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.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.
Lung image segmentation based on DRD U-Net and combined WGAN with Deep Neural Network

Luoyu Lian\textsuperscript{a,1,}\textsuperscript{*}, Xin Luo\textsuperscript{b,1}, Canyu Pan\textsuperscript{c}, Jinlong Huang\textsuperscript{a}, Wenshan Hong\textsuperscript{a}, Zhendong Xu\textsuperscript{a,1}

\textsuperscript{a} Department of Thoracic Surgery, Quanzhou First Hospital Affiliated to Fujian Medical University, 248-252 East Street, Licheng District, Quanzhou, Fujian 362000, China

\textsuperscript{b} Department of Cardiac and Thoracic Surgery, Longyan First Affiliated Hospital of Fujian Medical University, Longyan, Fujian 364000, China

\textsuperscript{c} Department of Medical Imaging, Quanzhou First Hospital Affiliated to Fujian Medical University, Quanzhou, Fujian 362000, China

\textbf{Article history:}
Received 12 April 2022
Revised 13 August 2022
Accepted 26 August 2022

\textbf{Keywords:}
Deep learning
DRD U-Net segmentation
Medical image segmentation
WGAN classification

\textbf{ABSTRACT}

\textbf{Purpose:} COVID-19 is a hot issue right now, and it's causing a huge number of infections in people, posing a grave threat to human life. Deep learning-based image diagnostic technology can effectively enhance the deficiencies of the current main detection method. This paper proposes a multi-classification model diagnosis based on segmentation and classification multi-task.

\textbf{Method:} In the segmentation task, the end-to-end DRD U-Net model is used to segment the lung lesions to improve the ability of feature reuse and target segmentation. In the classification task, the model combined with WGAN and Deep Neural Network classifier is used to effectively solve the problem of multi-classification of COVID-19 images with small samples, to achieve the goal of effectively distinguishing COVID-19 patients, other pneumonia patients, and normal subjects.

\textbf{Results:} Experiments are carried out on common X-ray image and CT image data sets. The results display that in the segmentation task, the model is optimal in the key indicators of DSC and HD, and the error is increased by 0.33\% and reduced by 3.57 mm compared with the original network U-Net. In the classification task, compared with SMOTE oversampling method, accuracy increased from 65.32\% to 73.84\%, F-measure increased from 67.65\% to 74.65\%, G-mean increased from 66.52\% to 74.37\%. At the same time, compared with other classical multi-task models, the results also have some advantages.

\textbf{Conclusion:} This study provides new possibilities for COVID-19 image diagnosis methods, improves the accuracy of diagnosis, and hopes to provide substantial help for COVID-19 diagnosis.

© 2022 Elsevier B.V. All rights reserved.

1. Introduction

Nowadays, the COVID-19 pandemic is sweeping the globe, taking a huge toll on the property and even lives in countries around the world. Currently, the main detection method, reverse transcription-polymerase reaction (RT-PCR), has many problems, such as time-consuming, high false-negative rate, insufficient sensitivity [1–3], and limited medical equipment. The rapid spread and high infection rates of COVID-19 make it urgent to tackle the problem. At the same time, imaging features (in X-ray and CT images) are listed as one of the three major clinical features of suspected COVID-19 cases [4], indicating that pulmonary imaging results can also be used as diagnostic criteria for COVID-19 cases. Recently, with the extremely rapid progress and rise of Artificial Intelligence (AI) technology and computer-aided diagnosis (CAD) technology [37,38], the related medical field is more and more widely used to improve diagnosis efficiency and speed up diagnosis. Among them, Convolutional Neural Network (CNN) shows excellent performance in the diagnosis of pulmonary diseases [5,6]. Therefore, the application of deep learning technology for image diagnosis of COVID-19 [39] has high research value and significance.

At present, as more COVID-19 diagnostic models based on deep learning are proposed, the diagnostic performance is improving. However, there are very few large-scale COVID-19 image data sets available in open source, and insufficient training data has become a major limitation of model training. To solve the problem of insufficient training samples, relevant studies in recent years focus on oversampling technology, which can solve the problem of data imbalance by adding oversampling to a small number of existing samples to generate simulation samples. However, data sets in many fields, including COVID-19 diagnosis, are not faced with data...
imbalance, but with small sample sizes, which makes it difficult to conduct training effectively [7].

With the further progress of Artificial Intelligence technology, a generative adversarial network (GAN) based on a Deep Neural Network (DNN) is proposed. It offers a whole new approach to solve the multi-classification problem of small samples represented by COVID-19 image diagnosis. At present, GAN is one of the most popular unsupervised deep learning models. It was proposed by Goodfellow et al. [8] in Generative Adversarial Networks. The framework trains both generative model G and discriminant model D. The function of the generation model is to capture data distribution, and the function of the discriminant model D is to estimate the probability of samples coming from training data. The game learning between the two models in the framework produces the desired output [9]. For the problem of insufficient labeled samples, GAN expands the size of labeled samples by generating simulated samples with the same distribution as real samples. Douzas and Bacao [10], respectively applied GAN and a variety of oversampling methods to generate simulated samples and compared them on different data sets. The results displayed that GAN was more efficient. Currently, computer-aided diagnostics for COVID-19 typically use deep learning or imagemonics. However, the training results of the deep learning method are not highly interpretable. The subjective bias of lesions in imaging is large, and the data with lesions labeling is relatively scarce and requires manual labeling.

Based on the above research status, this paper proposes a multi-classification model for COVID-19 diagnosis based on segmentation and classification multi-task. First, according to the imaging characteristics of COVID-19 patients in different periods, the lung lesion region was segmented through the end-to-end DRD U-Net (Dilated Residual and Deeply Supervised U-Net) segmentation task to extract features more efficiently. Using a classification task combining WGAN and DNN classifiers to learn the unique imaging features of COVID-19 patients, we can effectively distinguish COVID-19 patients, other pneumonia patients, and normal subjects. We hope to use this research to help substantively diagnose COVID-19.

2. Methods

2.1. Images of pneumonia

Firstly, we will introduce the lung images of various pneumonia patients and ordinary people obtained via X-ray and CT scanners in our hospital. We collected lung X-ray and CT images from 1034 normal subjects, 889 patients with bacterial pneumonia, 769 patients with viral pneumonia, and 245 patients with COVID-19. These images were used with the consent of the patient and the hospital, and the data sets were collected and manually annotated by professional imaging physicians. Ethics approval is granted by the institutional review board. Specific images and image analysis are shown below.

Fig. 1 shows X-ray and CT images of normal subjects, patients with bacterial pneumonia, patients with viral pneumonia, and patients with COVID-19, respectively. In X-ray images, COVID-19 lesions appear as single or multiple focal small patchy shadows with blurred edges. In contrast, other cases of pneumonia are characterized by large patchy and patchy fusion images, often distributed diffusely in both lungs. CT images can overcome the problem of X-ray image overlap and improve the detection rate of lesions. In CT images, the lesions of COVID-19 appear as small amounts of simple ground glass. However, CT images of common pneumonia are dominated by the shadows of consolidation of large lobes.

Fig. 2 shows CT images of normal subjects, patients with early COVID-19, and patients with severe COVID-19. In CT images, the early CT findings of COVID-19 show interstitial changes and multiple small patch shadows, with obvious peripheral lung and a small amount of simple ground glass shadow. During the severe late stage, it develops into infiltrating shadows and multiple ground-glass shadows of both lungs, and consolidation of lungs can occur in severe cases.

2.2. Multi-task learning algorithm

Currently, most machine learning uses single-task learning. When faced with some complex learning problems, single-task learning generally decomposes the problem into simple and independent sub-problems, solves these sub-problems first, and then merges the results to obtain the results of the initial complex problem [11]. The premise of doing this is that the sub-problems are independent of each other, so the correlation information between the problems is not noticed, which weakens the generalization effect of the model. Multi-task learning is a method of derivation and transfers learning, which improves the generalization effect of the main task by using training signals of related tasks [12].

This study used a multi-classification model for COVID-19 diagnosis based on segmentation and categorization multi-tasking (a single task only has classification or segmentation tasks). The classification task first extracted the whole lung region and used the network model combining WGAN and DNN classifiers. Diagnosis of COVID-19 can be achieved by learning the unique imaging features of COVID-19 patients to effectively distinguish between COVID-19 patients, other pneumonia patients, and normal subjects. The segmentation task further segmented the lung lesion region based on the pretreatment (extraction of the entire lung region), thus enabling more effective feature extraction and improving the diagnostic effect of COVID-19.

2.3. Segmentation task model

The segmentation task should be realized according to the imaging features of COVID-19 patients at different periods. Therefore, a large number of medical images with lesion labeling are required, but the data set with lesion labeling is significantly smaller than that without lesion labeling. This feature of the data set further confirms the necessity of using the combination of classification tasks and segmentation tasks to get a multi-task processing model. The brief processing flow of the segmentation task based on deep learning is shown in Fig. 3.

Ronneberger et al. [13] come up with the U-net network model in 2015, which is mainly used in the field of biomedical image segmentation due to its characteristics of small segmentation, good structure scalability, and good target effect. In this study, the end-to-end DRD U-Net network model is adopted. The following details the model and its key design elements.

2.3.1. U-Net model

U-Net model mainly includes an encoder, bottleneck module, and decoder. U-shaped network model structure enables it to speed up training by combining contextual information in the case of fewer data, which happens to coincide with the requirements of medical image segmentation technology, so it can be widely applied in the field of biomedical image segmentation [14].

The specific network structure of the U-Net model is as follows. Firstly, the encoder is composed of four parts, each of which contains two 3 × 3 cascaded convolutional layers, one maximum pooling layer, and a double channel number, which greatly reduces the number of arguments and image resolution, and meanwhile, more high-level semantic (pixel) features can be obtained. The decoder part corresponds to the encoder part one by one, but the difference is that the number of channels is doubled to halved, and the resolution of the image is restored through deconvolution. Fig. 4 shows
Fig. 1. Classification of lung images based on (a) Sample lung X-ray images and (b) Sample lung CT images.

Fig. 2. Sample lung CT images based on (a) Normal, (b) Early COVID-19 and (c) Severe COVID-19.

Fig. 3. Processing flow of segmentation task based on deep learning.
the basic model structure of the U-Net network. Although U-Net has achieved some good results, there are still areas for improvement. Because the structures of different organs are very different and the shapes of lesions are also diverse, it is difficult to meet the requirements of segmentation efficiency and precision only by using the basic U-Net network model to segment images. Therefore, the U-Net network model should be improved according to the specific situation.

2.3.2. DRD U-Net model

The DRD U-Net model used in this paper is an end-to-end network model directly constructed for medical images, and the following three aspects are improved on the basis of the traditional U-Net model. Firstly, residual modules are added to each layer of channel splitting of input images to ensure fast convergence and reduce the size of input signals in the following network layer while expanding the range of feature receptive fields. At the same time, extended convolution is used to expand the receptive field without changing the number of parameters, so as to obtain richer context feature information. In addition, multi-scale context feature extraction modules are used according to the size of segmentation targets. In the multi-output feature fusion module, to enhance the reuse capability of features of different depths in the network, the module of deep supervision [15] is used. Thus improving the accuracy of segmentation.

2.3.3. Residual module

Generally speaking, the expression ability and training effect of a neural network can be improved by increasing network width and depth. However, if we simply increase the network depth, we may encounter the problem of gradient extinction or gradient explosion. He et al. come up with a residual network for this reason [16]. The residual idea adds a hop link to every two layers of the network. The output of the residual block is obtained by the residual path directly superimposed on the input of the residual block, and the residual block learns and fits the residual to ensure that the network still has good expression ability with the increase of the network depth [17].

By integrating this residual idea into U-Net, the output convolved by the residual path is directly superimposed with the input of the convolution layer. To realize the purpose of extracting spatial features of different scales, making up the semantic gap between encoder and decoder, and enhancing network learning ability. Fig. 6 shows a residual module that combines the ReLU activation function with the Batch Normalization operation (BN). A series of $3 \times 3$ convolution kernels are used to simulate the enlarged receptive field, and the input of the convolution block is superposed with the output of the convolution after passing through the $1 \times 1$ convolution kernel and the residual path.

2.3.4. Multi-scale feature extraction module

The basic U-Net model uses the maximum pooling operation at the pooling layer, which dramatically decreases the number of arguments and increases the receptive field of the convolution operation, but also loses some important location space information. Therefore, in this study, the extended convolution operation was adopted to expand the receptive field to a greater extent without increasing the number of arguments, so as to acquire more context information. As shown in Fig. 7, the number of parameters of the conventional $3 \times 3$ convolution kernel is 9, the expansion rate is 1, and the receptive field is $3 \times 3$. When the expansion rate is 2, the number of parameters of the $3 \times 3$ convolution is also 9, but the size of the receptive field increases to $7 \times 7$.

At the same time, the hollow space pyramid pooling (ASPP) module is added to extract multi-scale features from context information. The spatial pyramid pool network proposed by He et al. [16] is the design source of ASPP module. The module acts as a bridge between encoding and decoding stages through these empty convolution kernels of different types. As shown in Fig. 8, ASPP consists of four parallel expansion convolutions, with expansion rates $[3, 5, 7, 9]$, respectively. ASPP can effectively extract the information of different scales on the same feature graph by parallel connecting the extended convolution with different expansion rates and splicing features, and making full use of the extracted multi-scale features.

2.3.5. Multi-output feature fusion module

In the U-Net network model, there are four jump connections for feature fusion. However, only the segmented image after prediction is output in the S4 stage [18]. In order to control the number of parameters, high-level features are compressed. Therefore, in the S4 phase, the feature information at the bottom level is integrated with the compressed high-level feature information, result-
2.4. Segmentation task model

The classification task model is used to extract the entire lung region and learn the unique imaging features of COVID-19 patients through the deep learning network, to effectively distinguish COVID-19 patients, other pneumonia patients, and normal subjects, realizing the purpose of efficient diagnosis of COVID-19 [19].

In order to effectively solve multi-classification problems under small samples such as COVID-19 image diagnosis, this study uses a network model based on the combination of WGAN and DNN.
classifier. The process of this model and its key design contents are described in detail below.

2.4.1. Process of the model

The process of this model mainly includes the following three steps.

Step 1: The original data set sample was divided into the training set and test set, and the WGAN model was trained with the training set sample to optimize model parameters. Two models were trained simultaneously: the generation model G, which captured the data distribution, and the discriminant model D, which estimated the probability that the samples came from the training data.

Step 2: Generate simulation samples through WGAN generator model, and filter simulation samples through WGAN discriminant model.
Step 3: Use filtered simulated samples to train DNN classifier. After the training, the test set samples were used to test the DNN classifier to verify the classification effect.

2.4.2. WGAN model

The traditional GAN [20] model first generates simulated samples through optimization generators, which are very close to real samples. The second step is to improve the discriminator. The purpose of this step is to distinguish whether the input image is the original sample or the simulated sample. Finally, the generator and discriminator are trained alternately, and the model is continuously improved through continuous games. Therefore, when the output probability of the discriminator to the sample is 0.5, the model reaches the optimal, that is, it cannot distinguish whether it is the original sample or the simulated sample. It is difficult to train the original GAN model, because in the process of game between generator and discriminator, generator loss has some problems, such as gradient instability, gradient disappearance, and even model collapse. These problems are mainly caused by unreasonable distance measurement method (JS/KL divergence) of equivalent optimization [21].

To solve these problems, this study used Wasserstein Generative Adversarial Network (WGAN) [22] to generate adversarial samples. The key of WGAN is to transform the JS/KL divergence of the traditional GAN loss calculation into Wasserstein distance, which is also known as EM(Earth Mover) distance and is defined as follows:

\[ W(P, P_g) = \inf_{\gamma \in \Pi(P, P_g)} \mathbb{E}_{x \sim \gamma}[\|x - y\|] \]  

(1)

Firstly, \(\Pi(P, P_g)\) is the set of whole feasible joint distributions combined with P_r and P_g. Then for every feasible distribution \(\gamma\), we are able to sample \((x, y) \sim \gamma\) to achieve a true sample \(x\) and a generated sample \(y\), and compute the distance \(\|x - y\|\) between the pair of samples. Then we will compute the expectation of the sample distance for the joint distribution \(\gamma\). In the whole feasible joint distributions, we will lower the expected value. This lower bound is regarded as the Wasserstein distance. The optimization function of WGAN is as follows:

\[ L = E_{x \sim P_r}[f_w(x)] - E_{x \sim P_g}[f_w(g(x))] \]  

(2)

By minimizing the distance L, WGAN narrows the distribution between real samples and simulated samples, so as to solve problems such as lack of diversity and model collapse existing in traditional GAN, and achieve faster training speed and more stable training process [23].

2.5. Datasets for training network

Firstly, we use open source data sets to train the deep learning model. The data types of COVID-19 images mainly include X-ray images and CT images [24]. Most of the research work carried out so far is in the range of tens to hundreds of datasets, with a small sample size and only a very small number of large open source COVID-19 image datasets. Samples in the dataset included Bacterial Pneumonia (BP), Viral Pneumonia (VP), COVID-19, other pneumonia and normal pneumonia categories. The above information suggests that the model for COVID-19 image diagnosis should target a small sample with multiple classifications. This feature of the dataset also provides the conditions for the classification model combining WGAN and DNN classifier. Table 1 shows six common COVID-19 open source datasets.

We then use the collected data set to test the model trained by the open source data set. The test sets were collected from lung images of 950 normal patients, 796 patients with common pneumonia, and 105 patients with COVID-19 provided by our hospitals. These data were used with the consent of the patients and the hospital. Three sets of data were collected and manually annotated by professional imaging physicians. The original data set contained 1345 normal lung X-ray images, 1075 ordinary pneumonia X-ray images, and 307 COVID-19 X-ray images. It also contains 1321 CT images of normal lungs, 1021 CT images of common pneumonia, and 587 X-ray images of COVID-19. Considering the limited size of the data set, we carried out data expansion and finally obtained 4251 image data through flipping, scaling, translation, rotation and other operations. Ethics approval is granted by the institutional review board.

2.6. Evaluation index

2.6.1. Evaluation index of segmentation task

There are seven commonly used evaluation indexes for segmentation tasks, which can be divided into region-based evaluation indexes for two-dimensional data and surface distance-based evaluation indexes for three-dimensional data [30]. Region-based evaluation indicators include: DSC (Dice Similarity Coefficient), Relative Volume Difference (RVD) and Volumetric Overlap Error (VOE). Evaluation indicators based on surface distance include: Root Mean Square Surface Distance (RMSD), Hausdorff Distance (HD) and Average Symmetric Surface Distance (ASD). These indexes will be used to evaluate the segmentation result of each model.

First, define the symbols \(V_{seg}\) and \(V_{gt}\). \(V_{seg}\) represents the predicted segmentation result, and \(V_{gt}\) represents the corresponding segmentation result of ground truth.

Dice Similarity Coefficient is the most commonly used evaluation index in segmentation tasks, and its expression is shown as follows:

\[ DSC = \frac{2 \times (V_{seg} \cap V_{gt})}{V_{seg} + V_{gt}} \]  

(3)

The relative volume difference RVD represents the volume difference between \(V_{seg}\) and \(V_{gt}\), and its expression is as follows:

\[ RVD = \frac{V_{seg}}{V_{gt}} - 1 \times 100\% \]  

(4)

The volume overlap error VOE is similar to the DSC. It replaces the intersection operation with the difference set operation to represent the error rate, and its expression is as follows:

\[ VOE = \frac{2 \times (V_{seg} - V_{gt})}{V_{seg} + V_{gt}} \]  

(5)

\(S(V_{gt})\) and \(S(V_{seg})\) represent the set of \(V_{gt}\) and \(V_{seg}\) surface elements, respectively. For any \(y \in S(V_{gt})\) to \(S(V_{seg})\), the distance is defined as the minimum Euclidean distance of all points from Y to \(S(V_{seg})\), and the distance formula is as follows:

\[ d(y, S(V_{seg})) = \min_{y \in S(V_{seg})} ||y - S(V_{gt})||_2 \]  

(6)

c represents the sum of the number of voxels on the surfaces of \(V_{gt}\) and \(V_{seg}\), and \(count()\) calculates the number of elements in the set. The calculation formula of the evaluation index based on the surface distance is as follows:

\[ RMSD = \sqrt{\frac{D^2(S(V_{gt}) \cdot S(V_{seg})) + D^2(S(V_{seg}) \cdot S(V_{gt}))}{c}} \]  

(7)

\[ HD = \max_{y \in S(V_{gt})} \max_{y \in S(V_{seg})} D(y, S(V_{seg})) \]  

(8)

\[ ASD = \frac{D^2(S(V_{gt}) \cdot S(V_{seg})) + D^2(S(V_{seg}) \cdot S(V_{gt}))}{c} \]  

(9)
2.6.2. Evaluation index of classification task

There are nine commonly used evaluation indicators for categorizing tasks. It includes Recall, Precision, Specificity, Accuracy, F1-score, Receiver Operating Characteristic curve (ROC), Area Under Curve (AUC) [31,32], F-measure and G-mean.

Recall, Precision, Specificity, Accuracy are shown as follows. In this expression, N is the total amount of samples. NTP indicates the number of positives that are truly classified and predicted to be positive. NFP represents the number of people who were truly classified as negative but predicted to be positive. NFN represents the number of negative truly classified and predicted to be negative. NTN represents the number of true categorizations that are positive and predicted to be negative.

\[
R_{\text{Recall}} = \frac{N_{TP}}{N_{TP} + N_{FN}} \tag{10}
\]

\[
P_{\text{Precision}} = \frac{N_{TP}}{N_{TP} + N_{FP}} \tag{11}
\]

\[
S_{\text{Specificity}} = \frac{N_{TN}}{N_{TN} + N_{FP}} \tag{12}
\]

\[
A_{\text{Accuracy}} = \frac{1}{N} (N_{TP} + N_{TN}) \tag{13}
\]

F1-score is the harmonic average of recall rate and accuracy rate. Compared with a single evaluation index, this index can analyze and evaluate diagnosis results more comprehensively.

\[
F_1 - \text{score} = \frac{2 \times P_{\text{Precision}} \times R_{\text{Recall}}}{P_{\text{Precision}} + R_{\text{Recall}}} \tag{14}
\]

ROC is a curve with False Positive Rate (FPR) as X-axis and True Positive Rate (TPR) as Y-axis. It represents the relation between trade-offs of model effects (represented by TPR) and losses (represented by FPR). The ideal case is TPR=1, FPR=0. Therefore, the closer the ROC curve is to the coordinate (0,1), the higher the accuracy of the classifier, the better the algorithm performance and the better the classification effect. AUC refers to the area under the ROC curve. According to the nature of ROC, the larger the value of AUC is, the closer it is to 1, the better the effect of the classifier will be.

F-measure is the harmonic mean of accuracy and recall [33], and G-mean is the geometric mean of recall of each category [34]. Parameter L represents the number of categories, and their calculation formula is as follows:

\[
F - \text{measure} = \frac{2 \times P_{\text{Precision}} \times R_{\text{Recall}}}{P_{\text{Precision}} + R_{\text{Recall}}} \tag{15}
\]

\[
G - \text{mean} = \left( \prod_{i=1}^{l} R_{i} \right)^{1/l} \tag{16}
\]

### 3. Results

According to the method proposed in this paper, experiments are carried out from three dimensions of segmentation task, classification task, and multi-task combined with segmentation and classification to obtain experimental results. Then we verify the model one by one and analyze the experimental results.

The environment of all experiments is a computer with 128 GB memory, 4-channel 2080Ti graphics card and E5-2678 V3 CPU. All network model algorithms are implemented in PyTorch framework, with a learning rate of 0.02 and a learning rate attenuation coefficient of 0.4. The optimizer uses SGD with momentum coefficient of 0.8 and weight attenuation of 0.000 1. In addition, cross-GPU synchronous BN operation is used. The preset number of training cycles epoch is 50 and batch-size is 20.

#### 3.1. Results and analysis of segmentation task

The first is qualitative analysis. In the same experimental environment, the original U-Net and DRD U-Net were used to segment the images of the test set, respectively, and the results are shown in Fig. 11. In the figure, the prediction area of U-Net is significantly larger than the corresponding DRD U-Net and manual segmentation, which indicates that the proposed network has a high segmentation accuracy for image segmentation. This is because the designed module increases the receptive field of convolution operation, obtains more context information, and greatly reduces the false positive and false negative areas. Therefore, the coherence and accuracy of the segmented area are improved.

Although the precision of manual segmentation in the figure above is very high, it can be seen from Fig. 12 that manual labeling will lead to human error. However, both U-NET and the proposed DRD U-NET can correctly segment the lesion region according to the image features, which has better consistency compared with manual segmentation.

Then it is quantitative analysis. In the same experimental environment, the DRD U-Net network model used in this study is compared with the basic U-Net, U-Net++, and Attention U-Net network models. The latter two algorithms are improvements on the basic U-Net. The parameters in the network are the same as those in the original paper, and all models in the experiment adopt hyperparameter settings. The specific values of the experimental results are shown in Table 2, with the optimal indexes in bold. It can be seen that the DRD U-Net model proposed in this paper has achieved the best results in 17 out of all 18 evaluation indicators in the three data sets. Compared with the traditional network, the U-Net coefficient is increased by 0.33%, and the error is reduced by 3.57 mm.

In addition, the ablation experiment of DRD U-Net model was also carried out. In the experiment, the influence of deleted modules on network performance was explored by deleting some modules, so as to verify that each improved algorithm of DRD U-Net model played a positive role in the results. Based on the basic U-Net model, the model with additional residual modules is called
Table 2
Comparison of different segmentation task models (mean ± variance).

| Dataset      | Model       | DSC/% (↑) | VOE/% (↓) | RVD/% (↓) | ASD/mm (↓) | RMSD/mm (↓) | HD/mm (↓) |
|--------------|-------------|-----------|-----------|-----------|------------|-------------|-----------|
| COVID-chestxray | U-Net       | 98.46±0.29| 3.03±0.57 | 0.72±0.54 | 0.23±0.07  | 0.52±0.3    | 8.13±7.03 |
|               | U-Net++     | 98.51±0.31| 2.93±0.59 | 0.64±0.51 | 0.21±0.06  | 0.46±0.16   | 7.28±5.29 |
|               | Att U-Net   | 98.46±0.31| 3.03±0.6  | 0.67±0.53 | 0.23±0.06  | 0.53±0.24   | 10.71±9.34|
|               | DRD U-Net   | 98.56±0.3 | 2.85±0.58 | 0.64±0.51 | 0.2±0.05   | 0.42±0.11   | 5.59±2.52 |
| MosMedData    | U-Net       | 84.96±4.13| 25.93±6.28| 10.22±6.8 | 0.24±0.08  | 0.53±0.23   | 5.22±2.17 |
|               | U-Net++     | 85.19±4.08| 25.58±6.22| 9.94±6.57 | 0.23±0.08  | 0.53±0.21   | 5.26±2.24 |
|               | Att U-Net   | 84.86±4.19| 26.08±6.35| 10.37±6.96| 0.24±0.08  | 0.54±0.23   | 5.23±2.08 |
|               | DRD U-Net   | 85.38±4.13| 25.28±6.27| 10.27±7.19| 0.23±0.07  | 0.51±0.2    | 5.04±2.14 |
| COVID-CT-set  | U-Net       | 89.1±2.51 | 19.57±4.02| 4.49±3.47 | 0.2±0.06   | 0.44±0.2    | 6.49±4.06 |
|               | U-Net++     | 89.42±2.58| 19.05±4.14| 4.11±3.1  | 0.2±0.06   | 0.43±0.19   | 6.27±3.77 |
|               | Att U-Net   | 89.15±2.63| 19.48±4.21| 4.27±3.3  | 0.2±0.06   | 0.45±0.21   | 6.72±5.63 |
|               | DRD U-Net   | 89.68±2.54| 18.62±4.12| 3.96±3.39 | 0.19±0.05  | 0.42±0.16   | 5.63±2.35 |

Table 3
Ablation experiments using the DRD U-Net model.

| Evaluation index | Res U-Net | DR U-Net | DRD U-Net |
|------------------|-----------|----------|-----------|
| DSC/% (↑)        | 92.85     | 92.94    | 93.11     |
| VOE/% (↓)        | 12.67     | 12.52    | 12.39     |
| RVD/% (↓)        | 3.93      | 3.82     | 3.87      |
| ASD/mm (↓)       | 0.23      | 0.20     | 0.19      |
| RMSD/mm (↓)      | 0.50      | 0.46     | 0.43      |
| HD/mm (↓)        | 6.79      | 5.66     | 5.21      |

Res U-Net, and the model with additional ASPP modules is called DR U-Net. Finally, the new model DRD U-Net proposed in this paper is further obtained. The experimental results are shown in Table 3. The results prove that each module improved by DRD U-Net model can gradually improve the segmentation accuracy.

3.2. Results and analysis of classification task

According to the model proposed in this paper, first of all, the MosMedData data set is divided into training set and test set, and WGAN is trained with the training set sample. Then, the trained WGAN is used to generate simulation samples. By comparing the results of several experiments, the best structural parameters of WGAN model were selected. The results show that the generator has 1 hiding layer, 32 correction linear units (ReLU) in the hiding layer, 42 SigmoID units in the output layer of the generator, and the dimension of noise vector Z is set to 16. The discriminator has a hidden layer, the hidden layer is 64 ReLUs, the output layer of the discriminator is 1 inactive function unit. The WGAN structure parameters corresponding to each type of training samples are the same. In each iteration cycle, the discriminator iterates 120 times and the generator iterates once more. After the simulation sample
is generated, the DNN classifier is trained with simulation samples and tested with test samples. DNN selects the following structural parameters after several experiments: there are 3 hidden layers, each containing 32 ReLUs, and the loss function is cross entropy, and the output layer is softmax function, and the number of iterations is 3000.

In order to prove the effectiveness of the model framework combined with WGAN and DNN in the application of small sample multi-classification COVID-19 diagnosis results, classical machine learning model and oversampling algorithm were used to compare the classification model algorithm and data generation, respectively. As the Random Forest (RF) algorithm has higher classification accuracy and better generalization performance compared with other statistical machine learning models, Naive Bayes (NB) algorithm has a simple principle, easy implementation, and stable classification performance. So RF and NB are used as classical statistical machine learning classification models. When the original small sample data is applied to the classical statistical machine learning algorithm, the classifier is first trained with the training sample, and then tested with the test sample. During the data oversampling method to generate the simulated sample, firstly, use SMOTE method to oversampling all kinds of original sample data, then use the SMOTE method to train the classification model and use the original sample test model. In the experiment of data oversampling method, classical statistical machine learning classification models RF and NB and deep learning classification model DNN were used for multi-classification models. In the generative adversarial network method, WGAN is used to generate a large amount of simulation data to train the classification model and the original samples are used to test the model. RF, NB and DNN are also used for the multi-classification model.

The experimental results are shown in Table 4. When no simulated samples are generated to expand the amount of data, the accuracy of RF algorithm and NB algorithm are 52.15% and 38.64%, F-measure is 54.32% and 11.98%, and G-mean is 46.21% and 0, respectively. All three evaluation indexes are relatively low. This represents a large number of samples from COVID-19 patients that have not been identified. All indicators of NB are much lower than RF, indicating that NB is more sensitive to sample number than RF in this MosMedData data set. The results display that it is difficult to train the classical machine learning model effectively in the environment of small sample and multi-classification data, which also proves the necessity of generating simulated samples.

| Strategies          | Accuracy   | F-measure | G-mean |
|---------------------|------------|-----------|--------|
| RF                  | 0.5215     | 0.5432    | 0.4621 |
| NB                  | 0.3864     | 0.1198    | 0      |
| RF with SMOTE       | 0.5965     | 0.6874    | 0.4657 |
| NB with SMOTE       | 0.5354     | 0.5527    | 0.5164 |
| DNN with SMOTE      | 0.6532     | 0.6765    | 0.6052 |
| RF with WGAN        | 0.3054     | 0.2835    | 0.2721 |
| NB with WGAN        | 0.6246     | 0.6354    | 0.6154 |
| DNN with WGAN       | 0.7384     | 0.7465    | 0.7437 |

The above experimental results show that both the end-to-end DRD U-Net network model segmentation task and the combination of GAN and DNN classifier classification task can achieve better results. Next, the classification task is combined with the segmentation task to compare the effects of the combination. Table 5 shows the experimental results on the MosMedData dataset using different multi-classification task network models. The results show that when U-Net is used as the segmentation strategy, the accuracy and F-measure values are the highest when using DNN and SMOTE as the classification strategy, and the G-mean is the highest when using DNN and WGAN. Therefore, for different segmentation strategies, the classification strategy combined with DNN and WGAN may not get the best result. However, when DRD U-Net model is used as the segmentation strategy and DNN and GAN model is used as the classification strategy, accuracy, F-measure and G-mean obtained from the MosMedData data set are all optimal, which also indicates that the classification and segmentation multi-task model involved in this study has a good effect. Thus, COVID-19 patients, other pneumonia patients and normal subjects can be effectively distinguished and COVID-19 images can be diagnosed.

3.3. The results and analysis of multi-task

Analysis of model tasks shows that the network structure of COVID-19 detection and diagnosis model includes segmentation network (U-Net, U-Net ++ Attention U-Net, DRD U-Net, etc.) and classification network (NB, RF, DNN with WGAN, etc.). According to the use of segmentation network, classification network, segmentation and classification network, it can be divided into the

4. Discussion

Analysis of model tasks shows that the network structure of COVID-19 detection and diagnosis model includes segmentation network (U-Net, U-Net ++ Attention U-Net, DRD U-Net, etc.) and classification network (NB, RF, DNN with WGAN, etc.). According to the use of segmentation network, classification network, segmentation and classification network, it can be divided into the
single-task method and multi-task method. Based on the imaging characteristics of COVID-19 patients at different times, feature extraction can be more efficient by using segmentation tasks to segment the lung lesion regions. By using the classification task to learn the unique imaging features of COVID-19 patients, the classification task is combined with the segmentation task to effectively distinguish COVID-19 patients, other pneumonia patients and normal subjects.

It can be seen from the analysis of image types that the results obtained from X-ray data sets are worse than those obtained from CT data sets. On the one hand, X-ray is less sensitive than CT images. On the other hand, CT image data is three-dimensional data, which contains more abundant feature information. CNN’s excellent performance in image processing derives from its powerful feature extraction ability in essence. Therefore, CNN can obtain more feature information from 3D CT image data, thus improving the diagnosis effect.

According to the above results, deep learning combined with WGAN has achieved good results in improving the accuracy of COVID-19 image recognition, and all indicators have been improved compared with traditional methods. This result has important significance for the study of COVID-19 image diagnosis. Most of the COVID-19 data sets are facing the problem of a small sample size, while deep learning needs a great deal of data sets for training, which leads to the inconvenience of using deep learning for COVID-19 image diagnosis research and further affects the rapid identification and diagnosis of COVID-19 images. GAN-based data enhancement is likely to help solve this problem. The proposed method is not only designed to solve the problem of COVID-19 image diagnosis and prediction, but also to provide ideas for solving the problem of multi-classification under a small sample. Meanwhile, based on the characteristics of deep learning, this method does not rely on precise knowledge of medical research field. Therefore, while ensuring the effectiveness, the obstacle of extending the method to other application fields is greatly reduced. It also provides an important idea for the field in which it is difficult to conduct deep learning modeling problems directly due to the small sample size.

In addition, some researchers also use auxiliary means such as FPN, attention mechanism, decision tree and Lasso while using the multi-task model. The results display that the auxiliary method can enhance the performance of the model effectively. Therefore, this can also be our next research direction, which is to make further improvement and perfection of the model. Three dimensional visualization [35] of the lung and chest model built after segmentation by deep learning may advance the diagnosis of the lung in future implementation. Extreme learning algorithms [36] may be implemented to improve segmentation procedures in future.

| Table 5: Comparison of evaluation indexes of different multi-task models. |
|---------------------------------------------------------------|
| Segmentation | classification          | Accuracy  | F-measure | G-mean  |
|---------------|-------------------------|-----------|-----------|---------|
| U-Net         | RF with SMOTE           | 0.5484    | 0.5673    | 0.4550  |
| U-Net         | NB with SMOTE           | 0.3226    | 0.3220    | 0.2232  |
| U-Net         | DNN with SMOTE          | 0.5806    | 0.6254    | 0.4811  |
| U-Net         | DNN with WGAN           | 0.5484    | 0.5660    | 0.5233  |
| Att U-Net     | RF with SMOTE           | 0.6452    | 0.6605    | 0.6332  |
| Att U-Net     | NB with SMOTE           | 0.2903    | 0.2931    | 0.2627  |
| Att U-Net     | DNN with SMOTE          | 0.6129    | 0.6260    | 0.6043  |
| Att U-Net     | DNN with WGAN           | 0.7097    | 0.7007    | 0.6839  |
| DRD U-Net     | RF with SMOTE           | 0.6135    | 0.6318    | 0.6321  |
| DRD U-Net     | NB with SMOTE           | 0.4484    | 0.4673    | 0.4550  |
| DRD U-Net     | DNN with SMOTE          | 0.7397    | 0.7657    | 0.7839  |
| DRD U-Net     | DNN with WGAN           | 0.7997    | 0.8057    | 0.7993  |

5. Conclusion

In this paper, a COVID-19 diagnostic model is constructed based on segmentation and classification multi-task. In the segmentation task, the end-to-end DRD U-Net model is used to segment the lung lesions to improve the ability of feature reuse and target segmentation. In the classification task, the model combining WGAN and DNN classifier is used to effectively solve the problem of multi-classification of COVID-19 images with small samples. Thus achieving the goal of effectively distinguishing COVID-19 patients, other pneumonia patients and normal people. Compared with traditional single-task methods and other multi-task methods, the accuracy of classification is improved to some extent, so that COVID-19 image recognition and classification can be carried out more efficiently, which is of great significance for COVID-19 image diagnosis. More importantly, this research provides a potentially efficient solution to the problem of small sample multiple classification in other application fields, thus improving the intelligence level of each fields. In the future, we will use auxiliary means such as attention mechanism and decision tree to further improve the classification accuracy of the model. It is expected that the model will continue to be improved in the future to enhance the applicability and accuracy of the method, and the details of the structure and details are shown in Figs. 5 through 10.

Funding

This research is not funded by any grants.

Ethics approval

All human subjects in this study have given their written consent for the participation of our research.

Declaration of Competing Interest

The authors declare no conflict of interest for this paper.

References

[1] Y. Fang, H. Zhang, J. Xie, et al., Sensitivity of chest CT for COVID-19:comparison to RT-PCR, Radiology 296 (2) (2020) E115–E117.
[2] T. Ai, Z. Yang, H. Hou, et al., Correlation of chest CT and RT-PCR testing in coronavirus disease 2019(COVID-19)in China: a report of 1014 cases, Radiology 296 (2) (2020) E32–E40.
[3] P. Huang, T. Liu, L. Huang, et al., Use of chest CT in combination with negative RT-PCR assay for the 2019 novel coronavirus but high clinical suspicion, Radiology 295 (1) (2020) 22–23.
[4] Y. A. N. G. Yaqi, J. I. A. Yuchi, A Study on the Translation Methods of Medical Terms in Diagnosis and Treatment Protocol for COVID-19 (Trial Version 7), Cross-Cultural Communication 18 (1) (2022) 78–82.
L. Riazpur, P. Irvin, J. Zhu K, et al. CheXNet: radiologist-level pneumonia detection on chest X-rays with deep learning[EB/OL]. [2021-01-25]. https://arxiv.org/pdf/1711.05225.pdf.

I.R. Zech, M.A. Badgley, M. Liu, et al. Variable generalization performance of a deep learning model to detect pneumonia in chest radiographs: a cross-sectional study. PLoS medicine 15 (11) (2018) 1–17.

Y.F. Liu, Y. Zhou, X. Liu, F. Dong, C. Wang, Z.H. Wang, Wasserstein GAN based on a new generation of AI small sample data enhancement: a case study of cancer staging data in the biological domain, Engineering 5 (01) (2019) 338–354.

L. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, et al. Generative adversarial nets. in: I. Chahramani, M. Welling, C. Cortes, N.D. Lawrence, K.G. Weinberger Eds.), Advances in Neural Information Processing Systems, La Jolla: Neural Information Processing Systems Foundation, Inc., 2014, pp. 2672–2680.

A. Creswell, T. White, V. Dumoulin, K. Arulkumaran, B. Sengupta, A.A. Bharath, Generative adversarial networks: an overview, IEEE Signal Process Mag. 35 (1) (2018) 53–65.

G. Douzas, F. Baco, Effective data generation for imbalanced learning using conditional generative adversarial networks, Expert Syst. Appl. 91 (2017) 464–471.

J. Peng, H.Y. Luo, G.S. Zhao, C.C. Lin, X.S. Yi, S.J. Chen, A review of medical image segmentation algorithms based on deep learning, Comput. Eng. Appl. 57 (03) (2021) 44–57.

W.Y. Shao, Y.F. Guo, Application of multi-task learning and convolutional neural network in face recognition, Comput. Eng. Appl. 52 (13) (2016) 32–37.

O. Ronneberger, P. Fischer, T. Brox, U-Net: convolutional networks for biomedical image segmentation, in: Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2015, pp. 234–241.

L.C. Chen, Y. Zhu, G. Papandreou, et al., Encoderdecoder with atrous separable convolution for semantic image segmentation, in: Proceedings of the European Conference on Computer Vision (ECCV), 2018, pp. 801–818.

T.Y. Lin, R.P. Doll, R. Girshick, et al., Feature pyramid networks for object detection, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 2117–2125.

K. He, X. Zhang, S. Ren, et al., Deep residual learning for image recognition, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 770–777.

Jianshe S, Yuguang Y, Daxin Z, Lianta S, Yifeng H, Jianlong H, Comparative Analysis of Pulmonary Nodules Segmentation using Multiscale Residual U-Net and fuzzy C-means Clustering, Computer Methods and Programs in Biomedicine [preprint], 103(10) (2020) 106323 (preprint).

S. Jianshe, Y. Yuguang, Z. Daxin, S. Lianta, H. Yifeng, and H. Jianlong, Automatic segmentation of cardiac magnetic resonance images based on multi-input fusion network, Comput. Methods Progr. Biomed. 209 (2021) 106323 (preprint).

Ye Y, Shi J, Huang Y, Zhu D, Su L, Huang J. Management of Medical and Health Big Data Based on Integrated Learning-based Health Care System: A Review and Comparative Analysis. Computer Methods and Programs in Biomedicine [preprint]. 10.1016/j. cmpb.2021.106293.

L. Goodfellow, J. Pouget-Abadie, M. Mirza, et al., Generative adversarial networks, in: Proceedings of the Conference and Workshop on Neural Information Processing Systems, 2014.

Z. Fang, et al., A novel multi-stage residual feature fusion network for detection of COVID-19 in chest X-ray images, IEEE Trans. Mol. Biol. Multi-Scale Commun. 8 (1) (2021) 17–27.