Classification of Pancreatic Tumors based on MRI Images using 3D Convolutional Neural Networks

Xuan Chen[1], Yufei Chen[1]*, Chao Ma[2], Xianhui Liu[1], Xin Tang[1]

1College of Electronic and Information Engineering, Tongji University, Shanghai 201804, China

2Department of Radiology, Changhai Hospital of Shanghai, Second Military Medical University, Shanghai 200433, China

april337@163.com

ABSTRACT

Computer aided diagnosis of pancreatic cancer can help doctors improve diagnostic efficiency and accuracy, which does not depend on the doctor's subjective judgment and experience. Existing pancreatic tumors classification methods suffer from the problem of partial automation and ignore spatial and temporal characteristics. In this paper, we used 3D versions of ResNet18, ResNet34, ResNet52 and Inception-ResNet for pancreatic magnetic resonance images (MRI) classification. In order to alleviate the effect of class imbalance, we proposed a weighted loss function. Two sets of comparative experiments were performed to compare the effect of the proposed loss function. Experimental results show with weighted loss function, the performances of the four models are mostly improved except the precision of ResNet52. Moreover, the false negatives are reduced. Among them, ResNet18 performs best and achieves the accuracy of 91%.

CCS CONCEPTS

•Computing methodologies→3D imaging

Keywords

Pancreatic tumors classification; 3D convolutional neural network; Magnetic resonance images; Weighted loss function.

1. INTRODUCTION

Pancreatic cancer is a common malignant pancreatic tumor. Its morbidity and mortality have increased significantly in recent years. Most traditional methods for automated classification of pancreatic tumors cannot work in an end-to-end way. Klawikowski *et al*. [1] used a research tool (IBEX) to do texture analysis and then manually modeled for classification. Li and Ngom [2] used high dimensional linear machine combined mass spectrometry technique to diagnose pancreatic cancer before carcinogenesis. J. Ko *et al*. [3] profiled the RNA cargo inside of these exosomes, and applied a machine learning algorithm to generate predictive panels that could identify samples derived from heterogeneous cancer-bearing individuals. With the advent of deep learning, Dmitriev *et al* [4] proposed an algorithm based on a Bayesian combination of a RF classifier and a Convolution Neural Network (CNN) to make use of both clinical information about the patient and fine imaging information from computed tomography (CT) scans. Their proposed method was reported a classification accuracy of 83.6%.

Above mentioned studies, problems of pancreatic tumors classification were that they were usually based on extraction of handcrafted features with training a classifier to distinguish malignant tumors or they performed 2D image analysis due to the small data sets collected but ignoring some global information.

In this paper, to determine the number of model layers that are appropriate for our MRI dataset, we use four different 3D CNN architectures which are the modifications of residual and inception residual CNN. The 3D residual CNN can tackle problems stated in the previous paragraph. Firstly, these networks can extract omics characteristics and advanced semantic features for classification. Secondly, residual network architectures relieve the headache of small training datasets. Also, we propose a weighted loss function to reduce false negatives and mitigate category imbalances. As a result, we apply an end-to-end deep learning model to 3D MRI images classification.

2. RELATED WORK

There are previous studies that used 3D CNNs for classifying other diseases and pancreas segmentation. These existing works, as follows, have proven the effectiveness of 3D CNNs in the field of medical image processing.

Hosseini-Asl *et al*. [5] used the Deeply Supervised Adaptive 3D-CNN (DSA-3D-CNN) which was initialized by training convolutional autoencoders for feature extraction and fine-tuning the network for Alzheimer's disease diagnostics on different domain images. Liu and Kang [6] explored a 3D multi-view CNN with both chain architecture and directed acyclic graph architecture, including 3D Inception and 3D Inception-ResNet, for the classification of lung nodules. Sergey Korolev *et al*. [7] proposed the modifications of a plain and residual CNN on the ADNI dataset for brain MRI scans classification. Mehrtash *et al*. [8] presented a 3D CNN method developed for the SPIE-AAPM-NCI PROSTATEx challenge, tailored for the task of classifying clinically significant prostate cancer findings in mpMRI.

Roth *et al*. [9] proposed a two-stage, coarse-to-fine approach that first used a 3D FCN to roughly define a candidate region, which would then be used as input to a second 3D FCN for the final segmentation. Zhu *et al*. [10] proposed a novel 3D-based coarse-to-fine framework for volumetric pancreas segmentation to tackle the limited amount of annotated 3D data but leveraged the rich spatial information along all three axes.

ISICDM 2018, October 13–14, 2018, Chengdu, China

ACM ISBN 978-1-4503-6533-8/18/10…$15.00
https://doi.org/10.1145/3285996.3286017

2.1 Method

We compare eight different approaches to pancreatic MRI classification with 3D CNNs. Taking into account the effect of the depth and width of the model, eight models are: ResNet18, ResNet34, ResNet52 and Inception-ResNet with binary crossentropy or the proposed weighted loss function, respectively. The workflow of our method is as follows in Figure 1:

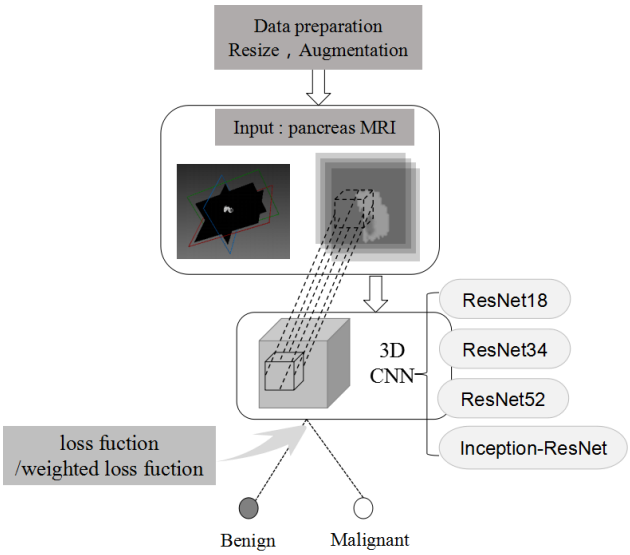


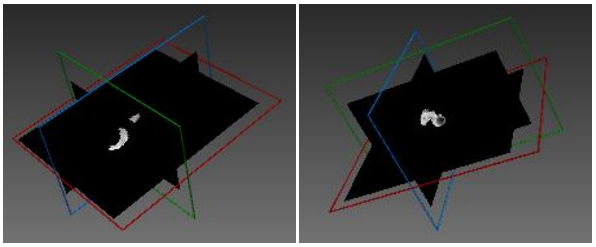
Figure 1. Workflow of our paper work

2.2 Data Preparation

Original MRI images are saved in NRRD format, as shown in Figure 2 and their size is different, as shown in Table 1.

Table 1: Original size of MRI images

320×260×64	288×234×40	512×512×116
256×192×52	256×192×48	256×208×44
512×512×72	512×512×25	512×512×80



(a) (b)

Figure 2. 3D visualization of pancreas (a) benign pancreas and (b) malignant pancreas

Firstly, we use data augmentation to amplify the data set. In our original data set, benign MRI images are more than twice as large as malignant MRI images. We use the nearest neighbor interpolation algorithm to interpolate in three directions to transform image resolution without changing intrinsic texture feature, the interpolated slices are showed in Figure 3. Then we use the zero-fill padding for adjusting to the uniform image size 128×128×128.

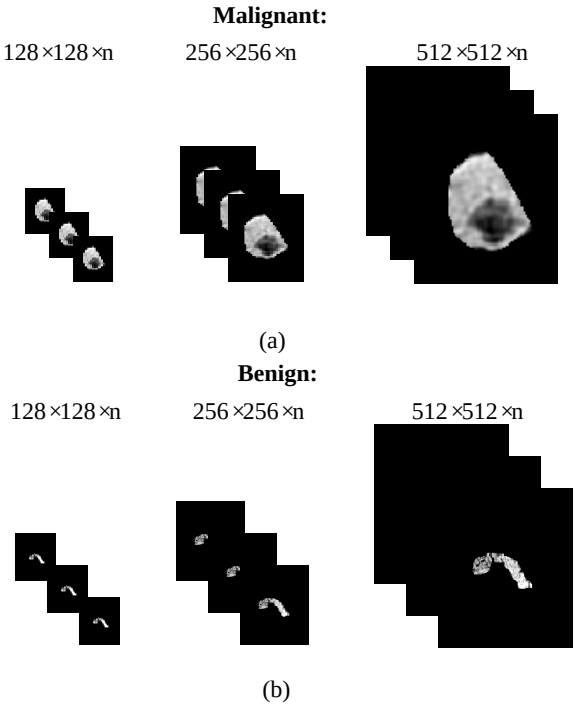


Figure 3. Visualization of pancreas after augmentation

2.3 3D Convolutional Neural Networks

3D ResNet Network

Residual neural networks [11] architecture won the ImageNet contest in 2015 and demonstrated possibility to greatly improve the depth of the network while having fast convergence. There were already publications that showed their results for 3D brain MRI image classification [7]. We use ResNet consisting of modules in Figure 4. As is shown in Figure 4(d), The output of the last ResNet module is sent to an average pooling layer to further reduce it to 1×1×1×512, followed by a flatten layer with 512 hidden units and a dense layer with an output for binary classification with sigmoid nonlinearity.

We use three ResNet architectures: ResNet18 contains 1 module (a), 4 modules (b), 3 modules (c) and 1 module (d); ResNet34 contained 1 module (a), 12 modules (b), 3 modules (c) and 1 module (d); ResNet52 contains 1 module (a), 20 modules (b), 3 modules (c) and 1 module (d). We train the final binary classification models using Adam with learning rate of 0.0001 and batchsize of 16 for 100 epochs.

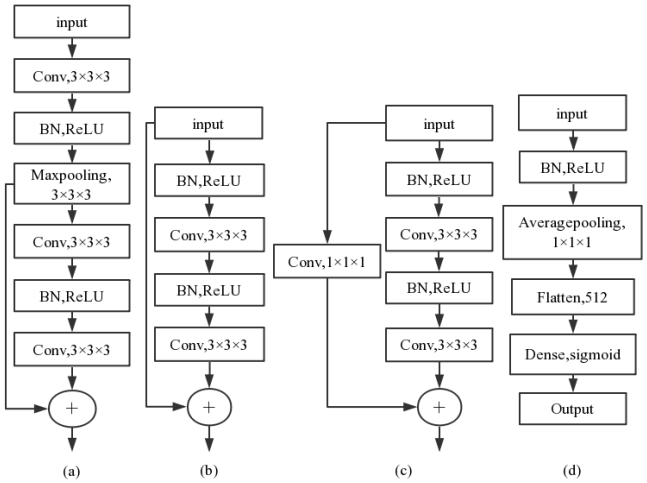
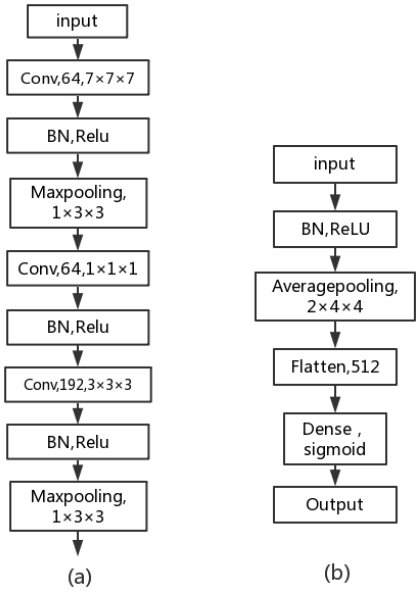


Figure 4. ResNet module: (a) is the first input module. (b) and (c) are the middle modules. (d) is the output module.

3D Inception-ResNet

Inception-ResNet was proposed by Google [12], in which the top-5 error rate for the ImageNet classification challenge was reduced to 3%, with an ensemble of three residual and one Inception-v4. We use Inception-ResNet architecture contains 1 module (a), 9 modules (c) and 1 module (b) (see in Figure 5). We insert maxpooling layer after the second and seventh inception3d modules to reduce the dimension of the layer output. As is shown in Figure 5(b), The output of the last Inception-ResNet module is sent to an average pooling layer to further reduce it to $15\times1\times1\times256$, followed by a flatten layer with 3840 hidden units and a dense layer with an output for binary classification with sigmoid nonlinearity.

We train the final binary classification models using Adam with learning rate of 0.0001 and batchsize of 8 for 100 epochs.



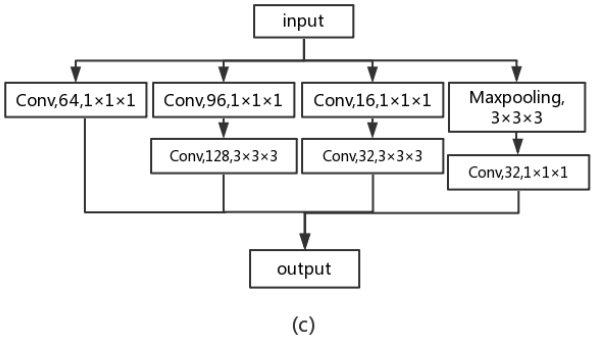


Figure 5. Inception-ResNet module (a) is the first input module. (b) is the output module. (c) is the inception_3d module.

2.4 A Weighted Loss Function

Considering that the data set used in the experiment is class-imbalanced. We give the sample with a smaller sample size a little more weight to weaken this imbalance problem on the basis of using data augmentation. We use the cross-entropy function of tensorflow and modify it to fit our model motivated by [13] which introduced a class-balancing weight on per-pixel term basis for image segmentation and [14] which optimized the cost of decision making. In our method, we just calculate the number of positive and negative categories every time the parameter is updated and iterated, then derive the weight coefficients for each category. The formula is as follows:

$$x1 = \frac{pos\_num}{neg\_num + pos\_num} \tag{1}$$

$$w1 = \frac{x1}{1 - x1} \tag{2}$$

$$loss = -(w_1 q \log p + (1-q)\log(1-p)), \tag{3}$$

where pos_num represents the number of positive samples each batch, neg_num represents the number of negative samples each batch, $w1$ represents the weight coefficient of the positive sample, q is the true value and p is the predict value.

3. EXPERIMENTAL RESULTS

3.1 Implement

Our models are implemented via Keras with the Tensorflow backend. We use early stopping to prevent overfitting. We train our models with 690 samples, and validate them on 173 samples. All experiments are accelerated using an NVIDIA GeForce GTX 1080 Ti (12 GB) GPU.

3.2 Data

MRI Images used in this work were collected from Changhai Hospital of Shanghai. The data set contains MRI images of 20 normal patients and 20 diseased patients. Each person has many groups of MRI images scanned by using different methods or different layer thickness, such as enhanced MRI and plain MRI. Thus, we have 77 benign MRI images and 38 malignant MRI images. Through data augmentation, the data we finally used for the experiment consist of 442 benign MRI images and 421 malignant MRI images.

3.3 Criterion

In order to evaluate the performance of the model, we use Accuracy, Precision, Recall, Area Under the Receiver Operating Characteristic Curve (ROC AUC Score) and F1 Score, as the evaluation criteria. They are calculated using the true positive (TP), true negative (TN), false negative (FN), and false positive (FP), as follows:

$$accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{4}$$

$$precision = \frac{TP}{TP + FP} \tag{5}$$

$$recall = \frac{TP}{TP + FN} \tag{6}$$

$$F1\ Score = \frac{2TP}{2TP + FP + FN} \tag{7}$$

The ROC AUC Score can better summarize the performance of the unbalanced sample classifier and become the criterion for many data mining competitions.

Especially, the recall score can tell us whether the number of the false negatives drop or not. We can learn from the variant of the formula (6), as the following:

$$TP + FN = N_{\text{malignant}}$$

$$recall = \frac{N_{malignant} - FN}{N_{malignant}}, \tag{8}$$

where $N_{\text{malignant}}$ is the total number of positive samples each batch, so if the recall score becomes larger, the FN will be reduced.

4. RESULTS

Since the original samples are unbalanced and we want to give more weight to the positive sample, we propose a weighted loss function mentioned in section 3.3. To prove the validity and rationality of the proposed loss function, we design two sets of comparison experiments. In each set of experiments, the effect of different depth and width on results is compared. The experimental results are shown in Figure 6 and Table 2.

Firstly, we classify the pancreatic MRI images with four models: ResNet18, ResNet34, ResNet52 and Inception-ResNet with binary crossentropy loss function. The classification results are measured by five criterions mentioned in section 4.3. The measurement is shown in Figure 6. It clearly indicates that the wider Inception-ResNet network performs the best, with the accuracy of 82%, followed by ResNet18. In our experiments, with the depth of residual network model deepened, the classification ability of the model first decreases and then increases, which indicates that the use of the original loss function leads to over-fitting and is not feasible in this experiment.

Then we carry out our experiments with the proposed loss function. The results are also shown in Figure 6, which are expressed in red. Note that the red part represents the increase in the original results. It can be clearly seen that each criterion of models has a certain degree of improvement, except the precision of ResNet52 reduces a little. The recall score close to 1.0 (see the formula (8)) means false negatives are reduced. Among them, ResNet18 performs the best with the accuracy of 91%, followed by ResNet34, Inception-ResNet ranks in third place and ResNet52 has the least performance improvement. The results show the proposed weighted loss function had a significant improvement on the classification results and also show that it is necessary to experiment multiple times to determine the appropriate depth and width of the model for our data set.

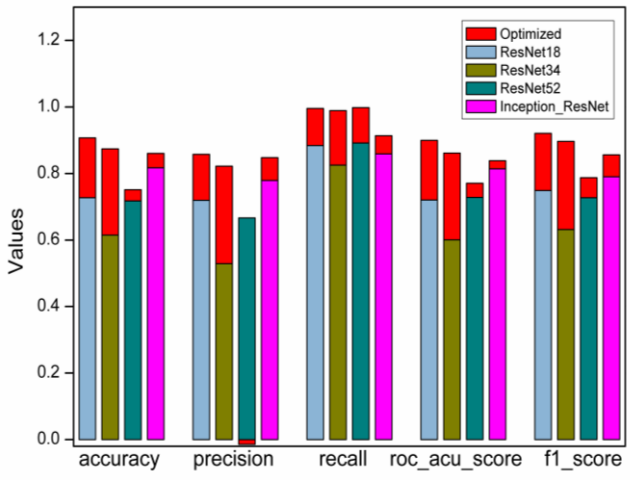


Figure 6: Overall evaluation of classification methods

A more detailed comparison is listed in Table 2. The standard deviation indicate that the proposed loss function with weight coefficient makes the model more robust.

Table 2. Scores of classification methods, * represents this method uses the proposed weighted loss function

Models	Accuracy	Precision	Recall	ROC AUC Score	F1 Score
ResNet18	0.73±0.14	0.72±0.18	0.88±0.27	0.72±0.14	0.75±0.2
ResNet18*	0.91±0.03	0.86±0.05	0.99±0.02	0.90±0.04	0.92±0.03
ResNet34	0.73±0.13	0.53±0.25	0.83±0.37	0.71±0.13	0.75±0.28
ResNet34*	0.87±0.04	0.82±0.05	0.98±0.02	0.86±0.04	0.90±0.03
ResNet52	0.72±0.12	0.67±0.17	0.89±0.26	0.73±0.12	0.73±0.19
ResNet52*	0.75±0.06	0.65±0.05	0.99±0.01	0.77±0.05	0.79±0.04
Inception-Res	0.82±0.13	0.78±0.22	0.86±0.27	0.81±0.13	0.79±0.23
Inception-Res*	0.86±0.10	0.85±0.11	0.91±0.20	0.84±0.12	0.86±0.14

In general, pancreatic cancer has individual differences. Some characteristics between benign and malignant are difficult to distinguish. The use of a new loss function improve performances on experimental results and reduces false negatives.

5. CONCLUSION

We used four 3D-CNN networks: ResNet18, ResNet34, ResNet52 and Inception-ResNet to do classification on pancreatic lesions. Experimental results and comparisons demonstrate that ResNet18 with the proposed weighted loss function method achieves the best results to classify the benign from the malignant on 3D MRI images, especially reducing false negatives which can reduce the rate of misdiagnosis in practical applications. In addition, when we use the appropriate weighted loss function, deepening the depth and width of the network cannot improve the classification accuracy of our small data sets. Our approach is a fully automated

classification method that is beneficial to auxiliary diagnosis of pancreatic cancer and takes advantage of the spatial information of the MRI image.

6. ACKNOWLEDGMENTS

This work was supported by the National Natural Science Foundation of China (No. 61573235), the Shanghai Innovation Action Project of Science and Technology (No. 17411952200), and the Fundamental Research Funds for the Central Universities.

7. REFERENCES

[1] S. Klawikowski, J. Christian, D. Schott, M. Zhang, and X. Li. 2016. Development of a CT Radiomics Based Early Response Prediction Model During Delivery of Chemoradiation Therapy for Pancreatic Cancer. Medical Physics 43, 6 (2016), 3350–3350.

[2] Yifeng Li and Alioune Ngom. 2012. Diagnose the premalignant pancreatic cancer using high dimensional linear machine. In *IAPR International Conference on Pattern Recognition in Bioinformatics*. 198–209.

[3] J. Ko, N. Bhagwat, S. S. Yee, N. Ortiz, A. Sahmoud, T. Black, N. M. Aiello, L. Mckenzie, M. O'Hara, and C. Redlinger. 2017. Combining Machine Learning and Nanofluidic Technology to Diagnose Pancreatic Cancer Using Exosomes. *Acs Nano* 11, 11 (2017), 11182.

[4] Konstantin Dmitriev, Arie E. Kaufman, Ammar A. Javed, Ralph H. Hruban, Elliot K. Fishman, Anne Marie Lennon, and Joel H. Saltz. 2017. Classification of Pancreatic Cysts in Computed Tomography Images Using a Random Forest and Convolutional Neural Network Ensemble. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*. 150–158.

[5] E Hosseini-Asl,MGhazal, AMahmoud, A Aslantas, A. M. Shalaby, M. F. Casanova, G. N. Barnes, G Gimel'Farb, R Keynton, and A El-Baz. 2016. Alzheimer's disease diagnostics by a 3D deeply supervised adaptable convolutional network. *Front Biosci* 23 (2016), 584–596.

[6] Kui Liu and Guixia Kang. 2017. Multiview convolutional neural networks for lung nodule classification. *Plos One* 12, 11 (2017), 12–22.

[7] Sergey Korolev, Amir Safiullin, Mikhail Belyaev, and Yulia Dodonova. 2017. Residual and plain convolutional neural networks for 3D brain MRI classification. In *IEEE International Symposium on Biomedical Imaging*.

[8] A Mehrtash, A Sedghi, M Ghafoorian, M Taghipour, C. M. Tempany, Wells Wm Rd, T Kapur, P Mousavi, P Abolmaesumi, and A Fedorov. 2017. Classification of Clinical Significance of MRI Prostate Findings Using 3D Convolutional Neural Networks. *Proc Spie Int Soc Opt Eng* 10134 (2017), 101342A.

[9] H. R. Roth, H. Oda, X. Zhou, N. Shimizu, Y. Yang, Y. Hayashi, M. Oda, M. Fujiwara, K. Misawa, and K. Mori. 2018. An application of cascaded 3D fully convolutional networks for medical image segmentation. *Computerized Medical Imaging & Graphics the Official Journal of the Computerized Medical Imaging Society* 66 (2018), 90.

[10] Zhuotun Zhu, Yingda Xia, Wei Shen, Elliot K. Fishman, and Alan L. Yuille. 2017. A 3D Coarse-to-Fine Framework for Automatic Pancreas Segmentation. (2017)

[11] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2015. Deep Residual Learning for Image Recognition. (2015), 770–778.

[12] Christian Szegedy, Sergey Ioffe, Vincent Vanhoucke, and Alex Alemi. 2016. Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning. (2016).

[13] Saining Xie and Zhuowen Tu. 2016. Holistically-Nested Edge Detection. In *IEEE International Conference on Computer Vision*. 3–18.

[14] Yufei Chen, Xiaodong Yue, Hamido Fujita, and Siyuan Fu. 2017. Three-way decision support for diagnosis on focal liver lesions. *Knowledge-Based Systems* 127, C (2017), 85–99.