Deep Model with Siamese Network for Viability and Necrosis Tumor Assessment in Osteosarcoma

Yu Fu

Abstract
Osteosarcoma is the most common primary malignant bone tumor, which has high mortality due to easy lung metastasis. Osteosarcoma is a highly anaplastic, pleomorphic tumor with a variety of tumor cell morphology, including fusiform, oval, epithelial, lymphocyte like, small round, transparent cells, etc. Due to the multiple patterns of osteosarcoma cell morphology, pathologists have differences in the classification (viable tumor, necrotic tumor, non-tumor) of osteosarcoma. Therefore, automatic and accurate recognition algorithms can help pathologists greatly reduce time and improve diagnostic accuracy. In recent years, deep learning technology has made great progress in the field of natural images and medical images, and has achieved excellent results beyond human performance in classification. In this paper, we propose a Deep Model with Siamese Network (DS-Net) for automatic classification in Hematoxylin and Eosin (H&E) stained osteosarcoma histology images. The proposed DS-Net architecture contains two parts: auxiliary supervision network and classification network. Auxiliary supervision network based on Siamese network is to increase the gap between classes and reduce the gap of intra-class. Classification network uses the information obtained by the auxiliary supervision network to classify. Compared to existing methods, our method can achieve the state-of-the-arts performance in classification with an average 97.13% accuracy. It shows that our proposed algorithm can effectively achieve the histological classification of osteosarcoma.

Keywords: Osteosarcoma, Classification, Deep Learning, Siamese Network

Introduction
Osteosarcoma which accounts for 19% of malignant bone tumors occurs mostly in adolescents with a high mortality rate although it is rare. Osteosarcoma has a high degree of heterogeneity which makes it difficult for pathologists to diagnose. Computer aided diagnosis (CAD) can help pathologists find lesions and improve the accuracy of diagnosis. Therefore, accurate osteosarcoma classification algorithms are very meaningful. Machine learning (ML) has been widely used in medical image analysis. ML algorithms identify manually extracted image features through various classification algorithms including K-means Clustering [1], Decision Tree Model (DTM) [2-3], Support Vector Machine (SVM) [4,5] and Multilayer Perceptron (MLP) [6], etc. Cuingnet R [7] proposed a method to detect differences at the group level in brain images based on SVM. Tamaki T [8] uses the SVM algorithm to detect lesions of colorectal tumors. Wong Lai [9] Wan proposed a novel SVM classifier for automated classification of emphysema, bronchiectasis and pleural effusion using optimized Gabor filter. Rajendran, P made [10] use of a accurate decision tree classification phase for classifying the brain images into normal, benign and malignant. Keir Bovis [11] extracted 70 features from mammograms and identified masses by MLP. The emergence of convolutional neural networks (CNN) has led to the application of deep learning in image analysis. Many network structures have been proposed on the classification of natural images, such as LeNet [12], AlexNet [13], VGG [14], GoogLeNet [15], ResNet [16], etc. In recent years, deep learning has been increasingly applied to the field of medical image analysis. Holger R [17] proposed methods improve CAD performance evaluated on three different datasets. Lihao Liu used ResNet with Siamese Network [18-19,31] for benign and malignant diagnosis of lung nodules [20].

Fig. 1. Examples of osteosarcoma histology images.

In this paper, we propose DS-Net for automatic classification of osteosarcoma images from TCIA [21]. Fig.1 shows some examples of data sets. There are some papers [22-25] about the histological classification of osteosarcoma. These algorithms are based on deep learning and have achieved good results. Our proposed algorithm uses Siamese Network for auxiliary supervision and achieves the best results.

Our main contributions are as follows:
1) We propose a novel deep learning-based algorithm for viable and necrotic tumor assessment in osteosarcoma. The algorithm combined with the auxiliary supervision network can effectively identify viable tumor, necrotic tumor and non-tumor.
2) We use the intra-class and inter-class information to train the auxiliary supervision network followed by a classification network, which greatly improves the convergence speed and performance of the model.
3) We validate our algorithm on the Osteosarcoma data from UT Southwestern and compare it with existing algorithms. The result shows that our algorithm achieves the best performance.

Method
In the data set of classification problems, there are not only direct category information, but also intra-class and inter-class information. This paper believes that these veiled messages can be used as weak supervision to influence the final classification results. Our proposed method consists of two tandem components. Firstly, the auxiliary supervision network is used to extract the primary features, which are intra-class similar features and inter-class differences. The auxiliary supervision network accepts two different data as
input at the same time. If the input data is of the same class, the given label is 0. If the input data is not of the same class, the given label is 1. Then the auxiliary supervision network is trained through the data pairs generated by the above strategies. Secondly, the classification network uses the features extracted by the auxiliary supervision network to classify. Experiments show that the classification effect can be improved by using the features extracted by the auxiliary supervision network, and the convergence speed will be accelerated at the model training stage. The entire network architecture is shown in Fig. 2.

### A. Auxiliary supervision network

Convolution neural networks (CNN) can effectively process image data which can be described as follows:

\[ y = \Phi(b + \omega x) \]  

(1)

where \( y \) is the output of the CNN, \( x \) is the input, \( \omega \) is the weight vector, \( b \) is the offset and \( \Phi \) is the activation function which we use is Rectified Linear Unit (ReLU).

\[ \Phi_{\text{ReLU}}(x) = \max(x, 0) \]  

(2)

Nowadays, there are many feature extraction network structures, and residual networks are widely used due to good performance. Since the residual network has a jump connection structure, not only can the information of different layers be merged, but also the gradient can be prevented from disappearing. Inspired by ResNet, this paper uses dilated convolution to construct an improved residual block called DRB. On the one hand, the dilated convolution can increase the receptive field and obtain more features. On the other hand, the model size remains unchanged. As a basic module of feature extraction, DRB can be cascaded to form a primary feature extraction network. As shown in Fig. 2, the feature maps obtained by ordinary convolution and dilated convolution operations are fused in depth, and then added directly to the input as the DRB output after passing through a layer of convolution. We build the DRB-Network beginning with a convolution by connecting a series of DRB in series and make a jump connection ending with a max-pooling layer. The size of the feature map will be halved after a DRB-Network. Auxiliary supervision network is mainly composed by three unequal scale DRB-Networks.

### B. Classification network

Classification network mainly consists of two DRB-networks, and replaces the fully connected layer with a global average pooling (GAP) layer which can let the network accept inputs of any size. The Classification network is trained by minimizing the cross-entropy loss of the auxiliary supervision network.

\[ \mathcal{L}_{AN} = -\sum_i \alpha \cdot y_i \cdot \log p_i(y_i|x; \theta) \]  

(3)

where \( \alpha \) is always equal to 1, then the formula (1) is the ordinary cross entropy (CE). Randomly select two inputs, the probability of the difference classes is twice the probability of same class for the three classifications. Therefore, when the input pairs are of the same class, let the parameter \( \alpha \) be 2 to balance the loss function.

\[ \alpha = \begin{cases} 1, & i = 1 \\ 2, & i = 0 \end{cases} \]  

(4)

Fig. 2. The architecture of DS-Net. A pair of input data is passed into the DS-Net through the auxiliary supervisory network. The blue line indicates that a pair of data sharing weights are processed simultaneously. Only in the auxiliary supervisory network, there is a pair of data transfer, and the other parts are not. AN_loss is the WCE loss of the auxiliary supervisory network and Loss is the total loss of DS-Net. N represents the number of DRBs in a DRB-Network. Notes that in the test, the model has only one input, and the data after the third DRB-Network is only passed to the fourth DRB-Network not to the convolution layer of the auxiliary supervisory network.
batch normalization (BN) layer after convolution and L2 regularization to the entire network. The specific configuration of the network is shown in the Table I. The whole loss function can be defined as:

$$L_{\text{total}} = \beta L_{\text{CN}} + \gamma L_{\text{AN}} + \lambda L_2$$

(5)

$$L_{\text{CN}} = -\sum y_i \log p_i(x; \theta)$$

(6)

where $\beta, \gamma, \lambda$ are three hyper parameters used balancing $L_{\text{CN}}, L_{\text{AN}}$ and L2 loss, $\theta$ is the parameter of the classification network and $x$ is the input.

### TABLE I

Detailed structure of DS-Net. DRB-Network3 is connected to Convolution1 and DRB-Network4 respectively, and serves as the input of the auxiliary supervisory network and the input of the classification network. “Number of DRB” represents the number of DRB in a DRB-Network.

| Net Name          | Subnetwork      | Number of DRB | Output Size   |
|-------------------|-----------------|---------------|---------------|
| **Auxiliary supervision network** |                 |               |               |
| DRB-Network1      | 4               | 128×128×8     |
| DRB-Network2      | 8               | 64×64×16      |
| DRB-Network3      | 8               | 32×32×32      |
| Convolution1      |                 | 32×32×2       |
| **Classification network** |                 |               |               |
| DRB-Network4      | 8               | 16×16×64      |
| DRB-Network5      | 4               | 8×8×128       |
| Convolution       |                 | 8×8×64        |
| max-pooling       |                 | 4×4×64        |
| Convolution2      |                 | 4×4×32        |
| Convolution3      |                 | 4×4×2         |
| GAP               |                 | 1×2           |

### Experiment

**A. Dataset and preprocessing**

The dataset (Osteosarcoma data) [26] we use was collected by a team of clinical scientists at University of Texas Southwestern Medical Center which is composed of H&E stained osteosarcoma histology images. The dataset consists of 1144 images of size 1024×1024 at 10X resolution with the following distribution: 536 (47%) non-tumor images, 263 (23%) necrotic tumor images and 345 (30%) viable tumor tiles. And it provides 66 ML features for each image. Among 345 viable tumor images, there have 53 both active tumors and necrotic tumor images, which will be removed in our experiments. So the final dataset contains 1091 images. As shown in Fig. 3, we visualize 1091 images using the t-SNE [27] algorithm. In the visualization, red points belong to class “0”, green points belong to class “1” and blue points belong to class “2”. T-SNE belongs to manifold learning [28-29] and is a method of nonlinear dimensionality reduction. It can map data in high-dimensional space to low-dimensional space and preserve the local characteristics. T-SNE algorithm is often used to reduce dimensionality of high-dimensional data and visualize it. As can be seen from Fig. 3, the original images are irregular and difficult to distinguish. There are some abnormal points in the figure, which may be caused by the large difference between these samples and the similar sample features. These samples cannot find a sample that is close to them. The existence of these samples increases the difficulty of classification. A good classification algorithm can effectively distinguish these samples and eliminate this phenomenon in visualization of the output results.

Data preprocessing is an important part of deep learning, which could improve model training speed and accuracy. Since the input data is a three-channel RGB images, we separately process the three channels separately by standardization, adaptive histogram equalization and gamma correction.

In order to reduce memory, we resized the original images to 256×256 resolution by using bilinear interpolation. Then we randomly flip it horizontally and vertically to achieve data augmentations.

**B. Experiment process**

DS-Net contains a total of 106 convolution layers and we initialize the parameters of the entire network by using the “Xavier” method. We used 60% of the dataset to train the model and 40% to test. Adam algorithm is used to update parameters during back propagation with a learning rate of le-4. The three hyper-parameters of the total loss function $\beta, \gamma, \lambda$ are set to 1, 1, 5e-2. In the model training phase, we set the batch size to 16. Our model is trained for 210 epochs in a single GTX-1080ti GPU. The 21 epochs were chosen because no more epochs accuracy would improve.
We use 53 (Angular Second Moment×4, Contrast×4, Correlation×4, Difference Entropy×4, Difference Variance×4, Entropy×4, Gabor, Information Measure×8, Inverse Difference Moment×4, Sum Average×4, Sum Entropy×4, Sum Variance×4, Variance ×4, Sum Variance×4) of 66 ML features to build SVM and MLP models.

![Fig. 5 MLP model](image)

Fig. 5. MLP model. “HLn” represents the nth hidden layer and “FC” represents full connection layer.

We constructed a perceptron with 5 hidden layers to model 53 features, as shown in Fig. 5. All activation functions in the MLP model are ReLU. To prevent overfitting, we add BN and dropout layers to the model.

C. Model evaluation

In order to verify the performance of our proposed algorithm, we calculate the accuracy (Acc), Sensitivity (Se), Specificity (Sp), Precision (Pr) and F1 score (F1). We also draw the Receiver Operator Characteristic (ROC) curves, Precision-Recall (PR) curves and calculate the area under the curve (AUC) as shown in Fig. 7 and Fig. 8. The above indicators are defined as follows:

\[
Acc = \frac{TP + TN}{TP + TN + FP + FN} 
\]

\[
Se = Recall = \frac{TP}{TP + FN} 
\]

\[
Sp = \frac{TN}{TN + FP} 
\]

\[
Pr = \frac{TP}{TP + FP} 
\]

\[
F1 = \frac{2 \cdot Pr \cdot Se}{Pr + Se} 
\]

where \(TP\), \(TN\), \(FP\), \(FN\), respectively denote the number of true positive, true negative, false positive and false negative samples.

Result

Our proposed method can accurately classify the original image. In order to display the classification results more intuitively, we use the t-SNE algorithm to visualize the original images and predictions of DS-Net as shown in Fig. 6. In Fig. 6, A is the distribution of DS-Net’s output and B is the original distribution of the test set. As can be seen from Figure 6, the original data is cluttered, and our algorithm can distinguish each class (viable tumor, necrotic tumor, non-tumor) very well. Relative to B, A eliminates the anomaly point, which indicates that the original anomaly point through DS-Net is no longer abnormal, and has the same characteristics as the same class. At the same time, we draw each category and the average ROC and P-R curves as shown in Figure 7 and Figure 8. As you can see, our algorithm can accurately distinguish each class of data.

![Fig. 6 ROC curves of each category and average](image)

Fig. 6. The visualization of the data distribution of the test set (437 images). A (left) represents the prediction of DS-Net and B (Right) represents the original images of test dataset.

![Fig. 7 ROC curves of each category and average](image)

Fig. 7. ROC curves of each category and average. “0” represents non-tumor, “1” represents necrotic tumor, “2” represents viable tumor and “AVE” represents the average of three categories.

![Fig. 8 P-R curves of each category and average](image)

Fig. 8. P-R curves of each category and average. “0” represents non-tumor, “1” represents necrotic tumor, “2” represents viable tumor and “AVE” represents the average of three categories.

In order to prove the effect of DS-Net, we compare it with Mishra’s method, MLP, SVM and no auxiliary supervision network architecture. The results (Acc, Se, Sp, Pr, F1) are shown in Table II. As can be seen from the Table II, the deep learning algorithm has better classification performance than the machine learning algorithm. The algorithm proposed in this paper is better than the algorithm of Mishra and the structure with auxiliary supervision network is better than the structure without it. As a weak supervision, the auxiliary supervision network can speed up the entire deep network training speed. Meanwhile, using the intra-class and inter-
class information, the model can obtain more labels and improve the classification performance. Mishra's method has fewer network layers and lacks residual blocks, resulting in lower accuracy.

We recorded the AUC of the test set for each epoch during training, and plotted the smoothed curve with a moving average of 8 as shown in Fig 9. As can be seen from the Fig 9, DS-Net is superior to the structure without auxiliary supervisory network in terms of model convergence speed and accuracy.

**Discussion and Conclusion**

In this paper, we proposed DS-Net for viability and necrosis tumor assessment in osteosarcoma. Convolutional networks can extract image features through layer-by-layer connections, which can be used to classify. Compared to traditional machine learning algorithms, deep learning does not require manual extraction of features to achieve end-to-end training of the model. Deep learning requires a large amount of training data to train the model, which makes it limited in the field of medical images. In the existing method, only a small number of tags are used, and the potential in-class and inter-class information is ignored. In this paper, we use the Siamese network as an auxiliary monitoring network to train the model. The Siamese network greatly increases the number of samples by randomly constructing input pairs, which could avoid overfitting and improve performance. The existence of the auxiliary supervisory network can reduce the depth of the network in some respects, and it is easier to pass the gradient in the back propagation process and speed up the model training. Figure 8 verifies the above views. The design of the auxiliary supervision network and

| class | Accuracy | Precision | Specificity | Sensitivity | F1 score |
|-------|----------|-----------|-------------|-------------|----------|
| Mishra | average | 0.924 | 0.97 | 0.83 | 0.94 | 0.95 |
| SVM 0 | 0.879 | 0.86 | 0.854 | 0.904 | 0.881 |
| SVM 1 | 0.906 | 0.833 | 0.955 | 0.743 | 0.785 |
| SVM 2 | 0.886 | 0.788 | 0.921 | 0.788 | 0.884 |

| SVM average | 0.890333 | 0.827 | 0.91 | 0.8116667 | 0.818 |
| SVM 0 | 0.888 | 0.865 | 0.874 | 0.903 | 0.884 |
| SVM 1 | 0.897 | 0.859 | 0.96 | 0.712 | 0.778 |
| SVM 2 | 0.899 | 0.792 | 0.915 | 0.858 | 0.824 |

| SVM average | 0.8946667 | 0.8386667 | 0.9163333 | 0.8243333 | 0.8286667 |
| Mishra's method | average | 0.922 | 0.911 | 0.909 | 0.936 | 0.923 |
| Mishra's method 0 | 0.922 | 0.911 | 0.909 | 0.936 | 0.923 |
| Mishra's method 1 | 0.924 | 0.934 | 0.985 | 0.703 | 0.802 |
| Mishra's method 2 | 0.924 | 0.81 | 0.918 | 0.941 | 0.871 |

| Mishra's method average | 0.922 | 0.885 | 0.9373333 | 0.86 | 0.8653333 |
| No_AN 0 | 0.918 | 0.91 | 0.909 | 0.927 | 0.918 |
| No_AN 1 | 0.952 | 0.877 | 0.961 | 0.921 | 0.899 |
| No_AN 2 | 0.961 | 0.963 | 0.987 | 0.89 | 0.925 |

| No_AN average | 0.9436667 | 0.9166667 | 0.9523333 | 0.9126667 | 0.914 |
| DS-Net 0 | 0.957 | 0.963 | 0.963 | 0.95 | 0.956 |
| DS-Net 1 | 0.975 | 0.959 | 0.988 | 0.931 | 0.945 |
| DS-Net 2 | 0.982 | 0.944 | 0.978 | 0.992 | 0.967 |

| DS-Net average | 0.971333 | 0.9553333 | 0.976333 | 0.957667 | 0.956 |

We recorded the AUC of the test set for each epoch during training, and plotted the smoothed curve with a moving average of 8 as shown in Fig 9. As can be seen from the Fig 9, DS-Net is superior to the structure without auxiliary supervisory network in terms of model convergence speed and accuracy.
the new residual module greatly improved the model classification performance and achieved the best results for tumor classification in osteosarcoma. Our proposed method could be used for pathologists as a tool in osteosarcoma analysis.

Another thing worth noting is that the auxiliary supervisory network itself also has the classification ability. By calculating the average distance between the test data and each type of data in the training set, the class with the smallest distance is the prediction category of the test data. Samples that do not work well for classification can be predicted by an auxiliary monitoring network. But this requires finding an evaluation criterion and searching for the original images in the training set is time consuming. Designing an efficient search algorithm and feature matching in low-dimensional space will make the above method possible.

References

[1] MacQueen, J. B. Some Methods for classification and Analysis of Multivariate Observations. Proceedings of 5th Berkeley Symposium on Mathematical Statistics and Probability. 1. University of California Press. pp. 281–297, 1967.
[2] Breiman, L., Friedman, J., Olshen, R. and Stone C. Classification and Regression Trees, Wadsworth, 1984. J.
[3] Ross Quinlan. C4.5: Programs for Machine Learning, Morgan Kaufmann, San Francisco, CA, 1993. J. Ross Quinlan. Induction of decision trees. Machine Learning, 1(1): 81-106, 1986.
[4] Cortes, C. and Vapnik, V. Support vector networks. Machine Learning, 20, 273-279, 1995.
[5] C.-C. Chang and C.-J. Lin. LIBSVM: a library for Support Vector Machines. ACM TIST, 2:27:1-27:27, 2011.
[6] Liu Jian-lin and Luo Xiong-lin, Research on Restricted Boltzmann Machines Vinod Nair.” International Conference on Learning Representations. 2015.
[7] Rémi Cuingnet, Rosso C, Chupin M, et al. Spatial regularization of SVM for the detection of diffusion alterations associated with stroke outcome[J]. Medical Image Analysis, 2011, 15(5):729-737.
[8] Tamaki, Toru, et al. "Computer-aided colorectal tumor classification in NBI endoscopy using local features." Medical image analysis 17.1(2013).
[9] Wan, Wong Lai, S. Rajarajan, and S. Pitchai. "A Novel Support Vector Machine Classifier Using Soft Computing Approach for Automated Classification of Emphysema, Bronchiectasis and Pleural Effusion Using Optimized Gabor Filter." Current Signal Transduction Therapy 11.2(2016):-.
[10] Rajendran, P., and M. Madheswaran. "Hybrid Medical Image Classification Using Association Rule Mining with Decision Tree Algorithm." Computer Science (2010).
[11] Bovis, Keir, et al. "Identification of masses in digital mammograms with MLP and RBF nets." Neural Networks, 2000. IJCNN 2000, Proceedings of the IEEE-INNS-ENNS International Joint Conference on IEEE, 2000.
[12] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner. Gradient-based learning applied to document recognition. Proceedings of the IEEE, november 1998.
[13] Alex Krizhevsky, Ilya Sutskever, Geoffrey E.Hinton. ImageNet Classification with Deep Convolutional Neural Networks. 2012.
[14] K. Simonyan and A. Zisserman. Very Deep Convolutional Networks for Large-Scale Image Recognition. international conference on learning representations. 2015.