J Sign Process Syst
DOI 10.1007/s11265-008-0180-z

Intravascular Ultrasound Tissue Characterization with Sub-class Error-Correcting Output Codes

Sergio Escalera · Oriol Pujol · Josepa Mauri · Petia Radeva

Received: 10 December 2007 / Revised: 14 March 2008 / Accepted: 3 April 2008

Abstract Intravascular ultrasound (IVUS) represents a powerful imaging technique to explore coronary vessels and to study their morphology and histologic properties. In this paper, we characterize different tissues based on radial frequency, texture-based, and combined features. To deal with the classification of multiple tissues, we require the use of robust multi-class learning techniques. In this sense, error-correcting output codes (ECOC) show to robustly combine binary classifiers to solve multi-class problems. In this context, we propose a strategy to model multi-class classification tasks using sub-classes information in the ECOC framework. The new strategy splits the classes into different sub-sets according to the applied base classifier. Complex IVUS data sets containing overlapping data are learnt by splitting the original set of classes into sub-classes, and embedding the binary problems in a problem-dependent ECOC design. The method automatically characterizes different tissues, showing performance improvements over the state-of-the-art ECOC techniques for different base classifiers. Furthermore, the combination of RF and texture-based features also shows improvements over the state-of-the-art approaches.

Keywords Intravascular ultrasound · Multi-class classification · Embedding of dichotomies · Sub-classes · Error-correcting output codes

S. Escalera (✉) · O. Pujol · P. Radeva
Centre de Visió per Computador, Campus UAB,
08193 Bellaterra Barcelona, Spain
e-mail: sergio@maia.ub.es

O. Pujol
e-mail: oriol@maia.ub.es

P. Radeva
e-mail: petia@maia.ub.es

S. Escalera · O. Pujol · P. Radeva
Department Matemàtica Aplicada i Anàlisi, UB,
Gran Via 585, 08007 Barcelona, Spain

J. Mauri
Hospital Universityari Germans Trias i Pujol,
Badalona, Spain
e-mail: jmauri.germanstrias@gencat.nett.es

1 Introduction

Cardiovascular diseases represents the first cause of sudden death in the occidental world [1]. Plaque rupture is one of the most frequent antecedent of coronary pathologies. Depending on the propensity to collapse, coronary plaque can be divided into stable and vulnerable plaque [2]. According to pathological studies, the main features of a stable plaque are characterized by the presence of a large lipid core with a thin fibrous cap. This last type of plaque can rupture generating thrombi followed by an intimal hyperplasia. Therefore, an accurate detection and quantification of plaque types represents an important subject in the diagnosis in order to study the nature and the plaque evolution to predict its final effect.

One of the most widely used diagnostic procedures consists of screening the coronary vessels employing intravascular ultrasound imaging (IVUS). This technique yields a detailed cross-sectional image of the vessel allowing coronary arteries and their morphology to be extensively explored. This image modality has become one of the principal tools to detect coronary plaque.

An IVUS study consists of introducing a catheter which shots a given number of ultrasound beams and collect their echoes to form an image. According with these echoes, three distinguishable plaques are considered in this type of images: calcified tissue (characterized by a very high echo-reflectivity and absorbtion of the ultrasound signal), fibrous plaque (medium echo-reflectivity and good transmission coefficient), and lipidic or soft plaque (characterized with very low reflectance of the ultrasound signal).

Despite the high importance of studying the whole coronary vessel, in clinical practice, this plaque characterization is performed manually in isolated images. Moreover, due to the variability among different observers, a precise manual characterization becomes very difficult to perform. Therefore, automatic analysis of IVUS images represents a feasible way to predict and quantify the plaque composition, avoiding the subjectivity of manual region classification and diminishing the characterization time in large sequences of images.

One of the main problems of the automatic tissue classification is the high variability in appearance of the same plaque in the images. This is mainly caused by the different image acquisition conditions of IVUS sequences. In order to enhance different IVUS regions, physicians typically change the image parameter set such as depth or region-based contrast of the imaging equipment. Furthermore, once the IVUS images are recorded, normalizing them with a unique parameter set (gain and offset) becomes impossible due to the loss of information and the non-linear transformation of the data in the image formation process. This lack of normalization hinders the automatic classification making coronary tissues from different case studies non comparable.

Given its clinical importance, automatic plaque classification in IVUS images has been considered in several research studies. The process can be divided in two stages, plaque characterization step which consist in extracting characteristic features in order to describe each tissue, and a classification step where a learning technique is used to train a classifier. In the first stage there are mainly two basic strategies: image-based approaches [19–21], and radio frequency (RF) signal analysis [22–24]. The main advantage of image-based methods is the availability of the images since they are the standard data source of the equipment. Additionally there is a high variety of descriptors which capture the spatial information of gray level values of a pixel together with its neighborhood in the image. However, this source suffers from a loss of information and the introduction of artifacts due to the reconstruction process. In Pujol et al. [19], authors propose to compute texture measurements such as local binary patterns to characterize the image, followed by the classification based on the AdaBoost learning technique with decision stumps. Korte et al. [21] use the difference of gray scale densities in the image with the k-means to classify fibrous and lipid plaque. Zhang et al. [20] use the gray scale values by means of co-occurrence matrices in order to segment and classify regions of different plaques, with the nearest neighbor learning technique. One of the main problems of these techniques is that they do not take into account the plaque size variation and they suffer from high variability among non-normalized DICOM sequences.

Characterization of RF signal has been proposed in [23] and [22] to take advantage of the of the raw IVUS signals. This data source represents a better the ultrasound data avoiding the introduction of artifacts from the pixel interpolation in the process of image formation. Due to the higher resolution of the unprocessed data, small regions of plaque could be distinguished. The approaches using this technique are: Kawasaki et al. [23] calculate the power of the acquired RF signal in decibels. On the other hand, Nair et al. in [22] obtain the power spectrum of a signal window using AR models and extract 8 measures from it (as maximum power, positive and negative slope, among others). However, this kind of data is not easily available and its processing is usually restricted to local spots [23]. In addition, the spatial information is lost when spectral measurements are calculated [22]. The preferred learning strategy for the plaque recognition is based on classification trees of the extracted features [22, 23].

In this paper, we base on texture-based features, RF signals, and combined features to characterize the different types of tissues. For the learning step, we focus on error-correcting output codes as a general framework to combine binary classifiers to deal with multi-class categorization problems. ECOC were born as a general framework to combine binary problems to address the multi-class problem. The strategy was introduced by Dietterich and Bakiri [3] in 1995. Based on the error correcting principles [3], ECOC has been successfully applied to a wide range of applications, such as face recognition [4], face verification [5], text recognition [6] or manuscript digit classification [7]. In this context, we propose a novel technique to address multi-class classification problems by means of error-correcting output codes. The new methodology is based on the splitting of the original set of tissue classes into different sub-tissues so that the base classifier applied is able to learn the data. In this sense, complex data sets containing overlapping data can be modelled by splitting the original set of classes into sub-classes and

embedding the binary problems in a problem-dependent ECOC design. The method shows to automatically characterize different tissues using different feature sets with high performance, obtaining significant performance improvements compared to previous state-of-the-art ECOC strategies for different base classifiers, such as discriminant analysis, AdaBoost, or support vector machines.

The paper is organized as follows: Section 2 overviews the ECOC framework and presents the new sub-class approach. Section 3 explains the acquisition of tissue features, and Section 4 shows the experimental results. Finally, Section 5 concludes the paper.

2 Problem-Dependent ECOC Sub-class

The ECOC technique can be broken down into two distinct stages: encoding and decoding. Given a set of classes, the coding stage designs a codeword[1] for each class based on different binary problems. The decoding stage makes a classification decision for a given test sample based on the value of the output code.

2.1 Error-Correcting Output Codes

Given a set of N_c classes to be learned, at the coding step of the ECOC framework, n different bi-partitions (groups of classes) are formed, and n binary problems (dichotomies) are trained. As a result, a codeword of length n is obtained for each class, where each bin of the code corresponds to a response of a given dichotomy. Arranging the codewords as rows of a matrix, we define a "coding matrix" M, where $M \in \{-1, 0, 1\}^{N_c \times n}$ in the ternary case. Joining classes in sets, each dichotomy is coded by $\{+1, -1\}$ according to their class set membership, or 0 if the class is not considered by the dichotomy. In Fig. 1 we show an example of a one-versus-one coding matrix M, where each dichotomy learns to split a pair of classes. The matrix is coded using 6 dichotomies $\{h_1, ..., h_6\}$ for a four-class problem (c_1, c_2, c_3, and c_4). The white regions are coded by 1 (considered as positive for its respective dichotomy, h_i), the dark regions by -1 (considered as negative), and the grey regions correspond to the zero symbol (not considered classes by the current dichotomy). For example, the first classifier (h_1) is trained to discriminate c_1 versus c_2 ignoring c_3 and c_4, the second one classifies c_1 versus c_3 ignoring c_2 and c_4, etc.

[1]The codeword is a sequence of bits of a code representing each class, where each bit identifies the membership of the class for a given binary classifier.

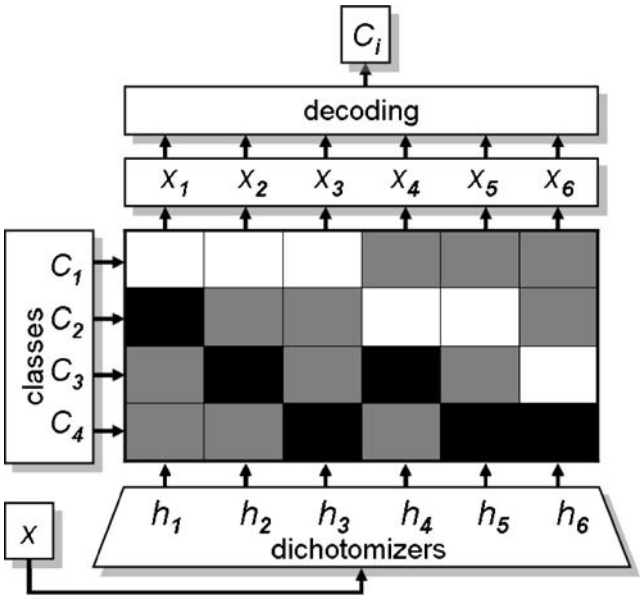

Figure 1 Example of ternary matrix M for a four-class problem. A new test codeword is classified using a decoding strategy.

During the decoding process, applying the n trained binary classifiers, a code x is obtained for each data point in the test set. This code is compared with the base codewords of each class $\{y_1, ..., y_4\}$ defined in the matrix M. And the data point is assigned to the class with the "closest" codeword [8, 9]. Although different distances can be applied, the Hamming (HD) and the Euclidean distances (ED) are the most frequently used. In Fig. 1, a new test input x is evaluated by all the classifiers and the method assigns label c_i with the closest decoding measure.

2.2 ECOC Sub-class

From an initial set of classes C of a given multi-class problem, the objective of the Sub-class ECOC strategy is to define a new set of classes C', where $|C'| > |C|$, so that the new set of binary problems is easier to learn for a given base classifier. For this purpose, we use a guided procedure that, in a problem-dependent way, groups classes and splits them into sub-sets if necessary.

Recently, the authors of [10] proposed a ternary problem-dependent design of ECOC, called discriminant ECOC (DECOC), where given N classes, a high classification performance is achieved with only $N - 1$ binary problems. The method is based on the embedding of discriminant tree structures derived from the problem domain. The binary trees are built by looking for the partition that maximizes the mutual information (MI) between the data and their respective class labels. Look at the three-class problem shown on the top of

Fig. 2a. The standard DECOC algorithm considers the whole set of classes to split it into two sub-sets of classes $\wp^+$ and $\wp^-$ maximizing the *MI* criterion on a sequential forward floating search procedure (*SFFS*). In the example, the first sub-sets found correspond to $\wp^+ = \{C_1, C_2\}$ and $\wp^- = \{C_3\}$. Then, a base classifier is used to train its corresponding dichotomizer h_1. This classifier is shown in the node h_1 of the tree structure shown in Fig. 2d. The procedure is repeated until all classes are split into separate sub-sets $\wp$. In the example, the second classifier is trained to split the sub-sets of classes $\wp^+ = C_1$ from $\wp^- = C_2$ because the classes C_1 and C_2 were still contained in a single sub-set after the first step. This second classifier is codified by the node h_2 of Fig. 2d. When the tree is constructed, the coding matrix M is obtained by codifying each internal node of the tree as a column of the coding matrix (see Fig. 2c).

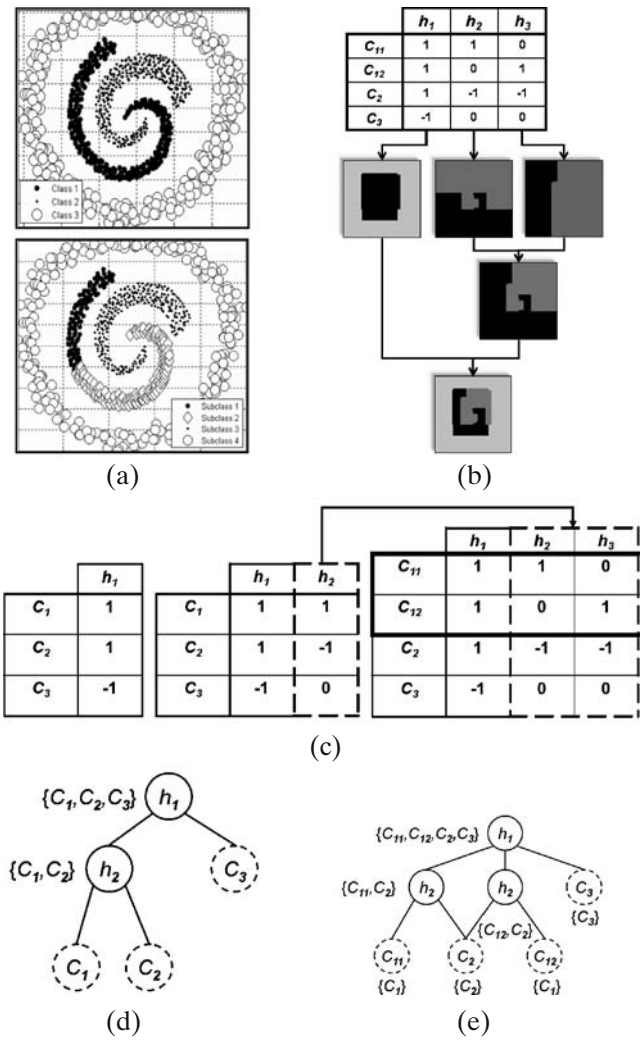

Figure 2 **a** *Top* original three-class problem. *Bottom* four sub-classes found. **b** Sub-class ECOC encoding using the four sub-classes using discrete AdaBoost with 40 runs of decision stumps. **c** Learning evolution of the sub-class matrix M. **d** Original tree structure without applying sub-class. **e** New tree-based configuration using sub-classes.

In our case, sequential forward floating search (*SFFS*) is also applied to look for the sub-sets $\wp^+$ and $\wp^-$ that maximizes the mutual information between the data and their respective class labels [10]. The encoding algorithm is shown in Table 1. Given a N-class problem, the whole set of classes is used to initialize the set L containing the sets of labels for the classes to be learned. At the beginning of each iteration k of the algorithm (*Step 1*), the first element of L is assigned to S_k in the first step of the algorithm. Next, $SFFS$ is used to find the optimal binary partition BP of S_k that maximizes the mutual information I between the data and their respective class labels (*Step 2*). The *SFFS* algorithm used is the one presented in [11], and the implementation details of the fast quadratic mutual information can be found in [10].

To illustrate our procedure, let us return to the example of the top of Fig. 2a. On the first iteration of the sub-class ECOC algorithm, $SFFS$ finds the sub-set $\wp^+ = \{C_1, C_2\}$ against $\wp^- = \{C_3\}$. The encoding of this problem is shown in the first matrix of Fig. 2c. The positions of the column corresponding to the classes of the first partition are coded by +1 and the classes corresponding to the second partition to -1, respectively. In our procedure, the base classifier is used to test if the performance obtained by the trained dichotomizers is sufficient. Observe the decision boundaries of the picture next to the first column of the matrix in Fig. 2b. One can see that the base classifier finds a good solution for this first problem.

Then, the second classifier is trained to split $\wp^+ = C_1$ against $\wp^- = C_2$, and its performance is computed. The separation of the current sub-sets is not a trivial problem, and the classification performance is poor. Therefore, our procedure tries to split the data $J_{\wp^+}$ and $J_{\wp^-}$ from the current sub-sets $\wp^+$ and $\wp^-$ into more simple sub-sets. At *Step 3* of the algorithm, the splitting criteria SC takes as input a data set $J_{\wp^+}$ or $J_{\wp^-}$ from a sub-set $\wp^+$ or $\wp^-$, and splits it into two sub-sets $J^+_{\wp^+}$ and $J^-_{\wp^+}$ or $J^+_{\wp^-}$ and $J^-_{\wp^-}$. On the experimental results chapter we discuss the selection of the splitting criterion. The splitting algorithm is shown in Table 2.

When two data sub-sets $\{J^+_{\wp^+}, J^-_{\wp^+}\}$ and $\{J^+_{\wp^-}, J^-_{\wp^-}\}$ are obtained, only one of both split sub-sets is used. We select the sub-sets that have the highest distance between the means of each cluster. Suppose that the distance between $J^+_{\wp^-}$ and $J^-_{\wp^-}$ is larger than between $J^+_{\wp^+}$ and $J^-_{\wp^+}$. Then, only $J_{\wp^+}$, $J^+_{\wp^-}$, and $J^-_{\wp^-}$ are used. If the new sub-sets improve the classification performance, new sub-classes are formed, and the process is repeated.

Table 1 Problem-dependent Sub-class ECOC algorithm.

Inputs: J, C, $\theta = \{\theta_{size}, \theta_{perf}, \theta_{impr}\}$ //Thresholds for the number of samples, performance, and improvement between iterations

Outputs: C', J', $\wp'$, M

[Initialization:]
Create the trivial partition $\{\wp_0^+, \wp_0^-\}$ of the set of classes $\{C_i\}$: $\{\wp_0^+, \wp_0^-\}$ $= \{\{\emptyset\}, \{C_1, C_2, ..., C_N\}\}$ $L_0 = \{\wp_0^-\}$;$J' = J$;$C' = C$;$\wp' = \emptyset$;$M = \emptyset$;$k = 1$

Step 1 S_k is the first element of L_{k-1}
$L'_k = L_{k-1} \backslash \{S_k\}$

Step 2 Find the optimal binary partition $BP(S_k)$:
$\{\wp_k^+, \wp_k^-\} = argmax_{BP(S_k)}(I(\mathbf{x}, d(BP(S_k))))$
where I is the mutual information criterion, $\mathbf{x}$ is the random variable associated to the features and d is the discrete random variable of the dichotomy labels[a], defined in the following terms,

$$d = d(\mathbf{x}, BP(S_k)) = \begin{cases} 1 & \text{if } \mathbf{x} \in C_i | C_i \in \wp_k^+ \\ -1 & \text{if } \mathbf{x} \in C_i | C_i \in \wp_k^- \end{cases}$$

Step 3 // Look for sub-classes
$\{C', J', \wp'\} = SPLIT\left(J_{p_k^+}, J_{p_k^-}, C', J', J, \wp', \theta\right)$[b]

Step 4 $L_k = \{L'_k \cup \wp_k^i\}$ if $|\wp_k^i| > 1$ $\forall i \in \{+, -\}$

Step 5 If $|L_k| \neq 0$
$k = k + 1$ **go to Step 1**

Step 6 Codify the coding matrix M using each partition $\{\wp_i^+, \wp_i^-\}$ of $\wp'$, $i \in [1, .., |\wp'|]$ and each class $C_r \in \wp_i = \{\wp_i^+ \cup \wp_i^-\}$ as follows:

$$M(C_r, i) = \begin{cases} 0 & \text{if } C_r \notin \wp_i \\ +1 & \text{if } C_r \in \wp_i^+ \\ -1 & \text{if } C_r \in \wp_i^- \end{cases} \quad (1)$$

[a]Use $SFFS$ of [11] as the maximization procedure and MI of [10] to estimate I.
[b]Using the splitting algorithm of Table 2.

In the example of Fig. 2, applying the splitting criteria SC over the two sub-sets, two clusters are found for $\wp^+ = C_1$ and for $\wp^- = C_2$. Then, the original encoding of the problem C_1 vs C_2 (corresponding to the second column of the matrix in the center of Fig. 2c) is split into two columns marked with the solid lines in the matrix on the right. In this way, the original C_1 vs C_2 problem is transformed to two more simple problems $\{C_{11}\}$ against $\{C_2\}$ and $\{C_{12}\}$ against $\{C_2\}$. Here the first subindex of the class corresponds to the original class, and the second subindex to the number of sub-class. It implies that the class C_1 is split into two sub-classes (look at the bottom of Fig. 2a), and the original three-class problem $C = \{C_1, C_2, C_3\}$ becomes the four-sub-class problem $C' = \{C_{11}, C_{12}, C_2, C_3\}$. As the class C_1 has been decomposed by the splitting of the second problem, we need to save the information of the current sub-sets and the previous sub-sets affected by the new splitting. The steps to update this information are summarized in the *Step 4* of the splitting algorithm. We use the object labels to define the set of sub-classes of the current partition $\wp_c$. If new sub-classes are created, the set of sub-classes C' and the data for sub-classes J' have to be updated. Note that when a class or a sub-class previously considered for a given binary problem is split in a future iteration of the procedure, the labels from the previous sub-sets $\{\wp^+, \wp^-\}$ need to be updated with the new information. Finally, the set of labels for the binary problems $\wp'$ is updated with the labels of the current sub-set $\wp' = \wp' \cup \wp_c$. In the example of Fig. 2, the dichotomizer h_1 considers the sub-sets $\wp_1^+ = \{C_1, C_2\}$ and $\wp_1^- = \{C_3\}$. Then, those positions containing class C_1 are replaced with C_{11} and C_{12}. The process is repeated until the desired performance is achieved or the stopping conditions are full-filled.

The conditions that guide the learning and splitting process are defined by the set of parameters $\theta = \{\theta_{\text{size}}, \theta_{\text{perf}}, \theta_{impr}\}$, where θ_{size} corresponds to the minimum size of a sub-set to be clustered, θ_{perf} contains the minimum error desired for each binary problem, and θ_{impr} looks for the improvement of the split sub-sets regarding the previous ones. The function $TEST\_PARAMETERS$ in Table 2 is responsible for testing the constraints based on the parameters $\{\theta_{\text{size}}, \theta_{\text{perf}}, \theta_{impr}\}$. If the constraints are satisfied, the new sub-sets are selected and used to recursively call the splitting

Table 2 Sub-class *SPLIT* algorithm.

Inputs: $J_{\wp^1}, J_{\wp^2}, C', J', J, \wp', \theta$ // C' is the final set of classes, J' the data for the final set of classes, and $\wp'$ is the labels for all the partitions of classes of the final set.

Outputs: $C', J', \wp'$

Step 1 Split problems:
$\{J^+_{\wp^+}, J^-_{\wp^+}\} = SC(J_{\wp^+})$[a]
$\{J^+_{\wp^-}, J^-_{\wp^-}\} = SC(J_{\wp^-})$

Step 2 Select sub-classes:
if $|\overline{J^+_{\wp^+}}, \overline{J^-_{\wp^+}}| > |\overline{J^+_{\wp^-}}, \overline{J^-_{\wp^-}}|$ // find the largest distance between the means of each sub-set.
$\{J^+_+, J^-_+\} = \{J^+_{\wp^+}, J_{\wp^-}\}$; $\{J^+_-, J^-_-\} = \{J^-_{\wp^+}, J_{\wp^-}\}$
else
$\{J^+_+, J^-_+\} = \{J^+_{\wp^-}, J_{\wp^+}\}$; $\{J^+_-, J^-_-\} = \{J^-_{\wp^-}, J_{\wp^+}\}$
end

Step 3 Test parameters to continue splitting:
if $TEST\_PARAMETERS(J_{\wp^1}, J_{\wp^2}, J^1_1, J^2_1, J^1_2, J^2_2, \theta)$// call the function with the new sub-sets
$\{C', J', \wp'\} = SPLIT(J^1_1, J^2_1, C', J', J, \wp', \theta)$
$\{C', J', \wp'\} = SPLIT(J^1_2, J^2_2, C', J', J, \wp', \theta)$
end

Step 4 Save the current partition:
Update the data for the new sub-classes and previous sub-classes if intersections exists J'.
Update the final number of sub-classes C'.
Create $\wp_c = \{\wp_{c^1}, \wp_{c^2}\}$ the set of labels of the current partition.
Update the labels of the previous partitions $\wp$.
Update the set of partitions labels with the new partition $\wp' = \wp' \cup \wp_c$.

[a] *SC* corresponds to the splitting method of the input data into two main clusters.

function (*Step 3* of the algorithm in Table 2). The constraints of the function $TEST\_PARAMETERS$ are fixed by default as follows:

- The number of objects in $J_{\wp^+}$ has to be larger than θ_{size}.
- The number of objects in $J_{\wp^-}$ has to be larger than θ_{size}.
- The error $\xi(h(J_{\wp^-}, J_{\wp^+}))$ obtained from the dichomomizer h using a particular base classifier applied on the sets $\{\wp^+, \wp^-\}$ has to be larger than θ_{perf}.
- The sum of the well-classified objects from the two new problems (based on the confusion matrices) divided by the total number of objects has to be greater than $1 - \theta_{impr}$.

θ_{size} avoids the learning of very unbalanced problems. θ_{perf} determines when the performance of a partition of classes is insufficient and sub-classes are required. And finally, when a partition does not obtain the desired performance θ_{perf}, the splitting of the data stops, preventing overtraining.

In the example of Fig. 2, the three dichotomizers h_1, h_2, and h_3 find a solution for the problem (look the trained boundaries shown in Fig. 2b), obtaining a classification error under θ_{perf}, so, the process stops. Now, the original tree encoding of the DECOC design shown in Fig. 2d can be represented by the tree structure of Fig. 2e, where the original class associated to each sub-class is shown in the leaves.

When the final set of binary problems is obtained, its respective set of labels $\wp'$ is used to create the coding matrix M (Eq. 1). The outputs C' and J' contain the final set of sub-classes and the new data for each sub-class, respectively. Finally, to decode the new sub-class problem-dependent design of ECOC, we take advantage of the recently proposed Loss-Weighted decoding design [12]. The decoding strategy uses a set of normalized probabilities based on the performance of the base classifier and the ternary ECOC constraints [12]. The decoding algorithm is described in [12].

2.3 Illustration Over Toy Problems

To show the effect of the sub-class ECOC strategy for different base classifiers, we used the previous toy problem of the top of Fig. 2a. Five different base classifiers are applied: Fisher linear discriminant analysis (*FLDA*), discrete AdaBoost, nearest mean classifier, linear *SVM*, and *SVM* with radial basis function kernel.[2] Using these base classifiers on the toy problem,

[2] The parameters of the base classifiers are explained in the experimental results section.

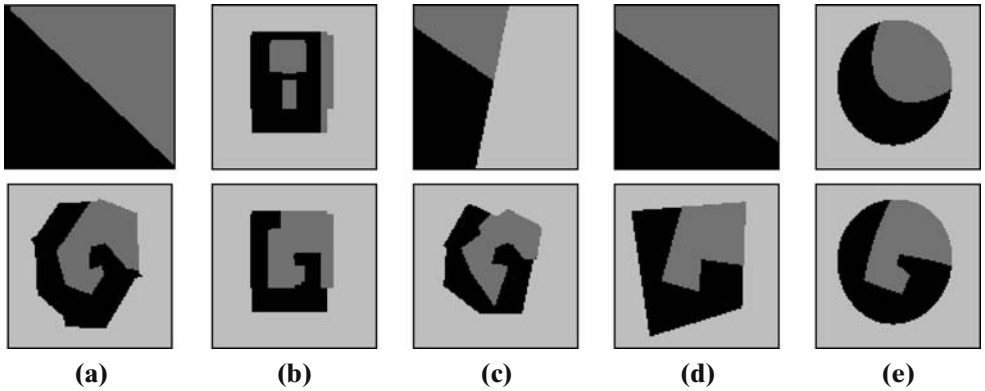

Figure 3 Sub-class ECOC without sub-classes (*top*) and including sub-classes (*bottom*): for *FLDA* (**a**), discrete AdaBoost (**b**), *NMC* (**c**), linear *SVM* (**d**), and *RBF SVM* (**e**).

the original DECOC strategy with the loss-weighted algorithm obtains the decision boundaries shown on the top row of Fig. 3. The new learned boundaries are shown on the bottom row of Fig. 3 for fixed parameters θ. Depending on the flexibility of the base classifier more sub-classes are required, and thus, more binary problems. Observe that all base classifiers are able to find a solution for the problem, although with different types of decision boundaries.

3 Feature Extraction

In this paper, we consider three types of features, the first ones obtained from RF signals, the second ones based on texture-based features from reconstructed images, and finally, the slope-based features proposed in [22].

3.1 RF Features

In order to analyze ultrasound images, the RF signals are acquired from the IVUS equipment with a sampling rate of at least two times the transducer frequency, and filtered using a band-pass filter with 50% gain centered at the transducer frequency [13]. Then, an exponential time gain compensation (TGC) is applied [13]. Once the RF signals have been acquired, filtered and exponentially compensated by the TGC, the power spectrum is obtained. Nair et al. in [22] show the modelling of the power spectrum using autoregressive models (ARM) as one of the most suitable and stable methods to analyze ultrasound signals [22]. It also represents an alternative to the Fourier transform since the ARM have been proved to be more stable when small signal windows are considered.

The ARM are defined as a linear prediction equation where the output x at a certain point t for each A-line is equal to a linear combination of its p previous outputs weighted by a set of parameters a_p [25]:

$$x(t) = \sum_{k=1}^{p} a_p(k)x(t-k),$$

where p is the ARM degree and the coefficients a_p are calculated minimizing the error of the modelled spectrum with respect to the original using the Akaike's error prediction criterium [25].

A sliding window is formed by n samples and m contiguous A-lines with a displacement of $n/4$ samples and $m/3$ A-lines in order to obtain an average AR model of a region. Only one side of the obtained spectrum is used because of its symmetrical properties. This spectrum is composed of h sampled frequencies ranging from 0 to $f_s/2$ [25].

In addition to the spectrum, two global measures are computed: the energy of the A-line and the energy of the window spectrum. All these features are compiled into a unique vector of $h+2$ dimensions which is used as a feature vector in the classification process.

3.2 Texture Features Extraction

Given that different plaques can be discriminated as regions with different grey-level distributions, it is a natural decision to use texture descriptors. In the bibliography, one can find a wide set of texture descriptors and up to our knowledge there are no optimal texture descriptors for image analysis in the general case. Our strategy is instead of trying to find out the optimal texture descriptor for our problem to gather several families of descriptors and apply multiple classifiers able to learn and extract the optimal features for the concrete problem.

Therefore, we employ three different texture descriptors: co-occurrence Matrix [26], local binary patterns [27] and Gabor filters [28, 29]. Additionally, taking into account that highly non-echogenic plaques

produce significant shade in the radial direction of the vessel, we include in the feature set the presence of shading in the image as a complementary feature.

The co-occurrence matrix is defined as the estimation of the joint probability density function of gray level pairs in an image [26]. The sum of all element values is:

$$P(i, j, D, \theta) = P(I(l, m) = i \otimes I(l + Dcos(\theta), m + Dsin(\theta)) = j),$$

where $I(l, m)$ is the gray value at pixel (l, m), D is the distance among pixels and θ is the angle between neighbors. We have established the orientation θ to be $[0^o, 45^o, 90^o, 135^o]$ [26, 31]. After computing this matrix, energy, entropy, inverse difference moment, shade, inertia and prominence measures are extracted [26].

Local binary patterns (LBP) are used to detect uniform texture patterns in circular neighborhoods with any quantization of angular space and spatial resolution [27]. LBP are based on a circular symmetric neighborhood of P members with radius R. To achieve gray level invariance, the central pixel g_c is subtracted to each neighbor g_p, assigning the value 1 to the result if the difference is positive and 0, otherwise. LBPs are defined as follows:

$$LBP_{R,P} = \sum_{p=0}^{P} a(g_p - g_c) \cdot 2^p$$

A Gabor filter is a special case of wavelets [28] which is essentially a Gaussian modulated by a complex sinusoid s. In 2D, it has the following form in the spatial domain:

$$h(x, y) = \frac{1}{2\pi\sigma^2} \exp\left\{-\frac{1}{2}\left[\left(\frac{x^2 + y^2}{\sigma^2}\right)\right]\right\} \cdot s(x, y)$$

$$s(x, y) = \exp[-i2\pi(Ux + Vy)] \qquad \phi = \arctan V/U$$

where σ is the standard deviation, U and V represent the 2D frequency of the complex sinusoid, and ϕ is the angle of the frequency.

According to [30], one of the main differences in the appearance of calcified tissue compared to the rest of tissue types is the shadow which is appreciated behind it. In order to detect this shadow, we perform an accumulative mean of the pixels gray values on the polar image from a pixel to the end of the column (the maximal depth considered). As a result of extracting the texture descriptors, we construct an n-dimensional feature vector where $n = k + l + m + 1$, k is the number of co-occurrence matrix measurements, l is the number of Gabor filters, m is the number of LPB and the last feature is the measure of the "shadow" in the image.

3.3 Data Set

To generate the data sets we used the RF signals and their reconstructed images from a set of 10 different patients with left descent artery pullbacks acquired in Hospital "German Trias i Pujol" from Barcelona, Spain. All these pullbacks contain the three classes of plaque. For each one, 10 to 15 different vessel sections were selected to be analyzed. Two physicians independently segmented 50 areas of interest per pullback. From these segmentations we took 15 regions of interest (ROI) of tissue per study randomly making a total of 5,000 evaluation ROIs. To build the data set, these selections were mapped in both RF signals and reconstructed images. In order to reduce the variability among different observers, the regions where both cardiologist agreed have been taken under consideration. Some samples from the data set are shown on the left of Fig. 4.

To generate the data set on texture features, the intersection between segmented images is mapped into a feature vector. Then, all the features collected are categorized by patient and each of the three possible plaques type. The image features are extracted by using the previous texture descriptors: co-occurrence matrix, local binary patterns, and Gabor filters. Those features are calculated for each pixel and gathered in a feature vector of 68 dimensions. An example of a manual and automatic texture-based segmentation for the same sample is shown on the right of Fig. 4.

To generate the data set of RF features, the RF signals have been acquired using a 12-bit acquisition card with a sampling rate of $f_s = 200$ Mhz. The IVUS equipment used is Galaxy II from Boston Scientific

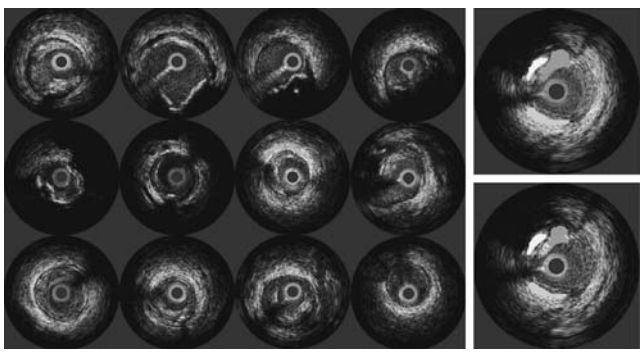

Figure 4 Left: IVUS data set samples. Right: (*top*) segmentation by a physician and (*down*) Automatic classification with texture-based features. The *white area* corresponds to calcium, the *light gray* area to fibrosis, and the *dark gray* area to soft plaque.

Table 3 Mean rank for each feature set.

Feature set	RF	Texture-based	RF+Texture-based	Slopes
Mean rank	2.94	2.28	1.72	2.83

with a catheter transducer frequency of $f = 40$ Mhz, and it is assumed a sound speed in tissue of 1565 m/s. Each IVUS image consists of a total of 256 A-lines (ultrasound beams), with a radial distance of $r = 0.65$ cm. The attenuation in tissue factor used is $\alpha = 1 Db/Mhz \times cm$. To analyze the RF signals, the sliding window is composed of $n = 64$ samples of depth and $m = 12$ radial A-lines, and the displacement is fixed in 16 samples and four A-lines. The power spectrum of the window ranges from 0 to 100 m/s and it is sampled by 100 points. Then, it is complemented with two energy measures yielding a 102 feature vector.

We also consider a third data set that concatenates the descriptors from the previous RF and texture-based features, obtaining a feature vector of length 170 features.

3.4 Slope-Based Features

Finally, the fourth data set considers the slope-based features proposed by in [22]. In particular, each sample is characterized by means of 14 slope-based features corresponding to: maximum power in DB from 20 to 60 MHz, frequency at the maximum power, negative slope in db/MHz between maximum and 60, minimum power in that slope, frequency corresponding to this negative slope, the estimated y intercept of this slope, the positive slope in db/Mhz between 20 and maximum, minimum power in that slope, frequency corresponding to this negative slope, the estimated y intercept of this slope, the mean power, the power at 0 MHz, power Db at 100 Mhz, and the power at the midband frequency (40 MHz) in DB [22].

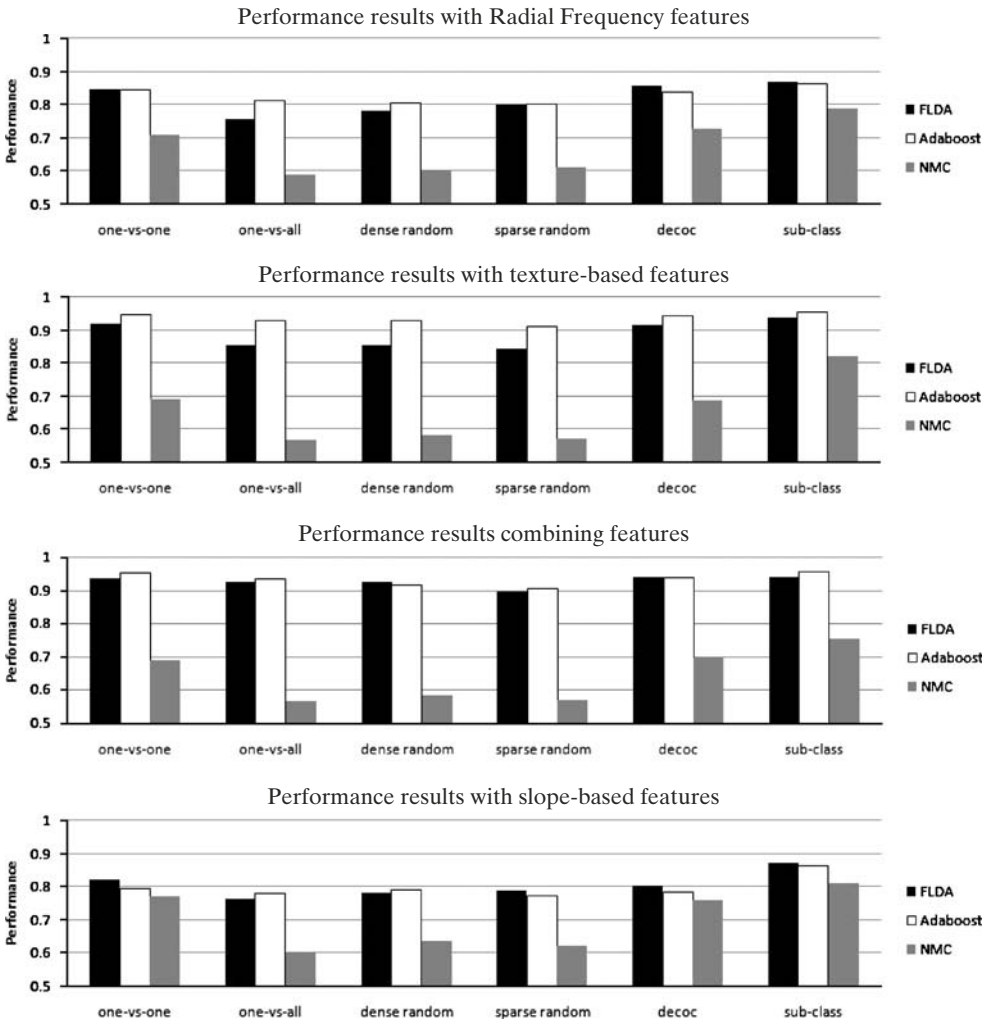

Figure 5 Performance results for different sets of features, ECOC designs and base classifiers on the IVUS data set.

Table 4 Mean rank for each ECOC design over all the experiments.

ECOC design	One-versus-one	One-versus-all	Dense random
Mean rank	2.33	5.08	4.25
ECOC design	Sparse random	Decoc	Sub-class
Mean rank	5.00	2.67	1.00

4 Results

Before the experimental results are presented, we comment the data, methods, and evaluation measurements.

- *Data* The data used for the experiments corresponds to the four data sets described at the previous section: RF, texture-based, combined RF and texture-based features data sets, and slope-based features.
- *Methods* We compare our method with the state-of-the-art ECOC coding designs: one-versus-one [14], one-versus-all [15], dense random [16], sparse random [16], and DECOC [10]. Each strategy uses the previously mentioned Linear Loss-weighted decoding to evaluate their performances at identical conditions. Three different base classifiers are applied over each ECOC configuration: nearest mean classifier (*NMC*) with the classification decision using the Euclidean distance between the mean of the classes, discrete AdaBoost with 40 iterations of decision stumps [17], and linear discriminant analysis implementation of the PR Tools using the default values (Faculty of Applied Physics, Delft University of Technology, The Netherlands, http://www.prtools.org/).
- *Evaluation measurements* To measure the performance of the different experiments, we apply leave-one-patient-out evaluation. Moreover, we use the statistical Friedman and Nemenyi tests to look for statistical significance among the methods performances [18].

4.1 IVUS Tissue Characterization

Applying the three different base classifiers over the set of ECOC configurations, the performance results for RF features, texture-based features, combined RF and texture-based features, and slope-based features are shown in Fig. 5.

Comparing the results among the different data sets, one can see that the worst performances are obtained by the RF and slope-based features, which obtain very similar results for all the base classifiers and ECOC configurations. The texture-based features obtain in most cases results upon 90%. Finally, the data set of combined RF and texture-based features slightly outperform the results obtained by the texture-based feature, though the results do not significantly differ.[3] This behavior is summarize on Table 3, where the mean rank obtained by each feature set is shown. The rankings are obtained estimating each particular ranking r_i^j for each problem i and each feature set j, and computing the mean ranking R for each feature set as $R_j = \frac{1}{N}\sum_i r_i^j$, where N is the total number of problems (3 base classifiers $\times$ 6 ECOC designs). Note that the best ranking corresponds to the combined set of features, and that the individual feature set that obtains the best results correspond to texture-based.

Concerning the classification strategies, observing the obtained performances in Fig. 5, one can see that independently of the data set and the ECOC design applied, the Sub-class ECOC approach always attains the best results. To compare these performances, the mean rank of each ECOC design considering the twelve different experiments is shown in Table 4. In this case, the rankings are obtained estimating each particular ranking r_i^j for each problem i and each ECOC configuration j, and computing the mean ranking R for each ECOC design as $R_j = \frac{1}{N}\sum_i r_i^j$, where N is the total number of problems (three base classifiers $\times$ four data sets). One can see that the sub-class ECOC attains the best position for all experiments. To analyze if the difference between methods ranks are statistically significant, we apply the Friedman and Nemenyi tests. In order to reject the null hypothesis that the measured ranks differ from the mean rank, and that the ranks are affected by randomness in the results, we use the Friedman test. The Friedman statistic value is computed as follows:

$$X_F^2 = \frac{12N}{k(k+1)}\left[\sum_j R_j^2 - \frac{k(k+1)^2}{4}\right] \quad (2)$$

In our case, with $k = 6$ ECOC designs to compare, $X_F^2 = 30.71$. Since this value is undesirable

[3] Due to the high similitude among slope-based and RF features results, the combination of texture-based and slope-based features has been omitted.

conservative, Iman and Davenport proposed a corrected statistic:

$$F_F = \frac{(N-1)X_F^2}{N(k-1)-X_F^2} \tag{3}$$

Applying this correction we obtain $F_F = 11.53$. With six methods and twelve experiments, F_F is distributed according to the F distribution with 5 and 55 *df*. The critical value of $F(5, 55)$ for 0.05 is 2.40. As the value of F_F is higher than 2.45 we can reject the null hypothesis. One we have checked for the for the non-randomness of the results, we can perform a post hoc test to check if one of the techniques can be singled out. For this purpose we use the Nemenyi test—two techniques are significantly different if the corresponding average ranks differ by at least the critical difference value (CD):

$$\mathrm{CD} = q_\alpha \sqrt{\frac{k(k+1)}{6N}} \tag{4}$$

where q_α is based on the Studentized range statistic divided by $\sqrt{2}$. In our case, when comparing six methods with a confidence value $\alpha = 0.10$, $q_{0.10} = 1.44$. Substituting in Eq. 4, we obtain a critical difference value of 1.09. Since the difference of any technique rank with the Sub-class rank is higher than the CD, we can infer that the Sub-class approach is significantly better than the rest with a confidence of 90% in the present experiments.

5 Conclusions

In this paper, we characterized intravascular ultrasound tissues based on different types of features, such as radial frequency, texture-based features, and combined features. We presented a Sub-class approach of error-correcting output codes that splits the tissue classes into different sub-sets according to the applied base classifier. In this sense, complex IVUS data sets containing overlapping data are solved by splitting the original set of classes into sub-classes, and embedding the binary problems in a problem-dependent ECOC design. The method automatically characterizes different tissues, showing performance improvements over the state-of-the-art ECOC techniques for *FLDA*, Discrete AdaBoost, *NMC*, and Linear and *RBF* SVM. In particular, the results shows higher performance when using texture-based features compared with RF signals and slope-based features, and slights improvements when the sets of features are combined.

Acknowledgements This work has been supported in part by TIN2006-15308-C02 and FIS ref. PI061290.

References

1. World Health Organization (2006). World Health Organization Statistics. http://www.who.int/entity/healthinfo/statistics/ .
2. Burke, A. P., Farb, A., Malcom, G. T., Smialek, J., & Virmani, R. (1997). Coronary risk factors and plaque morphology in men with coronary disease who died suddently. *The New England Journal of Medicine, 336*(18), 1276–1281.
3. Dietterich, T., & Bakiri, G. (1995). Solving multiclass learning problems via error-correcting output codes. *Journal of Artificial Intelligence Research, 2*, 263–282.
4. Windeatt, T., & Ardeshir, G. (2003). Boosted ECOC ensembles for face recognition. *International Conference on Visual Information Engineering*, 165–168.
5. Kittler, J., Ghaderi, R., Windeatt, T., & Matas, J. (2001). Face verification using error correcting output codes. *CVPR, 1*, 755–760.
6. Ghani, R. (2001). Combining labeled and unlabeled data for text classification with a large number of categories. *International Conference on Data Mining*, 597–598.
7. Zhou, J., & Suen, C., (2005). Unconstrained numeral pair recognition using enhanced error correcting output coding: A holistic approach. *Proceedings in Conference on Document Analysis and Records, 1*, 484–488.
8. Allwein, E., Schapire, R., & Singer, Y. (2002). Reducing multiclass to binary: A unifying approach for margin classifiers. *JMLR, 1*, 113–141.
9. Windeatt, T., & Ghaderi, R. (2003). Coding and decoding for multiclass learning problems. *Information Fusion, 1*, 11–21.
10. Pujol, O., Radeva, P., & Vitrià, J. (2006). Discriminant ECOC: A heuristic method for application dependent design of error correcting output codes. *PAMI, 28*, 1001–1007.
11. Pudil, P., Ferri, F., Novovicova, J., & Kittler, J. (1994). Floating search methods for feature selection with nonmonotonic criterion functions. *ICPR*, 279–283.
12. Escalera, S., Pujol, O., & Radeva, P. (2008). Loss-weighted decoding for error-correcting output coding. *International Conference on Computer Vision Theory and Applications, 2*, 117–122.
13. Karla, C., Joel, B., Oriol, P., Salvatella, N., & Radeva, P. (2006). In-vivo IVUS tissue classification: A comparison between RF signal analysis and reconstructed images. *Progress in Pattern Recognition* (pp. 137–146). Berlin: Springer
14. Hastie, T., & Tibshirani, R. (1998). Classification by pairwise grouping. *NIPS, 26*, 451–471.
15. Nilsson, N. J. (1965). Learning machines, McGraw-Hill.
16. Allwein, E., Schapire, R., & Singer, Y. (2002). Reducing multiclass to binary: A unifying approach for margin classifiers. *JMLR, 1*, 113–141.
17. Friedman, J., Hastie, T., & Tibshirani, R. (1998). Additive logistic regression: A statistical view of boosting. *The annals of statistics, 38*, 337–374.
18. Demsar, J. (2006). Statistical comparisons of classifiers over multiple data sets. *Journal of Machine Learning Research, 7*, 1–30.
19. Pujol, O., Rosales, M., & Radeva, P. (2003). Intravascular ultrasound images vessel characterization using adaBoost. *Functional Imaging and Modelling of the Heart: Lecture Notes in Computer Science*, 242–251.
20. Zhang, X., McKay, C. R., & Sonka, M. (1998). Tissue characterization in intravascular ultrasound images. *IEEE Transactions on Medicine, 17*(A), 889–898.
21. de Korte, C. L., PasterKamp, G., van der Steen, A. F. W., & Woutman, H. A. (1999). Characterisation of plaque components with IVUS elastography. *IEEE Ultrasonics Symposium Proceedings*, 1645-1648.

22. Nair, A., Kuban, B. D., Obuchowski, N., & Vince, G. (2001). Assesing spectral algorithms to predict atherosclerotic plaque composition with normalized and raw intravascular ultrasound data. *Ultrasound in Medicine & Biology, 27*, 1319–1331.
23. Kawasaki, M. (2002). In vivo quantitative tissue characterization of human coronary arterial plaques by use of integrated backscatter intravascular ultrasound and comparison with angioscopic findings. *Circulation, 105*, 2487–2492.
24. Murashige, A., Hiro , T., Fujii, T., Imoto, K., Murata, T., Fukumoto, Y., et al. (2005). Detection of lipid-laden atherosclerotic plaque by wavelet analysis of radiofrequency intravascular ultrasound signals. *Journal of the American College of Cardiology, 45*(12), 1954–1960.
25. Proakis, J., Rader, C., Ling, F., & Nikias, C. (1992). Advanced digital signal processing. *Mc Millan*.
26. Ohanian, P., Dubes, R. (1992). Performance evaluation for four classes of textural features. *Pattern Recognition, 25*, 819–833.
27. Ojala, T., Pietikainen, M., & Maenpaa, T. (2002). Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 24*, 971–987.
28. Daugman, J. G. (1985). Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters. *Journal of the Optical Society of America, 2*(A),1160–1169.
29. Bovik, A. C., Clark, M., & Geisler, W. S. (1990). Multichannel texture analysis using localized spatial filters. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 12*(1), 55–73.
30. Gil, D., Hernandez, A., Rodríguez, O., Mauri, F., & Radeva, P. (2006). Statistical strategy for anisotropic adventitia modelling in IVUS. *IEEE Transaction Medical Imaging, 27*, 1022–1030.
31. Randen, T., & Husoy, J. H. (1999). Filtering for texture classification: A comparative study. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 4*, 291–310.

Sergio Escalera received the BS and MS degrees from Universitat Autnoma de Barcelona in 2003 and 2005, respectively. He is currently working toward the PhD degree in Computer Science advised by Dr. Petia Radeva and Dr. Oriol Pujol. He is a collaborator professor at Universitat de Barcelona. His research interests include machine learning, statistical pattern recognition, and visual object recognition.

Oriol Pujol received the PhD degree in Computer Science from Universitat Autnoma de Barcelona in 2004. Currently, he is an associate professor at Universitat de Barcelona. His main research interest includes basic statistical machine learning techniques for object recognition and medical imaging analysis.

Josepa Mauri received the title of medical Doctor in 1982 at Universitat Autònoma de Barcelona. In 1992 she received the Laurea summa Cum Laude in Medicine the Universitat de Barcelona. Since 2000, she is the Director of Cardiac Catherization Laboratory in the Hospital Universitari "Germans Trias I Pujol de Badalona". From 2002 to 2005 she was the President of the Diagnostic Intracoronary Technics/ Intravascular Ultrasound Working Group of the Spanish Society of Cardiology. Since October 2006 she is the President of the Spanish Working Group in Cardiac Interventions of the Spanish Society of Cardiology. Her clinical research areas have been in coronary angioplasty and Dilated Cardiomyopathy, dilated Cardiomiopathy, coronary angioplasty, stents, endothelial disfunction and intravascular ultrasound studies, transradial intervention, left main, bifurcations and chronic total occlusions treatment.

Petia Radeva received her PhD at UAB on Development of physics-based models applied to image analysis. Currently, Petia Radeva is an associate professor in the Computer Science Department of the Universitat Autnoma de Barcelona. In 2007 she has joined the Applied Mathematics and Analysis Department of the Universitat de Barcelona. Her present research interest is concentrated on Development of physics-based and statistical approaches for object recognition, medical image analysis and industrial vision.