Classification of lungs infected COVID-19 images based on inception-ResNet

Yunfeng Chen [a,1], Yalan Lin [a,1], Xiaodie Xu [a], Jinzhen Ding [a], Chuzhao Li [a], Yiming Zeng [a,*], Weili Liu [e], Weifang Xie [b,c,d], Jianlong Huang [b,c,d]

[a] Department of Pulmonary Medicine, The Second Affiliated Hospital of Fujian Medical University, Quanzhou, 362000, China
[b] Faculty of Mathematics and Computer Science, Quanzhou Normal University, Quanzhou 362000, China
[c] Fujian Provincial Key Laboratory of Data Intensive Computing, Quanzhou 362000, China
[d] Key Laboratory of Intelligent Computing and Information Processing, Fujian Province University, Quanzhou 362000, China
[e] Software School, Xinjiang University, Urumqi 830091, China

ARTICLE INFO

ABSTRACT

Objective: Nowadays, COVID-19 is spreading rapidly worldwide, and seriously threatening lives . From the perspective of security and economy, the effective control of COVID-19 has a profound impact on the entire society. An effective strategy is to diagnose earlier to prevent the spread of the disease and prompt treatment of severe cases to improve the chance of survival.

Methods: The method of this paper is as follows: Firstly, the collected data set is processed by chest film image processing, and the bone removal process is carried out in the rib subtraction module. Then, the set preprocessing method performed histogram equalization, sharpening, and other preprocessing operations on the chest film. Finally, shallow and high-level feature mapping through the backbone network extracts the processed chest radiographs. We implement the self-attention mechanism in Inception-Resnet, perform the standard classification, and identify chest radiograph diseases through the classifier to realize the auxiliary COVID-19 diagnosis process at the medical level, all in an effort to further enhance the classification performance of the convolutional neural network. Numerous computer simulations demonstrate that the Inception-Resnet convolutional neural network performs CT image categorization and enhancement with greater efficiency and flexibility than conventional segmentation techniques.

Results: The experimental COVID-19 CT dataset obtained in this paper is the new data for CT scans and medical imaging of normal, early COVID-19 patients and severe COVID-19 patients from Jinyintan hospital. The experiment plots the relationship between model accuracy, model loss and epoch, using ACC, TPR, SPE, F1 score and G-mean to measure the image maps of patients with and without the disease. Statistical measurement values are obtained by Inception-Resnet are 88.23%, 83.45%, 89.72%, 95.53% and 88.74%. The experimental results show that Inception-Resnet plays a more effective role than other image classification methods in evaluation indicators, and the method has higher robustness, accuracy and intuitiveness.

Conclusion: With CT images in the clinical diagnosis of COVID-19 images being widely used and the number of applied samples continuously increasing, the method in this paper is expected to become an additional diagnostic tool that can effectively improve the diagnostic accuracy of clinical COVID-19 images.

© 2022 Published by Elsevier B.V.

1. Introduction

At the beginning of 2020, COVID-19 has become an epidemic worldwide. It is spreading in various regions of the world, seriously causing a threat to life and health. WHO has declared the disease a Public Health Emergency of International Importance (PHEIC), and the issue is considered a health emergency [1].

Many clinical trials have confirmed that the early diagnosis of COVID-19 infection can effectively reduce the incidence and mor-

* Corresponding authors at: 950 Eastsea street, Fengzhe District, Quanzhou City, Fujian Province, 362000, China.
E-mail addresses: 9199912007@fjmu.edu.cn (Y. Chen), zeng_yi_ming@126.com (Y. Zeng).
[1] Authors contributed equally to this work.

tality of COVID-19. Ifpatients can be screened early and treated with clinical diagnoses effectively, efficacy would be improved, medical costs would be reduced, and the survival of critically ill patients will be prolonged [2].

The biggest challenge to stopping the disease's spread is the absence of effective detection techniques. Although reverse transcription-polymerase chain reaction is a crucial tool for validating COVID-19, it requires numerous tests spaced out over several days and has a high false-negative rate. Additionally, it takes 4–6 hours to receive findings. In addition, there is a severe shortage of RT-PCR reagents in many areas with severe epidemics. In contrast, CT examination equipment is widely available in major hospitals [3]. CT is an efficient and safe method in clinical practice to identify COVID-19 through combined assistance of signs and clinical symptoms. Meanwhile, CT images could also play a role in the follow-up treatment and post-adjuvant treatment of COVID-19: light and common types of patients with lung imaging showed multiple small patches of lung and interstitial changes, with an apparent extrapulmonary band; severe and critical patients, more developed into multiple lung frosted glass performances and infiltration shadow, severe cases may appear pulmonary consolidation [4].

In conclusion, medical imaging diagnostic technology plays a crucial role in limiting the spread of the virus and treating COVID-19.

1.1. COVID-19 and clinical manifestations

Caused by SARS-CoV-2, COVID-19 is an infectious disease and could cause a severe acute respiratory syndrome, typical inflammatory response, vascular injury, microangiopathy, angiogenesis and extensive thrombosis [5].

Early COVID-19 usually presents with respiratory symptoms. Symptomatic individuals may occasionally develop mild but self-limiting infections with side effects, including cough, breathing problems, fever, shortness of breath, fatigue and sore throat [6,7].

Few patients might develop more severe symptoms, often requiring hospitalization for pneumonia. In some cases, people may experience severe life-threatening complications, such as those leading to ARDS, sepsis, and septic strictures, leading to multiple systemic organ dysfunction failures [8].

1.2. Medical multimodality imaging

Generally, computed tomography (CT) is used to present images of pneumonia, tuberculosis, emphysema, or other pleura diseases (membrane covering the lungs). Classification of lung CT images is a necessary initial step in lung image analysis. The classification's significance of algorithms is particularly evident due to the strength of homogeneity, artifacts, and the proximity of different soft tissue gray levels. In this paper, a fully convolutional neural network is designed to classify chest infected with COVID-19 CT images end-to-end automatically. For sizes of diseased area distributions in digital lung photography and the diversity of shapes, we present multiscale image information into a fully convolutional dense network structure, which improves classification accuracy.

Medical imaging techniques are commonly used in clinical practice, including computed tomography (CT), magnetic resonance imaging (MRI) and ultrasound imaging (US), etc. [9,10]. The more accurate imaging techniques and precise pictures allow for a better screening of disorders, which can give crucial data on conditions like COVID-19. CT has become a necessary technical means for observing treatment response and prognosis evaluation with technology development. The potential COVID-19 performance characteristics can be extracted from the image through the newest medical imaging information processing, effectively solving the prob-lem, which plays a significant role in routine clinical pathology sampling and analysis [11].

There are a series of parameters that can be modified in CNN, such as the number of processing units (neurons), the number of layers of the network model, the size of the convolution kernel, the size of the stride, the learning rate, and the activation function. After adjusting and changing various parameters, there are many typical networks based on CNN. The efficiency of CNN is significantly improved than the traditional image classification method,

2. Related work

Through the above background analysis, medical image diagnosis is the key to realizing the COVID-19 diagnosis analysis and treatment planning. At present, a large number of clinical studies and practices have proved that it is more effective to integrate multiple modal information to diagnose and treat diseases. And differences in physical imaging principles and their impact on image classification. From multiple dimensions of structure and function, medical image processing is a necessary technical means to analyze COVID-19 and carry out diagnosis, treatment planning, and prognosis analysis [12]. With the purpose of segmenting the normal lung organs and the lesions of COVID-19 in medical images, medical image classification is an indispensable step in medical image processing and analysis. The relevant features are extracted from the lung segmentation area in the analysis stage to provide a reliable basis for clinical diagnosis and treatment and prognosis monitoring of COVID-19. In the treatment planning stage, the divided COVID-19 diseased area information is used for treatment path planning and target dose calculation. Preoperative planning can ultimately assist doctors in making more accurate diagnoses, surgical navigation, prognosis analysis, and radiation therapy and drug chemotherapy plans [13].

In this section, we briefly introduce traditional CNN and ResNet.

2.1. Convolutional neural networks

At present, it is generally recognized in the industry that with good ability for feature extraction and recognition convolution neural network (CNN) is a deep artificial neural network and is the most widely used machine learning technology [14,15]. The sharing of weights is one of its most prominent features, which significantly reduces the number of weights and the complexity of the network model [16].

Convolutional neural network (CNN) minimizes the loss function by many convolution filters, nonlinear activation functions and pooling layers. Any component of the convolutional neural network is a simulation of biological neuronal activity, and all components have many inputs (denoted as $x_1, x_2, x_3, \ldots, x_n$), that is, the so-called characteristic matrix. When the characteristic matrix (i. e., input) is fixed, the convolution layer will convolution the characteristic matrix and each filter $\{W_i\}_k$ to obtain the pre-activated figure S. The specific Equations is as follows:

$$H_i = W_i \cdot X + b_i, \quad i = 1, \ldots, k \qquad (1)$$

Note: Symbol $\cdot$ represents convolution operation, b_i represents deviation parameter.

A typical CNN network structure usually includes convolutional layers and pooling layers (Pooling Layer), and the end is one or more fully-connected layers [17,18]. The network structure diagram is shown in Fig. 1.

After the convolution operation is completed, a bias needs to be added to the result of the convolution, and then a nonlinear excitation function (in a neural network, the excitation function is usually relu) is used to obtain the final output of the layer. Result [19]. The Eq. 2 illustrates this process. x_j^l represents the jth feature
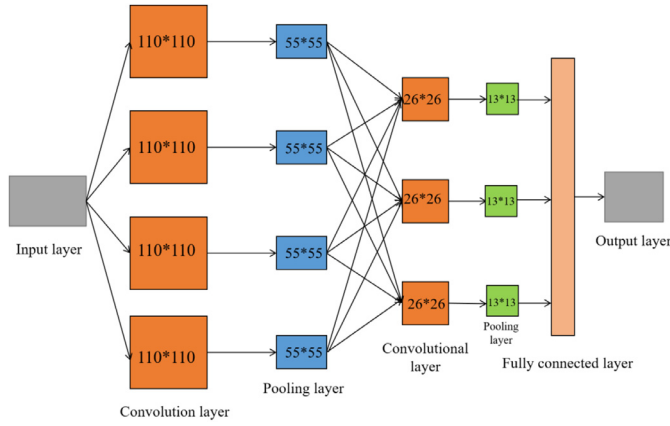
Fig. 1. Structure of CNN.

Fig. 2. ResNet network structure.

map of the lth layer, f represents the excitation function, M is the set of input feature maps, · represents the convolution operation, k represents the convolution kernel, and b represents the bias term [20].

$$x_j^l = f\left(\sum_{x \in M_j} x_j^{l-1} \cdot k_{ij}^l + b_i^l\right) \qquad (2)$$

2.2. Deep residual networks

ResNet is considered a continuation of deeper networks, in which optimal theory for training deeper networks is introduced [21]. The classic ResNet network has 50 layers, 101 layers and 152 layers. Among them, the 152-layer deep CNN won the 2015-ILSVRC championship. Furthermore, ResNet achieves a 28% progress on a prominent image recognition example dataset named cOco132 [22]. ResNet mainly exploits the idea of bypass channels in the "road network", using the following mathematical Eq. 3 and 4 [23]:

$$g(x_i) = f(x_i) + x_i \qquad (3)$$

$$f(x_i) = g(x_i) - x_i \qquad (4)$$

In Eq. 3, f is the transformed signal, and x is the original input. The original input is added to $f(x)$ via a bypass path. In Eq. 4, $g(x)$ is used to perform the residual operation. ResNet introduces shortcut channels within layers to realize connections between different layers, but these gates are data-independent and parameter-free compared to highway networks [24]. In a highway network, these layers represent non-residual functions when the shortcut path is closed. The submodule can be composed of two parts, a linear direct mapping $x \to x$, and a nonlinear mapping $F(x)$. If the direct mapping of $x \to x$, is optimal, the learning algorithm can easily set all the weight parameters of the nonlinear mapping $F(x)$ to 0. If there is no direct mapping, let the nonlinear mapping $f(x)$ learn a linear $x \to x$, mapping is difficult [25]. However, in ResNet, residual information is always passed, and the shortcut channel is never closed. Residual links (fast channel connections) accelerate the fusion of deep networks, allowing ResNet to avoid the gradient reduction problem [26]. A typical ResNet network structure is shown in Fig. 2.

ReLU is an activation function. The purpose of using the activation function is to prevent gradient dispersion, deepen the propagation depth of the function, and reduce the gradient attenuation caused by deep convolution. The Equation of ReLU is as follows.
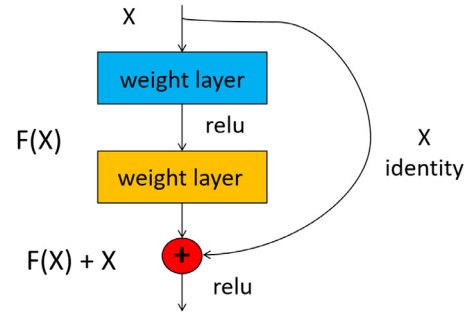
$$R(x) = max(0, x) \qquad (5)$$

When $x>0$, $R(x)=x$, the derivative is 1; when $x \leq 0$, $R(x)=0$, the derivative is 0.

3. Materials and methods

In this section, we will first discuss the selection and preprocessing of the dataset, then discuss the benefits of the Inception-ResNet method, further assess its performance, and finally discuss the selection of the self-attention mechanism to improve the effectiveness of Inception-ResNet image diagnosis.

3.1. Dataset

Firstly, we scanned the lung images of various pneumonia patients and healthy people obtained via X-ray and CT scanners at Jinyintan Hospital in Wuhan. We collected lung X-ray and CT images from 2045 normal subjects, 785 early patients, and 130 severe 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. The institutional review board grants ethics approval. A total of 2960 sample images were collected, and the sample images covered three main categories, marked as normal confirmed images(2045), images of early COVID-19(78) and images of severe COVID-19(130) [27,28]. All images were 299 × 299. The data distribution used in the experiment is shown in Table 1.

We segment the lesion area and combine the characteristics of COVID-19 and clinical cognition to extract the quantitative CT image features of many characteristics of the COVID-19 infection image, and then use feature extraction to change the neural network to construct a prediction model for the diagnosis of COVID-19, and screen out 20 radiomic features for the diagnostic value of COVID-19. This provides a non-invasive detection method for COVID-19 prediction.

3.2. Inception-ResNet architecture and model training

We use the ImageNet dataset, which contains more than 1 million images, to train the CNN model Inception-ResNet-v2. The network contains 164 layers so that it could classify about 1000 object categories [29]. Therefore, the network model can learn rich attribute representations for various images. The initial ResNet block incorporates convolutional filters and residual connections of multiple sizes [30]. The motivation chooses its architecture based on experimental results and comparative analysis with other popular deep learning models. Inception-ResNet-v2 has an excellent trade-off between resource requirements and model performance (accuracy). Other extensive and bulky models cannot be selected because this model is expected to work in an edge environment. Fig. 5 depicts the architecture of the custom model [31].

The final Inception-ResNet-v2 network structure is shown in the Fig. 3.

Table 1
The distribution of training, validation, and test sets used in the dataset.

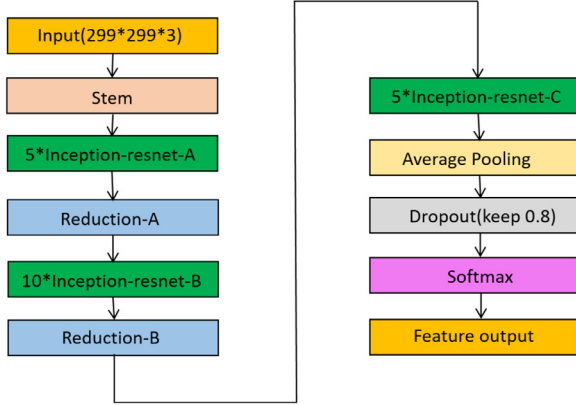| Type	Total number of images	Training set	Validation set	Test set
Normal	2045	1585	583	278
Early COVID-19	785	259	77	50
Severe COVID-19	130	46	24	18

Fig. 3. Structure of Inception-ResNet-v2 network.

The structure of the stem network is a shallow feature extraction module of the combined network. Through a variety of different types of convolution parallel combinations, some of the structures split the large convolution kernel into series small convolution kernels, for example, 7×7 large convolution kernels are split into 1×7 and 7×1 small convolution kernels. This asymmetric structure can extract more multi-level structural features and increase diversity. The stem network structure in the combined network is shown in Fig. 4.

Our process is image acquisition, image preparation and DL model building, and the final step is diagnosis and feedback. Data preparation is one of the most critical steps in many cases because it is pre-comparing the collected data efficiently that leads to accurate results [32]. It contains operations, including equalizing the number of images in each class, simple filtering and denoising, etc. Afterward, the used datasets are divided into two groups: the training set and the test set. We conduct many tuning experiments during the training process to obtain the optimized network parameters. The more the model learns, the more accurate the classification output will be. Then, after many experiments are aggregated and deployed, the model is tested on the remaining unseen images (the test set) and the DL network [33].

Fig. 5 shows more detailed information about the proposed system network.

3.3. Self-attention mechanism and 3D nature of CT images

In order to enhance the classification capability of Inception-ResNet further, we introduce a self-attention mechanism, in which we squeeze each channel of the output feature map, perform global average pooling, and obtain a feature vector [34]. Welearn the feature weight of each channel through this feature vector and convert the initial feature map into a weighted feature map. The performance of the network is further enhanced by this technique, which takes into account the relationships between various channels and adaptively modifies the relative feature intensities of the channels. The processing burden it introduces, according to experiments, is tolerable given the increased network performance [35].

In addition, CT images are natural 3D image sequences, and related studies have shown that better classification performance can

be obtained if lung nodule CT images are input as 3D cubes. This is because 3D CT images contain dependencies between consecutive slices, and we can obtain image features of lung nodules at a deeper level to achieve better classification. This study will also use 3D CNN to improve the accuracy of lung nodule classification.

3.4. Multi-scale feature extraction for SENet lung infection classification model

The underlying network architecture of this subsection is SENet, and based on this, a SENet lung infection classification method with multi-scale feature extraction (FB_SENet) is proposed. The method first preprocesses the original dataset, and the preprocessing process is divided into region of interest extraction, transformation of pixel values, and image enhancement. After the data are pre-processed, they are fed into FB_SENet for classification output.

The overall flow of the experiment is shown in Fig. 6.

FB_SENet is mainly composed of three 3×3 convolutions, one maximum pooling, four blocks, multi-branch feature extraction module and Softmax. In this paper, the extracted ROI is preprocessed and fed into FB_SENet. In the original SENet, the input layer is a 7×7 convolution, but in this paper, the 7×7 convolution is replaced by three 3×3 convolutions in order to capture more image features and obtain the same size of the perceptual field, which can reduce the number of network parameters and enhance the feature extraction capability of the network. Finally, feature extraction is performed by 4 Blocks and MFEM and the output is classified by Softmax.

3.4.1. Graded residuals module

In this paper, we use Res2net, a method to enhance the multi-scale representation of the network and expand the perceptual field of the network layers, which uses a series of smaller 3×3 convolutional groups to replace the single 3×3 convolution used in the original SENet. This is done by dividing an arbitrary input feature map $x_t \in R^{C \times H \times W}$ into s feature map subsets uniformly after 1×1 convolution, each of which is represented by a_i and $i \in \{1,2,3, ..., s\}$. Each feature subset a_i has the same spatial size as its input feature map x_t, but the number of channels is reduced to $1/s$ of x_t. To be able to reduce the number of parameters in the network and increase the reuse of features, each feature subset a_i is followed by its corresponding 3×3 convolution, except for a_1, which is represented by m_i. These convolution sets are connected by residuals, and the output of m_i is denoted by b_i. The feature subset a_i is aggregated with the output of m_{i-1} and fed into m_i, so that b_i is denoted as :

$$b_i = \begin{cases} a_i & i = 1 \\ m_i(b_{i-1} + a_i) & 1 < i \le s \end{cases} \tag{6}$$

Where i is an integer and $i \in \{1,2, ..., s\}$. In the FB_SENet proposed in this paper, s = 4 is taken to divide the feature map into 4 parts equally, and the best performance can be obtained at this time. The grouping of features can enhance the extraction ability of the network model for lesion region features, and can obtain more feature information at different scales and process them efficiently.

At this point, the output features are b1, b2, b3, and b4, respectively, and the features obtained from each branch are concatenated and passed through a 1×1 convolution as the output
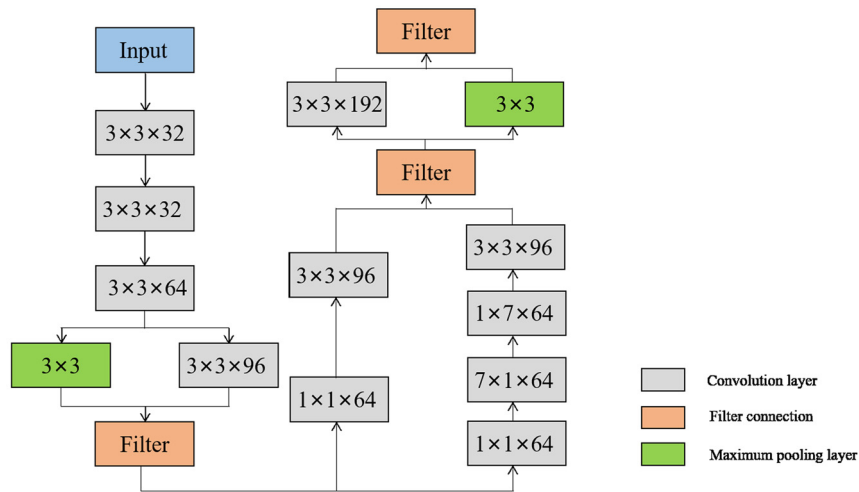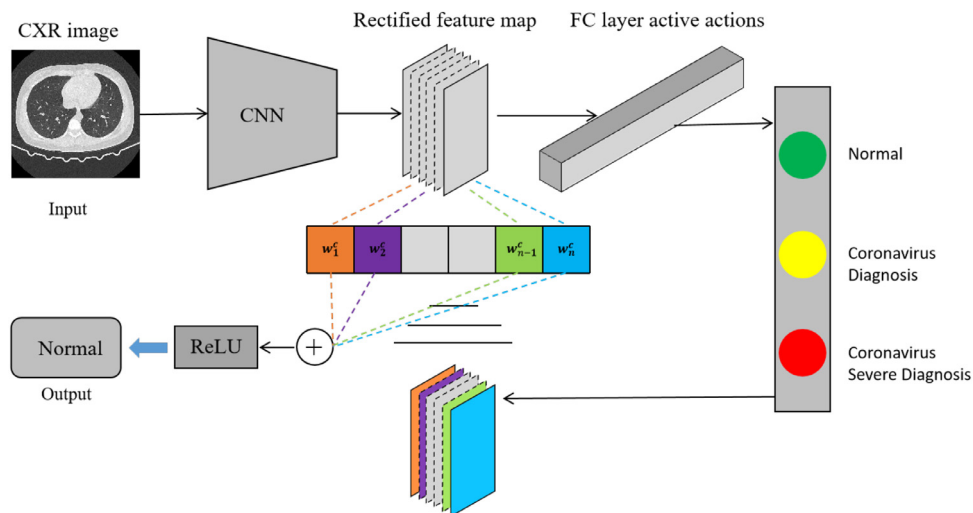
Fig. 4. Structure of stem.
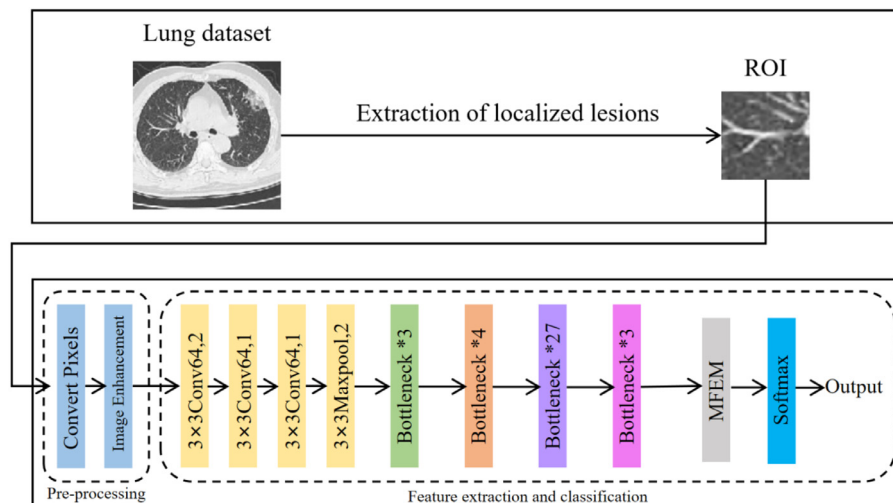

Fig. 5. The total model architecture.

Fig. 6. Overview of the proposed method.

of the graded residual module, i.e., the input of the dual-attention feature extraction module, as shown in Eq. 7.

$$\mathbf{F} = f^{1 \times 1}[b_1; b_2; b_3; b_4] \qquad (7)$$

The use of graded residual module in Eq 7 $\mathbf{F} \in R^{C \times H \times W}$ enhances the connection between the lesion region and the surrounding tissues, and enhances the feature extraction ability of the network while improving the multi-scale representation of the network.

3.4.2. Dual attention feature extraction module

In order to minimize the influence of background noise and spurious residual information, Woo et al. proposed an adaptive learning feature extraction method to extract important features from both channel dimension and spatial dimension. The method takes the middle layer feature map $F \in R^{C \times H \times W}$ as the input, and the 1D feature map $A_c \in R^{C \times 1 \times 1}$ computed by the channel attention module is dotted with the input feature map $F \in R^{C \times H \times W}$, and the 2D feature map $A_S \in R^{1 \times H \times W}$ computed by the spatial attention module, and finally the two are dotted to obtain the final output F' as shown in Eq. 8:

$$F\prime = A_S(A_C(F) \otimes F) \otimes A_C(F) \otimes F \qquad (8)$$

In equation $f^{1 \times 1}$ the convolution operation with kernel 1×1 and $\otimes$ represents element-wise point multiplication operation between elements. In the channel attention module, the extracted feature map is compressed by global maximum pooling and global average pooling in spatial dimension to obtain two $1 \times 1 \times C$ real numbers, mapped using multilayer perceptron, using Sigmoid activation function to obtain weight factor $A_C(F)$, and finally weighted with the original feature map F points, as shown in Eq. 9:

$$A_C(F) = \sigma(M(P_{GA}(F)) + M(P_{GM}(F))) \qquad (9)$$

Where, P_{GA} denotes global average pooling, P_{GM} denotes global maximum pooling, M denotes multilayer perceptron, and σ denotes Sigmoid activation function.

In the channel attention module, the proposed FB_SENet convolves the feature maps obtained by two pooling operations, and sums the output features to obtain the weight factors using the Sigmoid activation function. This not only solves the problem of losing spatial information in the input feature vector, but also enhances the nonlinear expression capability of the network. The weight factor is used to multiply and weight the original feature points, which can effectively avoid the interference of redundant information such as background, and thus enhance the recognition ability of the network for the lesion area, as shown in Eq. 10:

$$A_C(F) = \sigma\left(f^{1 \times 1}\left(f^{1 \times 1}(P_{GA}(F))\right) + f^{1 \times 1}\left(f^{1 \times 1}(P_{GM}(F))\right)\right) \qquad (10)$$

In the spatial attention module, the feature maps obtained after dot product are pooled on the channels using average pooling and maximum pooling to obtain two $1 \times H \times W$-dimensional pixel matrices $P_{GA}(G)$ and $P_{GM}(G)$, which are concatenated into a 7×7 convolution and aggregated using the Sigmoid activation function to obtain a two-dimensional weight matrix $A_S(G)$, which is finally weighted to the original input feature map. $F \in R^{C \times H \times W}$ to obtain the final output F'. The Equations are shown in (11) and (12):

$$G = A_C(F) \otimes F \qquad (11)$$

$$A_S(G) = \sigma(f^{7 \times 7}([P_{GA}(G); P_{GM}(G)]) \qquad (12)$$

where $f^{7 \times 7}$ denotes the convolution operation with kernel 7×7, and G denotes the output of the dot product between the channel attention feature map $A_c \in R^{C \times 1 \times 1}$ and the input feature map $F \in R^{C \times H \times W}$ corresponding to the pixel points. By using the dual-attention feature extraction module, we can enhance

the feature extraction ability of the network for the lesion region and improve the classification accuracy by adaptive refinement of the extracted features.

3.4.3. Multi-branch feature extraction module

To increase the feature resolution in ASPP, the cavity convolution is coupled in parallel. However, this technique does not properly make use of the deep semantic data that the network has previously gathered. In order to make the network model improve the classification accuracy without losing the image feature information, this paper improves it on this basis. Firstly, the feature map F'obtained after the dual-attention feature extraction module is convolved in parallel by averaging pooling and three nulls of 6, 12, and 18 nulls respectively. Next, 1×1 convolutions are used to enhance the extraction of lesion features after the parallel null convolution, then the average pooled features are fed into 1×1 convolutions, and the features obtained above are fused with the average pooled features to obtain a fixed size output of F'' by a 1×1 convolution. The network is then able to extract the global features of the lesioned region and deepen the semantic information of the lesioned region, while resampling the feature maps at different scales using multiple convolutions with different void rates can effectively improve the classification accuracy. The process can be represented by Eq. 13:

$$F\prime = \sigma\left(f^{1 \times 1}\left[f^{1 \times 1}(P_A(F\prime)); f^{1 \times 1}\left(f^{3 \times 3, d=6}(F\prime)\right); f^1\right]\right.$$
$$\left[\times 1\left(f^{3 \times 3, d=12}(F\prime)\right); f^{1 \times 1}\left(f^{3 \times 3, d=18}(F\prime)\right); P_A(F\prime)\right] \qquad (13)$$

where P_A is the average pooling and d denotes the dilation rate, i.e., the hole rate. In order to efficiently utilize the global and local feature information of the lesion area, the F'' computed by the multi-branch feature extraction module is element-wise dotted multiplication with the F' output obtained by using the two-attention feature extraction module, and finally it is summed with the original input to prevent the gradient from disappearing.

3.5. Performance evaluation indicator

Given the labeling and classification results, we use accuracy (ACC), sensitivity (true positive rate, Recall, TPR) and specificity (SPE) as several commonly used main indicators to evaluate the performance of the proposed Inception-ResNet. Specifically, we can use TP (true positives), FN (false negatives), FP (false positives) and TN (true negatives) to calculate ACC, recall TPR and positive class prediction error PPV. Related equations are listed in the following Eqs. 6–12:

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \qquad (14)$$

$$TPR = \frac{TP}{TP + FN} \qquad (15)$$

$$SPE = \frac{TN}{FP + TN} \qquad (16)$$

$$precision = \frac{TP}{TP + FP} \qquad (17)$$

$$F1\ score = \frac{2 \times precision \times recall}{precision + recall} \qquad (18)$$

$$TNR = \frac{TN}{TN + FP} \qquad (19)$$

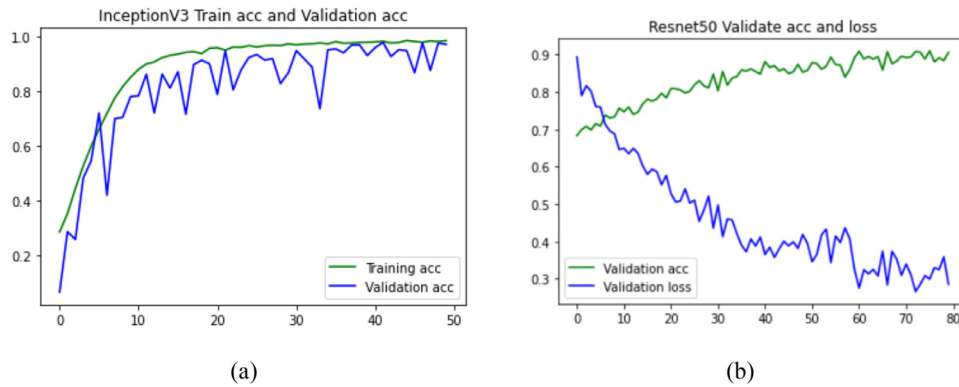$$G - mean = \sqrt{TPR \times TNR} \qquad (20)$$

(a)

(b)

Fig. 7. (a) The accuracy of Inception-ResNet (b) The loss of Inception-ResNet.

Table 2
Performance of different types of CNNs.

Model	ACC(%)	TPR(%)	SPE(%)	F1 score(%)	G-mean(%)
CNN	85.34	79.89	87.82	89.45	86.83
ResNet	87.56	82.41	88.94	90.63	84.71
Inception-ResNet	88.23	83.45	89.72	95.53	88.74

4. Experiments and analysis

4.1. Experimental details

We crop images of 64 × 64 and 3D volume data of 48 × 48 × 9 from CT scans of lung nodules based on the annotation center and divide them into two parts, 75% of which are used for training validation and the rest as test set. To overcome the limited sample of nodules, we augment the nodules with translation, rotation, and flip operations [36]. This enhancement also helped extract features from lung nodules' CT scans that did not change with the three procedures. In addition to the traditional initial ResNet, we introduce a self-attention mechanism and 3D convolution, resulting in our new networks: 3D SE-IRNet and 3D SE-CDNet. We employ mini-batch gradient descent to minimize the CNN's loss function and learn weights in our experiments. We randomly initialize the weights using a Gaussian distribution and update the network parameters using standard backpropagation during training. The learning rate is initially set to 0.1 then and decreases by 5% every 1500 epochs. Batch size and momentum are initialized to 64 and 0.9. We maintain a similar data distribution between training and test sets to avoid over-or under-expression of features due to imbalanced distributions. The hardware environment for our experiments is based on NVIDIA Tesla P100. The software environment is Keras 2.1.0 and Tensorflow GPU 1.12.0 based on Ubuntu 16.04 operating system. The language used is Python 3.6.7.

4.2. Experimental metrics and results

Based on Inception-ResNet, the CNN model is trained to classify the collected CT images of different periods.

Each pre-trained model is trained on grayscale images. The model accuracy and loss plots of the Inception-ResNet CNN are shown in Fig. 7 a and b.

From the model in Fig. 7, only 1200 COVID-19 images out of 1200 images were identified, and only 15 non-COVID-19 images out of 1230 images.

Our experimental indicator results are shown in Table 2, which presents the precision, recall, f1-score, and G-mean of Inception-ResNet for the dataset for detailed performance analysis [37].

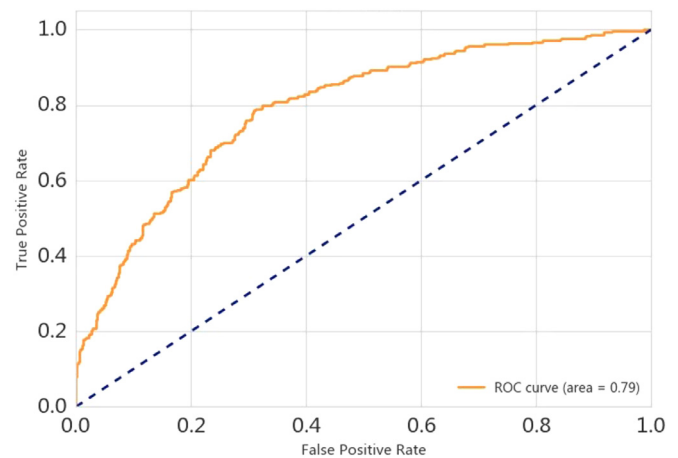The ROC curve of the experimental results is shown in Fig. 8.

Fig. 8. ROC curve of the experimental results.

Table 3
The effectiveness of the number of feature groups [%].

method	Accuracy	Recall	Precision	F1 Score
s=1	77.81	77.39	78.53	78.72
s=2	78.74	79.37	77.71	78.10
s=3	79.67	80.25	78.03	79.04
s=4	**80.60**	**82.67**	**79.96**	**80.53**
s=5	73.75	75.15	76.52	75.46

The results of normal images, early COVID-19 images and severe COVID-19 images are shown in Fig. 9.

The selection of the number of feature groups s in the graded residual module. In order to achieve the best classification effect, therefore, the selection of s was verified in SENet, and the experimental results are shown in Table 3.

When s=1, the network degenerates into an ordinary SENet, while when s=1, s=2, s=3, and s=4, all four evaluation indexes of the model improve to some extent, and when s=5, its accuracy decreases again. Therefore, it can be proved through experiments that when the number of feature groups s=4, the network can obtain the best performance at this time and its classification is the best.

Comparison with classical classification network models. Including: DenseNet, ResNet101, MnasNet, MobileNet2, ShuffleNetV2, SK_ResNet101, SENet. the classification results under the same data set are shown in Table 4. From the table, it can be seen that the FB_SENet proposed in this paper is higher than the above classical classification models under each evaluation index, which further confirms the superiority of the method in this paper.
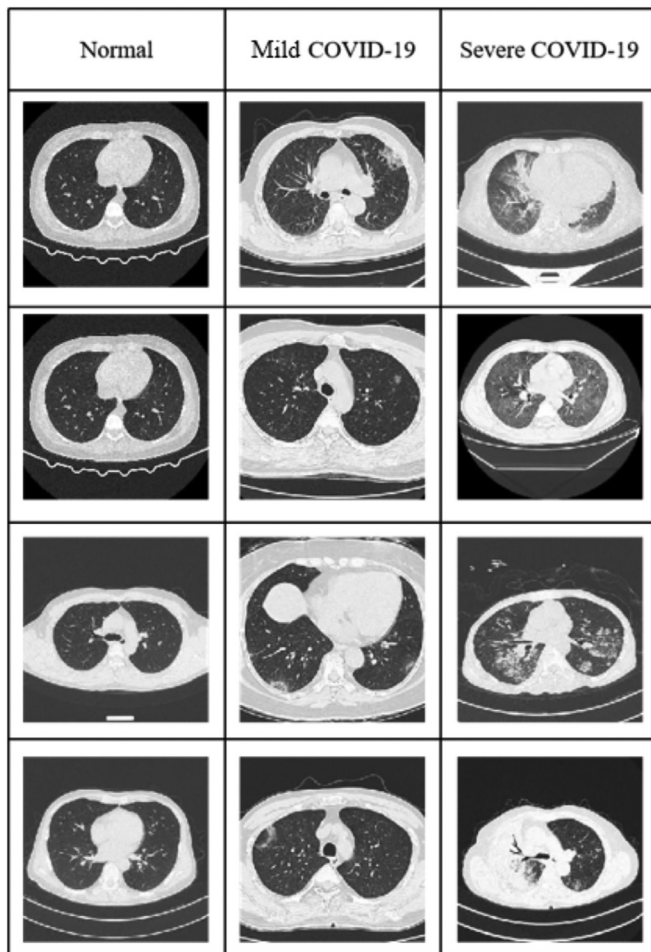
Fig. 9. The result of coronavirus pneumonia images prediction.

Table 4
Classification performance of the lung dataset.

Method	Accuracy	Recall	Precision	Fl score	AUC
DenseNet	79.53	79.17	78.33	79.50	93.00
ResNetl01	77.42	76.40	77.47	76.28	93.00
MnasNet	73.52	73.57	73.62	73.62	91.00
MobileNet2	81.26	81.05	80.39	80.31	95.00
ShuffleNetV2	76.07	76.56	76.89	76.89	91.00
SK_ResNetl01	81.32	82.22	81.03	81.50	91.00
SE_ResNetl01	77.82	77.78	78.75	78.02	91.04
FB_SENet	87.74	86.04	87.00	86.42	96.00

5. Discussion

To increase the feature resolution, ASPP connects the cavity convolution in parallel. The network has previously extracted deep semantic data, but this technique does not make good use of it. Recover spatial details from downsampled paths. The existence of 5 pooling layers in the network makes it difficult to identify small areas, so the number of pooling layers needs to be adjusted appropriately [38].

The best accuracy comes from a hybrid Inception-ResNet learned model. Image augmentation and self-attention mechanism help improve the accuracy of disease classification tasks. The average accuracy of this technique is 98.66%. Other TL models, such as CNN and ResNet, have accuracy rates of 97.33% and 90.33%.

The experimental results show that the classification accuracy of this method on the training set is 82.4%, and the classification accuracy on the test set is 77.7%. The COVID-19 imaging omics pre-diction model constructed in this paper was applied to the quantitative prediction of COVID-19 diagnosis and severe cases, assisting clinicians in diagnosis. The method in this paper has two advantages: 1) Non-invasive CT lung images were used to predict the diagnosis and severity of new COVID-19 patients from the whole lung level; 2) Lung CT images are easy to obtain, and the lung radiomics prediction model for benign and malignant evaluation is accessible to clinical application.

In this paper, a multi-scale feature extraction-based SENet lung classification method is proposed to address the problems in existing lung classification methods. This chapter proposes a multi-scale feature extraction-based SENet lung classification method. Firstly, a hierarchical residual module is used to strengthen the connection between the context of lesion regions and increase the perceptual field of the network; secondly, a dual-attention feature extraction module is used to strengthen the features of lesion regions while reducing the influence of background noise and its redundant information, and deeply focusing on useful lesion information; next, a multi-branch feature extraction module is constructed to sample features at different scales to enhance the feature utilization of the original image and Finally, the ordinary convolution is replaced by octave convolution to reduce the number of parameters and improve the classification effect. The method achieves the best performance under several evaluation metrics.

In practice, many other machine learning techniques, i.e. auto-encoder-extreme learning [39], ensemble learning [40], as well as multiscale residual U-Net [41] can be applied to the lung segmentation and analysis for COVID-19 diagnosis.It may be worthwhile noting that three dimensional visualization [42] of the lung lesions can also assist medical experts in the analysis of COVID-19.

6. Conclusion

The method in this paper is mainly aimed at diagnosing early and severe patients, and can also be applied to other pathological classification and prognosis prediction. In the follow-up work, the newly developed COVID-19 diagnosis and severe prediction software will be further improved in close connection with COVID-19 medicine and clinical knowledge.

Although some image systems base on neural network has been established in hospitals to help assistance, these systems still have excellent upside potential, and their performance can be further improved through the following aspects:

1) Artificial intelligence, as a data-driven discipline, collects more high-quality datasets, which will help to improve algorithm performance further.
2) Most existing AI diagnosis technologies are based on medical images, and the diagnosis is based on a single basis. Combining data such as clinical information and travel exposure history can make up for the insufficiency of medical imaging alone.
3) Since early COVID-19 and severe COVID-19 have many overlapping clinical features, how to better distinguish these two types is also one of the future research directions.

Ethics approval

The studies involving human participants were reviewed and approved by Jin Yin-tan Hospital Ethics Committee (KY-2020-51.01) and the Second Affiliated Hospital of Fujian Medical University Ethics Committee(2020-248).

Declaration of Competing Interest

The authors declare no conflict of interest for this paper.

Funding

This research is supported by Science and Technology Program of Quanzhou (no. 2021CT0010), Emergency Public Project of Fujian Medical University[grant number 2020YJ008]and Personnel training program of Fujian Respiratory Medicine Center, HXZX202208.

References

[1] F He, Y Deng, W. Li, Coronavirus disease 2019: What we know? J. Med. Virol. 92 (7) (2020) 719–725.

[2] S Felsenstein, J A Herbert, P S McNamara, et al., COVID-19: Immunology and treatment options, Clin. Immunol. 215 (2020) 108448.

[3] J. Li, L. Liu, S. Fong, R.K. Wong, S. Mohammed, J. Fiaidhi, Y. Sung, K.K.L. Wong, Adaptive Swarm Balancing Algorithms for rare-event prediction in imbalanced healthcare data, PLoS One (2017), doi:10.1371/journal.pone.0180830.

[4] Y. Ye, J. Shi, Y. Huang, D. Zhu, L. Su, J. Huang, Management of medical and health big data based on integrated learning-based health care system: a review and comparative analysis, Comput. Method. Program. Biomed. 209 (2021) 106293.

[5] C Stasi, S Fallani, F Voller, C. Silvestri, Treatment for COVID-19: an overview, Eur. J. Pharmacol. 889 (2020) 173644 Epub 2020 Oct 11. PMID: 33053381; PM-CID: PMC7548059, doi:10.1016/j.ejphar.2020.173644.

[6] A J Dickinson, C. Hintschich, Clinical manifestations, in: Graves' Orbitopathy, Karger Publishers, 2017, pp. 1–25.

[7] T Struyf, J J Deeks, J Dinnes, et al., Signs and symptoms to determine if a patient presenting in primary care or hospital outpatient settings has COVID-19, Cochrane Database System. Rev. (2) (2021).

[8] X Chen, S Laurent, O A Onur, et al., A systematic review of neurological symptoms and complications of COVID-19, J. Neurol. 268 (2) (2021) 392–402.

[9] L Martí-Bonmatí, R Sopena, P Bartumeus, et al., Multimodality imaging techniques, Contrast Media Molecul. Imaging 5 (4) (2010) 180–189.

[10] Jianshe Shi, Yuguang Ye, Daxin Zhu, Lianta Su, Yifeng Huang, Jianlong Huang, Automatic segmentation of cardiac magnetic resonance images based on multi-input fusion network, Comput. Methods Programs Biomed. 209 (2021) 106323.

[11] R Citro, G Pontone, M Bellino, et al., Role of multimodality imaging in evaluation of cardiovascular involvement in COVID-19, Trend. Cardiovasc. Med. 31 (1) (2021) 8–16.

[12] M J Horry, S Chakraborty, M Paul, et al., COVID-19 detection through transfer learning using multimodal imaging data, IEEE Access 8 (2020) 149808–149824.

[13] Jianshe Shi, Yuguang Ye, Daxin Zhu, Lianta Su, Yifeng Huang, Jianlong Huang, Comparative analysis of pulmonary nodules segmentation using multiscale residual U-Net and fuzzy C-means clustering, Comput. Method. Program. Biomed. 209 (2021) 106332.

[14] T J Brinker, A Hekler, J S Utikal, et al., Skin cancer classification using convolutional neural networks: systematic review, J. Med. Internet Res. 20 (10) (2018) e11936.

[15] R. Girshick, Fast r-cnn, in: Proceedings of the IEEE international conference on computer vision, 2015, pp. 1440–1448.

[16] H C Shin, H R Roth, M Gao, et al., Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning, IEEE Trans. Med. Imaging 35 (5) (2016) 1285–1298.

[17] G. Liu, J. Wu, D.N. Ghista, W. Huang, K.K.L. Wong, Hemodynamic characterization of transient blood flow in right coronary arteries with varying curvature and side-branch bifurcation angles, Comput. Biol. Med. 64 (2015) 117–126.

[18] S Hershey, S Chaudhuri, D P W Ellis, et al., CNN architectures for large-scale audio classification, in: 2017 ieee international conference on acoustics, speech and signal processing (icassp), IEEE, 2017, pp. 131–135.

[19] Y. Qin, J. Wu, Q. Hu, D.N. Ghista, K.K.L. Wong, Computational evaluation of smoothed particle hydrodynamics for implementing blood flow modelling through CT reconstructed arteries, J. X Ray Sci. Technol. 25 (2) (2017) 213–232.

[20] L Yu, B Li, B. Jiao, Research and implementation of CNN based on TensorFlow, IOP Conference Series: Materials Science and Engineering, 490, IOP Publishing, 2019.

[21] Z Lu, Y Bai, Y Chen, et al., The classification of gliomas based on a pyramid dilated convolution resnet model, Pattern Recognit. Lett. 133 (2020) 173–179.

[22] K He, X Zhang, S Ren, et al., Identity mappings in deep residual networks, in: European conference on computer vision, Cham, Springer, 2016, pp. 630–645.

[23] B Lim, S Son, H Kim, et al., Enhanced deep residual networks for single image super-resolution, in: Proceedings of the IEEE conference on computer vision and pattern recognition workshops, 2017, pp. 136–144.

[24] L Wen, X Li, L. Gao, A transfer convolutional neural network for fault diagnosis based on ResNet-50, Neur. Comput. Applica. 32 (10) (2020) 6111–6124.

[25] 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. 770–778.

[26] Zagoruyko S, Komodakis N. Wide residual networks. arXiv preprint arXiv:1605.07146, 2016.

[27] Cohen J P, Morrison P, Dao L. COVID-19 image data collection. arXiv preprint arXiv:2003.11597, 2020.

[28] Cohen J P, Morrison P, Dao L, et al. Covid-19 image data collection: Prospective predictions are the future. arXiv preprint arXiv:2006.11988, 2020.

[29] C Szegedy, S Ioffe, V Vanhoucke, et al., Inception-v4, inception-resnet and the impact of residual connections on learning, Thirty-first AAAI conference on artificial intelligence, 2017.

[30] S Peng, H Huang, W Chen, et al., More trainable inception-ResNet for face recognition, Neurocomputing 411 (2020) 9–19.

[31] Geoffery E. Hinton, RR Salakhutdinov, Reducing the dimensionality of data with neural networks, Science 313 (5786) (2006 Jul 28) 504–507.

[32] S Peng, H Huang, W Chen, et al., More trainable inception-ResNet for face recognition, Neurocomputing 411 (2020) 9–19.

[33] M Alruwaili, A Shehab, S Abd El-Ghany, COVID-19 diagnosis using an enhanced inception-ResNetV2 deep learning model in CXR images, J. Healthc. Eng. 2021 (2021 Jun 3) 6658058 PMID: 34188790; PMCID: PMC8195634, doi:10.1155/2021/6658058.

[34] D Oh, B Kim, J Lee, et al., Unsupervised deep learning network with self-attention mechanism for non-rigid registration of 3D brain MR images, J. Med. Imaging Health Inform. 11 (3) (2021) 736–751.

[35] Y Xie, J Zhang, C Shen, et al., Cotr: Efficiently bridging cnn and transformer for 3d medical image segmentation, in: International conference on medical image computing and computer-assisted intervention, Cham, Springer, 2021, pp. 171–180.

[36] S K Lakshmanaprabu, S N Mohanty, K Shankar, et al., Optimal deep learning model for classification of lung cancer on CT images, Futu. Gener. Comput. Syst. 92 (2019) 374–382.

[37] A Na, et al., Development of a computer-aided tool for detection of COVID-19 pneumonia from CXR images using machine learning algorithm, J. Radiat. Res. Appl. Sci. 15 (1) (2022) 32–43.

[38] D Kumar, A Wong, D.A. Clausi, Lung nodule classification using deep features in CT images, in: 2015 12th conference on computer and robot vision, IEEE, 2015, pp. 133–138.

[39] Zhenhao Tang, Shikui Wang, Xiangying Chai, Shengxian Cao, Tinghui Ouyang, Yang Li, Auto-encoder-extreme learning machine model for boiler NOx emission concentration prediction, Energy 256 (2022) 124552.

[40] Y. Lu, X. Fu, F. Chen, K.K.L. Wong, Prediction of fetal weight at varying gestational age in the absence of ultrasound examination using ensemble learning, Artif. Intell. Med. 102 (2020) Jan..

[41] Jianshe Shi, Yuguang Ye, Daxin Zhu, Lianta Su, Yifeng Huang, Jianlong Huang, Comparative analysis of pulmonary nodules segmentation using multiscale residual U-Net and fuzzy C-means clustering, Comput. Methods Programs Biomed 209 (2021) 106332.

[42] Chen Zhao, Jun Lv,and Shichang Du, Geometrical deviation modeling and monitoring of 3d surface based on multi-output gaussian process, Measurement 199:111569 (2022).