Deep model with Siamese network for viable and necrotic tumor regions assessment in osteosarcoma

Yu Fu and Peng Xue
Department of Mechanical, Electrical and Information Engineering, Shandong University, Weihai 264209, China

Huizhong Ji
Department of Mechanical, Electrical and Information Engineering, Shandong University, Weihai 264209, China

Wentao Cui[a)]
Department of Mechanical, Electrical and Information Engineering, Shandong University, Weihai 264209, China

Enqing Dong[a)]
Department of Mechanical, Electrical and Information Engineering, Shandong University, Weihai 264209, China

(Received 26 December 2019; revised 1 July 2020; accepted for publication 10 July 2020; published xx xxxx xxxx)

Purpose: To achieve automatic classification of viable and necrotic tumor regions in osteosarcoma, most of the existing deep learning methods can only design a simple model to prevent overfitting on small datasets, which leads to the weak ability of extracting image features and low accuracy of the models. In order to solve the above problem, a deep model with Siamese network (DS-Net) was designed in this paper.
Methods: The DS-Net constructed on the basis of full convolutional networks is composed of an auxiliary supervision network (ASN) and a classification network. The construction of the ASN based on the Siamese network aims to solve the problem of a small training set (the main bottleneck of deep learning in medical images). It uses paired data as the input and updates the network through combined labels. The classification network uses the features extracted by the ASN to perform accurate classification.
Results: Pathological diagnosis is the most accurate method to identify osteosarcoma. However, due to intraclass variation and interclass similarity, it is challenging for pathologists to accurately identify osteosarcoma. Through the experiments on hematoxylin and eosin (H&E)-stained osteosarcoma histology slides, the DS-Net we constructed can achieve an average accuracy of 95.1%. Compared with existing methods, the DS-Net performs best in the test dataset.
Conclusions: The DS-Net we constructed can not only effectively realize the histological classification of osteosarcoma, but also be applicable to many other medical image classification tasks affected by small datasets. [https://doi.org/10.1002/mp.14397]

Key words: auxiliary supervision network, classification network, deep learning, osteosarcoma classification, Siamese network

1. INTRODUCTION

Osteosarcoma has been reported to be the third most common cancer in adolescence[1] and it is relatively rare compared to lung cancer, stomach cancer, and liver cancer. However, osteosarcoma has a higher mortality rate due to prone metastasis and rapid deterioration. For a long time, operation is the first choice for the treatment of osteosarcoma. In recent years, many scholars have carried out research on neoadjuvant chemotherapy combined with multiple drugs, which has improved the comprehensive treatment effect of osteosarcoma. Neoadjuvant chemotherapy refers to the operation after adequate and regular preoperative chemotherapy, to determine the necrosis rate of the tumor removed by the operation, and to determine the postoperative chemotherapy plan. The survival rate of patients after tumor resection has increased from 30% before neoadjuvant chemotherapy to 60% −70%, and the limb rescue rate has also increased significantly.[2] Histopathological assessment of chemotherapy is an important index to judge the prognosis. So, the pathological evaluation of tumor necrosis after chemotherapy is very necessary in the treatment of osteosarcoma.

Osteosarcoma is a highly anaplastic, pleomorphic tumor with a variety of tumor cell morphology, including fusiform, oval, epithelial, lymphocyte like, small round and transparent cells, etc. Due to the multiple patterns of osteosarcoma cell morphology, the diagnosis is easily affected by the experience and knowledge level of doctors. With the development of technology, osteosarcoma histological slides can be transformed into digital images. Digital histopathology images can be obtained by scanning hematoxylin and eosin (H&E)-stained microscopic slides. At the same time, machine learning (ML) methods are good at processing digital images, which makes it possible to evaluate osteosarcoma necrosis by computer.

ML methods have been widely used in medical image analysis during the past decades, most of which are based on handcrafted features. Common ML methods include K-means clustering,[3] decision tree model (DTM),[4,5] support vector

machine (SVM),[6,7] and multilayer perceptron (MLP),[8] etc. Rémi[9] proposed a method to detect differences in brain images based on SVM. Tamaki[10] used SVM algorithm to detect lesions of colorectal tumors. Wong[11] proposed a novel SVM classifier for automated classification of emphysema, bronchiectasis, and pleural effusion by using an optimized Gabor filter. Rajendran[12] used DTM for classifying the brain images into normal, benign, and malignant. Bovis[13] extracted 70 features from mammograms and identified masses by MLP. In addition, some research literature[14,15] have applied ML methods to the digital pathology field.

In recent years, deep learning has made great progress in the field of natural images and medical images, and has achieved excellent results in classification.[16–19] Deep learning provides a unified classification framework for feature extraction, thus getting rid of troublesome manual image feature extraction. However, deep learning needs to rely on a large amount of data to build models. The larger the amount of data, the higher the accuracy of the model. In addition, a large number of data can also prevent the model from overfitting. As the collection and analysis of medical images also involves privacy, ethics and other issues, it is difficult to form a large-scale medical image dataset. In order to solve the problems, one possible solution is data enhancement, including random flip, crop, zoom, radiation transform, noise perturbation, color change, and contrast transform, etc. Another solution is transfer learning.[20-22] Transfer learning improves the performance of a model by pre-training the model on larger datasets and fine-tuning it on small datasets. The biggest shortcoming of transfer learning is the inability to train the model end-to-end. Pre-training models on huge datasets may take a lot of time, which increases unnecessary time of training models. Generative adversarial networks (GANs)[23,24] can also be used as a data enhancement method. However, these effective methods for natural images are generally limited in medical images. Whether it is transfer learning or generating false data, it will reduce the medical credibility of the network representation.

Over the past decades, Siamese network[25,26] has been applied to various perception tasks, such as signature verification,[27] face verification,[28] and natural language processing.[29] These applications can be viewed as classification tasks in essence. Siamese network extends the richness of the samples through pair-wise learning, so it is good at dealing with small sample set problems. In Ref. [18], a Siamese network with a margin ranking loss for automated lung nodule analysis was designed elaborately. In Ref. [19] a hybridization method between transfer learning and GANs for the classification of healthy cells and cancer cell lines acquired by quantitative phase imaging was proposed. In Ref. [30] a Siamese network was trained and yielded better retrieval task results on Camelyon16 dataset than existing ImageNet based and generic self-supervised feature extraction methods. In Ref. [31] a deep learning-based reverse image search tool for histopathology images was introduced: Similar Medical Images Like Yours (SMILY), which improved the efficiency of searching large archives of histopathology images.

To overcome the problem caused by small datasets, in the paper, a deep model with Siamese network (DS-Net) was designed to automatically classify osteosarcoma images from TCIA.[32] In recent years, some research literatures[33-37] have proposed some methods for histological classification in osteosarcoma using deep learning methods. It should be noted that the method in Ref. [35] is the same as the deep learning method in Ref. [37]. However, due to the limitations of small sample sets, these deep learning-based methods can only design smaller networks, resulting in weaker ability to extract network features and unsatisfactory performance. Therefore, we constructed an auxiliary supervision network (ASN) based on Siamese network to solve the problem of a small training set. In this way, the DS-Net has a large-scale network by using the ASN.

2. MATERIALS AND METHODS

2.A. Framework of the DS-Net

A dataset used for classification problems contains a lot of interclass and intraclass information. Usually we only pay attention to some of the main interclass label information that is helpful for classification, while ignoring some other interclass and intraclass information. Therefore, we believe that these neglected interclass and intraclass information can be used as weakly supervised information to improve the classification performance of a model. In order to make full use of these neglected interclass and intraclass information, we constructed the DS-Net based on the Siamese network with full convolutional layers. The entire architecture of the DS-Net is shown in Fig. 1. The DS-Net is composed of an ASN and a classification network (CN).

The ASN based on Siamese network can be used to extract the primary features that can represent intraclass similarity and interclass variation of osteosarcoma. The ASN is required to accept two different data as input at the same time during training, and only one data as input during testing. We need to specify the following label definition strategy during training. If the input data belongs to the same class, the label is defined as 0; otherwise, the label is defined as 1. Then, the ASN is trained through the data pairs generated by the above label definition strategy. The CN can use these features extracted by the ASN to classify osteosarcoma. It is worth noting that the DS-Net we constructed is a series-integrated network structure composed of two sub-networks, in which the two sub-networks ASN and CN need to update parameters simultaneously. Experiments show that by introducing the constructed ASN, the classification effect of the DS-Net model can be improved, and the convergence speed of the model can be accelerated during the model training stage.

2.B. Auxiliary supervision network

In recent years, with the emergence of convolutional neural networks (CNN), deep learning techniques have been widely used in the field of image processing. In the field of

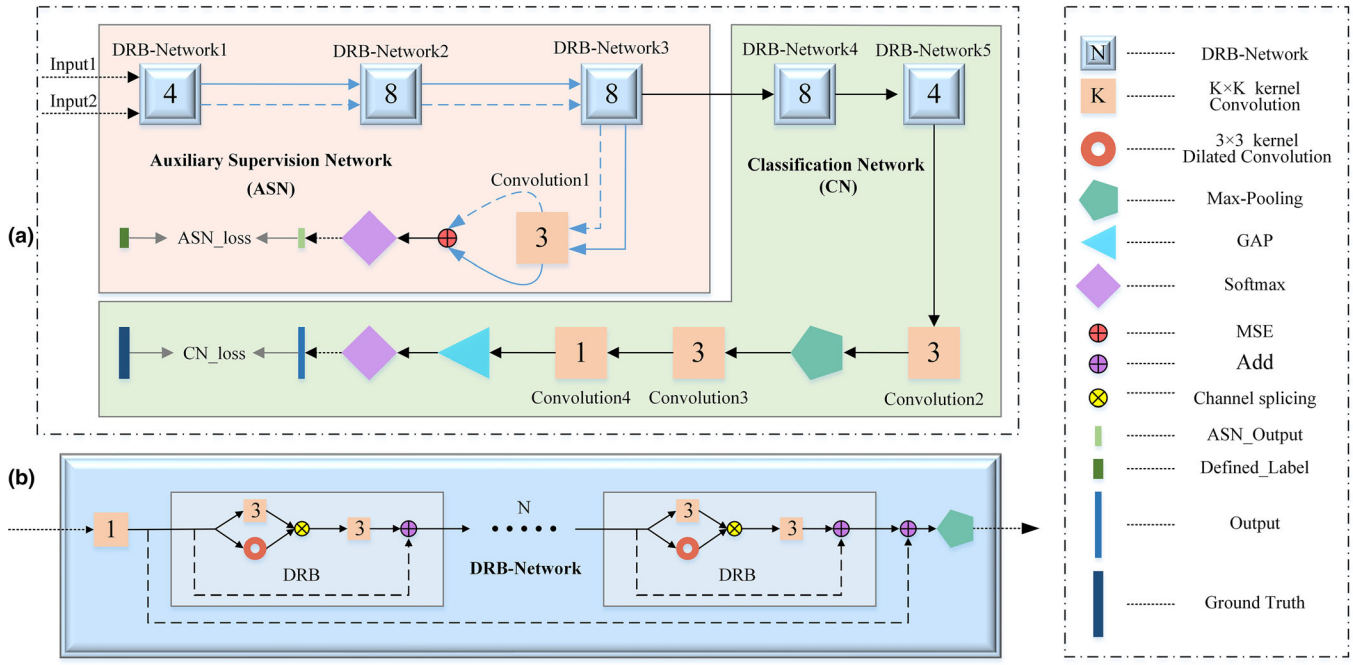

FIG. 1. The architecture of the deep model with siamese network (DS-Net). The blue line indicates that a pair of data sharing weights is processed simultaneously. Only in the auxiliary supervision network (ASN), there is a pair of data transfer, and the other parts of the DS-Net are not. ASN_loss is the weighted cross-entropy (WCE) loss of the ASN and CN_loss is the cross entropy (CE) loss of classification network. "N" represents the number of dilated convolution residual block (DRB) in one DRB-Network.

natural image classification, many network structures have been proposed, such as LeNet,[38] AlexNet,[39] VGG,[40] GoogLeNet,[41] and ResNet,[42] etc. CNN are generally composed of multiple convolutional layers and pooling layers. In a convolutional layer, different convolutional kernels can extract different features. A convolution layer operation can be described as follows:

$$\boldsymbol{y} = \Phi(b + \boldsymbol{\omega} \otimes \boldsymbol{x}) \tag{1}$$

where $\boldsymbol{y}$ is the output of the convolution layer operation, $\boldsymbol{x}$ is the input, $\boldsymbol{\omega}$ is the weight vector of the convolutional kernel, "$\otimes$" represents convolution operation, b is the offset, and Φ is the activation function. Common activation functions used in neural networks include sigmoid, Rectified Linear Unit (ReLU), and softmax. Set $\boldsymbol{\xi} = b + \boldsymbol{\omega} \otimes \boldsymbol{x}$, in this paper, we used ReLU and softmax, which are defined as follows:

$$\Phi_{ReLU}(\xi) = \max(\xi, 0) \tag{2}$$

$$\Phi_{softmax}(\xi_i) = \frac{exp(\xi_i)}{\sum_j^n exp(\xi_j)} \tag{3}$$

The learning process of the convolutional layer is to continuously update kernel $\boldsymbol{\omega}$ and bias b.

Nowadays, there are many feature extraction network structures, and residual networks (ResNet) are widely used because of the high accuracy achieved in classification tasks. Since residual networks have a skip connection, not only can the information of different layers be merged, but also the gradient can be prevented from disappearing. Inspired by ResNet, we used dilated convolution[43] to construct an improved residual block called dilated convolution residual block (DRB). As shown in Fig. 1(b), for a DRB, first, the feature maps obtained by ordinary convolution and dilated convolution of the input of the DRB are fused in depth. Second, the fusion result is subjected to an ordinary convolution. Finally, the residual features obtained by adding the output of this ordinary convolution and the input of the DRB are taken as the output of the DRB. The role of DRB not only allows the dilated convolution to increase the receiving field and obtain more features, but also the structure of the dilated convolution keeps the parameters of the convolution layer unchanged. We built a DRB-Network by connecting a series of DRBs, and established a data skip connection in the entire DRB-Network. The size of feature maps will be halved after a DRB-Network.

The ASN is mainly composed of three DRB-Networks with different scales [Fig. 1(a), N = 4, 8, 8]. A pair of inputs passes through three DRB-Networks to obtain a pair of feature maps. In the ASN, the pair of feature maps are convolved through a convolution kernel with a size of 3 × 3 to generate a pair of two-channel outputs, and the mean square error (MSE) between the two outputs values is calculated, then the ASN output is obtained after passing a softmax classifier, and ASN loss is calculated with the defined label by using the weighted cross-entropy (WCE) loss. Weighted cross-entropy loss can be defined as follows:

$$\mathscr{L}_{ASN} = -\sum_i \alpha \cdot \boldsymbol{l}_i \cdot log\boldsymbol{p}_i(\boldsymbol{l}_i|\boldsymbol{x};\theta) \tag{4}$$

$$\alpha = \begin{cases} 1, & i = 1 \\ 2, & i = 0 \end{cases} \tag{5}$$

where $\boldsymbol{x}$ denotes the input data, $\boldsymbol{l}_i \in \{0,1\}$ denotes the defined label, $\boldsymbol{p}_i \in [0,1]$ denotes the predicted probability value, and θ denotes the parameters of the ASN. If α is always equal to 1, then (4) is the ordinary cross entropy (CE). When we randomly select two inputs, for a three-category problem, the probability that these inputs belong to different categories is twice that of the same category. Therefore, when the input pairs belong to the same category, let the parameter α be 2 to balance the loss function.

Assuming that there are k samples in a dataset, C_k^2 data pairs can be randomly formed. For the same dataset, the number of data samples is greatly increased by using input pairs, which is conducive to improving the robustness of the DS-Net. However, input pairs may make the model difficult to fit. Therefore, when training the model, a parameter γ should be reasonably selected to balance the impact of ASN on the entire DS-Net. This parameter γ will be described in the next section.

2.C. Classification network

The CN is mainly composed of two DRB-Networks, a set of feature maps is arbitrarily selected from a pair of feature maps generated by the third DRB-Network in the ASN as the input feature maps of the CN [Fig. 1(a)]. We replaced the fully connected layer with a global average pool (GAP) layer. Global average pool makes the entire network avoid a fixed input size, and can also greatly reduce the size of the DS-Net. At the same time, GAP can also prevent the model from overfitting. In the ASN and CN, we used 1 × 1 convolution kernel to change the number of feature maps. The activation functions of CN are ReLU except the last layer is softmax. The CN is trained by minimizing the cross-entropy loss of the predicted value and ground truth. In the back propagation process,[44] WCE loss only affects the ASN, while the cross-entropy loss affects the entire DS-Net. However, due to the existence of ASN_loss, it should be pointed out that the ASN mainly relies on WCE loss to update parameters, and cross-entropy loss has little effect on the ASN. ASN_loss makes the output of the ASN's third DRB-Network has the ability to identify intraclass similarity and interclass variation. In order to prevent overfitting, we added batch normalization (BN)[45] layer after convolution and L_2 regularization. BN makes the output conform to the normal distribution, which can not only speed up the model training speed, but also improve the accuracy of the model. The specific configuration of the DS-Net network is shown in the Table I. The whole loss function of the DS-Net can be defined as:

$$\mathscr{L}_{total} = \beta \mathscr{L}_{CN} + \gamma \mathscr{L}_{ASN} + \lambda L_2 \tag{6}$$

$$\mathscr{L}_{CN} = -\sum_i \boldsymbol{c}_i \cdot log \boldsymbol{p}_i(\boldsymbol{c}_i|\boldsymbol{\upsilon};\psi) \tag{7}$$

where β, γ, and λ are the three hyperparameters used to balance the losses of $\mathscr{L}_{CN}$, $\mathscr{L}_{ASN}$, and L_2. $\boldsymbol{c}_i \in \{0,1,2\}$ denotes the ground truth. ψ is the parameter of CN, and υ is the input feature map of the CN. γ is the parameter mentioned in section 2.B. and is used to balance ASN and the entire network.

3. EXPERIMENTS

3.A. Dataset and preprocessing

The dataset (Osteosarcoma data)[46] we used 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. Detailed information about the dataset can be found in Ref. [37]. The dataset consists of 1144 images with a size of 1024 × 1024, the specific distribution is as follows: 536 (47%) non-tumor images, 263 (23%) necrotic tumor images, and 345 (30%) viable tumor tiles. These 1144 images were collected from four different patients. At the same time, the dataset also contains 53 kinds of textural features generated by an automated image processing tool and 8 kinds of programmable expert-guided features generated by the guidance of two experienced pathologists for each image. A more detailed introduction of these features can be found in Ref. [37]. In order to ensure the degree of automation of an algorithm in practical applications, we abandoned the expert-guided features in the experiment and only used 53 texture features. These 53 textural features include 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, Variance × 4, Sum Variance × 4, respectively. "4" represents the Gray-level co-occurrence matrix (GLCM) in four different angular directions (0°, 45°, 90°, and 135°). Among 345 viable tumor images, 53 images with both viable tumors and necrotic tumors have been removed in our experiment.

TABLE I. Detailed structure of the deep model with siamese network (DS-Net). Dilated convolution Residual Block (DRB)-Network3 is connected to Convolution1 and DRB-Network4, respectively, and serves as the input of the classification network. "Number of DRB" represents the number of DRB in one DRB-Network.

Net name	Subnetwork	Number of DRB	Output size
Auxiliary supervision network (ASN)	DRB-Network1	4	128×128×8
	DRB-Network2	8	64×64×16
	DRB-Network3	8	32×32×32
	Convolution1	-	32×32×2
Classification network (CN)	DRB-Network4	8	16×16×64
	DRB-Network5	4	8×8×128
	Convolution2	–	8×8×64
	max-pooling	–	4×4×64
	Convolution3	–	4×4×32
	Convolution4	–	4×4×2
	GAP	–	1×2

Therefore, the final dataset used for the experiment contains 1091 images. As shown in Fig. 2, when the classification labels are known in the dataset, we used the t-SNE[47] algorithm to visually classify 1091 images with 53 features. It can be seen that the original images are irregular and difficult to distinguish. A good classification algorithm can effectively distinguish these samples in the visualization of the classification results.

Data preprocessing is an important part of deep learning, which can improve the speed and accuracy of model training. We used 60% (654) of the dataset as the training set, 20% (218) as the validation set, and 20% (219) as the test set. First, we randomly generated 50 image patches of size 256 × 256 from each of 654 images in the training set, and the 50 extracted image patches will partially overlap, which can increase the diversity of the training data, and obtained 32 700 image patches as a new training set for model training. It should be noted that during the verification and testing process, we generated 16 non-overlapping image patches with a size of 256 × 256 from each verification and test image. Second, we used standardization, adaptive histogram equalization, and gamma correction to process the RGB three channels of these patches separately. Finally, we randomly flipped these patches horizontally and vertically to achieve data enhancement. Figure 3 shows the process of data preprocessing.

3.B. Experiment process

The DS-Net contains a total of 105 convolutional layers, and we used the "Xavier" algorithm to initialize the parameters of the entire network. The Adam algorithm was used to update the parameters of DS-Net at the learning rate of 1×10^{-4}. The three hyperparameters of the total loss function β, γ, and λ are set to 1, 1, and 0.1, respectively. Through a large number of experiments, we selected the hyperparameters with the best model performance on the validation set as the final hyperparameters of the DS-Net. In the model training phase, we set the batch size to 16, and the DS-Net was trained 310 epochs in a single NVIDIA GTX-1080ti GPU (11 GB). The programming implementation of the DS-Net was based on Tensorflow.[48] It took about 22.45 ms to test one patch on average. The results of the 16 patches were averaged as the corresponding tile result.

In this paper, in order to further verify the validity of the constructed DS-Net model, we also used the 53 texture features of each image to construct SVM and MLP models for comparative analysis. We adopted a one-to-one method to construct a multi-class SVM model. The specific method was to design the SVM between any two types of samples, so for k categories of samples, we needed to design $k(k-1)/2$ SVM models. In this experiment, because we dealt with a three-classification problem, we needed to design three SVM models. We also constructed a perceptron with 4 hidden layers to model 53 textural features, as shown in Fig. 4. All activation functions in the MLP are ReLU except that the last layer is softmax. To prevent overfitting, we added BN and Dropout (Dropout rate = 0.5)[49] layers to the MLP.

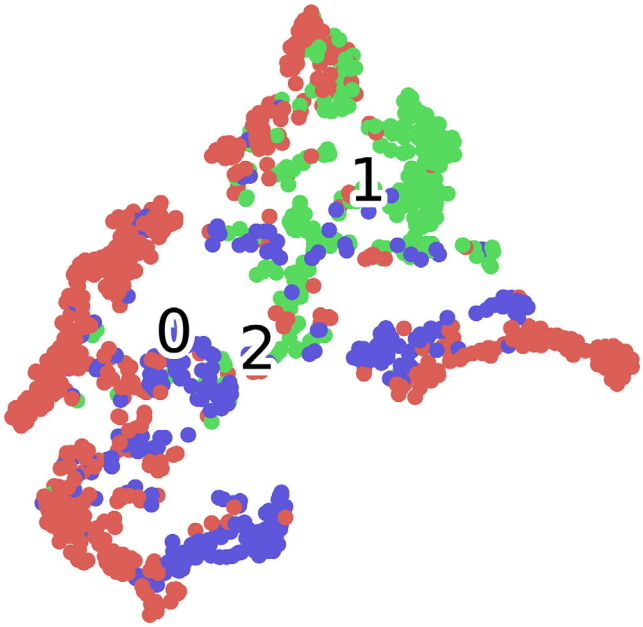

FIG. 2. Classification and visualization of 1091 images of an osteosarcoma dataset with 53 texture features. "0" (red) represents non-tumor, "1" (green) represents necrotic tumor, and "2" (blue) represents viable tumor.

3.C. Model evaluation

Although accuracy (Acc) is the main evaluation indicator, the performance of a model cannot be comprehensively evaluated only by accuracy. In order to verify the performance of the DS-Net, we calculated accuracy (Acc), sensitivity (Se), specificity (Sp), precision (Pr), and F1 score (F1). The above indicators are defined as follows:

$$Acc = \frac{TP + TN}{TP + TN + FP + FN} \tag{8}$$

$$Se = Recall = \frac{TP}{TP + FN} \tag{9}$$

$$Sp = \frac{TN}{TN + FP} \tag{10}$$

$$Pr = \frac{TP}{TP + FP} \tag{11}$$

$$F1 = \frac{2 \cdot Pr \cdot Se}{Pr + Se} \tag{12}$$

where TP, TN, FP, and FN denote the number of true positive, true negative, false positive, and false negative samples, respectively. Acc represents the probability that all samples are classified correctly. Se reflects the probability that positive samples are correctly classified. Sp reflects the proportion of negative classes that are predicted to be negative. Pr represents the proportion of the true positive samples to the predicted positive samples.

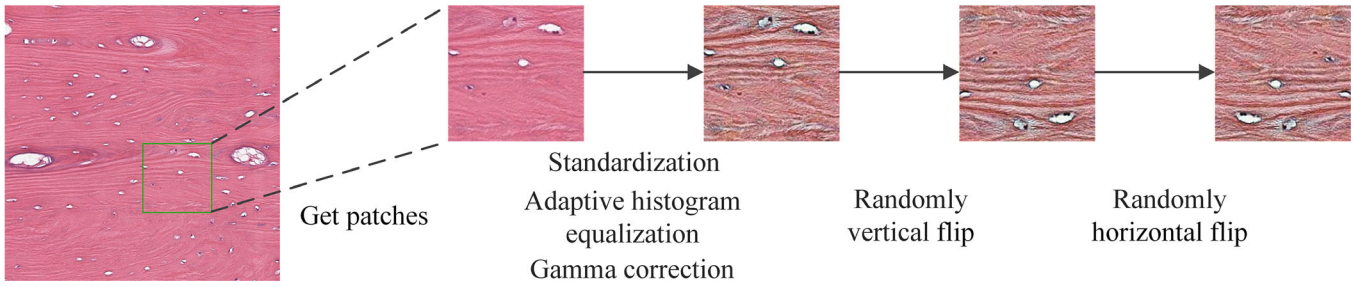

FIG. 3. The process of data preprocessing.

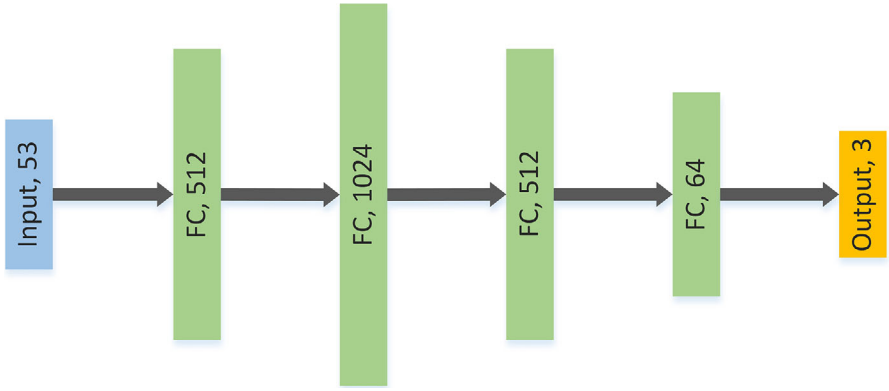

FIG. 4. Multilayer perceptron model. "FC" represents fully connected layer. The number after "FC" represents the number of neurons in corresponding layer.

4. RESULTS

In this paper, we used a test set of 219 tiles to verify the performance of the constructed DS-Net. In order to display the classification results more intuitively, Fig. 5 is a visual comparison between the classification results of the DS-Net [Fig. 5(a)] and the classification results of the original test images with 53 features using the t-SNE algorithm [Fig. 5(b)]. It can be seen from Fig. 5 that the classification result obtained by the t-SNE algorithm [Fig. 5(b)] has many interclass intersections, and the classification is not good. On the contrary, the classification result of the DS-Net [Fig. 5(a)] has little interclass intersection and high intraclass aggregation.

At the same time, we calculated the confusion matrix to further illustrate the classification effect of the constructed DS-Net model. As shown in Fig. 6, it is very consistent with the use of confusion matrix analysis, the larger the value of the main diagonal, the better the classification performance of the model. Figure 7 shows the receiver operating characteristic (ROC) curves [Fig. 7(a)] and precision-recall (P-R) curves [Fig. 7(b)] for each category. This shows that the DS-Net can accurately distinguish the types of data.

In order to further verify the effect of the constructed DS-Net model, we compared it with Mishra's method,[35] MLP, SVM, and DS-Net without ASN. In the literature,[35] because Mishra's method used the data of the old version of the dataset, for the consistency of comparison, in this experiment, the new version of the dataset for the comparison experiment was used in all methods. However, we still used Mishra proposed CNN architecture: eight learning layers, three feature extraction layers (including convolutional layers and maximum merge layers), and two fully connected layers. Figure 8 is a comparison chart of the average evaluation indicators of the five classification methods. The detailed results of the evaluation indicators of the five methods are shown in Table II. As can be seen from Fig. 8, the DS-Net is significantly better than the other four methods. More objectively, the DS-Net without ASN takes the second place, followed by Mishra's method, which shows that the proposed DS-Net framework has advantages. As can be seen from Table II, the deep learning methods as a whole have better classification performance than the machine learning methods. As a weak supervision strategy, the ASN can use intraclass and interclass information, and the DS-Net can obtain more training samples due to the use of sample pairs to improve classification performance.

Table III shows the comparison results of the accuracy of the estimated classification between our DS-Net method and Arunachalam's method using fivefold cross-validation. It can be seen from Table III that the DS-Net is superior to Arunachalam's deep learning method.

In order to verify the practicality of the proposed DS-Net model, we did another experiment. The data used in the experiment came from four different patients. Table IV shows the distribution of pathology types contained in the images collected from each patient. As can be seen from Table IV, the distribution of pathology of patient 2 is more uniform than that of the other three patients.

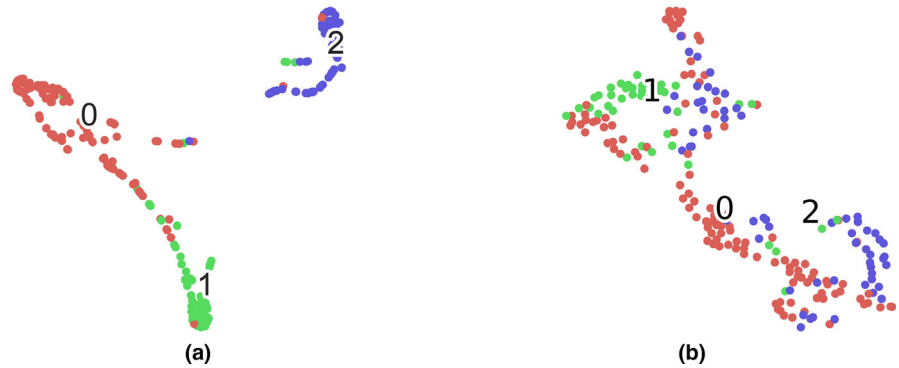

FIG. 5. Visual comparison between the classification result of the deep model with siamese network (a) and the classification result of the original test images with 53 features (b).

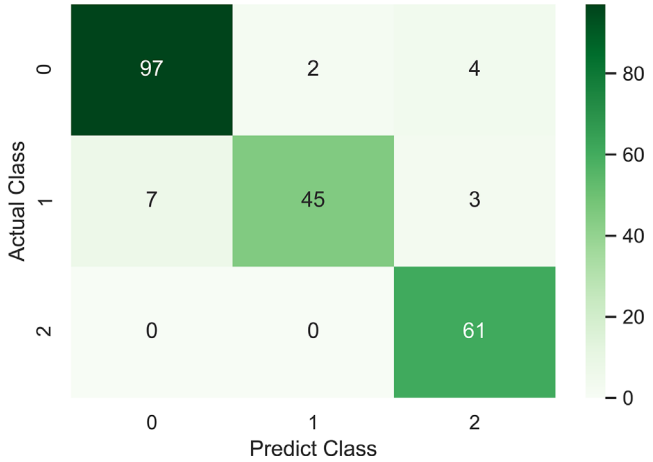

FIG. 6. Confusion matrix calculated from the classification result predicted by the constructed deep model with siamese network model using the test set data. "0" represents non-tumor, "1" represents necrotic tumor, and "2" represents viable tumor.

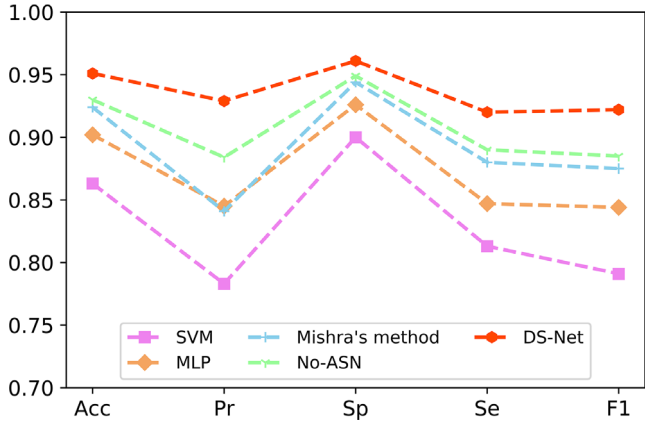

FIG. 8. Comparison of the average evaluation indicators of five classification methods on Acc, Se, Sp, Pr, and F1 indicators. No-auxiliary supervision network (ASN) represents the deep model with siamese network without ASN.

The best experiment should be fourfold cross-validation. However, the data distributions of the three patients (1, 3, and 4) are extremely uneven. Therefore, we used the images of patient 1, patient 3, and patient 4 as the training set, and the image of patient 2 as the test set. 80% of the images in the training set were used to train the DS-Net model, and the other images were used to verify the model and adjust parameters. Experimental results are shown in Table V.

To verify the effectiveness of the ASN, we compared the DS-Net with the DS-Net without ASN. We recorded the AUC of the validation set for each epoch during training, there were 310 epochs in total, and used the moving average algorithm to draw a smooth curve, as shown in Fig. 9. It can be seen that the DS-Net is superior to the structure without ASN in terms of the convergence speed and accuracy.

Finally, we evaluated the performance of the ASN. The ASN accepts a pair of data as input, and outputs two data to

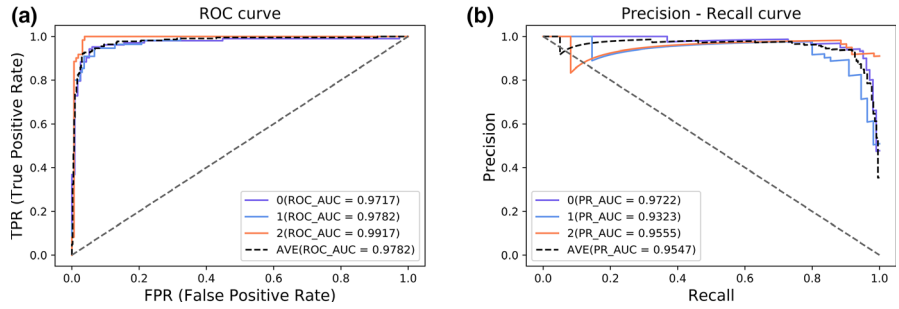

FIG. 7. Receiver operating characteristic (ROC) and P-R curves for each category. (a) represents ROC curves, (b) represents P-R curves. "0" represents non-tumor, "1" represents necrotic tumor, "2" represents viable tumor, and "AVE" represents the average of three categories.

TABLE II. Detailed results of five methods for evaluating indicators of osteosarcoma classification.

	Category	Accuracy	Precision	Specificity	Sensitivity	F1 score
SVM	Non-tumor	0.850	0.756	0.870	0.808	0.781
	Necrotic tumor	0.845	0.904	0.932	0.743	0.815
	Viable tumor	0.895	0.690	0.897	0.889	0.777
	Average	0.863	0.783	0.900	0.813	0.791
MLP	Non-tumor	0.899	0.909	0.922	0.874	0.891
	Necrotic tumor	0.909	0.843	0.951	0.782	0.811
	Viable tumor	0.899	0.783	0.905	0.885	0.831
	Average	0.902	0.845	0.926	0.847	0.844
Mishra's method	Non-tumor	0.932	0.849	0.957	0.903	0.925
	Necrotic tumor	0.918	0.849	0.951	0.818	0.833
	Viable tumor	0.922	0.824	0.924	0.918	0.868
	Average	0.924	0.841	0.944	0.880	0.875
No-ASN	Non-tumor	0.936	0.959	0.966	0.902	0.93
	Necrotic tumor	0.922	0.865	0.957	0.818	0.841
	Viable tumor	0.931	0.829	0.924	0.951	0.885
	Average	0.930	0.884	0.949	0.890	0.885
DS-Net	Non-tumor	0.941	0.933	0.940	0.942	0.937
	Necrotic tumor	0.945	0.957	0.988	0.818	0.882
	Viable tumor	0.968	0.897	0.956	1.000	0.946
	Average	**0.951**	**0.929**	**0.961**	**0.920**	**0.922**

The bold values represents the best average value among the five methods.

TABLE III. Accuracy of the estimated classification between DS-Net method and Arunachalam method.

Indicator	Method	Non-tumor	Necrotic tumor	Viable tumor	Average
Patch accuracy	Arunachalam's	0.919	**0.927**	0.953	0.933
	DS-Net	**0.936**	**0.927**	**0.954**	**0.939**
Tile accuracy	Arunachalam's	0.895	0.915	0.926	0.912
	DS-Net	**0.922**	**0.936**	**0.977**	**0.945**

The bold values represents the best value achieved by Arunachalam's method and DS-Net at the patch and tile levels, respectively.

represent the probability of belonging to the same category or different categories. Therefore, the ASN can be regarded as a CN that can perform two classifications. We randomly selected 10 000 pairs of patches to test the ASN and obtained 0.915 AUC (ROC). This shows that the ASN can effectively extract the common features of each category and determine whether the two inputs belong to the same category.

TABLE IV. Distribution of pathological types contained in the images collected from each patient.

Patient	Non-tumor	Necrotic tumor	Viable tumor
1	110	171	3
2	78	90	87
3	136	2	202
4	212	0	0

In addition, in order to verify the performance of the DS-Net on other datasets, we carried out two additional experiments. We used the skin cancer dataset[50-52] from ISIC 2019. ISIC 2019 is a competition event whose goal is to classify dermoscopic images among nine different diagnostic categories: melanoma, melanocytic nevus, basal cell carcinoma, actinic keratosis, benign keratosis (solar lentigo/seborrheic keratosis/lichen planus-like keratosis), dermatofibroma, vascular lesion, squamous cell carcinoma, and none of the others. Since the test set labels have not been published, we split the training set of ISIC 2019 into training set (60%), verification set (20%), and test set (20%). The distribution of the original training set of ISIC 2019 is shown in Table VI.

We split the original training dataset to construct two new datasets suitable for the four-classification problem. The two new datasets are Dataset A (MEL + NV+BCC + BKL, total 23 344 images) and Dataset B (AK + DF+VASC + SCC,

TABLE V. Results of cross-validation for the DS-Net.

	Category	Accuracy	Precision	Specificity	Sensitivity	F1 score
DS-Net	Non-tumor	0.890	0.857	0.944	0.769	0.811
	Necrotic tumor	0.886	0.808	0.885	0.889	0.847
	Viable tumor	0.925	0.895	0.946	0.885	0.890
	Average	0.900	0.853	0.925	0.848	0.849

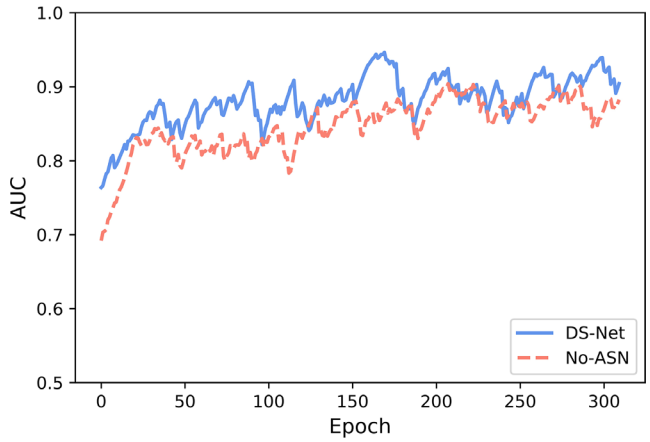

FIG. 9. Compare the AUC of the deep model with siamese network and No-auxiliary supervision network (ASN) during the training for each epoch on the verification set. No-ASN represents the method without ASN and DS-Net represents our proposed method.

total 1987 images). We performed experiments on Dataset A (big) and Dataset B (small), respectively. Since the number of these two new datasets is relatively moderate, there is no need to do preprocessing like randomly extracting 50 patches from each image in section 3.A. Due to the different resolutions of the images in the original dataset, it is necessary to use a bilinear interpolation algorithm to resample the images to a resolution of 256 × 256.

Table VII shows the performance comparison of the DS-Net with/without ASN on Dataset A and Dataset B. It should be noted that this experiment is to verify the performance of the model constructed whether using ASN on datasets of different sizes. As can be seen from Table VII, compared with the large dataset (Dataset A), the average values of several evaluation indicators on the small dataset (Dataset B) of the DS-Net model using ASN both are higher than that of the model without ASN. On the contrary, they are not very high on large datasets.

5. DISCUSSION AND CONCLUSION

In this paper, the constructed DS-Net composed of an ASN and CN is mainly used for viable and necrotic tumor regions assessment in osteosarcoma. The ASN based on Siamese network extracts primary features by reducing intraclass gaps and increasing interclass gaps. During network training, the ASN and the CN participate in the training simultaneously. The loss of the ASN only affects its own subsequent training, while the loss of the classification network affects the entire DS-Net.

TABLE VI. Data distribution of the training set of ISIC 2019.

Category	MEL	NV	BCC	AK	BKL	DF	VASC	SCC	UNK
Number	4522	12875	3323	867	2624	239	253	628	0

The DRB we proposed is called a new residual block. Its main purpose is to increase the receiving field and enrich the features by using dilated convolution, without increasing the number of parameters of the convolution kernel. A DRB-Network can be formed by multiple DRBs. How many DRBs form a DRB-Network is determined by the experiment. Generally, a multiple of 2 is selected. This DRB-Network can be used as a sub-network of the ASN and CN.

In order to solve the problem of the limited application of deep learning methods in small medical datasets, we constructed the ASN based on Siamese network. Siamese network can greatly increase the number of samples by randomly constructing input pairs, which can avoid overfitting and improve performance of the model. Experiments show that for the complex osteosarcoma classification problem, for the three-classification, only 1091 samples of the new constructed date set in section 3.A. can meet the DS-Net deep learning model with a network depth of 105 convolutional layers. Since the label strategy we defined was used for training in the ASN, and the labels of the dataset itself were used for training in the CN, this would invisibly increase the diversity of the entire dataset, thus avoiding the overfitting of the entire DS-Net. To compare and test the five methods, it is shown that the DS-Net is significantly better than the other four methods.

In the experiment in section 3.A., although the training set consisting of 32 700 (654 tiles × 50 patches/ tile) patches looks large enough, it is actually expanded from a small sample set consisting of 654 tiles. There is inevitably a great degree of similarity between the samples in the constructed large sample set. The generation of such a large sample set is actually to meet the needs of training deep learning models for the number of samples. The number of the original sample set is actually limited (654 tiles), which belongs to the category of small sample sets. Similarly, since a total of 32 700 patches in the training set can randomly generate more than 5.34×10^8 sample pairs, therefore, even with 310 epochs, the ASN still has an AUC of 0.915 when all data pairs cannot be traversed at once. While the accuracy in Table V is lower than that in the previous experiment with the different data split (accuracy is 5% lower and F1-score is 7% lower), the accuracy is still high, indicating that the DS-Net has the potential to effectively distinguish three histological tumor regions of unseen patients. Therefore, after several experimental verifications, the proposed DS-Net model is very suitable for the model construction of small sample sets. After testing two new datasets with different sizes constructed using the ISIC 2019 training dataset, the results show that the advantage of adding ASN to a smaller dataset to improve the performance of the DS-Net model is more obvious than that of a larger dataset.

The focus of the next research is to further explore how many DRB-Networks combinations in the ASN and CN are more suitable for smaller datasets, so that DS-Net can accurately classify. Finally, in summary, our main contributions are as follows: (a) We designed a novel deep

TABLE VII. The performance comparison of the DS-Net with/without ASN on dataset A and dataset B.

		Category	Accuracy	Precision	Specificity	Sensitivity	F1 score
Dataset A	No-ASN	MEL	0.866	0.703	0.942	0.556	0.621
		NV	0.847	0.827	0.771	0.911	0.867
		BCC	0.933	0.740	0.951	0.822	0.779
		BKL	0.913	0.650	0.968	0.475	0.549
		Average	**0.890**	0.730	0.908	0.691	0.704
	DS-Net	MEL	0.866	0.719	0.949	0.529	0.609
		NV	0.843	0.827	0.773	0.900	0.862
		BCC	0.941	0.777	0.96	0.831	0.803
		BKL	0.91	0.602	0.953	0.569	0.585
		Average	**0.890**	**0.731**	**0.909**	**0.707**	**0.715**
Dataset B	No-ASN	AK	0.829	0.790	0.861	0.780	0.785
		DF	0.912	0.671	0.922	0.855	0.752
		VASC	0.940	0.727	0.948	0.889	0.800
		SCC	0.801	0.726	0.905	0.566	0.582
		Average	0.871	0.729	0.909	0.773	0.730
	DS-Net	AK	0.834	0.766	0.828	0.843	0.802
		DF	0.942	0.898	0.985	0.710	0.793
		VASC	0.975	0.879	0.980	0.944	0.911
		SCC	0.806	0.696	0.873	0.656	0.675
		Average	**0.889**	**0.810**	**0.917**	**0.788**	**0.795**

The bold values represents the best value achieved by No-ASN and DS-Net on Dataset A and Dataset B, respectively.

learning-based framework (DS-Net) for the assessment of viable and necrotic tumor regions in osteosarcoma. The ASN is the core part of the DS-Net, which can be regarded as weakly supervised learning or pretraining network for transfer learning. We used a pair of training strategies to train the DS-Net, which greatly improved the convergence speed and performance of the DS-Net model. (b) We designed a new feature extraction block, which can use a small size convolution kernel to achieve the feature extraction capability of a large size convolution kernel. (c) Using osteosarcoma data from UT Southwestern, our results show that our method achieves higher classification accuracy over state-of-the-art existing approaches.

ACKNOWLEDGMENT

This work was supported by the Fundamental Research Funds for the Central Universities (China), National Natural Science Foundation of China under Grant 81671848 and 81371635, Key Research and Development Project of Shandong Province under Grant 2019GGX101022.

CONFLICT OF INTEREST

The authors have no conflict to disclose.

[a)]Author to whom correspondence should be addressed. Electronic mails: enqdong@sdu.edu.cn, wentaocui@sdu.edu.cn.

REFERENCES

1. Anja L, Paul AM, Ian L et al Osteosarcoma treatment – where do we stand? A state of the art review. *Cancer Treatm Rev*. 2014;40:523–532.
2. Negin S, Durdi Q, Tooba Y et al The regulatory functions of circular RNAs in osteosarcoma. *Genomics*. 2020;112:2845–2856.
3. Huang Z. Extensions to the k-means algorithm for clustering large data sets with categorical values. *Data Min Knowl Disc*. 1998;2:283–304.
4. Loh WY. Classification and regression trees. *WIREs Data Min Knowl Disc*. 2011;1:14–23.
5. Salzberg SL. C4.5: programs for machine learning by J. Ross Quinlan. Morgan Kaufmann Publishers, Inc., 1993. *Mach Learn*. 1994; 16:235–240.
6. Cortes C, Vapnik V. Support-vector networks. *MachLearn*. 1995;20: 273–297.
7. Chang CC, Lin CJ. LIBSVM: a library for support vector machines. *ACM Trans Intell Syst Technol*. 2011;2:1–27.
8. Liu JW, Shen FL, Luo XL. Research on perceptron learning algorithm. *Comput Eng*. 2010;36:190–192.
9. Rémi C, Charlotte R, Marie C et al Spatial regularization of SVM for the detection of diffusion alterations associated with stroke outcome. *Med Image Anal*. 2011;15:729–737.
10. Tamaki T, Yoshimuta J, Kawakami M et al Computer-aided colorectal tumor classification in NBI endoscopy using local features. *Med Image Anal*. 2013;17:78–100.
11. Wong LW, Somasundaram R, Saravanan P. A novel support vector machine classifier using soft computing approach for automated classification of emphysema, bronchiectasis and pleural effusion using optimized gabor filter. *Curr Sign Transduct Therapy*. 2016;11:121–129.
12. Rajendran P, Madheswaran M. Hybrid Medical Image Classification Using Association Rule Mining with Decision Tree Algorithm; 2010; arXiv:1001.3503.
13. Bovis K, Singh S, Fieldsend J et al Identification of masses in digital mammograms with MLP and RBF Nets. In: Amari S, Giles CL, Gori M, Piuri V, eds. Ijcnn 2000: Proceedings of the IEEE-Inns-Enns International Joint Conference on Neural Networks, Vol I. Los Alamitos: IEEE Computer Soc; 2000:342–347.

14. Alex Skovsbo J, Anders M, Niels K et al Using cell nuclei features to detect colon cancer tissue in hematoxylin and eosin stained slides. *Cytom Part A*. 2017;91:785–793.
15. Chen J, Qu A, Liu W et al Computer-aided prognosis for breast cancer based on hematoxylin & eosin hiistopathology image. *J Biomed Eng*. 2016;33:598–603.
16. Khoshdeli M, Cong R, Parvin B. Detection of Nuclei in H&E Stained Sections Using Convolutional Neural Networks. IEEE EMBS International Conference on Biomedical & Health Informatics (BHI); 2017:105–108.
17. Roth HR, Lu L, Liu JM et al Improving computer-aided detection using convolutional neural networks and random view aggregation. *IEEE Trans Med Imaging*. 2016;35:1170–1181.
18. Liu L, Dou Q, Chen H et al Multi-task deep model with margin ranking loss for lung nodule analysis. *IEEE Trans Med Imaging*. 2020;39:718–728.
19. Rubin M, Stein O, Turko NA et al TOP-GAN: stain-free cancer cell classification using deep learning with a small training set. *Med Image Anal*. 2019;57:176–185.
20. Razavian AS, Azizpour H, Sullivan J et al CNN Features off-the-shelf: an Astounding Baseline for Recognition. In: 2014 IEEE Conference on Computer Vision and Pattern Recognition Workshops. https://doi.org/10.1109/cvprw.2014.131 New York: IEEE; 2014:512–519.
21. Yosinski J, Clune J, Bengio Y et al How transferable are features in deep neural networks? 2014. arXiv:1411.1792.
22. Pan SJ, Yang QA. A survey on transfer learning. *IEEE Trans Knowl Data Eng*. 2010;22:1345–1359.
23. Goodfellow IJ, Pouget-Abadie J, Mirza M. Generative Adversarial Networks. Advances in Neural Information Processing Systems; 2014. arXiv:1406.2661.
24. Radford A, Metz L, Chintala S. Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks; 2015. arXiv:1511.06434.
25. Norouzi M, Fleet DJDDJ, Salakhutdinov R et al Hamming distance metric learning. *Adv Neu Inform Process Syst*. 2012;2:1061–1069.
26. Nair V, Hinton GE. Rectified Linear Units Improve Restricted Boltzmann Machines Vinod Nair. Paper presented at: Proceedings of the 27th International Conference on Machine Learning (ICML-10), June 21-24, 2010, Haifa, Israel; 2010.
27. Bromley J, Bentz JW, Bottou L et al Signature verification using a ldquoSiameserdquo time delay neural network. *Int J Pattern Recognit Artif Intell (Singapore)*. 1993;7:669–688.
28. Chopra S, Hadsell R, LeCun Y. Learning a similarity metric discriminatively, with application to face verification. In: Schmid C, Soatto S, Tomasi C, eds. 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Vol 1, Proceedings. Los Alamitos: IEEE Computer Soc; 2005:539–546.
29. Mueller J, Thyagarajan A, Aaai. *Siamese Recurrent Architectures for Learning Sentence Similarity.* Palo Alto: Assoc Advancement Artificial Intelligence; 2016:2786–2792.
30. Gildenblat J, Klaiman E. Self-Supervised Similarity Learning for Digital Pathology; 2020. arXiv: 1905.08139.
31. Hegde N, Hipp JD, Liu Y et al Similar image search for histopathology: SMILY. *npj Digital Medicine*. 2019;2:9.
32. Clark K, Vendt B, Smith K et al The Cancer Imaging Archive (TCIA): maintaining and operating a public information repository. *J Digit Imaging*. 2013;26:1045–1057.
33. Mishra R, Daescu O, Leavey P et al Histopathological diagnosis for viable and non-viable tumor prediction for osteosarcoma using convolutional neural network. In: Cai Z, Daescu O, Li M, eds. *Bioinformatics Research and Applications*. Cham: Springer International Publishing; 2017;10330:12–23.
34. Arunachalam HB, Mishra R, Armaselu B et al Computer aided image segmentation and classification for viable and non-viable tumor identification in osteosarcoma. In: Altman RB, Dunker AK, Hunter L, Ritchie MD, Murray T, Klein TE, eds. *Pacific Symposium on Biocomputing 2017*. Singapore: World Scientific Publ Co Pte Ltd; 2017:195–206.
35. Mishra R, Daescu O, Leavey P et al Convolutional neural network for histopathological analysis of osteosarcoma. *J Comput Biol*. 2018;25:313–325.
36. Leavey P, Arunachalam HB, Armaselu B et al Implementation of computer-based image pattern recognition algorithms to interpret tumor necrosis; a first step in development of a novel biomarker in osteosarcoma. *Pediatr Blood Cancer*. 2017;64:26–29.
37. Arunachalam HB, Mishra R, Daescu O et al Viable and necrotic tumor assessment from whole slide images of osteosarcoma using machine-learning and deep-learning models. *PLoS One*. 2019;14:e0210706.
38. Lecun Y, Bottou L, Bengio Y, Haffner P. Gradient-based learning applied to document recognition. *Proc IEEE*. 1998;86:2278–2324.
39. Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. *Commun ACM (USA)*. 2017;60:84–90.
40. Simonyan K, Zisserman A. Very Deep Convolutional Networks for Large-Scale Image Recognition; 2014. arXiv:1409.1556.
41. Szegedy C, Liu W, Jia YQ et al Going Deeper with Convolutions. In: 2015 IEEE Conference on Computer Vision and Pattern Recognition. New York: IEEE; 2015:1-9. https://doi.org/10.1109/cvpr.2015.7298594
42. He K, Zhang X, Ren S, Jian S. Deep residual learning for image recognition. *IEEE Conf Comput Vis Pattern Recogn*. 2016;1:770–778.
43. Yu F, Koltun V, Funkhouser T. Dilated residual networks. *IEEE Conf Comput Vis Pattern Recogn*. 2017;1:636–644.
44. LeCun Y, Boser B, Denker JS et al Backpropagation applied to handwritten zip code recognition. *Neural Comput*. 1989;1:541–551.
45. Ioffe S, Szegedy C. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift; 2015. arXiv:1502.03167.
46. Leavey P, Sengupta A, Rakheja D et al Osteosarcoma data from UT Southwestern/UT Dallas for Viable and Necrotic Tumor Assessment. The Cancer Imaging Archive. https://doi.org/10.7937/tcia.2019.bvhjhdas
47. Van der Maaten L, Hinton G. Visualizing data using t-SNE. *J Mach Learn Res*. 2008;9:2579–2605.
48. Abadi M, Barham P, Chen J et al Tensorflow: A system for large-scale machine learning; 2016; arXiv:1605.08695.
49. Hinton G, Srivastava N, Krizhevsky A et al Improving neural networks by preventing co-adaptation of feature detectors; 2012. arXiv:1207.0580.
50. Tschandl P, Rosendahl C, Kittler H. The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. *Sci Data*. 2018;5:180161.
51. Codella NCF, Gutman D, Celebi ME et al Skin Lesion Analysis Toward Melanoma Detection: A Challenge at the 2017 International Symposium on Biomedical Imaging (ISBI), Hosted by the International Skin Imaging Collaboration (ISIC); 2017. arXiv:1710.05006
52. Combalia M, Codella NCF, Rotemberg V et al BCN20000: Dermoscopic Lesions in the Wild; 2019; arXiv:1908.02288.