Learning Fourier Descriptors for Computer-Aided Diagnosis of the Supraspinatus[1]

Oliver van Kaick, MSc, Ghassan Hamarneh, PhD, Aaron D. Ward, PhD, Mark Schweitzer, MD, Hao Zhang, PhD

Abstract

Rationale and Objectives. Supraspinatus muscle disorders are frequent and debilitating, resulting in pain and a limited range of shoulder motion. The gold standard for diagnosis involves an invasive surgical procedure. As part of a proposed clinical workflow for noninvasive computer-aided diagnosis (CAD) of the condition of the supraspinatus, we present a method to classify 3D shapes of the muscle into relevant pathology groups, based on magnetic resonance (MR) images.

Materials and Methods. We obtained MR images of the shoulder from 72 patients, separated into five pathology groups. The imaging protocol ensures that the supraspinatus is consistently oriented relative to the MR imaging

plane for each scan. Next, we compute the Fourier coefficients of 2D contours lying on parallel imaging planes and integrate the corresponding frequency components across all contours. To classify the shapes, we learn the Fourier coefficients that best distinguish the different classes.

Results. We show that our method leads to significant improvement when compared to previous work. We are able to distinguish between normal shapes and shapes that possess a pathology with an accuracy of almost 100%. Moreover, we can differentiate between the different pathology groups with an average accuracy of 86%.

Conclusion. We confirm that analyzing the 3D shape of the muscle has potential as a form of diagnosis reinforcement to assess the condition of the supraspinatus. Moreover, our proposed descriptor based on Fourier coefficients is able to distinguish the different pathology groups with accuracies higher than those obtained by previous work, indicating its potential application to support a system for CAD of the supraspinatus.

Key words: shoulder pathology, supraspinatus, magnetic resonance image, computer aided diagnosis, shape analysis

---

1. Introduction

The supraspinatus muscle originates from the supraspinatus fossa of the scapula and runs along the top of the shoulder blade. This muscle is part of the rotator cuff, which is a group of muscles and tendons responsible for shoulder movement and stabilization. Disorders affecting the rotator cuff can cause pain and reduce patient mobility [1], and their occurrence is frequent, with a reported rate of 30% of individuals past 60 years of age in a cadaveric

study [2]. Supraspinatus disorders involving tendon tearing can be accompanied by muscle retraction, atrophy, or both. The standard procedure for the diagnosis of rotator cuff disorders is shoulder arthroscopy, which is a surgery involving the insertion of an optical camera. However, diagnosis based on magnetic resonance (MR) images is a preferred noninvasive alternative. Additionally, the impact of a supraspinatus tendon tear on the overall body of the muscle has prognostic value and is visible on MR images, but is not visible during arthroscopy [3, 4].

The long-term goal of our research is to develop a tool for noninvasive computer-aided diagnosis (CAD) of the supraspinatus, based on 3D shapes extracted from MR images. The realization of this goal would lead to several important benefits and results. Firstly, CAD can be helpful to provide a second opinion or to serve as a form of diagnosis reinforcement for the physician, and it can also be valuable when other clinical data (e.g., palpation or range-of-motion exams, in the context of musculoskeletal disorders) do not provide a clear indication of the pathology. Secondly, understanding the relationship between shape and pathology can be helpful in providing evidence for etiological or epidemiological studies. Lastly, the impact of a torn tendon on the shape of the supraspinatus, which has prognostic significance, can be assessed more accurately by analyzing the 3D shape of the muscle. We are therefore motivated to perform an automated analysis of the 3D shape of the supraspinatus, which would allow us to improve upon the diagnosis based on 2D MR images or even shoulder arthroscopy. Previous work concluded that the shape of the supraspinatus is helpful in the diagnosis of rotator cuff disorders, but a more effective shape analysis procedure is necessary to achieve

a pathology classification with 100% accuracy [5]. From a computational viewpoint, our goal is: given a set of shapes and their diagnoses (which were determined by a physician), learn the relationship between the shape of the supraspinatus and its pathologies, and employ such a relationship in CAD.

There have been several studies focusing on learning the relationship between the shapes of anatomical structures and their pathological conditions, such as in the case of cerebral structures [6, 7, 8, 9, 10, 11], but there are relatively fewer works relating shape properties to musculoskeletal pathologies [5, 12]. These works make use of global shape properties (e.g., area and volume) [7, 5], or representations such as the Fourier transform [11, 12], level sets [13], spherical harmonics [8, 9], and spherical wavelets [10]. However, measures such as volume have limitations in capturing changes that appear in higher frequency detail or which are of a more localized nature. On the other hand, spherical harmonics or spherical wavelets, which allow for localization, require the data to have some form of correspondence (a bijection matching anatomically-related points between the two shapes). The computation of an accurate and repeatable correspondence is in general a difficult problem and particularly in our case, where the given muscle shapes lack reliably identifiable anatomic landmarks [14]. Moreover, methods based on spherical representations also require a parameterization onto a sphere, which can introduce considerable distortion in the data [15].

In this paper, we propose a shape representation that does not require a correspondence (in the spirit of works for 2D such as [16, 17]), but that is able to detect variations in the shape at all scales, from coarse to fine. We take advantage of the fact that the clinical imaging protocol encourages a consis-

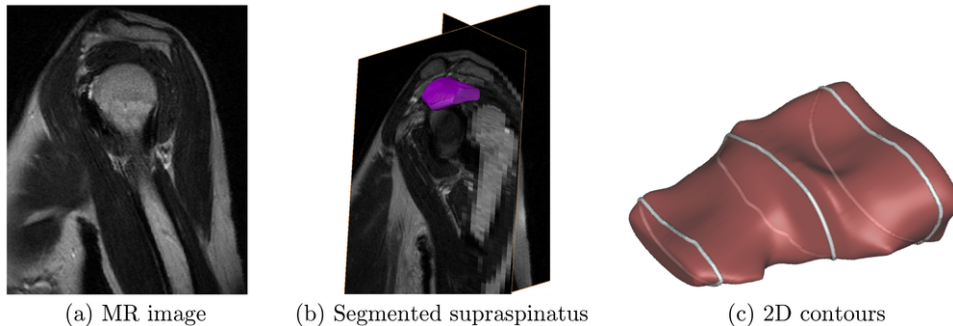|		
(a) MR image	(b) Segmented supraspinatus	(c) 2D contours

Figure 1: Input to the representation used in this work: (a) MR images of the shoulder of a patient are acquired (only one image is shown), (b) The 3D shape of the supraspinatus muscle is segmented from the MR images, (c) The final shape of the muscle is captured as a set of 2D contours (only 3 contours are shown for illustration purposes).

tent alignment of the shapes, where the orientations of the sagittal imaging planes are consistent with respect to the orientation of the supraspinatus across all subjects. This orientational consistency permits the use of our proposed correspondence-free shape representation. Our representation is generated as follows (Figure 1 shows the input to our representation). First, patients are imaged with a protocol that ensures the consistent orientation of the supraspinatus. Next, the 2D MR images are segmented by a radiology resident, yielding a set of parallel 2D contours that define the 3D shape of the muscle. Subsequently, for the analysis of the shapes, we apply the Fourier transform to each contour and obtain a set of coefficients. This is motivated by the fact that the Fourier transform has well-known properties of decomposing a signal into all of the frequency components that provide a complete, coarse-to-fine description of the shape. Finally, we integrate each frequency coefficient for all the contours, which yields a descriptor that is coherent across shapes because the contours lie on planes oriented along a consis-

tent direction. The integration of coefficients also ensures coherence when muscles are described by different numbers of contours. Although this approach slightly reduces the discriminative power of the descriptors, it avoids the requirement of having the contours correspond to each other across subjects. Since we can manipulate the components obtained with the Fourier transform to obtain descriptors that are invariant to rotations, translations, and scale, we see this transform as the best choice for our representation. Utilizing other decompositions (e.g., the level set method) would require a consistent alignment of the shapes along all dimensions and a one-to-one correspondence between contours of different shapes. Learning the geometric variations between the shapes is then posed as the problem of finding the descriptor components and the decision rule that best distinguish the different classes. Our contribution in this aspect is that, since it is not possible to know beforehand which is the best descriptor for the classification, we propose to learn it by selecting components from our representation, thus allowing us to make use of all of the geometric information that helps in distinguishing the pathologies.

2. Materials and methods

Our dataset consists of MR images of the shoulders of 72 patients (a sample is shown in Figure 1(a)). This dataset was initially acquired and utilized in the work of Ward et al. [5], and the images were taken as part of routine clinical studies. The MR images were acquired at 1.5T, with the following imaging parameters: repetition time 4,000-5,500, echo time 36, field of view 14, matrix 256×256. The in-plane (sagittal) resolution of the data

ranges from 0.3 to 0.6 mm and the slice thickness ranges from 3 to 5 mm. As part of the routine clinical imaging protocol, the patients were consistently imaged in the supine position, relaxed, in minimal external rotation, and with consistent elevation of the arm, to minimize the effects of pose and gravity on the shape of the muscle. Great care was taken in the consistent positioning of the shoulder of each patient, and subjects were instructed not to move after the arm had been positioned. This imaging protocol ensures the consistent orientation of the supraspinatus fossa (part of the shoulder blade) and the humeral head (top of the upper arm bone) across all subjects. Since these are the anchoring points of the supraspinatus muscle, the imaging protocol also ensures the consistent orientation of this muscle across all subjects. This study was conducted under a waiver from the Institutional Review Board.

Each muscle was segmented from the images by a radiology resident in a slice-by-slice manner and the set of contours representing each shape was constructed (Figure 1(b) and (c)). Manual segmentation of contours was performed (as opposed to an automatic method) because of the low contrast that the supraspinatus muscle has with surrounding tissues on many slices, which requires the use of expert knowledge for the segmentation. We follow the same segmentation approach as in [5]. In the software tool used for the segmentation, the resident selects points lying on the boundary of the supraspinatus on each slice. Next, the tool fits a parametric cubic spline to these points, generating a smooth curve that segments the muscle. Adjustments can be made to the curve interactively (by adding, moving or removing points). A previous study has shown that the intra- and interobserver variability in supraspinatus contouring is less than 5% [18], while another study

Table 1: Pathology groups considered in the study.

Pathology	Abbreviation	Number of shapes
No pathology (normal)	N	14
Tear	T	19
Tear and atrophy	TA	13
Tear and retraction	TR	15
Tear and atrophy and retraction	TAR	11

concluded that the interobserver variability in determining the volume of the supraspinatus by 3D contouring in MR images is 3.63% [19]. It should be noted that the resident responsible for the segmentation was not aware of the diagnoses during this procedure, since the diagnoses were made only after the segmentation was complete.

Finally, the supraspinatus of each patient was diagnosed in conjunction by a radiologist and a radiology resident, based on the MR images. The correctness of the diagnosis was further confirmed by shoulder arthroscopy. According to the study in [20], the inter- and intra-observer variability in discriminating normal from torn rotator cuff muscles is low, with accuracies ranging from 89% to 98%. The supraspinatus of each patient was assigned to one of the following classes: normal (N), tendon tear (T), tendon tear and muscle atrophy (TA), tendon tear and tendon retraction (TR), and tendon tear with muscle atrophy and tendon retraction (TAR). The number of shapes in each class is presented in Table 1. The study was composed of patients suffering different disease severities (e.g., full vs. partial thickness tear), assuming that some of these patients would have visible muscle

shape changes and some would not. It should be noted that, out of necessity to obtain ground truth, this dataset consists of a selected, non-random, surgically-proven population.

After concluding the acquisition and segmentation procedures, the shape of each muscle is captured as a set of 2D contours. The contours are uniformly distanced from each other according to the slice thickness, and the first and last contours correspond to the anatomic ends of the muscle as determined by the radiology resident during the segmentation. Since the slice thickness ranges from 3 to 5 mm, the outcome of the segmentation step is a set of 3D points which are dense within the (sagittal) imaging planes, but sparse along the orthogonal direction of the planes. To obtain a set of points that is dense along all axes, we perform an interpolation of the contours in 3D. For this step, we use the interpolation module of the segmentation editor in the Amira software (Mercury Computer Systems, Inc), based on an implicit level set shape representation, similar to the work by Turk and O'Brien [21]. Finally, we extract uniformly spaced 2D contours from the sagittal planes of the interpolated 3D shape and use them as the input to our method.

Now, we describe in more detail the steps of our approach, as illustrated in Figure 2. In the formulations that follow, we denote scalars in regular typeface (e.g., n or N) and vectors in bold (e.g., $\mathbf{x}$ or $\mathbf{F}$), where the i-th entry of the vector $\mathbf{x}$ is written as $\mathbf{x}_i$.

2.1. Descriptor computation

Contour decomposition: Firstly, all of the contours are oriented in a consistent direction (clockwise or counterclockwise) and individually resampled to possess the same number of vertices and equidistant vertex positions
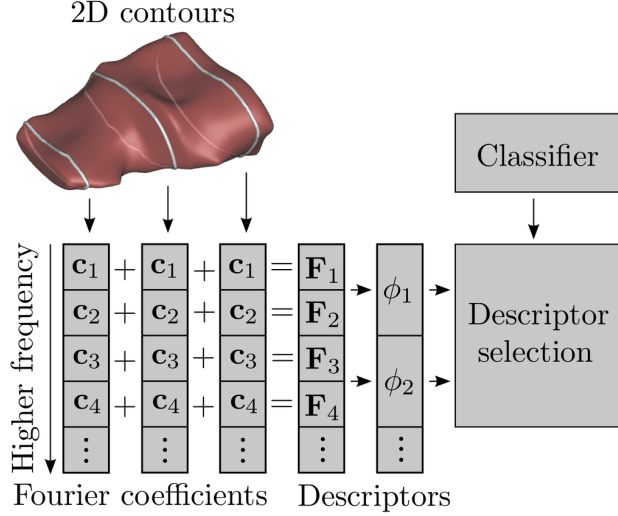
Figure 2: Proposed method: starting from the segmented muscle, the Fourier transform is applied to each individual contour (only 3 contours are shown for illustration purposes). The resulting normalized coefficients $\mathbf{c}_i$ are integrated for all the contours, yielding a vector of descriptors $\mathbf{F}$. The vector is then partitioned into sets ϕ_i and the best combination of sets is selected according to a classifier. The set size is 2 in this example for illustration purposes only.

within each contour. We set the target number of vertices as the maximum possible number of vertices in all of the contours, so that no geometric details are lost in this process. Next, we denote a 2D contour as a polygonal curve (x_n, y_n), where $n \in [0, 1, \ldots, N - 1]$, with N being the number of vertices in the polygon. We define a complex vector using the coordinates of the vertices and denote it as $\mathbf{x} = (x_0 + i\, y_0,\ x_1 + i\, y_1,\ \ldots,\ x_{N-1} + i\, y_{N-1})$. We then apply the discrete Fourier transform to this vector according to

$$\mathbf{f}_k = \sum_{n=0}^{N-1} \mathbf{x}_n \, \exp\left(-\frac{2\pi\, i}{N} kn\right), \tag{1}$$

10

where $k \in [0, 1, \ldots, N-1]$ and $\mathbf{f}$ is a vector of complex coefficients. The Fourier transform decomposes a signal into a set of frequency components that provide a complete, coarse-to-fine description of the signal. The coefficient $\mathbf{f}_k$ corresponds to the amount of frequency k that is present in the original signal. The frequencies k are increasing multiples of the sampling frequency.

Obtaining invariance: To obtain a vector of descriptors that is invariant to rotations of the contours in their 2D planes, we discard the phase component and make use only of the magnitude $|\mathbf{f}_k|$. When the original signal is interpreted as a combination of waves, the phase component captures the initial angle of the corresponding wave present in the signal, while the magnitude captures the amplitude or size of the wave. Discarding the phase component also provides invariance to selecting different starting points for the contours. Moreover, by discarding the DC-components ($|\mathbf{f}_0|$), we obtain invariance to translations of the contours in their 2D planes. The DC-components, which are the first coefficients given by the Fourier decomposition, correspond to the mean value of the signal's shape. We also seek invariance to scaling to account for the variation in the size of the muscles of different patients. So, we first rescale each muscle to fit in a unit cube while preserving its aspect ratio. Next, we divide each coefficient by the maximum magnitude associated to the first frequency of all the contours in the muscle, which has the effect of further normalizing the descriptors to any scale changes happening between the shapes of different patients. The rationale behind this procedure is that, if two contours have the same coefficients $\mathbf{f}_k$, but multiplied by different scales s_1 and s_2, then clearly, dividing all of the coefficients by one

specific non-zero coefficient cancels out changes in scale, i.e.,

$$\frac{s_1 \mathbf{f}_k}{s_1 \mathbf{f}_l} = \frac{s_2 \mathbf{f}_k}{s_2 \mathbf{f}_l}, \quad \forall k, \tag{2}$$

where $\mathbf{f}_l$ is a specific non-zero component. Since we discard the DC-components (set $|\mathbf{f}_0|$ to zero), we use the first frequency components $|\mathbf{f}_1|$ for the aforementioned scale normalization. Therefore, the normalized coefficients are given by $\mathbf{c}_k = |\mathbf{f}_k|/|\mathbf{f}_1^*|$, where $|\mathbf{f}_1^*| = \max_l |\mathbf{f}_1^l|$, with $\mathbf{f}_k^l$ denoting the coefficient associated to frequency k for contour l in the muscle.

Final descriptor: Each contour l has now an associated vector of descriptors $(\mathbf{c}_1^l, \mathbf{c}_2^l, \ldots, \mathbf{c}_{N-1}^l)$. We obtain our final shape descriptor $\mathbf{F}$ by integrating the different frequency magnitudes for all the contours,

$$\mathbf{F} = \left(\sum_l \mathbf{c}_1^l, \sum_l \mathbf{c}_2^l, \ldots, \sum_l \mathbf{c}_{N-1}^l \right). \tag{3}$$

The descriptor component $\mathbf{F}_k$ (referred to simply as *component* hereafter), indicates the total contribution of frequency k to the shape, measured on slices along a specific axis. The resolution within each slice is between 0.3 and 0.6 mm, so the frequencies analyzed are multiples of the sampling frequency, i.e. $k\frac{10}{6}$ to $k\frac{10}{3}$ cycles/mm. Since the contours lie on planes aligned in a consistent direction in all shapes and they are described by rotation-, translation-, and scale-invariant components, the final descriptor $\mathbf{F}$ is coherent across different shapes and hence the need to establish point correspondence is avoided. The descriptor implicitly captures the 3D shape of the muscle by recording the changes in scale between the contours of a given muscle and also the variations in the shapes of the contours at different frequencies. These frequency variations are integrated for all the contours in the muscle.

12

2.2. Descriptor selection and classification

The final step in our approach is to learn how components of the shape descriptors can best distinguish the different classes. Such a task can be carried out more efficiently by first selecting the components that best distinguish each pair of classes, and then combining the pairwise results to obtain the final classification [22]. Given a pair of classes and the shapes belonging to these two classes, the descriptors are selected according to their discriminatory power. That is, we train a classifier for each possible combination of components and verify its accuracy in distinguishing the two classes. The most accurate classifier is then chosen for the given pair of classes. Next, we combine the best pairwise classifiers using a voting method to obtain the final classification. That is, given a shape, we classify it into one of two classes according to all of the pairwise classifiers, and then we assign the shape to the class with the largest number of positive classifications. In this manner, given a muscle shape acquired from a new patient, we can assign it to one of the 5 pathology classes.

Partitioning and set combination: Moreover, to reduce the complexity of component selection and also to avoid the selection of single components that could be potential outliers, we propose a partitioning scheme. Notice that techniques for feature selection that exist in the machine learning literature could be alternatively used for this task [23]. Given a set size parameter S, we divide the descriptor vector $\mathbf{F}$ of each shape into contiguous sets $\Phi = [\phi_0, \phi_1, \ldots, \phi_{M-1}]$, where the j-th set is given by $\phi_j = [\mathbf{F}_{jS}, \mathbf{F}_{jS+1}, \ldots, \mathbf{F}_{jS+(S-1)}]$ and $M = \lfloor \frac{N}{S} \rfloor$. Each ϕ_j corresponds to a group of components associated with similar frequencies. Next, for each

13

possible combination of descriptor sets, e.g. $[\phi_0, \phi_1, \phi_3]$, we train a classifier and verify its accuracy in distinguishing the shapes (this procedure is described in more detail later). We keep the set combination that gives the best classification results in terms of accuracy.

In the experiments, we evaluate the performance of our approach with $S = 5$. This chosen set size is large enough to avoid selecting outlier components that can be unstable across different shapes. On the other hand, $S = 5$ is small enough so that the partitioning of the components is not too coarse, allowing the descriptor selection to find components that provide high accuracies in the classification.

The number of all possible combinations of sets is given by $\sum_{k=1}^{M} \binom{M}{k} = 2^M - 1$. To attenuate the exponential explosion of this search, we consider only components associated to frequencies below a threshold, since high frequency coefficients tend to be sensitive to noise. That is, given a frequency limit λ, we only consider ϕ_j that contain elements $\mathbf{F}_k$ where $k < \lambda$.

We choose λ in our experiments as 10% of the length of the contours. Although it might appear that much information is lost with this choice, when reconstructing the contours using only 10% of their Fourier coefficients, the maximum Hausdorff distance between any contour in the dataset and its reconstruction is 0.009 (less than 1%). The Hausdorff distance intuitively gives the maximum error between an original and a reconstructed contour, and is defined as

$$\mathrm{dist}_H(A, B) = \max_{a \in A} \min_{b \in B} \mathrm{dist}(a, b) \qquad (4)$$

where A and B are two contours, a and b are vertices on the contours, and $\mathrm{dist}(a, b)$ is the Euclidean distance between two vertices.

14

Moreover, we also restrict the maximum number of sets that can be combined during the search, since we verified empirically that aggregating a large number of sets does not improve classification accuracies over sparser combinations. To determine that, we conducted several tests where all possible combinations were considered and observed that combinations of three or four sets would have similar or superior accuracy to combinations with many more sets. So, the term $\sum_{k=1}^{M} \binom{M}{k}$ becomes $\sum_{k=1}^{L} \binom{M}{k}$, where L is the combination limit. We evaluate the performance of our approach using $L = 5$.

Classification: We make use of a support vector machine (SVM) [24] for classification. Given a set of input points and their corresponding class labels, an SVM constructs a hyperplane that best separates the data points into the different classes. The hyperplane can then be utilized to classify new data points. One reason for choosing SVMs in our work is that this type of classifier possesses the maximum margin property, which ensures maximum separation between the classes and low generalization error (the expected error of the classifier for previously unseen data). More specifically, we utilize a soft-margin non-linear version of SVM, which requires a parameter C that assigns a penalty to errors, and a parameter γ defining the width of a radial basis function (RBF). The RBF is used as a kernel to obtain the non-linearity of the SVM. Given the shape descriptors, a standard logarithmic grid search procedure is used to determine the optimal values for the parameters C and γ. We verify whether there is overfitting in the learned model by evaluating the classification accuracy, which gives an indication of the generalization ability of the model. The accuracy of the classification is evaluated according to a leave-one-out scheme. In this scheme, for each shape in a pair of classes, we

retain the shape for validation and train the pairwise classifier with all the remaining shapes. Next, we verify the accuracy in classifying the retained shape. Finally, we compute the average accuracy for all the validations. Notice that this procedure is repeated twice. First, leave-one-out is used to determine the best set of parameters C and γ for a pairwise classifier. Next, the best parameters are fixed and leave-one-out is utilized again to evaluate the accuracy that is reported for the pairwise classifier.

Multi-class scenario: After having obtained the best classifiers to distinguish between each pair of classes, we combine these pairwise classifiers using a voting method to address the multi-class case. The class for each shape is given by

$$\delta = \arg\max_i \sum_{j, j \neq i} p_{ij}, \tag{5}$$

where p_{ij} is the posterior probability that the shape belongs to class i according to the classifier trained to distinguish between classes i and j. In this work, $1 \leq i \leq 5$ and $1 \leq j \leq 5$.

3. Results

We present here the results of applying the proposed method for pairwise classification and then to the multi-class case. We are interested in confirming that it is possible to learn a classifier to infer the pathology of a muscle from its shape. Therefore, we compare the proposed method against previous work by comparing the accuracy of the classifiers, in order to determine how well each group of pathologies can be classified by the different methods. The results are evaluated according to a number of different measures: accuracy,

16

recall, precision, F-measure, G-mean [25], and the area under the receiver operating characteristic (ROC) curve [26]. We choose this specific set of measures because they possess different properties, e.g., the F-measure is not invariant to asymmetry (if there is a bias towards a specific class, it will be reflected in a low value). See [27] for details on these invariances. First, we formulate these measures and then we present the results.

Let us denote TP_{ij}, FP_{ij}, TN_{ij}, and FN_{ij} as the true positives, false positives, true negatives, and false negatives that appear in the classification when considering class i over class j, respectively. The accuracy of classification is then given by $\mathrm{A}_{ij} = (\mathrm{TP}_{ij} + \mathrm{TN}_{ij})/\mathrm{N}$, where N is the number of shapes in both classes. We define recall (R_{ij}), precision (P_{ij}), and the F-measure (F_{ij}) as

$$\mathrm{R}_{ij} = \frac{\mathrm{TP}_{ij}}{\mathrm{TP}_{ij} + \mathrm{FN}_{ij}}, \quad \mathrm{P}_{ij} = \frac{\mathrm{TP}_{ij}}{\mathrm{TP}_{ij} + \mathrm{FP}_{ij}}, \quad \text{and} \quad \mathrm{F}_{ij} = \frac{2\mathrm{R}_{ij}\mathrm{P}_{ij}}{\mathrm{R}_{ij} + \mathrm{P}_{ij}}. \tag{6}$$

When considering only two classes, we have that the recall R_{ij} is also known as the *sensitivity* for class i, while the recall R_{ji} is the *specificity* for class i. Moreover, the F-measure is a way of combining into a single number the recall and precision values. Since we are performing pairwise classification, we compute the average of the F-measure for the two classes involved, denoting it as $\overline{\mathrm{F}}_{ij} = (\mathrm{F}_{ij} + \mathrm{F}_{ji})/2$. An assessment of the overall classification is given by the G-mean, which is the geometric mean of recall values for all classes. It is formally defined as

$$G = \left(\prod_{i=1}^{K} \prod_{j=1,\, j \neq i}^{K} \mathrm{R}_{ij} \right)^{1/(K(K-1))}, \tag{7}$$

where K is the number of classes.

17

Finally, we also evaluate the results in terms of ROC curves, which show the change in trade-off between the false positive (FP) and true positive (TP) rates when varying the classifier parameters. The area under such a curve (AUC) is a summary of the overall accuracy of the classifier, with larger values implying higher accuracy. We denote the area under each curve in our experiments as AUC_{ij}. The dashed diagonal line in the ROC graphs corresponds to the *line of no discrimination*, i.e., to a classifier that uses random guesses to assign a shape to a class and has an AUC of 0.5.

For our experiments, we vary the C and γ parameters of the SVM (according to a logarithmic grid search from 2^{-8} to 2^8) and measure its performance. Basically, for each pair of parameters, we train a classifier and compute its FP and TP rates, which correspond to one operating point $(x, y) = (\mathrm{FP}, \mathrm{TP})$ in the graph. Next, we select the set of optimal operating points to constitute the ROC curve. An operating point (x, y) is optimal if there is no other point (x', y') such that $x' < x$ and $y' \geq y$. That is, if two operating points have the same TP rate, then we select the one with the lowest FP rate as being optimal. The ROC curves created in this manner inform the expected performance of the system [26].

We compare the results of three methods, described and labeled as follows.

- G: we consider the results obtained by previous work on the same dataset [5], where three global measures were utilized as shape descriptors: surface area, volume, and the ratio of the area to the volume. We refer to these three measures as *the global descriptors* from now on.

- G_{Opt}: an implementation of the same global descriptors, but when the classifier parameters C and γ are optimized to obtain the best

18

classification results.

- F: our method based on selection of Fourier descriptors.

Pairwise classification: Tables 2 to 6 present the values obtained for the assessment of the results. Each cell in a table corresponds to a classifier that distinguishes between the two indicated groups. A value of 100 denotes perfect classification results. Figures 3 and 4 show the ROC curves obtained for each pairwise classifier.

Multi-class scenario: To demonstrate the generalization of the proposed method to the multi-class case, we combine the pairwise classifiers using the voting method described in the previous section. In the case of the global descriptors (G_{Opt}), we obtain classification accuracies of 79%, 74%, 31%, 27%, and 45% for the groups N, T, TA, TR, and TAR, respectively, with an overall accuracy of 51%. Moreover, for the proposed method (F), we obtain classification accuracies of 100%, 95%, 85%, 80%, and 82% with respect to the same groups, with an overall accuracy of 88%.

Robustness to misalignments: Although the utilized descriptors are invariant to scale (implying invariance to muscles of different sizes) and invariant to rotations and translations of the contours on the slice planes, the accuracy of the method can be affected if there are errors in the consistent alignment of the muscles. To present some evidence that the method is robust to reasonable amounts of misalignment, we simulate cases where the muscles are not well aligned and apply the proposed method.

Basically, we create sets of misaligned muscles by taking each shape in our dataset and rotating it around the x, y, and z axes by a random angle

Table 2: Accuracy (A_{ij}) results for pairwise classification. Improvements in relation to G are marked with red color. Please refer to Section 3 and Table 1 for the meaning of the row and column labels.

	G				G_{Opt}				F			
	N	T	TA	TR	N	T	TA	TR	N	T	TA	TR
T	70				79				100			
TA	81	72			85	81			100	84		
TR	79	44	82		90	65	82		97	82	89	
TAR	76	73	50	73	88	73	67	73	100	87	92	81

Table 3: Recall (R_{ij}) results for pairwise classification. Improvements in relation to G_{Opt} are marked with red color. Please refer to Section 3 and Table 1 for the meaning of the row and column labels.

	G_{Opt}					F				
	N	T	TA	TR	TAR	N	T	TA	TR	TAR
N		64	93	93	93		100	100	100	100
T	89		95	68	79	100		95	89	89
TA	77	62		62	54	100	69		92	100
TR	87	60	100		100	93	73	87		87
TAR	82	64	82	36		100	82	82	73	

Table 4: Precision (P_{ij}) results for pairwise classification. Improvements in relation to G_{Opt} are marked with red color. Please refer to Section 3 and Table 1 for the meaning of the row and column labels.

	G_{Opt}					F				
	N	T	TA	TR	TAR	N	T	TA	TR	TAR
N		82	81	87	87		100	100	93	100
T	77		78	68	79	100		82	81	89
TA	91	89		100	78	100	90		86	87
TR	93	60	75		68	100	85	93		81
TAR	90	64	60	100		100	82	100	80	

Table 5: Average F-measure ($\overline{F}_{ij}$) and G-mean results for pairwise classification. Improvements in relation to G_{Opt} are marked with red color. Please refer to Section 3 and Table 1 for the meaning of the row and column labels.

	G_{Opt}				F			
	N	T	TA	TR	N	T	TA	TR
T	77				100			
TA	85	79			100	83		
TR	90	64	81		97	82	89	
TAR	88	71	66	67	100	86	91	80
	G-mean = 75				G-mean = 90			

21

Table 6: Area under ROC curve (AUC_{ij}) results for pairwise classification. Improvements in relation to G_{Opt} are marked with red color. Please refer to Section 3 and Table 1 for the meaning of the row and column labels.

	G_{Opt}				F			
	N	T	TA	TR	N	T	TA	TR
T	79				100			
TA	93	78			100	81		
TR	95	75	83		99	81	88	
TAR	87	75	81	72	100	87	96	82

of α_x, α_y, and α_z degrees, respectively. The x, y, and z axes correspond to the coronal, transverse, and sagittal axes of the muscles, respectively. By changing the maximum angle that α_x, α_y, and α_z can assume, we obtain misaligned datasets of increasing rotation and, therefore, of increasing difficulty. Next, we evaluate the accuracy of the proposed method on these misaligned datasets. However, to obtain a method that is robust to this form of noise, we train the classifiers with two misaligned datasets (corresponding to random rotations of 2 and 4 degrees), in addition to the original dataset. The rationale behind this approach is that we need to present some misaligned examples to the classifiers, so that they can learn which descriptors are more robust in this context.

Figure 5 presents the average accuracy for all the pairwise classifiers applied to the misaligned datasets, as well as the minimum and maximum accuracies among all the classifiers.
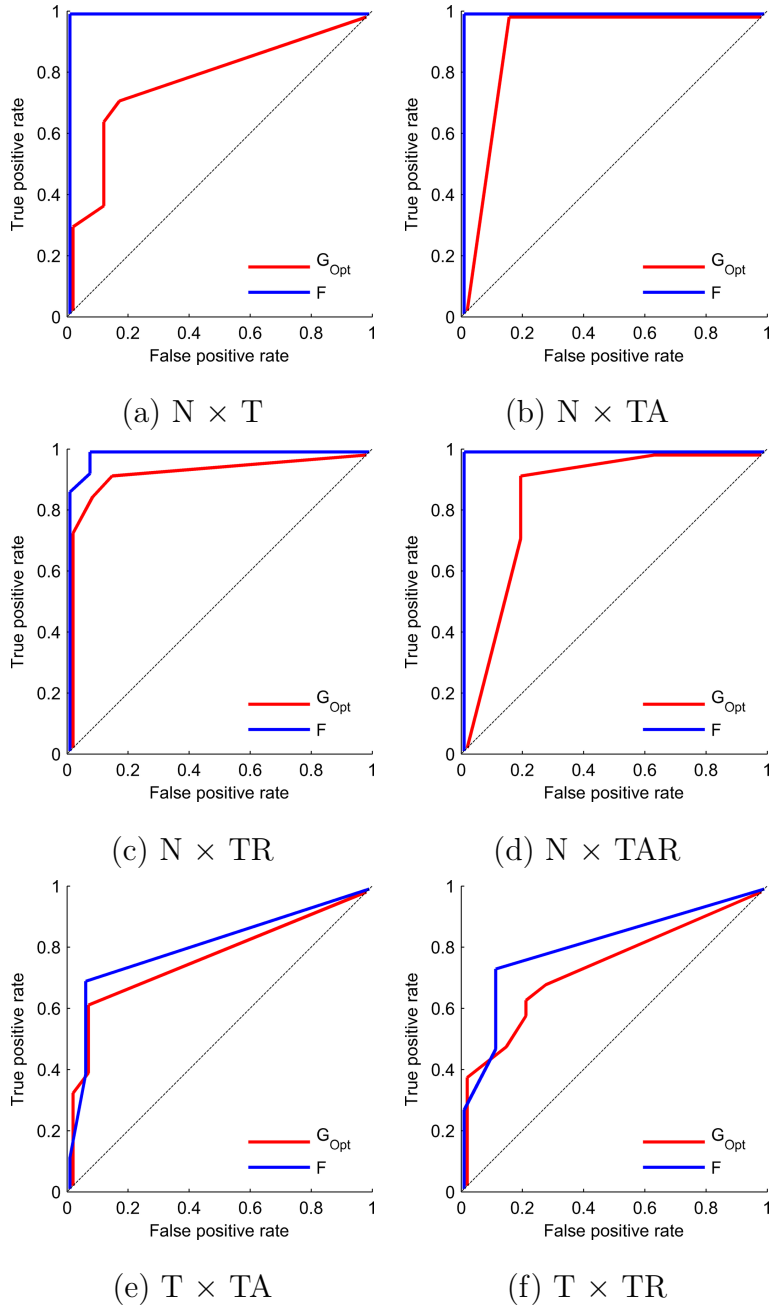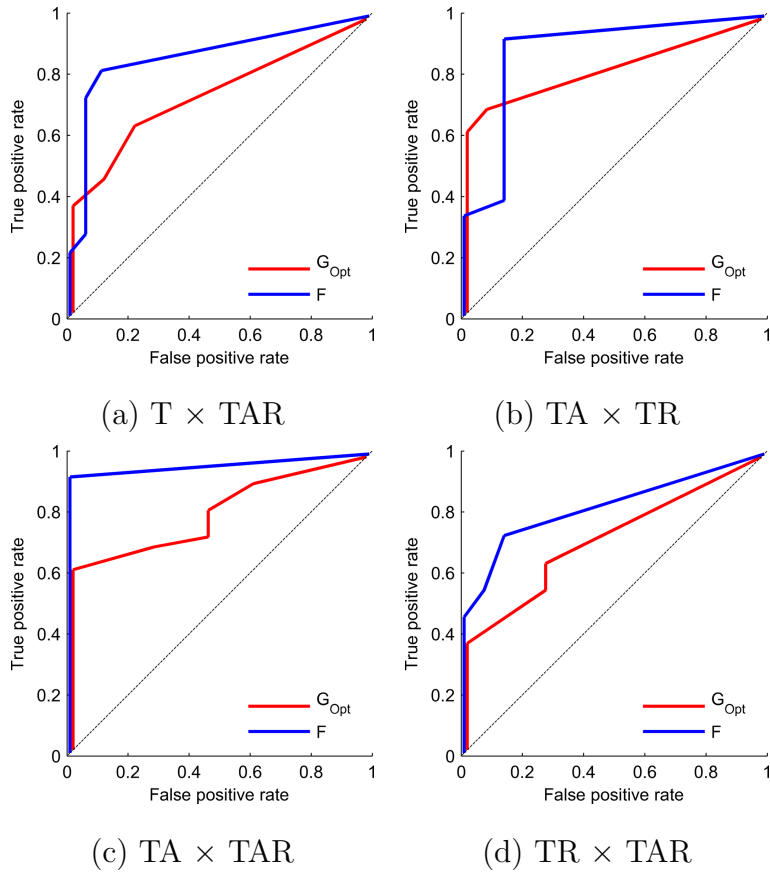
Figure 3: ROC curves for the pairwise classifiers. The red curve denotes the global descriptors (G_{Opt}), while the blue curve denotes the proposed method (F): (a) N × T, (b) N × TA, (c) N × TR, (d) N × TAR, (e) T × TA, (f) T × TR.

(a) T × TAR

(b) TA × TR

(c) TA × TAR

(d) TR × TAR

Figure 4: ROC curves for the pairwise classifiers. The red curve denotes the global descriptors (G_{Opt}), while the blue curve denotes the proposed method (F): (a) T × TAR, (b) TA × TR, (c) TA × TAR, (d) TR × TAR.
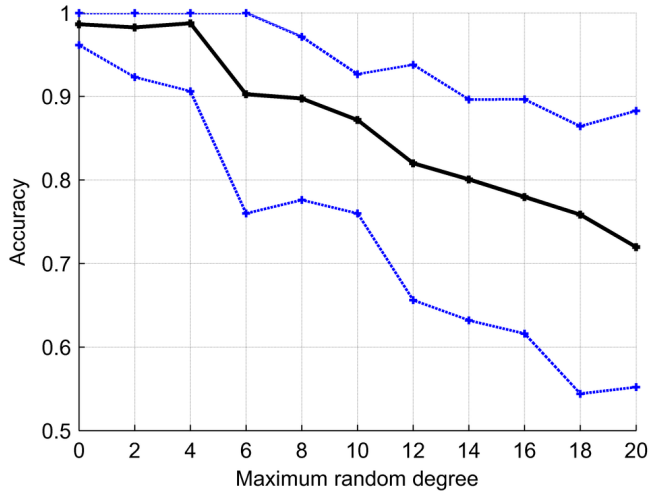
Figure 5: Evaluation of the robustness of the method to misalignments in the data: the black solid line shows the average accuracy for all the pairwise classifiers, while the blue dotted lines show the minimum and maximum accuracy among all the pairwise classifiers. We can see that, although the accuracy generally decreases for large angles, the minimum accuracy is still above 75% for misalignments created with rotations of up to 10 degrees.

Timing: The descriptor selection for each pairwise classifier takes approximately 40 minutes when using a 1 GHz AMD Opteron processor with 8 Gbytes of memory, since we run a full cross-validation experiment with classifier optimization for every combination of subsets. This timing is for the parameters mentioned before ($N = 1,000$, $S = 5$, and $L = 5$). Thus, the time required to obtain the best pairwise classifiers is approximately 7 hours. Once the results of this learning stage are available, the time required to classify one new muscle is on average 3 seconds (not including the time required for the initial steps of acquisition and segmentation).

4. Discussion

Pairwise classification: When considering the accuracy in Table 2, we see that although there is already some improvement from G to G_{Opt} by optimizing the classifier, we obtain the most significant improvement by utilizing the proposed method (F). We notice that the accuracy for distinguishing shapes in the normal class from shapes in the pathology groups is almost 100%, while the average accuracy for distinguishing between shapes that belong to the different pathology groups is 86%. This discrepancy in the classification accuracy could be attributed to the different forms of variation that the shapes possess. That is, the shapes of pathological muscles have clear differences when compared to normal muscles (e.g., differences in volume and area [5]). On the other hand, the difference between two muscles with different pathologies is more subtle and reflected in higher frequency detail, which also renders it more sensitive to noise and could account for the lower classification accuracy in this case.

Moreover, by analyzing the other evaluation measures in Tables 3 to 6, we notice a similar pattern of improvement. For the proposed method, we can see that the pairwise classifiers are well balanced and do not possess a significant bias towards any specific class, as seen in the precision, recall, and F-measure matrices. The values of the areas under the ROC curves also confirm these results.

Multi-class scenario: Moreover, we again see substantial improvement for the proposed method, with an increase in the overall accuracy from 51% (G_{Opt}) to 88% (F).

Robustness to misalignments: When considering Figure 5, we see that

the method has high accuracy for the three datasets that it was trained with (0, 2, and 4 degrees). Furthermore, we can also see that the average accuracy of the method still stays above 85% for the next three datasets (6 to 10 degrees), while the minimum accuracy is 75%. This experiment implies a robustness to 6 degrees of error in the misalignment. Additionally, for large misalignments, the maximum and average accuracies still stay above 85% and 70%, respectively.

Selected descriptors: When analyzing the results of descriptor selection, we observe a tendency for descriptors in the center of the set of frequencies lying below our chosen λ to be selected to distinguish between the non-normal pathology groups (related to frequency indices $k \in [30, \ldots, 60]$, for contours with $N = 1,000$ vertices). On the other hand, lower frequencies ($k \in [1, \ldots, 30]$) are preferred to differentiate between shapes in the normal group vs. shapes in the aggregation of all pathology groups. We speculate that this could correspond to the fact that there is a clear change in shape between normal and pathological shapes, but that the distinguishing factor between the shapes of different pathology groups is present in higher frequency details.

5. Conclusion and future work

We presented a method to classify 3D shapes of the supraspinatus muscle into several pathology groups, obtaining significant improvement when compared to previous work. For distinguishing shapes in the normal class from shapes in the pathology groups, we obtained an accuracy of almost 100%, which compares favorably with radiologists' accuracies of 89% to 98% in

distinguishing these same rotator cuff pathologies on MR images [20]. Moreover, for distinguishing between shapes that belong to the different pathology groups, we obtained an average accuracy of 86%. We have also demonstrated the robustness of the proposed representation to reasonable amounts of misalignment in the data. This evidence is motivational of future work examining the performance of our method on additional clinical data, with emphasis on measuring the augmentation of radiologists' diagnostic performance when provided with advice from our classifier. The results presented in this work suggest that such a future study would reveal the potential of our method as part of a clinically useful CAD system for the rotator cuff.

Our method is based on computing Fourier descriptors of a set of parallel 2D contours that capture the shape of a muscle. Although we applied the approach to the specific case of the supraspinatus, we would like to extend our method to serve as a general tool for CAD. This method can be potentially applied for the analysis of shapes that can be consistently aligned and decomposed into parallel contours, and where we do not know beforehand which are the distinguishing geometric variations in the shapes. Since the routine clinical workflow of MR image acquisition for the diagnosis of many musculoskeletal disorders includes the positioning of the patients' limbs to normalize for pose and gravity, our representation is not restricted to controlled research studies, but rather has potential in the analysis of the relationship between shape and pathology for other muscles.

Moreover, in relation to the set of techniques that constitute our method, the investigation of methods for automatic segmentation of the supraspinatus muscle is a topic of further research, allowing for a comparison between super-

vised and unsupervised approaches. Moreover, transforms that decompose 2D contours which have different properties when compared to the Fourier transform can be investigated (e.g., the wavelet transform). To address the exhaustive combination of sets for descriptor selection, we plan to make use of methods for feature selection [23] or spectral band selection [28] that exist in the literature. Finally, we obtained improved results for the multi-class case by making use of a voting method, but even better results may be obtained with more elaborate coupling methods, e.g., by solving an optimization problem that takes into consideration the difference between the pairwise probability estimates [22].

References

[1] S. Fuchs, C. Chylarecki, A. Langenbrinck, Incidence and symptoms of clinically manifest rotator cuff lesions, Int. J. Sports Medicine 20 (1999) 201–205.

[2] C. Lehman, F. Cuomo, F. J. Kummer, J. D. Zuckerman, The incidence of full thickness rotator cuff tears in a large cadaveric population, Bull. Hosp. Joint Dis. 54 (1995) 30–31.

[3] M. Zanetti, C. Gerber, J. Hodler, Quantitative assessment of the muscles of the rotator cuff with magnetic resonance imaging, Invest. Rad. 33 (3) (1998) 163–170.

[4] Y. Morag, J. A. Jacobson, B. Miller, M. D. Maeseneer, G. Girish, D. Jamadar, MR imaging of rotator cuff injury: What the clinician needs to know, Radiographics 26 (4) (2006) 1045–1065.

[5] A. Ward, G. Hamarneh, R. Ashry, M. Schweitzer, 3D shape analysis of the supraspinatus muscle: A clinical study of the relationship between shape and pathology, Academic Radiology 14 (10) (2007) 1229–1241.

[6] J. Ashburner, J. G. Csernansk, C. Davatzikos, N. C. Fox, G. B. Frisoni, P. M. Thompson, Computer-assisted imaging to assess brain structure in healthy and diseased brains, The Lancet Neurology 2 (2) (2003) 79–88.

[7] L. Wang, J. S. Swank, I. E. Glick, M. H. Gado, M. I. Miller, J. C. Morris, J. G. Csernansky, Changes in hippocampal volume and shape across time distinguish dementia of the Alzheimer type from healthy aging, NeuroImage 20 (2) (2003) 667–682.

[8] B. Gutman, Y. Wang, L. M. Lui, T. F. Chan, P. M. Thompson, Hippocampal surface analysis using spherical harmonic function applied to surface conformal mapping, in: Int. Conf. Pattern Recognition, Vol. 3, 2006, pp. 964–967.

[9] M. K. Chung, B. M. Nacewicz, S. Wang, K. M. Dalton, S. Pollak, R. J. Davidson, Amygdala surface modeling with weighted spherical harmonics, Lecture Notes in Computer Science 5128 (2008) 177–184.

[10] D. Nain, M. Styner, M. Niethammer, J. J. Levitt, M. E. Shenton, G. Gerig, A. Bobick, A. Tannenbaum, Statistical shape analysis of brain structures using spherical wavelets, in: Int. Symp. on Biomedical Imaging, 2007, pp. 209–212.

[11] S. Wang, M. K. Chung, K. M. Dalton, A. L. Alexander, R. J. Davidson,

Automated diagnosis of autism using Fourier series expansion of corpus callosum boundary, in: Human Brain Mapping Conf., 2007.

[12] S. Y. Y. Chen, P. E. Lestrel, W. J. S. Kerr, J. H. McColl, Describing shape changes in the human mandible using elliptical Fourier functions, Europ. J. of Orthodontics 22 (3) (2000) 205–216.

[13] A. Tsai, W. M. Wells, S. K. Warfield, A. S. Willsky, An EM algorithm for shape classification based on level sets, Medical Image Analysis 9 (5) (2005) 491–502.

[14] T. Heimann, H.-P. Meinzer, Statistical shape models for 3D medical image segmentation: A review, Medical Image Analysis 13 (4) (2009) 543–563.

[15] C. Gotsman, X. Gu, A. Sheffer, Fundamentals of spherical parameterization for 3D meshes, ACM Trans. on Graphics (Proc. SIGGRAPH) 22 (3) (2003) 358–363.

[16] L. H. Staib, J. S. Duncan, Boundary finding with parametrically deformable models, Trans. in Pattern Analysis and Machine Intelligence 14 (11) (1992) 1061–1075.

[17] G. Hamarneh, T. Gustavsson, Statistically constrained snake deformations, in: Proc. Int. Conf. on Systems, Man, and Cybernetics, Vol. 3, 2000, pp. 1610–1615.

[18] J. T. Lehtinen, M. J. Tingart, M. Apreleva, et al., Practical assessment of rotator cuff muscle volumes using shoulder MI, Acta Orthop Scand 74 (2003) 722–729.

[19] M. J. Tingart, M. Apreleva, J. T. Lehtinen, B. Capell, W. E. Palmer, J. J. P. Warner, Shoulder magnetic resonance imaging in quantitative analysis of rotator cuff muscle volume, Clinical Orthopaedics and Related Research 415 (2003) 104–110.

[20] P. L. Robertson, M. E. Schweitzer, D. G. Mitchell, F. Schlesinger, R. E. Epstein, B. G. Frieman, J. M. Fenlin, Rotator cuff disorders: Interobserver and intraobserver variation in diagnosis with MR imaging, Radiology 194 (1995) 831–835.

[21] G. Turk, J. F. O'Brien, Shape transformation using variational implicit functions, in: Proc. ACM SIGGRAPH, 1999, pp. 335–342.

[22] T.-F. Wu, C.-J. Lin, R. C. Weng, Probability estimates for multi-class classification by pairwise coupling, J. of Machine Learning Research 5 (2004) 975–1005.

[23] I. Guyon, A. Elisseeff, An introduction to variable and feature selection, J. of Machine Learning Research 3 (2003) 1157–1182.

[24] V. Vapnik, Statistical Learning Theory, Wiley, 1998.

[25] Y. Sun, M. S. Kamel, et al., Boosting for learning multiple classes with imbalanced class distribution, in: Int. Conf. on Data Mining, 2006, pp. 592–602.

[26] T. Fawcett, An introduction to ROC analysis, Pattern Recognition Letters 27 (2006) 861–874.

[27] M. Sokolova, Assessing invariance properties of evaluation measures, in: Proc. of NIPS Workshop on Testing Deployable Learning and Decision Systems, 2006.

[28] M. Skurichina, S. Verzakov, et al., Effectiveness of spectral band selection/extraction techniques for spectral data, Lecture Notes in Computer Science 4109 (2006) 541–550.