MAJOR PAPER

Combined Image Enhancement, Feature Extraction, and Classification Protocol to Improve Detection and Diagnosis of Rotator-cuff Tears on MR Imaging

Chen-Chiang LIN^{1,2}, Chih-Nan WANG², Yang-Kun OU², and Jachih FU^{2*}

¹*Department of Orthopedics, National Taiwan University Hospital (Yunlin Branch)
No. 579, Sec. 2, Yunlin Rd.; Douliu City, Yunlin County; 640 Taiwan (R.O.C.)*

²*Computer-aided Measurement and Diagnostic Systems Laboratory, Department of Industrial Engineering and Management, National Yunlin University of Science and Technology
No. 123, Sec. 3, University Rd.; Douliu City, Yunlin County; 640 Taiwan (R.O.C.)*

(Received August 7, 2013; Accepted January 29, 2014; published online July 2, 2014)

Purpose: The diagnosis of most rotator cuff tears (RCTs) relies upon magnetic resonance (MR) imaging, but direct capture of MR images without enhanced image processing leads to poor image contrast and potential misdiagnosis. Therefore, we developed a 2-stage model for the detection and diagnosis of injury of the supraspinatus tendon.

Methods: The first stage used coupled weighted histogram separation (WHS) to improve image enhancement, and the second stage extracted suspicious texture, features of both spatial and spectral domains, and sequential floating forward selection (SFFS) selected features conducive to classification of RCTs. We then tested injuries of the supraspinatus tendon using the classifier.

Results: The extraction of features by SFFS can increase detection of supraspinatus injury by reducing the input vector by 57.78% from the enhanced input images. The receiver operating characteristic (ROC) curve indicated an azimuth (Az) value of 84.38% when SFFS selected 76 features to construct a support vector machine (SVM) classifier from the enhanced images, compared with 56.94% when all 180 features from the raw input images were used for the construction.

Conclusions: The performance of the classifier constructed by SFFS-selected features is superior to that using all features. These findings can serve as references to improve diagnosis and treatment of supraspinatus injuries.

Keywords: *data mining, feature extraction, image enhancement, magnetic resonance imaging, support vector machine, supraspinatus*

Introduction

The rotator cuff comprises the subscapularis, supraspinatus, infraspinatus, and teres minor muscles and is responsible for the elevation and rotation of the shoulder joint. Currently, the diagnosis of rotator cuff tears (RCTs) relies on magnetic resonance (MR) imaging. Generally, The T₂-weighted imaging sequence is used to identify injuries of the supraspinatus tendon, but visual interpretation of images without image enhancement can result in misdiagnosis.

Accurate diagnosis relies upon the quality of the MR imaging. Advances in computer technology have led to the wide use of digital computer image processing in the diagnosis of a variety of diseases,^{1–4} producing a consistent output that enables more reliable diagnosis. Enhancement is intended to make images more suitable for visual inspection and/or subsequent computer-aided analysis. Image contrast is essential to improve visual quality and is characterized by a histogram.

In image analysis, the histogram of an image is a graphical representation that illustrates the number of pixels in an image at each intensity value. Because the histogram is intuitive and straightforward, histogram-based techniques for image enhancement

*Corresponding author, Phone: +886-972-655-417, Fax: +886-5-537-9742, E-mail: drlincc@yahoo.com.tw

are popular. Pei and associates⁵ proposed a WHS to enhance image contrast to eliminate the shortcomings of histogram equalization. Weighted histogram separation (WHS) is the most widely used histogram-based image enhancement technique. The flexible use of weighted parameters permits adjustment of the WHS to the user's satisfaction. Zeng and Liao⁶ proposed a revised WHS to reduce the block effect in a WHS image.

The identification of feature indicators is the most important component of a detection system. Feature extraction allows conversion of an image into a feature vector so that normal and abnormal areas may be effectively differentiated. Features are extracted by analysis of texture, spatial domain, and spectral domain.

Methods to analyze texture describe the features of texture to be extracted from the image. Texture is generally defined as the group of primitive lines that form an image according to certain arrangement rules.⁷ Christoyianni and colleagues^{8,9} used gray-level histograms to detect tumors, but the features of such histograms ignore the spatial correlation in pixels. Haralick's group¹⁰ proposed a co-occurrence matrix (also known as a gray-tone spatial-dependence matrix) to convert images to obtain features. Use of a gray-value co-occurrence matrix permits calculation of a number of features between pixels at different angles and distances in space to describe the changes in texture within an image and improve identification performance and diversification.

The method to extract features in the spatial domain calculates features directly from the original image. Zheng and colleagues¹¹ extracted the variances of pixel intensity and energy from an original image to ascertain the degree of change in its gray-level distribution. Yu and Guan¹² used the spatial domain feature to extract feature vectors related to calcification shape in original breast radiographs.

Compared to extraction of features in the spatial domain, the algorithm used to extract features in the spectral domain converts information to the spectral domain and then extracts the feature vector information. Zheng's team¹¹ used discrete cosine transformation to calculate spectral domain information and then extracted block activity, spectral entropy, and other features from an original image.

In the diagnosis of injury of the supraspinatus tendon, the main purpose of the classifier from the extracted feature vectors is to distinguish a supraspinatus tear from a rupture. Past studies have shown that the support vector machine (SVM), a common classification system, can convert the cor-

responding relationships of independent and dependent variables from low to high dimension to determine the optimal classification hyperplane. The SVM exhibits excellent efficacy in solving various classification problems and is now widely used throughout the medical field.¹³⁻¹⁵

The purpose of this study was to construct an effective computer-aided diagnostic system to evaluate injury of the supraspinatus tendon using a 2-stage approach. The first stage involved image preprocessing through a coupled WHS to provide image enhancement and increase the contrast intensity of the image. The second stage applied feature extraction to the optimized images to determine the optimal feature combination and used an SVM classifier to classify injuries of the supraspinatus muscle.

Patients and Methods

Patients

Our institutional review board approved this study, and all patients gave their written informed consent prior to participation.

Forty-eight patients (24 with partial tears, 24 with complete ruptures of the supraspinatus) underwent MR imaging of the shoulder at National Taiwan University Hospital (Yunlin Branch) from August 1, 2005 to December 31, 2009. Rupture was defined as a full-thickness tear of the supraspinatus tendon and tear, as a partial-thickness tear. Patients were included if MR imaging showed either a rupture or tear, patients with rupture accepted surgery, and patients with a tear accepted rehabilitation and medical therapy. We excluded patients with rupture who did not undergo surgery. The mean age of patients with ruptures was 64.3 ± 9.0 years (range, 49 to 85 years), and 45.8% were male. The mean age of patients with tears was 66.0 ± 11.9 years (range, 40 to 92 years), and 41.7% were male. In all cases, rupture was confirmed at surgery.

MR imaging studies were performed with a 1.5-tesla MR imaging system (Signa Excite, GE Healthcare, Fairfield, CT, USA) using T₂ fast relaxation fast spin echo (FRFSE) sequence, repetition time (TR), 4300 ms, and echo time (TE), 85.5 ms to capture shoulder images. We used an i5-450M processor (Intel Corp., Santa Clara, CA, USA) with 2GB of RAM; software was written in MATLAB R2006a (Mathworks, Inc., Natick, MA, USA).

System framework

Figure 1 illustrates the research framework. We performed image enhancement followed by feature extraction and classification. For MR image en-

hancement, we adopted the WHS formula of Zeng and Liao⁶ to reduce the block effect and improve enhancement. We then performed the second stage, feature extraction and data mining. The physician selected a region of interest (ROI) on a 32×32 block MR image from which to extract the features of texture, spectral domain, and spatial domain. In this study, there were 180 extracted features, including 176 texture features, 2 spatial domain features, and 2 spectral domain features. We employed sequential floating forward selection (SFFS) to select the input feature vectors for classification and support vector machines (SVMs) to classify injuries of the supraspinatus muscle by type. Figure 1 shows the flow chart of the proposed system.

We used a cross-validation technique to verify performance of the classifiers. One round of cross-validation partitioned the data into complementary

subsets (training dataset, test dataset). The training dataset was used to construct the classifier, and the test dataset, to verify the classifier's performance. To reduce variability, multiple rounds of cross-validation were performed using different partitions. For N -fold cross-validation, N rounds of cross-validation are performed. The validation results are averaged over the rounds. We employed a 5-fold cross-validation method in which the data were randomly assigned into 5 groups. For each round of cross-validation, 4 groups were chosen as the training dataset and the fifth group was used as the test dataset. In this study, we employed the area under the curve (AUC) of a receiver operating characteristic (ROC) curve as the index to evaluate classification performance. We detail the performance evaluation in a later section.

Weighted histogram separation, WHS

The WHS used in this study, as described by Pei and associates,⁵ is based on the extended hierarchical histogram separation unit (HSU)^{16,17} (Fig. 2).

The WHS algorithm, which comprises 3 steps, selects the appropriate threshold for the division of an image into 2 parts prior to enhancement. First, a weighted parameter (w) is decided. Each HSU separates the histogram into 2 subhistograms via a threshold (τ). The threshold (τ) for dividing the histogram into 2 subhistograms is then calculated as:

$$\tau = \arg \min_{0 \leq t \leq h} \left| w - \frac{1}{M} \sum_{i=0}^t H(i) \right|, \quad [1]$$

in which h is the dimension of the histogram, $H(i)$ is the value of the i -th segment of histogram H , W is the weighted vector parameters for control of the image separation, and M is the total number of pixels in the entire image.

Based on the threshold value determined by Eq. [1], the image is divided into 2 subhistograms. Eq. [2] defines the histogram separation process. H_0 is that part of the histogram that is less than

Fig. 1. Flow chart showing the experimental design.

Fig. 2. Graphical representation of the weighted histogram separation (WHS) algorithm (ref. 5, Pei *et al.*, 2006).

or equal to τ , and Eq. [2] defines the histogram greater than τ , H_1 :

$$H_0(i) = \begin{cases} H(i), & \text{if } i \leq \tau \\ 0, & \text{otherwise} \end{cases}$$

and

$$H_1(i) = \begin{cases} H(i), & \text{if } i \leq \tau \\ 0, & \text{otherwise} \end{cases}, \quad [2]$$

where $i \in \{0, 1, 2, \dots, h-1\}$.

If the intensity, i , is less than or equal to τ , it is grouped into the subhistogram H_0 . Otherwise, it is grouped into the subhistogram H_1 . H_0 and H_1 can be propagated forward to form the next histogram separation unit. For the k^{th} level of WHS, a total number of $2k-1$ HSUs are required to obtain $2k$ subhistograms. The hierarchical HSUs constitute the WHS. Finally, the weight value according to the attributes is given to each section of the image after separation. To quantify the pixels' gray level in every subhistogram to the same reconstructed value, the images are combined to produce a new enhanced image, as described by:

$$\hat{H}(i) = \left\{ \sum_{j=0}^{s-1} H_{i,j}(j) | i = \lfloor (k+0.5) \cdot s/2^l \rfloor, \right. \\ \left. 0 \leq k \leq 2^l \right\}. \quad [3]$$

Zeng and Liao⁶ developed a post-process to reduce block effects by combining the original and enhanced images. Their method was called coupled WHS because it coupled the original images with the images enhanced by WHS. The output image, L_F , of the coupled WHS is defined as:

$$L_F(i, j) = L_E(i, j) + \alpha \times L(i, j), \quad [4]$$

where L_E is the image enhanced by WHS, L is the original MR image i, j is the pixel's location at the i^{th} row and the j^{th} column, and α is the weighting factor.

However, Zeng and Liao⁶ did adjust the parameter α in accordance with the maximum grayscale value of the original image. In the enhanced images, the highest intensity is always set as the upper limit of the histogram; for this paper, it is one. Because the highest intensity varies in the original MR images, α is used to compensate for such variations. In such a case, α is defined as:

$$\alpha = \frac{1}{\max(L)}, \quad [5]$$

where $\max(L)$ is the highest intensity in the original raw image, L . Figure 3 shows an example of the original raw MR image (Fig. 3a) and the output image enhanced by WHS (Fig. 3b).

Fig. 3. (a) Original image and (b) image after conversion using weighted histogram equalization ($W = 0.5$).

Table 1. Feature extraction

Feature Type		Number of Features
Texture analysis	GTSDM 0° , $d = 1, 3, 5, 7$	44
	GTSDM 45° , $d = 1, 3, 5, 7$	44
	GTSDM 90° , $d = 1, 3, 5, 7$	44
	GTSDM 135° , $d = 1, 3, 5, 7$	44
Spatial domain features	Pixel intensity variance	1
	Energy variance	1
Spectral domain features	Block energy	1
	Spectral entropy	1
Total		180

d : distance parameter to calculate the co-occurrence matrix
 GTSDM: gray-tone spatial dependency matrix

Feature extraction

The authors performed feature extraction and data mining from ROIs. Feature extraction was performed to convert the 2-dimensional image into a feature (or feature vector). Extracted features included 176 texture analysis, 2 spatial domain, and 2 spectral domain features (Table 1).

Texture analysis

Haralick's group¹⁰ proposed a co-occurrence matrix to convert images to obtain texture features. They aimed to define the relationship between different gray-level values between pixels and calculate the number of pixels in 2 relative positions (at different distances and angles) to obtain a co-occurrence matrix. We extracted 11 features (Appendix) from this matrix, converted the original imaging data into a co-occurrence matrix, and then used the correlation of matrix-represented spatial gray-level intensities to describe the features. Different detection angles for the relative position of each

Fig. 4. Co-occurrence matrix description (ref. 7, Haralick, 1979).

pixel created different corresponding co-occurrence matrices, so each image had 4 co-occurrence matrices of 0° , 45° , 90° , and 135° at the same time. As Fig. 4 shows, for example, in image matrix I , which contains 4 different gray-level values (with $\theta = 0^\circ$ and $d = 1$), i and j represented the gray-level values of each of 2 pixels that satisfied the positional relationship, $h(i, j)$, for the total number that meets this positional relationship, and d represented the distance between the 2 co-occurrence pixels in the original image.

The symbol θ represented the relative angle between 2 pixels, and $\theta = 0^\circ$, 45° , 90° , and 135° and $d = 1, 3, 5$, and 7 . The 4 values of the angle parameter, θ (0° , 45° , 90° , 135°), include all possible angular directions in the digital images. For the distance parameter (d), the direct neighbor ($d = 1$) of the reference pixel was conventionally used. To span the range of the solution space without sacrificing computational efficiency, we employed the 4 values ($d = 1, 3, 5$, and 7). Experimental results show the contribution of all attributes of both parameters (θ and d) to the input vector of the pattern classification. It followed that 176 features could be extracted from each input image (11 types of feature $\times$ 4 different angle combinations $\times$ 4 different distance combinations = 176 features).

Spatial domain features

We derived the features of spatial domain, which characterize the spatial property of the input images, directly from the raw images. Zheng and associates¹¹ proposed extracting 2 features of the spatial domain from the original image. Their method involves first calculating the average pixel intensity and average energy of the pixel intensity for each block image, $x(m, n)$:

$$AVG = \frac{1}{M \times N} \sum_{m=1}^M \sum_{n=1}^N x(m, n) \quad [6]$$

and

$$EAVG = \frac{1}{M \times N} \sum_{m=1}^M \sum_{n=1}^N x^2(m, n), \quad [7]$$

where $x(m, n)$ is the gray level of the pixel located in the m^{th} row and the n^{th} column of the window. The variances of pixel intensity and energy are then calculated as:

$$VAR = \sum_{m=1}^M \sum_{n=1}^N [x(m, n) - AVG]^2 \quad [8]$$

and

$$EVAR = \sum_{m=1}^M \sum_{n=1}^N [x(m, n)^2 - EAVG]^2. \quad [9]$$

Spectral domain features

In analyzing images, features of the spectral domain describe the characteristics of frequency of the input image rather than those of space. The discrete cosine transform (DCT) is a common conversion technique in domain analysis that is used in the texture spectrum method for spectral domain features. Zheng's group¹¹ proposed converting spatial domain information into spectral domain information before extracting features. The spectral domain information $X(i, j)$ was extracted from the original image, $x(m, n)$, by the DCT, and features of block activity and spectral entropy were then extracted.

The features of block activity and spectral entropy were calculated as:

$$A = \sum_{i=1}^M \sum_{j=1}^N |X(i, j)| \quad [10]$$

and

$$E = - \sum_{i=1}^M \sum_{j=1}^N \bar{X}(i, j) \ln[\bar{X}(i, j)], \quad [11]$$

where $X(i, j)$ is the DCT coefficient in the i^{th} row and the j^{th} column of the DCT matrix,

$$\bar{X}(i, j) = \frac{|X(i, j)|}{A}.$$

Sequential floating forward selection

Pudil and associates¹⁸ proposed the sequential floating forward selection (SFFS) algorithm, a search mechanism that combines sequential forward selection (SFS) and sequential backward selection (SBS) algorithms. Table 2 shows the procedures of the algorithm.

In Table 2, Y is the set consisting of the extracted features; D , the number of extracted features; and X , the feature set selected from Y . $J(X)$ was the feature selection criterion function of the feature set X .

Table 2. Sequential floating forward selection (SFFS) algorithm

Input: $Y = \{y_i | i = 1, \dots, D\}$ where D is the number of extracted features

Output: $X_k = \{x_i | i = 1, \dots, k, x_i \in Y\}$, $k = 0, 1, \dots, D$

Feature selection criterion function of the feature set X : $J(X)$

Initial value: $X_0 = \emptyset$; $k = 0$

Termination condition: when k equals the predefined number of features

Step 1

$$x^+ := \arg \max_{x \in Y - X_k} J(X_k + x)$$

$$X_{k+1} = X_k + x^+; k = k + 1$$

Step 2

$$x^- := \arg \max_{x \in X_k} J(X_k - x)$$

If $J(X_k + \{x^-\}) > J(X_{k-1})$ then

$$X_{k-1} := X_k - x^-;$$

$$k := k - 1$$

Go to *Step 2*

Else

Go to *Step 1*

In brief, completion of each sequential forward selection by SFFS was immediately followed by sequential backward selection. During Step 1, the SFS identified the most significant feature among unselected features, x^+ , which was derived using the equation $x^+ := \arg \max_{x \in Y - X_k} J(X_k + x)$ in Table 2.

During Step 2, the SBS identified the least significant feature in the current feature set, x , which was derived using the equation $x^- := \arg \max_{x \in X_k} J(X_k - x)$

in Table 2. If the new $k-1$ dimension found by SBS was better than the last $k-1$ dimension, SBS was continued; if it was worse, SBS was discontinued. Instead, the next SFS with the k -dimensional combination after SFS was used. The SFFS improved the “nesting” effect caused by the SFS and SBS—features either permanently stay in the selected feature set, or are permanently removed.

In this study, the purpose of feature selection was to choose the subset of features that minimized the classification error. The classification error was measured by the mean square error (MSE) of the target output and the classifier’s output. The square error (SE) was the sum of the square of the difference between the target output and the classifier’s output. In such case, maximizing system performance ($\max J(X)$ in Steps 1 and 2) meant minimizing MSE. The first step selected the feature that mini-

mized the MSE among the unselected features and the second step, from among the current feature set.

In the original design, selection iteration stopped when the number of features in the set, X , equaled a predefined number of features, K . However, an arbitrarily chosen K did not guarantee global minimization. In this paper, we applied an exhaustive search mechanism to achieve global minimization. SFFS was sequentially conducted under each possible selection of K , where $K = 1, 2, \dots, D$. Each predefined number of features (K) generated its own optimal set, X , associated with classification error. Then, the set X with the globally best performance was chosen as the final output of the SFFS.

Classification

The purpose of the classifier was to match the feature vector extracted from input to the actual lesion phenomenon; in this study, the lesion that corresponded with output was a supraspinatus tear or rupture. The advantage of the SVM is that it converts the actual problem by nonlinear transformation to high dimensional feature space. By constructing the linear discriminant function in high dimensional space (to achieve the nonlinear discriminant function of the original space), the SVM can accurately classify complex multidimensional information and reduce problems of over-learning. At present, the SVM is widely used in the field of machine learning for classification in data mining.^{19,20} This algorithm primarily uses a separating hyperplane to differentiate 2 or more different classes of data. The SVM algorithm first defines boundaries of the separating hyperplane. $d_+(d_-)$ is the shortest distance between the categories labeled $+1$ and the hyperplane (the shortest distance of categories labeled -1 and the hyperplane). This type of data must meet the following conditions: all the x of the hyperplane must be met, $w \cdot x + b = 0$, where w is the normal vector of the hyperplane and b is the offset. So, $f(x) = w \cdot x + b$ can be called the decision function. When inputting test data, it can be classified according to the value of the decision function. If $f(x) > 0$, the data are classified as $+1$; if $f(x) < 0$, the data are classified as -1 . It is hoped that the SVM can separate the hyperplane of the maximum margin for different types of data. As Fig. 5 shows, the hyperplanes H_0 , H_1 , and H_2 can achieve the effect of classification where H_0 is optimal because H_0 has the maximum distance with boundary A and boundary B.

In linear indivisible problems, a kernel is usually used to conduct conversion, project data to a high dimensional space, and then solve the problem us-

Fig. 5. The optimal separating hyperplane.

ing linear classification. The radial basis function (RBF) kernel used in this study creates a classification system. Because the RBF can classify nonlinear and high dimensional data, it is the preferred option in selecting the kernel, k , where:

$$k(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2), \quad [12]$$

and γ is a constant. The grid search algorithm was applied to optimize the parameters of the SVM classifiers.

Performance evaluation

Sensitivity and specificity are the basic measures of accuracy of a diagnostic test. Sensitivity is the ability to detect disease correctly, and specificity is the ability to avoid classifying normal tissue as diseased. However, they depend on the cut point of the classifier used to define “positive” and “negative” test results. As the cut point shifts, sensitivity and specificity will shift. The receiver operating characteristic (ROC) curve is a plot of the sensitivity of a test versus its false-positive rate for all possible cut points. First developed in signal-detection theory analysis, the ROC curve is commonly used to evaluate classification performance. Obuchowski²¹ detailed the advantages of the ROC curve as a means of defining the accuracy of a test. In this study, we employed a ROC curve to obtain the area under the curve (AUC), which we used to assess the overall classification performance, and to compare the discriminative ability of the 2 classification systems. In calculating the ROC curve, the x-axis is the false positive fraction (FPF), also known as 1-specificity, and the y-axis is the true positive fraction (TPF), also known as sensitivity. For a given threshold, a ROC curve can be plotted, and the area under the ROC curve (azimuth [Az] value) can be used as the

index to evaluate system identification performance (Excellent: 0.9 ~ 1.0; Good: 0.80 ~ 0.9; Fair: 0.70 ~ 0.80; Poor: 0.60 ~ 0.70; Failed: 0.50 ~ 0.60).

The formulas to measure the x-axis (1-specificity) and y-axis (sensitivity) are:

$$1\text{-Specificity} = FPF = \frac{FP}{TN + FP} \quad [13]$$

and

$$\text{Sensitivity} = TPF = \frac{TP}{FN + TP}, \quad [14]$$

where: true positive (TP) indicates that the measured area is the fracture occurrence zone, and the system detects it as the fracture occurrence zone; false negative (FN) indicates that the measured area is the fracture occurrence area, but the system misjudges it as a tear occurrence zone; false positive (FP) indicates that the measured area is the tear occurrence zone, but the system misjudges it as the fracture occurrence zone; and true negative (TN) indicates that the measured area is the tear occurrence zone, and the system detects it as the tear occurrence zone.

The area under the ROC curve (AUC, also called Az value), which indicates the average sensitivity over all possible specificities, is a widely used index to characterize machine learning performance. Az values, ranging from 0 to one, characterize the tradeoff between the sensitivity and specificity of the diagnostic test. ROC curves are calculated from the prediction scores of the SVM, which are acquired by the relaxation of the binary output of the SVM generated by thresholding of the hyperplane.¹³

Results

In the SVM classification, performance comparisons were divided into 2 parts, the first using all 180 features and the second using features selected by SFFS. All features and SFFS-selected features were taken as the input features of the SVM classifier used to compare differences in the AUC by the ROC curve (Fig. 6).

Once SFFS was employed, the Az values in the test dataset were increased from 68.75 to 84.38%. The results showed that the Az values of all features were lower than those of the 76 SFFS-captured features for both the training and test datasets. This indicated that selecting features to classify supraspinatus tendon injuries was better using SFFS than using all 180 features.

In addition, SFFS feature selection drastically reduced the dimension of the input feature vector.

Fig. 6. Receiver operating characteristic (ROC) curves for (a) all features in the training set; (b) all features in the test set; (c) features selected by sequential floating forward selection (SFFS) in the training set, and (d) features selected by SFFS in the test set.

Table 3. Features selected by sequential floating forward selection (SFFS)

Type	Amount	Feature number by discriminatory power
Textural information	75	92,99,94,71,95,97,98,90,93,100,96,103,101,123,126,128,142,102,118,150,155,104,106,115,161,134,138,148,175,119,131,151,55,40,42,62,72,59,60,74,83,81,70,75,85,86,87,88,133,172,158,107,141,169,129,111,105,108,114,130,125,137,112,140,113,143,136,139,144,124,147,132,149,164,27
Spatial domain	0	NA
Spectral domain	1	180

NA, not applicable

This means that SFFS can successfully filter out the features that are noisy in terms of pattern classification. The total number of features after feature extraction through SFFS was reduced to 57.78% [(180–76)/180], which implies a reduction in both computing time and storage space. SFFS selected 76 of 180 features, including texture features and spectral domain features (Table 3).

In the extraction of 75 texture features, $\theta = 90^\circ$ was the highest proportion selected (Table 4). This was the case in 22 of the top 30 with respect to classification performance. This result showed that supraspinatus muscle injury points may occur in the particular texture analysis of $\theta = 90^\circ$. In addition, a feature was extracted in the spectral domain; this feature was the spectral entropy. However, no

feature was captured in the spatial domain, so supraspinatus tendon injuries cannot be effectively represented using spatial domain features.

To verify the effect of image enhancement, the experiment was replicated using the raw images instead of the WHS-enhanced images as input images. Table 5 shows the comparison of classification performance between the raw and enhanced images.

Based on the results in Table 5, we have drawn the following conclusions.

Under either feature selection criterion (no selection/SFFS selection), the enhanced images outperformed the raw images. For all 180 features used as input vectors, the test dataset yielded Az values of 68.75% for enhanced images and 56.94% for raw

Table 4. Feature distribution of support vector machine (SVM) classifiers by sequential floating forward selection (SFFS) feature selection

Angle and distance of co-occurrence matrix	Feature Index										
	Angular second moment	Contrast	Correlation	Sum of square variance	Inverse different moment	Sum average	Sum variance	Sum entropy	Entropy	Difference variance	Difference entropy
$\theta = 0^\circ, d = 1$	1	2	3	4	5	6	7	8	9	10	11
$\theta = 0^\circ, d = 3$	12	13	14	15	16	17	18	19	20	21	22
$\theta = 0^\circ, d = 5$	23	24	25	26	27	28	29	30	31	32	33
$\theta = 0^\circ, d = 7$	34	35	36	37	38	39	40	41	42	43	44
$\theta = 45^\circ, d = 1$	45	46	47	48	49	50	51	52	53	54	55
$\theta = 45^\circ, d = 3$	56	57	58	59	60	61	62	63	64	65	66
$\theta = 45^\circ, d = 5$	67	68	69	70	71	72	73	74	75	76	77
$\theta = 45^\circ, d = 7$	78	79	80	81	82	83	84	85	86	87	88
$\theta = 90^\circ, d = 1$	89	90	91	92	93	94	95	96	97	98	99
$\theta = 90^\circ, d = 3$	100	101	102	103	104	105	106	107	108	109	110
$\theta = 90^\circ, d = 5$	111	112	113	114	115	116	117	118	119	120	121
$\theta = 90^\circ, d = 7$	122	123	124	125	126	127	128	129	130	131	132
$\theta = 135^\circ, d = 1$	133	134	135	136	137	138	139	140	141	142	143
$\theta = 135^\circ, d = 3$	144	145	146	147	148	149	150	151	152	153	154
$\theta = 135^\circ, d = 5$	155	156	157	158	159	160	161	162	163	164	165
$\theta = 135^\circ, d = 7$	166	167	168	169	170	171	172	173	174	175	176
Spatial + spectral	177	178	179	180							

θ : direction parameter to calculate the co-occurrence matrix

d : distance parameter to calculate the co-occurrence matrix

Table 5. Classification performance with and without image enhancement

Input images	Raw images		Enhanced images	
	No. of features	Az value of test set	No. of features	Az value of test set
All features	180	56.94%	180	68.75%
Selected by SFFS	84	65.28%	76	84.36%

Az value: area under the receiver operating characteristic curve

SFFS: sequential floating forward selection

images. With input features selected by SFFS, the corresponding Az values were 84.38% for the enhanced images and 65.28% for the raw images.

Whether or not the input images underwent enhancement, SFFS feature selection improved classification performance. For the raw input images,

the Az values increased from 56.94 to 65.28%. For the enhanced input images, the Az values increased from 68.75 to 84.38%.

During the SFFS feature selection, image enhancement reduced the dimension of the input vector. The dimension of the input vector was 84 for the raw input images and 76 for the enhanced images.

Combining SFFS feature selection and WHS image enhancement, the Az values increased from 56.94 to 84.38%, and the dimension of the input vectors decreased from 180 to 76.

Discussion

This 2-phase computer-aided RCT model based on MR imaging produced a detection performance ROC Az value of 84.38% using 76 features compared to 56.94% using all 180 features and without image enhancement. Applying SFFS feature selection to the enhanced images reduced the dimension

of the input vector by 57.78%. In addition to reducing computing time and data storage capacity, it also improved diagnostic performance.

At present, the evaluation of supraspinatus tear using MR imaging involves either simple image enhancement performed manually or the input of data into a classifier for identification after feature extraction.^{15,22} However, in this study, we combined 2 detection modes. In the first phase, MR images were enhanced to strengthen information regarding suspicious features, and the data were then entered into the classifier to perform the identification. The enhancement in the first phase significantly reduced noise, thereby reducing misjudgment and the probability of false positives. In this study, coupling the image enhancement mode, which differed from such traditional strengthening methods as histogram equalization, with WHS effectively reduced the block effect.

The classifier required to strengthen the classification can significantly reduce system complexity and training to improve the efficiency of system calculations. With respect to the diagnosis of supraspinatus injury, the main purpose of the classifier from the extracted feature vectors was to distinguish a tear from a rupture. The SVM exhibits excellent efficacy in addressing various types of classification problems and is now widely used in the medical field.^{13–15} Fu and associates¹³ applied the SVM to the detection of breast microcalcifications. In the first stage, to detect the texture features of a calcification point and conduct feature selection by SFS, the SVM performed better than the generalized regression neural network [GRNN] in classifying benign versus malignant calcifications. Han's team¹⁴ proposed an ultrasonic image detection system to effectively improve the rate of cancer diagnosis by integrating texture and clinical features with SVM classifiers. In contrast, Horng¹⁵ used ultrasonic images of the supraspinatus muscle and 4 different methods of texture analysis to capture the different features; 5 different multi-class support vector machines (MCSVMs) were applied to detect supraspinatus muscle injuries and were reported to achieve 90% of the classification results.

In detection systems, the developmental trend of feature vectors is to combine the feature extraction modes of texture, spatial, and spectral domains. The advantage of this combination lies in its ability to reduce the possibility of missing significantly influential features,¹⁵ and applications to MR imaging have been developed in many different medical fields.^{23–28} Research on extraction of texture features in medical images focuses on finding the best features and best texture parameters according to

the image features. Most current research in feature selection focuses on reducing data dimensions to reduce computational complexity. Common methods applied in the past included SFS and SBS, but both of these were associated with problems of nesting effect. Pudil and colleagues¹⁸ proposed the SFFS algorithm, which combines the search features of SFS and SBS, thus significantly reducing the possibility of a nesting effect.

In addition to the detection mode, a major factor that affects the detection performance of an imaging system is whether the elements of the input vector of the classifier are filtered. Our experimental results showed that combining the SFFS algorithm with WHS image enhancement resulted in a 57.78% reduction in vector dimension and increased the Az value by 27.44% (84.38 to 56.94%). The SFFS algorithm can also provide a significant level of classification from highest to lowest, the most suitable choice of feature vectors for follow-up study.

We developed a 2-stage detection mode and verified it as a viable model. We also utilized the SFFS algorithm that can self-adapt for different types of input data to select the most suitable feature combination for the input database. To maintain optimum system performance, it is suggested that the SFFS algorithm be used to select the most suitable feature vectors in the system design stage (before retraining the classifier parameters).

Conclusions

Currently, diagnosis of most supraspinatus injuries relies upon MR imaging. However, visual interpretation of images by radiologists can result in misdiagnosis. Digital computer image processing can improve diagnosis, but direct capture of the original MR images without enhanced processing leads to poor image contrast. This 2-phase computer-aided model of the RCT used WHS technology to enhance MR images in the first phase and extracted the features in the area of interest to establish an SVM classification model in the second phase. The model's detection performance improved after selection of features with SFSS compared to that without SFFS selection. Moreover, SFFS significantly reduced vector dimension, computing time, and data storage capacity. Therefore, the proposed model can improve the diagnosis and treatment of supraspinatus injuries.

Acknowledgements

The authors would like to extend their gratitude

to the National Science Council for partially funding this research (NSC 101-2221-E-224-029 & NSC 102-2221-E-224-068).

Appendix

1. Angular Second Moment:

$$F_1 = \sum_{i=0}^{L-1} \sum_{j=0}^{L-1} [P(i, j)]^2 \quad (1)$$

2. Contrast:

$$F_2 = \sum_{n=0}^{L-1} n^2 \left[\sum_{i=1}^L \sum_{j=1}^L P(i, j) \right] \quad (2)$$

3. Correlation:

$$F_3 = \frac{\sum_{i=0}^{L-1} \sum_{j=0}^{L-1} i j P(i, j) - \mu_x \mu_y}{\sigma_x \sigma_y} \quad (3)$$

4. Sum of Squares Variance:

$$F_4 = \sum_{i=0}^{L-1} \sum_{j=0}^{L-1} (i - \mu)^2 P(i, j) \quad (4)$$

5. Inverse Difference Moment (IDM):

$$F_5 = \sum_{i=0}^{L-1} \sum_{j=0}^{L-1} \frac{1}{1 + (i - j)^2} P(i, j) \quad (5)$$

6. Sum Average:

$$F_6 = \sum_{i=0}^{2L-2} i P_{x+y}(i) \quad (6)$$

7. Sum Variance:

$$F_7 = \sum_{i=0}^{2L-2} (i - F_6)^2 P_{x+y}(i) \quad (7)$$

8. Sum Entropy:

$$F_8 = - \sum_{i=0}^{2L-2} P_{x+y}(i) \log[P_{x+y}(i)] \quad (8)$$

9. Entropy:

$$F_9 = - \sum_{i=0}^{L-1} \sum_{j=0}^{L-1} P(i, j) \log(P(i, j)) \quad (9)$$

10. Difference Variance:

$$F_{10} = \text{Variance of } P_{x-y} \quad (10)$$

11. Difference entropy:

$$F_{11} = - \sum_{i=0}^{L-1} P_{x-y}(i) \log(P_{x-y}(i)) \quad (11)$$

References

- Díaz F, Fdez-Riverola F, Glez-Peña D, Corchado JM. Applying GCS networks to fuzzy discretized microarray data for tumour diagnosis. *Lect Notes Comput Sci* 2006; 4224:1095–1102.
- Hung SY, Chen CY. Mammographic case base applied for supporting image diagnosis of breast lesion. *Expert Syst Appl* 2006; 30:93–108.
- Latifoglu F, Şahan S, Kara S, Güneş S. Diagnosis of atherosclerosis from carotid artery Doppler signals as a real-world medical application of artificial immune systems. *Expert Syst Appl* 2007; 33:786–793.
- Moon N, Bullitt E, van Leemput K, Gerig G. Automatic brain and tumor segmentation. *Lect Notes Comput Sci* 2002; 2488:372–379.
- Pei SC, Zeng YC, Ding JJ. Color images enhancement using weighted histogram separation. *IEEE ICIP* 2006; 2889–2891.
- Zeng YC, Liao HY. Video enhancement based on saturation adjustment and contrast enhancement. *IEEE ISCAS* 2008; 3550–3553.
- Haralick RM. Statistical and structural approaches to texture. *Proc IEEE* 1979; 67:786–804.
- Christoyianni I, Dermatas E, Kokkinakis G. Fast detection of masses in computer-aided mammography. *Signal Processing Magazine IEEE* 2000; 17:54–64.
- Christoyianni I, Dermatas E, Kokkinakis G. Neural classification of abnormal tissue in digital mammography using statistical features of the texture. *IEEE Int Conf Electron Circuits Syst* 1999; 1:117–120.
- Haralick RM, Shanmugam K, Dinstein I. Textural features for image classification. *IEEE Trans Syst Man Cybern* 1973; 3:610–621.
- Zheng B, Qian W, Clarke LP. Digital mammography mixed feature neural network with spectral entropy decision for detection of microcalcifications. *IEEE Trans Med Imag* 1996; 15:589–597.
- Yu S, Guan L. A CAD system for the automatic detection of clustered microcalcifications in digitized mammogram films. *IEEE Trans Med Imaging* 2000; 19:115–126.
- Fu JC, Lee SK, Wong ST, Yeh JY, Wang AH, Wu HK. Image segmentation feature selection and pattern classification for mammographic microcalcifications. *Comput Med Imag Graph* 2005; 29:419–429.
- Han SM, Lee HJ, Choi JY. Computer-aided prostate cancer detection using texture features and clinical features in ultrasound image. *J Digit Imaging* 2008; 21 Suppl 1:S121–S133.
- Horng MH. Multi-class support vector machine for classification of the ultrasonic images of supraspinatus. *Expert Syst Appl* 2009; 36:8124–8133.
- Nilsson M, Dahl M, Claesson I. The successive mean quantization transform. *IEEE ICASSP* 2005; 4:429–432.
- Nilsson M, Dahl M, Claesson I. Gray-scale image enhancement using the SMQT. *IEEE ICIP* 2005; 1: 933–936.
- Pudil P, Novovičová J, Kittler J. Floating search methods in feature selection. *Pattern Recogn Lett* 1994; 15:1119–1125.
- Vapnik V. Statistical learning theory. Oxford: John

- Wiley, 1998.
20. Muller KR, Mika S, Ratsch G, Tsuda K, Scholkopf B. An introduction to kernel-based learning algorithms. *IEEE Trans Neural Networks* 2001; 12: 181–201.
 21. Obuchowski NA. Receiver operating characteristic curves and their use in radiology. *Radiology* 2003; 229:3–8.
 22. Horng MH. Performance evaluation of multiple classification of the ultrasonic supraspinatus images by using ML, RBFNN and SVM classifiers. *Expert Syst Appl* 2010; 37:4146–4155.
 23. Pérez de Alejo R, Ruiz-Cabello J, Cortijo M, et al. Computer-assisted enhanced volumetric segmentation magnetic resonance imaging data using a mixture of artificial neural networks. *Magn Reson Imaging* 2003; 21:901–912.
 24. Saeed N, Puri BK. Cerebellum segmentation employing texture properties and knowledge based image processing: applied to normal adult controls and patients. *Magn Reson Imaging* 2002; 20:425–429.
 25. Kovalev V, Kruggel F. Texture anisotropy of the brain's white matter as revealed by anatomical MRI. *IEEE Trans Med Imaging* 2007; 26:678–685.
 26. Ghazel M, Traboulsee A, Ward RK. Semi-automated segmentation of multiple sclerosis lesions in brain MRI using texture analysis. 2006 IEEE International Symposium on Signal Processing and Information Technology, Vancouver, Canada, 2006; 6–10.
 27. Zhang J, Tong L, Wang L, Wang L, Li N. Texture analysis of multiple sclerosis: a comparative study. *Magn Reson Imaging* 2008; 26:1160–1166.
 28. Jiráček D, Dezortová M, Taimr P, Hájek M. Texture analysis of human liver. *J Magn Reson Imaging* 2002; 15:68–74.
-