distribution:

$$D_M(\hat{z}_i;\mu,\Sigma) = (\hat{z}_i-\mu)^T\Sigma^{-1}(\hat{z}_i-\mu)$$

Each distance is a point estimate for the corresponding model input. The samples in test set will be compare by the train samples through KL divergence.

3.3. Classification of CT Images into COVID or NonCOVID

3.3.1. Data Augmentation

Data augmentation is an effective way to expand the size of data samples. Deep learning is a method based on big data, and we currently hope that the larger and higher quality the data is, the better the model can have better generalization ability. Especially for medical images, where the size of the data set itself is small, it is necessary to enhance the data set before the actual training to improve the classification accuracy of the model.

In this paper, we mainly use the data augmentation methods that come with pytorch and our own data augmentation methods built on the pytorch framework.

Methods in pytorch

- Application class
  Not transformation of the image itself, but processing for the transformation operation itself.
- Basic transformation class
  Perform basic essential transformations.
- Selection class
  The image itself is not transformed, only a different selection of the original image content is made.
- Complex transformation class
  Includes whitening transformation, H and S channel perturbation in HSV color space, luminance contrast perturbation, affine transformation, perspective transformation.

Other methods based on pytorch

- HEDJitter
  Using the color deconvolution method, the HE image is transformed from RGB space to HED space, then random perturbations are added for each channel in HED space, and finally transformed to RGB space again. HED spatial channel perturbation strategy is:

  $$s' = \alpha * s + \beta$$

- RandomElastic
  Perturbation deviation of x and y for each pixel point separately.
- RandomAffineCV2
  The affine transformation is essentially the same as RandomAffine in torchvision, except that the implementation is the difference between PIL.Image and cv2. Using the opencv method, the transformed content-free boundary region can be mirrored and filled.
- RandomGaussBlur

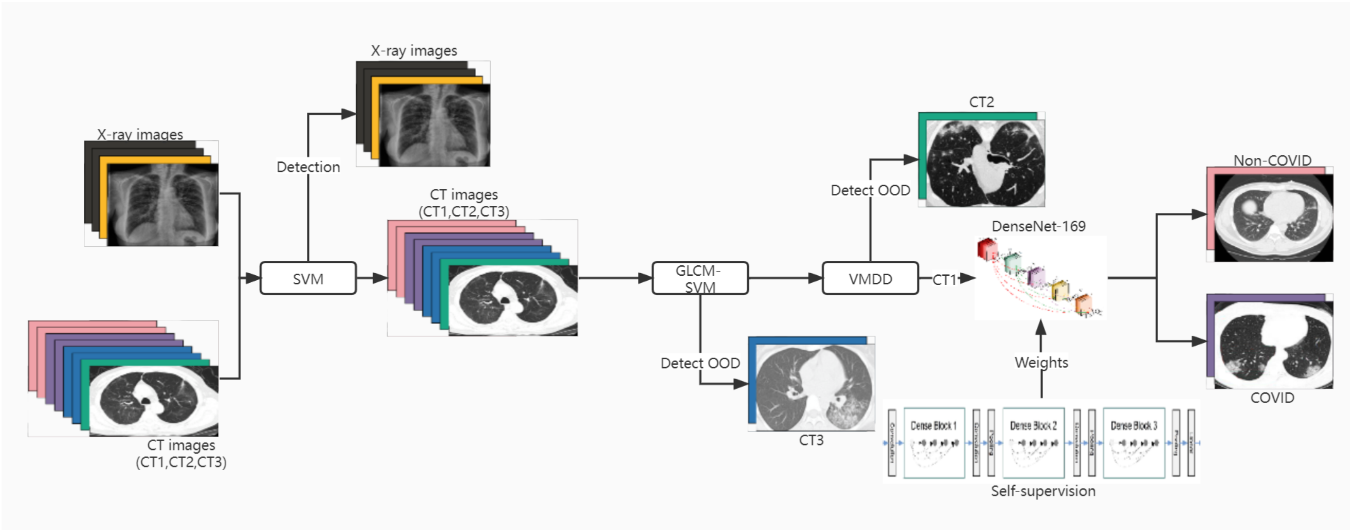


Figure 1: The whole workflow of our system

Based on Gaussian filtering method, the image is blurred to different degrees, which is suitable for the case that some areas in the pathology full-scan section images are not focused enough.

3.3.2. Transfer Learning

Transfer learning focuses on storing a solution model of an existing problem and leveraging it on other different but related problems. It is to use a pre-trained image classification model and transform it into a target detection model, or a key point regression model. Due to the small amount of data in the medical image CT domain, we use transfer learning to transfer the knowledge learned in the case of sufficient data volume to an environment with small data volume. If we train directly with these small amount of images without transfer learning, the results will not achieve the desired accuracy and may cause overfitting. Therefore, using ImageNet pretrained models for transfer learning is important to improve the generalization ability and accuracy of the models. It can also reduce the training time.

3.3.3. Self-supervised Learning

Since to apply supervised learning in deep neural networks, we need enough labeled data. But manually labeling data manually is time-consuming and expensive. Getting enough data for the medical field is a challenge in itself. The supervised information for self-supervised learning is not manually labeled, but the algorithm automatically constructs supervised information in large-scale unsupervised data to perform supervised learning or training.

Therefore, we apply and optimize a contrast-based self-supervised learning approach for lung CT medical images: we use a pretext task to self-supervised information of CT images from large-scale unsupervised data, and train the network with this constructed supervised information so that it can learn a valuable representation for downstream tasks(classification COVID from NonCOVID) valuable representations. It does this by combining: the location size of diseased CT features such as bilateral pulmonary atelectasis, with diffuse ground glass-like atelectasis, interstitial thickening, ground glass-like lesions, interstitial septal thickening and peripheral fibrosis-like changes, and we set some kinds of auxiliary tasks to identify the CT images. It pre-trains the network based on the contrast constraint by learning to encode the similarity or dissimilarity of two things to construct representations to complete the self-supervised task. To achieve the pretext task, the network needs to learn to understand the lungs. In turn, pre-training is completed. We then transfer the parameter weights of the self-supervised network into the training deep neural network model. We refer to the latest loss function based on contrastive self-supervised learning, using the loss function InfoNCE, into a better completion of contrastive self-supervised learning.

4. Experiments

4.1. Classification of CT and X_ray(first OOD)

Dataset

We extracted X-ray dataset from COVID-19 related medical image dataset[12], also extracted CT dataset from other assays' datasets[5][16][17].

	CT	X_ray
train	810	324
test	184	10

(a) CT sample (b) X_ray sample

Experimental procedure

Logistic regression(LR), SVM, and KNN methods were used to classify these two types of images, and the parameters of the specific methods were adjusted to obtain the following experimental results.

	running time(s)	accuracy(%)
LR(liblinear)	0.275	74.312
LR(newton-cg)	0.364	96.330
LR(lbfgs)	0.325	96.330
LR(sag)	13.108	95.413
LR(liblinear,regularized in L1)	0.062	91.743
SVM(rbf)	0.067	99.541
SVM(linear)	0.038	93.578
SVM(poly)	0.146	77.523
SVM(sigmoid)	0.066	79.358
KNN(k=2)	0.014	93.119

Based on the above results, the LR and KNN methods were basically excluded, after which a network search was performed for SVM(poly) and SVM(rbf) with relevant parameters, and the final results are as follows.

	running time(s)	accuracy(%)
SVM(rbf, gramma=auto, C=0.1)	0.164	100
SVM(poly, degree=2, C=0.1)	0.207	99.083

After using the optimal SVM classification method to remove all X_ray images, complete the first ood step.

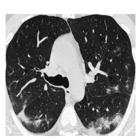
4.2. Detection of Specific CT images(second OOD)

Dataset

We extracted three different kinds of CT images from three different hospitals and taken by different kinds of CT machines. The target CT images are called CT1 is taken from Xingyi et al.[5] have 246 images consist of both COVID-19 and NonCOVID-19. We add other kind of CT images, the set called CT2 from Laradji et al.[16] has 98 images and the set called CT3 from Maftouni et al.[17] has 190 images .

	CT1	CT2	CT3
train	150	50	150
test	96	48	40

(c) CT1 sample



(d) CT2 sample

(e) CT3 sample

Experimental setup

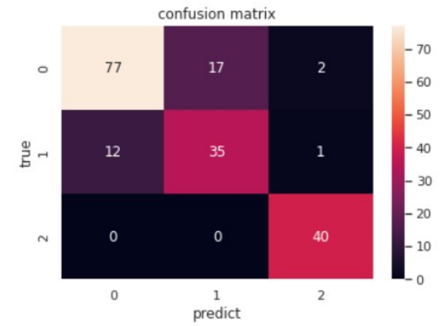
In the first step of classification detection with GLCM statistics as features, the parameters of the GLCM are chosen as the level is 256, the displacement offset is 5, and the angle is chosen as 0 degrees.

In detection classification, we used support vector machines as classifiers, where the kernel function is a polynomial sum function with a maximum number of terms of 2, $\gamma = 0.02$, $C = 10$.

Experimental procedure

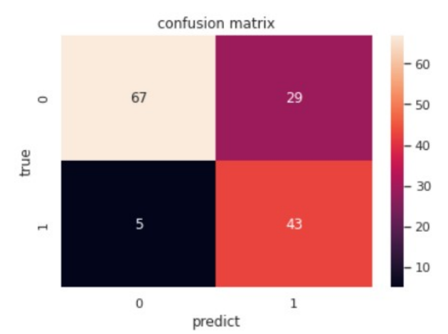
In the classification detection of GLCM, we output the confusion matrix and the accuracy matrix as follows.

	precision	recall	f1-score	support
0.0	0.87	0.80	0.83	96
1.0	0.67	0.73	0.70	48
2.0	0.93	1.00	0.96	40
accuracy			0.83	184
macro avg	0.82	0.84	0.83	184
weighted avg	0.83	0.83	0.83	184



As can be seen from the results, at this point, the model can completely detect the CT3 images that are not suitable for use in the deep learning model, with an overall accuracy of 83%, while the remaining CT1 and CT2 images will then go into the image features obtained by VGGNet, using the features to estimate the Gaussian distribution of the features, based on the Gaussian distribution of the Mahalanobis distance, and finally the classification detection based on the test images and the KL scatter of the Gaussian distribution.

	precision	recall	f1-score	support
1.0	0.93	0.70	0.80	96
2.0	0.60	0.90	0.72	48
accuracy			0.76	144
macro avg	0.76	0.80	0.76	144
weighted avg	0.82	0.76	0.77	144



As can be seen from the results, in the second step of the comprehensive evaluation system, the accuracy of the classification is 76%, but it is possible to completely detect the CT2 images that are not applicable, even if some CT1 images are considered inappropriate.

The overall accuracy of the comprehensive evaluation system ended up being 81%, which was calculated by the following formula.

$$Accuracy = \frac{CP}{P}$$

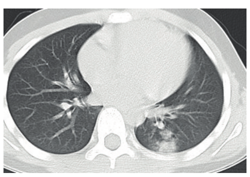
Where CP represents the correct prediction and P represents the whole test samples.

4.3. Classification of CT Images into COVID or NonCOVID

Dataset

The data we final used is called COVID-CT-Dataset[5], we used all the images in this dataset and divided them into train set, validation set and test set.

	COVID	NonCOVID
Train	250	250
Validation	50	50
Test	45	45

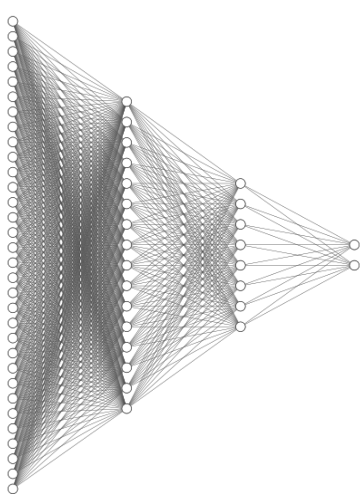


(f) COVID CT sample

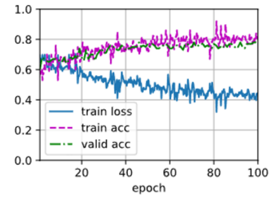
(g) NonCOVID sample

Experimental procedure

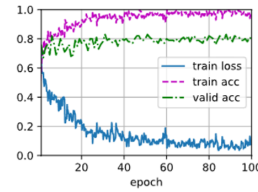
The network we used for final classification is DenseNet169, while retaining most of the network structure, only the final classifier layer has been changed as follows: add multiple linear layers, using dropout and relu activation functions in between, the final number of classification output is 2.



First of all, verify the effectiveness of the pre-trained model for our used dataset: load the pre-trained model in pytorch, the model is trained by ImageNet.



(h) Unload



(i) Load

From the initial point position of the training and the size of the final val accuracy achieved, it can be seen that loading the pre-trained model can significantly improve the fitting ability of the model and can reach the optimal value of the model more quickly.

Secondly, we use the self-supervised learning to obtain pre-trained model to fine-tuning of ImageNet pre-trained models, use the initial images and the transformed images to do experiment, a total of four experiments were set up(all transformations of the image do not change the image itself, so there is no loss of information in the image).

1) EX1:
   Flipped horizontally
2) EX2:
   Flipped vertically
3) EX3:
   Rotated 30 degrees clockwise
4) EX4:
   Rotated 30 degrees counterclockwise

Other experimental settings of these four experiments are the same: train set's size is 168, the val set's size is 24; the size of the dynamic dictionary was set to 224 $\times$ 224; do normalized operations for all images; stochastic gradient descent(SGD) was used as the optimizer with a minibatch size of 10, a weight decay of 0.0001, a momentum of 0.9, and an initial learning rate of 1e-4.

	train accuracy	val accuracy
EX1	72.67%	79.12%
EX2	70.67%	75%
EX3	69.56%	76.8%
EX4	60.41%	65.29%

Afterwards, the obtained network parameters were substituted into the actual final classification model, and the following results were obtained:

	train accuracy	val accuracy	test accuracy
EX1	51.2%	48%	56.25
EX2	51.4%	49%	54.17
EX3	52%	46%	55.21
EX4	51.6%	51%	55.31

Combining the above results, it can be seen that the effect of image classification is best when the pretext task is to determine whether the CT image is horizontally flipped. However, if the obtained weights and parameters are directly substituted into the actual classification model, the accuracy of each dataset will be low, which is of course obvious, so it is necessary to retrain in the new dataset and fine-tune the parameters to finally get a better classification result.

Then, experiments related to the data augmentation methods for the actual classification model were conducted in the hope of obtaining the optimal combination of data augmentation methods(all experiments were performed with parameters load after self-supervised learning).

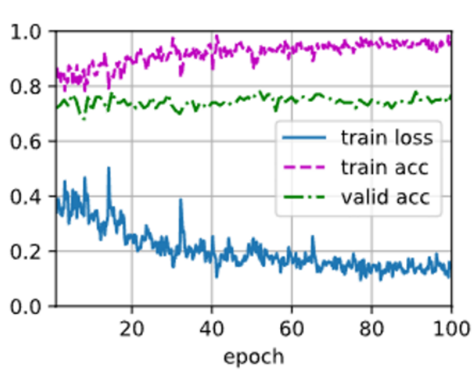
1) EX1:
   Flipped horizontally
2) EX2:
   Flipped horizontally + Flipped vertically
3) EX3:
   Flipped horizontally + Flipped vertically + HED-Jitter(theta=0.05)
4) EX4:
   Flipped horizontally + Flipped vertically + RandomAffine(degrees=0, shear=(0, 45))
5) EX5:
   Flipped horizontally + Flipped vertically + HED-Jitter(theta=0.05) + + RandomAffine(degrees=0, shear=(0, 45))

Other experimental settings of these four experiments are the same: train set's size is 500, the val set's size is 100; the test set's size is 90; the size of the dynamic dictionary was set to 224 × 224; do normalized operations for all images; stochastic gradient descent(SGD) was used as the optimizer with a minibatch size of 10, a weight decay of 0.0001, a momentum of 0.9, and an initial learning rate of 1e-4. At the same time, the new network parameters will be trained again while doing the transfer learning by reading the previous self-supervised learning's parameters, and the final classification will be done accordingly.

	train accuracy	val accuracy	test accuracy
EX1	96.4%	78%	64.44%
EX2	98%	84%	87.78%
EX3	96.2%	77%	65.56%
EX4	95.6%	81%	75.56%
EX5	92%	81%	66.67%

Combining the above experimental results, a combination of horizontal flip and vertical flip was chosen as the final data augmentation method. Afterwards, several times of self-supervised learning and secondary training were performed, and finally the optimal network structure parameters were optimized, and the optimal classification results were obtained.

The figure below shows training process:



The figure below shows the best accuracy obtained by the best params:

```
Train Accuracy:  0.996
Val Accuracy:    0.85
Test Accuracy:   0.944
```

The figure below shows the other indicators:

	precision	recall	f1-score	support
0	0.98	0.91	0.94	45
1	0.92	0.98	0.95	45
accuracy			0.94	90
macro avg	0.95	0.94	0.94	90
weighted avg	0.95	0.94	0.94	90

5. Results analysis

Since no similar paper was found about the OOD section using the same dataset, comparison and variance analysis with other results is not possible. So in this section only the DenseNet model we finally constructed can be compared with other papers' results using the same dataset.

	accuracy	f1-Score	recall	precision
Densenet169[6]	0.8621	0.8704	/	/
ViT[20]	0.776	0.76	/	/
COVID-Net[7]	0.7869	0.8378	0.7971	0.7802
Densenet169[5]	0.891	0.896	/	/
Densenet169(Ours)	0.944	0.94	0.945	0.94

Through comparison, we found that our accuracy is better than other papers. We analyze this as follows: we reasonably use transfer learning to pre-train the network structure to improve accuracy and prevent overfitting, we have image pre-processing and data augmentation methods to reasonably expand our dataset, we also use contrastive based self-supervised learning to improve our accuracy, and we adjust the network structure to find the optimal parameters. These steps and methods make us superior to other papers.

6. Conclusion

In this paper, we constructed a medical image triage and classification system based on COVID-19 CT and X_ray scan datasets. Among our system, two-step ood detection classifies CT and X_ray images with 100% accuracy and different types of CT images with 81% accuracy; our final classification model based on DenseNet169 classifies of CT Images into COVID or NonCOVID with 94.4% accuracy. Therefore, the system we built can be applied in real-life situations to serve the hospital's diagnostic system and improve the efficiency of the hospital.

References

[1] Agarap, Abien Fred. (2017). An Architecture Combining Convolutional Neural Network (CNN) and Support Vector Machine (SVM) for Image Classification.

[2] Guo Y, Jia X, Paull D. Effective Sequential Classifier Training for SVM-Based Multitemporal Remote Sensing Image Classification. IEEE Trans Image Process. 2018 Feb 22. doi: 10.1109/TIP.2018.2808767. Epub ahead of print. PMID: 29994442.

[3] Dang, Y., Jiang, N., Hu, H., Ji, Z., Zhang, W. (2018). Image Classification Based on Quantum KNN Algorithm. ArXiv, abs/1805.06260.

[4] Camila Gonzalez,Karol Gotkowski, Andreas Bucher, Ricarda Fischbach, Isabel Kaltenborn, Anirban Mukhopadhyay. Detecting when pre-trained nnU-Net models fail silently for Covid-19 lung lesion segmentation. arXiv preprint arXiv:2107.05975v2, 2021.

[5] Xingyi Yang, Xuehai He, Jinyu Zhao, Yichen Zhang, Shanghang Zhang, Pengtao Xie. COVID-CT-Dataset: A CT Scan Dataset about COVID-19. arXiv preprint arXiv:2003.13865v3, 2020.

[6] Nicolas Ewen, Naimul Khan. Targeted Self Supervision for Classification on a Small COVID-19 CT Scan Dataset. arXiv preprint arXiv:2011.10188v1, 2020.

[7] Wang Z, Liu Q, Dou Q. Contrastive Cross-Site Learning With Redesigned Net for COVID-19 CT Classification. IEEE J Biomed Health Inform. 2020 Oct;24(10):2806-2813. doi: 10.1109/JBHI.2020.3023246. Epub 2020 Sep 10. PMID: 32915751.

[8] Mingming He, Dongdong Chen, Jing Liao, Pedro V. Sander, Lu Yuan. Deep Exemplar-based Colorization. arXiv preprint arXiv:1807.06587v2, 2018.

[9] Jheng-Wei Su, Hung-Kuo Chu, Jia-Bin Huang. Instance-aware Image Colorization. arXiv preprint arXiv:2005.10825v1, 2020.

[10] Aaron van den Oord, Yazhe Li, Oriol Vinyals. Representation Learning with Contrastive Predictive Coding. arXiv preprint arXiv:1807.03748v2, 2019.

[11] Kendall, A., Gal, Y.: What uncertainties do we need in bayesian deep learning for computer vision? Advances in Neural Information Processing Systems 30, 5574– 5584 (2017) 14

[12] COVID-19 Image Data Collection: Prospective Predictions Are the Future Joseph Paul Cohen and Paul Morrison and Lan Dao and Karsten Roth and Tim Q Duong and Marzyeh Ghassemi arXiv:2006.11988, https://github.com/ieee8023/covid-chestxray-dataset, 2020

[13] Liang, S., Li, Y., Srikant, R.: Enhancing the reliability of out-of-distribution image detection in neural networks. In: International Conference on Learning Representations (2018)

[14] Lee, K., Lee, H., Lee, K., Shin, J.: Training condence-calibrated classiers foe detecting out-of-distribution samples. In: International Conference on Learning Representations (2018)

[15] Lee, K., Lee, K., Lee, H., Shin, J.: A simple unied framework for detecting out-of distribution samples and adversarial attacks. In: Advances in Neural Information Processing Systems. pp. 7167-7177 (2018)

[16] Laradji, Issam Rodriguez, Pau Manas, Oscar Lensink, Keegan Law, Marco Kurzman, Lironne Parker, William Vázquez, David Nowrouzezahrai, Derek. (2021). A Weakly Supervised Consistency-based Learning Method for COVID-19 Segmentation in CT Images. 2452-2461. 10.1109/WACV48630.2021.00250.

[17] Maftouni, M., Law, A.C, Shen, B., Zhou, Y., Yazdi, N., and Kong, Z.J. "A Robust Ensemble-Deep Learning Model for COVID-19 Diagnosis based on an Integrated CT Scan Images Database," Proceedings of the 2021 Industrial and Systems Engineering Conference, Virtual Conference, May 22-25, 2021.

[18] Singh, Nripendra Raza, Khalid. (2020). Medical Image Generation using Generative Adversarial Networks.

[19] Krishna, Arjun Bartake, Kedar Niu, Chuang Wang, Ge Lai, Youfang Jia, Xun Mueller, Klaus. (2021). Image Synthesis for Data Augmentation in Medical CT using DeepReinforcement Learning.

[20] Xiaohong Gao, Yu Qian, Alice Gao. COVID-VIT: Classification of COVID-19 from CT chest images based on vision transformer models. arXiv preprint.arXiv:2107.01682, 2021.