

Enhancing In-Vitro IVUS Data for Tissue Characterization

Francesco Ciompi¹, Oriol Pujol¹, Oriol Rodriguez Leor³, Carlo Gatta²,
Angel Serrano Vida⁴, and Petia Radeva¹

¹ Dep. of Applied Mathematics and Analysis, University of Barcelona, Spain

² Computer Vision Center, Campus UAB, Bellaterra, Barcelona, Spain

³ Hospital Universitari “Germans Trias i Pujol”, Badalona, Spain

⁴ Hospital General de Granollers, Granollers, Barcelona, Spain

f.ciompi@cvc.uab.es

Abstract. Intravascular Ultrasound (IVUS) data validation is usually performed by comparing post-mortem (*in-vitro*) IVUS data and corresponding histological analysis of the tissue, obtaining a reliable *ground truth*. The main drawback of this method is the few number of available study cases due to the complex procedure of histological analysis. In this work we propose a novel semi-supervised approach to enhance the *in-vitro* training set by including examples from *in-vivo* coronary plaques data set. For this purpose, a *Sequential Floating Forward Selection* method is applied on *in-vivo* data and plaque characterization performances are evaluated by *Leave-One-Patient-Out* cross-validation technique. Supervised data inclusion improves global classification accuracy from 89.39% to 91.82%.

Keywords: Intravascular Ultrasounds, Plaque characterization, Semi-supervised learning.

1 Introduction

Coronary plaque rupture is one of the most frequent case of acute coronary syndromes and it can end in myocardial infarction or sudden cardiac death [2][8]. An accurate analysis of *in-vivo* plaque composition represents an important task in diagnosis and detection of vulnerable atheroma before plaque rupture.

IVUS is a powerful imaging technique that gives a detailed cross-sectional image of the vessel allowing to explore both coronary arteries morphology and composition. Automatic plaque composition have been performed by texture analysis on IVUS images [15][17] as well as spectral analysis on raw Radio Frequency (RF) signals [6][11][12][13].

However automatic analysis is hindered by uncontrolled data acquisition with different imaging system parameters, thus data acquired in different cases are not comparable. In this work we exploit the RF access to reconstruct IVUS images with a unique and well controlled parameter set: this process gives us the chance

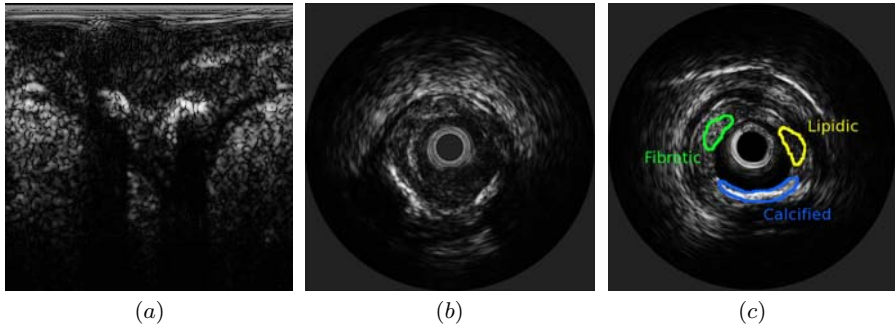


Fig. 1. (a) *In-vivo* IVUS image obtained from RF data and (b) corresponding image in cartesian coordinates; (c) *In-vitro* labeled plaques

to extract normalized features to use in the plaque characterization process, to discriminate among *fibrotic*, *lipidic* and *calcified* tissue.

To achieve a correct IVUS data validation based on tissue-plaque ground truth, a histological analysis of *in-vitro* coronary arteries should be performed. This methodology has the main problem of complicated procedure to obtain *in-vitro* data: scarce artery availability, frequent tissue spoiling during analysis and possible mismatching between IVUS and histological image. As counterpart, provided data are highly reliable. On the other hand, collecting *in-vivo* data is a relatively easy task but the plaques segmentation, performed by expert physicians, cannot be histologically validated.

We propose a method to select plaques from *in-vivo* cases to feed the *in-vitro* training dataset in a supervised manner. The inclusion criterion is based on *Sequential Floating Forward Selection* (SFFS) algorithm [18] and results are computed by *Leave-One-Patient-Out* (LOPO) [4] cross-validation technique. The multiclass classification problem is here solved by adopting the *Error Correcting Output Code* (ECOC) technique [1][7] using AdaBoost on Decision Stumps as basic classifier.

2 Plaque Characterization in Hybrid Data

Since our goal is to fuse *in-vivo* and *in-vitro* data, it is important to consider that the presence of *blood* into *in-vivo* data may modify Ultrasounds response respect to *in-vitro* cases. Our hypothesis is that the two data types could share areas in the feature space, and that *in-vivo* data inclusion, after a proper selection process, could enhance the validated dataset improving classification performances. This fact assumes even more importance since the final application shall detect plaques in *in-vivo* real cases.

Data inclusion process is here formulated as a *semi-supervised learning* problem [20], since *in-vivo* plaques segmentation cannot be validated by histology.

2.1 Data Processing

IVUS RF data (acquired by either *in-vivo* or *in-vitro* cases) are processed to extract textural and spectral features. First, the spatial US attenuation is compensated via *Time Gain Compensation* (TGC). Then the IVUS image is obtained by Band Pass filtering, envelope recovering, normalization, logarithmic compression and Digital Development Process [3]. Classical IVUS image is produced by converting polar data (Figure 1a) in cartesian coordinates (Figure 1b), using linear interpolation and Gaussian smooth filtering.

In order to extract spectral features, we compute the power spectrum by *Auto Regressive Models* (ARM) on the TGC-compensated RF signal. As suggested in [13], the used order is 10. The power spectrum corresponding to each point is then computed by a sliding window as in [3][13]. A typical power spectrum is showed in Figure 2(a).

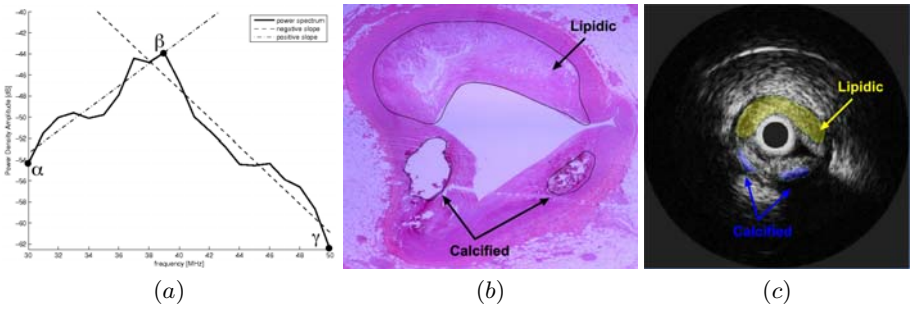


Fig. 2. (a) Spectral Features in band *B* of power spectrum; (b) Atherosclerotic plaques in histological image and (c) corresponding plaques in IVUS image

2.2 Features Extraction

Given the IVUS reconstructed image, we extract a wide set of visual features by applying Gabor filters, Local Binary Patterns and by computing the shading of the polar image, as in [3].

Given the central frequency of the catheter $f_C = 40$ MHz, we consider a 10 MHz band-side, limiting the spectral information to a band $B \in [f_L, f_H]$ ($f_L = 30$ MHz, $f_H = 50$ MHz). Being $S(f)$ the power spectrum function and defining $f_{MAX} = \arg \max_{f \in B}(S(f))$, we compute the straight lines fitting the curves in $[\alpha, \beta]$ and $[\beta, \gamma]$ (Figure 2(a)) and extract corresponding slopes and y-axis-intercept values. We also extract: $S(f_{MAX})$, $S(f_L)$, $S(f_H)$, $S(0)$ values and two other global measures: the *energy of A-line* and the *averaged amplitude of spectral component in B*.

The final feature vector is constructed concatenating textural and spectral features thus providing a description of both spatial and spectral IVUS data.

2.3 Enhancing In-Vitro Data for Tissue Characterization

Sequential Floating Forward Selection has been proposed in [18] as a sub-optimal features selection algorithm. In our case the problem can be formulated as follow: let us call $X = \{x_1, \dots, x_N\}$ the final training set, initially formed only by plaques from *in-vitro* cases, $Y = \{y_1, \dots, y_m\}$ the set of frames of *in-vivo* plaques and J_k a parameter related to the classification performance at iteration k .

The algorithm individually includes *in-vivo* plaques from frames $y \in Y$ into the set X while J increases. If $J_{k^*} < J_{k^*-1}$, already added plaques are temporary rested, one by one with re-insertion, and J is computed again. A combination of plaques returning a higher J is kept and temporary rested plaques are permanently discarded. The algorithm stops when $Y = \{\emptyset\}$ or when a certain number K of maximum iterations is reached.

In our method, J is defined as the weighted sensitivity over classes, to avoid the algorithm to improve (or keep constant) global classification quality by growing in sensitivity for some classes while getting worse other classes. For this reason, each weight is inversely proportional to the sensitivity of each class computed at the actual algorithm step k . In this way the class penalized at current iteration will have more weight at next step:

$$J_k = \sum_{l=1}^{N_c} w_k^l S_k^l,$$

where N_c is the number of classes, S_k^l and w_k^l are the sensitivity and the weighting of class l at iteration k , defined as

$$w_k^l = \frac{\log\left(\frac{1}{S_k^l}\right)}{\sum_{l=1}^{N_c} \log\left(\frac{1}{S_k^l}\right)}.$$

3 Experimental Results

Data acquisition. The IVUS equipment used in this work is a *Galaxy II IVUS Imaging System* (Boston Scientific) with a catheter *Atlantis SR Pro 40 MHz* (Boston Scientific), installed in the Hospital "Germans Trias i Pujol" of Badalona (Spain). RF data are collected using a *12-bit Acquiris acquisition card* ($f_s = 200MHz$). In this work we used a rotational catheter, generating 256 A-lines of 1024 samples each one.

In-vivo data are extracted during an hemodynamic intervention procedure. Once the procedure is over an *IVUS Pullback* is acquired by storing data while the probe is *pulled-back* at constant speed¹. The result is a sequence of images describing the cross-sectional vessel structure in which *in-vivo* plaques can be

¹ The catheter is connected to the IVUS equipment by a motorized tool and its position is constantly monitored by X-ray analysis.

labeled by expert physicians according on their professional experience. *In-vivo* data have been taken from 9 patients, plaques have been labeled by two expert physicians resulting in 49 *fibrotic*, 37 *lipidic* and 35 *calcified*: only regions in which both physicians agreed have been considered.

In-vitro data are extracted from a *post-mortem* case. The artery, separated from the heart, is first fixed on a mid-soft plane and filled (by a catheter) with physiological saline solution at constant pressure (around 120 *mmHg*), simulating blood pressure. The probe is then introduced and RF data are acquired in correspondence of plaques: the sampled position is marked on the external part of the artery as reference for histological analysis. When the acquisition is over, the artery can be studied by histological analysis. Observed plaques are then labeled in the corresponding IVUS images. *In-vitro* data have been extracted from 9 *ex-vivo* cases with 26 *fibrotic*, 14 *lipidic* and 31 *calcified* plaques.

Validation methodology. The *in-vitro* data set, enhanced by *in-vivo* selected data, is used to train an Adaboost classifier with up to 50 Decision Stumps [4], in the multiclass framework of Error Correcting Output Codes [1][7] parameterized using a *one-versus-one* coding with Euclidean decoding [5]. Classification performances - plaque sensitivity S and global accuracy A - are evaluated using the *Leave-One-Patient-Out* cross validation technique. Let us call N_p the number of clinical cases: to cross-validate the method, the training set is built taking at each time all *in-vivo* selected frames and all *in-vitro* cases, except one, used as test. The process is repeated N_p times and the confusion matrix C_k is computed at each round k to evaluate classification performance parameters. Global results are computed by averaging the N_p folds.

In our study the amount of *in-vivo* data included in the *in-vitro* dataset is varied from 10% up to 50%. Further inclusion is avoided for the sake of the training set consistency. Including more than 50% in fact could affect too much the validated dataset.

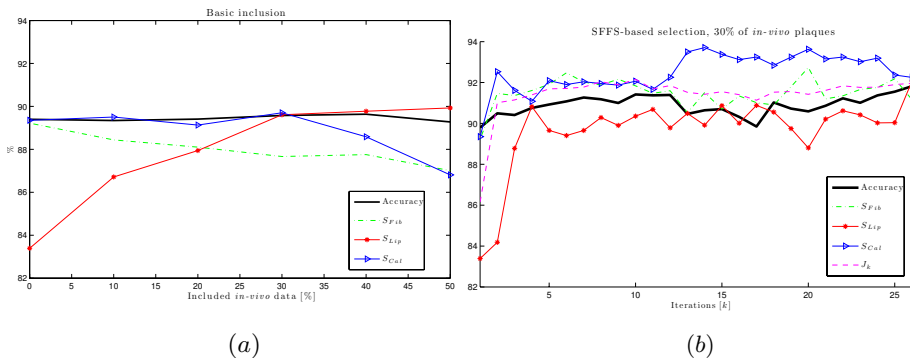


Fig. 3. (a) Basic-inclusion and (b) SFFS-based inclusion results

Comparisons. The *in-vivo/in-vitro* IVUS data fusion is a novel technique. As such, there are no related methods that directly address this problem. Thus, we compare the method with two plausible approaches. The first one is a *basic inclusion* of the whole *in-vivo* into the *in-vitro* dataset, according to labels given by physicians, at different percentages.

Since the idea of the SFFS-based data fusion is close to the *semi-supervised* learning, the second approach consists in defining an extended training set using the labels provided by means of the *Low Density Separation* (LDS) [19] technique. Based on cluster assumption, it essentially trains a *Transductive Support Vector Machine* (TSVM) by gradient descent on a graph-distance derived kernel. The graph is constructed by assigning to each edge a weight given by the Euclidean distance with each other feature point. The main idea is that two points in the same cluster present a continuous connecting curve passing through regions of high density, while two points of different clusters are connected with curves traversing a density valley. The similarity of two points is then defined by maximizing over all continuous connecting curves the minimum density along the connection. MATLAB implementation² of the method has been used in this paper.

Both *in-vitro* and *in-vivo* data have been used to construct the distance graph and the TSVM has been trained. *In-vivo* points are added (30%) to the training set according to classification results and performances have been computed by means of the *Leave-One-Patient-Out*. LDS parameters have been set as $(C, \rho) = (850, 0.05)$ by cross-validation.

Figure 3(a) shows the performance measures using *basic inclusion*. Observe that the sensitivity for the lipidic class and global accuracy increase with the amount of *in-vivo* data. However, the classification of *fibrotic* and *calcified* class is poorer. Note that the best performance is achieved when the amount of new included data is around 30%.

Figure 3(b), show the sensitivities and global accuracy for the SFFS-based inclusion method as described in section 2.3 (*in-vivo* plaques are included in the *in-vitro* training set according to segmentation of physicians). The curves in the figure are shown up to the iteration 26, where maximum accuracy is achieved. We can observe how, given initial conditions, plaque sensitivity measure increases up to a stable condition. Note that the lipidic class characterization performance, the most critical one at the initial point, grows as new plaques are added. Different percentages have been tested and 30% resulted to perform better. Observe that the same result is also reported in basic-inclusion approach.

Finally, Table 1 reports classification accuracy and sensitivity for the analyzed approaches, compared with the initial case (*NI*). The *basic-inclusion* only reports poor and unpredictable improvements, mainly due to the unsupervised nature of the data inclusion. The *LDS-based* method (LDS) only results in an improvement in the lipidic class sensitivity. This behavior can be explained by the fact that LDS looks for a minimum density path to cluster data. Since the lipidic plaque is underrepresented in the training set, the improvement in the corresponding

² <http://www.kyb.tuebingen.mpg.de/bs/people/chapelle/lds/>

Table 1. Performance of analyzed methods

	<i>NI</i>	<i>BI</i>	<i>LDS</i>	<i>SFFS</i>
<i>A</i>	89.39% (7.33)	89.57% (5.82)	84.96% (7.86)	91.82% (4.60)
<i>S_{fib}</i>	89.22% (9.38)	87.66% (8.58)	87.68% (9.53)	91.73% (7.52)
<i>S_{lip}</i>	83.38% (14.21)	89.61% (10.86)	89.87% (13.19)	92.54% (10.51)
<i>S_{cal}</i>	89.35% (12.85)	89.70% (9.33)	75.98% (22.39)	92.37% (3.19)

sensitivity is expected. However, this improvement comes at the cost of hindering *calcified* and *fibrotic* performances. The proposed *SFFS-based* approach improved all performance parameters of all the classes and outperforms other approaches in all the cases.

4 Conclusions

A framework for IVUS image reconstruction and power spectrum computation from raw radio frequency data extracted and validated from *in-vitro* cases has been presented as well as both textural and spectral features extraction process for robust tissue characterization.

In-vivo data acquisition process has been described and feature extraction has been performed.

It has been demonstrated that the inclusion of selected *in-vivo* data in the *in-vitro* training set by means of SFFS-based method improves both individual sensitivities and overall accuracy.

It has been also showed that the basic formulation of LDS algorithm is not useful to solve our problem, since the cluster assumption seems to be not appropriate to our dataset.

Given the promising results of SFFS-based algorithm, it is worth to study more in depth its potentials by including different classifiers and by comparing it with more semi-supervised learning methods.

Acknowledgment

This work was supported in part by a research grant from projects TIN2006-15308-C02, FIS-PI061290, MI1509-2005 and CONSOLIDER - INGENIO 2010 (CSD2007-00018).

References

1. Allwein, E.L., et al.: Reducing multiclass to binary: A unifying approach for margin classifier. *Journal of Machine Learning Research* 1, 113–141 (2000)
2. Burke, A.P., et al.: Coronary risk factors and plaque morphology in men with coronary disease who died suddenly. *The New England Journal of Medicine* 336(18), 1276–1282 (1997)

3. Caballero, K.L., et al.: Using Reconstructed IVUS images for Coronary Plaque Classification. In: Proceedings of the 29th Annual International Conference of the IEEE EMBS
4. Caballero, K.L.: Coronary Plaque Classification Using Intravascular Ultrasound Images and Radio Frequency Signals. Universitat Aut3noma de Barcelona (2007)
5. Caballero, K.L., Barajas, J., Pujol, O., Salvatella, N., Radeva, P.I.: In-Vivo IVUS Tissue Classification: A Comparison Between RF Signal Analysis and Reconstructed Image. In: Mart3nuez-Trinidad, J.F., Carrasco Ochoa, J.A., Kittler, J. (eds.) CIARP 2006. LNCS, vol. 4225, pp. 137–146. Springer, Heidelberg (2006)
6. DeMaria, A.N., et al.: Imaging vulnerable plaque by ultrasound. *Journal of the American College of Cardiology* 47(8), C32–C39 (2006)
7. Dietterich, T.G., Bakiri, G.: Solving multiclass learning problems via error-correcting output codes. *Journal of Artificial Intelligence Research* 2, 263–286 (1995)
8. Ehara, S., et al.: Spotty calcification typifies the culprit plaque in patients with acute myocardial infarction: An intravascular ultrasound study. *Circulation* 110, 3424–3429 (2004)
9. Filho, E.D., et al.: A study on intravascular ultrasound image processing (2005)
10. Friedman, J., Hastie, T., Tibshirani, R.: Additive logistic regression: a statistical view of boosting. *Annals of Statistics* 28(2000) (2000)
11. Moore, M.P., et al.: Characterization of coronary atherosclerotic morphology by spectral analysis of radiofrequency signal: in vitro intravascular ultrasound study with histological and radiological validation. *Heart* 79, 459–467 (1998)
12. Murashige, A., et al.: 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 (2005)
13. Nair, A., et al.: Coronary plaque classification with intravascular ultrasound radiofrequency data analysis. *Circulation* 106, 2200–2206 (2002)
14. Proakis, J., Rader, C., Ling, F., Nikias, C.: *Advanced Digital Signal Processing*. McMillan, NYC (1992)
15. Pujol, O.: A Semi-Supervised Statistical Framework and Generative Snakes for IVUS Analysis. Universitat Aut3noma de Barcelona (2004)
16. Schapire, R.E.: The boosting approach to machine learning: An overview (2002)
17. Zhang, X., McKay, C.R., Sonka, M.: Tissue characterization in intravascular ultrasound images. *IEEE Transaction on Medical Imaging* 17(6), 889–899 (1998)
18. Pudil, P., Ferri, F.J., Kittler, J.: Floating search methods for feature selection with nonmonotonic criterion functions. *IAPR*, 279–283 (1994)
19. Chapelle, O., Zien, A.: Semi-Supervised Classification by Low Density Separation. In: Proceedings of the Tenth International Workshop on Artificial Intelligence and Statistics, pp. 57–64 (2005)
20. Zhu, X.: Semi-supervised learning literature survey. Computer Sciences Technical Report TR 1530, University of Wisconsin-Madison (2008)