Optimization of arterial age prediction models based in pulse wave

A G Scandurra¹, G J Meschino¹, L I Passoni¹, A L Dai Pra², A R Introzzi¹ and F M Clara¹

¹ Bioengineering Laboratory, Electronic Department, Mar del Plata University, Argentine
² Engineering Aplied Artificial Intelligence Group, Mathematics Department, Mar del Plata University, Argentine

E-mail: scandu@fi.mdp.edu.ar

Abstract. We propose the detection of early arterial ageing through a prediction model of arterial age based in the coherence assumption between the pulse wave morphology and the patient's chronological age. Whereas we evaluate several methods, a Sugeno fuzzy inference system is selected. Models optimization is approached using hybrid methods: parameter adaptation with Artificial Neural Networks and Genetic Algorithms. Features selection was performed according with their projection on main factors of the Principal Components Analysis. The model performance was tested using the bootstrap error type .632E. The model presented an error smaller than 8.5%. This result encourages including this process as a diagnosis module into the device for pulse analysis that has been developed by the Bioengineering Laboratory staff.

1. Introduction
The most traditional variables to assess cardiovascular risk are the arterial pressure, life habits and blood lipids analysis. Without discarding these, the analysis of pulse wave is another factor to consider in the evaluation and early detection of cardiovascular illnesses [1].

The design of the equipment for acquisition and storage of the pulse wave was previously published and also the clustering of normal patterns in groups according with ageing were presented [2]-[7].

Changes of pulse wave shape with natural ageing and certain hypertensive processes has been demonstrated [8]. By analyzing pulse wave morphology, patient “arterial age” can be estimated and also it could be compared with its “chronological age.”

Instead of wave classification schemes, we propose in this paper the design of an arterial age prediction model.

The features set that feed the prediction model are extracted from the pulse wave shape. In order to reduce the dimension of this descriptor ensemble, the Principal Components Analysis (PCA) helped to pick the most relevant variables.

Different techniques for predictor design were evaluated: Linear Models, Artificial Neural Networks and Fuzzy Inference Systems. The models optimization was performed through hybrid
approaches such as parameters adaptation with Artificial Neural Networks (ANN) and Genetic Algorithms (GA).

The estimation of the arterial age according to a coherence assumption between the pulse wave morphology and the patient chronological age was performed. The data set became from a group of normotensive males between second and seventh decade.

In order to evaluate the model performance different error algorithms such as the Bootstrap, Leave-One-Out and.632E Error [9]-[11] were computed. The optimized model showed an error less than 8.5%. This result encourages including this process as a diagnosis module into the device for pulse analysis that has been developed by the Bioengineering Laboratory staff.

2. Materials and Methods

2.1. Data Acquisition and Storage
A non-invasive method to acquire and register the diameter variation of the radial artery (ADV) was used. The sensor is a capacitive transducer attached to a bandage wrist applied to the palpation zone of the radial pulse. The capacity changes produced by radial arterial diameter variations are converted through an analog to digital interface and then acquired by a personal computer system.

The morphological analysis of the signals was performed jointly with a medical field expert. A group of 11 features was considered: 9 features became from morphological characteristics of pulse waves and both remaining were the Systolic (SP) and Diastolic pressure (DP).

The next signal features were extracted: $BE_h$ (systolic pulse wave width measured at 70% to the systolic peak), $CD_h$ (systolic wave width measured at 90% to the systolic peak), $AF_v$ (vertical distance from horizontal axis to the beginning of reflected systolic wave), $AH_v$ (vertical distance from horizontal axis to incisure point), $AI_v$ (vertical distance from horizontal axis to diastolic peak), $FG_v$ (vertical distance from the beginning to the peak of reflected systolic wave), $HI_h$ (horizontal distance from the beginning to the peak diastolic pulse wave), $(HI_v$ (vertical distance from the beginning to the peak diastolic pulse wave), pod (ascending slope of the diastolic wave). The “h” indicates values measured in the horizontal axe and “v” those measured in the vertical axe.

In the figure1 the points which define the extracted features can be observed.

![Figure 1. Relevant points for the morphological analysis.](image)
2.2. Fuzzy Model
One of the most important applications in the Fuzzy Set Theory is the rule based systems. In the literature two different classes from these systems were presented, the Mamdani [12] and the Takagi-Sugeno-Kang (TSK) [13], [14]. The main difference between them is the consequent structure. The interpretability or the accuracy defines the used of one or another model.

In the TSK fuzzy model the rules consequents are not based on linguistic variables (such as in a Mamdani type fuzzy system). Instead of, they are linear functions of input variables. The rules syntaxes are as follows:

\[ R_j : \text{if } x_1 \text{ is } c_{1j} \quad \text{and } x_2 \text{ is } c_{2j} \quad \text{and} \ldots \text{and } x_n \text{ is } c_{nj} \quad \text{then } y \text{ is } f_j(x_1, \ldots, x_n) \]

where \( j \) is the rule number, \( x_1, x_2, \ldots, x_n \) are the input variables and \( y \) is the output variable. Each rule describes fuzzy areas of the input space through \( f_j(x_1, \ldots, x_n) \) functions. The fuzzy areas are characterized by membership values to fuzzy sets. They are defined by \( \mu_i \) functions, mapping \( x_i \) in the \([0,1]\) real range. The \( y \) output value is obtained through average weight of the \( f_j(x_1, \ldots, x_n) \).

The subspaces are defined through fuzzy clustering methods. In this particular work, fuzzy clusters were optimized with genetic algorithms.

2.3. Genetic Algorithms
The AG’s are search algorithms based on the biological evolution principles. The AG’s can be applied for solving an optimization problems in which the objective function is discontinuous, non differentiable, stochastic or highly nonlinear.

GA maintains a population of individuals that evolve according to rules of selection and genetic operators, such as reproduction, crossover and mutation. GA begins with a population that consists in randomly created individuals (possible solutions) and repeatedly modifies this population "evolving" towards an optimal solution.

Each individual in the population is assigned a measure of its fitness in the environment. Reproduction focuses its attention on high fitness individuals, thus exploiting the available fitness information. Crossover and mutation perturb those individuals, providing general heuristics for exploration. Although simplistic from a biologist's viewpoint, these algorithms are complex enough to provide robust (good performance across a variety of problem types) and powerful adaptive search mechanisms. The adaptive behavior of the GA depends on this feedback to drive the population towards better overall performance [15, 16].

Therefore, considering a particular problem, an ad-hoc evaluation or fitness function must be devised. As already known, GAs’ performance is a function of parameter settings [17].

3. Results

3.1. Signal processing
For this work, 55 registers corresponding to normotensive male individuals was processed. They were between 18 and 67 years old (second and seventh decades). The study was approved by Bioethics Committee of the Mar del Plata National University.

The selected patients were chosen by the field expert. The selection was based on clinical data, chronological age and coherency between chronological age and pulse wave morphology.

3.2. Features extraction
Principal Component Analysis (PCA) was used in order to determine the subset of arterial age prediction model features. This analysis was done using SPAD (Systeme Portable Pour L’analyse des Donnees) software.

Figure 2 shows the scatter plot of the two first principal components belonging to the complete data set. Cases exhibit spatial distribution according to age. The symbol size is proportional to age value.
By observing the scatter PCA plot it can be inferred that Factor 1 explains sufficiently the age variation of individuals. Numerical results show that 45.77 % of the variance is reached by the first principal component. Percentages corresponding to remain components are presented at Table 1.

Analyzing the projections of each feature onto the factors, figure 3, it was possible to select a set with those that shows the higher projections over factor 1. Hence the PCA analysis was useful to select the features to be included on the arterial age prediction model.

Table 1. Principal Component Analysis.

| # Factor | Explanation % | Cumulative % |
|----------|---------------|--------------|
| 1        | 45.77         | 45.77        |
| 2        | 14.18         | 59.95        |
| 3        | 13.27         | 73.22        |
| 4        | 11.09         | 84.30        |
| 5        | 7.48          | 91.79        |
| 6        | 3.96          | 95.75        |
| 7        | 2.32          | 98.07        |
| 8        | 1.17          | 99.24        |
| 9        | 0.62          | 99.86        |
| 10       | 0.14          | 100.00       |
| 11       | 0.00          | 100.00       |

The features that will feed the model were addressed observing figure 3 (variables projections onto factors): AF_v, AH_v, BE_h, CD_h (Factor 1 positive axe), HI_h, HI_v, pod (Factor 1 negative axe). Features named FG_v, PD y PS had no appreciable projection over Factor 1 and consequently, they were not selected.
The mentioned variables and their numerical Factor 1 projections can be observed at Table 2. Additionally, the selection state is indicated at third column.

**Table 2. Features with highest projections over factor 1.**

| Variable | Coordinate Factor 1 | Selected variable |
|----------|---------------------|-------------------|
| BE_h     | 0.81                | YES               |
| AF_v     | 0.94                | YES               |
| AH_v     | 0.91                | NO                |
| CD_h     | 0.66                | YES               |
| HI_h     | -0.76               | NO                |
| pod      | -0.76               | NO                |
| HI_v     | -0.85               | YES               |

By looking at figure 3 it can be deduced that HI_h, HI_v and pod are strongly correlated, as their vectors at the plane are practically overlapping. Only one of these features was selected, since the other two should only give redundant information. Also, we found a strong correlation between AF_v and AH_v, and again we discard one of them.

The PCA projection analysis jointly with the experts’ knowledge let us conclude that the set features of the prediction model will be: BE_h, CD_h, AF_v, HI_v.

### 3.3. Fuzzy model and optimization

A TSK fuzzy Inference prediction model was designed using the selected features. The subtractive clustering method [19] was used at the initial stage. Thus we obtained a model with four rules and Gaussian membership functions to define input fuzzy sets.

By evaluating the performance of this initial model, we found a relative mean absolute error (RMAE) of 10.6%. To improve it a model parameter optimization was proposed, keeping the model design.

At the optimization stage, genetic algorithms was used, in order to tune the centers and $\sigma$ of input Gaussian memberships. The model was build with 4 inputs; each input has a fuzzy set for each rule.
(16 fuzzy sets) and each fuzzy set is determined through 2 parameters. So, the GA solution is a 32 values vector.

The GA was configured with a population size of 50 individuals. The chosen selection function was stochastic uniform. In the successive generations, 70% of new individuals were generated by crossover and 30% by mutation.

The fitness function of the GA tries to minimize the .632E error [9]. To achieve this aim, the next steps are implemented:

**Resubstitution error computation**: a TSK model optimized with the parameters (membership functions centers and spreads) given by the GA was evaluated using the whole data set. Model error value was computed using the RMAE.

**Bootstrap error computation**: a different TSK model was generated, using the same parameters of the previous step, into a 100 times cycle. To find these models, a data set was generated by sampling with replacement. Each model was tested with the data not included at this sampling. Model error value was computed using the RMAE and the values obtained by 100 cycles was averaged.

**.632E error computation**: it was done used the next expression:

\[ .632E \text{ Error} = 0.368 \times \text{ResubstitutionError} + 0.632 \times \text{BootstrapError} \]

.632 Error gives a downwardly biased estimate of error with low variability, critically improving the resubstitution error, which is badly biased downwards estimator.

It has been demonstrated that .632E error is a good choice for model comparison and selection [11].

Performance of models based on different paradigms is shown for comparison at table 3: linear model, artificial neural network (multilayer perceptron), fuzzy inference systems and hybrid systems like ANFIS (Adaptive-Network-based Fuzzy Inference Systems) [20]. Error values are presented.

| Model                                                       | Error .632E |
|-------------------------------------------------------------|-------------|
| Non optimized TSK Model, 4 rules                            | 10.6 %      |
| Multilayer Perceptron, one hidden layer (Back-propagation)  | 10.5 %      |
| Multiple Linear Regression                                  | 10.4 %      |
| ANFIS (Adaptive Network-based Fuzzy Inference Systems)      | 10.4 %      |
| TSK Model, 4 rules with GA optimization                     | 8.4 %       |
| TSK Model, 5 rules with GA optimization                     | 8.2 %       |

We selected, as the best model, the optimized Fuzzy TSK with 4 rules, given that the increase of the quantity of rules produces no improvement in the obtained error.

Testing the model with data not being included in the training set, there were cases found where the predicted arterial age value differed in more than 20% from the chronological age. In these situations the medical expert assigned specific reasons for this early arterial ageing, therefore the model usefulness was validated.

### 4. Conclusions

In this work, we present a model for detecting early arterial ageing, based on a GA tuned TSK fuzzy model. We compare its performance with other models.

Features used as model input variables were obtained from pulse wave morphology. A small subset of them was selected using Principal Component Analysis.

The .632E Error was used as model selection criteria. The GA tuned model presented the best performance over the evaluated methods.
The obtained error allows us concluding that the model is adequate to estimate the patient arterial age. The comparison of the estimated value with chronological age can constitute a proper method to asymptomatic cardiovascular disease diagnosis assistance.

5. Acknowledgments
The authors thank Dra. Clotilde Ubeda (Instituto Nacional de Epidemiología), for her useful contribution on biostatistical themes.

6. References
[1] O’Rourke M, Pauca A and Giang X 2001 Pulse wave analysis J Clin Pharmacol 51 507-22
[2] Passoni L I, Meschino G J, Scandurra A G, Clara F M and Introzzi A R 2001 Redes neurales auto-organizadas en la exploración de relaciones de los cambios en la distensión de la arteria radial con el envejecimiento. In: XIII Congreso Argentino de Bioingeniería, (Tafi del Valle, Tucumán, Argentina)
[3] Passoni L I, Meschino G J, Scandurra A G, Clara F M and Introzzi A R 2001 Assessing features of the radial artery diameter variations using self-organizing maps. In: XIX Reunión de Trabajo en Procesamiento de la Información y Control (RPIC), (Santa Fe, Argentina pp 328-33)
[4] Passoni L I, Meschino G J, Scandurra A G, Clara F M and Introzzi A R 2002 Evaluación de las características de la variación del diámetro de la arteria radial por medio de mapas autoorganizados Revista Argentina de Bioingeniería 8 30-7
[5] Clara F M, Passoni L I, Scandurra A G, Meschino G J and Introzzi A R 2003 Detección de riesgo cardiovascular basada en registros de variación de diámetro arterial. In: XIV Congreso Argentino de Bioingeniería, (Córdoba, Argentina)
[6] Clara F M, Cayrol M L, Scandurra A G, Meschino G J, Garzillo M G, Moyano E and Introzzi A R 2005 La técnica de análisis de onda de pulso en la determinación de riesgo cardiovascular Revista de la Federación Argentina de Cardiología 34 213-20
[7] Clara F M, Scandurra A G, Meschino G J, Passoni L I and Introzzi A R 2005 Evaluación de la aumentación radial en base a registros de variación de diámetro arterial. In: XV Congreso Argentino de Bioingeniería, (Paraná, Argentina p 29)
[8] Introzzi A R, Passoni L I, Scandurra A G, Meschino G J and Clara F M 2005 Relación entre el envejecimiento y la morfología de los registros de variación de diámetro arterial: Clasificación con redes neuronales artificiales Revista Argentina de Cardiología 73 352-7
[9] Efron B and Tibshirani R 1997 Improvements on cross-validation: The .632+ bootstrap method Journal of the American Statistical Association 92 548-60
[10] Efron B 1983 Estimating the error rate of a prediction rule: Improvement on cross-validation Journal of the American Statistical Association 78 316-31
[11] Twomey J M and Smith A E 1998 Bias and variance of validation methods for function approximation neural networks under conditions of sparse data IEEE Transactions on Systems, Man, and Cybernetics, Part C 28 417-30
[12] Mamdani E H 1974 Application of fuzzy algorithms. For control of simple dynamic plant Proceedings of the Institution of Electrical Engineers 121 1585-8
[13] Takagi T and Sugeno M 1985 Fuzzy identification of systems and its application to modeling and control IEEE Transactions on Systems, Man, and Cybernetics 15 116-32
[14] Sugeno M and Kang G T 1988 Structure identification of fuzzy model Fuzzy Sets and Systems 28 15-33
[15] Koza J 1992 Genetic programming: On the programming of computers by mean of natural selection (Cambridge, MA)
[16] Michalewicz Z. 1992 Genetic algorithms + data structures = evolution programs (Berlin: Springer-Verlag)
[17] Schaffer J D, Caruana R, Eshelman L J and Das R 1989 A study of control parameters affecting online performance of genetic algorithms for function optimization In: 3rd International Conference on Genetic Algorithms, pp 51-60
[18] Jolliffe I T 2002 Principal component analysis. Springer.
[19] Chiu S 1994 Fuzzy model identification based on cluster estimation Journal of Intelligent & Fuzzy Systems 2 267-78
[20] Jang J S R 1993 Anfis: Adaptive network based fuzzy inference system IEEE Transactions on Systems, Man and Cybernetics 23 665-85