33rd Annual International Conference of the IEEE EMBS
Boston, Massachusetts USA, August 30 - September 3, 2011

Feature selection applied to ultrasound carotid images segmentation

Samanta Rosati, Filippo Molinari, *Member, IEEE,* and Gabriella Balestra, *Member, IEEE*

Abstract—**The automated tracing of the carotid layers on ultrasound images is complicated by noise, different morphology and pathology of the carotid artery. In this study we benchmarked four methods for feature selection on a set of variables extracted from ultrasound carotid images. The main goal was to select those parameters containing the highest amount of information useful to classify the pixels in the carotid regions they belong to.**

Six different classes of pixels were identified: lumen, lumen-intima interface, intima-media complex, media-adventitia interface, adventitia and adventitia far boundary. The performances of QuickReduct Algorithm (QRA), Entropy-Based Algorithm (EBR), Improved QuickReduct Algorithm (IQRA) and Genetic Algorithm (GA) were compared using Artificial Neural Networks (ANNs).

All methods returned subsets with a high dependency degree, even if the average classification accuracy was about 50%. Among all classes, the best results were obtained for the lumen. Overall, the four methods for feature selection assessed in this study return comparable results. Despite the need for accuracy improvement, this study could be useful to build a pre-classifier stage for the optimization of segmentation performance in ultrasound automated carotid segmentation.

I. INTRODUCTION

Carotid artery (CA) intima-media thickness (IMT) is commonly deemed as one of the risk marker for cardiovascular diseases (CVDs) [1].

Ultrasound imaging is the most used procedure in order to assess the IMT. In the past years several approaches for ultrasound image analysis were developed, both in terms of segmentation techniques and classification methods [2].

The main problem encountered in ultrasound CA imaging is represented by the images variability introduced by noise, vessel morphology and pathology. Particularly, ultrasound images are principally affected by speckle noise, an interference caused by multiple scattering of ultrasound wave, which perhaps represents the most prominent limit to image perception and numerical processing [3].

All these factors complicate the automated tracings of carotid layers, thus making user interaction often required in order to improve the segmentation performance. In this context it becomes a crucial point the individuation of a limited number of pixels features, which can be used to automatically identify the carotid regions the pixels belong to.

Feature selection is a procedure allowing dimensional reduction of multivariate data, deleting the redundant attributes, in order to extract from a high-dimensional dataset the features with most significant information. Moreover, a too large number of features does not necessarily allow increasing the classification accuracy: several attributes may be irrelevant or, even worse, may introduce some kind of noise which decreases the classifier performances.

There are several different approaches to perform feature selection. All methods need an appropriate and well-defined criterion to measure the relevance of the chosen features. However, as the number of initial features is usually large, it is computationally impossible to test all possible subsets of them, even if the criterion is simple to evaluate. A heuristic procedure is than applied to find a good set of features in a reasonable amount of time. Another important difference among the possible feature selection methods is the hypothesis that the system is linear that some of them require in order to give good results. As most real situations are non linear, it may happen that two features that are useless taken individually will become highly predictive used together. When there is not information on linearity it is important to choose a method that does not require it.

The main idea of the study presented here is to start calculating a large and overabundant amount of parameters extracted from ultrasound carotid images and then select the most important ones for the pixels classification through a feature selection method. In particular we describe and compare the performances of four different feature selection methods applied to the extracted parameters.

II. MATERIALS AND METHODS

We used a dataset of 300 B-Mode longitudinal carotid images. One hundred images were acquired at the Neurology Division of Nicosia (Cyprus) from 100 healthy subjects (age: 54±24, 60 males) and 200 from the Neurology Dept. of the Gradenigo Hospital of Torino (Italy) from 150 patients (age: 69±16, 97 males). All the images were discretized on 8 bits (grayscale from 0 to 255) and digitally sent to a computer. The conversion factor for Nicosia images was 0.06 mm/pixel, that of Torino was 0.0625 mm/pixel. Our database comprised both normal carotids (*i.e.*, with an IMT lower than 0.85 mm and no plaques) as well as pathologic vessels. Also, we had straight and horizontal, curved, and inclined arteries. Finally, we had images with a good signal-to-noise ratio and also images with a high degree of blood backscattering.

S. Rosati, F. Molinari, and G. Balestra are with the BioLab, Department of Electronics, Politecnico di Torino, Corso Duca degli Abruzzi 24, 10129, Torino, Italy

(G. Balestra is the corresponding Author; phone: +39-11-564-4135; fax: +39-11-564-4217; e-mail: gabriella.balestra@polito.it)

978-1-4244-4122-8/11/$26.00 ©2011 IEEE

A. Feature Extraction

In order to build the dataset used for feature selection, we started identifying six different classes of pixels, according to their physiological meaning: lumen, lumen-intima interface, intima-media complex, media-adventitia interface, adventitia and far adventitia boundary.

Among the whole set of available images, 50 of them were randomly selected and, from every image, ten pixels per class were randomly chosen, for a total of 3000 pixels. For each single pixel, other than its intensity, we considered as features different parameters essentially based on the intensity of the pixels around it and belonging to two categories: statistical moments estimates and texture features. Texture gives important information that humans use in analyzing a scene [4]. Particularly, texture features are a set of digital parameters based on the spatial displacement of the intensity levels in an image. They are based on the *Grey Level Co-occurrence Matrix* (GLCM) [5].

A list of the image descriptors used in this work is given below:

1) Intensity of the single pixel.
2) Statistical moments: mean value, standard deviation, skewness, and kurtosis.
3) Spatial Gray Level Dependence Method (SGLDM) [6] with a displacement $\delta=(0,1)$: energy, contrast, homogeneity, entropy, and first, third and fourth order moments.
4) Gray Level Difference Method (GLDM) [6] with a displacement $\delta=(0,1)$: contrast, angular second moment, entropy, and mean.
5) Gray Level Run Length Matrix (GLRLM) [6] in directions θ equal to 0°, 45°, 90° and 135°: short run emphasis, long run emphasis, gray level distribution, run length distribution, and run percentage.

Each of the above described features was calculated on four different areas centered on the selected pixel, with sizes 7x15, 15x7, 7x3, and 3x7 pixels. In this way we obtained a total of 141 features for all analyzed pixels.

B. Feature Selection

Four feature selection methods were applied to our data set, all employing the *Rough-Set Theory* (RST) concepts. RST, as defined by Pawlak [7], is a powerful tool to model imperfect and incomplete knowledge, which does not require any a-priori information or model assumptions about data. The basic principle of RST says that if two objects are indiscernible with respect to a certain variable, then they should be classified in the same class. RST defines a *decision system* as composed of a nonempty set of objects (the Universe *U*) and a nonempty set of attributes *A*. The latter is made up of a certain number of *conditional attributes C*, which represent the input features, and a *decision attribute D*, which is the class the objects belong to. According to RST, each $X \subseteq U$ can be divided into two disjoint subsets, named *lower* and *upper approximations*, using only the information contained in P. The *P-lower approximation* of *X* ($\underline{P}X$) is the complete set of objects certainly belonging to the target set *X,* according to the information carried on *P*, while the *P-upper approximation* of *X* ($\overline{P}X$) includes the objects of *U* which may possibly belong to *X*. The couple ($\underline{P}X$, $\overline{P}X$) defines a rough set.

In the past years RST has found wide and different areas of application, such as machine learning [8], knowledge acquisition [9], [10], decision analysis [11], [12], pattern recognition [13], knowledge discovery from databases and expert systems [14]. Recently, feature selection has been one of the most important fields in which RST has been employed, with very satisfactory results.

As the RST can only work with discrete data and in this study the dataset was made up of continuous values, a discretization strategy was needed. For each variable, different intervals of values have been identified, enabling the passage from continuous values to a number of discrete elements. The discretized dataset has then been used for the feature selection and evaluated by means of the dependency degree measure. The *dependency degree* $\gamma_C(D)$ measured between a decision attribute *D* and the subset of conditional features *C* is equal to 1 if all values from *D* are uniquely determined by values of attributes *C* [15]. In this case the dataset is defined as consistent. Real data set are usually not consistent so the maximum value is less than 1, in our case the maximum value is 0.99.

QuickReduct Algorithm (QRA), introduced and depicted in [16], is a basic tool allowing to resolve reduct search problems without generating all the possible subsets. It is based on the *dependency degree* $\gamma_C(D)$ measured between a decision attribute *D* and the subset of conditional features *C*. The main idea is to add to the reduct subset those attributes producing a larger increase in the dependency degree.

A further easy feature selection algorithm, with structure similar to the QRA, is the *Entropy-Based Reduction* (EBR) developed from [17]. It is based on the measure of *conditional information entropy* $H(D|A)$ produced by an attribute *A* with respect to the decision feature *D*. The algorithm structure results similar to QRA in which those features resulting in a higher decrease of entropy are added to the current subset.

Improved QuickReduct Algorithm (IQRA) [18] is based on a generalization of the standard RST named *Variable Precision Rough Set* (VPRS) theory and introduced by Ziarko in [19]. This new approach allows surpassing some of the RST limits admitting a certain degree of objects misclassification. IQRA basic idea is similar to QRA: it adds to the reduct subset those attributes which induce the greatest increment in the dependency degree. If there is no increase of dependency during an iteration, the dependency degree with tolerance β ($\gamma_{C,\beta}(D)$) is calculated. Moreover the algorithm takes into account the redundant elements, deleting them from the analyzed dataset.

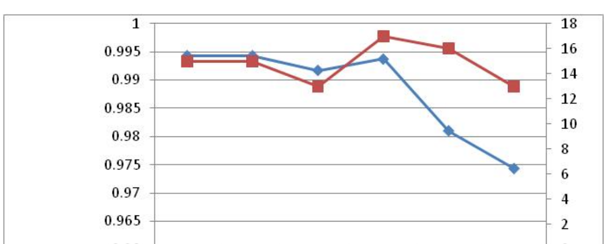
Genetic Algorithms (GAs), introduced by Holland [20] and belonging to the evolutionary algorithms, are a class of metaheuristics which wants to mime the natural evolutionary process of species. In GAs, each individual (*chromosome*) represents a possible solution of the problem and a fitness function associated to each individual represents the solution goodness. The main aim of GA is to optimize the population so that there is an increment in the fitness value during iterations. In this way a final solution most similar to the best one can be found. In [21] the *Genetic Rough-Set Attribute Reduction* (*GenRSAR*) is introduced, that is an algorithm employing a genetic search strategy in order to determine rough set reduct. It uses a standard GA structure in which the fitness function considers both the size of subset R and its suitability in terms of dependency degree.

MATLAB environment was used in order to implement all procedures for feature selection. As for the GA algorithm, an initial population of 100 random generated individuals, a probability of mutation and crossover sets to 0.4 and 0.6 respectively, and a number of generations equal to 100 were set, as suggested in [21].

C. Testing the different approaches

The performances of the different subset were compared using an artificial neural network (ANN). The basic idea was that a good procedure of feature selection allows removing redundant features so that the reduct provides the same quality of classification of the original set [22] or even improve it.

Specifically we built a similar network for each method. The number of input neurons was equal to the number of selected features, then the inputs were their values. About the ANN structure, we chose two hidden layers with a number of neurons approximately equal to 2/3 and 1/3 of the input neurons. As for the neuron activation functions, we used a logarithmic sigmoid function for the hidden layers and a linear function for the output layer. Back-propagation was chosen as the learning algorithm and the mean squared error was used as performance function. The initial values of interconnection weights were set randomly. As we only wanted to have a tool to compare the performances of the two feature selection methods we did not optimized the parameters of the ANNs.



	QRA	EBR	IQR (13feat)	IQR (17feat)	GA 10cp	GA 1cp
Gamma	0.9943	0.9943	0.9917	0.9937	0.9810	0.9743
No of features	15	15	13	17	16	13

Fig. 1. Feature selection results. The blue line represents the γ_C value for each extracted reduct while the red one shows the number of features selected by every procedure.

A second data set, built in the same way as the first, was used as testing set and to compare the performances of the methods with different training sets. We also used a training data set made of the sum of the samples of the two dataset to investigate the influence of the dimension of the data set on the performances of the network.

III. Results and Discussion

The feature selection results in terms of γ_C and number of features selected are reported in Fig. 1.

All methods return a number of selected features between 13 and 17 and all the reducts allow obtaining a very high dependency degree, even if GA has a lower performance. In two cases, for QR and EBR, the dependency is equal to the maximum obtainable value. The number of selected features does not correlate with the dependency degree.

Looking at the selected features it can be observed that they belong to all the different areas, and there is reasonable agreement among the selected subsets in the discarded features: 93 features out of 141 are never selected.

Fig. 2 presents the results of applying the ANNs based on the different features subsets in terms of percentage of correct classification. The results are independent from the data set and from its dimension. Also using one data set as training set and the other as test set gives approximately the same results.

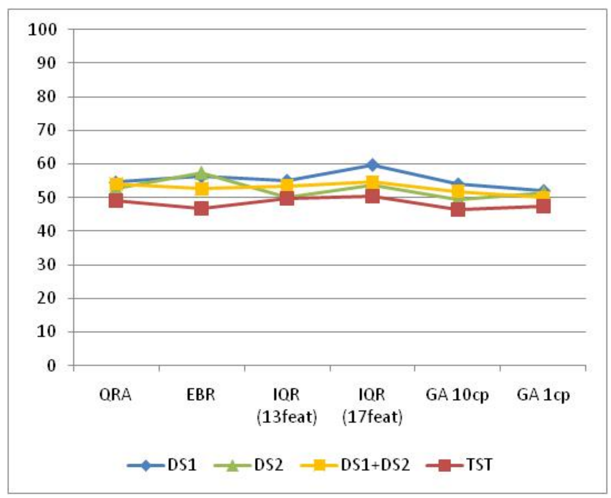


Fig. 2. Results of applying the ANNs based on the different features subsets in terms of percentage of correct classification.

The overall average classification accuracy is low. This is due to the great difference that there is in the classification results of each class.

Fig. 3 shows the difference in the correct classification of pixels belonging to the different classes using the training example of the first data set. It can be noted that while more than 90% of lumen pixels are classified in the right class, the percentage of correct classification decreases to around 50% for the other classes. This result is encouraging, because a wide number of automated segmentation algorithms perform lumen detection in order to *i)* locate the carotid artery in the

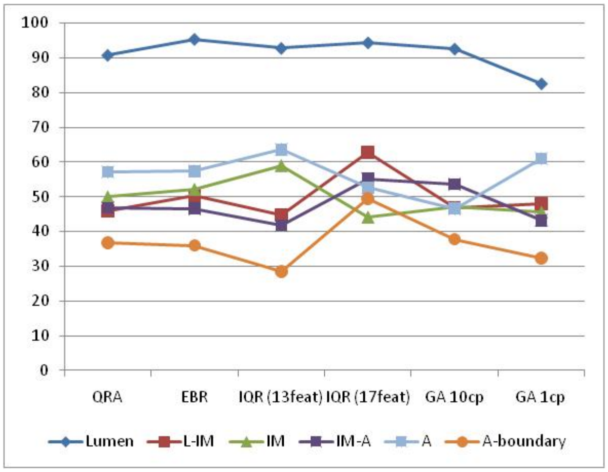


Fig. 3. Percentage of correct classification of pixels belonging to the different classes using the training example of the first data set.

image frame [23] or *ii)* to drive the automated segmentation of the lumen-intima boundary (which is the most critical, because characterized by a lower intensity of the gradient and usually by a low signal-to-noise ratio) [24]. Also, the refined and automated lumen detection could be exploited in plaque imaging, since the automated recognition and segmentation of atherosclerotic plaques in ultrasound images is still difficult.

The overall system performance shows that the output of different selection approaches are comparable in terms of selected features. There is still an overt difference among the classification accuracies of pixel belonging to different classes. We are now extending this study by investigating both the possibility of adding new features to the database and using different type of classifiers.

Nevertheless, this study could pave the way for the development of an automated image pre-classifier, which could be used for optimizing the segmentation strategy. In fact, noise plays a different role in different segmentation techniques and the choice of the optimal technique for a given image is believed a possible way to bring the performance of fully automated segmentation techniques close to those of user-driven methods.

REFERENCES

[1] P. J.Touboul, M. G. Hennerici, S. Meairs, et al., "Mannheim intima-media thickness consensus," *Cerebrovascular Diseases,* vol. 18, pp. 346-349, 2004.

[2] E. C. Kyriacou, M. Pattichis, C. Loizou, C. Christodoulou, S. K. Kakkos and A. Nicolaides, "A review of noninvasive ultrasound image processing methods in the analysis of carotid plaque morphology for the assessment of stroke risk," *IEEE Transactions on Information Technology in Biomedicine*, vol. 14, pp. 1027-1038, July 2010.

[3] F. Molinari, G. Zeng, J. S. Suri, "A state of art review on intima-media thickness (IMT) measurement and wall segmentation techniques for carotid ultrasound," *Computer Methods and Programs in Biomedicine,* vol. 100, pp. 201-221, Dec. 2010.

[4] M. Amadasun and R. King, "Textural features corresponding to textural properties," *IEEE Transactions on Systems, Man and Cybernetics*, vol. 19, pp. 1264-1273, Sep. 1989.

[5] J. H. Tan, E. Y. K. Ng, U. Rajendra Acharya and C. Chee, "Study of normal ocular thermogram using textural parameters," *Infrared Physics and Technology,* vol. 53, pp. 120-126, Mar. 2010.

[6] R. W. Conners and C. A. Harlow, "A Theoretical comparison of texture algorithms," *IEEE Transactions on In Pattern Analysis and Machine Intelligence,* vol. 2, pp. 204-222, May 1980.

[7] Z. Pawlak, "Rough sets," *International Journal of Computer and Information,* vol. 11, pp. 341-356, Oct. 1982.

[8] H. Moradi, J. W. Grzymala-Busse and J. A. Roberts, "Entropy of english text: experiments with humans and a machine learning system based on rough sets," *Information Sciences*, vol. 104, pp. 31-47, Jan. 1998,

[9] L. Feng, G. Y. Wang and X. X. Li, "Knowledge acquisition in vague objective information systems based on rough sets," *Expert Systems*, vol. 27, pp. 129-142, May 2010.

[10] Y. Matsumoto and J. Watada, "Knowledge acquisition from time series data through rough sets analysis," *International Journal of Innovative Computing, Information and Control*, vol. 5, pp. 4885-4897, Dec. 2009.

[11] S. Greco, B. Matarazzo, and R. Slowinski, "Rough sets theory for multicriteria decision analysis," *European Journal of Operational Research*, vol. 129, pp. 1-47, Feb. 2001.

[12] Z. Pawlak and R. Sowinski, "Rough set approach to multi-attribute decision analysis," *European Journal of Operational Research*, vol. 72, pp. 443-459, Feb. 1994.

[13] R. W. Swiniarski and A. Skowron, "Rough set methods in feature selection and recognition," *Pattern Recognition Letters*, vol. 24, pp. 833-849, March 2003.

[14] S. Tsumoto, "Automated extraction of medical expert system rules from clinical databases based on rough set theory," *Information Sciences*, vol. 112, pp. 67-84, 1998.

[15] R. Jensen and Q. Shen, *Computational intelligence and feature selection: rough and fuzzy approaches.* Hoboken : Wiley, 2008.

[16] Q. Shen, and A. Chouchoulas, "Modular approach to generating fuzzy rules with reduced attributes for the monitoring of complex systems," *Engineering Applications of Artificial Intelligence*, vol. 13, pp. 263-278, June 2000.

[17] R. Jensen and Q. Shen, "A rough set–aided system for sorting www bookmarks," *Lecture Notes in Computer Science,* vol. 2198, pp. 95-105, 2001.

[18] P. S. V. S. Sai Prasad and C. Raghavendra Rao, "IQuickReduct: An Improvement to Quick Reduct Algorithm," *Lecture Notes in Computer Science*, vol. 5908, pp. 152-159, 2009.

[19] W. Ziarko, "Variable precicion rough set model," *Journal of computer and system sciences,* vol. 46, pp. 39-59, Feb. 1993.

[20] J. Holand, *Adaptation in natural and artificial systems.* Ann Arbor: University of Michigan Press, 1975.

[21] R. Jensen and Q. Shen, "Finding rough set reducts with ant colony optimization," in *Proceeding of the 2003 UK Workshop on Computational Intelligence*, pp. 15-22.

[22] Y. Chen, D. Miao, R. Wang, and K. Wu, A rough set approach to feature selection based on power set tree, Knowl-Based Syst, 24, pp. 275-281 (2011).

[23] A. C. Rossi, P. J. Brands, and A. P. Hoeks, "Automatic recognition of the common carotid artery in longitudinal ultrasound B-mode scans," *Med Image Anal,* vol. 12, no. 6, pp. 653-65, Dec, 2008.

[24] F. Molinari, W. Liboni, P. Giustetto, S. Badalamenti, and J. S. Suri, "Automatic computer-based tracings (ACT) in longitudinal 2-D ultrasound images using different scanners," *Journal of Mechanics in Medicine and Biology*, vol. 9, no. 4, pp. 481-505, 2009.