Multi-class Classification of Ultrasonic Supraspinatus Images
based on Radial Basis Function Neural Network

Ming-Huwi Horng1,* Shu-Min Chen2

1Department of Computer Science and Information Engineering, National Ping Tung Institute of Commerce, Pingtung 900, Taiwan, ROC
2Department of Physical Medicine and Rehabilitation, National Cheng Kung University Hospital, Tainan 704, Taiwan, ROC

Abstract
This article reports a study on applying texture analysis methods for classifying the different rotator cuff disease groups which are normal, tendon inflammation, calcific tendonitis and rotator cuff tear by using ultrasonic images. In conventional diagnosis, physicians observe the micro/macro structures of ultrasonic tendon images to judge the severity of rotator cuff disease. The accuracy of visual observation depends on the expertise of physicians. It is often not reliable. The supraspinatus is usually involved in the above-mentioned diseases progression categories. Four texture analysis methods, gray-level co-occurrence matrix, texture spectrum, fractal dimension and texture feature coding method, were used to extract features of tissue characteristic of supraspinatus. In the feature selection stage, two different criteria, the mutual information selection and the F-score measurement were independently used to select powerful features and then compare them. It revealed that features selected by the two different methods were not significantly different, and could potentially be reliable for classification. Meanwhile, radial basis function networks were also designed to discriminate test images into one of the four disease groups in the classification stage. The percentage of correct classification was more than 92.5% using this proposed automatic computer system. Experimental results show that the proposed method performed very well for the classification of ultrasonic supraspinatus images.

Keywords: Rotator cuff, Supraspinatus, Ultrasound images, Mutual information, Radial basis function network

1. Introduction
Clinical history and examination are essential for determining the origin of the diagnosis of shoulder pain, a common orthopedic problem. One of the most common causes of shoulder pain is rotator cuff disease, which can result from impingement syndrome or tearing into the rotator cuff tendons. The rotator cuff is made up of four muscles (the supraspinatus, infraspinatus, subscapularis and teres minor) and the associated tendons that surround the humeral head of the shoulder and originate at the scapula. Among the four muscles, the supraspinatus of the rotator cuff is easiest to injure or tear during energetic exercise. For this reason, clinical physicians routinely observe injuries to it.

Neer’s classification system recently became a popular method for separating different diseases of supraspinatus into three stages in clinical diagnosis [1]. Inflammation manifestations, such as edema or hemorrhage, usually exhibit in supraspinatus of Stage 1; thus, all ultrasonic images of this stage are collected for the Inflammation Group for subsequent classification. The supraspinatus of Stage 2 is more serious and considered irreversible. Fibrosis and calcification always appear, hence, the images of this stage are regarded as the Calcific Group. Stage 3 generally occurs in patients over 50 years old and frequently involves a tendon rupture or tear; if the tendon rupture arises in the supraspinatus, it may require further repair. The images of Stage 3 are referred to as the Tear Group.

Ultrasongraphy (US) has been proven to be a useful diagnostic tool in patients with shoulder pain and/or limited range of motion. Iagnocco et al. explained how to use US for the careful qualitative assessment of a wide range of changes in different anatomic structures of rotator cuff tendons such as tendonitis, tendon tears and calcific deposits [2]. To date, most studies continue to use a subjective evaluation in terms of a visual inspection of the images or measurements of muscle disorders. Arsian et al. used a 7.5-MHz linear-array transducer to grab images of patients with physical examination suggestive of rotator cuff injury under longitudinal view [3]. The study demonstrated that the bursa fluid and biceps effusion were highly correlated with symptoms of rotator cuff injury. In addition, Chiu et al. proposed diagnostic criteria for the classification of full/partial supraspinatus tears in patients

* Corresponding author: Ming-Huwi Horng
Tel: +886-8-7238700; Fax: +886-8-7223962
E-mail: horng@npic.edu.tw
with shoulder pain [4]. The sensitivity and specificity of experiments in applying the diagnostic criterion were 0.98 and 0.87. Furthermore, Marcello et al. used specific criteria—(a) one or more cuff tendon(s) being not visible, (b) focal non-visibility of one the tendons, and (c) defect of well-defined discontinuity of the tendons—to diagnose rotator cuff tear [5]. Finally, Al-Shawi et al. studied the detection of full-thickness rotator cuff tears using ultrasound [6]. The accuracies for the detection of large and massive tears, moderate tears and small tears were 96.5 percent, 88.8 percent and 91.6 percent, respectively.

Quantitative evaluations have included the distribution of the gray-level scales of the pixel in the image to describe tissue characteristics. However, muscles with different muscle structure in terms of spatial distributions can be determined through quantitative ultrasonic image analysis. Nielsen et al. proposed the so-called “blob analysis” method that was related to the higher-order gray-level statistics of the image for quantitative ultrasound tissue characterization of the discrepancy between the supraspinatus muscle and the right vastus lateralis muscle [7,8]. The authors found that first-order histogram features were effective in the classification of the shoulder and thigh muscles. However, this approach failed to provide accurate data for the discussion of the classification of other impingement syndromes such as inflammation and tendon calcification of the rotator cuff. In addition, in our past studies, we have proposed a comparative article using the various multi-class support vector machines to classify ultrasonic supraspinatus images into the four disease groups [9]. The experimental results revealed that the one-against-all fuzzy support vector machine was the most powerful for classifying ultrasonic supraspinatus images, resulting in a classification rate of 90 percent.

This paper uses an alternative classification method, radial basis function network, to classify the four disease groups from the ultrasonic supraspinatus images. The feature extraction stage comprises four texture analysis methods—the gray-level co-occurrence matrix [10], texture spectrum [11], fractal dimension method [12-14] and texture feature coding method [15]—that are used to compute adequate texture features. In order to verify the reliability of selected features, two different feature selection methods, the mutual information criteria [16,17] and F-score [18] ranking methods, were independently applied to identify the powerful features among the information generated from the four texture analysis methods for comparison. Finally, the radial basis function network was trained and then used to classify test images for disease classification [19,20]. The experiments compared the classification results with those of previous studies by using the multi-class support vector machine (SVM).

2. Materials and methods

This section will briefly describe methodology adopted in this study supporting the feature extraction method, feature selection method, classification method and performance evaluation.

2.1 Data acquisition and system equipment

All the ultrasonic images used were recorded from 2004 to 2007, and the ages of patients ranged from 30 to 65 years. In all, 120 shoulders in 120 patients with shoulder pain who had undergone preoperative and subsequent arthroscopy were identified. The arthroscopy diagnosis was thickness tear in 30, tendon inflammation in 30, calcific tendon in 30 and normal in 30. A longitudinal view of an ultrasound image of each shoulder was acquired using an HDI Ultramark 5000 Ultrasound system (ATL Ultrasound, Stockton, CA, USA) fitted with a 5.0 MHz dynamic focusing transducer (C5-40 5.0 MHz Curved Linear Array, ATL Ultrasound, Stockton, CA) from National Cheng Kung University Hospital based on current the clinical setting for ultrasound examination. The captured images were digitized into 256 × 256 pixels with 256 gray levels via a frame grabber and then stored on a disk.

Figure 1. Image (a) is a normal supraspinatus case. Image (b), tendon inflammation, is a sample of tendon inflammation. The images (c) and (d) are calcific tendonitis and supraspinatus tear, respectively.

The ultrasonic system settings were standardized for all of the participants and kept constant during the image acquisition. We used a depth setting of 3.0 cm. The depth-gain compensation was built into the ultrasound machine. The acoustic signal received by the ultrasonic transducer was digitized by eight-bit intensity values to make the ultrasonic image. Each image consisted of pixels that were 0.1172 mm × 0.1172 mm = 0.0137 mm². For each image, a region of interest (ROI) with 30 × 60 pixels at a depth of approximately 1.5 to 2.5 cm from the body surface was manually selected by ultrasound musculoskeletal radiologists with five years of experience in shoulder examination to extract texture features for subsequent classification. In general, radiologists avoided including the areas of ruptured tendons in the ROI selection process. All programs were implemented in Visual C++ associated with QuickRBF [19] Toolbox of Matlab software on a personal computer with a 2.4-GHz CPU and 1G RAM using the Window XP operating system. The execution time was 0.238 seconds for classifying a supraspinatus image using this proposed method. Figure 1 provides sample images of normal, tendon inflammation, calcific tendonitis and rotator cuff tears. Among the 120 acquired images, 40 supraspinatus
images equally divided into the four classes were selected as the training data to search for powerful features and then to establish the RBF network for classification. The remaining 80 supraspinatus images were used as the test images for subsequent classification.

2.2 Feature extraction methods

Texture-based measures have been applied to ultrasound images for over a decade. Horng et al. compared the effectiveness of texture descriptors that include the gray-level co-occurrence matrix (GLCM), the fractal dimension (FD), the texture spectrum (TS), the statistical feature matrix, and the texture feature coding method (TFCM) in classifying the chronic liver diseases (i.e., normal liver, hepatitis and cirrhosis) [15,21]. Features generated from the gray-level co-occurrence matrix and texture feature coding method were effective for classifying the three liver states. In the current studies, we also adopted GLCM, FD, TS and TFCM to extract features for classifying ultrasonic supraspinatus images.

A co-occurrence matrix is generally referred to as a gray-level co-occurrence matrix whose entries are transitions between all pairs of two gray levels (not necessarily distinct). The gray-level transitions are calculated based on two parameters: displacement \( d \) and angular orientation \( \theta \). More precisely, let \( i \) and \( j \) be two gray-levels, while \( N_{d,\theta}(i,j) \) denotes the number of transitions between two pixels whose gray levels are \( i \) and \( j \), are \( d \) pixels apart and have angular orientation \( \theta \). In other words, \( N_{d,\theta}(i,j) \) is the number of pixel-pairs at locations \((x, y)\) and \((w, z)\) satisfying the following conditions:

\[
G(x, y) = i, G(w, z) = j \quad \| (x, y) - (w, z) \|_{\infty} = (d, \theta) \tag{1}
\]

where \( \| (x, y) - (w, z) \|_{\infty} = (d, \theta) \) is a distance measure to describe the distance between two pixels of spatial locations at \((x, y)\) and \((w, z)\) that are \( d \) pixels apart and have angular orientation \( \theta \). Normalizing \( N_{d,\theta}(i,j) \) yields the probability or the relative frequency of gray-level transitions.

\[
p(i, j) = \frac{N_{d,\theta}(i,j)}{N} \tag{2}
\]

where \( N \) is the number of total gray-level transitions in the co-occurrence matrix.

Haralick et al. suggested that 14 texture features be extracted from the co-occurrence matrices [10]. These feature measures were used in experiments for comparison where \( d = 1, 2, 3 \) and 4 pixels under all angular orientations. The features are described in Appendix I.

The concept of FD introduced by Mandelbrot provided a good representation of the roughness of natural surfaces [12]. The variation method and box-counting method are used to extract the FD for lateral analysis. Dubuc et al. proposed evaluation of FD by computing the maximum variation, \( V(\varepsilon) \), of an image intensity surface \( f(x, y) \) in a square \( \varepsilon \) [13]. The maximum variation is defined as

\[
V(\varepsilon) = \int_0^{\varepsilon} \int_0^{\varepsilon} |f(x, y) - f(x, y + \varepsilon)| \, dx \, dy \tag{3}
\]

where \( f(x, y) = \sup|f(x, y - 1) - f(x, y + 1)| \) and the supreme is taken over all pairs \((x_1, y_1), (x_2, y_2)\) for which \( \max(|x_1 - x_2|, |y_1 - y_2|) \leq \varepsilon \). The FD can be shown to satisfy the following equation.

\[
FD = \lim_{\varepsilon \to 0} \frac{\ln V(\varepsilon)}{\ln \varepsilon} \tag{4}
\]

That is, the FD is estimated as three minus the slope of the least square fit of the data: \( \{\ln \varepsilon, \ln (V(\varepsilon))\} \) as \( \varepsilon \).

In the box-counting method, the rate of growth of the number of boxes needed to cover a volume of the gray-level surface is computed as the FD. Let \( N(L) \) denote the number of boxes of size \( L \) needed to cover the gray-level volume. Keller et al. estimated FD based on the following equation [14]:

\[
FD = -\frac{\ln N(L)}{\ln L} \tag{5}
\]

In other words, the FD can be estimated for the least square linear fit of \( \ln N(L) \) and \(-\ln L\).

He and Wang first proposed the texture spectrum, which considers a so-called texture with the central pixel as \( v_0 \), designated as the pixel currently being examined and its eight neighboring pixels \( v_i, i > 0 \) [11]. Three values \( \{0, 1, 2\} \) are assigned to \( v_i \), respectively, according to Eq. (6) as follows:

\[
E_i = \begin{cases} 0 & \text{if } |v_i - v_0| < \Delta \\ 1 & \text{if } |v_i - v_0| \leq \Delta \\ 2 & \text{if } |v_i - v_0| > \Delta \\ \end{cases} \quad i = 1, 2, \ldots, 8
\]

\[
N_{E_i} = \sum_{j=1}^{8} E_j \cdot 3^{j-1} \tag{6}
\]

In Eq. (6), \( \Delta \) denotes the tolerance of variation. In experiments, \( \Delta \) is assigned as 3. Obviously, there are \( 3^8 = 5611 \) combinations for the \( E_i \) in Eq. (6). \( N_{E_i} \) denotes texture number, while the distribution of occurrence of all texture numbers, \( S(N_{E_i}) \), is the texture spectrum. As the texture unit represents the local texture information of a given pixel and its neighborhood, the statistics of all texture units in an image reveal its global texture aspects. The eight features generated from the texture spectrum are listed in Appendix II.

The TFCM is a coding and rotation-invariance scheme that transforms an original gray-scale image into a texture feature image whose pixels are represented by a texture feature number (TFN). The TFN of each pixel \( X \) is generated according to the gray-level changes in its eight surrounding pixels of the \( 3 \times 3 \) texture unit (TU). The TFCM method generates 55 TFNs of each TU. In this case, we can define a TFN histogram by

\[
p_n = \frac{N_n}{N}, \quad n \in \{0,1,2,\ldots,54\} \tag{7}
\]
where $N_c(n)$ is the frequency of occurrence of the texture feature number $n$, $N$ is the total number of pixels in the feature image and $\Delta$ is a predefined parameter. Furthermore the probability of transitions of TFN $i$ and TFN $j$ in the displacement $(d, \theta)$ is defined in Eq. (8).

$$p_s(i,j|d, \theta) = \frac{N_{i,j,d}(i,j)}{N}, i,j \in \{0,1,2,\ldots,54\} \tag{8}$$

where $N_{i,j,d}(i,j)$ is defined similarly as in Eq.(2), $i$ and $j$ are TFNs rather than gray-levels and $N_i$ is the total number of TFN transitions. In experiments, displacement $d$ is chosen to be one and two pixels under angular orientation $\theta$ covering all eight orientations of 0, 45, 90, 135, 180, 225, 270 and 315°. In addition, the $\Delta$ is assigned as 3. The seven texture features are extracted from the TFN histogram and TFN co-occurrence matrix, as shown in Appendix III.

In summary, each ROI $R$ of the supraspinatus muscle can extract 80 features that are generated by the abovementioned four texture analysis methods detailed herein. In these features, 56 features were generated from GLCM, eight features from TS, four texture analysis methods detailed herein. In these features, extract 80 features that are generated by the abovementioned method. The universal algorithms of feature selection are often divided into two groups: wrapper and filter approaches. The wrapper model consists of two phases, feature subset selection phase and learning phase, while the filter approach is built on the intrinsic properties of the data, not on a particular classifier. A filter model of feature selection also consists of two phases: 1) feature selection, which uses certain measures such as mutual information and F-score measurement as search criteria, and 2) the classifier is learned on the training data with the selected features [22]. The two feature selection methods (i.e., mutual information and F-score measure) can be used as search criteria for independently searching for features for subsequent comparison.

First, we describe the mutual information criteria for feature selection. Given a set of $n$ training samples $(x_i, c_i)$ of $d$-dimensional continuous features $X, x_i = (f_{i,1}, f_{i,2}, \ldots, f_{i,d}) \in R^d$; and class labels as samples of discrete-valued random variable $C, c_i \in \{1,2,\ldots,N\}$. In experiments, the class number $N$ is assigned to 4, while $c_4$ is the normal disease group, $c_2$ is calcific tendonitis, and $c_3$ is calcific tendonitis, and $c_4$ is rotator cuff tear. The objective of mutual information (MI) is to find the subset $S \subseteq X$ with $k$-dimensional features ($k < d$) that maximizes the MI of $I(S,C)$. More precisely, the feature $s_j$ of $S$ can be represented in the formula of $s_j = \{f_{1,j}\sigma_1, f_{2,j}\sigma_2, \ldots, f_{d,j}\}$. The details of the computation are described below.

In classification problems, the class $c_i$ has discrete values, while the input features are usually continuous variables. In this case, the mutual information $I(S,C)$ between the selected features $S$ and class $C$ can be expressed as follows:

$$I(S;C) = H(C) - H(C \mid S), \tag{9}$$

$$H(C) = -\sum_{c \in C} p(c) \log p(c), \tag{10}$$

$$H(C \mid S) = -\int s p(s) \sum_{c \in C} p(c | s) \log p(c | s) ds, \tag{11}$$

where $s$ is the instances of selected features.

$H(C)$ is the entropy function; it can be easily calculated using Eq. (10). However, the computation of the conditional entropy $H(C|S)$ is very difficult because $p(s)$ and $p(c|s)$ are hard to calculate. Kwak and Choi proposed the Parzen window method for estimating the probabilities of $p(s)$ and $p(c|s)$ [16]. Battiti adopted a greedy forward sequential search method to search for the most relevant features from the input feature set [23]. In this scheme, starting from the empty set of selected features, we add the best available feature to the selected feature set one by one until the mutual information between selected features and their corresponding class labels is maximized.

Feature ranking approaches use a principal or auxiliary mechanism to select the best feature set for classification. Because of their simplicity and scalability, the approaches have been widely applied. The F-score ranking method is one feature ranking approach. The larger the F-score of a feature is, the more likely this feature is to be more discriminative. Given training features $x_k, k = 1, 2, \ldots, n$, if the number of positive and negative instances are $n_+$ and $n_-$, respectively, then the F-score of the $i^{th}$ feature is defined as follows:

$$F(i) = \frac{1}{n_+ - 1} \sum_{k=1}^{n_+} (x_{i,k}^{(+)} - \bar{x}_i^{(+)})(x_{i,k}^{(+)} - \bar{x}_i^{(-)}) + \frac{1}{n_- - 1} \sum_{k=1}^{n_-} (x_{i,k}^{(-)} - \bar{x}_i^{(-)})(x_{i,k}^{(-)} - \bar{x}_i^{(+)}) \tag{12}$$

In Eq. (12), the $\bar{x}_i^{(+)}$, $\bar{x}_i^{(-)}$, and $\bar{x}_i$ are the average of the $i^{th}$ feature of the whole, positive and negative data sets, respectively; $x_{i,k}^{(+)}$ is the $i^{th}$ feature of the $k^{th}$ positive instance, and $x_{i,k}^{(-)}$ is the $i^{th}$ feature of the $k^{th}$ negative instance. In the experiments, we calculated the average of F-scores that are obtained by computing between two different groups in order to analyze the discrimination of each texture feature. Chen and Lin proposed this procedure for feature selection [24].

### 2.4 Classification method

A radial basis function (RBF) network can be considered a special three-layered network [20]. The input nodes pass the input values to the internal nodes that formulate the hidden layer. Each unit of hidden layer implements an activation function called radial basis function. The nonlinear responses of hidden nodes are weighted in order to calculate the final
outputs of network in the output layer. The input layer of this network has \( l \) units for \( l \)-dimensional input vectors. The input units are fully connected to \( n_{h} \) hidden layer units, which are in turn fully connected to the \( n_{o} \) output layer units, where \( n_{o} \) is the number of output classes.

The activation functions of the hidden layer are chosen to be Gaussians and are characterized by their mean vectors \( \mu_{i} \) and covariance matrices \( C_{i} \), \( i = 1,2,\ldots,n_{h} \). For simplicity, it is assumed that the covariance matrices that are in the form \( C_{i} = \sigma_{i}^{2}I \), \( i = 1,2,\ldots,n_{h} \). Then the activation function of the \( i^{th} \) hidden unit for an input vector \( x_{j} \) is given by:

\[
\phi_{i}(x_{j}) = \exp\left[-\frac{\|x_{j} - \mu_{i}\|^{2}}{2\sigma_{i}^{2}}\right]
\]

The \( \mu_{i} \) and \( \sigma_{i}^{2} \) are estimated using a suitable clustering algorithm such as \( k \)-means clustering. The number of activation functions in the network and their spread influence the smoothness of the mapping. The number of hidden units is empirically determined and it is assumed that \( \sigma_{i}^{2} = \sigma^{2} \), in which \( \sigma^{2} \) is given in Eq. (14).

\[
\sigma^{2} = \frac{d\eta^{2}}{2}
\]  

(14)

In this equation, \( d \) is the maximum distance between the chosen centers, and \( \eta \) is an empirical scale factor that serves to control the smoothness of the mapping function. Therefore, Eq. (13) can be rewritten as:

\[
\phi_{i}(x_{j}) = \exp\left[-\frac{\|x_{j} - \mu_{i}\|^{2}}{d\eta^{2}}\right]
\]

(15)

The hidden layer units are fully connected to the \( n_{o} \) output layer units through weight \( w_{ij} \). The response of the \( k^{th} \) output unit for an input vector \( x_{j} \) is given by Eq. (16).

\[
y_{i}(x_{j}) = \sum_{j=0}^{n_{o}} w_{ij} \phi_{i}(x_{j}), \quad k = 1,\ldots,n_{o}
\]

(16)

where \( \phi_{i}(x_{j}) = 1 \).

Training the RBF network includes two stages [21,25]. First, the basis functions must be established using an algorithm to cluster data in the training set. Conventionally, the unsupervised \( k \)-means clustering algorithm can be applied to find the cluster means \( \mu_{i} \). Parameter \( d \) is then computed by finding the maximum distance between cluster means.

Next, it is necessary to determine the weights \( w_{ij} \), between the hidden and output layer. Given that the centers and widths of activation function form \( n_{h} \) training vectors. Eq. (16) may be written in matrix form as

\[
Y = GW
\]

(17)

where \( Y \) is a \( n_{o} \times n_{h} \) matrix with elements \( Y_{ij} = y_{i}(x_{j}) \), \( G \) is a \( n_{h} \times (n_{h} + 1) \) matrix with elements \( G_{ij} = y_{i}(x_{j}) \), and \( W \) is a \( (n_{h} + 1) \times n_{o} \) matrix of unknown weights. \( W \) can be determined from the standard least square solution:

\[
W = (G^{T}G)^{-1}G^{T}Y
\]

(18)

To solve for \( W \) from Eq. (18), \( G \) is specified from the clustering results. The elements of \( Y \) are computed as

\[
y_{i} = \begin{cases} 1 & \text{if } x_{i} \in \text{class } j, \\ 0 & \text{otherwise} \end{cases}
\]

(19)

2.5 Performance evaluation

Several options exist for evaluating the performance of classification algorithms. In general, measures of quality of classification are built from a confusion matrix which records correct and incorrect recognition, such as the true positive (TP), false positive (FP), false negative (FN) and the true negative (TN), in binary classification. In order to extend the usage of the confusion matrix, we defined TP, FP, FN and TN in this paper as follows.

TP: TP denotes the number of patients having the supraspinatus diseases in question and being diagnosed as positive.

FP: FP denotes the number of patients that are classified with more severe diseases than actual diagnosis.

FN: FN denotes the number of patients that are classified with less severe disease than actual diagnosis.

TN: TP denotes the number of patients having no the supraspinatus diseases in question and being diagnosed as negative.

The definitions of the sensitivity, specificity and accuracy of classification are as follows.

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

(20)

\[
\text{Sensitivity} = \frac{TP}{TP + FN}
\]

(21)

\[
\text{Specificity} = \frac{TN}{FP + TN}
\]

(22)

The accuracy measure assesses the overall effectiveness of the algorithm, while the other two indices (sensitivity and specificity) estimate the classifier’s performance in different classes. These two indices are often employed in medical applications.

The receiver operating characteristics (ROC) analysis is based on statistical decision theory and has been applied extensively to the evaluation of classification methods [26]. The ROC curve can manifest the relationship between the true-positive fraction (TPF) and false-positive fraction (FPF) with the variations in decision threshold [27,28]. In general, the area under the ROC curve (AUC), \( A_{c} \), is a powerful index for assessing the classification performance of the classifier. In general, a large value of AUC is desirable as AUC values greater than 0.9 suggest that the corresponding diagnosis system is very effective [29].

3. Results and discussion

3.1 Comparison of the selected features

In the experiment, each training image could be extracted
to a total of 80 features that were generated from the abovementioned four texture analysis methods. All extracted features were first required to be normalized to zero mean and unit standard deviation which ensured the larger value input features so as not to overwhelm smaller value inputs and to reduce errors before the feature selection and classification. Two search criteria, mutual information criteria and F-score ranking measure, were adopted to conduct the dimension reduction. Table 1 lists the selected features by independent usage of the two criteria. Method I used mutual information criterion to extract features from the 80 texture features, while Method II adopted the F-score measure. The selected texture features of Method I included sum average (GLCM), sum variance (GLCM), mean convergence (TFCM), contrast (GLCM) and difference variance (GLCM); meanwhile, the selected textures of Method II were sum variance (GLCM), sum average (GLCM), mean convergence (TFCM), code variance (TFCM) and contrast (GLCM). The results of the texture feature selection were that all of the selected features were generated from the GLCM and TFCM. This result may reveal that the discriminative capability of the selected features generated from the GLCM and TFCM methods are superior to those of other methods.

Table 1. Texture feature selection by using the four texture analysis methods.

<table>
<thead>
<tr>
<th>Methods</th>
<th>Feature selected ($d$: displacement)</th>
</tr>
</thead>
<tbody>
<tr>
<td>I. Four methods + MI selection</td>
<td>Sum average (GLCM, $d=2$), Sum variance (GLCM, $d=2$), Mean convergence (TFCM, $d=2$), Contrast (GLCM, $d=3$), Difference variance (GLCM, $d=2$)</td>
</tr>
<tr>
<td>II. Four methods + F-scoring ranking</td>
<td>Sum variance (GLCM, $d=2$), Sum average (GLCM, $d=2$), Mean convergence (TFCM, $d=2$), Code variance (TFCM, $d=2$), Contrast (GLCM, $d=3$)</td>
</tr>
</tbody>
</table>

3.2 Performance evaluation

In order to further investigate the performance of classification using the two previously selected feature sets in the RBF classifier, the accuracy, sensitivity and specificity were computed (see Tables 2 and 3). The classification accuracy of Method I was 92.5 percent, which is better than that of Method II’s accuracy. The sensitivity and specificity of Method I were 0.931 and 0.909, respectively, which were also better than those of Method II. This revealed that the usage of the Method I was relatively effective in the disease classification. In addition, an effective classification method should decrease the possibility of misclassification, especially for the false-negative rate. A high false-negative rate represents the danger to underestimate the disease severity in a patient when the clinical doctor uses the classification system; therefore, the false-negative rate may be considered as an index for evaluating the performances of the RBF networks of the two different selected feature sets. The four performance indices of the two different selected feature sets are listed in Table 4. The false-negative rate by using the features of Method I was 0.05. The satisfactory results by using the features of Method I revealed that it is promising to develop a diagnosis tool using this for clinical application.

Table 2. Classification results based on the features of Method I using RBF network classifier.

<table>
<thead>
<tr>
<th>Predicted results</th>
<th>Actual results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Predicted Normal</td>
<td>Actual Normal</td>
</tr>
<tr>
<td>Predicted Inflammation</td>
<td>Actual Inflammation</td>
</tr>
<tr>
<td>Predicted Calcific</td>
<td>Actual Calcific</td>
</tr>
<tr>
<td>Predicted Tear</td>
<td>Actual Tear</td>
</tr>
</tbody>
</table>

Table 3. Classification results based on the features of Method II using RBF network classifier.

<table>
<thead>
<tr>
<th>Predicted results</th>
<th>Actual results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Predicted Normal</td>
<td>Actual Normal</td>
</tr>
<tr>
<td>Predicted Inflammation</td>
<td>Actual Inflammation</td>
</tr>
<tr>
<td>Predicted Calcific</td>
<td>Actual Calcific</td>
</tr>
<tr>
<td>Predicted Tear</td>
<td>Actual Tear</td>
</tr>
</tbody>
</table>

Table 4. The four performance indices, which are accuracy, sensitivity, specificity and the corresponding false-negative rate (FNR) with fixed scale factor $\eta = 1.0$ of RBF network classifier.

<table>
<thead>
<tr>
<th>Performance index</th>
<th>Method I</th>
<th>Method II</th>
<th>Radiologist’s observation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy</td>
<td>0.925</td>
<td>0.875</td>
<td>92.5%</td>
</tr>
<tr>
<td>Sensitivity</td>
<td>0.931</td>
<td>0.911</td>
<td>0.932</td>
</tr>
<tr>
<td>Specificity</td>
<td>0.909</td>
<td>0.791</td>
<td>0.904</td>
</tr>
<tr>
<td>FNR</td>
<td>0.05</td>
<td>0.0625</td>
<td>0.05</td>
</tr>
</tbody>
</table>

The corresponding ROC curves for the two different texture feature sets are depicted in Fig. 2. Table 5 shows that the AUCs of ROC curves of the two methods were 0.948 and 0.923, respectively. This result showed that the trained RBF network with features of Method I had considerable success in the classification of ultrasonic supraspinatus images. In order to compare the classification results of proposed Method I with the results of diagnosis of clinical musculoskeletal radiologists, the acquired images were evaluated in consensus by two radiologists with five and 10 years of experience in the shoulder ultrasound examination, respectively. This classification result is shown in Table 6. The sensitivity and specificity of Table 6 are 0.931 and 0.904. In Table 6, the classification of images with supraspinatus tears is perfectly classified and is superior to Table 2 of the usage of the Method I. This may reveal that it is still an interesting viable direction to improve the effectiveness of the classification of images with supraspinatus tears for further research.

Table 5. The $A_z$ of ROC curves of different texture features by RBF neural network under different scale factors.

<table>
<thead>
<tr>
<th>Area of ROC Curve</th>
<th>Method I</th>
<th>Method II</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A_z$ value</td>
<td>0.948</td>
<td>0.923</td>
</tr>
</tbody>
</table>

Table 6. The classification result based on the radiologist’s observation.

<table>
<thead>
<tr>
<th>Actual results</th>
<th>Predicted results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Predicted Normal</td>
<td>Actual Normal</td>
</tr>
<tr>
<td>Predicted Inflammation</td>
<td>Actual Inflammation</td>
</tr>
<tr>
<td>Predicted Calcific</td>
<td>Actual Calcific</td>
</tr>
<tr>
<td>Predicted Tear</td>
<td>Actual Tear</td>
</tr>
</tbody>
</table>
there is still a need for the improvement in the classification of supraspinatus tear images in further research. The investigation of how to integrate the merit of other texture analysis methods, such as wavelet transforms [30] and Gabor transforms [31], may be one direction for further research. Furthermore, other search methods, such as the support vector machine-based recursive feature elimination [32] and the evolutionary-based feature selection approaches [33], are alternatives in future research for improving the classification of ultrasonic supraspinatus images.

Acknowledgement

The authors would like to thank the National Science Council, ROC, under Grant No. NSC 97-2213-E-251-001 for support of this work.

Appendix I: Texture features of the gray-level co-occurrence matrix notation

Notation

\( p(i,j) \) \( (i,j) \)th entry in a normalized GLCM matrix

\( p_r(i) \) \( i \)th entry in the marginal-probability matrix obtained by summing the rows of \( p(i,j) \)

\( N_g \) Number of gray-levels of the supraspinatus image.

\( P_{xy} \) \( x \) and \( y \) the means and standard deviations of \( p_r \) and \( p_g \)

\( H_{XY} \) the variance of \( p_r \)

\( H_{X} \) the variance of \( p_g \)

\( I \) the inverse difference moment

\( F \) the difference entropy

\( TPF \) the true positive fraction

\( TNR \) the true negative rate

\( 
\begin{align*}
\text{Actual results} & \quad \text{Predicted results} \\
\text{Normal} & \quad 18 & 2 & 0 & 0 \\
\text{Inflammation} & \quad 2 & 18 & 0 & 0 \\
\text{Calcific} & \quad 0 & 1 & 18 & 1 \\
\text{Tear} & \quad 0 & 0 & 1 & 19 \\
\end{align*}
\end{align*}
\)

Appendix II: Texture features of the gray-level co-occurrence matrix notation

\( p(i,j) \) \( (i,j) \)th entry in a normalized GLCM matrix

\( p_r(i) \) \( i \)th entry in the marginal-probability matrix obtained by summing the rows of \( p(i,j) \)

\( N_g \) Number of gray-levels of the supraspinatus image.

\( p_g(k) \) \( k \)th entry in the marginal-probability matrix obtained by summing the columns of \( p(i,j) \)

\( p_{xy} \) \( x \) and \( y \) the means and standard deviations of \( p_r \) and \( p_g \)

\( H_{XY} \) the variance of \( p_r \)

\( H_{X} \) the variance of \( p_g \)

\( I \) the inverse difference moment

\( F \) the difference entropy

\( TPF \) the true positive fraction

\( TNR \) the true negative rate

\( 
\begin{align*}
\text{Actual results} & \quad \text{Predicted results} \\
\text{Normal} & \quad 18 & 2 & 0 & 0 \\
\text{Inflammation} & \quad 2 & 18 & 0 & 0 \\
\text{Calcific} & \quad 0 & 1 & 18 & 1 \\
\text{Tear} & \quad 0 & 0 & 1 & 19 \\
\end{align*}
\end{align*}
\)

Figure 2. ROC curves of the two selected feature sets.
where
\[ HXY = \sum_{i,j} p(i,j) \log p(i,j), \quad HXY1 = \sum_{i,j} p(i,j) \log \{p(i) p(j)\}, \]
\[ HXY2 = \sum_{i,j} p(i,j) \log \{p(i) p(j)\} \]

(14) Maximal correlation coefficient:
\[ f_{si} = (\text{Second largest eigenvalue of } Q_i)^{\frac{1}{2}}, \quad \text{where } Q_i = \sum_{j} p(i,j) p(j,i) \]

Appendix II: Texture features of the texture spectrum

Notation
- \( S(i) \): The texture spectrum, \( i \) is the texture number
- \( S(i,j) \): The texture spectrum of the corresponding texture unit
- \( P(V_i, V_j) \): The number of the same value of \( V_i \) and \( V_j \)
- \( K(i) \): The number of the same value in the number pair \((V_i, V_j), (V_i, V_j), (V_j, V_j) \) and \((V_j, V_j)\)

Texture features

(1) Black-white symmetry:
\[ BWS = \frac{1}{100} \left[ 1 - \frac{\sum_{i=1}^{128} S(i) \cdot s(6560-i)}{\sum_{i=1}^{128} S(i)} \right] \]

(2) Geometric symmetry:
\[ GS = \frac{1}{100} \left[ 1 - 0.25 \times \frac{\sum_{i=1}^{128} S(i) - S(i+6)}{2 \times \sum_{i=1}^{128} S(i)} \right] \]

(3) Degree of direction:
\[ DD = \frac{1}{6} \times \left[ \sum_{i=1}^{128} \frac{S(i) \cdot S(i+6)}{2 \times \sum_{i=1}^{128} S(i)} \right] \]

(4) Orientation features:
\[ MHS = \sum_{i=1}^{128} S(i) \times P(V_i, V_j, V_k) \times P(V_i, V_j, V_k) \]
\[ MYS = \sum_{i=1}^{128} S(i) \times P(V_i, V_j, V_k) \times P(V_i, V_j, V_k) \]
\[ MDS1 = \sum_{i=1}^{128} S(i) \times P(V_i, V_j, V_k) \times P(V_i, V_j, V_k) \]
\[ MDS2 = \sum_{i=1}^{128} S(i) \times P(V_i, V_j, V_k) \times P(V_i, V_j, V_k) \]

(5) Central symmetry:
\[ CS = \frac{1}{128} \cdot \text{number of } S(i) \cdot K(i)^2 \]

Appendix III: Texture features of the texture feature coding method

Notation
- \( p_x(n) \): The probability of texture feature number (TFN)
- \( p_x(i,j,d,\theta) \): The probability of transitions of TFN \( i \) and TFN \( j \) in the displacement \((d,\theta)\)
- \( p(i,j,x,y) \): The joint probability of TFN \( i \) of pixel \((x, y)\) in the TFN-co-occurrence matrix with \( x = 0 \) and TFN \( j \) of the same pixel \((x, y)\) in the TFN-co-occurrence matrix with \( x = 3 \)

Texture features

(1) Variance:
\[ \text{Var} = \sum_{x} p_x(n) \cdot (n - \mu_x)^2 \]

(2) Homogeneity:
\[ \text{Hom} = \sum_{x} p_x(0) \]

(3) Mean convergence:
\[ \text{MC} = \sum_{x} p_x(n) - \mu_x \]

References
