

Medical Image Classification of Pulmonary Nodules Based on Convolutional Neural Network

Dingwen Wang

Hangzhou High School International Division, Hangzhou, China

Hzhc_wjdi@163.com

Keywords: deep learning, Convolutional Neural Network (CNN), pulmonary nodules, lung CT images, LIDC-IDRI data set, lung cancer

Abstract: A pulmonary nodule is a large amount of tissue located in the lungs. Identifying whether it has lung CT medical imaging is of great significance for the discovery of underlying diseases, particularly lung cancer or other risk factors. In real life, radiologists or doctors are prone to misjudgment in the diagnosis of lung CT images due to fatigue and other factors, and traditional machine learning and identification medical images have the problems of time-consuming and labor-intensive manual design features and incomplete feature design. In recent years, with the great success of deep learning technology in the fields of image, voice, video, etc., the author proposes whether the CT image of the lung exists based on the Convolutional Neural Networks (CNN) combined with the LIDC-IDRI data set. The lung nodules are classified. The experimental results show that the classifier reaches 93.25% on the verification set, and the accuracy rate reaches 85.28% on the test set of 1622 pictures, the recall rate is 80.14%, the specificity is 97.46%, and the AUC value under ROC is 0.97 (The AUC value is between 0-1 and the larger the classifier is, the better the effect is). It achieves the purpose of automatically and accurately identifying lung nodules in lung CT images.

1. Introduction

A pulmonary nodule is a large amount of tissue located in the lungs. Although about 95% of lung nodules are benign, they may be cancer, especially among the elderly and smokers. Many people with this condition do not have any symptoms, but some do have a wheezing, shortness of breath or persistent cough. Treatment depends on the size of the growth, whether it is cancer, and whether it is difficult for people to breathe.

Although pulmonary nodules are generally quite common, some types of people are more likely to get lung nodules than others. This includes people over the age of 50, people who smoke, people with a family history of lung cancer, and people who use chemicals at work. Non-cancerous pulmonary nodules can be caused by a variety of conditions, including bacterial infections such as tuberculosis and histoplasmosis. Inflammatory diseases such as rheumatoid arthritis can also cause these growths, such as some birth defects, parasitic infections, and tissue abnormalities such as lung cysts and hamartomas. Malignant nodules are usually caused by lung cancer, but may also be caused by cancer in other parts of the body: for example, breast and colon cancer often spread to the lungs.

Since pulmonary nodules usually do not show obvious symptoms, pulmonary nodules are usually found only when a person performs a chest X-ray or computed tomography (CT) scan for other purposes. If the lung nodules are small or contain calcium, healthcare providers usually recommend observation waits, especially if the person does not have any underlying disease or other risk factors.

For those who are particularly large or asymmetrical in shape, he or she usually recommends a biopsy to determine if they are cancerous. This is usually done by taking a small incision in the chest to remove a sample of the cell in the lung nodule, or by passing the mouth or nose through a thin flexible tube and then removing the cells there. After the biopsy is completed, the sample is examined under a microscope for diagnosis.

Benign lung nodules usually require treatment only if they cause symptoms that is severe enough to affect breathing. Often, healthcare providers treat potential causes and then monitor growth to determine if they are growing or becoming abnormal. If a person's lung nodules are large enough to disturb the breathing, they are usually surgically removed. Although the prognosis of benign nodules is usually good, their size and location may vary.

Cancerous nodules are usually surgically removed and can also be treated with chemotherapy or radiation. Like other types of lung cancer, the survival rate of malignant nodules is very low, especially if the cancer has spread to other parts of the body. Despite this, this can be significantly increased by timely treatment, making early diagnosis very important.

2. Method

2.1 Artificial neural networks

Artificial neural network is an algorithm mathematical model that imitates the behavior characteristics of biological neural networks and performs distributed parallel information processing, but does not need to describe the mathematical model accurately. The various mathematical tools we use today are described in precise mathematical models, but in many cases it is difficult to get an accurate mathematical description. The artificial neural network is a simulation of the human brain neural network. This network relies on the complexity of the system to adjust the relationship between a large numbers of internal nodes to achieve the purpose of processing information. It has a series of intrinsic non-computing problems such as learning, associative memory, synthesis, and other capabilities, such as speech, image recognition, understanding, knowledge processing, and combination optimization calculation.

Understanding the artificial neural network from the overall framework, it consists of three main parts: the neuron model, the network model, and the learning rules of the network. Therefore, different neuron models, network models, and learning rules are combined to produce a variety of artificial neural networks. The combined neural networks also have their own characteristics.

The basis of all neurons is the MP model proposed by McCulloch and Pitts in 1943 [5], and each has its own characteristics. The difference between them is mainly in two aspects: First, the operation relationship between input and weight, such as perceptron, ADLINE, BP network are the accumulation of input and weight, while the radial basis function neural network is not; The difference in transfer function, such as step function, symbol function, linear function, saturated linear function, logarithmic sigmoid function, hyperbolic tangent sigmoid function, etc., makes the neural network more colorful. That is because the transfer function is different, which also causes the network topology of some neurons to be limited. For example, the transfer function in the original perceptron is a step function, and its output is a binary function of 1 or 0, so it cannot as an input to the further network, the original perceptron is just a single layer network.

The network topology of the neural network generally has topological structures such as forward type, feedback type, self-organizing competition type and random type network (the division method of different network structures at the starting point is also different). The forward-oriented artificial neural network has functions such as learning and classification. The feedback artificial neural network has functions such as associative memory and optimization calculation. The self-organizing competitive artificial neural network has functions such as clustering, searching, self-organization and self-learning. . Our well-known perceptrons, ADLINE, supervised Hebb learning, multi-layer forward BP network [8] and other networks are all part of the forward network. The number of layers in the network is different. For a multi-layer network, it is obviously more complicated than a network with fewer layers, but the functions will be more advanced. To give a counterexample, for example, a single-layer perceptron has a simple classification function, and it must be a linearly separable task.

Each learning algorithm has its own unique characteristics on the overall neural network. Almost all neural network learning rules can be seen as variants of the Hebb learning rules, all based on the

Hebb rule. The Hebb rule assumes that when two cells are excited at the same time, the strength of the connection between them should be increased. This rule is consistent with the "conditional reflex" doctrine and was later confirmed by neurocytology. Various learning rules are concentrated in the change of weights. Some learning rules are based on differences, some based on competition, some based on relevance, and so on. According to the learning rules, there is no need for training samples to have expected output. We can also divide the neural network into a tutor (supervised) and unsupervised (unsupervised).

2.2 Convolutional Neural Network

Convolutional neural networks and artificial neural networks are essentially indistinguishable, except that they have a convolutional layer in the neural network, that is, a convolution operation on image processing. Convolutional neural networks usually consist of convolutional layers, nonlinear processing (such as ReLu), pooling layers, fully connected layers (classifiers), etc. The name of the convolutional neural network comes from the convolution operation. The main purpose of convolution is to extract features from the input image. Convolution can learn the characteristics of an image from a small piece of data entered, and can preserve the spatial relationship between pixels.

A convolution kernel is called a "filter" or a "kernel" or a "feature detector". By sliding the filter over the image and calculating the point multiplication, the matrix is called a convolved feature or "Activation Map" or "Feature Map". Its role in the image is to play this feature check.

The role of the detector, and these feature detectors can detect different features of the image, such as edges, curves, etc.

The value in the feature map generated after convolution is closer to 1 indicating that it is more related to the feature, and the closer to -1 is, the less relevant, and when we perform feature extraction, in order to make the data less and the operation is more convenient, we directly discard it. Drop out data that is not associated.

Rectified Linear Unit (ReLU), also known as modified linear unit, is an activation function commonly used in artificial neural networks. It usually refers to a nonlinear function represented by a ramp function and its variants. An important feature of ReLu is sparsity.

Currently, a clear goal of deep learning is to dissociate key factors from data variables. Raw data (mainly based on natural data) is often entangled with highly dense features. However, if the complex relationship between features is unwrapped and converted to sparse features, the features are robust (without extraneous noise). Sparse features do not require the network to have a strong linearity inseparable mechanism. Then in deep networks, the degree of dependence on nonlinearity can be reduced. Once the neurons and neurons become linearly activated, the nonlinear part of the network comes only from the selective activation of the neurons.

Compared with the 95% sparseness of brain work, there is still a big gap between the existing computational neural networks and biological neural networks. Fortunately, ReLu only has a negative value to be sparse, that is, the introduced sparsity can be adjusted and dynamically changed. As long as the gradient training is performed, the network can automatically adjust the sparse ratio in the direction of error reduction to ensure a reasonable number of non-zero values in the activation chain.

Spatial Pooling (also known as sub-acquisition or down sampling) reduces the dimensions of individual feature maps but preserves most of the important information. There are several ways to space pooling: maximization, averaging, summing, and so on.

For Max Pooling, we define a spatial neighborhood (for example, a 2x2 window) and take the largest element from the modified feature map in the window. In addition to taking the largest element, we can also average (Average Pooling) or sum the elements in the window. In practice, the maximum pooling proved to be better.

The main role of pooling is to make the input representation (feature dimension) smaller, and the number of parameters and calculations in the network is more controllable, so that over-fitting can be controlled; making the network more Small changes, redundancy, and transformations become

invariant (small redundancy of input will not change the pooled output - because we use maximal/average operations in local neighborhoods; making the network for input images Smaller variations, redundancy, and transformations become invariant (small redundancy of inputs will not change the pooled output—because we use maximal/average operations in local neighborhoods.

The fully connected layer is a traditional multi-layer perceptron, and the Softmax activation function is used at the output layer (other classifiers like SVM can also be used, usually using Softmax). The word "Fully Connected" indicates that all neurons in the front layer are connected to all neurons in the next layer. The purpose of the fully connected layer is to use these features to classify the input image based on the training data set.

In addition to classification, adding a fully connected layer is also (generally) a simple way to learn the nonlinear combination of these features. Most of the features derived from the convolution and pooling layers may be valid for classification tasks, but combinations of these features may be better.

The sum of the output probabilities obtained from the fully connected layer is 1. This can be guaranteed at the output layer using Softmax as the activation function. The Softmax function enters a vector of any value greater than 0 and converts them to a value vector between zeros, the sum of which is one.

Finally, by combining them and training using back propagation algorithms, we can classify our images.

3. Experiment design

3.1 Data Set Handle

The data we used to train the neural network came from the Lung Image Database Consortium and the Infectious Disease Research Institute (LIDC/IDRI) [2]. Since the LIDC/IDRI database was too large, I decided to use the formatted data set "LUNA16" [3]. The data set contained CT images from 888 annotations containing pulmonary nodules, from 888 CT scans.

The size of the CT examination image is $512 \times 512 \times n$, and n is the number of axial inspections of the CT. The data set has about 200 axial directions per CT examination, that is, 200 images.

The entire data set [1] had a total of 551,065 markers, of which 1351 were labeled positive for pulmonary nodules and the rest were marked as negative for pulmonary nodules. Because of the extreme imbalance between the number of markers with and without pulmonary nodules, in order to deal with this imbalance, we decided to rotate a small number (positive pulmonary nodules) and sample most (negative lung nodules). select.

Since the training of the whole picture consumes too high computational power, we intercept and mark 50×50 grayscale images for each marker according to the coordinates for training and testing [4].

At the same time, in order to expand the data set, we adopt a data enhancement method, that is, one image can be generated by rotating, panning, flipping, etc., as shown in the following figure.

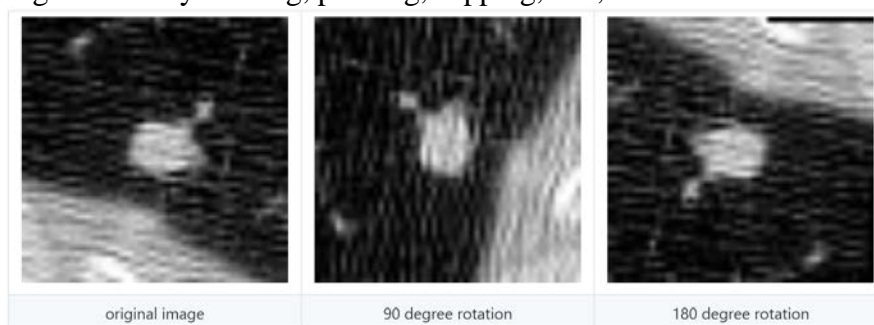


Fig. 1. Data Augmentation by rotating sample

3.2 Network Design

The model we designed has a total of 9 layers. The first layer is the input layer, and the image matrix is input. The input image is a 50*50*1 grayscale image. The second layer is a convolutional layer. The 32*5*5 size convolution kernel is used to generate a 50*50*32 feature map. The padding mode is “SAME”, that is, the image size after convolution is consistent with the original image size. The third layer is the pooling layer, and the feature map of the second layer is down sampled, the sampling size is 2*2, and the sampling mode is the maximum sampling, that is, the maximum pixel value point is taken in the sampling area of 2*2, and the function is the feature. The graph is reduced in dimension, reducing the model calculation time, and the output image is 25*25*32. The fourth layer is a convolutional layer. It uses 64 5*5-sized convolution kernels to output a more abstract feature map of 25*25*64. The padding method is “SAME”, same as above. The fifth layer is a convolutional layer. The feature map generated by the fourth convolution kernel continues to generate a 25*25*64 feature map using 64 3*3 convolution kernels, and the padding mode is “SAME”. The sixth layer is the pooling layer, the pooling window is 2*2, the pooling mode is the maximum pooling mode, and the data dimension is halved. The seventh layer is followed by a fully connected layer with a total of 512 neural units. The eighth layer is a random inactivation layer. The neurons applied to the seventh layer allow a certain proportion of neurons to be randomly inactivated to prevent over-fitting of the neural network. Here we set the deactivation ratio to 0.5, that is, halving Inactivated during training. The ninth layer applies the Softmax activation function, and the result is the second classification prediction output. The model architecture is as follows:

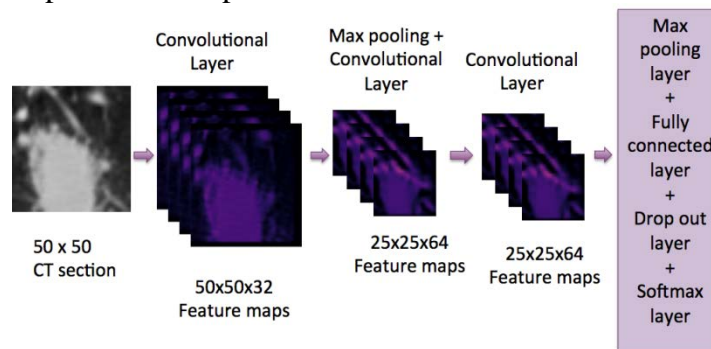


Fig. 2. Network Architecture

In summary, our model is mainly divided into three parts: the input layer is used to receive the image matrix input; the hidden layer is the middle three convolutional layers, which are used to extract the higher-level and more abstract layer-by-layer features of the input image. The main principle is to use the convolution kernel with weight parameters as the sliding window to slide on the image to generate a new feature map; the last part is a classifier composed of the fully connected layer + deactivated layer + Softmax layer for input image Category predictive output.

3.3 Hyper parameter setting

The Hyper parameters we set when training the model are as follows:

Table 1. The hyper parameters of our experiment

ITEM	VALUE	NOTES
optimizer	'adam'	Adaptive Moment Estimation
learning_rate	0.001	
EPOCHS	70	
BATCH_SIZE	96	

4. Data analysis

First of all, our data set is divided into 2 categories, with nodules and no nodules. We set no nodules to 0, that is, negative classes, and nodules are 1, that is, positive classes. Through data enhancement, our dataset implements two types of distributions of roughly 80-20. That is to say, one out of every six photos in the entire data set has a pulmonary nodule.

The total number of images in the data set is 8106, which is divided into training set (5187 sheets), verification set (1297 sheets), and test set (1622 sheets). At the beginning of the experiment, we set the batch sample size to 96, the number of rounds to 70, and save the model parameters every 100 steps. The training process is as follows:

```
Training Step: 52 | total loss: 0.35513 | time: 94.087s
| Adam | epoch: 001 | loss: 0.35513 - acc: 0.8546 -- iter: 4992/5187
Training Step: 53 | total loss: 0.35831 | time: 96.083s
| Adam | epoch: 001 | loss: 0.35831 - acc: 0.8515 -- iter: 5088/5187
Training Step: 54 | total loss: 0.37499 | time: 97.821s
| Adam | epoch: 001 | loss: 0.37499 - acc: 0.8519 -- iter: 5184/5187
Training Step: 55 | total loss: 0.36397 | time: 106.169s
| Adam | epoch: 001 | loss: 0.36397 - acc: 0.8537 | val_loss: 0.33945 - val_acc: 0.8604 -- iter: 5187/5187
--
Training Step: 56 | total loss: 0.41465 | time: 0.201s
| Adam | epoch: 002 | loss: 0.41465 - acc: 0.7805 -- iter: 0096/5187
Training Step: 57 | total loss: 0.39692 | time: 2.503s
| Adam | epoch: 002 | loss: 0.39692 - acc: 0.8109 -- iter: 0192/5187
Training Step: 58 | total loss: 0.40021 | time: 5.020s
| Adam | epoch: 002 | loss: 0.40021 - acc: 0.8125 -- iter: 0288/5187
```

Fig. 3. Network training process

After training the model, we load our log output through the Tensor Board that comes with Tensor Flow, and get the following output:

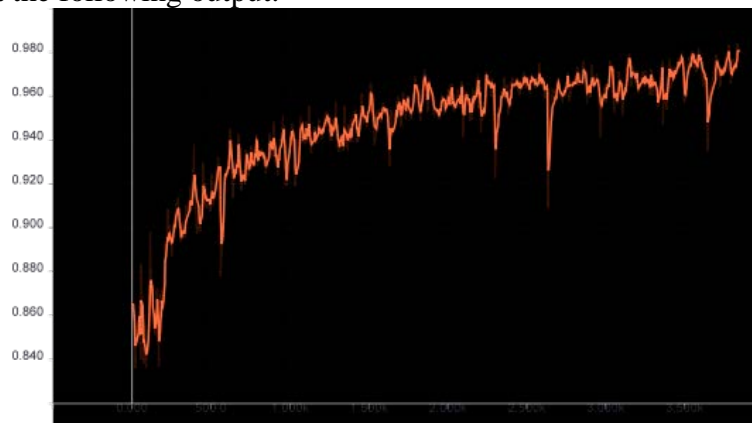


Fig. 4. Accuracy on the train set

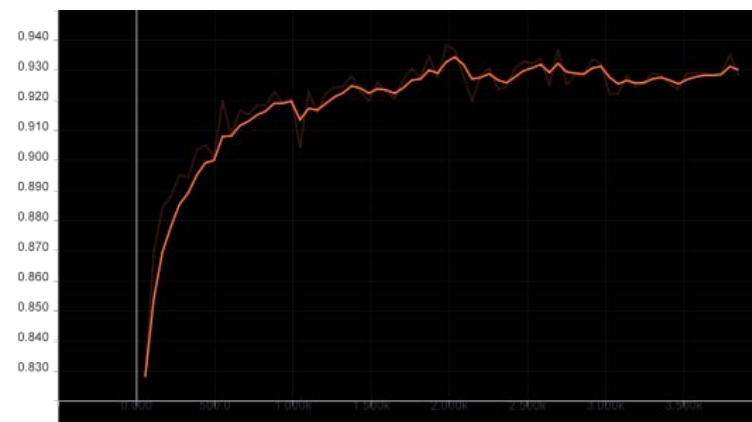


Fig. 5. Accuracy on the test set

The overall trend of Accuracy values on the training and validation sets is increasing as the number of steps increases, starting to grow faster and then gradual. After 70 rounds of training, the Accuracy on the training set was 0.9801 and the Accuracy on the verification set was 0.9283.

The overall trend of the Loss values on the training and validation sets is decreasing as the number of steps increases, starting to grow faster and then gradual. The Loss on the validation set exploded after approximately 2695 steps and then stabilized. After the 70 rounds of training, the Loss on the training set was 0.06325, and the Loss on the verification set was 0.2680.

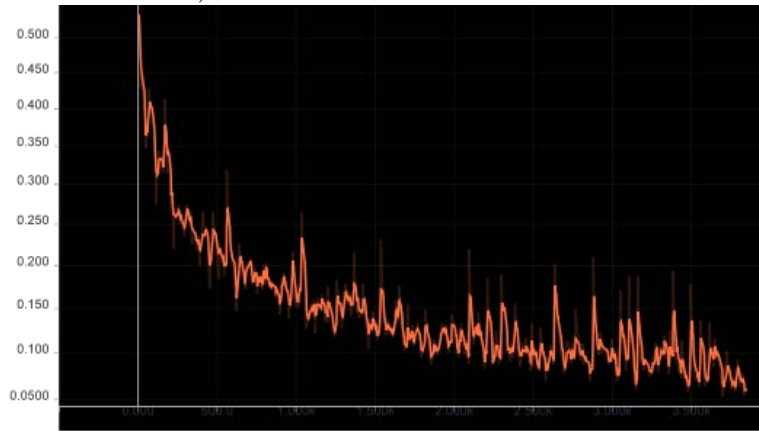


Fig. 6. Loss on the train set

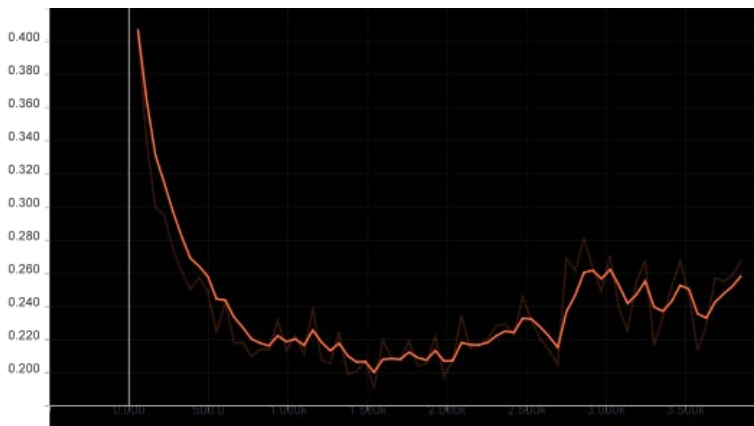


Fig. 7. Loss on the validation set

The confusion matrix of our experimental results is shown below:

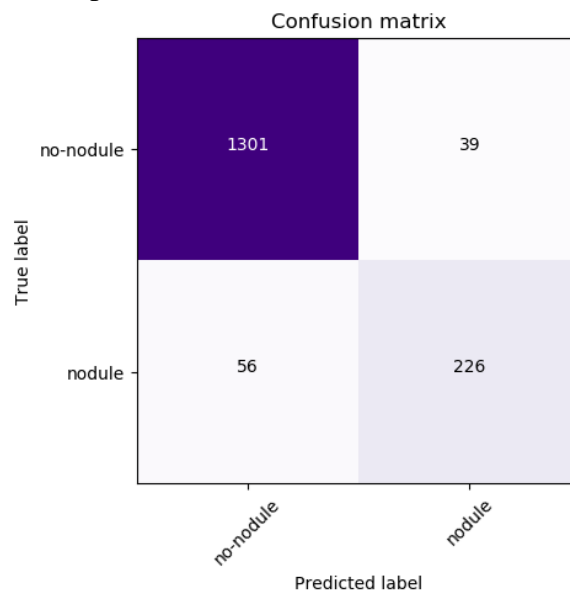


Fig. 8. Confusion marix of results

From above the matrix, the model has a precision of 85.28% and ecall of 80.14%. The model has a specificity of 97.09%.

The ROC curve of our experimental results is shown below:

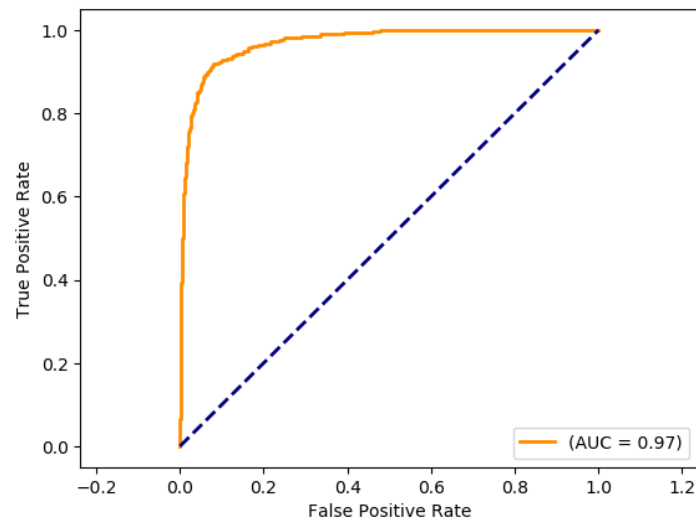


Fig. 9. The ROC curve of our experimental results

5. Conclusion

According to this Fig.9, the AUC value is 0.97. First, the AUC value is a probability value. When you randomly select a positive sample and a negative sample, the probability that the current classification algorithm ranks the positive sample in front of the negative sample according to the calculated Score value is the AUC value. The larger the AUC value, the more likely the classification algorithm is to place a positive sample in front of the negative sample, so that it can be better classified. As we can see from the Fig.9, our binary classifier works well.

References

- [1] K. Murphy, B. van Ginneken, A. M. R. Schilham, B. J. de Hoop, H. A. Gietema, and M. Prokop, "A large scale evaluation of automatic pulmonary nodule detection in chest CT using local image features and k-nearest-neighbour classification," *Medical Image Analysis*, vol. 13, pp. 757–770, 2009.
- [2] C. Jacobs, E.M. van Rikxoort, T. Twellmann, E. T. Scholten, P. A. de Jong, J. M. Kuhnigk, M. Oudkerk, H. J. de Koning, M. Prokop, C. Schaefer-Prokop, and B. van Ginneken, "Automatic detection of subsolid pulmonary nodules in thoracic computed tomography images," *Medical Image Analysis*, vol. 18, pp. 374–384, 2014. National Institute of Mental Health, "Any Anxiety Disorder," November 2017.
- [3] A. A. A. Setio, C. Jacobs, J. Gelderblom, and B. van Ginneken, "Automatic detection of large pulmonary solid nodules in thoracic CT images," *Medical Physics*, vol. 42, no. 10, pp. 5642–5653, 2015.
- [4] E. M. van Rikxoort, B. de Hoop, M. A. Viergever, M. Prokop, and B. van Ginneken, "Automatic lung segmentation from thoracic computed tomography scans using a hybrid approach with error detection", *Medical Physics*, vol. 4236 no. 10, pp.2934-2947, 2009.
- [5] McCulloch, Warren S.; Pitts, Walter. A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics*. 1943-12-01, 5(4): 115–133.
- [6] Graves, Alex; and Schmidhuber, Jürgen; Offline Handwriting Recognition with Multidimensional Recurrent Neural Networks, in Bengio, Yoshua; Schuurmans, Dale; Lafferty, John; Williams, Chris K. I.; and Culotta, Aron (eds.), *Advances in Neural Information Processing*

Systems 22 (NIPS'22), 7–10 December 2009, Vancouver, BC, Neural Information Processing Systems (NIPS) Foundation, 2009, pp. 545–552.

[7] Hinton, G.E., “Deep Belief Nets,” 2017 NIPS Tutorials Notes, Neural Information Processing Systems Conference, Vancouver, BC, December 2007.

[8] Bengio, Y. and Lecun, Y, “Scaling Learning Algorithms Towards AI,” In Large Scale Kernel Machines, 2007a.

[9] Delalleau, O. and Bengio, Y., “Shallow vs. Deep Sum-Product Networks,” In NIPS, 2011.

[10] LeCun, Y., Jackel, L.D., Boser, B., Denker, J.S., Graf, H.P., Guyon, I., Henderson, D., Howard, R.E., and Hubbard, W., “Handwritten Digit Recognition: Application of Neural Network Chips and Automatic Learning,” IEEE Communication Magazine, 27(11), 41-46, 1989.

[11] Chen, B., Ting, J. A., Marlin, B. M., and de Freitas, N., “Deep Learning of Invariant Spatial-Temporal Features from Video,” NIPS*2010 Deep Learning and Unsupervised Feature Learning Workshop, 2010.

[12] Kingma, D. and Ba, J., “Adam: A Method for Stochastic Optimization,” arXiv preprint arXiv:1412.6980, 2014.