

Automated Search for Arthritic Patterns in Infrared Spectra of Synovial Fluid Using Adaptive Wavelets and Fuzzy C-Means Analysis

Jie Cui, Student *Member*, *IEEE*, John Loewy, and Edward J. Kendall

Abstract—Analysis of synovial fluid by infrared (IR) clinical chemistry requires expert interpretation and is susceptible to subjective error. The application of automated pattern recognition (APR) may enhance the utility of IR analysis. Here we describe an APR method based on the fuzzy C-means cluster adaptive wavelet (FCMC-AW) algorithm, which consists of two parts: one is a FCMC using the features from an M -band feature extractor adopting the adaptive wavelet algorithm and the second is a Bayesian classifier using the membership matrix generated by the FCMC. A FCMC – crossvalidated quadratic probability measure (FCMC-CVQPM) criterion is used under the assumption that the class probability density is equal to the value of the membership matrix. Therefore, both values of posterior probabilities and selection criterion M_{FQ} can be obtained through the membership matrix. The distinctive advantage of this method is that it provides not only the ‘hard’ classification of a new pattern, but also the confidence of this classification, which is reflected by the membership matrix.

Index Terms—adaptive wavelets, pattern recognition, arthritis, infrared spectrum.

I. INTRODUCTION

ARTHRITIS remains a difficult disease to diagnose [1;2]. It is particularly difficult to assess its early stage of development, partially due to the overlap of normal and abnormal joint physiology.

The analysis of synovial fluid (SF) plays a major role in the diagnosis of arthritis [3;4], which usually starts with recording the volume and the gross appearance of the fluid (*i.e.* color, clarity, viscosity and mucin clot), or gross analysis. Other procedures include microscopic tests, microbiologic tests, 1serologic tests, and chemical tests [1]. However, the conventional approaches usually make diagnosis subjective, as well as

potentially labor intensive and costly. Because usually no single test can distinguish among the various possible disorders, it may take several months or years before the clinical symptoms become sufficiently distinctive to make a definitive diagnosis, by which time irreversible damage to the joint may occur.

Some indirect measures of arthritic activity exist: C-reactive protein and erythrocyte sedimentation rate correlate with disability [3]. However, these measures reflect the inflammatory response rather than the synovial involvement. Recently, some evidence has been presented that suggests joint fluid might contain early indications of degenerative joint disease [5-7]. IR spectral features were found to specifically correlate with early versus late rheumatoid arthritis. We questioned if these features were accessible to automated pattern recognition techniques that potentially enhance the sensitivity and specificity of diagnosis.

Infrared spectrum (IRS) provides a molecular fingerprint based on the constituent functional groups characteristic vibrational energy [8]. For a biological fluid or tissue each of the major biomolecules present will contribute to the spectrum. The IRS of SF thus should be highly characteristic. If the disease process in the arthritic joint results in changes in the composition of the SF, this will alter the molecular fingerprint obtained by infrared spectroscopy, and the change in the fingerprint should be diagnostic. In addition, subtle biochemical differences may be detected through changes in the fingerprint of SF before there is significant physical manifestation of the disease process, allowing diagnosis at an earlier stage [7].

The potential value of semi-automatic IRS analysis of arthritis has been explored and demonstrated in a series of studies that exploited the near infrared ($4,000\text{ cm}^{-1} - 13,300\text{ cm}^{-1}$) and middle infrared ($400\text{ cm}^{-1} - 4,000\text{ cm}^{-1}$) regions [5-7;9;10]. The advantages of their methods are that combined application of infrared spectroscopy and pattern recognition principles provided a rapid (of the order of minutes), non-subjective method for the diagnosis of the major forms of arthritic disorders. An important technical advantage is that the method requires minute volumes (5-20 μl) of synovial fluid. Although these studies demonstrated that IR analysis of SF could aid in differential diagnosis of arthritic disorders, some limitations exist. Firstly, they adopted correct rate (CR), the ratio of the number of correctly classified samples to the total number, as the sole

Manuscript received March 2, 2005. This work was supported in part by a National Sciences and Engineering Research Council (Canada) operating grant to E. J. Kendall.

criterion to classify the data. The CR criterion proved quite variable, possibly due to the relatively small cohort size. Secondly, they provided “crisp” or “hard” classification at the completion of the algorithm; while an alternative “soft” classification approach provides a confidence measure when a new pattern is assigned to a class.

Published approaches to feature extraction for SF characterization have capitalized on the relative magnitude and position of absorption peaks. The SF IR “fingerprint” is characterized by a series of absorption peaks in the spectrum. On one hand, the position of the peaks is a characteristic of a functional group. In signal processing, these peaks are selected as instantaneous changes and contribute mainly to the high frequency components in signals. These components are not well characterized by traditional Fourier transform (FT). On the other hand, the shape and the width of an absorption peak (combined with the shift information of the peak) are due to a series of peak absorptions in a functional group frequency neighborhood. These absorptions reflect the chemical environment of the specific function groups. Previous papers [5;6] arduously searched spectra to identify discrete points as features. By ignoring the shape of the spectrum, the cost for feature extraction was very high.

In this paper, we present an approach to automating analysis of SF IRS with the purpose of making IRS a routine lab procedure that a non-specialist can handle. Our approach emphasizes two points: First, we adopt M -band adaptive wavelet filters to extract features. With this approach, both the position and the shape of the peaks, or the low and high frequency components, in a spectrum are evaluated in the process of feature extraction. Second, we employed fuzzy C-means classification [11] that gives the probability confidence of the classification.

II. BACKGROUND

A. System Outline

In principle, non-subjective diagnosis with the assistance of the IR analysis includes two distinct steps: obtaining high quality spectra within the desired IR segment, and the design of an effective algorithm of classification.

Our hypothesis is that retrievable characteristics of each disease variant will be encoded in the spectrum. In order to form effective and robust coefficients of the classifier, the spectra in training phase must be correctly classified by a physician beforehand.

We focused on mid-IR region since studies [5;7] have indicated that the C-H stretching absorption band ($2,800 \text{ cm}^{-1} - 3,050 \text{ cm}^{-1}$) might have extraordinary significance in classification. Unfortunately, water presents a strong absorption in mid-IR too. While our ultimate objective is to reduce or eliminate processing in these experiments, in order to minimize the potentially dominant water absorption band, we em-

ployed a thin dry film of SF.

In the following section, we describe the principles of the adaptive wavelet algorithm for feature extraction and the fuzzy C-means cluster (FCMC) [11] in sequence.

B. The Adaptive Wavelet Algorithm for Feature Extraction

A theory of wavelet analysis has been well developed by a number of researchers, *e.g.* [12-14]. The adaptive wavelet algorithm used in this paper is based on the work of [15-19]. Important definitions and results are presented below. A more detailed explanation is presented in [20].

Denote $c_{n,k}$ as the k -th scaling coefficients at level n , and $d_{n,k}$ as the k -th wavelet coefficients at level n . When we consider a wavelet matrix as a filter bank, the filtering operations from level n to $n + 1$ are then described by [16]

$$\begin{aligned} c_{n+1,k} &= \sum_{i=0}^{mg-1} l_i c_{n,mk+i} \\ d_{n+1,k}^s &= \sum_{i=0}^{mg-1} h_i^s c_{n,mk+i} \end{aligned}, \quad (1)$$

for $s = 1, 2, \dots, m-1$, and $n = 0, 1, \dots, L$, where L is the desired level of decomposition, and m and g are the rank and genus [17] of the wavelet matrix respectively. It is clear that the number of coefficients is reduced to $1/m$ when moving from one level to the next. This is also the standpoint for reducing the dimensionality of the feature vector that is one segment of these coefficients.

Define coefficients vector by

$$\mathbf{x}_n = \begin{bmatrix} c_{n,k} \\ d_{n,k}^1 \\ \vdots \\ d_{n,k}^{m-1} \\ c_{n,k} \end{bmatrix}_{M \times 1} \quad (2)$$

$$\text{for } k = 1, 2, \dots, N / (m^n - 1), n = 1, 2, \dots, L$$

and define the feature vector at level n by

$$y(n, t) = \mathbf{x}_n(t) = \begin{cases} c_{n,k}, & \text{if } t = 0 \\ d_{n,k}^t, & \text{if } t = 1, 2, \dots, m-1 \end{cases}. \quad (3)$$

It is assumed that the original discrete signal represents the scaling coefficients at level zero, $c_{0,k}$, for $k = 0, 1, \dots, N - 1$, where N is the dimension of the signal. Apparently, if L levels of the decomposition are required, then the dimensionality of the data should be $N \geq m^L$. Furthermore, when periodic boundary conditions are applied to the decomposition (1), we have

$$\begin{aligned} c_{n,k} &= c_{n, N / m^n + k} \\ d_{n,k}^s &= d_{n, N / m^n + k}^s \end{aligned}. \quad (4)$$

Before applying the algorithm, values for m , g , L and t must be specified.

C. Fuzzy C-Means Classification

1) Choice of the Discriminant Criterion

The analysis of FCMC reflects the information on the membership values of samples in different classes [21]. However, further effort is needed to optimize the classification of these samples. One approach is to convert the fuzzy set into a crisp set, and then chose a criterion based on correct rates or CR. This method has been adopted widely and the simple explanation credits its advantage.

The disadvantages of this criterion are equally apparent: first, when the sample size is small, the value of CR may vary greatly. For example, if the size of the sample is less than 100, every misclassification will result in a rate change greater than 1%; second, CR is a discrete measure which may not reveal the fundamental characteristics of the sample set; third, the use of CR assumes that all the samples in the set have the same credit in classification. Thus it ignores the fact that some samples may be closer to the center of the cluster and should be given more credits and some others locate on the edge of the cluster, which are perhaps of less importance in the evaluation of the performance of classifier.

For these reasons, a statistical approach, based on a variant of the cross-validated quadratic probability measure (CVQPM) [22], was taken to optimize selection criteria. This variant obeys the principles of Bayesian classification, takes advantage of the membership matrix and assesses the trustworthiness of class predictions made by the discriminant model.

2) Bayesian Classification

If one is interested in predicting the class membership of an observation, then one should aim to minimize the risk of misclassifying the object. This may be achieved by using a classification theory based on Bayes decision rules [23]. Suppose that there are c possible classes $\pi_1, \pi_2, \dots, \pi_c$, and N p -dimensional feature vectors, n_1, n_2, \dots, n_N , in the feature space. Consider the classification problem of assigning the object $n_k, k = 1, 2, \dots, N$, into one of c predefined classes. The Bayes rule assigns n_k to the class $\pi_i, i = 1, 2, \dots, c$, which maximizes the posterior probability [24]

$$P(p_i | n_k) = \frac{p(n_k | p_i)P(p_i)}{\sum_{i=1}^c p(n_k | p_i)P(p_i)}, \quad i = 1, 2, \dots, c \quad (5)$$

where $P(\pi_i)$ is the prior probability of belonging to class π_i , and $p(n_k | \pi_i)$ is the class probability density, *i.e.* the probability of object n_k arising from class π_i , so it requires that

$$\sum_{i=1}^c p(n_k | p_i) = 1, \quad 1 \leq k \leq N. \quad (6)$$

In practice, the class probability densities are often assumed to follow a multivariate normal distribution and can be written as

$$p(n_k | p_i) = \frac{1}{\sqrt{2\pi}^p \sqrt{|S_i|}} e^{-\frac{1}{2}(n_k - \bar{n}_i)^T S_i^{-1} (n_k - \bar{n}_i)}. \quad (7)$$

If n_k^i indicates that n_k is from class π_i and n_i^* the total number of samples in π_i in training set, then

$$S_i = \frac{1}{n_i^*} \sum_{k=1}^{n_i^*} (n_k^i - \bar{n}_i)(n_k^i - \bar{n}_i)^T \quad (8)$$

is the mean vector of class π_i , and

$$\bar{n}_i = \frac{1}{n_i^*} \sum_{k=1}^{n_i^*} n_k^i \quad (9)$$

is the mean vector of class π_i . Typically, \bar{n}_i and S_i are calculated from the training data.

If Bayesian linear discriminant analysis is employed, it assumes the class covariance matrices in (7) are equal, and are replaced with a pooled covariance matrix [25]

$$S_{pool} = \frac{1}{N} \sum_{i=1}^c n_i^* S_i, \quad (10)$$

where $N = \sum_{i=1}^c n_i^*$, to give

$$p(n_k | p_i) = \frac{1}{\sqrt{2\pi}^p \sqrt{|S_{pool}|}} e^{-\frac{1}{2}(n_k - \bar{n}_i)^T S_{pool}^{-1} (n_k - \bar{n}_i)}. \quad (11)$$

Bayesian linear discriminant analysis is a special case of Bayesian discriminant analysis, which is particularly suitable for small sample size.

3) FCMC-CVQPM Discriminant Criterion

The basic form of most probabilistic discriminatory measures is

$$P = \frac{1}{N} \sum_{k=1}^N a(n_k), \quad (12)$$

where $a(n_k)$ is the appreciation score for n_k . The simplest probabilistic function is the average probability that an object is assigned to the correct class, which has the appreciation score

$$a(n_k^{i'}) = P(p_{i'} | n_k^{i'}), \quad i' = 1, 2, \dots, c \quad (13)$$

where $P(p_{i'} | n_k^{i'})$ is the posterior probability for the true class of $n_k^{i'}$. If it is assumed that the class probability densities follow a multivariate normal distribution and the class covariance matrices are equal, the posterior probabilities can be calculated by substituting (11) into (5).

The quadratic appreciation score used is formulated as follows,

$$a_Q(n_k^{i'}) = \frac{1}{2} + P(p_{i'} | n_k^{i'}) - \frac{1}{2} \sum_{i=1}^c P(p_i | n_k^{i'})^2. \quad (14)$$

When substituting (14) into (12), we obtain quadratic probability measure

$$M_Q = \frac{1}{N} \sum_{k=1}^N a_Q(n_k). \quad (15)$$

Consider the properties of the quadratic probability measure in two extreme cases: if samples are completely classified into correct classes, *i.e.* $P(p_{i'} | n_k^{i'}) = 1$ and $P(p_i | n_k^{i'}) = 0, i \neq i'$, then $a(n_k^{i'}) = 1$. If completely misclassified, *i.e.* $P(p_{i'} | n_k^{i'}) = 0$ and $P(p_i | n_k^{i'}) = 1, \forall i \neq i'$,

then $a(n_k^{i'}) = 0$. Thus, CVQPM ranges from zero to one.

It has been shown that generally the quadratic probability measure is superior to CR criteria when used for assessing the performance of discriminant models [18;26]. Usually, larger values of the CVQPM are preferred, since this infers the classes can be differentiated with a higher degree of certainty.

However, the common approach above is based on the assumption of multivariate normal distribution of the class probability and of the equal class covariance, which is not always guaranteed. Furthermore, the calculation of the class probability densities by (7) through (11) is quite complicated and could occupy a lot of computing resources. On the other hand, because we have already obtained information from the analysis of FCMC and the membership matrix providing with the confidence of classification for every sample, can we take advantage of it?

An FCMC-CVQPM criterion that combines these features has been proposed. The basic idea is to use the information included in the membership matrix (cf. TABLE I). $u_{i,k}$ is the membership value of sample n_k belonging to class π_i , which satisfies

$$\sum_{i=1}^c u_{i,k} = 1, \quad 1 \leq k \leq N. \quad (16)$$

Recall the property of the class probability density (6), $u_{i,k}$ and $p(n_k | p_i)$ are equal mathematically. Under this assumption, replacing $p(n_k | p_i)$ in (5) by $u_{i,k}$, we have

$$P_{FQ}(p_i | n_k) = \frac{u_{i,k} P(p_i)}{\sum_{i=1}^c u_{i,k} P(p_i)}, \quad i = 1, 2, L, c \quad (17)$$

and subsequently $a_{FQ}(n_k^{i'})$ and M_{FQ} . Obviously, if all the prior probabilities of classes are equal, then

$$P_{FQ}(p_i | n_k) = u_{i,k}. \quad (18)$$

For Bayesian classification, the class probability density function $p(n_k | \pi_i)$, the probability of object n_k arising from class π_i , must satisfy condition (6). Similarly, the membership value $u_{i,k}$, a measure of the degree of the belonging of object n_k to class π_i , must satisfy similar conditions:

$$u_{i,k} \in [0, 1], \quad 1 \leq i \leq c; \quad 1 \leq k \leq N \quad (19)$$

$$\sum_{i=1}^c u_{i,k} = 1, \quad 1 \leq k \leq N \quad (20)$$

$$\sum_{k=1}^N u_{i,k} > 0, \quad 1 \leq i \leq c \quad (21)$$

where c is the number of classes and N the number of objects to be classified. We use the membership value to approximate the probability density functions in Bayesian classifier.

The principal advantages of the measure of FCMC-CVQP are that it doesn't waste the information from FCMC and the complexity of calculation is less than the traditional one. Furthermore, not only does it have a indicator of total quality of classification, M_{FQ} (the function of the optimizer), but also it provides quantitative values, $u_{i,k}$, reflecting the confidence of

every sample belonging to each class. Finally, in principle, the assumption of multivariate normal distribution of the class probability and of the equal class covariance is not necessary. This should enhance the robustness of the algorithm when operating on real data where these criteria may not be met.

III. METHODS AND RESULTS

A. Experiments using Simulated Test Figures

Simulated data were constructed as described below (22) – (24). The ramp and rectangular functions are intended to simulate the broad spectral features that we eventually intend to classify in the real spectra. The 'rectangular' and 'ramp' waveforms shown in the simulation feature both continuous areas and abrupt changes. Abrupt changes are characteristic of infrared spectra (e.g. Fig. 7). The large absorptions peaks within the Amide band ($1273 \text{ cm}^{-1} - 1800 \text{ cm}^{-1}$) or the small peaks in the C-H stretching band ($2778 \text{ cm}^{-1} - 3149 \text{ cm}^{-1}$) feature sharp discontinuities. The simulation is intended to determine the behavior of the algorithm on well defined signals.

The dimensionality N of the simulated data, i.e. the number of variables, is 128. There are totally three classes. A single spectrum from each class is generated as follows, for $j = 1, \dots, 128$,

$$\text{Class 1: } x_j^{(1)} = (6 + h) \times c_{[a,b]}(j) + e(j), \quad (22)$$

$$\text{Class 2: } x_j^{(2)} = (6 + h) \times c_{[a,b]}(j)(j - a) / (b - a) + e(j), \quad (23)$$

$$\text{Class 3: } x_j^{(3)} = (6 + h) \times c_{[a,b]}(j)(b - j) / (b - a) + e(j), \quad (24)$$

Here a and b are two random values initialized by $a \sim U_z(16, 32)$ and $b \sim U_z(64, 80)$, where U_z denotes the integer-valued uniform distribution. $c_{[a,b]}(j)$ is the character function on the interval $[a, b]$, i.e.

$$c_{[a,b]}(j) = \begin{cases} 1, & j \in [a, b] \\ 0, & \text{otherwise} \end{cases}. \quad (25)$$

h and $e(j)$ are also random values that abide by normal distribution $N(0,1)$. The four random variables a , b , h and $e(j)$ function as added noise on the figure. In order to study the influence of noise on the ability of the algorithm, noise was added in nine incremental levels ranging from zero to eight.

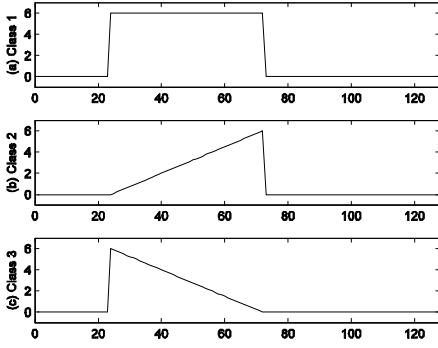


Fig. 1. Simulated test figures with noise level zero.

For example, Fig. 1 illustrates the spectra without noise, where $a = 24$ and $b = 72$ are the middle numbers of their interval. Fig. 2 and Fig. 3 show the spectra with noise level 5 and level 8 respectively. In each of these two figures, a totally nine spectra are collected with three spectra in each class.

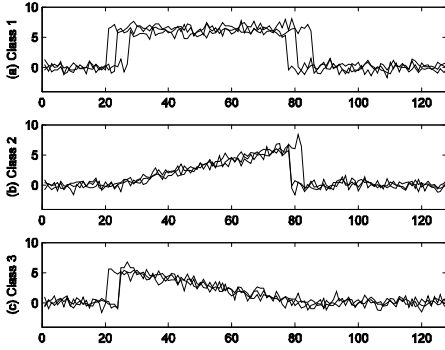


Fig. 2. Simulated test figures with noise level five. The average signal to noise ratio is 12.8 dB.

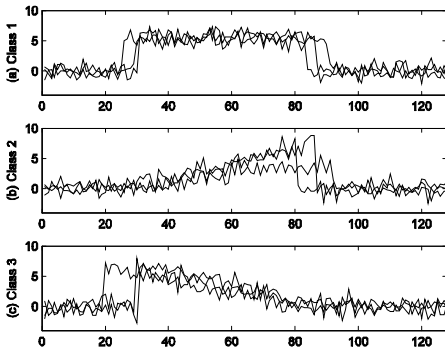


Fig. 3. Simulated test figures with noise level eight. The average noise level is 5.2 dB.

The increase in vagueness of the figures between different classes can be quantitatively reflected in the membership matrix.

The experiments were carried out with different combination of the parameters z (fuzzy index), m , g , L and t , and then the

values of M_{FQ} and correct rate (CR) were collected. However, to calculate CR, we need to know to which predefined class the spectrum is assigned. Usually we can only know the distribution of samples in the cluster, but not know to which predefined class the cluster correspondents after FCMC. To do this, a “marker figure” with known classification was defined. Recall that a figure will be assigned into a class if it acquires the highest posterior probability for that class. Thus, by definition, the marker figure obtains the highest probabilities for their predefined classes. In practice, for some kinds of clustering, if more than one marker obtains the highest posterior probability for a cluster, CR could not be calculated and this results in a ‘failed cluster’.

Since we are interested in studying the case of small sample set (fewer than 100 samples, typically), the choice of the parameters will be considered under this assumption.

B. Simulation Experiment Results and Discussion

TABLE I shows a typical membership matrix with three clusters and nine sample spectra at the initialization of the FCMC-AW (adaptive wavelet) algorithm. The noise level was eight and the parameters were $z = 2$, $m = 4$, $g = 3$, $L = 2$ and $t = 0$. In this table, $\pi_1 - \pi_3$ stand for three predefined classes, and $mk_1 - mk_3$ are three marker spectra indicating three classes respectively as is shown in Fig. 3. For simplicity, the first three samples $n_1 - n_3$ came from class π_1 , $n_4 - n_6$ from π_2 , and $n_7 - n_9$ from π_3 . Since it was assumed that the prior probabilities of different classes were the same, the membership values were equal to the posterior probabilities of the samples. Shaded blocks point out the highest posterior probabilities of different samples, which also show the classification of the spectra, according to Bayesian discriminator. Thus the CR was one. The measure of M_{FQ} can be obtained from the table, which was 0.8027.

TABLE I
FCMC-AW ALGORITHM: INITIAL MEMBERSHIP MATRIX^{1,2,3}

	Marker Figures			Test Figures								
U	mk_1	mk_2	mk_3	n_1	n_2	n_3	n_4	n_5	n_6	n_7	n_8	n_9
π_1	0.43	0.04	0.04	0.99	0.94	0.85	0.03	0.01	0.06	0.01	0.01	0.03
π_2	0.22	0.75	0.21	0.01	0.03	0.06	0.89	0.96	0.77	0.03	0.04	0.11
π_3	0.35	0.21	0.75	0.01	0.03	0.09	0.08	0.03	0.17	0.96	0.95	0.86

¹Test data with noise level 8

² $z = 2$, $m = 4$, $g = 3$, $L = 2$ and $t = 0$.

³ π_1, π_2, π_3 are three predefined classes

TABLE II
FCMC-AW ALGORITHM: OPTIMIZED MEMBERSHIP MATRIX¹

	Marker Figures			Test Figures								
U	mk_1	mk_2	mk_3	n_1	n_2	n_3	n_4	n_5	n_6	n_7	n_8	n_9
π_1	0.55	0.04	0.06	0.98	0.93	0.97	0.01	0.01	0.08	0.00	0.01	0.05
π_2	0.14	0.86	0.11	0.01	0.04	0.01	0.98	0.98	0.86	0.00	0.01	0.05
π_3	0.31	0.10	0.83	0.01	0.04	0.02	0.01	0.02	0.07	1.00	0.98	0.90

¹Test data with noise level 8

At the completion of the algorithm, the quadratic measure M_{FQ} was 0.9967 and CR was still one. The membership matrix was presented in TABLE II. It can be observed that the posterior probabilities of the samples were enhanced.

TABLE III
FCMC-AW ALGORITHM: INITIAL MEMBERSHIP MATRIX¹

	Marker Figures			Test Figures								
U	mk_1	mk_2	mk_3	n_1	n_2	n_3	n_4	n_5	n_6	n_7	n_8	n_9
π_1	0.63	0.14	0.04	1.0	1.0	0.9	0.04	0.07	0.03	0.08	0.07	0.04
π_2	0.12	0.57	0.05	0.02	0.02	0.04	0.91	0.88	0.94	0.08	0.07	0.06
π_3	0.26	0.29	0.91	0.02	0.02	0.03	0.05	0.05	0.03	0.83	0.86	0.90

¹Test data with noise level 5

TABLE III shows a membership matrix upon initialization of the algorithm with the same parameters as is shown in TABLE II except that the noise level is 5. The M_{FQ} was 0.9925. Note that the M_{FQ} for TABLE III was 23.7% higher than that for TABLE I. Thus M_{FQ} can reflect the degree of the fuzziness of the classification.

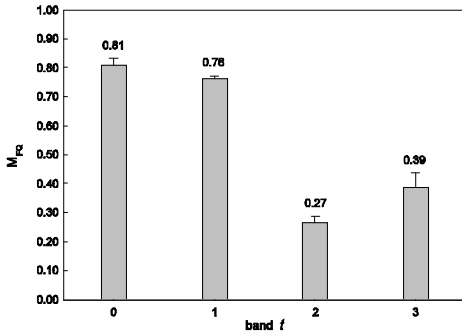


Fig. 4. Average and standard deviation of M_{FQ} at the initialization of the algorithm at the band of $t = 0, 1, 2, 3$. The noise level is 8.

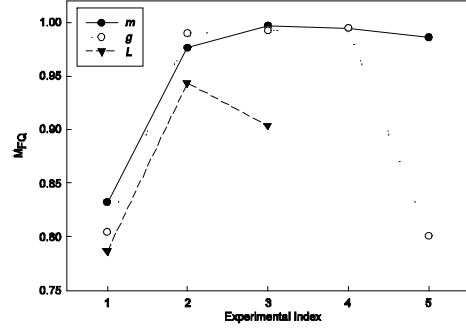


Fig. 5. Optimization of M_{FQ} using m, g, L parameters at noise level 8.

TABLE IV
INITIAL AND FINAL MEMBERSHIP VALUES WITH DIFFERENT VALUES OF m, g , AND L AT NOISE LEVEL EIGHT

	Values	Initialization	Completion
m	2	0.65	0.83
	3	0.80	0.98
	4	0.79	1.00
	5	0.80	0.99
	6	0.81	0.99
g	1	0.80	0.80
	2	0.99	0.99
	3	0.99	0.99
	4	0.99	0.99
	5	0.80	0.80
L	1	0.78	0.79
	2	0.93	0.94
	3	0.85	0.90

With fuzzy index z set to 2 and the other parameters set as follows: $m=4, g=3, L=2$, we consider the choice of band t for $t = 0, 1, 2$ and 3, based on the highest M_{FQ} in the algorithm. The average and standard deviation of 10 tests of M_{FQ} at the initialization with noise level 8 is shown in Fig. 4. The results shown in Fig. 5 indicate that the M_{FQ} at $t = 0$ has the highest value and this was carried forward in the optimization procedure. For other values of m, g , and L , band $t = 0$ still had the highest M_{FQ} at the initialization.

We next considered the choice of other parameters: m, g and L . The optimization procedure did not produce significant better results using the test figures at noise level less than five. However, significant improvement was achieved using the test figures with higher noise levels (TABLE IV).

Optimal values for m, g and L , based on the M_{FQ} calculation were obtained (Fig. 5). M_{FQ} attained its maximum value when $m = 4, g = 3$ and $L = 2$.

Recall that we are interested in classifying small samples. The recommended ratio of feature vector to the sample size dimension is 1 to 10. The length of the original simulated data was 128 and consequently for $L = 2$ and $m = 4$, the dimension of feature vector was $128/4^2 = 8$, a satisfactory value. Although the M_{FQ} reached a maximum when g was 4, $g = 3$ was chosen to reduce calculation complexity. Thus the choices of parameters

for simulated sample set were $z = 2$, $t = 0$, $m = 4$, $g = 3$ and $L = 2$. For this setting of parameters, 90 test figures were tested after training with 9 predefined simulated samples. The value of M_{FQ} at the completion of the algorithm for the figures with noise level 5 was 0.9925 ± 0.0021 and the CR was 100%; at noise level 8 M_{FQ} was 0.8839 ± 0.091 and the CR was 98.67%, indicating an average mis-classification of 1.2 spectra.

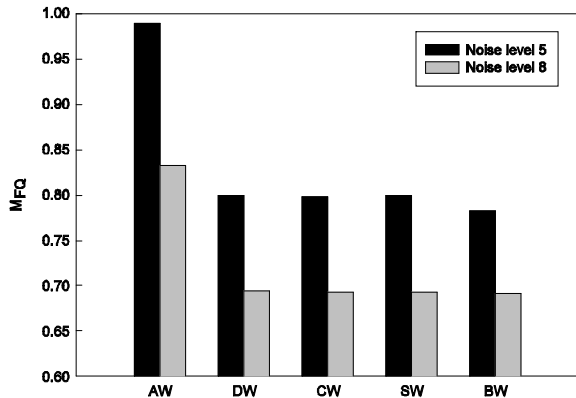


Fig. 6. M_{FQ} for adaptive and fixed wavelets using test figures with noise levels 5 and 8. AW, Adaptive wavelet; DW, Daubechies Wavelets, CW, Coiflet Wavelets; SW, Symlet Wavelets and BW, Biorthogonal Wavelets

Finally, we should highlight the advantage of using an adaptive wavelet as opposed to a predefined wavelet. It should be noted that the filter bands of the fixed wavelets was a fixed number $m = 2$. Fig. 6 presents the values of M_{FQ} at the completion of the algorithm for the 90 simulation spectra with noise levels 5 and 8 respectively. The number of filter bands of the adaptive wavelets was $m = 2$.

It is demonstrated that the values of M_{FQ} calculated by the adaptive wavelet algorithm were significantly higher than those by fixed wavelets when the number of the filter band is $m = 2$. Since adaptive wavelets algorithm can search the highest effective classification when m is more than two it has a far greater potential for further improving the results.

C. Experimental Results Using Synovial Fluid (SF) and Cerebral Spinal Fluid (CSF)

Physicians at Royal University Hospital aspirated SF samples during surgery on knee and hip joints. CSF samples were aspirated by the physicians prior to myelography. All diagnoses were made by the rheumatologist in attendance, based upon established criteria, using standard clinical and laboratory tests.

Infrared spectra were recorded on a Digilab FTS 60 FT-IR spectrometer equipped with a liquid nitrogen cooled mercury cadmium telluride detector and continuously purged with dry air. For each sample, 256 mid-IR interferograms were recorded, signal averaged, and Fourier-transformed to generate a spectrum with a nominal resolution of 2 cm^{-1} . To facilitate in-

ter-sample comparison the spectra were normalized to maximum intensity.

1) Classification of SF Samples

We had a total of seven SF specimens with diagnostic information, including three fluid specimens drawn from patients suffering from rheumatoid arthritis (RA), three from patients suffering from osteoarthritis (OA), and one from a patient suffering from trauma without evidence of arthritis. The latter specimen was used as control in the classification of SFs. The three RA spectra were denoted as RA1 – RA3, OA spectra as OA1 – OA3.

Since the size of the sample set was very limited, it was not practical to divide the set into a training set and a testing set. The leave-one-out (LOO) [27] method was employed to test the effectiveness of the algorithm. Under cross-validation, the available data are divided into k disjoint sets; k models are then trained, each on a different combination of $k-1$ partitions and tested on the remaining partition. Cross-validation makes good use of the available data since each pattern is used as both as training and testing data. Cross-validation is especially useful where the amount of data is too small to form the usual training and testing data sets. The most extreme form of cross-validation, where k is equal to the sample number is known as leave-one-out cross validation. Following this idea, in our experiments, every specimen was sampled with two spectra, and the algorithm was trained using one of the spectra. After the optimal wavelet matrix A was found, another spectrum of the same specimen was applied to the discriminator to determine its classification.

The experiments revealed that the algorithm obtained the highest M_{FQ} in the training phase when the parameters were set as: $m = 4$, $g = 2$, $L = 3$ and $t = 0$. Three approaches to training the algorithm were investigated. The first way was to use the whole spectrum ($1,273 \text{ cm}^{-1} - 3,982 \text{ cm}^{-1}$) to train the algorithm (cf. Fig. 7). The CR of classification was 100% and the value of M_{FQ} was 0.9501. The second way was to use only the Amide band of the spectrum ($1273 \text{ cm}^{-1} - 1800 \text{ cm}^{-1}$) to train the algorithm. The value of M_{FQ} was reduced to 0.9050 and the CR was 85.71% with one misclassified spectrum. The third way was to use only the C-H stretching band of the sample spectrum ($2778 \text{ cm}^{-1} - 3149 \text{ cm}^{-1}$) in the training of the algorithm. Although only 384 points of the spectrum were involved in training, the value of M_{FQ} was enhanced to 0.9815 and the correct rate was 100%. These measures are collected in TABLE V.

TABLE V
THE VALUES OF M_{FQ} AND CORRECT RATE (CR) USING DIFFERENT RANGES OF THE SPECTRA

M_{FQ}	CR (%)	Spectral Range	Wave-number cm^{-1}
0.9050	85.7	Amide Band	1273 – 1800
0.9501	100.0	Whole Spectrum	1273 – 3976
0.9815	100.0	C-H Stretching Band	2778 – 3079

TABLE VI
SYNOVIAL FLUID CLASSIFICATION MATRIX USING THE AMIDE BAND

<i>U</i>	<i>Control</i>	<i>RA1</i>	<i>RA2</i>	<i>RA3</i>	<i>OA1</i>	<i>OA2</i>	<i>OA3</i>
<i>Control</i>	1.00	0.03	0.07	0.02	0.01	0.03	0.39
<i>RA</i>	0.00	0.91	0.85	0.95	0.03	0.07	0.53
<i>OA</i>	0.00	0.06	0.08	0.03	0.96	0.90	0.09

TABLE VII
SYNOVIAL FLUID CLASSIFICATION MATRIX USING THE C-H STRETCHING BAND

<i>U</i>	<i>Control</i>	<i>RA1</i>	<i>RA2</i>	<i>RA3</i>	<i>OA1</i>	<i>OA2</i>	<i>OA3</i>
<i>Control</i>	1.00	0.00	0.08	0.01	0.00	0.03	0.00
<i>RA</i>	0.00	0.99	0.82	0.73	0.06	0.16	0.02
<i>OA</i>	0.00	0.01	0.11	0.25	0.94	0.81	0.98

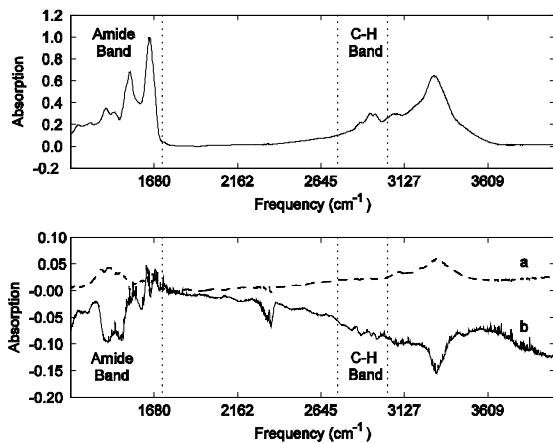


Fig. 7. Synovial fluid marker spectra for control and the spectral difference between RA, OA and Control. Top: control spectrum; Bottom: spectral difference. a: RA-Control, b: OA-Control.

The table indicates that the C-H stretching band of the spectrum has the highest significance in classification. Because only 384 points was used and with $m = 4$, $L = 3$, the feature vector contained only 8 points. This represents a significant reduction in complexity when compared to the original signal. The sample classification matrices using the Amide and C-H stretching bands are presented in TABLE VI and TABLE VII.

In these tables the membership values are identical to the posterior probabilities, since the prior probabilities of predefined classes were the same. Since the membership value of OA3 increased significantly, it compensated for losses elsewhere in the level of agreement with the predefined classification. However, the posterior probabilities in TABLE VI may indicate that OA3 may be better classified into the RA class.

TABLE VII provides comparable results using the C-H stretching band (Fig. 7). Here, the posterior probabilities suggest that all samples are correctly classified.

TABLE VIII
CONFUSION MATRIX OF SF CLASSIFICATION

	Control	RA	OA	Total
Control	1	1	0	2
RA	0	1	0	1
OA	0	1	3	4
Total	1	3	3	7

TABLE IX
SYNOVIAL FLUID CLASSIFICATION USING AMID BAND

	<i>Control</i>	<i>RA1</i>	<i>RA2</i>	<i>RA3</i>	<i>OA1</i>	<i>OA2</i>	<i>OA3</i>
M_{FQ}	0.9809	0.9834	0.8979	0.863	0.9874	0.9471	0.9756
<i>CR</i>	100%	100%	85.71%	85.71%	100%	100%	100%

A confusion matrix generated using the LOO technique was computed to cross-validate the SF classification (TABLE VIII). The value of Kappa [28;29], κ , was 0.60, which indicated that the agreement between the clinical diagnosis and the classification by the algorithm was significantly higher than the value by random chance.

The CR was not quite satisfied (TABLE IX), which was mainly due to the very limited size of the sample set, since, for seven samples, every misclassified sample will contribute as much as 14.29% variations to the correct rate. Nevertheless, it may be more informative to compare M_{FQ} with a correct rate. Note that the M_{FQ} values were less sensitive to small sample size than the CR metric. Furthermore, even for a given correct rate the M_{FQ} values may differ, providing the physician additional information towards alternative classification. For example, the CRs of RA1 and RA3 are the same, but corresponding M_{FQ} are not. On the other hand, no such information can be derived from the CR alone.

2) Classification of CSF

CSF spectra (Fig. 8) were employed as a reference to compare with the classification effectiveness of the algorithm on SF. The seventeen CSF spectra were collected from 10 patients diagnosed with OA, and 7 patients considered free of OA.

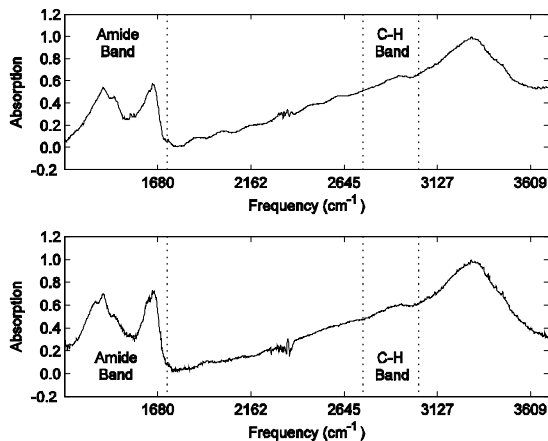


Fig. 8. Marker Cerebral Spinal Fluid (CSF) Spectra of Control and OA. (Normalized absorption versus frequency (cm^{-1})).

TABLE X
CSF CLASSIFICATION MATRIX FOR THE TRAINING PHASE USING
THE AMIDE BAND

U	mk_1	mk_2	Cont-r ol_1	Cont-r ol_2	Cont-r ol_3	OA_1	OA_2	OA_3	OA_4
Control	0.90	0.08	0.84	0.07	0.95	0.55	0.27	0.08	0.99
OA	0.10	0.92	0.16	0.93	0.05	0.45	0.73	0.92	0.01

TABLE XI
CONFUSION MATRIX OF CSF CLASSIFICATION

	Control	OA	Total
Control	3	5	8
OA	1	1	2
Total	4	6	10

The sample set was divided into a training set (3 control and 4 OA specimens) and a testing set (4 control and 6 OA specimens). In the training phase, the parameters were set the same as for the SF classification ($z = 2$, $m = 4$, $g = 2$, $L = 3$ and $t = 0$). For CSF, the Amide band was selected for classification, since the absorption of C-H stretching band was very weak in the CSF. The marker spectra for these two classes were the average spectra from these two classes, shown in Fig. 8. The posterior probabilities (TABLE X) suggest that the OA samples are as likely to be classified as control as OA. The training phase for CSF classification achieved an $M_{FQ} = 0.6779$ and a CR = 57.14% indicating that classification was coincident with random distribution. The test set confirmed this observation (TABLE XI). Here, the value of κ from the LOO confusion matrix was around zero. This confirms that no correlation could be found between OA and infrared spectra using CSF spectra.

IV. DISCUSSION

Previous IR correlation studies determined that features of SF spectral features were positive indicators of arthritic disorders [5-7;9]. However, the analytical approach placed limitations on the utility of the technique. First, they adopted CR as the only criterion to evaluate the results of classification. Generally their results were obtained for the case of a small sample set. As a result, the CR value proved quite variable. In this study, the FCMC-CVQPM (measured by M_{FQ}) instead of CR was employed, and therefore the evaluation of the algorithm was less influenced by the sample size. Earlier studies adopted “crisp” or “hard” classification. Here we found that even for the same CR, the ‘goodness’ of a classification might differ. Our algorithm provided this additional information as a M_{FQ} index. In addition, the entries of the membership matrix reflected the classification confidence of each new pattern when it was assigned to a predefined class. This information may provide a confidence metric for physicians in reporting a diagnosis.

The “fingerprint” of a sample is characterized by a series of absorption peaks in the spectrum. In terms of signal processing, these peaks are the instantaneous changes, contributing mainly to the high frequency components in bio-signals, which are not well characterized by traditional FT. On the other hand, the shape and the width of an absorption peak (combined with the shift information of the peak) are due to a series of peak absorptions in its frequency neighborhood, which reflect its chemical environment. This important information mainly contributes to the low frequency components of the signal and may not be well characterized in values of discrete points. Previous reports [5;6] emphasized signal position and amplitude and arduously searched the spectra to identify discrete features. Thus they ignored the shape of the spectrum and the cost for feature extraction was high.

In this paper, our approach was based on the algorithm of the M -band adaptive wavelet filters. Wavelet technology has been widely used in signal processing to characterize instantaneous or non-stationary signals, *e.g.* [30]. The M -band filters used divided the signal into one low and several high frequency bands; the feature vectors were searched among these bands. Therefore, both the position and the shape of the peaks, or low and high frequency components, in a spectrum could be evaluated in the process of feature extraction.

According to the experimental results, we found that the feature vector at band $t = 0$ was more effective in the SF classification than vectors at the other bands $t = 1, 2, \dots, m-1$. We know that band $t = 0$ consists mainly of the low frequency components and contributes to the envelope of the signal. Our results are different from the previous reports [5;6]. Here we found the “shape”, or the low frequency part of the signal rather than the “change”, or the high frequency part had gained more credits in classification.

The fit statistics of the experimental results were calculated through the method of the confusion matrix and the index κ . This method was first used in psychological research to study

the agreement between two judgments on the same questions made by two independent subjects [28]. The advantage of this method is that no statistical distribution of the samples is required as, for example, is the case in Student t-test. The value for κ fluctuates between 0 and 1, where zero indicates that the coincidence of the two judgments is only by random chance and one indicates that the agreement is met without chance. The value for the SF classification ($\kappa = 0.6$), indicates that the correlation is significantly higher than chance. In contrast, that κ was zero for the CSF classification indicates no correlation. These results demonstrate that the FCMC-AW algorithm is consistent with expected outcomes.

The sample size was limited by the number of recruits available in our 18 month window of opportunity. We adapted the validation techniques to the small sample sizes. It should be pointed out that although the sample size in the project was small, the technology of AW showed great potential in IR analysis. To our knowledge, this is the first time that the AW has been used in IR analysis. The algorithm has the impartial ability to search features in the whole frequency spectrum, including both low and high frequency bands.

V. CONCLUSION

The APR system described in this paper has been proposed on the basis of the fingerprint characteristic of the infrared spectrum, and under the hypothesis that the information regarding each variant of arthritis is reflected in a distinguishable spectral pattern. The fuzzy C-means cluster adaptive wavelet (FCMC-AW) algorithm employed consists of two parts: one is an adaptive wavelet based M -band feature extractor, and the other is a Bayesian classifier using the information from the membership matrix generated during fuzzy C-means clustering.

The spectral simulations emphasized two characteristics of the FCMC-AW algorithm. The first was to show that the algorithm can provide a measure of confidence of classification which may have practical significance for physicians. The second demonstrated that adaptive wavelets were more efficient classifiers than predefined wavelets. The simulation also provided an elementary approach to determining starting parameters for the algorithm.

The biological experiments showed that the value of M_{FQ} was a more favorable index than the value of correct rate (CR) alone, especially in the case of a small sample size. The experiments also demonstrated that the 'shape' or the low frequency components of the spectrum gained more credits in classification, a finding that contrasts with earlier reports (e.g. [6]). In addition, the algorithm could produce additional information besides CR by the value of M_{FQ} and the entries of the membership matrix, reflecting the classification confidence of a new pattern when it was assigned to one of the predefined classes.

Clearly, the sample size was small, representing the lower limit of practical application. As a result, we are careful not to

draw conclusions that relate to pathology. Our interest is in the development of a tool to assist in the diagnosis and management of this debilitating disease. There exists no simple method to evaluate joint fluid. Often joint swelling is not accompanied by radiological findings of joint degradation. Developing an indirect measure to assist in the early detection of degenerative disease could assist in staging the disease.

ACKNOWLEDGMENT

The authors thank the medical staff of the Royal University Hospital for assistance in biological sample collection.

REFERENCES

- [1] W. J. Koopman, *Arthritis and allied conditions: A textbook of rheumatology*, Baltimore: Williams & Wilkins, 1997.
- [2] H. Visser, S. le Cessie, K. Vos, F. C. Breedveld, and J. M. W. Hazes, "How to diagnose rheumatoid arthritis early - A prediction model for persistent (erosive) arthritis," *Arthritis Rheum.*, vol. 46, pp. 357-365, Feb. 2002.
- [3] C. R. Kjeldsberg and J. A. Knight, "Synovial fluid," in *Body fluids, laboratory examination of cerebrospinal, synovial, and serous fluids: A textbook atlas*, Chicago: Educational Products Division, American Society of Clinical Pathologists, 1982, pp. 48-60.
- [4] A. Swan, H. Amer, and P. Dieppe, "The value of synovial fluid assays in the diagnosis of joint disease: a literature survey," *Ann. Rheum. Dis*, vol. 61, pp. 493-498, June 2002.
- [5] R. A. Shaw, S. Kotowich, H. H. Eysel, M. Jackson, G. T. D. Thomson, and H. H. Mantsch, "Arthritis diagnosis based upon the near infrared spectrum of synovial fluid," *Rheumatol. Int.*, vol. 15, pp. 159-165, 1995.
- [6] H. H. Eysel, M. Jackson, A. Nikulin, R. L. Somorjai, G. T. D. Thomson, and H. H. Mantsch, "A novel diagnostic test for arthritis: multivariate analysis of infrared spectra of synovial fluid," *Biospectroscopy*, vol. 3, pp. 161-167, 1997.
- [7] J. M. G. Canvin, S. Bernatsky, C. A. Hitchon, M. Jackson, M. G. Sowa, J. R. Mansfield, H. H. Eysel, H. H. Mantsch, and H. S. El-Gabalawy, "Infrared spectroscopy: shedding light on synovitis in patients with rheumatoid arthritis," *Rheumatology*, vol. 42, pp. 76-82, 2003.
- [8] B. Stuart, *Infrared spectroscopy: fundamentals and applications*, Chichester, West Sussex, England: J. Wiley, 2004.
- [9] M. Jackson and H. H. Mantsch, "FTIR spectroscopy in the clinical sciences," in *Advances in Spectroscopy: Biomedical Applications*, Chichester: John Wiley & Sons, Inc., 1996, pp. 185-251.
- [10] C. M. Ziegler, P. Kircher, and S. Hassfeld, "Analysis of temporomandibular joint synovial fluid using Fourier transform/infrared spectroscopy," *J. Oral Maxillofac. Surg.*, vol. 60, pp. 1302-1306, Nov. 2002.
- [11] J. C. Bezdek, *Pattern recognition with fuzzy objective function algorithms*, New York: Plenum Press, 1981.
- [12] C. K. Chui, *An introduction to wavelets*, Boston: Academic Press, 1992.
- [13] I. Daubechies, *Ten lectures on wavelets*, Philadelphia, PA: Society for Industrial and Applied Mathematics, 1992.
- [14] S. G. Mallat, *A wavelet tour of signal processing*, 2nd ed., San Diego: Academic Press, 1999.
- [15] J. Kautsky and R. Turcajova, "Pollen product factorization and construction of higher multiplicity wavelets," *Linear Algebra Appl.*, vol. 222, pp. 241-260, June 1995.
- [16] G. Strang and T. Nguyen, *Wavelets and filter banks*, Rev. ed., Wellesley, MA: Wellesley-Cambridge Press, 1997.
- [17] H. L. Resnikoff and R. O. Jr. Wells, *Wavelet analysis : the scalable structure of information*, New York: Springer-Verlag, Inc., 1998.
- [18] Y. Mallet, D. Coomans, J. Kautsky, and O. DeVel, "Classification using adaptive wavelets for feature extraction," *IEEE Trans. Pattern Anal. Mach. Intel.*, vol. 19, pp. 1058-1066, Oct. 1997.

- [19] H. Brunzell and J. Eriksson, "Feature reduction for classification of multidimensional data," *Pattern Recogn.*, vol. 33, pp. 1741-1748, Oct. 2000.
- [20] J. Cui, "Automated search for patterns of disease in infrared spectra of body fluids," M.S. thesis, Univ. Saskatchewan, Saskatoon, Canada, 2002.
- [21] J. C. Bezdek, *Analysis of fuzzy information*, Boca Raton, Fla: CRC Press, 1987.
- [22] D. Coomans and L. Broeckaert, *Potential pattern recognition in chemical and medical decision making*, Letchworth, Hertfordshire, England: Research Studies Press, John Wiley and Sons, 1986.
- [23] H. L. Van Trees, *Detection, estimation, and modulation theory*, New York: Wiley, 2001.
- [24] A. Papoulis and S. U. Pillai, *Probability, random variables, and stochastic processes*, 4th ed., Boston: McGraw-Hill, 2002.
- [25] S. Aeberhard, D. Coomans, and O. DeVel, "Comparative-Analysis of Statistical Pattern-Recognition Methods in High-Dimensional Settings," *Pattern Recogn.*, vol. 27, pp. 1065-1077, Aug. 1994.
- [26] J. D. F. Habbema, J. Hilden, and B. Bjerregaard, "Measurement of performance in probabilistic diagnosis .1. problem, descriptive tools, and measures based on classification matrices," *Methods Inf. Med.*, vol. 17, pp. 217-226, 1978.
- [27] G. C. Cawley and N. L. C. Talbot, "Efficient leave-one-out cross-validation of kernel Fisher discriminant classifiers," *Pattern Recogn.*, vol. 36, pp. 2585-2592, 2003.
- [28] D. A. Bloch and H. C. Kraemer, " 2×2 Kappa-coefficients - measures of agreement or association," *Biometrics*, vol. 45, pp. 269-287, Mar. 1989.
- [29] J. F. Reed, "Homogeneity of kappa statistics in multiple samples," *Comput. Meth. Prog. Bio.*, vol. 63, pp. 43-46, Aug. 2000.
- [30] G. Sarty and E. Kendall, "Self-diffusion maps from wavelet de-noised NMR images," *J. Magn. Reson. Series B*, vol. 111, pp. 50-60, Apr. 1996.