Detection of Carotid Plaque Symptoms Using Ultrasound Imaging*

David Afonso[1,2,**] and João Miguel R. Sanches[1,2]

[1] Institute for Systems and Robotics
dafonso@isr.ist.utl.pt
[2] Department of Bioengineering,
Instituto Superior Técnico, Technical University of Lisbon

Abstract. Carotid plaques are one of the commonest causes of neurological symptoms due to embolization of plaque components or flow reduction. The classification of plaques vulnerability is then a relevant clinical issue, and a technical challenge. Recently, several atherosclerotic plaque characterization methods were proposed based on plaque morphology assessed through 2D ultrasound. One of these methods, proposed by Seabra *et al* [1] presents a measure with clinical significance, known as enhanced activity index (EAI), that the clinician then uses to classify the plaque. The present paper builds upon that work and by using machine learning, proposes an ensemble classifier that shows promising results outperforming both the gold medical standard degree of stenosis and the EAI score. Results are obtained on a real clinical database of 146 plaques. Future work will investigate the predictive performance of the proposed classifier, *i.e.*, how well does the classifier identify stable lesions at high risk of becoming symptomatic.

1 Introduction

The arteries that supply our brains, the carotids, are prone to develop atherosclerotic plaques that reduce blood flow. More dangerous than that, they are vulnerable to rupture or break-away and block smaller vessels causing ischemia (death) to the surrounding tissues. Carotid bifurcation disease is actually responsible for one-third of acute cerebrovascular events, hence it has a major clinical and social impact.

A stable Carotid plaque is usually benign with a stroke risk around 3% annually, but a more vulnerable plaque migth easilly cause myocardial infarction, stroke and lower limb ischemia. Correct characterization of the carotid disease is then vital for an accurate decision to surgically remove the plaque (carotid endarterectomy) or not. The major premise here is that a vulnerable plaque contains predictive information for future cardiovascular events. Hence, its detection might play a major role in the treatment decision that has important clinical, social and economical consequences.

* This work was supported by the FCT project [PEst-OE/EEI/LA0009/2011].
** Corresponding author.

J.M. Sanches, L. Micó, and J.S. Cardoso (Eds.): IbPRIA 2013, LNCS 7887, pp. 584–591, 2013.

The degree of stenosis (arterial lumen narrowing) is up to now considered the most important features for determining the plaque vulnerability. This metric, together with other patient information such as age, health and clinical history are features clinicians usually use to subjectively decide upon endarterectomy. Even though, several numerous studies [2,3,4,5] report that plaque morphology is also an important ultrasound marker that positively correlates with symptoms.

Ultrasound (US) is a suitable imaging technique to assess this pathological condition mostly because it provides real-time visualization and interpretation of the carotid plaques, it is non-invasive, does not involve ionizing radiation, it is cheap and is very common in clinical facilities.

There are several recently proposed methods [2,3,4,1], that make use of plaque morphology to characterize carotid plaques. In general it is argued that an optimal method for identifying vulnerable lesions should include morphological and textural features, extracted from pixel intensity information, and clinical information regarding plaque structure and appearance (e.g. stenosis, evidence of surface disruption and presence of echogenic cap) given by experienced physicians. The combination of this information is expected to produce a more comprehensive description of the profile of a vulnerable plaque.

J.Seabra method [1] makes use of very interesting texture features, by not discarding the so called "image noise", but rather considering this as an important information source of tissue texture, translated into US speckle. It then proposes a measure with clinical significance, known as enhanced activity index (EAI), a risk score in $[0, 100]$ range, that the clinician then uses to make a final decision (classification).

Although this method might be comfortable for the clinicians, because it provides them with a score that they are used to having (such as the degree of stenosis, age, etc), it restricts the classifier structure and hence it's performance.

In this work we propose a full machine learned classifier that makes a binary detection of plaque vulnerability, and that outperforms both the degree of stenosis (DS) gold-standard and the EAI score method.

2 Problem Formulation and Data

We wish to develop a classification method to decide if a carotid lesion (plaque) is associated to neurological events (symptoms). A plaque was considered symptomatic when amaurosis fugax (AF) or focal (transitory, reversible or established) neurological symptoms in the carotid territory were observed in the previous 6 months. We then have a binary detection problem, where we define the positive class (P) as symptomatic and the negative class (N) as asymptomatic. The available dataset contains 146 ultrasound b-mode images of real carotid arteries with plaques, from a cross-sectional study of 99 patients (75 males and 24 females with a mean age of 68 years (41-88)) acquired at *Instituto Cardiovascular de Lisboa* and *Department of Vascular Surgery, Hospital de Santa Maria, Lisboa*. The ground truth of this database is $N = 102(70\%)$ and $P = 44(30\%)$.

3 EAI and DS Presentation

In this section we present an overview of the DS and EAI methods [1], which results are compared against the proposed method. Both methods are characterized by a final positive score to which a linear threshold is applied for classification.

The criterion used to define this threholds leads to different performance results. Seabra proposed the "sensitivity = specificity", criteria and the consequent performance of both AI and EAI is used for comparison with the proposed method in 3.

Although the DS feature constitutes a score by itself, EAI is a score extracted from several features (see section 4.1). First it follows a statistical (*Mann-Whitney*) feature selection. Reference values for the mean (μ) and variance (σ^2) of each selected feature across all training objects are computed for symptomatic (P) and asymptomatic (N) plaques. The score is then similar to a Bayes factor:

$$EAI = \frac{R_P}{R_N}, \text{ where } R_K = \sum_i N(\mu_i(k), \sigma_i^2(k)), \quad k = \{P, N\} \tag{1}$$

giving the clinician a risk score, as does DS, upon which a final classification is performed, either with unsubjective threshold criteria (such as the one mentioned above), or by taking into consideration extra variables the clinician might find useful.

4 Methods

We employed a three-step method sketched in Fig. 1. The first step (*i*) consists of feature extraction from the US images. The second (*ii*) step consists of training and evaluating several classifiers in the full dataset feature space. In the third (*iii*) stage a final classifier ensemble is build and evaluated on a reduced feature space.

4.1 Feature Extraction *(i)*

After image acquisition, the clinician is required to manually identify a small area both in the lumen and adventitia (for image normalization) and segment the plaque ROI, since this an important step in the method. The final dataset (Fig. 2) is comprised of two subsets of features: *A)* the 4 features obtained by the clinician (existence of fibrous cap, surface disruption, plaque texture homogeneity and DS) which are mainly binary and *B)* 110 features automatically extracted from the images, via a series of post-processing steps (histogram features, Rayleigh mixture models features, Rayleigh parameters, textural features, morphological features, etc) [1] consisted of real numbers. Features extracted take into account the nature of US images and their, namely the ones based on the Rayleigh mixture models [6] and the ones extracted from the images that result from decomposing the B-mode image into its anatomical and textural tissue components [7].

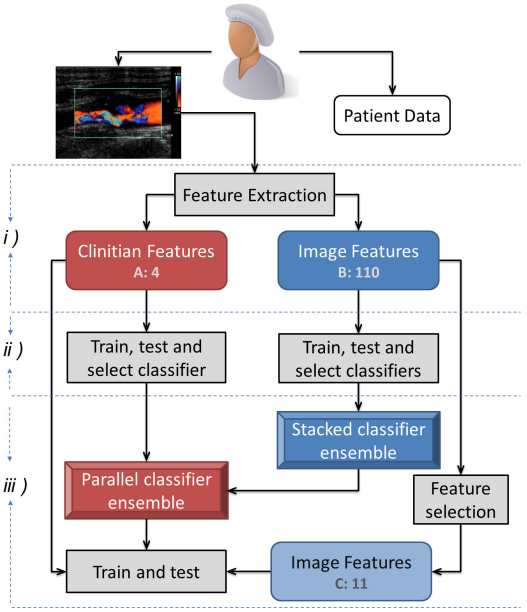

Fig. 1. Scheme of the phases (*i*, *ii*, *iii*) implemented to build and evaluate the proposed classifier

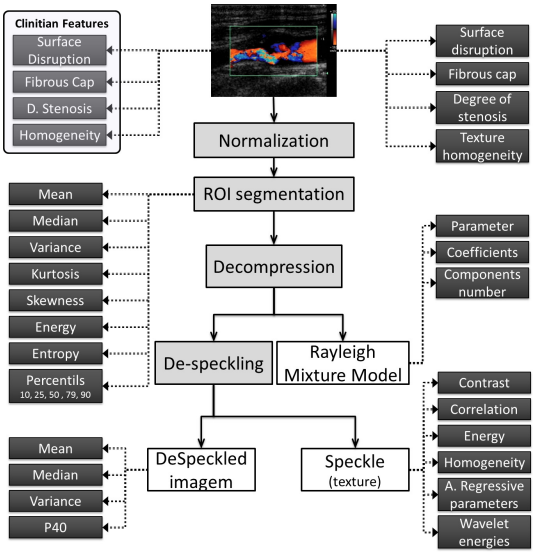

Fig. 2. Scheme of the post-processing steps applied to the b-mode US image representing the carotid plaque and the corresponding extracted features

4.2 Classifier Evaluation and Selection *(ii)*

We use the PRTools Matlab® toolbox to train and test 18 different classifiers (see Table 1) on the separate subsets A and B, using a bootstrap approach. This dataset separation lead to better results, presumably because of their different nature (see section 4.1). Almost unbiased metrics of Probability of error, Sensitivity, Specificity and Precision are computed using the 0.632+ bootstrap bootstrap estimate:

$$metric_{632} = 0.368 \times metric_{Apparent} + 0.632 \times metric_{Jack-knifing} \tag{2}$$

where the apparent (optimistic) metric estimate has equal training and test sets, and Jack-knifing is a pessimistic estimator where training and sets are disjoint. All sets are randomly selected, and metrics averaged over 100 bootstrap repetitions.

The overall criteria adopted in this work to evaluate classifier performance is the lowest difference between sensitivity and specificity and below average probability of error (see table 2). I.e:

$$Pe(k) < \frac{1}{18}\sum_{n=1}^{18} Pe(n) \ \cap \ \underset{k}{\operatorname{argmin}} \left(|Sen(k) - Spec(k)|\right) \ , k = 1:18 \tag{3}$$

where Pe is the Probability of error, Sen sensibility and $Spec$ the specificity, all of them 0.632+ corrected. Besides the intrinsic logic of the "sensitivity = specificity" criteria, the main reason for this criteria is that the method in [1] also proposes the same criteria (see section 3) to set the final threshold over the EAI score for classification purposes, and we wish to have a fair comparison as much as possible.

As expected, several classifier have similar performance on dataset A, due to its low dimensionality, including the Support Vector Classifier, and stacking them in ensembles also did not produce better performance. On the other hand on dataset B the classifiers performances diverge (see Table 2). We then construct a stacked ensemble of the top 5 classifiers on B (see table 1) according to criteria (3).

The classifiers combination rule follows the class which yields the highest value of the product of the classifiers posterior probabilities.

4.3 Feature Selection and Classifier Ensemble *(iii)*

To avoid classifier overfitting, feature selection reduced the size of subset B from 110 to 11 features (subset C). This dimension was optimally determined by the forward-search procedure as the feature space that lead to the lowest probability of error, taking the *1-Nearest Neighbor (1-NN)* criteria. Because 1-NN was the classifier used for feature selection, the *K-Nearest Neighbor classifier* was removed from the classifiers (Table 2) poll to avoid overfitting.

Table 1. List of tested classifiers. *rp* means "regularization parameter"; *sp* means "smoothing parameter";

Classifier	**Parameters**
• Support Vector	linear kernel; $rp = 1$
• Voted Perceptron	10 sweeps
• Linear Bayes Normal	no regularization
• Subspace Classifier	1 dimension
• Quadratic Bayes Normal	no regularization
• Decision Stump Classifier	purity criterion; 1 node
• Uncorrelated Normal Based Quadratic Bayes	
• Linear Perceptron	learning rate = 0.1;
• K-Nearest Neighbor	K is optimized with leave-one-out error
• Naive Bayes	10 bins
• Fisher's Least Square Linear	
• Optimisation of the Parzen	*sp* estimated from dataset
• Nearest Mean	
• Quadratic Discriminant	no regularization
• Parzen density based	no smoothing
• Logistic Linear	
• Discriminative Restricted Boltzmann Machine	50 hidden units; no regularization
• Support Vector Classifier: NU algorithm	linear kernel; 1-NN error ≤ 0.01

The final classifier is a parallel ensemble of the expert stacked classifier (see Table 2) on dataset C and the expert Support Vector Classifier on dataset A. The combination follows the class with the highest vote of the base classifiers by using:

$$D(k, j) = (v + 1)/(n + c) \tag{4}$$

Table 2. Individual classifier performances on the subsets A and B. All metrics are 0.632+ corrected for a more unbiased estimate. Only one classifier for A is shown because several classifiers displayed very similar behavior.

Classifier	***Set***	***Prob. Error***	***Sensitivity***	***Specificity***	***Precision***
Quadratic Discriminant	B	30.53 ± 0.31	65.97 ± 1.11	70.97 ± 0.47	55.47 ± 0.37
Fisher	B	30.91 ± 0.27	65.51 ± 1.27	70.67 ± 0.43	55.39 ± 0.36
Logistic	B	29.26 ± 0.28	66.90 ± 1.11	72.44 ± 0.42	57.37 ± 0.38
Parzen	B	26.98 ± 0.20	60.23 ± 2.14	78.96 ± 1.30	57.55 ± 0.97
Linear Bayes	B	27.88 ± 0.28	55.99 ± 1.02	79.13 ± 0.44	54.62 ± 0.67
Support Vector	A	16.04 ± 0.19	73.36 ± 1.11	88.11 ± 0.25	72.96 ± 0.75

in which v is the number of votes object k receives for class j, n is the total number of classifiers, and c the number of classes (two). The training and evaluating procedures are the same as described in section 4.2.

5 Experimental Results

In this document, for space sake, we do not present the results for all these 18 classifiers (Table 1) plus the created ensembles (see section 4.3) but only the "best ones" according to the criteria (3) presented in section 4.2.

The performance results of the individual classifiers on both subsets A and B, using the bootstrap approach discussed in section 4.2, are display on Table 2

The performance of the classifiers on dataset B is clearly not as good as on A, as expected, considering that the later includes the DS. The stacked classifier ensemble help then to dilute their errors and obtain a better predictive performance, as we can see in Fig. 3.

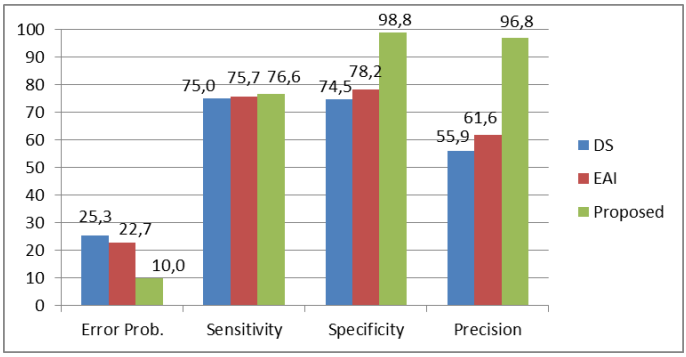

Fig. 3. Proposed classifier performance compared to the standard criteria (DS) and the EAI score. Error bars are too small to be graphically visible and relevant.

These final results (Fig. 3) show a relative reduction of error probability of 61% over DS and 56% over EAI. Precision also displays better performance with an increase of 41% and 36% over DS and EAI, respectively. Although Specificity also shows a performance with a 21% relative increase over both DS and EAI, the increase in Sensitivity is marginal, with $1,3\%$ over EAI and 2.2% over DS.

6 Conclusions and Future Work

We present a work-in-progress that shows promising results in the detection of asymptomatic v.s. symptomatic (vulnerable) carotid plaques, using a machine learned classifier ensemble over a set of features extracted from ultrasound images by both the clinician (A) and computer processing (B). We approach the problem by building two expert classifiers on both (A) and (B) datasets, and

ensemble them in a parallel classifier. To build these experts we search for good performance individual classifiers out of 18 on these subsets, and ensemble them in a stacked approach for dataset (B).

After feature selection, to avoid overfitting, final results show a significant increase in performance on error probability, specificity and precision when compared to the degree-of-stenosis (which is the medical standard) and the EAI method proposed in [1]. But there is still a marginal increase in Sensibility which does not allow us a clear statement of over performance, and that will be addressed on future work.

This detection procedure allows for a characterization of the plaque vulnerability. Future work will provide an estimate of the predictive power (identify asymptomatic plaques that develop symptoms down the line) of this characterization on a longitudinal study.

References

1. Seabra, J., Pedro, L.M., Fernandes e Fernandes, J., Sanches, J.: Ultrasound plaque enhanced activity index for predicting neurological symptoms. In: Vitrià, J., Sanches, J.M., Hernández, M. (eds.) IbPRIA 2011. LNCS, vol. 6669, pp. 184–191. Springer, Heidelberg (2011)
2. Elatrozy, T., Nicolaides, A., Tegos, T., Griffin, M.: The objective characterization of ultrasonic carotid plaque features. Eur. J. Vasc. Endovasc. Surg. 16, 223–230 (1998)
3. Pedro, L., Pedro, M., Goncalves, I., Carneiro, T., Balsinha, C.: Computer-assisted carotid plaque analysis: characteristics of plaques associated with cerebrovascular symptoms and cerebral infarction. Eur. J. Vasc. Endovasc. Surg. 19, 118–123 (2000)
4. Baroncini, L., Filho, A., Junior, L., Martins, A., Ramos, S.: Ultrasonic tissue characterization of vulnerable carotid plaque: correlation between videodensitometric method and histological examination. Cardiovasc Ultrasound 4, 32 (2006); Comparative Study
5. Park, A., McCarthy, W.J., Pearce, W.H., Matsumura, J.S., Yao, J.S.: Carotid plaque morphology correlates with presenting symptomatology. J. Vasc. Surg. 27(5), 872–878, discussion 878–879 (1998)
6. Seabra, J.C.R., Ciompi, F., Pujol, O., Mauri, J., Radeva, P., Sanches, J.M.: Rayleigh mixture model for plaque characterization in intravascular ultrasound. IEEE Trans. Biomed. Engineering 58(5), 1314–1324 (2011)
7. Seabra, J., Sanches, J.: On estimating de-speckled and speckle components from B-mode ultrasound images. In: Proceedings of IEEE International Symposium on Biomedical Imaging, Rotterdam, The Netherlands, pp. 284–287. IEEE Engineering in Medicine and Biology Society (April 2010)