Research on Application of Improved Random Forest in Medical Ultrasound Image Classification

Cui Min¹ and Zhu Haijiang*¹

¹College of Information Science and Technology, Beijing University of Chemical Technology, 15 BeiSanhuan East Road, ChaoYang District, Beijing 100029, China

*E-mail: 2018210475@mail.buct.edu.cn; zhuhj@mail.buct.edu.cn

Abstract. Gray-scale ultrasound imaging methods are commonly used to assess synovitis in Rheumatoid Arthritis (RA) in clinical practice. This paper proposed an improved classification algorithm for ultrasound image of metacarpophalangeal joint in Rheumatoid Arthritis(RA). Using several features of gray scale co-occurrence matrix and improved Random Forest method to achieve automatic classification of Rheumatoid Arthritis(RA). Three grading experiments were carried out in the experiment: the first one is for the binary classification of grade 0 (normal) and grade 3 (lesion) ultrasound images of 80 × 50 pixels, 130 × 70 pixels, and 185 × 90 pixels. The second is for the binary classification of grade 0 (normal) and grade 1/2/3 (lesion) ultrasound images of 80 × 50 pixels, 130 × 70 pixels, and 185 × 90 pixels. The third is for four-grade classification of ultrasound images at grade 0(normal), grade 1(mild), grade 2(moderate), and grade 3(severe) . Conclusion shows that use the method of combing gray scale co-occurrence matrix and improved Random Forest algorithm can achieve automatic classification of Rheumatoid Arthritis(RA), which has a high Classification Accuracy.

1. Introduction
Rheumatoid arthritis is a kind of diseases with joint tissue, which could cause stiffness, swelling, and ache of joint arthritis[1].Consequently, it damages skeletal system and reduces body function of joints[2]. Rheumatoid arthritis has affected many people around the world. Rheumatoid arthritis is clinically manifested as recurrent symmetric small arthritis, and joints of advanced stages may show varying degrees of rigidity and deformity, causing bone atrophy and is a disease with a high disability rate. Clinically, the severity of RA is divided into four grades.

Most of the related research on the severity of RA is focused on exploration of joint or recognition by X-ray, and ultrasonic imaging technology is a detection method more suitable for RA than X-ray[2]. The assessment of the severity of RA could be regarded as a question to achieve medical image classification . Nowadays, statistical classifiers have begun to combine image features with data training for classification, such as gradient histogram (HOG), and scale-invariant feature transformation[3].

The traditional single image classification algorithm is to achieve a classification model that is nearest to actual classification function in a function space with diverse possibilities. Including Decision-making tree, support vector machines, K-nearest neighbor , and so on. However, single classifiers often have some shortcomings, such as overfitting and difficulty in solving large or small data set classification problems. Therefore, the integrated learning algorithm came into being. The idea of the integrated learning classifier is to first construct several base classifiers from training samples,
and then classify through voting on prediction results of every base classifier[4]. Integrated learning is mainly used to To improve the efficiency of the classifier, it has a wide range of applications and good performance in the field of medical image classification[5]. Multiple single classifiers can be combined to classify the samples to make better accuracy of classification model.

The technique of using multiple classifiers to combine classification in integrated learning has been diffusely used in the range of machine learning[6,7]. A single classifier cannot guarantee that all feature descriptions can be accurately mapped. For the combined classifier, it can effectively avoid the shortage of a single classifier. By selecting the best classifier combination and more comprehensive feature selection could greatly make better classification accuracy of classifier[7]. The decision classification tree is a typical single classifier, and the random forest algorithm is an excellent combined classifier, which is composed of N independent and identically distributed different decision classification trees, and finally through each decision[8]. The comprehensive vote of the tree determines the final category of the new sample. This can effectively improve the performance of the classifier. Therefore, based on the characteristics of ultrasound images, this paper proposes a method to automatically classify rheumatoid arthritis by combing statistical characteristics of gray level co-occurrence matrix and improved random forest algorithm.

2. Experiment process and method

2.1 Feature extraction

In this chapter, the algorithm steps of this paper are mainly given. The first is the extraction of grayscale features of ultrasound images.

The gray level co-occurrence matrix is based on two-dimensional probability-density function of estimated image \( f \). Probability, corresponding matrix element is recorded as \( p(i, j \mid d, \theta) \). For the Image \( f \), Gray level co-occurrence matrix whose direction is \( \theta \) and distance is \( d \), which can be defined as:

\[
p(i, j \mid d, \theta) = \# \{ (k, l) \times (m, n) \in (Z_r \times Z_r) \times (Z_c \times Z_c) \mid d_\theta = \sqrt{(k-m)^2 + (l-n)^2}, f(k, l) = i, f(m, n) = j \}
\]

(1)

\( Z_c \) and \( Z_r \) respectively represent horizontal and vertical spatial domain of texture-image, \((k, l)\) and \((m, n)\) are pixel coordinate values in the texture image, where \( d_\theta \) is the distance in direction of \( \theta \), and \# represents the number of pixel pairs that make the braces stand. In order to make the calculation of gray level co-occurrence matrix and Image texture independent of the direction, the gray level co-occurrence matrix and corresponding characteristic parameters of every Image block are ordinarily found in four fixed directions of \( 0^0, 45^0, 90^0 \) and \( 135^0 \).

\[
p(i, j \mid d, 0^0) = \# \{ (k, l) \times (m, n) \in (Z_r \times Z_r) \times (Z_c \times Z_c) \mid k - m = 0, l - n = d, f(k, l) = i, f(m, n) = j \}
\]

(2)

\[
p(i, j \mid d, 45^0) = \# \{ (k, l) \times (m, n) \in (Z_r \times Z_r) \times (Z_c \times Z_c) \mid k - m = \sqrt{2}d, l - n = \sqrt{2}d, f(k, l) = i, f(m, n) = j \}
\]

(3)

\[
p(i, j \mid d, 90^0) = \# \{ (k, l) \times (m, n) \in (Z_r \times Z_r) \times (Z_c \times Z_c) \mid k - m = d, l - n = 0, f(k, l) = i, f(m, n) = j \}
\]

(4)
$$P(i, j | d, l, 135^0) = \# \{ (k, l) \times (m, n) \in (Z_r \times Z_c) \times (Z_r \times Z_c)$$

$$| k - m| = \sqrt{2d_2'}, | l - n | = \sqrt{2d_2'}, f(k, l) = i, f(m, n) = j \}$$

(5)

According to gray level co-occurrence matrix, some corresponding characteristic values can be calculated, and the corresponding characteristic values can be used to characterize the texture information of image. This paper extracts four parameters with strong description ability, namely: angle two matrix, entropy, moment of inertia, correlation,

$$ASM = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} p(i, j)^2$$

(6)

The angular two matrix is the sum of squares of element value in gray-level co-occurrence matrix, which represents uniformity of gray level distribution and thickness of texture. The ASM value will be larger when the distribution of elements in co-occurrence matrix is relatively concentrated.

$$H = - \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} p(i, j)^2 \ln p(i, j)$$

(7)

Entropy is considered as a measure of the amount of information which an image has. The entropy will be larger when the distribution of elements in co-occurrence matrix is more dispersed.

$$CON = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} (i \times j)^2 p(i, j)$$

(8)

The contrast represents sharpness of image and depth of image texture grooves, which the CON will be larger when grayscale difference becomes larger.

$$COR = \left[ \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} (i \times j) \times p(i, j) - \mu_1 \mu_2 \right] - \sigma_1 \sigma_2$$

(9)

$$\mu_1$$ and $$\mu_2$$ are considered as mean value of rows and columns directions of the gray-level co-occurrence matrix, and mean square deviation of rows and columns in gray level co-occurrence matrix, respectively.

2.2 Modeling based on random forest

The decision tree is a constituent unit of the random forest. The decision tree is a tree structure classifier, consisting of root nodes[8]. Given sample data $$x \in X$$ and a single decision tree $$t \in T$$, the prediction function $$h(x | t) : X \to Y$$ of the decision tree $$t$$ for the sample data $$x$$ can be expressed as:

$$\begin{cases} 
\text{Non-leaf node: } h(x | N(\psi, t_1, t_2)) = \begin{cases} 
h(x | t_1), \text{if} \psi(x) = 0 \\
h(x | t_2), \text{if} \psi(x) = 1 
\end{cases} & (10) \\
\text{Leaf node: } h(x | L_{\pi t}) = \pi & 
\end{cases}$$

Where $$\psi(x)$$ is the split function of every node of the decision trees, which determines the of the left and right sub-trees in the process of sample data from the root-node to leaf-node. $$\pi$$ is category information of the leaf node.

Random forest is a strong classifier, which is constructed with a great deal of weak classifiers (Decision Trees). The classification results are calculated by results of average member decision trees [9]. For classification results of Random Forest F, voting mechanism is generally accustomed for decision-making, which is expressed as follows:

$$y^* = \arg \max_{y \in Y} \sum_{t \in F} I(h(x | t) = y)$$

(11)
3. Improved random forest classification algorithm

3.1 Algorithm theory

The higher classification accuracy of the decision tree will make the random forest perform better intuitively. However, the classification accuracy of the random forest is affected by the previous correlation of the decision tree. Therefore, most of the research is to optimize the integration algorithm by reducing the correlation of the decision tree. In order to make the optional new feature attributes have a wider information domain, the correlation of the decision tree is further reduced.

Randomly select \( L \) feature attributes at each split point of the decision tree to combine, which is equivalent to extending the optional attributes from the original \( N \)-dimensional attributes to \( \gamma = C_N^1 + C_N^2 + \lambda + C_N^N \) dimensions, so that each split node of the decision tree can be performed on a wider range of attributes. Selection, the probability of each node selecting the same attribute is lower, which further reduces the similarity between decision trees.

The improved algorithm is expressed as follows:

Definition: \( F_N \) is the number of characteristic attributes of sample, \( D \) is the sample data:

1. **bootstrap**\((D)\): Resample the sample data to acquire training samples of Decision Tree.

2. At a node to be split in the decision tree, \( F = \left\lfloor \text{rand}(0, F_N) \right\rfloor \) attributes \((s_1, s_2, \lambda, s_F)\) are randomly selected from the sample attribute set as attributes to be combined, which \( \left\lfloor \cdot \right\rfloor \) is an upward rounding operation.

3. Select \( L = \left\lfloor \text{rand}(0, \text{int}(lbF + 1)) \right\rfloor \) weight variables \((X_1, X_2, \lambda, X_L)\), where \( X_i \) is a vector obtained by randomly taking real samples \( F \) times in \((-1, +1)\), which is \( X_i = (x_{i1}, x_{i2}, \lambda, x_{ip})\), \( i = 1: L \).

4. The \( L \) new features \((s_n, s_n, \lambda, s_n, \lambda, s_n, \lambda, s_n, \lambda)\) of Decision Tree at the split node are selected by linear weighting to obtain \( s_n = s_1 \times x_{i1} + s_2 \times x_{i2} + \lambda + s_F \times x_{ip}; i = 1: L \).

5. Select optimal new characteristic as splitting attribute of the node through entropy gain or GINI index.

6. Recursively construct every node until node sample has merely a single category to ensure the decision trees grow completely.

7. Repeat the process of constructing the decision tree (1)-(6) to produce a random forest of size \( N \).

Finally, the test samples are classified and verified according to the trained random forest classification model, which are described by feature vectors composed of gray features corresponding to each region of interest just like the training samples. In order to assess the algorithm in this paper, we evaluated the experiment result with evaluation index.

3.2 Evaluation Index of Experiment Results

We used confusion matrix, classification accuracy, AUC (Area Under Curve), accuracy, recall, and F1-measure to describe performance of the classifier. The closer the curve is to the Y axis, we calculated the evaluation index AUC. The transformation range of AUC is \([0, 1]\), the larger the AUC value, the better the classification effect of the classifier. Confusion matrix is shown in Table 1.

**Table 1.** Confusion matrix

|             | Forecast is positive | Forecast is negative |
|-------------|----------------------|----------------------|
| Actually positive | TP                   | FN                   |
| Actually negative | FP                   | TN                   |

TP represents number of positive categories predicted as positive categories. TN represents number of positive categories, which is predicted as negative categories. FP represents number of negative
categories, which is predicted as positive categories. TN represents number of negative categories, which is predicted as negative categories.

Classification accuracy rate represents the probability of being correctly predicted in all samples. The accuracy rate reflects proportion of positive samples actually predicted to be positive. The recall rate reflects proportion of positive samples which are actually predicted to be positive. F1-measure reflects harmonic average of precision and recall.

\[
\text{Accuracy} = \frac{TP + TN}{TP + FP + FN + TN} \quad (12)
\]

\[
\text{Precision} = \frac{TP}{TP + FP} \quad (13)
\]

\[
\text{Recall} = \frac{TP}{TP + FN} \quad (14)
\]

\[
F1\text{-measure} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (15)
\]

4. Experimental results and analysis

4.1 Gray feature extraction

The original joint (inflammation) ultrasound images were used in the experiment: 88 cases at level 0, 95 cases at level 1, 120 cases at level 2, and 97 cases at level 3. The original B-mode ultrasound image of the fourth-level metacarpophalangeal joint is shown in Figure 1, where the red arrow marks the ultrasound images of the finger bones, and the ultrasound images of the joints between the two bones.

![ultrasound images](image)

Figure 1. Original ultrasound images: (a) grade 0, (b) grade 1, (c) grade 2, and (d) grade 3.

4.2 Classification Results

Three grading experiments were carried out in this paper: the first one is only for the second classification of 0-grade (normal) and 3 (lesion) ultrasound images; the second is for the 0-grade (normal) and 1/2/3 (lesion) in the second category of ultrasound images; the third category is for the four categories of ultrasound images at grade-0, grade-1, grade-2, and grade-3. In the first classification, for ROIs of 80 × 50 pixels, 130 × 70 pixels, and 185 × 90 pixels, the classification accuracy rates are 97.16%, 93.85%, and 96.84% respectively. In the second classification, for ROIs of 80 × 50 pixels, 130 × 70 pixels, and 185 × 90 pixels, the classification accuracy is 97.33%, 97.81%, and 96.60% respectively. In the third classification, for ROIs of 80 × 50 pixels, 130 × 70 pixels, and 185 × 90 pixels, the classification accuracy is 91.33%, 92.18%, and 91.09% respectively. The classification model accuracy is shown in Table 2.
Table 2. Classification Model Accuracy

| Classification Accuracy | Grade 0 vs Grade 3 | Grade 0 vs Grade 123 | Grade 0 vs Grade 1 vs Grade 2 vs Grade 3 |
|-------------------------|-------------------|---------------------|-----------------------------------------|
| 80×50 pixels            | 97.16%            | 97.33%              | 91.33%                                  |
| 130×70 pixels           | 93.85%            | 97.81%              | 92.18%                                  |
| 185×90 pixels           | 96.84%            | 96.60%              | 91.09%                                  |

The AUC values of two-class model were also analyzed and evaluated so as to further evaluate the performance of the proposed method classification model.

The AUC values of the two classification models at different ROI sizes are shown in Table 3.

The precision, recall rates, and F1-measure values for each classification model at different ROI sizes are shown in the Table 4 – Table 6.

The Accuracy curve of the classification model in different number of trees is shown in Figure 2–Figure 4.

Table 3. AUC Value of Binary Classification

| Classification types   | ROI size  | AUC  |
|------------------------|-----------|------|
| Grade 0 vs Grade 3     | 80x50 pixels | 0.9939 |
|                        | 130x70 pixels | 0.9310 |
|                        | 185x90 pixels | 0.9949 |
| Grade 0 vs Grade 1/2/3| 80x50 pixels | 0.9993 |
|                        | 130x70 pixels | 0.9889 |
|                        | 185x90 pixels | 0.9881 |

Table 4. Precision, Recall and F1-measure of 80×50 pixels

| 80×50 pixels          | Grade 0 vs Grade 3 | Grade 0 vs Grade 123 | Grade 0 vs Grade 1 vs Grade 2 vs Grade 3 |
|-----------------------|-------------------|---------------------|-----------------------------------------|
| Precision             | 0.9733            | 0.9784              | 0.9483                                  |
| Recall                | 0.9650            | 0.9667              | 0.9500                                  |
| F1-measure            | 0.9689            | 0.9717              | 0.9464                                  |

Table 5. Precision, Recall and F1-measure of 130×70 pixels

| 130×70 pixels         | Grade 0 vs Grade 3 | Grade 0 vs Grade 123 | Grade 0 vs Grade 1 vs Grade 2 vs Grade 3 |
|-----------------------|-------------------|---------------------|-----------------------------------------|
| Precision             | 0.9426            | 0.9751              | 0.9565                                  |
| Recall                | 0.9401            | 0.9790              | 0.9000                                  |
| F1-measure            | 0.9382            | 0.9770              | 0.9217                                  |

Table 6. Precision, Recall and F1-measure of 185×90 pixels

| 185×90 pixels         | Grade 0 vs Grade 3 | Grade 0 vs Grade 123 | Grade 0 vs Grade 1 vs Grade 2 vs Grade 3 |
|-----------------------|-------------------|---------------------|-----------------------------------------|
| Precision             | 0.9649            | 0.9762              | 0.9286                                  |
| Recall                | 0.9692            | 0.9000              | 0.9348                                  |
| F1-measure            | 0.9670            | 0.9322              | 0.9270                                  |

Figure 2. From left to right: Grade (0 vs 3) \ ((0 vs 1,2,3) \ (0 vs 1 vs 2 vs 3)) of 80×50 pixels
5. Conclusion
In the classification of a small number of RA ultrasound images has always been a difficult problem. This paper proposes an automatic grading method based on gray-scale features of images for classification of Rheumatoid Arthritis (RA) based on ultrasound images. Firstly, this method obtains the region of interest of the rheumatoid arthritis ultrasound image, then extracts grayscale features of the regions of interest, and then uses the improved random forest method to train the grading model. Finally, the experimental samples are tested according to the training model. In the experiment, the two-level and four-level experiments of RA grading were carried out, which the RA grading effects of different sizes of regions of interest were compared. The results show that integrated learning has good classification effect and accuracy for small sample classification problems. The experiment results indicate that the improved random forest algorithm proposed in the paper has a good effect on automatic classification of ultrasound images of rheumatoid arthritis.

References
[1] Vapnik V N. The nature of statistical learning theory[M]. New York: Springer Verlag,(1995):19-23.
[2] Szkudlarek, Marcin, et al. Interobserver agreement in ultrasonography of the finger and toe joints in rheumatoid arthritis. Arthritis & Rheumatism 48.4 (2003): 955-962.
[3] Zikic D, Glocker B, Konukoglu E, et al. Decision forest for tissue-specific segmentation of high-grade gliomas in multichannel MR[C].International Conference on Medical image Computing and Computer-Assisted Intervention. Berlin:Spring, Berlin, Heidelberg, 2012, 7512: 369-376.
[4] Colebatch, Alexandra N, et al. EULAR recommendations for the use of imaging of the joints in the clinical management of rheumatoid arthritis. Annals of the rheumatic diseases (2013): annrheumdis-2012.
[5] Tsoumakas G, Vlahavas I. Random k-labelsets: An ensemble method for multilabel classification[M]/Machine Learning: ECML 2007. Berlin Heidelberg: Springer, 2007, 406-411.
[6] Litjens, Geert, et al. A survey on deep learning in medical image analysis. Medical image analysis 42 (2017): 60-88.
[7] Rodríguez-Galiano V F, Ghimire B, Rogan J, et al. An assessment of the effectiveness of a random forest classifier for land-cover classification[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2012, 34(1): 93-104.
[8] BreimanL, Random fores[J]. Machine Learning, 2001, 45(1): 5-32.
[9] LeCun, Yann, Yoshua Bengio, and Geoffrey Hinton. Deep learning. nature 521.7553 (2015):436.