American Scientific Research Journal for Engineering, Technology, and Sciences (ASRJETS)

ISSN (Print) 2313-4410, ISSN (Online) 2313-4402

http://asrjetsjournal.org/

Five Classifications of Mammography Images Based on Deep Cooperation Convolutional Neural Network

TANG Chun-ming[a]*, CUI Xiao-mei[b], YU Xiang[c], YANG Fan[d]

[a]*School of Artificial Intelligence Institute, Tianjin Polytechnic University, Tianjin, 300387,CHINA*

[b]*School of Electronics and Information Engineering, Tianjin Polytechnic University, Tianjin, 300387,CHINA*

[c]*Country Center for Engineering Practice and Training, Tianjin Polytechnic University, Tianjin, 300387, CHINA*

[d]*Union hospital Tongji Medical College Huazhong University of Science and Technology, Wuhan, 430030, CHINA*

[a]*Email: tangchunming@tjpu.edu.cn*

[b]*Email: 1731095297@stu.tjpu.edu.cn*

[c]*Email: 1922489185@qq.com*

[d]*Email: fyang@vip.163.com*

Abstract

Mammography is currently the preferred imaging method for breast cancer screening. Masses and calcification are the main positive signs of mammography. Due to the variable appearance of masses and calcification, a significant number of breast cancer cases are missed or misdiagnosed if it is only depended on the radiologists' subjective judgement. At present, most of the studies are based on the classical Convolutional Neural Networks (CNN), which uses the transfer learning to classify the benign and malignant masses in the mammography images. However, the CNN is designed for natural images which are substantially different from medical images. Therefore, we propose a Deep Cooperation CNN (DCCNN) to classify mammography images of a data set into five categories including benign calcification, benign mass, malignant calcification, malignant mass and normal breast. The data set consists of 695 normal cases from DDSM, 753 calcification cases and 891 mass cases from CBIS-DDSM. Finally, DCCNN achieves 91% accuracy and 0.98 AUC on the test set, whose performance is superior to VGG16, GoogLeNet and InceptionV3 models. Therefore, DCCNN can aid radiologists to make more accurate judgments, greatly reducing the rate of missed and misdiagnosis.

Keywords: Mammography; Breast cancer screening; CNN; Deep Cooperation CNN.

---

* Corresponding author.

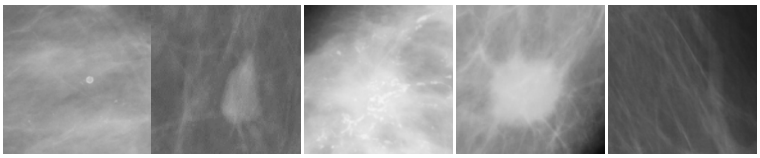
1. Introduction

Breast cancer is the most common cancer and it is also one of the main causes of female death [1]. Smoking, drinking, obesity, excessive stress and other factors are the main causes of breast cancer in women. In China, the incidence of breast cancer is increasing year by year, about 270,000 new cases appearing each year [2]. Breast cancer screening helps people detect potential breast cancer early. Therefore, early breast cancer screening increases the recovery chances for breast cancer patients. Breast cancer screening methods include ultrasonography, mammography, and MRI. Mammography is the internationally recommended method for diagnosing early breast cancer in recent years [3].The result of breast cancer screening shows that screening mammography can reduce breast cancer mortality by 38–48% [4]. However, classification of lesions in mammography images is a challenging job, so mammography images are usually examined by one or two experienced radiologists to detect whether malignant lesions exist. Suspicious cases will then be recalled for further diagnostic evaluation. Mammography reading is boring, time consuming, and expensive. According to studies [5, 6], the average diagnostic error rate of radiologists is 30%. Therefore, it is important to improve the accuracy of breast cancer examination. In order to help radiologists improve the accuracy of early breast cancer screening, researchers have done a lot of related work. Prathibha and his colleagues [7] perform Multi-resolution transform on the Region Of Interest(ROI) extracted from the DDSM data set, and they design a CNN, which consists of two convolutional layers and four fully connected layers, with images size of 128×128. Its accuracy is up to 85.4%. Levy and his colleagues [8] use transfer learning in AlexNet and GoogLeNet to solve this classification problem. AlexNet consists of five convolutional layers and three fully connected layers, and its accuracy is up to 89%. GoogLeNet consists of 22 layers and uses the inception module to fuse different scale features, whose accuracy is up to 92.9%. Soriano and his colleagues [9] use transfer learning in the InceptionV3 model. They adjust the following parameters: batch size, the number of frozen layers, learning rate, three optimizers (Adadelta, RMSProp and SGD) and two loss functions (categorical cross entropy and mean squared error). It achieves 85% accuracy. Chougrad and his colleagues [10] implement a series of experiments on three models: VGG16, RestNet50, and InceptionV3 via transfer learning on the public dataset DDSM. The architecture of VGG16 is deep and simple, but it is very expensive in terms of memory and computational cost. RestNet50 can be trained deeper after it adds shortcut connections that bypass few convolutional layers at a time. The shortcut connections create residual blocks, where the output of the convolutional layers is added to the block's input tensor. The best AUC value on the DDSM dataset is obtained on the InceptionV3, its value is up to 0.98. However, the mentioned studies just classify the masses in the breast lesions. Mass-type breast cancer in clinical practice is easier to diagnose, but non-mass type breast cancer is often missed. Calcification is usually an important feature of non-mass breast cancer. Therefore, we classify the mammography images into five categories: benign calcification, benign mass, malignant calcification, malignant mass and normal breast. It has a great value in clinic. Only one published paper [11] we found studying on five classifications for mammography images. Based on the standard VGG16 architecture, Quan and his colleagues extract the ROI from the Curated Breast Imaging Subset of DDSM (CBIS-DDSM). Transfer learning is performed with network weight initialized with the ImageNet pretrained VGG16 network, the best AUC is 0.88 on the test set. In this paper, the mammography images are also classified into five categories. A novel Deep Cooperation Convolutional Neural Network (DCCNN) is designed, which does not require manual

design features. In addition, it uses parallel structures to automatically learn images' features from low to high levels. For the two branches of DCCNN, they do not share weights, so the extracted features are different. Fully connected layers connect the two different output features may improve the classification accuracy, especially to calcification, which is difficult to a single network. This paper is organized as following: in section 2, we introduce the data set and preprocessing process. In section 3, we present the details of the network design, parameter analysis and evaluation indicators. In section 4, we give the experimental results, and evaluate the proposed method via comparison. A brief discussion is given in Section 5.

2. Data sets and preprocessing

The data sets we used are the Digital Database of Screening Mammography (DDSM) [12] and the Curated Breast Imaging Subset of DDSM (CBIS-DDSM) [13]. The negative data from DDSM and includes 695 normal cases. Positive data from CBIS-DDSM includes 753 calcification cases and 891 mass cases. The images from DDSM are randomly cropped into size of 598×598, then resized to 299×299. The images from CBIS-DDSM were extracted using a mask with a small amount of padding to provide context, then each ROI was randomly cropped three times to a size of 598×598 image, and then resized to 299×299. We separate these data into training set (90%) and testing set (10%) randomly. Training in supervised learning requires a large number of labeled dataset. Therefore, we use the data enhancement method [14] to expand the training set via randomly flipping and rotating. Figure 1 shows the samples of benign calcification, benign mass, malignant calcification, malignant mass and normal breast after enhanced.



(a) (b) (c) (d) (e).

Figure 1: The five categories of the data sets are: (a) benign calcification, (b) benign mass, (c) malignant calcification, (d) malignant mass, (e) normal breast

3. Network construction and parameter analysis

3.1. Network structure design

We design a Deep Cooperation Convolution Neural Network (DCCNN) for the classification of mammography images. As shown in Figure 2, DCCNN consists of two subnetworks. The first is the Feature Extraction Network, which consists of two independent Deep Convolutional Neural Networks (DCNN) with the same structure. The second is the Fully Connected Network, which connects the features extracted from two branches and contains four fully connected layers. The sizes of the fully connected layers are: 1024, 512, 32 and 5. The final layer is a softmax layer for five classifications. One advantage of this model is: in the parallel network, the parameters in the two branches are different after training, because the initialization of weights in each branch is random. Therefore, the fitting degree to features is different which can extract more various features than a

single network. It may improve the classification accuracy further. The other advantage is that parallel structure network reduces the computational burden of computer while maintains the high level of performance. As well as the aim of parallel structure with a computational budget balanced between its depth and width.

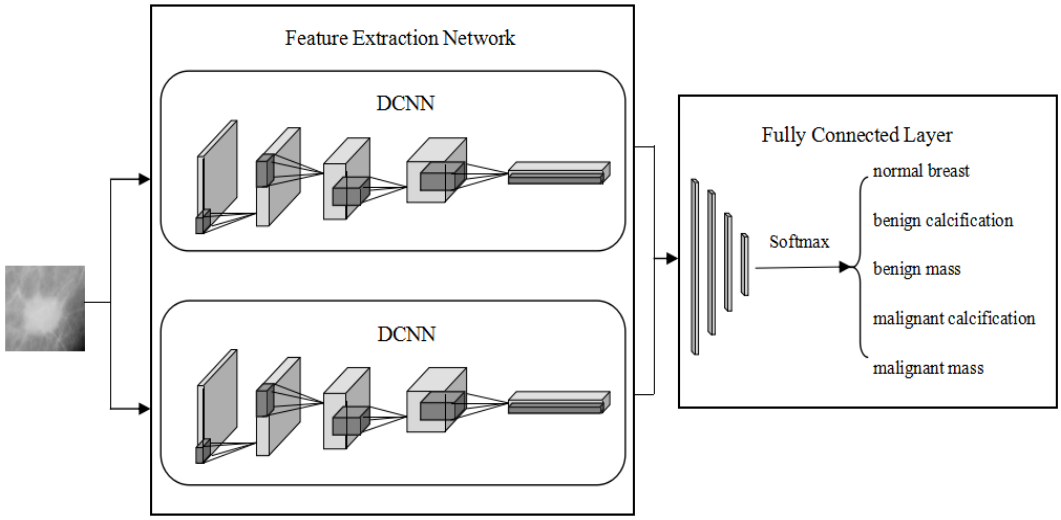


Figure 2: the framework of DCCNN

The structure of DCNN is shown in Figure 3. The input is 299×299 ROIs of mammography images. The ROI is a grayscale image. The sizes of the convolutional kernel and max pooling layer are designed respectively according to the sizes of the lesion areas, especially to very small calcification point. Therefore, the first part of our network only uses 3×3 convolution kernels and 2×2 pooling layers. The 3×3 convolution kernel extracts detail information of edges, texture shapes, and contours.

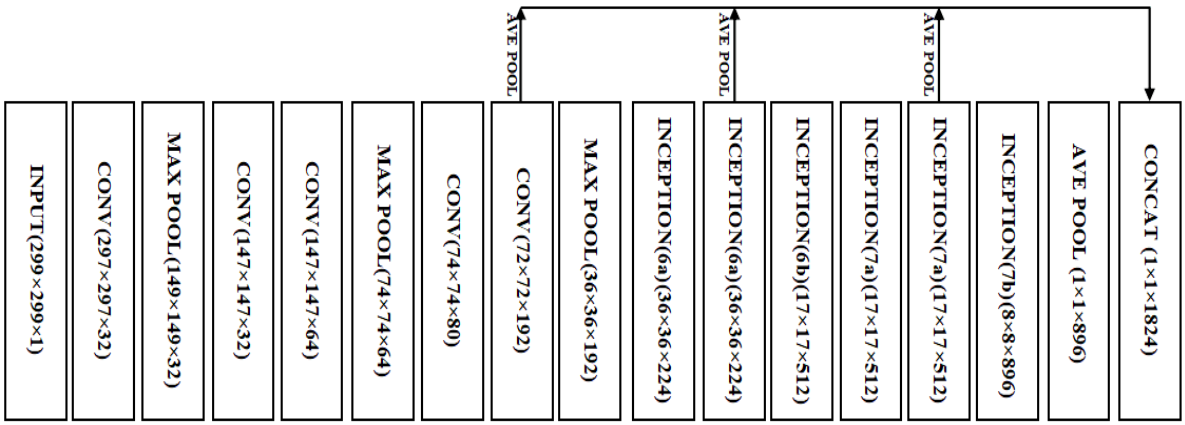


Figure 3: the structure and parameters of DCCNN. Concat layer fuses the middle-level information extracted from the fifth convolutional layer, inception (6a), and inception (7a) with high-level information of images

The main role of the 2×2 pooling layer is to reduce the feature map size, thereby reducing the computation time. At the same time, 1×1 convolution kernel is used to increase the nonlinearity of the network, so that the network can express more complex features. In DCNN, we also use the multi-scale fusion inception modules. Inception

(6a) uses two 3×3 convolution kernels instead of 5×5 convolution kernels, which reduces the number of parameters. Inception (7a) further uses 1×3, 3×1 instead of 5×5. In this way, calculation efficiency is further improved. Inception (6b) and inception (7b) fuse feature information with multiple scales and reduce feature maps size, thus reducing network overfitting. The DCNN not only fuses information on different scales in the same layer, but also fuses the middle-level information extracted from the fifth convolutional layer, inception (6a), and inception (7a) into high-level information. The details of the network are shown in Figure 4.

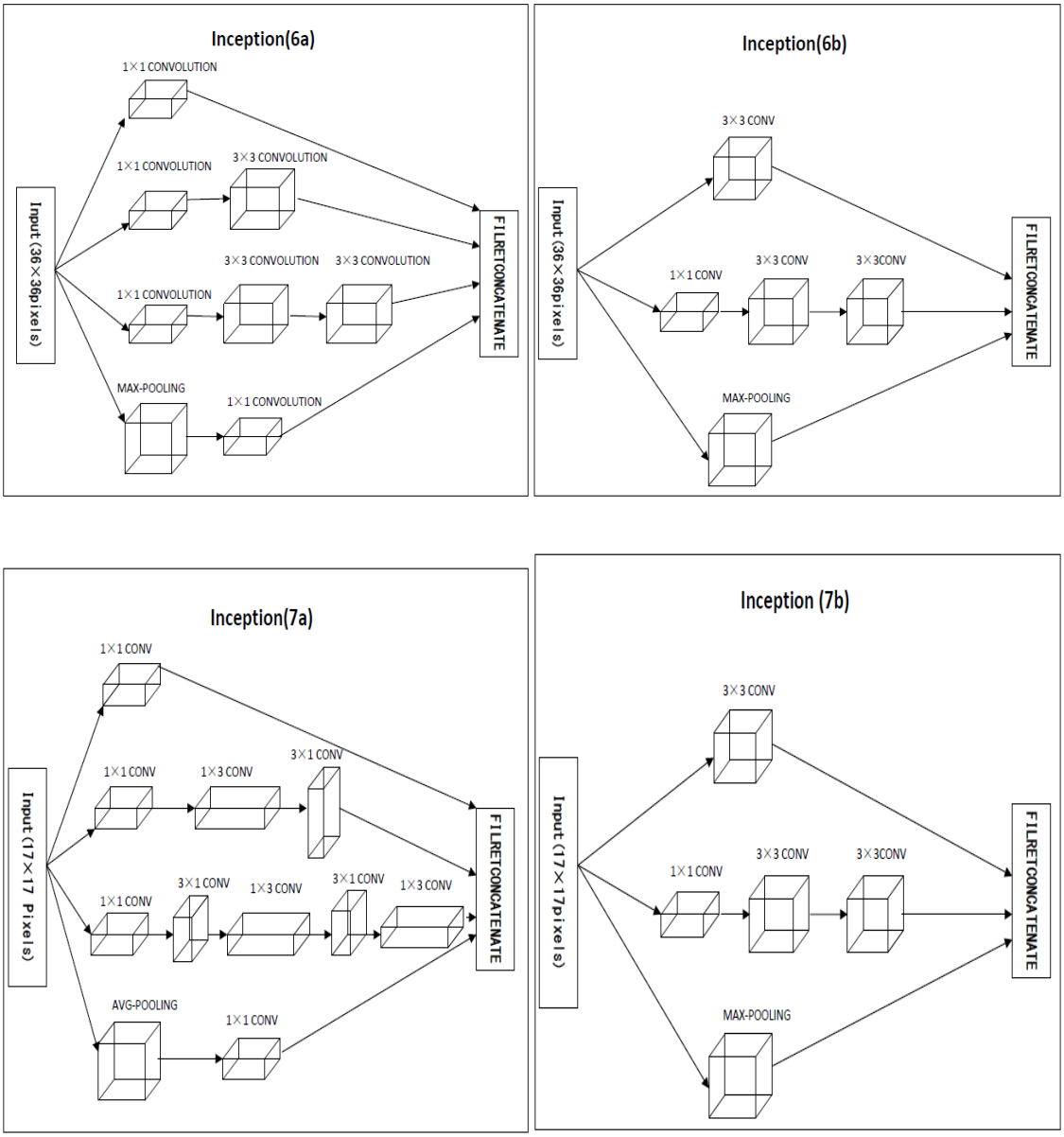


Figure 4: details of inception (6a), inception (6b), inception (7a), inception (7b) in DCNN

3.2. Analysis of network parameters

We additionally use batch normalization [15], 299×299×1 as an input, and initialize weights with truncated normal distribution. The learning rate is set to 1e-4. Loss function is categorical cross entropy [16], which is used to calculate the percentage of error when the neural network model tests the data. Adam [17] is used as the

optimizer. Rectified linear unit (ReLU) is used as activation function, which is defined by Equation 1:

$$F(x) = \begin{cases} x, & x \geq 0 \\ 0, & x \leq 0 \end{cases} \quad (1)$$

To ensure that our model generalizes well, L2 regularization (weight decay) is used to penalize large weights and prefer smaller ones. It reduces model complexity as much as possible. L2 regularization penalty operates on weight matrix W and is written as: $R(W) = \sum_i \sum_j W_{i,j}^2$, setting its parameter to 1e-6 ,because it gives the best result when tested.

3.3. Model performance evaluation metric

Model performance is the accuracy rate of the final classification. We choose two indicators to evaluate it: accuracy (ACC) and area under the receiver operating characteristic (ROC) curve (AUC). They are calculated using a confusion matrix which is a 2x2 array including four parameters: false positive (FP), false negative (FN), true positive (TP), and true negative (TN) [18] [19]. The accuracy is given by Equation 2. The horizontal and vertical coordinates of the ROC curve are the false positive rate (FPR) and the true positive rate (TPR), respectively, which are given by Equation 3 and Equation 4.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \times 100\% \quad (2)$$

$$False \quad positive \quad rate = \frac{FP}{TN + FP} \quad (3)$$

$$True \quad positive \quad rate = \frac{TP}{TP + FN} \quad (4)$$

4. Experimental results

4.1. DCCNN and DCNN, DCNN_deep test results

In addition to training the Deep Cooperation Convolutional Neural Network (DCCNN), we also train the single branch Deep Convolutional Neural Network (DCNN) in DCCNN and the DCNN's deepened model named DCNN_deep (which is formed by adding two modules with the same structure of inception (7a) based on DCNN). In order to test the influence of network structure on classification accuracy. The number of iterations is 15000. Figure 5 shows the ROC curves obtained by the three networks on the test set. It also shows the AUC of each class and the AUC of the five classes' average.

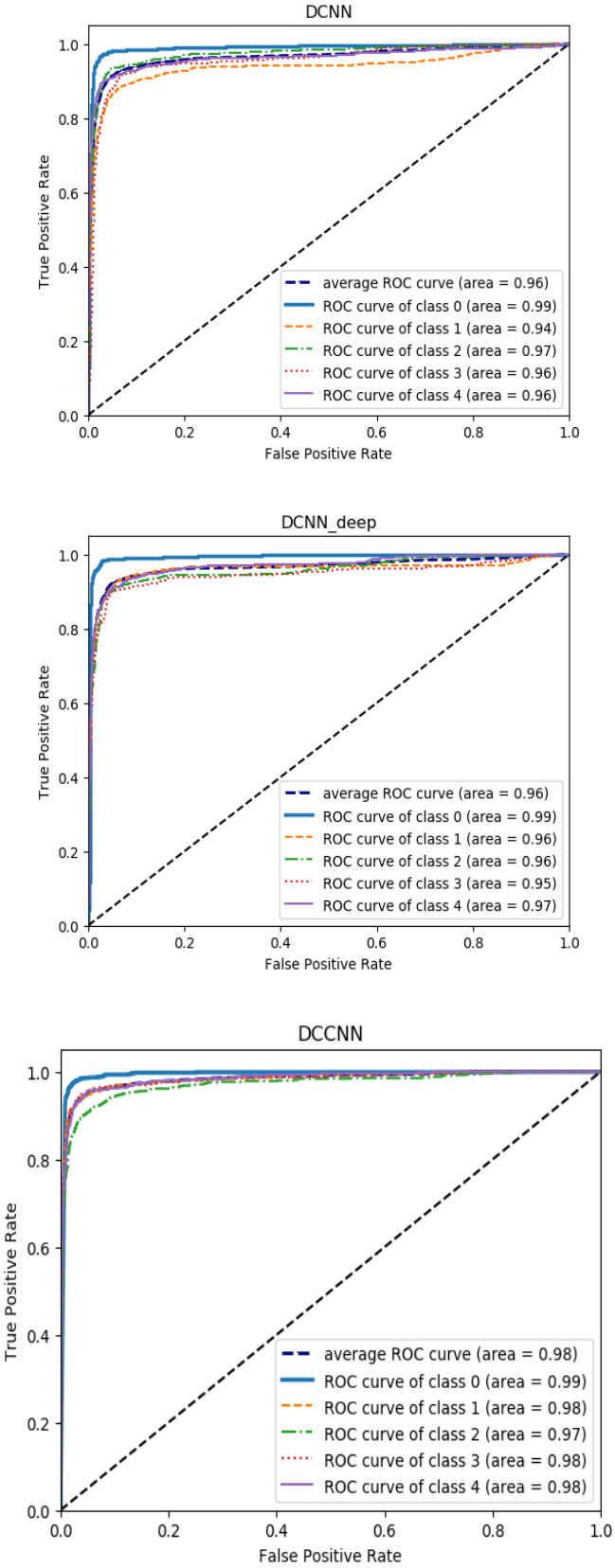


Figure 5: ROC curves of the DCNN, DCNN_deep and DCCNN. Class 0-4 relatively stand for normal breast, benign calcification, benign mass, malignant calcification and malignant mass. There are two solid lines, the thicker line is the ROC curve of class 0, the thinner line is the ROC curve of class 4.

It can be seen from Figure 5 that the average AUC value of the five types of DCNN reaches 0.96. Generally

speaking, the deeper the network, the better the classification's effect, so the DCNN is further deepened to obtain the DCNN_deep. Compared with the test results of DCNN, the DCNN_deep's AUC values of benign calcification and malignant mass are improved, but the classification effect of the other two types of lesions are reduced. Finally, the average AUC of the five categories is tested by DCNN_deep is still 0.96, which is not improved. Our aim is to improve the classification accuracy of the all five types of lesions as much as possible. From the experimental results of the DCCNN, the average AUC reaches 0.98, which is 2% higher than DCNN and DCNN_deep. The experimental result further verifies our idea: although the DCNN structures of the two branches in DCCNN are the same, but the fitting degree to data is different. And as a result, the fusion features extracted by the parallel structure can improve the classification accuracy. The improvement of experimental accuracy is determined by the structure of the network, rather than simply deepening or widening the network.

4.2. Test result of the classic network

We also compare our proposed model with other three classical networks: VGG16, GoogLeNet and InceptinV3. The output is just replaced by output 5 classes in these networks. And we also remove two auxiliary classifiers from GoogLeNet because these classifiers are found to effect on the experimental results. The number of iterations of the three networks are 15000. Figure 6 shows the ROC curves obtained by the three models on the test set. It also shows the AUC of each class and the AUC of the five classes' average.

Among the three classical networks, the AUC values obtained by VGG16, GoogLeNet and InceptionV3 on the test set are 0.95, 0.96 and 0.94 respectively. Obviously, GoogLeNet has the best classification results. Through the experiment, we can draw two conclusions: (1) For this data set, it is not the deeper the network is, the higher the classification accuracy will be.

InceptionV3's results are worse than GoogLeNet's. (2) Classical networks do not achieve the best classification results for all datasets. Our Designed DCCNN clearly outperforms GoogLeNet for the classification of mammography images.

The classical network is designed for natural images classification. Mammography images are more complex than natural images, and they requires specially designed network. Therefore, we design the DCCNN, which improves the classification accuracy.

Figure 7 shows the classification accuracy of VGG16, GoogLeNet, InceptionV3 and DCCNN on the test set. Their accuracy are: 81%, 86%, 87% and 91% respectively. Model training and testing were performed on the NVidia GeForce GTX 1080 graphics card.

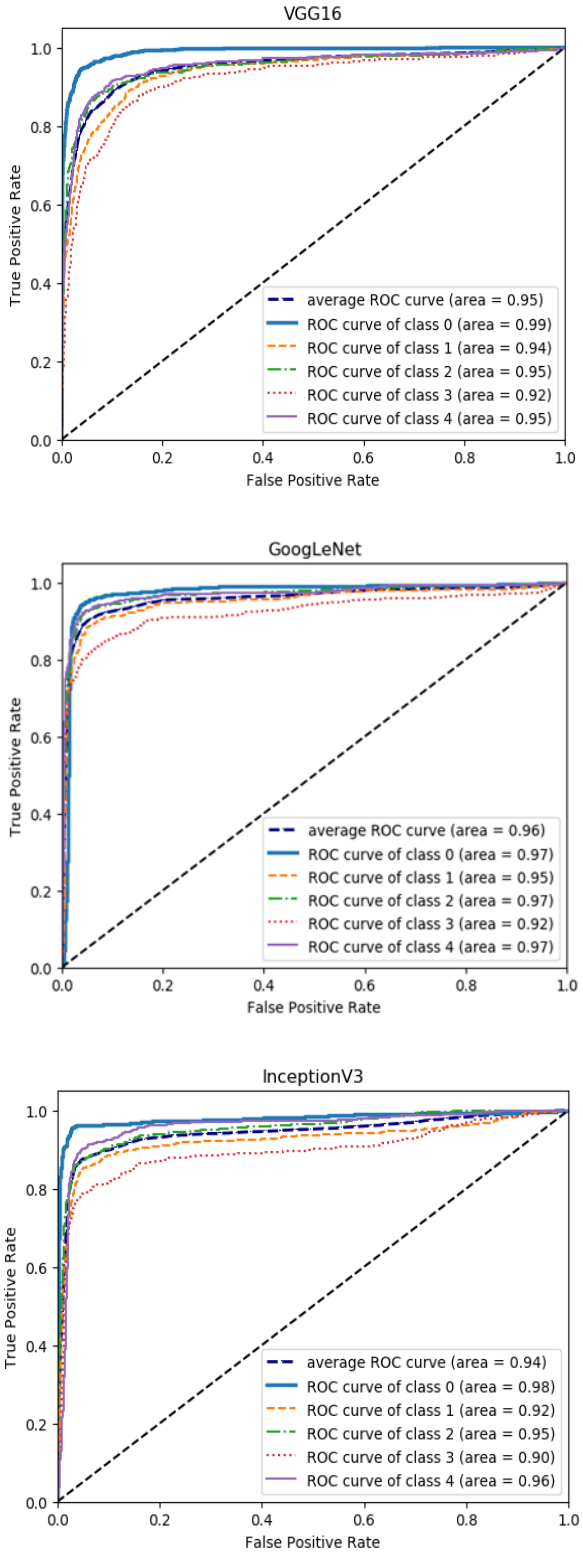


Figure 6: ROC curves of the VGG16, GoogLeNet and InceptinV3. Class 0-4 relatively stand for normal breast, benign calcification, benign mass, malignant calcification and malignant mass. There are two solid lines, the thicker line is the ROC curve of class 0, the thinner line is the ROC curve of class 4.

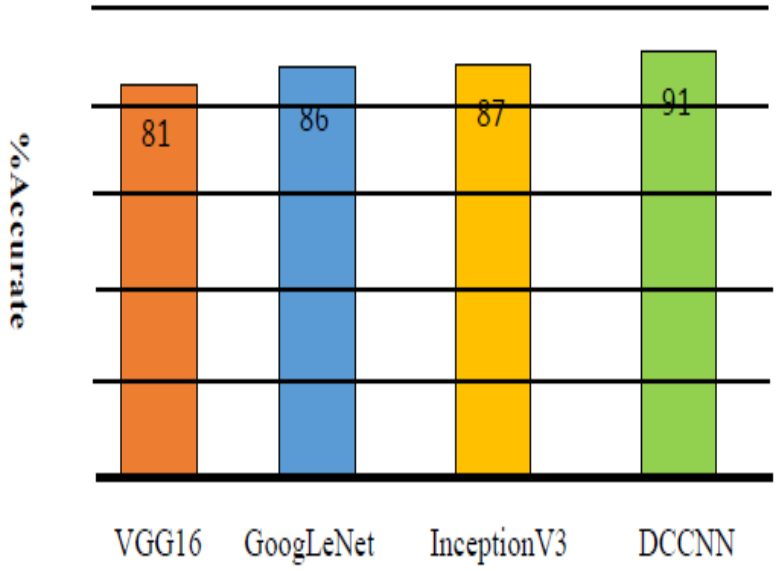


Figure 7: the classification accuracy of VGG16, GoogLeNet, InceptionV3 and DCCNN

5. Conclusion

The DCCNN structure designed in this paper achieves 91% accuracy and 0.98 AUC in test data set, compare with other popular CNN network structures, DCCNN has the highest accuracy and reliable performance. The results of experiment indicate that it can learn high level features those have large number of easily identifiable information. Therefore, DCCNN can aid radiologists to make more accurate judgments, greatly reducing the rate of missed and misdiagnosis.

6. Recommendations

We explore a variety of network structures and settings, and construct an end-to-end network structure for the classification of mammography images. We have learned that breasts are divided into four categories, namely: fat type, a small amount of gland type, a large number of gland type, and dense type. Among them, the fat type has the highest detection rate. The dense breast tissue and mass are white, and this situation is difficult for classification. Therefore, in the future work, we need to explore other structures and extend our approach to new mammography images data sets, selecting challenging pictures to learn, or adding other modes of images, such as: MRI (Magnetic Resonance Imaging), then it can learn the characteristics of cases with high breast density in many aspects. Aiding radiologists make more accurate judgments.

References

[1] Ferlay, Jacques, Clarisse Héry, Autier P , Clarisse Héry, Autier P. Global Burden of Breast Cancer. Breast Cancer Epidemiology, 2010, pp. 1–19.

[2] Chen, W., Zheng, R., Baade, P. D., Zhang, S., Zeng, H., & Bray, F., et al. "Cancer statistics in china,

2015,"Ca Cancer J Clin, vol. 66, no. 2, pp. 115-132, 2016.

[3] Schreer, Ingrid. "Dense Breast Tissue as an Important Risk Factor for Breast Cancer and Implications for Early Detection," Breast Care, vol. 4, no. 2, pp. 89-92, 2009.

[4] Moss, S. M., Nystrom, L., Jonsson, H., Paci, E., Lynge, E., & Njor, S., et al. "The impact of mammographic screening on breast cancer mortality in Europe: a review of trend studies," Journal of Medical Screening 19.Supplement 1, pp.26-32, 2012.

[5] Kerlikowske, and Karla. "Performance of Screening Mammography among Women with and without a First-Degree Relative with Breast Cancer," Annals of Internal Medicine, vol. 133, no. 11, pp.855, 2000.

[6] Berlin, and Leonard. "Radiologic errors, past, present and future." Diagnosis, vol. 1, no. 1, pp.79-84, 2014.

[7]Prathibha,G., Dr.Chandra,B. "Mammograms classification using Multiresolution Transforms and Convolution Neural Networks." In 9th International Conference on Computing, Communication and Networking Technologies, 2018.

[8] Lévy, Daniel, and A. Jain. "Breast Mass Classification from Mammograms using Deep Convolutional Neural Networks.", 2016.

[9] Danny Soriano, Carlos Aguilar, Ivan Ramirez-Morales, Eduardo Tusa, Wilmer Rivas, Maritza Pinta. "Mammogram Classification Schemes by Using Convolutional Neural Networks,"2017.

[10] Chougrad, Hiba, H. Zouaki, and O. Alheyane . "Deep Convolutional Neural Networks for Breast Cancer Screening." Computer Methods and Programs in Biomedicine, 2019.

[11] Quan, C., Jinze L., Kyle L., Xiaofei Z., XiaoqinW. "Transfer deep learning mammography diagnostic model from public datasets to clinical practice: a comparison of model performance and mammography datasets,"in 4th International Workshop on Breast Imaging, 2018.

[12] Heath, M., Bowyer, K., Kopans, D., Moore, R., & Kegelmeyer, P. "The digital database for screening mammography,"2001.

[13] Lee, R.S., Gimenez, F., Hoogi, A. and Rubin, D. "Curated Breast Imaging Subset of DDSM,"The Cancer Imaging Archive, 2016.

[14] Cubuk, E. D., Zoph, B., Mane, D., Vasudevan, V., & Le, Q. V. "Autoaugment: learning augmentation policies from data,"2018.

[15] Ioffe, Sergey, and C. Szegedy. "Batch normalization: accelerating deep network training by reducing

internal covariate shift," International Conference on International Conference on Machine Learning JMLR.org, 2015.

[16] Trujillano, Javier. "Aproximación a la metodología basada en árboles de decisión (CART). Mortalidad hospitalaria del infarto agudo de miocardio," Gaceta Sanitaria, vol. 22, no. 1, pp. 65-72, 2008.

[17] Kingma,D. P., & Ba, J. " Adam: a method for stochastic optimization. Computer Science, "2014.

[18] Leal,Y.,Gonzalez-Abril, L.,Ruiz, M.,Lorencio, C.,Bondia, J.,Vehi, J. "Un nuevo enfoque para detectarmediciones de glucosaerroneasen los sistemas de monitorizacioncont inuos de glucose,", JARCA 2012, vol. 15, pp. 17, 2012.

[19] Fernandez,P. C.,Taladriz,C. C.,Alcaide,F. G.,Sacchi,L., Bellazzi,R., Aguilera, E, J, G. "Extraccion de reglas temporales complejas para la deteccion de fallos del tratamiento antiretroviral,"2008.