An Ultrasonographic Risk Score For Detecting Symptomatic Carotid Atherosclerotic Plaques

David Afonso, José Seabra, *Member, IEEE*, Luís M Pedro, J. Fernandes e Fernandes, and J. Miguel Sanches, *Senior Member, IEEE*

***Abstract*—this paper proposes a risk score computed from ultrasound data that correlates to plaque activity. It has the twofold purpose of detecting symptomatic plaques and estimating the likelihood of the asymptomatic lesion to become symptomatic. The proposed *ultrasonographic activity index* (UAI) relies on the plaque active profile, which is a combination of the most discriminate ultrasound parameter associated with symptoms. These features are extracted by the automatic algorithm and also by the physician from the ultrasound images and from some transformations on it, such as monogenic decomposition, which is a novelty in this clinical problem. This information is used to compute a risk score from the conditional probabilities of either symptomatic or asymptomatic groups. Symptom detection performance is evaluated on a transversal dataset of 146 plaques, where UAI obtained 83.5% accuracy, 84.1% sensitivity, and 83.7% specificity. Performance is also assessed on a longitudinal study of 112 plaques, where UAI shows a significant improvement over the gold standard degree of stenosis, demonstrating higher power at predicting which asymptomatic plaques developed symptoms in an average follow-up of ten months. Results suggest that this score could have a positive impact on early stroke prevention and treatment planning.**

***Index Terms*—Atherosclerotic carotid disease, computer-aided detection and diagnosis, machine learning, pattern recognition and classification, probabilistic and statistical methods, risk score, receiver–operator curve (ROC) analysis, ultrasound, vessels.**

I. Introduction

CAROTID bifurcations are prone to develop atherosclerotic plaques that reduce blood flow and are vulnerable to rupture or to break-away causing embolization. The atherosclerotic disease of the carotid is responsible for one-third of acute cerebrovascular events, hence, it has a major clinical, social, and economical impact.

A stable carotid plaque is usually benign, but a vulnerable carotid plaque has high risk of stroke. A correct characterization of the disease is then vital to help identifying and assigning vulnerable plaques to surgery. Here, medical information from the plaque as well as clinical information from the patient are used to access the stability of the plaque and quantify risk of symptoms.

The *degree of stenosis* (DS)—arterial lumen narrowing—is considered the most important feature for determining plaque vulnerability. This measure, together with other patient information such as age, health, and clinical history are used to subjectively decide upon endarterectomy. But this data alone provide uncertain benefits in asymptomatic and moderate obstruction cases [1], potentially submitting the patient to unnecessary surgery risks. This clearly motivates the need for new, more accurate, strategies for plaque risk prediction.

The inexpensive, wide availability, noninvasive, nonionizing nature of *Ultrasound* (US) make this image modality a suitable and practical method to asses carotid disease via real-time visualization and interpretation of the carotid plaque. Although DS and blood flow measures are the standard for US atherosclerotic disease assessment, several recently proposed methods show the benefit of using morphological information.

In one of the earlier textural works [2], the influence of ultrasonic carotid plaque morphology on the incidence of ipsilateral hemispheric symptoms was studied on 96 plaques of which 25 are positive (P) for symptoms and 71 are negative (N). Other authors also showed [3] how simple ultrasound analysis can have a good correlation with symptoms and histological analysis. Pedro *et al.* [4], [5] correlated echostructural characteristics of carotid plaques with neurological symptoms and cerebral infarcts in 105 plaques, and proposed the *activity index* (AI) method. Seabra *et al.* [6] developed this approach further by incorporating several textural features and a Bayesian classification method and proposing an enhanced AI.

Wijeyaratne *et al.* studied [7] the discriminative sensitivity of a multiple cross-sectional analysis using *gray scale median* (GSM) and GSM heterogeneity on 74 plaques ($41N$, $33P$). Some studies [8], [9] incorporated laws' textural features in their classification, while [8] also explore textural local binary patterns obtaining 83% accuracy. Different neurological symptoms were also correlated with specific textural parameters [10].

The range of classifiers employed in this area is also considerable, with *support vector machine* (SVM) classifiers being some of the most used and successful. Kyriacou *et al.* used [11] SVM on multilevel binary and gray scale morphological features on 274 plaques ($137N$, $137P$) to obtain an average accuracy of 73.7%; while Tsiaparas *et al.* used [12] SVM to compare the discriminative power of the dual-tree complex wavelet, the finite ridgelet and the fast discrete curvelet transforms, to access plaque symptoms on 20 plaques ($9N$, $11P$) obtaining an accuracy of $84,9\%$. Modular neural network composed

Manuscript received March 26, 2014; revised July 23, 2014; accepted September 4, 2014. Date of publication September 19, 2014; date of current version July 23, 2015. This work was supported by the FCT Project [PEst-OE/EEI/LA0009/2013]

D. Afonso, J. Seabra, and J. M. Sanches are with the Department of Bioengineering, Institute for Systems and Robotics, Instituto Superior Técnico, Technical University of Lisbon, 1049-001 Lisbon, Portugal (e-mail: dafonso@isr.ist.utl.pt; jseabra@isr.ist.utl.pt; jmrs@ist.utl.pt).

L. M Pedro is with the Medical Faculty of Lisbon, University of Lisbon, 1649-004 Lisbon, Portugal (e-mail: lmendespedro@clix.pt).

J. F. e Fernandes is with the Lisbon University Medical School, Hospital de SantaMaria, 1649-035 Lisbon, Portugal (e-mail: jff@me.com).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/JBHI.2014.2359236

of self-organizing map classifiers have been used with an average accuracy of 73.1% on 230 plaques using ten textural features [13].

It is our belief that to obtain an accurate characterization of the vulnerable plaque, one should consider a supervised features space that takes into account the plaque structure and appearance as judge by experienced physicians, together with morphological and textural computed automatically from the ultrasound image. We also believe that ultrasound speckle [14] can provide useful information for tissue characterization, although most studies filter and discard it as noise [15], [16].

This paper introduces a score for plaque diagnosis, which combines discriminant clinical and image based features. For the latter, we make use of echo-morphological and textural descriptors by decomposing the *B-mode ultrasound imaging* (BUS) image into noiseless and speckle image components, according to [14]. In this method, the echogenic (intensity) contents are estimated on the despeckle image, while the spatial texture arrangement is retained on the speckle image.

Although this study is a progression of several other publications, appropriately cited in this text, this paper introduces two main novelties with respect to the state of the art: $1)$ textural characterization of the plaque from features based on the monogenic description of the speckle field and $2)$ a new weighted linear discriminant classifier, where the prior importance of the features are taken into account and coded in the weights. It is shown that the proposed method outperforms the previous scores described in the literature mainly due to the increased ability of the monogenic approach to capture the textural features of the plaque.

In this paper, the proposed score is used on a dataset from a transversal study to detect symptomatic versus asymptomatic plaques. The method's robustness and its prediction power is also assessed on a dataset from a longitudinal study conducted in asymptomatic subjects.

II. Materials and Methods

In this section, the data used in this study are described and characterized and the method approach is presented.

A. Data

This study includes two datasets.

Dataset A is a transversal (cross sectional) study of 146 carotid bifurcation plaques obtained from 99 patients, 75 males, and 24 females, with an average age of 68 years (41–88). Dataset contains 44 symptomatic (P) and 102 asymptomatic plaques (N) with an average DS of 65% and 44%, respectively. Patients were observed through a medical consultation at *Cardiovascular Institute of Lisbon* and at the *Department of Vascular Surgery, Lisbon Academic Medical Center*. The examination included an assessment with color-flow Duplex-scan of both carotids, performed with ATL-HDI 3000 or 5000 scanner (Philips Medical Systems, Bothell, WA, USA). Equipment setup included 5–12 MHz broadband linear-array transducer, 60-dB dynamic range and postprocessing linear maps. Image acquisition included the most representative longitudinal cut in color and B-mode as selected by the medical expert (L. M. Pedro) who performed the studies.

A plaque was considered P when amaurosis fugax (AF) or focal (transitory, reversible, or established) cerebral events in the carotid territory occurred in the previous six months.

Dataset B is a longitudinal study from the asymptomatic carotid stenosis and risk study [17] with 112 asymptomatic plaques with $\mathrm{DS} > 50\%$. The DS distribution as an average value of $75\%(50-99)$ and the number of plaques with $\mathrm{DS} > 70\%$ is 80. Patients were followed for possible occurrence of symptoms for a mean time interval of 37.1 weeks. At the end of the study, 13 out of 112 patients (11.6%) had developed symptoms (three AF, six transient ischemic attacks, four strokes).

B. Methods

In this section, we describe how the risk score is built. The following method is also implemented as a *computer-aided diagnosis* (CAD) standalone software, which consists in an updated version of [18].

1) Preprocessing: Small regions at the lumen and adventitia areas are identified in the BUS images by the physician for image normalization purposes. In this procedure, the image intensity is linearly rescaled setting the median intensity (gray values) of these areas to 5 and 190, respectively. The plaque is segmented by the clinical expert using the provided interface, resulting in one or more *regions of interest* (ROIs). The original envelope *radio frequency* (RF) image of the ROI is retrieved by reverting [14] the nonlinear compression performed by the ultrasound equipment. This decompression operation is important for the following preprocessing steps, but it is also useful to compensate for the acquisition and equipment parameters variability.

Here, we decompose the envelope ROI image (see Fig. 1) into a despeckled image ($\Lambda = \{\eta_{i,j}\}$), carrying information about the local tissue echogenicity, and a speckle field ($\Gamma = \gamma_{i,j}$), related to the structure and the characteristic pattern of the tissue using the method in [19]. This multiplicative decomposition is supported by the understanding that US speckle is known to depend on the intrinsic acoustic properties of tissues [20]. Different tissues, with their textural variation, are composed of elements with different refractive indexes, that cause propagating sound waves to scatter differently. This differentiation is here explored for texture analysis purposes.

All of these preprocessing steps are sketched in Fig. 2.

2) Feature Extraction: A total of 212 features were considered (see overview in Fig. 2) in this paper. 5 of them are extracted by the physician upon visual inspection of the images, such as DS and surface disruption. The remaining 155 features are computed automatically from the preprocessed images. The preprocessing steps as well as a subset of these features are described in [6]. Here, an extended set of features is used such as described next:

1) *BUS morphological features* provided by an experimented physician during BUS examination. The five-element vector of morphological parameters include: 1) evidence of plaque disruption, defined by an interruption

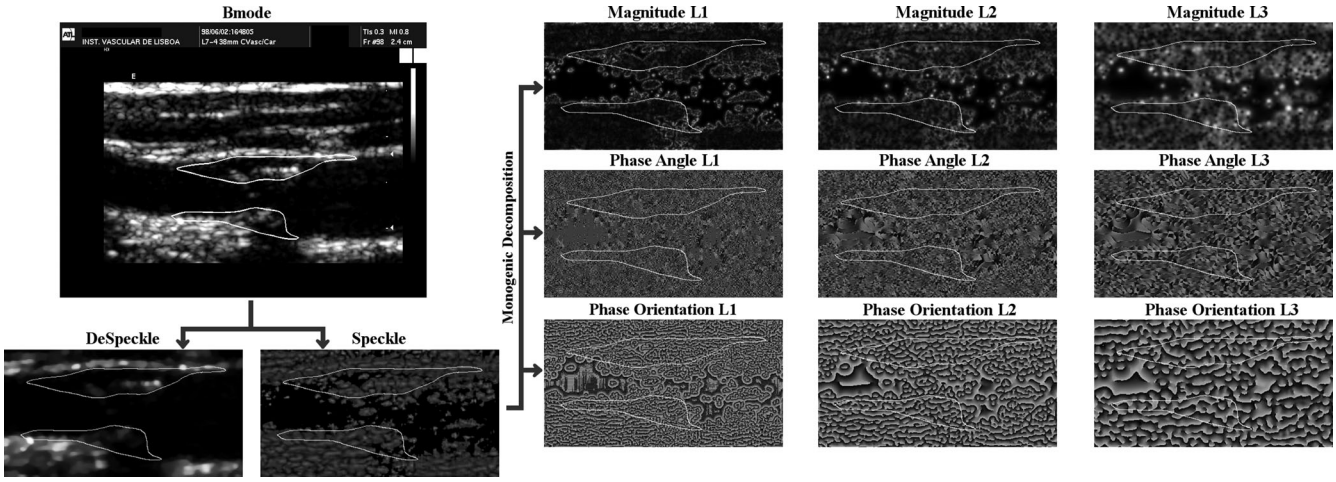


Fig. 1. Decomposition example of Bmode into speckle (η) and despeckle fields and the monogenic decomposition of η into magnitude, phase angle, and phase orientation, using three levels of detail.

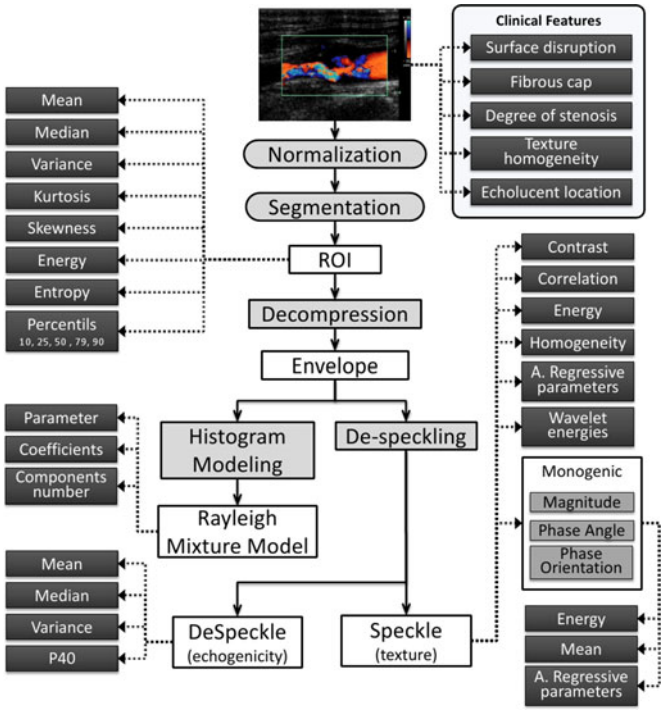


Fig. 2. Scheme of the postprocessing steps applied to the BUS image representing the carotid plaque and the features extracted from them. Rectangular gray: automatic methods; round-shaped gray: semi automatic methods; rectangular white: internal outputs; dark gray: features.

in the echogenic surface of the plaque; 2) presence of echogenic cap, identified as an equivalent of a thick fibrous cap and characterized by an echogenic line over the visible structure of the plaque; 3) the DS, quantified using cross-sectional area measurement combined with hemodynamic assessment; 4) plaque echo-structure appearance, where uniform plaques are defined as homogeneous as opposed to heterogeneous plaques; and 5) location of the echolucent region in heterogeneous lesions (central or juxta-luminal), identified as a dark region inside the plaque.

2) *Histogram features* extracted from the histogram of normalized pixel intensities inside the plaque. A total of 13 histogram features are estimated, including the mean and median gray values (GSM), percentage of pixels with gray value lower than 40 ($PP40$) [4], standard deviation of gray values, kurtosis, skewness, energy, and entropy, 10, 25, 50, 75, and 90 percentiles.
3) *Rayleigh Mixture Model (RMM) features* are extracted from the modeling of plaque echo-morphology contents by a mixture of Rayleigh distributions according with [21]. A 13-element maximum feature vector is obtained consisting of up to six mixture coefficients and six Rayleigh parameters plus the effective nonnull number of RMM coefficients.
4) *Rayleigh features* extracted from the Γ field consist on average theoretical estimators of the Rayleigh distribution, whose parameters are given by the pixels on the despeckled [19] image, which form the plaque. The Rayleigh features include the mean $\mu_\Gamma = \overline{\gamma_{i,j}}\sqrt{\frac{\pi}{2}}$, median $\upsilon_\Gamma = \overline{\gamma_{i,j}}\sqrt{2\log(2)}$, variance $\sigma_\Gamma = \overline{\gamma_{i,j}}\sqrt{\frac{4-\pi}{2}}$ of Rayleigh values, and percentage of pixels with Rayleigh value lower than 40 $PP40_\Gamma = 100 - \exp\left(-\frac{40^2}{2\,\overline{\gamma_{i,j}}^2}\right)$
5) *Texture features* involve the study of the spatial distribution of gray levels inside the plaque region extracted from the speckle image [19]. These features are estimated from *gray level co-occurrence matrices* (GLCM) *autoregressive* (AR) models, Wavelet models, and monogenic signals.

The elements, $q_{d,\theta}(i,j)$ of the GLCMs [22] are the relative occurrence frequencies, $P(i,j|d,\theta)$, of intensity pairs (i,j) of neighboring pixels at distance d and angle θ. 16 GCL images are computed for $d = \{1,2,3,4\}$ and $\theta = \{0,45,90,135\}$. The following four features are

extracted from these 16 GLCM's leading to a 64 length feature vector:

$$\text{Contrast}: \quad \sum_{i,j} |i-j|^2 \cdot q_{d,\theta}(i,j)$$

$$\text{Correlation}: \quad \sum_{i,j} \frac{(i-\mu_i)(j-\mu_j) \cdot q_{d,\theta}(i,j)}{\sigma_i \cdot \sigma_j}$$

$$\text{Energy} \quad \sum_{i,j} q_{d,\theta}(i,j)^2$$

$$\text{Homogeneity} \quad \sum_{i,j} \frac{q_{d,\theta}(i,j)}{1+|i+j|}$$

μ_k and σ_k, with $k = i, j$ are the mean and standard deviation, respectively, of pixel value in the k direction.

Here, *Contrast* measures the local variations in the GLCM, while Correlation gives the joint probability occurrence of the specified pixel pairs. The *Energy* provides the sum of squared elements in the GLCM, and finally, *Homogeneity* measures the closeness of the distribution of elements in the GLCM to the GLCM diagonal.

Additionally a first-order 2-D AR model is used to describe the spatial correlation between pixels. In this model, each pixel from the speckle field (η) is expressed as the following linear combination:

$$\eta_{i,j} = \sum_{n,m} a_{n,m}\, \eta_{i-m,j-n} + \varepsilon_{i,j} \tag{1}$$

where $(n,m) \in \{(0,1);(1,0);(1,1)\}$. $a_{n,m}$ are the AR coefficients to be estimated and $\varepsilon_{i,j}$ are the residues.

Plaque texture is also studied using multilevel 2-D wavelet decomposition and monogenic decomposition signal analysis.

(6) *Wavelets* are used to decompose the speckle image (η) into four levels, resulting in 12 detail and one approximation images. Here, the energies of these fields are used as textural features.

Energy, mean, and AR 2-D coefficients (1) features are extracted from each of these images, resulting in the total of 65 textural features.

Monogenic features are obtained by decomposing the speckle image (η) in the monogenic signal framework [23]. In this framework, a *Riesz* transform, a 2-D generalization of the 1-D *Hilbert* transform [24], allows the decomposition of the input image into three components: magnitude, phase angle, and phase orientation. Here, we set the decomposition parameters to the following default values proposed by Felsberg: 1) minimum wavelength of 4 pixels; 2) scaling factor of 2 between successive filters; 3) 0.65 of ratio for the standard deviation of the Gaussian describing the log Gabor filter's transfer function in the frequency domain to the filter centre frequency.

This decomposition is performed on three band-pass filtered versions of the speckle image allowing for a multiscale analysis (see Fig. 1).

Energy, mean, and AR 2-D coefficients feature (same as extracted for wavelets decomposition) are extracted from the monogenic components of each of the three levels of details, resulting in the total of 45 textural features.

In the end, each plaque is described by a feature vector x of 5 (Clinical) + 13 (Histogram) + 13 (RMM) + 4 (Rayleigh) + 177 (Texture) = 212 features.

TABLE I
LIST OF OVERALL COMPARED FEATURE SELECTION METHODS AND THE RESULTANT FINAL ACCURACY IN %

Feature selection method	Accurracy
Sequential Floating Forward [25], minimum Redundancy Maximum Relevance [26]	84%
particle swarm optimization [27],	81%
Plus-l-takeaway-r [28]	79%
Genetic algorithm [27]	78%
Rough set theory [27]	
Backward feature selection [28]	76%
Branch and bound [28]	72%
Forward feature selection [28]	68%

C. Feature Scaling and Selection

The set of C total features are organized by columns in a $R \times C$ matrix $D = \{d_{i,j}\}$, where rows correspond to plaques.

Nonbinary features are normalized to the maximum range of [0, 1] according to

$$\Phi(d_{i,j}) = \left(1 + e^{-\frac{d_{i,j}-\mu_j}{\sigma_j}}\right)^{-1} \tag{2}$$

where μ_j and σ_j are the mean and standard deviation, respectively, of jth column of D. This transformation removes the influence of the dynamic range of each featured in final classification. The binary features are not normalized since they are already in the same range.

Feature selection was performed to reduce the number of features to less than 5% of the total number of samples on the training dataset A, in order to avoid over fitting.

This procedure is performed in two steps. 1) A preselection based on the discriminative power of each feature and 2) a finer selection with the *Pudil* sequential floating forward feature selection (SFFS) [25], selected after a comparison test of several feature selection methods (see Table I).

Step 1 is required to reduce the computational burden of step 2. In this step, an hypothesis testing is used for determining if the statistical properties of a given feature significantly differ from the P and N classes, by means of the Mann–Whitney $U-$ test [29] that performs a two-sided rank sum test. We select the features with $p-$ value, $\alpha < 0.05$ (probability of rejecting the null hypothesis) resulting in 49 of the starting 212 features.

In step 2, a *linear Bayes normal classifier* [30] with no regularization is used as the wrapped criteria, with a maximum limit of seven features, less than 5% of C.

D. Ultrasonic Activity Index Measure

From the selected features, a single probabilistic measure is devised for plaque characterization. This real valued score, scaled from 0 to 100, is correlated with plaque activity [5]

TABLE II
FEATURES SELECTED IN ORDER OF THEIR SFFS INCLUSION, THEIR MAN–WHITNEY P-VALUE, AND THE CORRESPONDENT WEIGHTING VARIABLE υ (SEE SECTION D)

Name	*p-value* ($\boldsymbol{\alpha}$)	$\boldsymbol{\upsilon}$
Surface Disruption	8.7E-12	0.38
DS	2.8E-8	0.26
$a(1,1)$ AR parameter from the third-level monogenic energy response	2.9E-2	0.05
Intensity correlation to $d = 3$, $\theta = 90$ neighbor of the speckle GLCM	9.3E-3	0.07
Intensity contrast to $d = 1$, $\theta = 0$ neighbor of the speckle GLCM	1.5E-2	0.06
$a(1,1)$ AR parameter from the first-level monogenic phase orientation response	5.6E-3	0.08
Intensity contrast to $d = 2$, $\theta = 45$ neighbor of the speckle GLCM	9.0E-3	0.07

and allows for a relative measure of disease severity. Resuming the proposed *ultrasonic activity index* (UAI) can be seen as a modification of a Bayesian classifier, with multivariate variable distribution and with a new weighting approach.

$$\mathrm{UAI}(\mathbf{x}) = G(\Lambda(\mathbf{x})) \tag{3}$$

where $G(.)$ is a logistic function, $\Lambda(\mathbf{x})$ is the Bayes Factor, and $\mathbf{x} = \{\mathbf{b}, \mathbf{r}\}$ is a vector of $C = m + n$ selected features with m binary components ($\mathbf{b} \in \{0,1\}^m$) and n real components ($\mathbf{r} \in \mathbb{R}^{\mathbf{n}}$) normalized by (2).

$$\Lambda(\mathbf{x}) = l_P(\mathbf{x}) - l_N(\mathbf{x}) \tag{4}$$

is the Bayes factor associated with the binary classification process into N and P classes. $l_\tau(\mathbf{x}) = -\log[p(\mathbf{x}|\omega_\tau)]$ is the symmetric of the log-likelihood function, $\tau = N, P$ and $p(\mathbf{x}|\omega_\tau)$ can be interpreted as the probability of a sample described by feature vector $\mathbf{x}$ to belong to a group whose features are described by ω_τ.

G is a monotonic function used to rescale $\Lambda(\mathbf{x})$ into a maximum range of [0, 100]

$$G(y) = a \,.\, e^{b.e^{c\,\cdot\, y}} \tag{5}$$

known as *Gompertz function* [31], where $y \in \Re$, $a = 100$ is the maximum asymptote, b sets the y displacement, and c sets the growth rate. These parameters are estimated by fitting (5) to the accumulated normalized UAI histogram obtained from the training dataset. The rescaling method provides a more intuitive evaluation of the plaque risk by the physician that typically performs a diagnosis based on the DS $\in [0 - 100]$.

We assume that the m binary components ($\mathbf{b}$) of the feature vector ($\mathbf{x}$) are independent, while the n real components ($\mathbf{r}$) follow a multivariate Gaussian distribution

$$p(\mathbf{x}|\omega_\tau) = p(\mathbf{r}|\omega_\tau) \,\cdot\, p(\mathbf{b}|\omega_\tau) = N(\mu_\tau, \Sigma_\tau) \,\cdot\, \prod_{i=1}^{m} p\left(\mathbf{b}(i)|\omega_\tau\right) \tag{6}$$

where the last factors, evolving the binary features $\mathbf{b}(i)$, are estimated as the relative frequency of the (ith) binary feature value within class τ in the training dataset

$$p(\mathbf{b}(i)|\omega_\tau) = f_\tau(i) \,\cdot\, \mathbf{b}(i) + (1 - f_\tau(i)) \,\cdot\, (1 - \mathbf{b}(i)) \tag{7}$$

where $f_\tau(i)$ is the relative frequency of positive value ($\mathbf{b}(i) = 1$) within class τ in the training dataset.

To take into account the influence of each binary feature on the final score, we introduce in (7) a weighting variable $\upsilon_b(i)$, described next, quantifying the relative importance of each feature, such that

$$\begin{aligned} q(\mathbf{b}(i)|\omega_\tau) = \; & \vartheta_\tau(i) \Big[f_\tau(i)^{\upsilon_b(i)} \,\cdot\, \mathbf{b}(i) \\ & + (1 - f_\tau(i))^{\upsilon_b(i)} \,\cdot\, (1 - \mathbf{b}(i)) \Big] \end{aligned} \tag{8}$$

where $\vartheta_\tau(i) = [f_\tau(i)^{\upsilon_b(i)} + (1 - f_\tau(i))^{\upsilon_b(i)}]^{-1}$ is a normalizing factor to make $q(\mathbf{b}(i)|\omega_\tau)$ a true probability function.

Taking into consideration the binary nature of subset $\mathbf{b}$, the log of (8), to be used in the computation of (4), can be written as

$$\log\left[q(\mathbf{b}(i)|\omega_\tau)\right] \;=\; \upsilon_b(i) \,\cdot\, \log\left[p(\mathbf{b}(i)|\omega_\tau)\right] + \Delta_\tau(i) \tag{9}$$

where $\Delta_\tau(i)$ is a constant.

The symmetric of the log of the first factor on (6), a multivariate Gaussian distribution, is

$$\log[p(\mathbf{r}|\omega_\tau, \mu_\mathbf{r}, \tilde{\Sigma}_\mathbf{r})] = \frac{1}{2}(\mathbf{r} - \mu_\tau)^T \,\cdot\, \tilde{\Sigma}_\tau^{-1} \,\cdot\, (\mathbf{r} - \mu_\tau) + \tilde{K}_\tau \tag{10}$$

where $\tilde{K}_\tau = \log[\frac{1}{\sqrt{(2\pi)^n\,|\tilde{\Sigma}_\tau|}}]$ is the Gaussian normalizing factor, $\mu_\tau \in \Re^n$ is the expectation vector for class τ, and $\tilde{\Sigma}_\tau \in \Re^{n\times n}$ a weighted version of the covariance matrix (Σ_τ) for class τ. This weighting of the covariance matrix narrows/widens the Gaussian distribution changing its discriminative power and is implemented as

$$\tilde{\Sigma}_\tau^{-1} = \left(S_\tau W S_\tau^T\right) \tag{11}$$

where $W = \mathrm{diag}(\boldsymbol{\upsilon_r})$ a diagonal matrix containing the weights associated with relative importance of each real feature ($\mathbf{r}$) and S_τ is the principal square root matrix of the covariance matrix inverse (Σ_τ^{-1}) obtained by the Cholesky decomposition [32] such that $S_\tau S_\tau^T = \Sigma_\tau^{-1}$.

By using (9) and (10), the log likelihood functions in (4) are as follows:

$$\begin{aligned} l_\tau(\mathbf{x}) = \; & \frac{1}{2}(\mathbf{r} - \mu_\tau)^T \,\cdot\, S_\tau W S_\tau^T \,\cdot\, (\mathbf{r} - \mu_\tau) \\ & + \sum_{i=1}^{m} \left[\, \upsilon_b(i) \,\cdot\, \log\left(q(\mathbf{b}(i)|\omega_\tau)\right)\right] + C_\tau \end{aligned} \tag{12}$$

where C_τ combines all the constants from all terms.

Finally, the weighting factors ($\boldsymbol{\upsilon}$) used previously are

$$\boldsymbol{\upsilon} = \frac{-\log(\boldsymbol{\alpha})}{\sum_{k=1}^{N} \sqrt{\left(-\log(\alpha_k)\right)^2}} \tag{13}$$

TABLE III
PERFORMANCE RESULTS FOR DIFFERENT CRITERIA FOR SELECTING CUTOFF VALUE(t) (IN %) FOR THE UAI SCORE ON DATASET A

Criteria	Acc	Sens	Spec	PPV	NPV
max [Acc]	89.4	59.0	84.5	75.3	63.9
max [Gmean]	83.3	83.7	84.1	90.1	66.2
min [|Sens − Spec|]	**83.5**	**84.1**	**83.7**	**91.2**	**65.4**

changing the influence of both binary ($\boldsymbol{v_b}$) and real valued ($\boldsymbol{v_r}$) features on the final score. This change is dependent on each feature discriminative power (see Section II-C) given by the Man–Whitney p-value (α).

The transversal dataset A (see Section II-A) is used to estimate the relative frequencies ($f_\tau(i)$) in (9), the multivariate parameters ($\boldsymbol{\mu_\tau}$, Σ_τ) in (10), the weighting parameters ($\boldsymbol{v}$) in (13), and the score scaling parameters (b and c) in (5).

E. AI Measure

The AI [5], used in this study for comparison purposes, is based on the set of morphological measures extracted by medical expert and the GSM and P40 histogram feature, all described earlier in Section II-B2). In short terms, the AI is a stepwise measure, scaled from 0 to 100, that consists on a linear sum of the probabilities of each significant variable causing symptoms.

F. Classification

The classification strategy, based on any of the real valued scores (DS, AI, and UAI), provides an estimation of the observation class (N or P), $\hat{\tau}$, according to the following simple binary decision rule:

$$\hat{\tau} = \begin{cases} P, & \text{if } score(\mathbf{x}) > t \\ N, & \text{otherwise} \end{cases} \tag{14}$$

where t is a cutoff value that depends on optimal criteria [33]. Also, different criteria lead to different performance results (see Table III). For example, t can be lowered to limit false negatives; on the other hand if false positives are to be avoided, t should be increased. This score, as it is designed, is based on information obtained from US image of the carotid plaque. It is our belief that this tool can be further adapted to incorporate extra variables considered useful such as age, clinical history, surgery costs, hereditary risk factors, etc.

III. EXPERIMENTAL RESULTS

In this section, the classification performance is accessed for the three scores (DS, AI, and UAI) addressed and the results are compared. As noted in Section II-A, we have two datasets used for training and testing. The transversal dataset A, of higher sample size, will be used to capture the plaque characteristics (features) associated with the existence or nonexistence of symptoms. The classifier trained and tested with dataset A, is also tested on a second experiment, with the longitudinal dataset B. Here, we access the predictive power of the classifier to identify plaques that although asymptomatic at the acquisition time developed symptoms afterwards in a time frame of approximately ten months. Considering the significant differences between datasets, this last performance test also allows to access the robustness of the tested methods.

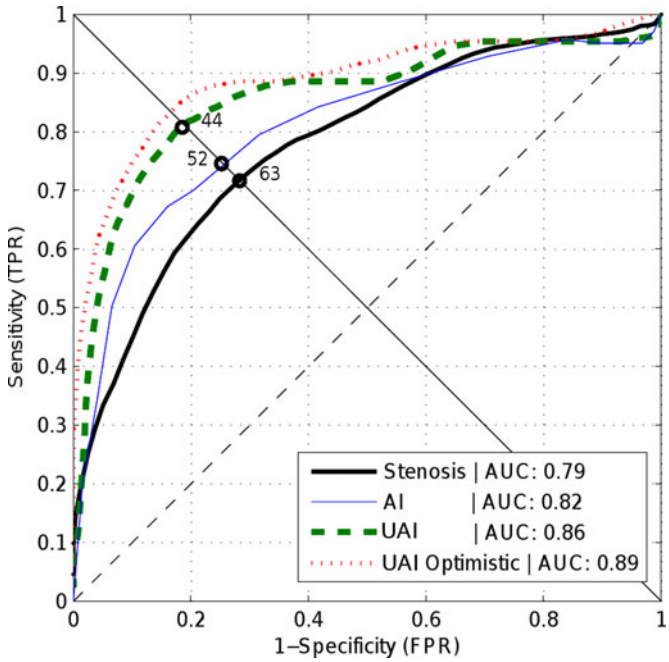


Fig. 3. ROC of the proposed score compared to the standard criteria (DS) and the AI score. Threshold scores for each measure marked with circles. Legend displays AUC for each measure. Point (0, 0) corresponds to highest scores and point (1, 1) to lowest scores.

The performance figures of merit used are accuracy (Acc), sensitivity (Sens), specificity (Spec), and positive and negative predictive value (PPV and NPV)[1].

A. Classification Results

The classifier training (see Section II) is performed with a leave-one-out cross-validation approach [30]. For overfitting evaluation purposes only, an optimistic training is also performed by using the full dataset.

The *receiver–operator curve* (ROC) [34] in Fig. 3 displays the different behavior of DS, AI, UAI, and optimistic UAI. As an overall performance metric, the area under the curve (AUC) is also displayed.

Several optimal criterion to select cut-offs values [threshold t on (14)] on the DS, AI, UAI scores were tested and the results are in Table III. Among these, the common maximization of accuracy, minimizing the Gmean ($\sqrt{\text{Sens} \times \text{Spec}}$) and the chosen simple criteria of the lowest difference between sensitivity and specificity. The later criterion can be graphically represented by intercepting the ROCs with the descending diagonal line (DDL) as shown in Fig. 3. We adopted the threshold values provided

[1]TP = True Positives; TN = True Negatives; FP = False Positives; FN = False Negatives; $\text{Acc} = \frac{\text{TP}+\text{TN}}{\text{TP}+\text{FN}+\text{TN}+\text{FP}}$; $\text{Sens} = \frac{\text{TP}}{\text{TP}+\text{FN}}$; $\text{Spec} = \frac{\text{TN}}{\text{TN}+\text{FP}}$; $\text{PPV} = \frac{\text{TP}}{\text{TP}+\text{FP}}$; $\text{NPV} = \frac{\text{TN}}{\text{TN}+\text{FN}}$.

TABLE IV
PERFORMANCE RESULTS (IN %) FOR THE DS, AI, AND UAI ON DATASET A

Score	AUC	Acc	Sens	Spec	PPV	NPV
DS	79	74.7	75.0	74.5	87.3	55.9
AI	82	75.6	74.6	76.5	88.0	52.7
UAI	**86**	**83.5**	**84.1**	**83.7**	**91.2**	**65.4**
UAI_{s_1}	80	75.3	74.9	77.5	88.2	53.1
UAI_{s_2}	77	75.0	74.6	76.7	87.3	54.2
UAI_{s_3}	82	76.1	75.2	77.1	88.5	53.6

Results for UAI operating on feature subsets (s) are also displayed: s_1 —DS and surface disruption; s_2 —automatically extracted features; s_3 —without monogenic features.

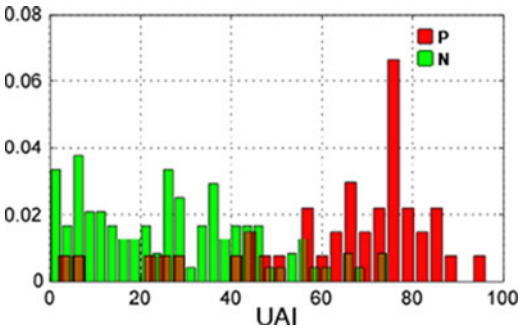


Fig. 4. Normalized histogram of the UAI score for positive and negative classes, in dataset A.

TABLE V
PERFORMANCE RESULTS OBTAINED WITH DS, AI, AND UAI FOR LONGITUDINAL STUDY (DATASET B)

Score	**AUC**	**Cutoff**	**Acc**	**Sens**	**Spec**	**PPV**	**NPV**
DS	65.0	>63	35.3	92.3	29.2	14.5	96.5
AI	73.3	>52	50.4	90.2	44.8	17.4	96.4
UAI	94.6	>44	77.1	100	67.7	33.2	100

by the DDL criteria, due to performance and simplicity, leading to the classification results on Table IV. The proposed method is able to deal with unbalanced datasets, where high accuracies are obtained simultaneously with high specificities and sensitivities. When the data base is unbalanced accuracy may not be an accurate measure of the classification procedure.

The features influence in the final UAI performance is not directly proportional to their v (13) weighting in the classifier. To help assess this impact, Table IV also displays the UAI performance operating on the following subsets of features (see Table II): s_1 includes only the features provided by the clinician (DS and surface disruption); s_2 includes only the automatically computed features, i.e., all features excluding the ones in s_1; s_3 includes all features that do not come from the monogenic decomposition.

The discriminative power of the proposed UAI is also accessed visually in Fig. 4 by observing that the normalized histogram on the symptomatic (P) and asymptomatic (N) plaques are well separated by means of the UAI.

Table V shows results for the second experiment, where the classifier is trained with dataset A and tested with dataset B.

IV. DISCUSSION

Although it would be highly useful to compare our method performance against results from the literature, having different datasets, features, classification methods, and metrics makes this comparison very difficult and not very meaningful. Still the achieved performance is comparatively very interesting, and more than justifies being currently used in the central surgical cardiovascular center in Lisbon.

In this particular classification problem, generalization is a central issue, because severe clinical consequences can result from the decrease of performance. Therefore, we constraint the maximum number of selected features to 5% of the sample size. This upper bound corresponds to the lowest limit commonly accepted in the literature ($[5\%, 20\%]$) to avoid overfitting.

The resulting set of selected features (see Table II) is composed of two morphological features provided by the physician (DS and existence of surface disruption) plus five image-bases automatically extracted features, all of them encoding texture information. Not surprisingly, the binary indicator of plaque heterogeneity given by the physician is not selected in favor of the computed textural features.

Also, although wavelet and monogenic features encode similar information, the automatic selection of the later indicates that monogenic textural information might carry more asymptomatic versus symptomatic discriminative power than wavelets.

Regarding optimal cutoff criteria selection, it should be noted that sensitivity is of particular importance in this application. While maximizing accuracy led to only 59% sensitivity, the chosen criteria provides a significant increase in both sensitivity and specificity for a reasonable tradeoff in accuracy (6%).

For purposes of classification, several classifiers, ensembles of classifiers, and mixture of these ensembles were tested. In the end, the advantage of using a probabilistic classification method provided higher medical usefulness since the probability result can be seen as an AI of the plaque [5]. This real value in the $[0, 100]$ interval allows medical doctors to have a relative idea on disease severity for each patient, which is not provided by classifiers that return a binary decision. In qualitative terms: 1) the lower the score in comparison with the threshold value (t_{UAI}=44), the higher the confidence in classifying it as asymptomatic and 2) the higher the score in comparison with the threshold value, the higher the confidence in classifying it as symptomatic. The opposite of both 1 and 2 are also true.

It should be noted that dataset B has low number of positive cases. With such low statistical relevant data, it is not hard to have a classifier that gives high performance, but suffers from overfitting (generalization problems). Hence, we train and test UAI in dataset A and use B for testing purposes only.

Regarding the performance in dataset A (see Table IV), the proposed UAI clearly outperforms the medical standard DS and the previously proposed AI score. Specifically, UAI shows 9.1% and 7.9% more accuracy than the standard DS and AI, respectively. The results on the subsets of features s_1, s_2, and s_3 clearly show that it is the combination of information provided by the clinician along with texture information automatically extracted that renders the best performance. It also indicates that

every subset of features has relevant impact to the overall UAI performance, with special mention for the proposed monogenic features (s_3).

Regarding results on dataset B (see Table V), UAI delivers a greater ability to characterize and identify asymptomatic plaques that became symptomatic, in an average followup of ten months, being 38.7% and 26.7% more accurate than DS and AI, respectively. Also, UAI was the only method able to identify all TP cases. The longitudinal dataset (B) is rather different from the transversal one (A), naturally resulting in the decrease of all FOM's for the three scores (AI, DS, and UAI). Still, this performance reduction was more significant on AI and DS than on UAI, the later showing higher generalization robustness. Also, datasets A and B were acquired by different operators, with different equipment and image settings, on separate countries. This reinforces the robustness of the proposed method. Although it requires further testing, it can be argued that the rotation and intensity invariant nature of monogenic signals contribute to such generalization robustness.

In absolute terms, the 77.1% accuracy of UAI on dataset B is rather significant considering it was trained with dataset A. This suggests that the method is capturing textural characteristics of symptomatic plaques (positive cases in A) that also exist in asymptomatic plaques that develop symptoms in a near future (positive cases in B).

This study is implemented in an CAD tool used in research [18], which allows the user to register relevant patient information, load, and process images according to the work flow in Fig. 2, extract the image features and calculate the final score. The automatic processing required runs in an average of 10s for each plaque on an 1.7-GHz Intel core i3 laptop. Although user dependant, overall plaque assessment takes around 1 min to complete.

V. Conclusion and Future Work

In summary, the proposed UAI method allows 1) an objective characterization of atherosclerotic disease of the carotid artery, 2) automatic classification into symptomatic versus asymptomatic, and 3) a promising approach to identify plaque vulnerability.

It is a nonexpensive, easy to carry out method that runs on nearly real-time obtained with precessing methodology supervised by experts including both clinical and image-based features.

Plaque assessment based on UAI clearly out performs the characterization based on the DS (which is the medical standard) and the AI method proposed in [5].

UAI uses monogenic decomposition-based features, which to our knowledge have never been applied in the scope of carotid atherosclerotic disease diagnosis. Additionally, a new weighting classification method is proposed to encode the relative discriminative power of each feature on the score.

It is our goal to expand this current method with plaque morphological 3-D features and clinical features such as smoking habits, cholesterol levels, hereditary conditions, medication, etc. Still, the presented results suggest that this score could have a significant impact on stroke prediction and treatment planning.

References

[1] European Carotid Surgery Trialists Collaborative Group, "Randomised trial of endarterectomy for recently symptomatic carotid stenosis: Final results of the MRC European Carotid Surgery Trial (ECST)," *Lancet*, vol. 351, pp. 1379–1387, 1998.

[2] T. Elatrozy, A. Nicolaides, T. Tegos, and M. Griffin, "The objective characterization of ultrasonic carotid plaque features," *Eur. J. Vasc. Endovasc. Surg.*, vol. 16, pp. 223–230, 1998.

[3] L. Baroncini, A. Filho, L. Junior, A. Martins, and S. Ramos, "Ultrasonic tissue characterization of vulnerable carotid plaque: Correlation between videodensitometric method and histological examination," *Cardiovasc. Ultrasound*, vol. 4, p. 32, 2006.

[4] L. Pedro, M. Pedro, I. Goncalves, T. Carneiro, and C. Balsinha, "Computer-assisted carotid plaque analysis: Characteristics of plaques associated with cerebrovascular symptoms and cerebral infarction," *Eur. J. Vasc. Endovasc. Surg.*, vol. 19, pp. 118–123, 2000.

[5] L. Pedro, J. Fernandes, M. Pedro, I. Gonçalves, and N. Dias, "Ultrasonographic risk score of carotid plaques," *Eur. J. Vasc. Endovasc. Surg.*, vol. 24, p. 492–498, Dec. 2002.

[6] J. Seabra, L. Pedro, J. Fernandes, and J. Sanches, "Ultrasound plaque enhanced activity index for predicting neurological symptoms," in *Pattern Recognition and Image Analysis* (Lecture Notes in Computer Science). Berlin, Germany: Springer-Verlag, 2011, pp. 184–191.

[7] S. M. Wijeyaratne, S. Jarvis, L. A. Stead, S. G. Kibria, J. A. Evans, and M. J. Gough, "A new method for characterizing carotid plaque: multiple cross-sectional view echomorphology," *J. Vasc. Surg.*, vol. 37, no. 4, pp. 778–784, Apr. 2003.

[8] U. R. Acharya, S. Vinitha Sree, F. Molinari, L. Saba, S. Y. S. Ho, A. T. Ahuja, S. C. Ho, A. Nicolaides, and J. S. Suri, "Atherosclerotic risk stratification strategy for carotid arteries using texture-based features," *Ultrasound Med. Biol.*, vol. 38, no. 6, pp. 899–915, 2012.

[9] S. G. Mougiakakou, S. Golemati, I. Gousias, A. N. Nicolaides, and K. S. Nikita, "Computer-aided diagnosis of carotid atherosclerosis based on ultrasound image statistics, laws' texture and neural networks," *Ultrasound Med. Biol.*, vol. 33, no. 1, pp. 26–36, 2007.

[10] S. K. Kakkos, A. N. Nicolaides, E. Kyriacou, S. S. Daskalopoulou, M. M. Sabetai, C. S. Pattichis, G. Geroulakos, M. B. Griffin, and D. Thomas, "Computerized texture analysis of carotid plaque ultrasonic images can identify unstable plaques associated with ipsilateral neurological symptoms," *Angiology*, vol. 62, no. 4, pp. 317–328, 2011.

[11] E. Kyriacou, M. S. Pattichis, C. S. Pattichis, A. Mavrommatis, C. Christodoulou, S. Kakkos, and A. Nicolaides, "Classification of atherosclerotic carotid plaques using morphological analysis on ultrasound images," *Appl. Intell.*, vol. 30, no. 1, pp. 3–23, 2009.

[12] N. Tsiaparas, S. Golemati, I. Andreadis, J. Stoitsis, I. Valavanis, and K. Nikita, "Assessment of carotid atherosclerosis from b-mode ultrasound images using directional multiscale texture features," *Meas. Sci. Technol.*, vol. 23, no. 11, p. 114004, 2012.

[13] C. Christodoulou, C. S. Pattichis, M. Pantziaris, and A. Nicolaides, "Texture-based classification of atherosclerotic carotid plaques," *IEEE Trans. Med. Imag.*, vol. 22, no. 7, pp. 902–912, Jul. 2003.

[14] J. Sanches, A. Laine, and J. Suri, *Ultrasound Imaging: Advances and Applications*. New York, NY, USA: Springer, 2012.

[15] O. Michailovich and A. Tannenbaum, "Despeckling of Medical Ultrasound Images," *IEEE Trans. Ultrason., Ferroelectr. Freq. Control*, vol. 53, no. 1, pp. 64–78, Jan. 2006.

[16] Y. Yu and S. Acton, "Speckle Reducing Anisotropic Diffusion," *IEEE Trans. Image Process.*, vol. 11, no. 11, pp. 1260–1270, Nov. 2002.

[17] A. Nicolaides, S. Kakkos, M. Griffin, M. Sabetai, S. Dhanjil, T. Tegos, D. J. Thomas, A. Giannoukas, G. Geroulakos, N. Georgiou, S. Francis, E. Ioannidou, and C. J. Dor, "Severity of asymptomatic carotid stenosis and risk of ipsilateral hemispheric ischaemic events: Results from the ACSRS Study," *Eur. J. Vasc. Endovasc. Surg.*, vol. 30, no. 3, pp. 275–284, Sep. 2005.

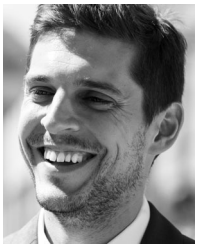
[18] D. Afonso, J. Seabra, J. S. Suri, and J. M. Sanches, "A CAD system for atherosclerotic plaque assessment," in *Proc. Annu. Int. Conf. Eng. Med. Biol. Soc.*, 2012, pp. 1008–1011.

[19] J. Seabra and J. Sanches, "On estimating de-speckled and speckle components from B-mode ultrasound images," in *Proc. IEEE Int. Symp. Biomed. Imag.*, Rotterdam, The Netherlands, Apr. 2010, pp. 284–287.

[20] C. Sehgal, "Quantitative relationship between tissue composition and scattering of ultrasound," *Acoustical Soc. Amer. J.*, vol. 94, p. 1944–1952, Oct. 1993.
[21] J. C. R. Seabra, F. Ciompi, O. Pujol, J. Mauri, P. Radeva, and J. M. Sanches, "Rayleigh mixture model for plaque characterization in intravascular ultrasound," *IEEE Trans. Biomed. Eng.*, vol. 58, no. 5, pp. 1314–1324, May. 2011.
[22] R. Haralick, K. Shanmugam, and I. Dinstein, "Textural features for image classification," *IEEE Trans. Syst., Man Cybern.*, vol. SMC-3, no. 6, pp. 610–621, Nov. 1973.
[23] M. Felsberg and G. Sommer, "The monogenic signal," *IEEE Trans. Signal Process.*, vol. 49, no. 12, pp. 3136–3144, Dec. 2001.
[24] R. N. Bracewell and R. Bracewell, *The Fourier Transform and its Applications*. New York, NY, USA: McGraw-Hill, 1986, vol. 31999.
[25] P. Somol, P. Pudil, J. Novovicov, and P. Paclk, "Adaptive floating search methods in feature selection," *Pattern Recog. Lett.*, vol. 20, no. 11–13, pp. 1157–1163, 1999.
[26] H. Peng, F. Long, and C. Ding, "Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 8, pp. 1226–1238, Aug. 2005.
[27] N. Omar, F. Jusoh, R. Ibrahim, and M. Othman, "Review of feature selection for solving classification problems," *J. Inf. Syst. Res. Innovation*, vol. 3, 2013.
[28] I. Guyon and A. Elisseeff, "An introduction to variable and feature selection," *J. Mach. Learn. Res.*, vol. 3, pp. 1157–1182, 2003.
[29] J. Gibbons and S. Chakraborti, *Nonparametric Statistical Inference*. Boca Raton, FL, USA: CRC Press, May. 1992.
[30] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, 2nd ed. New York, NY, USA: Wiley, 2001.
[31] J. Kenney and E. Keeping, *Mathematics of Statistics*. New York, NY, USA: Van Nostrand, 1947.
[32] M. Hazewinkel, *Encyclopaedia of Mathematics*, vol. 9. New York, NY, USA: Springer, 1993.
[33] L. M. Pedro, J. M. Sanches, J. Seabra, J. S. Suri, and J. F. E Fernandes, "A new clinical tool for assessing neurological stroke risk of asymptomatic carotid disease: A 146 plaque study (accepted)," *Echocardiography*, vol. 31, pp. 353–361, 2013.
[34] S. M. Kay, *Fundamentals of Statistical Signal Processing (Detection Theory), Vol. 2*. Englewood Cliffs, NJ, USA: Prentice Hall, 1998.

David Afonso received the M.Sc. degree from the Faculdade de Mediciana de Lisboa, Instituto Superior Técnico, Lisboa, Portugal, in 2002, where he is currently working toward the Ph.D. degree in biomedical engineering.

He is specialized in signal processing and machine learning, with focus on medical imaging such as ultrasound functional MRI. He also worked for Siemens and Arthur D. Little. His research interests include medical image processing, computer aided diagnosis, machine learning, and app and web development



José Seabra received the M.Sc. and Ph.D. degrees from the Technical University of Lisbon, Lisbon, Portugal, in 2007 and 2011, respectively.

His thesis investigated different methologies to model and characterize the carotid artery disease using ultrasound images. Currently, he is working as an R&D Engineer for the iMagX project—a partnership between the Universit catholique de Louvain and Ion Beam Applications, world leader in proton therapy—focused in the research and development of complete imaging solutions for patient positioning in radiotherapy and proton therapy. His research interests include medical image processing, computer aided diagnosis, reconstruction, denoising, segmentation, and machine learning.

Luís Mendes Pedro received the M.D. and Ph.D. degrees from the Medical School of the University of Lisbon, Lisbon, Portugal, in 1987 and 1993, respectively.

He is currently an Associate Professor and a Consultant of Vascular Surgery with the Cardiovascular Institute and the Lisbon University Medical School, Hospital de Santa Maria, Lisbon. His main research interests include ultrasound plaque characterization, carotid surgery, and endovascular surgery.

José Fernandes e Fernandes received the M.D. and Ph.D. degrees.

He is the Chief of Vascular Surgery, and the Chairman and a Full Professor of Surgery at the University of Lisbon Medical School, Hospital de Santa Maria, Lisbon, Portugal. His current research interests include vascular hemodynamics, noninvasive vascular evaluation, and morphological studies of arterial wall in atherosclerosis, vascular surgery, and endovascular reconstructions in both occlusive and aneurysmal disease.

João Sanches J. Miguel Sanches teaches at the Instituto Superior Técnico, University of Lisbon, Lisbon, Portugal, at the Bioengineering Department in the scientific area of biosignals. His research activity, developed in the Institute of Systems and Robotics (ISR), has been focused in biomedical engineering, namely, in biological and medical image processing and statistical signal processing of physiological and behavioral data. Ultrasound image processing for the diagnosis of the atherosclerotic disease of the carotid and diffuse diseases of the liver are central issues as well as fluorescence images of microscopy for biological quantification purposes. Almost all of his research work is in collaboration with medical and biological institutions, and several publications and patents were already produced in the scope of this collaborative work.

Mr. Sanches has been a Senior Member of the IEEE Engineering in Medicine and Biology Society since 2011 and a Member of the Bio Imaging and Signal Processing Technical Committee of the IEEE Signal Processing Society.