

## Research Article

# Fusion of FDG-PET Image and Clinical Features for Prediction of Lung Metastasis in Soft Tissue Sarcomas

Jin Deng , Weiming Zeng , Yuhu Shi , Wei Kong , and Shunjie Guo 

College of Information Engineering, Shanghai Maritime University, 1550 Haigang Ave., Shanghai 201306, China

Correspondence should be addressed to Weiming Zeng; zengwm86@163.com

Received 1 February 2020; Accepted 4 April 2020; Published 5 May 2020

Guest Editor: Lin Lu

Copyright © 2020 Jin Deng et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Extracting massive features from images to quantify tumors provides a new insight to solve the problem that tumor heterogeneity is difficult to assess quantitatively. However, quantification of tumors by single-mode methods often has defects such as difficulty in features extraction and high computational complexity. The multimodal approach has shown effective application prospects in solving these problems. In this paper, we propose a feature fusion method based on positron emission tomography (PET) images and clinical information, which is used to obtain features for lung metastasis prediction of soft tissue sarcomas (STSs). Random forest method was adopted to select effective features by eliminating irrelevant or redundant features, and then they were used for the prediction of the lung metastasis combined with back propagation (BP) neural network. The results show that the prediction ability of the proposed model using fusion features is better than that of the model using an image or clinical feature alone. Furthermore, a good performance can be obtained using 3 standard uptake value (SUV) features of PET image and 7 clinical features, and its average accuracy, sensitivity, and specificity on all the sets can reach 92%, 91%, and 92%, respectively. Therefore, the fusing features have the potential to predict lung metastasis for STSs.

## 1. Introduction

Sarcomas are a highly heterogeneous group of tumors classified according to the similar adult tissue types in tissue occurrence [1]. It is characterized by invasive or destructive growth that can recur and by distant metastasis [2]. As one of the sarcomas, soft tissue sarcomas (STSs) can occur anywhere in the body, and 59% of which originate in an extremity [3]. Unfortunately, 10%-20% of the patients with sarcomas or STSs have distant metastasis at the time of diagnosis. The metastasis rate is approximately 30%-40% in the course of follow-up, of which lung metastasis accounts for about 90% [4-6]. Moreover, there is a great deficiency in the cognition of the prognostic factors of lung metastatic tumor resection and the recurrence rate after resection is high [7]. Therefore, early screening and prediction of lung metastasis can help patients with STSs find corresponding self-treatment measures at an early stage and improve the survival rate of patients.

The most common method to evaluate the risk of lung metastasis is to study the heterogeneity of tumors from histopathological samples, while the biological relationships

between different clonal subgroups or clones and microenvironments in solid tumors such as STSs are still unclear, so that the information obtained from the samples is affected by the sampled region, which is not necessarily representative [8]. Therefore, it is hard to study the heterogeneity of tumors from the point of view at the molecular level because solid cancers are spatial and temporal heterogeneous. Lambin et al. have suggested that extracting a large number of features from medical images can solve this problem because the radiomic feature has the ability to capture intratumoural heterogeneity in a noninvasive [9]. Several studies have predicted the effect of lymph node metastasis and adjuvant radiotherapy or chemotherapy in preoperative colorectal cancer by using radiomic features [10, 11]. Corino et al. performed radiomic analysis of STSs to distinguish moderate and high lesions [12]. Valliã Res et al. extracted a large number of texture features from 2-deoxy-2-[18F]fluoro-D-glucose (FDG) positron emission tomography (PET) and magnetic resonance imaging (MRI) data for the construction of a STS lung metastasis prediction model [13]. However, the cost of acquiring multiple modal images at the same time is

high and may not be affordable for some patients. Also, the image acquisition of different modes is complex, and different image information sets tend to obtain several overlapping feature information. In addition, the acquisition of a large amount of information increases the complexity of the model construction and the time complexity of the operation. The features obtained from only a single image are limited, and more other accessible modal data may be overlooked such as clinical data. Actually, the fusion of image and other modal information can help us to obtain features from multiple aspects that are used to build a more accurate and stable model. There are a few studies that have used the multimodal method to achieve great results; for example, Aerts et al. quantified the tumor microenvironment by fusing imaging, gene, and clinical information to quantify tumor gene heterogeneity in the early stage [14]. Tingdan et al. developed and validated a clinical radiomics nomogram for preoperative prediction of lung metastasis in colorectal cancer patients [15].

PET image is a kind of reaction molecular metabolism imaging. When the disease is changing in the early stage of the molecular level and the morphology of the lesion area has not been abnormal, the lesion can be found by PET examination. Compared with other types of images, PET has the characteristics of high sensitivity, high specificity, and better security [16]. Therefore, this study extracted the features from two kinds of data including PET image and clinical data, and then the feature fusion was performed, which was applied to the subsequent prediction model construction. For the problem of model construction, it is expected that the predictive model can be as simple as possible and has a good predictive effect. The aim of the simple model is actually to expect it to be used as a representative feature as possible for the model construction. Hence, it is particularly important to choose higher importance features from a large number of features. There are two functions for feature selection including reducing the number of features and lowering the dimension of features, which are both used to make the model generalization more powerful by reducing the overfitting and by enhancing the understanding of features. Taking into account that the random forest algorithm has the ability to analyze the features of complex interaction classification and has good robustness for noise data or data with missing values, its variable importance measure can be used as a feature selection tool for high-dimensional data [17]. Therefore, the random forest method is applied to extract higher-contribution features from multimode data. Then, they are used as the input for a back propagation (BP) neural network with the superior ability of nonlinear mapping, self-learning, self-adaptive, generalization, and fault tolerance to construct the lung metastasis prediction model [18]. The results showed that the model constructed by combining the features of image and clinical data has a better performance in all data sets. Furthermore, it could be found that only the top PET features and clinical features achieved a higher accuracy rate of more than 90%. These features are strongly correlated with lung metastasis and may be used as a label for lung metastasis prediction of STSs.

TABLE 1: Patient and tumor characteristics.

Clinical parameters	
Age, years (mean $\pm$ SD)	54.8 $\pm$ 16.0
Gender, $n$ (%)	
Male	24 (47.1)
Female	27 (52.9)
Histology, $n$ (%)	
Malignant fibrous histiocytomas	17 (33.3)
Liposarcoma	11 (21.6)
Leiomyosarcoma	10 (19.6)
Synovial sarcoma	5 (9.8)
Extraskeletal bone sarcoma	4 (7.8)
Fibrosarcoma	1 (2.0)
Other	3 (5.9)
Grade, $n$ (%)	
High	28 (54.9)
Intermediate	15 (29.4)
Low	5 (9.8)
Unknown	3 (5.9)
Metastases, $n$ (%)	
Lung	19 (37.3)
Other	5 (9.8)
None	27 (52.9)
Time, days (mean $\pm$ SD)	
Diagnosis to outcome	285.7 $\pm$ 252.3
Diagnosis to last follow-up	849 $\pm$ 447.4
Status, $n$ (%)	
No evidence of disease	26 (51.0)
Alive with disease	9 (17.6)
Dead	15 (29.4)

Note: SD: standard deviation;  $n$ : number; diagnosis to outcome: days elapsed between the date of diagnosis of primary STS (biopsy) and the date of diagnosis of recurrence or metastases; diagnosis to last follow-up: days elapsed between the date of diagnosis of primary STS (biopsy) and the date of last-follow-up (or death, if applicable).

## 2. Methods

**2.1. Data Sources and Preprocessing.** The FDG-PET imaging data of 51 patients with STSs were included in this study, and the corresponding clinical data were downloaded from The Cancer Imaging Archive (TCIA). All patients underwent pretreatment FDG-PET scanned between November 2004 and November 2011, during which 19 patients developed lung metastases [13]. The FDG-PET slice thickness resolution was 3.27 mm for all patients and the median inplane resolution was  $5.47 \times 5.47$  mm<sup>2</sup> (range: 3.91-5.47 mm). The details of samples could be found in Table 1.

We divided these samples into two groups, *LungMets* and *NoLungMets*. *NoLungMets* were patients that did not develop lung metastases and *LungMets* were patients that eventually developed lung metastases. The FDG-PET imaging data also contained contours of the 3D tumor region of each

sample drawn by an expert radiation oncologist. To better understand the difference of images between the two groups, the region of interest (ROI) is a tumor area which is extracted according to the lesion contour mask mapped onto the original image. Therefore, ROI volumes of 51 patients are obtained for further analysis.

**2.2. Feature Extraction.** Considering that the standard uptake values (SUV) as a semiquantitative index in PET can be effectively applied in the evaluation of benign and malignant tumors and evaluation of curative effect, the FDG-PET volume is first converted to SUV maps and then a square root transform is applied to help stabilize the PET noise in the image.

These features are extracted from three aspects, and all of which are derived from the ROI regions. In Table 2, there are 67 features including 5 SUV metrics features, 5 types of texture features, and 6 types of clinical features. The study uses five types of texture features, namely, global, gray-level co-occurrence matrix (GLCM) [19], gray-level run-length matrix (GLRLM) [20–23], gray-level size zone matrix (GLSZM) [20–23], and neighborhood gray-tone difference matrix (NGTDM) [24].

The corresponding feature vector is calculated and obtained for each feature. SUV-related features can be obtained by simple mathematical calculations. Each extraction method of texture feature has a corresponding calculation formula, which can be obtained by corresponding references [19–24]. Then, we calculated all the texture feature values corresponding to each sample by the formula of each texture feature. Furthermore, considering that the most clinical features are presented in text form, a coding method called one-hot is applied to extract text features [25]. For example, the *Sex* feature includes male and female, the male is represented as 1 0 and the female is denoted as 0 1. Similarly, feature *Status* is coded as Alive (1 0 0), Alive with disease (0 1 0) and Dead (0 0 1). These feature vectors make up a feature matrix with a size of  $51 \times 67$ . Row denotes the sample and each column is a feature vector.

**2.3. Feature Selection Based on Random Forest.** A random forest is an integrated classifier with a set of decision tree classifiers that can be expressed as  $\{h(X, \theta_k), k = 1, 2, \dots, K\}$ , where  $\{\theta_k\}$  is a random vector obeying independent and identical distribution.  $K$  represents the number of decision trees in random forest. The optimal classification result is determined by the voting of each decision tree classifier when given an argument  $X$  [17].

The variable importance assessment is a significant feature of random forest algorithms. In this study, we use a variable importance measure based on the classification accuracy of out-of-bag (oob) data. The evaluation criterion of this method is the average reduction of the classification accuracy after the slight disturbance of the independent variable of the oob data and classification accuracy before the disturbance.

Assuming that there are bootstrap samples  $b = 1, 2, \dots, B$ , where  $B$  represents the number of training samples, the

variable importance  $\overline{D}_j$  of feature  $X_i$  based on classification accuracy is calculated by the following steps: Firstly, the decision tree  $T_b$  is constructed based on training samples after setting the value of  $b$  to 1, and then the oob data is defined as  $L_b^{oob}$ . After that, the oob data is classified by using the  $T_b$ , and the number of correct classifications is calculated as  $R_b^{oob}$ . For feature  $X_j (j = 1, 2, \dots, N)$ , the value of the feature  $X_j$  in  $L_b^{oob}$  is disturbed, and the data set after disturbance is defined as  $L_{bj}^{oob}$ . The number of  $L_{bj}^{oob}$  is classified by  $T_b$ , and the count of correct classifications is calculated as  $R_{bj}^{oob}$ . The same steps are performed on other features. Finally, the importance  $\overline{D}_j$  of feature  $X_j$  is calculated by the formula

$$\overline{D}_j = \frac{1}{B} \sum_{b=1}^B (R_b^{oob} - R_{bj}^{oob}). \quad (1)$$

**2.4. Back Propagation (BP) Neural Network Model.** The principle of BP neural network is that the gradient descent method is used to adjust the weights and thresholds, so that the mean square error value of the actual output of the network and the expected output is minimal. The training simulation process is presented as follows.

Firstly, the BP neural network structure is determined and the input layer to the implicit layer weight value  $w_{ij}$ , the implicit value to the output layer weight  $v_{ij}$ , the implicit layer threshold  $\theta_j$ , and the output layer threshold  $\gamma_t$  are assigned. Then the training samples  $(P^k, R^k)$  are randomly selected to provide to the network. After that, the input of each element of the implicit layer  $S_j$  is calculated by using the output sample  $P^k$ , the connection weight value  $w_{ij}$ , and the threshold value  $\theta_j$ , and then the output  $B_j$  of the implicit layer unit is calculated by using the  $S_j$  through the transfer function as follows:

$$S_j = \sum_{i=1}^m w_{ij} p_i^k - \theta_j, \quad (2)$$

$$B_j = f(S_j).$$

Next, the output  $L_t$  of the output layer units is calculated using the output  $B_j$ , weight value  $v_{jt}$ , and threshold  $\gamma_t$  of the middle layer, and then the response  $C_t$  of the output layer unit is calculated by the transfer function so that the following formulas are obtained:

$$L_t = \sum_{j=1}^n v_{jt} B_j - \gamma_t, \quad (3)$$

$$C_t = f(L_t).$$

The generalization error  $d_t$  of the output layer can be calculated using the expected output  $R^k$  and the network actual output  $C_t$ , and the generalization error  $e_j$  of each unit in the

TABLE 2: SUV metrics features and Clinical features used in this study.

Type	Name	Description
SUV metrics	SUVmax	Maximum SUV of the tumour region
	SUVpeak	Average of the voxel with maximum SUV within the tumour region and its 26 connected neighbors
	SUVmean	Average SUV value of the tumour region
	aucCSH	Area under the curve of the cumulative SUV volume histogram describing the percentage of total tumour volume above a percentage threshold of maximum SUV
	PercentInactive	Percentage of the tumour region that is inactive. A threshold of $0.005 \times (\text{SUVmax})^2$ followed by closing and opening morphological operations were used to differentiate active and inactive regions on FDG-PET scans
Textures	Global	Variance
		Skewness
		Kurtosis
		Energy
		Contrast
	GLCM	Entropy
		Homogeneity
		Correlation
		SumAverage
		Variance
		Dissimilarity
		AutoCorrelation
		Textures
LRE (long run emphasis)		
GLN (gray-level nonuniformity)		
RLN (run-length nonuniformity)		
RP (run percentage)		
LGRE (low gray-level run emphasis)		
HGRE (high gray-level run emphasis)		
SRLGE (short run low gray-level emphasis)		
SRHGE (short run high gray-level emphasis)		
LRLGE (long run low gray-level emphasis)		
LRHGE (long run high gray-level emphasis)		
GLV (gray-level variance)		
RLV (run-length variance)		
GLSZM	SZE (small-zone emphasis)	
	LZE (large-zone emphasis)	
	GLN (gray-level nonuniformity)	
	ZSN (zone-size nonuniformity)	
	ZP (zone percentage)	
	LGZE (low gray-level zone emphasis)	
	HGZE (high gray-level zone emphasis)	
	SZLGE (small-zone low gray-level emphasis)	
	SZHGE (small-zone high gray-level emphasis)	
	LZLGE (large-zone low gray-level emphasis)	
LZHGE (large-zone high gray-level emphasis)		
GLV (gray-level variance)		
ZSV (zone-size variance)		

TABLE 2: Continued.

Type	Name	Description
		Coarseness
		Contrast
	NGTDM	Busyness
		Complexity
		Strength
	Age	Age
	Sex	Male
		Female
		Radiotherapy + surgery + chemotherapy
	Treatment	Radiotherapy + surgery
		Surgery + chemotherapy
		Alive
	Status	Alive with disease
		Dead
Clinical		High
	Grade	Intermediate
		Low
		Liposarcoma
		Leiomyosarcoma
		Synovial sarcoma
	MSKCC type	Malignant fibrous histiocyctomas
		Extraskelatal bone sarcoma
		Fibrosarcoma
		Other

middle layer can be also calculated based on three parameters including  $v_{jt}$ ,  $d_t$ , and  $B_j$ .

$$\begin{aligned} d_t &= (r - C_t) \cdot C_t(1 - C_t) \\ e_j &= \left[ \sum_{t=1}^p d_t \cdot v_{jt} \right] B_j(1 - B_j) \end{aligned} \quad (4)$$

Then, the connection weight  $v_{jt}$  and threshold  $\gamma_t$  are corrected by using the generalization error  $d_t$  of the output layer units and the output  $B_j$  of each unit in the middle layer.

$$\begin{aligned} v_{jt}(N+1) &= v_{jt}(N) + \alpha \cdot d_t \cdot B_j, \\ \gamma_t(N+1) &= \gamma_t(N) + \alpha \cdot d_t. \end{aligned} \quad (5)$$

Therefore, the fixed connection weight  $w_{ij}$  and threshold  $\theta_j$  can be obtained as follows:

$$\begin{aligned} w_{ij}(N+1) &= w_{ij}(N) + \beta \cdot e_j \cdot p_i, \\ \theta_j(N+1) &= \theta_j(N) + \beta \cdot e_j, 0 < \beta < 1. \end{aligned} \quad (6)$$

Furthermore, the next train sample is randomly selected to provide to the network according to the previous method of training until the training samples are fully trained.

### 3. Results

**3.1. Feature Selection Based on Random Forest.** Feature selection can not only improve the performance of the model but also help us to understand the characteristics of the data and the underlying structure, which plays a significant role in the further improvement of model and algorithm. In this study, the number of trees in random forest is set to 250, and then there are 67 features including 5 SUV metrics features, 5 types of texture features, and 6 types of clinical features. In order to explore the contribution of these features to the prediction model, random forest was applied to sort these features. 36 features were selected whose importance values were more than 0.01. Moreover, a  $T$  test is used to verify the performance of random forest method in this study. 25 significant features were retained in these features by using the  $T$  test with a confidence level of 95% which are shown in Figure 1.

As shown in Figure 1, these selected features contain 6 clinical features and 19 image features including 5 SUV features and 14 types of texture features. Furthermore, 6 of the top 10 features are clinical features, including age, status, treatment, and MSKCC type, and the other features belong to image features namely SUV features, including SUVpeak, SUVmax, aucCSH, and PercentInactive. Therefore, there is no any texture feature.

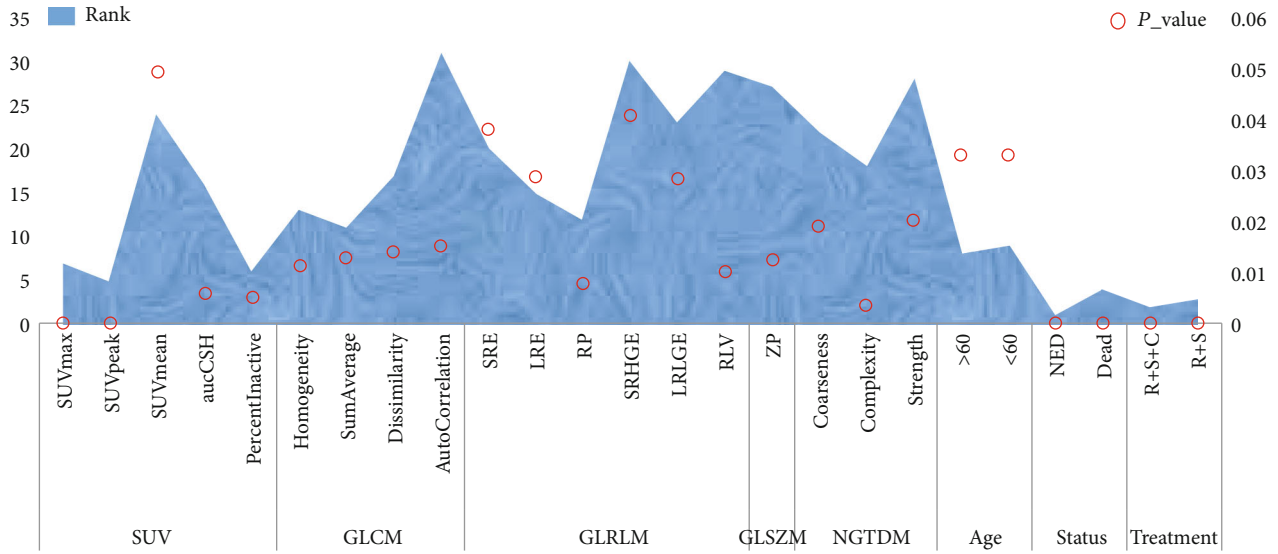


FIGURE 1: The 25 significant features based on random forest and  $T$  test. The transverse axis represents different characteristics, the main longitudinal axis represents the ranking of the feature, and the secondary longitudinal axis represents the significance level of the feature.

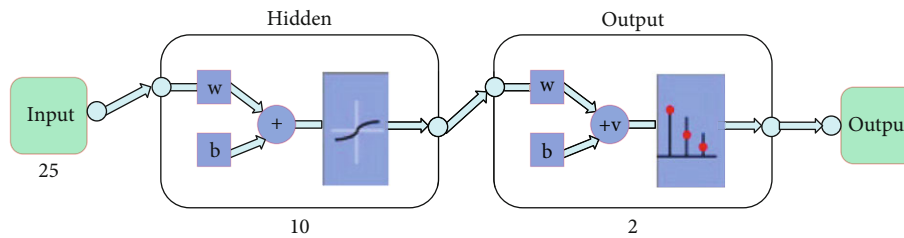


FIGURE 2: Neural network model. The numbers in the figure represent the number of neurons in each layer.

3.2. *BP Neural Network.* After obtaining those features, the BP neural network model was constructed, including 1 input layer, 1 hidden layer, and 1 output layer as shown in Figure 2.

For this neural network model, there are 25 most important features so that the number of neurons for input layer is 25. In order to make the model as simple as possible and the time complexity lower, the model in this paper is a simple three-layer model that is the hidden layer is single. The number of neurons for the output layer is 2 because the output layer contains two groups: *LungMets* and *NoLungMets*.

In this study, the *sigmoid* function was applied to be the activation function and the method of adaptive learning rate adjustment used to avoid local optimization and overfitting [26]. Then, 51 samples were randomly disrupted and divided into three types according to the proportion of 70%, 15%, and 15%, which included 35 training samples, 8 test samples, and 8 validation samples. When the number of iterations reached 1000 times or the gradient value is less than 0.001, the training model was considered to have been trained. Furthermore, in order to overcome the impact of small sample volume and sample specificity on the model, the samples were randomly divided 10 times and then repeated the above process in our study.

In order to measure the performance of the model, three indicators were used in this study, including accuracy, speci-

ficity, and sensitivity, as shown in Figure 3(a). In addition, the best validation performance of a randomly selected model is shown in Figure 3(b).

It is expected that the model has a high specificity and is sensitive on the basis of high accuracy. In other words, we hope that the model will have a better effect on both *LungMets* and *NoLungMets*. In terms of the total performance of the model, the average accuracy is 92%, and the specificity and sensitivity are 89% and 94%, respectively. Moreover, the model can achieve a good performance in training and validation set as expected. In fact, the results of the test set are more concerned because the model construction is based on the training set and the validation set; the test set is actually not involved in the construction of the model and completely independent of other sets. Therefore, the result of the test set is the standard of model performance evaluation; it can be seen from Figure 3(a) that the average accuracy, specificity, and sensitivity of the test set reached 90%, 87%, and 92%, respectively. In addition, the best validation performance is shown in Figure 3(b); the model has been trained after iterating 43 times, and the gradient mean square error is less than 0.001. At the same time, it can be seen on the validation set that the overall trend of the curve is also in the gradient drop. These evidences not only show that this model has good stability but also show that the

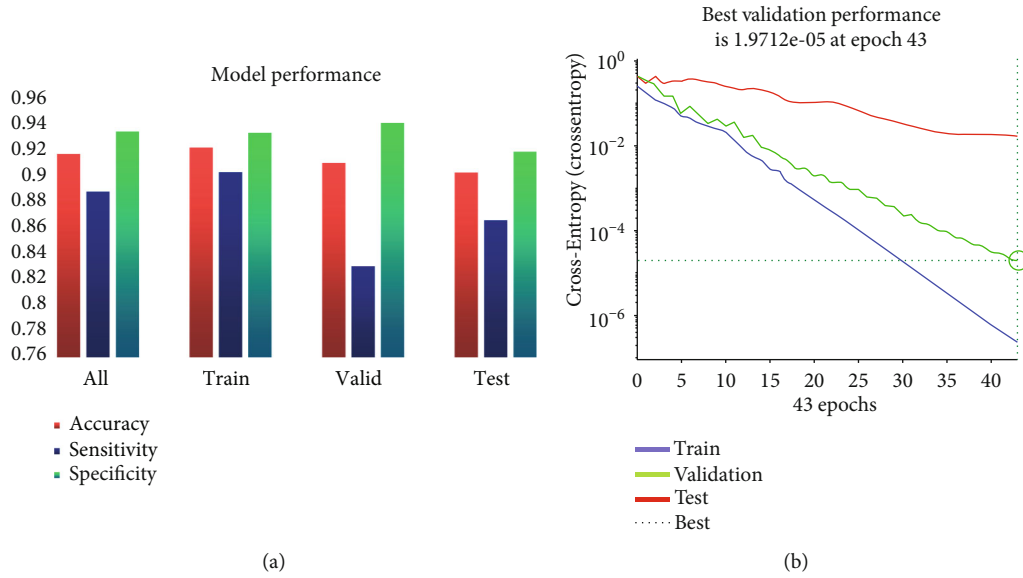


FIGURE 3: (a) The performance of model including accuracy, sensitivity, and specificity from different data sets. (b) Best validation performance of the neural network.

selected features can predict the lung metastasis in the soft tissue sarcomas.

#### 4. Discussion

In order to further confirm whether the prediction effect after feature fusion is really better than that of the single type of features and to avoid the occurrence of chance, this study merely compares the final results of feature fusion with those of image features or clinical features as shown in Figure 4.

Compared with the original features, the effect of prediction is obviously improved based on the selected features by using random forest. For the test set, the prediction accuracy based on the original feature is 83%, while the model prediction accuracy after feature extraction reaches 90%, and the sensitivity of the model is increased by 16%. Therefore, the random forest method can effectively extract the features of higher contribution to the prediction model from the original features and can greatly improve the specificity and sensitivity of the prediction model. In addition, in order to verify the necessity of feature fusion based on multimodal data, the features of single modal data are used to establish a prediction model. The evaluation results of model according to the three measurement indices show that the prediction performance of the feature fusion model is better than that using a single class of modal data at the level of all data sets. For the test set, the multimodal features can obtain higher accuracy and specificity, and although image features also reveal a high average value of sensitivity, it cannot characterize its high sensitivity because the variance is actually too large. Moreover, the sensitivity is not even exceeding 70% in the training and test set because the image features alone are used to construct the model so that the model cannot be trained very well.

It is worth mentioning that the top 10 features of the 25 features include 3 SUV features and 4 types of clinical

features, but these do not contain texture features. Therefore, the prediction model with 10 features as the model input was constructed in this study, and it was compared with the previous model with 25 features as shown in Figure 5.

It can be seen from Figure 5(a) that the top ten features of the contribution ranking are mainly 5 types, of which the image features include three SUV features and the clinical features include the *Status*, *Treatment*, *Age*, and *MSKCC type*. In Figure 5(b), it can be found that the prediction accuracy of the model without texture features decreased by less than 1% compared to the model with 25 features, and the sensitivity increased significantly although the specificity does not seem to be as ideal. These results suggest that the effect of the 10 features is similar to that of the 26 features, which means that it may be not necessary to do a lot of complex texture feature calculations to obtain the same good prediction effect, while the basic clinical features and SUV features are easier to obtain to get than texture features.

For PET images, SUV is widely used in the identification and prognosis prediction of benign and malignant tumors. *SUVmax* is always used as the initial index of benign and malignant tumor due to the characteristics of simple operation, good repeatability and not affected by the sketch area of interest. Compared to *SUVmax*, *SUVpeak* overcomes the problem of insensitive to image noise but is sensitive to regions of interest, and *PercentInactive* denotes the percentage of the inactive tumour region. A threshold of  $0.005 \times (SUVmax)^2$  followed by closing and opening morphological operations is used to differentiate active and inactive regions on FDG-PET scans. For patients with lung metastasis, a vast majority of tumors are at a low differentiation stage, and these characteristics of SUV are indicators to distinguish the low differentiation of tumors. Furthermore, we calculated the correlation between these top features and lung metastasis events, which is the label shown in Figure 6(a). In order to verify the relationship between the most contributing feature

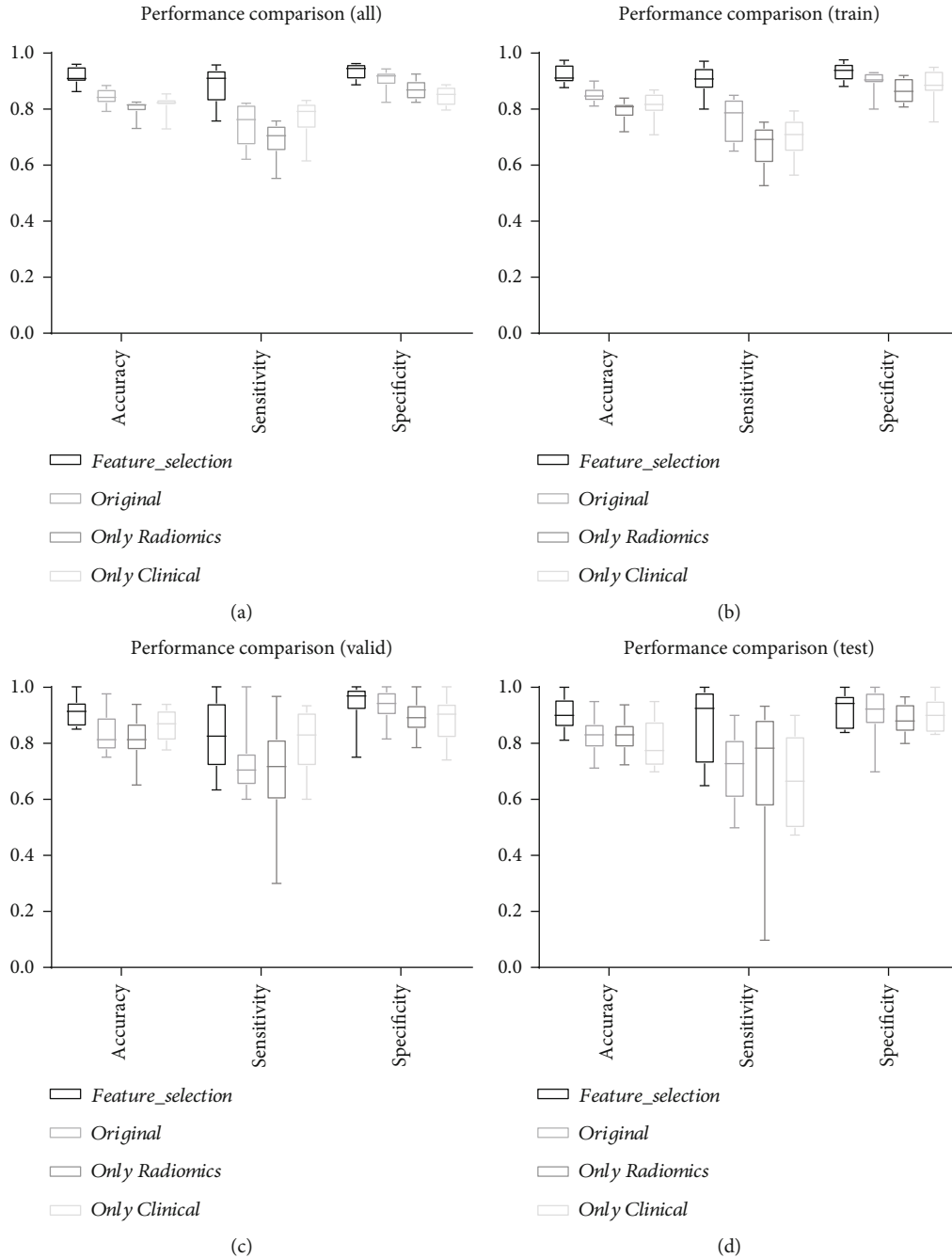


FIGURE 4: Performance comparison of feature selection. (a) represents the overall performance from the perspective of all data sets. (b–d) denote the performance from the three types of data sets including training, validation, and test set, respectively. *Feature\_selection* represents 24 features selected by random forest and  $T$  test methods. *Original* denotes 67 features including 48 image features and 16 clinical features. *Only Radiomics* represents 48 image features, and *Only Clinical* denotes 16 clinical features.

*Status* and lung metastasis, the clinical data of 254 sarcoma samples with complete information was downloaded from the TCGA database, including 220 samples of *NoLungMets* and 34 *LungMets*, and then these sets of information were applied to calculate the survival curve of sarcoma patients, as shown in Figure 6(b).

It can be seen from Figure 6(a) that there is over medium even strong correlation between most features and label, especially for the *Status*, *Treatment*, and *SUV* features. More-

over, the correlation between features is weak, except for the features of the same types that are very strong, such as *SUVmax*, *SUVpeak*, *Age*, and *Status*. It is also completely understandable, such as the feature *Age* is either greater than 60 or less than 60. Therefore, these types of nonrelated features are highly representative and can be used as a feature of the prediction model effectively. Figure 6(b) shows that the data of *LungMets* and *NoLungMets* has a significant difference in patient survival time, which demonstrates that



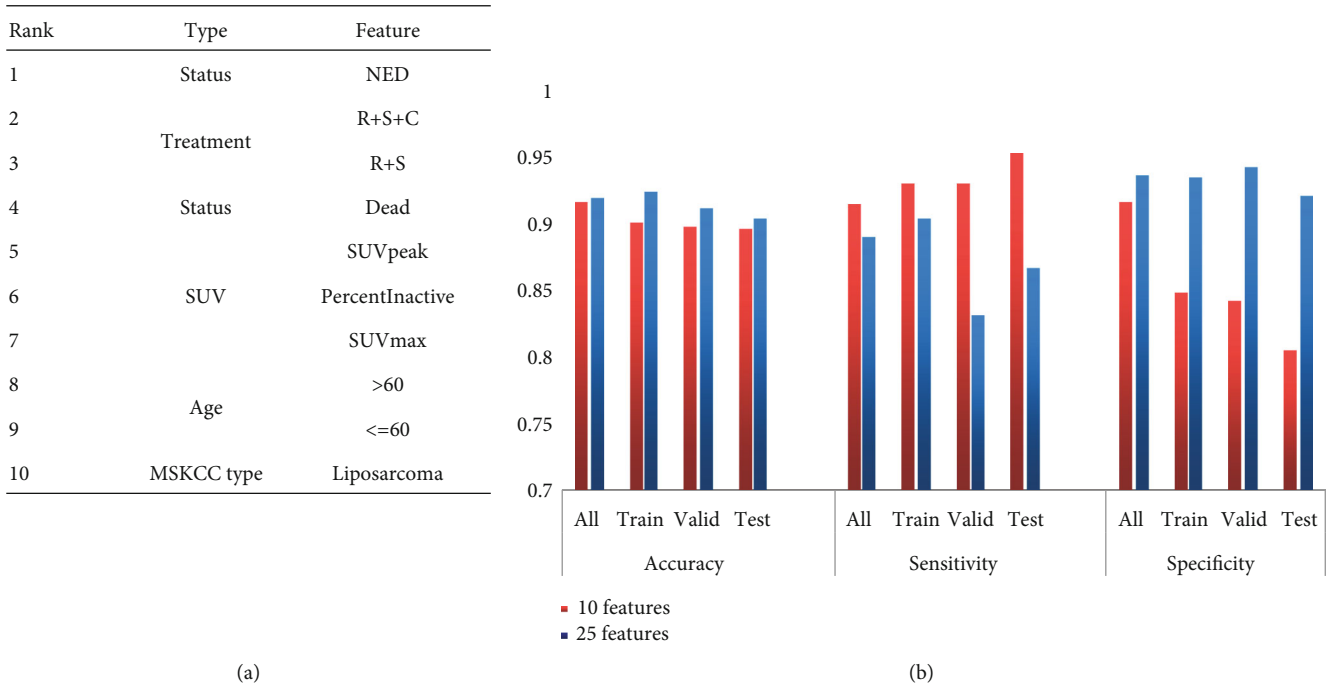


FIGURE 5: (a) The 10 top features of the contribution degree to prediction model. (b) Comparison of model performance based on different features.

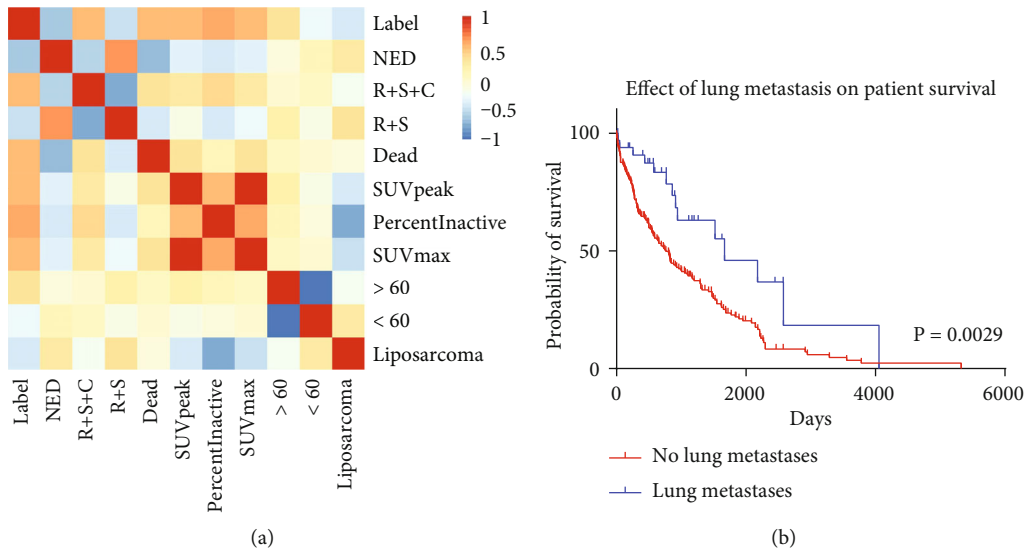


FIGURE 6: (a) Correlation between features and label. (b) Effect of lung metastasis on patient survival. The blue and red lines represent the survival probability of patients with *LungMets* and *NoLungMets*, respectively. The x-axis shows survival time of the patients, and the y-axis shows the survival probability of patients.

the correlation between survival state and lung metastasis is strong. In other words, our results suggest that the feature *Status* is very helpful for the prediction of lung metastasis, and it is easy to obtain the information about the state of survival in practical clinical.

With the development of comprehensive treatment of tumors and the prolongation of survival of cancer patients, the incidence of lung metastatic tumors is increasing. How-

ever, in the past, there were few studies on lung metastasis prediction of soft tissue tumors. Valliã`Res et al. used PET and MRI image data to construct different prediction models, and obtained a considerable prediction accuracy rate by selecting the optimal model [13]. Their study mainly used a large number of texture features with a large number of different parameters. There were more than 9000 texture features extracted from PET data according to different

parameters that resulted in excessive time complexity. In our study, the texture features under the optimal parameters were selected, which were merged with SUV features to construct the prediction model for the lung metastasis prediction as shown in Figure 5. It can be drawn from the figure that the performance of models constructed with only image features is significantly lower than that of images and clinical fusion features. Moreover, compared with the best model proposed in [13] based on PET data, the specificity of model was significantly improved, which overcomes the problem of using a large number of texture features. In fact, each texture feature corresponds to a texture feature algorithm, which was complex in the implementation of feature acquisition.

## 5. Conclusion

Based on the complementarity between different modal data features, extracting features from images and clinical data separately provides a new idea to construct predictive models. In this study, the texture and SUV features were extracted from the PET image, and the features of age, gender, and others were extracted from the clinical data. Then, all features were sorted by random forest and two-sample  $T$  test. The selected features were constructed using the BP neural network to predict the model. The results showed that the performance of multimodal feature fusion was better than that of the image data or clinical data alone. At the same time, the study further analyzed the top 10 significant features, and these features were applied to construct predictive model. It was found that the performance of the model could still achieve the previous effects without the presence of texture features, which were hard to obtain. Furthermore, the method proposed in this study could effectively select high-performance features to construct a prediction model of lung metastasis in STSs, and a high predictive performance was achieved in all data sets. In the future, we hope that this method can integrate more modal data to construct a more effective model to achieve better results, including molecular data such as genes and proteins. At the same time, this method can be extended to other prediction problems such as tumor staging and degree of tumor differentiation.

## Data Availability

The data used to support the findings of this study are from previously reported studies and public database, which have been cited.

## Conflicts of Interest

The authors declare no conflict of interests.

## Acknowledgments

This work was supported by the National Natural Science Foundation of China (Nos. 31870979 and 61906117), Natural Science Foundation of Shanghai (No. 18ZR1417200), and Shanghai Sailing Program (No. 19YF1419000). We also would like to acknowledge the individuals and institutions

that have provided data for this collection: McGill University, Montreal, Canada—special thanks are due to Martin Vallières of the Medical Physics Unit.

## References

- [1] L. A. Doyle, "Sarcoma classification: an update based on the 2013 World Health Organization classification of tumors of soft tissue and bone," *Cancer*, vol. 120, no. 12, pp. 1763–1774, 2014.
- [2] W. D. Tap, Z. Papai, B. A. van Tine et al., "Doxorubicin plus evofosfamide versus doxorubicin alone in locally advanced, unresectable or metastatic soft-tissue sarcoma (TH CR-406/SARC021): an international, multicentre, open-label, randomised phase 3 trial," *Lancet Oncology*, vol. 18, no. 8, pp. 1089–1103, 2017.
- [3] J. N. Cormier and R. E. Pollock, "Soft tissue sarcomas," *British Medical Journal*, vol. 54, no. 2, pp. 94–109, 2010.
- [4] A. J. Chou, E. S. Kleinerman, M. D. Krailo et al., "Addition of muramyl tripeptide to chemotherapy for patients with newly diagnosed metastatic osteosarcoma: a report from the Children's oncology group," *Cancer*, vol. 115, no. 22, pp. 5339–5348, 2010.
- [5] N. C. Daw, A. J. Chou, N. Jaffe et al., "Recurrent osteosarcoma with a single pulmonary metastasis: a multi-institutional review," *British Journal of Cancer*, vol. 112, no. 2, pp. 278–282, 2015.
- [6] K. Giuliano, T. Sachs, E. Montgomery et al., "Survival following lung metastasectomy in soft tissue sarcomas," *Thoracic and Cardiovascular Surgeon*, vol. 64, no. 2, pp. 150–158, 2016.
- [7] T. Treasure, M. Milošević, F. Fiorentino, and F. Macbeth, "Pulmonary metastasectomy: what is the practice and where is the evidence for effectiveness?," *Thorax*, vol. 69, no. 10, pp. 946–949, 2014.
- [8] D. L. Longo, "Tumor heterogeneity and personalized medicine," *The New England Journal of Medicine*, vol. 366, no. 10, pp. 956–957, 2012.
- [9] P. Lambin, E. Rios-Velazquez, R. Leijenaar et al., "Radiomics: extracting more information from medical images using advanced feature analysis," *European Journal of Cancer*, vol. 48, no. 4, pp. 441–446, 2012.
- [10] Y. Q. Huang, C. H. Liang, L. He et al., "Development and validation of a Radiomics Nomogram for preoperative prediction of lymph node metastasis in colorectal Cancer," *Journal of Clinical Oncology*, vol. 34, no. 18, pp. 2157–2164, 2016.
- [11] K. Nie, L. Shi, Q. Chen et al., "Rectal Cancer: assessment of Neoadjuvant Chemoradiation outcome based on Radiomics of multiparametric MRI," *Clinical Cancer Research*, vol. 22, no. 21, pp. 5256–5264, 2016.
- [12] V. D. A. Corino, E. Montin, A. Messina et al., "Radiomic analysis of soft tissues sarcomas can distinguish intermediate from high-grade lesions," *Journal of Magnetic Resonance Imaging*, vol. 47, no. 3, pp. 829–840, 2018.
- [13] M. Vallières, C. R. Freeman, S. R. Skamene, and I. el Naqa, "A radiomics model from joint FDG-PET and MRI texture features for the prediction of lung metastases in soft-tissue sarcomas of the extremities," *Physics in Medicine & Biology*, vol. 60, no. 14, pp. 5471–5496, 2015.
- [14] H. J. W. L. Aerts, E. R. Velazquez, R. T. H. Leijenaar et al., "Decoding tumour phenotype by noninvasive imaging using

- a quantitative radiomics approach,” *Nature Communications*, vol. 5, no. 1, article 4006, 2014.
- [15] T. Hu, S. P. Wang, L. Huang et al., “A clinical-radiomics nomogram for the preoperative prediction of lung metastasis in colorectal cancer patients with indeterminate pulmonary nodules,” *European Radiology*, vol. 29, no. 1, pp. 439–449, 2019.
- [16] S. N. Reske and J. Kotzerke, “FDG-PET for clinical use. Results of the 3rd German Interdisciplinary Consensus Conference, “Onko-PET III”, 21 July and 19 September 2000,” *European Journal of Nuclear Medicine*, vol. 28, no. 11, pp. 1707–1723, 2001.
- [17] L. Breiman, “Random forests,” *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.
- [18] L. U. Yan, “Optimization and application research of BP neural network,” *Journal of Beijing University of Chemical Technology*, vol. 28, no. 1, pp. 67–69, 2001.
- [19] R. M. Haralick, K. Shanmugam, and I. Dinstein, “Textural features for image classification,” *IEEE Transactions on Systems, Man, and Cybernetics*, vol. SMC-3, no. 6, pp. 610–621, 1973.
- [20] M. M. Galloway, “Texture analysis using gray level run lengths,” *Computer Graphics and Image Processing*, vol. 4, no. 2, pp. 172–179, 1975.
- [21] A. Chu, C. M. Sehgal, and J. F. Greenleaf, “Use of gray value distribution of run lengths for texture analysis,” *Pattern Recognition Letters*, vol. 11, no. 6, pp. 415–419, 1990.
- [22] B. V. Dasarathy and E. B. Holder, “Image characterizations based on joint gray level—run length distributions,” *Pattern Recognition Letters*, vol. 12, no. 8, pp. 497–502, 1991.
- [23] G. Thibault, B. Fertil, C. Navarro et al., “Shape and texture indexes application to cell nuclei classification,” *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 27, no. 1, article 1357002, 2013.
- [24] M. Amadasun and R. King, “Textural features corresponding to textural properties,” *IEEE Transactions on Systems, Man and Cybernetics*, vol. 19, no. 5, pp. 1264–1274, 1989.
- [25] P. Cerda, G. Varoquaux, and B. Kégl, “Similarity encoding for learning with dirty categorical variables,” *Machine Learning*, vol. 107, no. 8-10, pp. 1477–1494, 2018.
- [26] G. Song, J. Zhang, and Z. Sun, “the research of dynamic change learning rate strategy in BP neural network and application in network intrusion detection,” in *2008 3rd International Conference on Innovative Computing Information and Control*, Dalian, Liaoning, China, June 2008.