Breast Tumors Multi-classification Study Based on Histopathological Images with Radiomics Approach

Shuang Zhao\(^{a}\), Guohui Wei\(^{b}\), Zhiqing Ma\(^{*}\), Wenhua Zhao\(^{c}\)

College of Science and Technology, Shandong University of Traditional Chinese Medicine, Jinan, China

*Corresponding author e-mail: mazhq126@163.com, \(^{a}\)zs2899011142@163.com, \(^{b}\)bmie530@163.com, \(^{c}\)zhaowh0621@163.com

Abstract. Breast cancer is the most common malignant tumor in women. It has important clinical significance for the automatic classification of breast tumors, and the current research mainly focuses on the benign and malignant classification of breast tumors. In this paper, we proposed a radiomics method for multi-classification of breast cancer. By the radiomics method, 212 features were extracted for quantifying breast tumor images' intensity, color and texture and a multi-classification diagnosis model of breast tumors was constructed by support vector machines (SVM). The breast tumors were divided into eight categories, these eight categories include adenosis, fibroadenoma, phyllodes tumor, tubular adenoma, ductal carcinoma, lobular carcinoma, mucinous carcinoma, and papillary carcinoma. Final, the classification accuracy reached 90.3%. The radiomics approach provides an auxiliary role for developing the best treatment plan for breast tumors.

1. Introduction
The World Health Organization (WHO) World Cancer Report 3 shows that breast cancer is the highest morbidity and mortality among women worldwide, and it is increasing year by year [1]. Therefore, the treatment of breast tumors is particularly important, and the optimal treatment of breast tumors depends on the fine multi-classification. Depending on the clinical manifestations and prognosis of different types of breast tumors, doctors can early control the metastasis of tumor cells and develop precise treatment options. In addition, the histopathological image is a gold standard for identifying breast cancer compared with other medical imaging, e.g., mammography, magnetic resonance (MR), and computed tomography (CT). Nevertheless, classification for breast tumors histopathological images is a major challenge. At present, the classification of artificial pathology images in clinical diagnosis is laborious and time consuming, and the diagnosis results are susceptible to subjective human factors. Therefore, the use of computer-assisted breast tumor multi-classification is very urgent and important, which can reduce the heavy workload of pathologists and help to avoid misdiagnosis [2].

The existing researches mainly focus on the benign and malignant classification of breast tumors. For example, Wang et al [3] extracted the shape and texture features from 68 breast tumor pathology images and achieved an accuracy of 96.19% by using SVM. Spanhol et al. [4] developed and used breast cancer histopathological image dataset (BreaKHIs), which was classified by different texture...
features and different machine learning classifiers, they achieved an accuracy range of 80% to 85%. Bayramoglu et al. [5] used a magnification independent deep learning method to classify on the same dataset, the accuracy reached 83%. Spanhol et al. [6] further used AlexNet [7] on BreaKHis to classify breast tumors with benign and malignant, and the classification results were 6% higher than machine learning methods. In recent years, some scholars have studied the multi-classification of breast tumors. For example, Araújo et al. [8] used convolutional neural networks to divide breast tumor pathological images into normal tissue, benign lesion, in situ carcinoma and invasive carcinoma in datasets developed by the Israel Institute of Technology and the accuracy reached 77.8%. Brancati et al. [9] made further improvements on the same dataset, using the deep residual convolutional neural network for four classifications, achieving an accuracy of 86%.

The emergence of radiomics has tapped the potential of medical images. It extracts high-throughput features from medical images to quantify diseases such as tumors, and shows great advantages in tumor phenotyping, treatment options and prognosis analysis [10]. At present, radiomics is mainly used in mammography images, MR images and CT images, which are rarely used in pathological images. In this paper, the breast tumors were divided into eight categories by radiomics method, these eight categories include adenosis, fibroadenoma, phyllodes tumor, tubular adenoma, ductal carcinoma, lobular carcinoma, mucinous carcinoma, and papillary carcinoma. First, we extracted the features of breast tumor pathology images and selected the features, and then constructed a multi-category diagnosis model, and obtained good classification results.

2. Materials And Methods

2.1. Dataset

This paper uses the public dataset BreaKHis [4], which was established by Spanhol et al. in collaboration with the P&D laboratory in Brazil. The BreaKHis dataset includes 7909 microscopic images of breast tumors (700*460 pixels, RGB three channels, 8-bit depth per channel, PNG format), which were obtained by using different magnifications (40 ×, 100 ×, 200 × and 400 ×) from breast tumor tissue of 82 patients. Pathologists have divided breast tumors into eight different types, these eight categories include adenosis, fibroadenoma, phyllodes tumor, tubular adenoma, ductal carcinoma, lobular carcinoma, mucinous carcinoma, and papillary carcinoma. Figure 1 shows the pathological images of eight types of breast tumors from the BreaKHis dataset. Table1 shows benign and malignant distribution of breast tumor images at different magnifications.

![Fig 1. Eight classes of breast cancer histopathological images from BreaKHis dataset: (a) adenosis; (b) fibroadenoma; (c) phyllodes tumor; (d) tubular adenoma; (e) ductal carcinoma; (f) lobular carcinoma; (g) mucinous carcinoma; (h) papillary carcinoma.](image)
Tab. 1 Benign and malignant distribution of breast tumor images at different magnifications

| Magnifications factors | benign | malignant | total |
|------------------------|--------|-----------|-------|
| 40x                    | 625    | 1370      | 1995  |
| 100x                   | 644    | 1437      | 2081  |
| 200x                   | 623    | 1390      | 2013  |
| 400x                   | 588    | 1232      | 1820  |
| patient                | 24     | 58        | 82    |

2.2. Feature extraction

In this paper, 212 radiomics features are extracted to quantify breast tumors’ gray, color features, and texture.

The gray features are obtained by calculating the mean, standard deviation, smoothness, third moment, consistency and entropy of the grayscale image, yielding a 6-D feature vector.

Color features include color moments and color auto-correlograms. Because the HSV color space is in line with the human eye's subjective consciousness [11], the color moment is expressed by the first moment (mean), second moment (variance), and third moment (slope) of each color component in the HSV color space. Color moments of three-color components form a 9-D feature vector.

The color auto-correlogram is obtained by means of the color correlogram, which depicts the proportion of the number of pixels in a particular color to the whole image, and also reflects the spatial correlation between different color pairs [12]. Since the color correlogram is very complicated and bulky, and the color auto-correlogram represents the spatial relationship between pixels having the same color, the space complexity is much reduced. This paper extracted a 64-D feature vector of the color correlogram when the pixel space distance is equal 1.

Texture features include Haralick texture features, LBP features, and Gabor features.

The Haralick texture features are obtained by the Gray-level Co-occurrence Matrix (GLCM) [13]. The gray level co-occurrence matrix is obtained by calculating the pixel relative distance (d=1) and four different directions (θ=0°, 45°, 90°, 135°) and quantizing the gray level of the original image to 16 levels. Calculate 13 Haralick texture features [14]: angular second moment, contrast, correlation, sum of squares, variance, inverse difference moment, sum average, sum variance, sum entropy, entropy, difference variance, difference entropy, information measures of correlation 1, and information measures of correlation 2. Finally, we obtain 26 feature vectors by taking the mean and standard deviation of the feature vectors in the four directions.

The LBP features use the "Uniform Pattern" of the LBP operator [15]. In this mode, when there is a maximum of two transitions from 0 to 1 or from 1 to 0 in one of the LBP loop binary numbers, the type is reserved; when the number of transitions exceeds 2 times, it is classified into one class. Finally, a 59-D feature vector is obtained for 8 sampling points in the 3×3 neighborhood.

Tab. 2 The radiomics feature

| Name                                  | Number |
|---------------------------------------|--------|
| Gray features                         | 6      |
| Color moment features                 | 9      |
| Color auto-correlogram features       | 64     |
| Haralick features                     | 26     |
| LBP features                          | 59     |
| Gabor features                        | 48     |
| Total features                        | 212    |

The Gabor features are formed by scale expansion and the selection of direction of the two-dimensional Gabor filter [16]. According to the characteristics of the Gabor filter, a set of Gabor filter
banks with 4 scales and 6 directions is designed. The input images are convoluted with the real part of each filter. Finally, a 48-D feature vector is obtained by the mean and variance of the image gray value. A description of radiomics features are shown in Table 2.

2.3. Feature selection

Feature selection has been widely used in many fields such as data mining and pattern recognition. Its fundamental task is to find the most effective features from many features and reduce the dimension of the dataset and improve the accuracy of the model [17]. The ReliefF algorithm is currently recognized as a better performance method for feature selection. Its original algorithm Relief is limited to dealing with two categories of problems, and can’t deal with the lack of data and the existence of noise. Kononenko extended it and proposed the ReliefF algorithm [18]. When dealing with multi-classification problems, the ReliefF algorithm randomly takes out a sample R from the training sample set, and then finds the k neighbor samples of R from the R-like sample set, and also finds k neighbor samples from a different set of samples of each R.

The process of the ReliefF algorithm is as follows:

Input: Training sample set D, sample sampling m, nearest neighbor sample k.
Output: Feature Weight Vector Collection W
Initialize feature weights;
for i = 1 to m
Randomly select a sample Ri from D;
for each class C = class( Ri)
Find the k nearest neighbor samples Hj similar to Ri;
for each class C≠class( Ri)
Find k nearest neighbor samples Mj(C) of different classes from Ri;
for A = 1 to all feature

\[ W[A] = \frac{\sum_{i=1}^{m} \text{diff}(A, R_i, H_j)}{mk} + \sum_{C=\text{class}(R_i)} \frac{P(C)}{1-P(\text{class}(R_i))} \sum_{j=1}^{k} \frac{\text{diff}(A, R_i, M_j(C))}{mk} \]  

(1)

Where the diff function represents the difference between two samples of a given attribute, calculated as follows:

If the sample \( I_x \) has missing values, then:

\[ \text{diff}(A, I_x, I_y) = 1 - P(\text{value}(A, I_x) | \text{class}(I_x)) \]

(2)

If the samples \( I_x \) and \( I_y \) have missing values, then:

\[ \text{diff}(A, I_x, I_y) = 1 - \sum_{V} \frac{P(V | \text{class}(I_y)) \times (P(V | \text{class}(I_x)))}{P(V | \text{class}(I_y))} \]

(3)

Where \( V \) represents the value of feature A in all samples.

In this paper, the ReliefF algorithm was selected for feature selection. The sample sampling times m was equal to 100. To evaluate the stability of feature ranking, we set k to be an odd number from 1 to 29, selected the optimal neighbor sample k, and then removed the features with lower weight.

2.4. Multi-classification

The feature data after feature selection was randomly divided into a training set (70%) and a test set (30%). The feature data of the training set and the test set were normalized by "min-max normalization", and then the multi-classification model of the breast tumor was established based on the training set by support vector machine, and predicted in the test set. The flow chart of multi-classification model construction is shown in Figure 2.
The support vector machine [19] is a two-class classifier, but in order to be able to classify multiple types, the main method is to train multiple two classifiers. There are two main methods for multi-classification [20]:

1) One-to-many, for a given m categories, m classifiers need to be trained. In the training, the samples of a certain category are classified into one class, and the other remaining samples are classified into another class. Therefore, training m categories of samples requires constructing m two classifiers. Classify the data with an classifier i for an unknown sample. If a positive result is obtained, the sample belongs to class i (i ∈ m). If a negative result is obtained, the sample does not belong to class i. Repeatedly and repeatedly get the category attribute of the sample.

2) One-to-one, for a given m categories, one classifier is trained for two of the m categories. Therefore, x requires the number of two classifiers to be m(m-1)/2. For an unknown sample, it needs to be identified by all classifiers to determine the category it belongs to.

In this paper, we used the libsvm toolkit in Matlab to select one-to-one method for classification, and used the grid search and 5-fold cross-validation method to obtain the optimal penalty factor c and the kernel function parameter g.

3. Results and discussion
In this paper, we proposed a radiomics method for multi-classification of breast tumor histopathological images. Firstly, 212 features were extracted for quantifying breast tumor images’ intensity, color and texture. And then these features were selected to find the most effective features from many features and reduce the dimension.

The ReliefF algorithm was selected for feature selection. The main program was run 100 times, and the effect of k value on the performance of classification results was tested. It was found that the overall trend of classification accuracy increased first and then decreased with the increase of k value, and the maximum value is obtained when k=13, but the fluctuation was small overall, and the feature ranking is stable when k is 11~15. Finally, we selected the neighbor sample k as 13. Figure 3 shows the effect of the k-value on the performance of the classification results.
Then, according to the optimal $k$, the mean value of each feature weight was obtained, and all the feature parameters greater than 0 were selected to create the optimal feature parameter subset. It was found that the 112-dimensional feature parameter weight was greater than 0, and these features were mainly distributed in the texture feature and the color feature. Figure 4 shows the feature parameter weights.

For the SVM, the classifier parameters were optimized to improve the accuracy of classification results. Different kernels were tried; we retained the Gaussian kernel which produced the best results. The kernel parameters $c$ and $g$ were empirically defined through a grid search and fivefold cross-validation using the training set. The SVM optimal parameters $c$ and $g$ are shown in Table 3.
Finally, the multi-classification model was evaluated, and the approximate range of the results was calculated by the statistical method of the mean accuracy and standard deviation.

| Tab.3 Optimal c and g |
|-----------------------|
| c         | g         |
| 10.721    | 0.075     |

Tab.4 Classification results

| Magnification factor | accuracy       |
|----------------------|----------------|
| 40×                  | 90.30%±2.31%   |
| 100×                 | 89.65%±1.79%   |
| 200×                 | 87.55%±1.93%   |
| 400×                 | 87.63%±2.08%   |

We analyzed the classification results of the models constructed with different magnifications, and found that the classification results of 40× were generally higher than those of other magnifications, with an average of 90.30%. It can be seen that 40× is the most informative magnification, and contains more tumor information. It also shows that this model has certain feasibility for multi-classification of breast tumors; by analyzing the standard deviation of multi-classification results, it is found that the standard deviation is small, and the classification results are mainly concentrated in 85%~93%, and the deviation from the mean is small, indicating that the model has good robustness.

In addition, from the accuracy of the classification results, this model still has a lot of room for improvement. Improvements can be made from the extraction of pathological image features to extract features that clearly distinguish between benign and malignant images.

This article is mainly for multi-classification of breast tumors, but also for multi-classification of tumors in other parts. In addition, the method of this paper has certain limitations. Experiments based solely on breast cancer cases in 82 patients did not represent all cases. Only the type of tumor can be diagnosed, and the specific stage of development of the tumor cannot be given. I hope that the future work can improve the classification accuracy rate and can be refined to a certain stage of development of a certain type of tumor.

4. Conclusion

On the basis of computer-aided diagnosis, this paper proposed a radiomics method for multi-classification of breast tumor histopathological images. Firstly, 212 features were extracted for quantifying breast tumor images’ intensity, color and texture. And then these features were selected to find the most effective features from many features and reduce the dimension, and all feature parameters greater than 0 are selected. A multi-classification model was established by the subset of the optimal feature parameters, and the classification accuracy was 90.30%. It provides reliable help for clinicians to propose precise treatment options.

Acknowledgments

The corresponding author of this paper is Professor Zhiqing Ma. This work is funded by the Project of Shandong Province Medical and Health Technology Development Program (No. 2018WS206), the Project of 2016 Shandong University Humanities and Social Sciences Research Program (No. J16YB11), the Project of Shandong University of Traditional Chinese Medicine 2017 Postgraduate Teaching Reform Project (No. JG 2017015)

References

[1] S. McGuire, “World Cancer Report 2014. Geneva, Switzerland: World Health Organization, International Agency for Research on Cancer, WHO Press, 2015,” Advances in Nutrition,
W. B. Demartini, C. D. Lehman, S. Peacock, and M. T. Russell, “Computer-Aided Detection Applied to Breast MRI: Assessment of CAD-Generated Enhancement and Tumor Sizes in Breast Cancers Before and After Neoadjuvant Chemotherapy,” Academic Radiology, vol. 12, no. 7, pp. 806-814, 2005.

P. Wang, X. Hu, Y. Li, Q. Liu, and X. Zhu, “Automatic cell nuclei segmentation and classification of breast cancer histopathology images,” Signal Processing, vol. 122, pp. 1-13, 2016.

F. A. Spanhol, L. S. Oliveira, C. Petitjean, and L. Heutte, “A dataset for breast cancer histopathological image classification, IEEE Transactions on Biomedical Engineering,” vol. 63, no. 7 pp. 1455-1462, 2016.

Bayramoglu, J. Kannala, and Janne Heikkilä. “Deep learning for magnification independent breast cancer histopathology image classification,” International Conference on Pattern Recognition, 2017, pp. 2440-2445.

F. A. Spanhol, L. S. Oliveira, C. Petitjean, and L. Heutte, “Breast cancer histopathological image classification using Convolutional Neural Networks,” International Joint Conference on Neural Networks, 2016, pp. 2560-2567.

A. Krizhevsky, I. Sutskever, and G. E. Hinton, “Imagenet classification with deep convolutional neural networks,” In Advances in neural information processing systems, 2012, pp. 1097-1105.

T. Araújo, G. Aresta, E. Castro, J. Rouco, P. Aguiar, C. Eloy, A. Polónia, and A. Campilho, Classification of breast cancer histology images using Convolutional Neural Networks,” Plos One, vol. 12, no. 6, 2017.

N. Brancati, M. Frucci, and D. Riccio, “Multi-classification of Breast Cancer Histology Images by Using a Fine-Tuning Strategy,” Image Analysis and Recognition, pp. 771-778, 2018.

P. Lambin, E. Riosvelazquez, and R. Leijenaar, “Radiomics: extracting more information from medical images using advanced feature analysis,” European Journal of Cancer, vol. 48, no. 4, pp. 441-446, 2012.

M. A. Stricker, and M. Orengo, “Similarity of Color Images,” Proceedings of SPIE - The International Society for Optical Engineering, vol. 2420, pp. 381-392, 1995.

J. Huang, S. Ravi Kumar, M. Mitra, W. J. Zhu, and R. Zabih, “Spatial Color Indexing and Applications,” Proc. of the 6th International Conference on Computer Vision. Bombay, India, IEEE Press, 1998

R. M. Haralick, K. Shanmugam, I. Dinstein, “Textural Features for Image Classification,” Systems Man & Cybernetics IEEE Transactions on, vol. 3, no. 6, pp. 610-621, 1973.

R. M. Haralick, “Statistical and structural approaches to texture,” Proceedings of the IEEE, vol. 67, no. 5, pp. 786-804, 2005.

T. Ojala, M. Pietikäinen, and T. Määnpää, “Multiresolution Gray-Scale and Rotation Invariant Texture Classification with Local Binary Patterns,” IEEE Transactions on Pattern Analysis & Machine Intelligence, vol. 24, no. 7, pp. 971-987, 2002.

D. Gabor, “Theory of communication,” Iee Proc London, vol. 93, no. 73, p. 58, 1946.

K. Kir, and L. A. Rendell, “The feature selection problem: traditional methods and a new algorithm,” Tenth National Conference on Artificial Intelligence, 1992, pp. 129-134.

I. Kononenko, “Analysis and extension of RELIEF,” In Proceedings of European Conference on Machine Learning, 1994, pp. 171-182.

M. A. Hearst, S. T. Dumais, E. Osman, J. Platt, and B. Scholkopf, “Support vector machines,” IEEE Intelligent Systems & Their Applications, vol. 13, no. 4, pp. 18-28, 1998.

P. Lingras, and C. Butz, “Rough set based 1-v-1 and 1-v-r approaches to support vector machine multi-classification,” Information Sciences, vol. 177, no. 18, pp. 3782-3798, 2007.