A Decision Support System for the Diagnosis of Coronary Artery Disease

M.G. Tsipouras, T.P. Exarchos, D.I. Fotiadis
Unit of Medical Technology and Intelligent Information Systems, Dept. of Computer Science, University of Ioannina, GR 451 10 Ioannina, Greece
markos@cs.uoi.gr, me01238@cc.uoi.gr, fotiadis@cs.uoi.gr

A. Kotsia, A. Naka, L.K. Michalis
Michaelideion Cardiology Center, GR 45110 Ioannina, Greece
Dept. of Cardiology, Medical School, University of Ioannina, GR 45110 Ioannina, Greece
hcsanna@cc.uoi.gr, anaka@cc.uoi.gr
lmihalis@cc.uoi.gr

Abstract

A rule-based Decision Support System is presented for the diagnosis of Coronary Artery Disease. The generation of the decision support system is realized automatically using a three stage methodology: (a) induction of a decision tree from a training set and extraction of a set of rules; (b) transformation of the set of rules into a fuzzy model and (c) optimization of the parameters of the fuzzy model. The system is evaluated using 199 subjects, each one characterized by 19 features, including demographic and history data, as well as laboratory examinations. Ten fold cross validation was employed and the average sensitivity and specificity obtained was 80% and 65% respectively. Our approach provides diagnosis based on easily acquired features and, since it is rule based, is able to provide interpretation for the decisions made.

1. Introduction

Coronary artery disease (CAD) is the development of atherosclerotic plaques in the coronary arteries, resulting in coronary luminal narrowing and subsequently occlusion. CAD is the leading cause of death in western countries. Coronary angiography (CA), the “gold standard” method for the diagnosis of CAD, is an invasive and costly procedure that cannot be used for the diagnosis of coronary atherosclerosis and screening of large populations. Therefore, several alternative methods have been proposed in the literature for the diagnosis of CAD. They can be divided into various categories, based on the type of data they use for subject characterization: (a) methods which employ the electrocardiogram (ECG) of the patient, extracting features from it, like the ST segment [1], the QT interval [2], the amplitude of T [3] or the R wave [4] and the Heart Rate Variability (HRV) [5]; (b) methods based on heart sounds associated with coronary occlusions [6,7]; (c) methods using medical images, such as, scintigraphy [8,9], (d) methods based on arterio-scillography [10]; (e) methods employing demographic, history and laboratory data (subject’s data) [11-13]; (f) methods that employ more than one type of data, such as ECG, scintigraphy and subject’s data [14,15].

Most of the methods are based on the analysis of data, obtained by examinations, such as stress echocardiography and myocardial scintigraphy, which are expensive, not widely available and suffer from technical limitations [16]. Exercise tolerance test is inexpensive and widely available but has a low sensitivity and specificity. Only few methods are based on the analysis of subject’s data, which can be easily obtained. As far as the methods used for the analysis of data, most of the proposed approaches are based on neural networks, thus they cannot provide interpretation for the decisions made. In addition CA is not used for the initial annotation (i.e. presence or absence of CAD), thus an uncertainty is imposed in the obtained results. Therefore, a method able to predict non-invasively the presence of CAD and able to provide interpretation for the classification decisions, would be of great clinical value.

In the current work we propose a decision support system (DSS) for the diagnosis of CAD. The DSS is automatically generated with a data-driven method, employing only demographic, history and laboratory data, using a three stage methodology (Fig. 1). In the first stage, a decision tree is induced from the dataset

Proceedings of the 19th IEEE Symposium on Computer-Based Medical Systems (CBMS'06)
0-7695-2517-1/06 $20.00 © 2006 IEEE

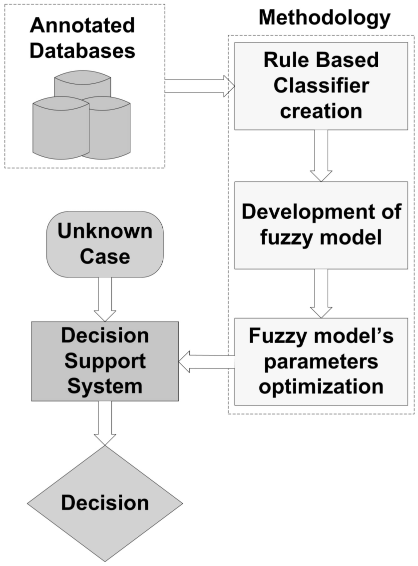

Figure 1. The three-stage methodology used to create the Decision Support System

using the C4.5 algorithm, and a set of rules is extracted from it. This set of rules forms a crisp rule-based classifier. The second stage involves the transformation of the crisp set of rules into a fuzzy model, using a fuzzy membership function and fuzzy equivalents of the binary operators. Finally, in the third stage, the parameters entering the fuzzy model are optimized, using a global optimization technique.

The system is evaluated using a dataset consisting of 199 subjects. CA, which is considered as the safest and most reliable examination for the assessment of a subject concerning CAD, was conducted to all subjects and their clinical status (presence or absence of CAD) was determined by two experienced cardiologists.

2. Materials and methods

The three stage methodology used to generate the DSS involves: (a) creation of a rule-based classifier; (b) development of a fuzzy model; (c) optimization of the fuzzy model's parameters.

2.1. Creation of a rule-based classifier

In the first stage of the methodology we construct a rule based classifier from the training records [17]. A rule based classifier is a technique for classifying records using a collection of "if…then…" rules. The rules for the model are crisp and form the set of rules, which is represented in a disjunctive normal form (DNF), $\left(r_1 \vee r_2 \vee ... r_k\right)$, where r_i are the classification rules or disjuncts. Each classification rule is expressed as: $r_i : \left(Cond_i\right) \rightarrow y$, where y is the predicted class. The left-hand side of the rule is the rule antecedent or precondition. It contains a conjunction of feature tests: $Cond_i = \left(a_1\ op\ \theta_1\right) \wedge \left(a_2\ op\ \theta_2\right) \wedge ... \wedge \left(a_m\ op\ \theta_m\right)$, where $\left(a_j, \theta_j\right)$ is an feature-value (threshold) pair and op is a comparison operator chosen from the set $\left(=, \neq, <, >, \leq, \geq\right)$. Each feature test $a_j\ op\ \theta_j$ is a conjunct. The right-hand side of the rule is the rule consequent, which contains the predicted class y_i.

A rule r covers a record x if the precondition of r matches the features of x. r is also said to be fired or triggered whenever it covers a given record.

The quality of a classification rule can be evaluated using measures such as coverage and accuracy. Given a data set D and a classification rule r, the coverage of the rule is defined as the fraction of records in D triggering rule r. On the other hand, its accuracy is defined as the fraction of records triggered by r whose class labels are equal to y.

There are two important aspects to consider when constructing the set of rules of a rule based classifier. The set of rules should be composed by: (a) mutually exclusive rules. The rules in a set are mutually exclusive if a record triggers only one rule. This property ensures that every record is covered by at most one rule in the set. (b) Exhaustive rules. A set of rules has exhaustive coverage if there is a rule for each combination of feature values. This property ensures that every record is covered by at least one rule in the set. Together, these properties ensure that every record is covered by exactly one rule.

A well known method to construct a set of rules that satisfies the two above aspects is to extract rules from a decision tree. Decision trees are a widely used classification technique. They represent the acquired knowledge in the form of a tree. The tree can be easily transformed to a set of rules with mutually exclusive and exhaustive rules.

The construction of the decision tree is implemented using the C4.5 inductive algorithm [18]. The essence of the algorithm is to construct a decision tree from the training data. Each internal node of the tree corresponds to a principal component, while each outgoing branch corresponds to a possible range of that component. The leaf nodes represent the class to be assigned to a sample. The C4.5 algorithm applies to a set of data and generates a decision-tree, which minimizes the expected value of the number of tests for the classification of the data.

The most important factor in the C4.5 algorithm is its ability to automatically select the feature which is appropriate at each node. The feature of each node is selected in order to divide input samples effectively. *Information gain* is used as a measure of effectiveness.

After the induction of the decision tree, we apply a pruning method to reduce the tree's size and complexity. There are two common methods for pruning: prepruning and post-pruning. In our problem we prefer the post-pruning method [18]. Post-pruning tends to give better results than prepruning since it makes pruning decisions based on a fully grown tree, unlike prepruning, which can suffer from early termination of the tree growing process. Post-pruning in our case is performed by replacing a subtree with a new leaf node whose class label is determined from the majority class of records associated with the subtree (subtree replacement). The subtree replacement was performed by calculating the pessimistic error. The confidence value was set to 25% and the minimum number of records per leaf was set to 2.

The crisp set of rules is created from the final decision tree as follows: (a) One condition is created for every leaf of the tree, by parsing the tree from the root node to the leaf. The tests encountered along the path form the conjuncts of the condition while the class label at the leaf node is assigned to the rule consequent:

$$Cond_i(A,\Theta) = c_{root}(a_{root},\theta_{root}) \wedge c_{n_j}(a_{n_j},\theta_{n_j}) \wedge \ldots \wedge c_{n_k}(a_{n_k},\theta_{n_k}), \quad (1)$$

where $Cond_i$ is a condition, $A=\{a_1,a_2,\ldots,a_{nf}\}$ is the feature vector, $\Theta=\{\theta_1,\theta_2,\ldots,\theta_{nf}\}$ is a vector containing all thresholds, nf is the number of features characterizing a record, a_j and θ_j are the feature and the threshold used in the conjunct j and $c_j(a_j,\theta_j)$ is the respective conjunct, expressed as: $c_j(a_j,\theta_j)=g_c(a_j,\theta_j)$, where g_c is the crisp membership function, defined as:

$$g_c^{inc}(a,\theta)=\begin{cases}0 & a\le\theta\\ 1 & a>\theta\end{cases} \text{ (increasing) or}$$
$$g_c^{dec}(a,\theta)=\begin{cases}1 & a\le\theta\\ 0 & a>\theta\end{cases} \text{ (decreasing),} \quad (2)$$

(b) A general rule is created for each class, using all the conditions $Cond_i(A,\Theta)$ having as consequent this class:

$$R_y(A,\Theta) = Cond_{j_1}(A,\Theta) \vee Cond_{j_2}(A,\Theta) \vee \ldots \vee Cond_{j_n}(A,\Theta), \quad (3)$$

where y is the class. These general rules comprise the crisp set of rules (CSR), which is exhaustive and mutually exclusive. Therefore, for each feature vector A, one of the general rules is true, thus defining its class.

2.2. Development of a fuzzy model

The crisp set of rules is transformed into a fuzzy model using a fuzzy membership function instead of the crisp one, and fuzzy equivalents of the binary *AND* ($\wedge$) and *OR* ($\vee$) operators [19]. The sigmoid function, defined as:

$$g_s^{inc}(a,\theta_1,\theta_2)=\frac{1}{1+e^{\theta_1(\theta_2-a)}} \text{ (increasing) or}$$
$$g_s^{dec}(a,\theta_1,\theta_2)=\frac{1}{1+e^{\theta_1(a-\theta_2)}} \text{ (decreasing),} \quad (4)$$

is used as fuzzy membership function, while the minimum and maximum operators are used as fuzzy equivalents for the binary *AND* and *OR*. According to these, the crisp conjuncts are transformed to fuzzy ones as: $c_j^f(a_j,\theta_{1,j},\theta_{2,j})=g_s(a_j,\theta_{1,j},\theta_{2,j})$, and the crisp conditions are transformed to fuzzy ones as:

$$Cond_i^f(A,\Theta^f)=min\begin{Bmatrix} c_{root}^f(a_{root},\theta_{1,root},\theta_{2,root}),\\ c_{n_j}^f(a_{n_j},\theta_{1,n_j},\theta_{2,n_j}),\ldots,\\ c_{n_k}^f(a_{n_k},\theta_{1,n_k},\theta_{2,n_k})\end{Bmatrix}, \quad (5)$$

where $\Theta^f=\{\theta_{1,1},\theta_{2,1},\theta_{1,2},\theta_{2,2},\ldots,\theta_{1,nf},\theta_{2,nf}\}$ is a vector containing all parameters used in the fuzzy model.

The resulting set of rules is no longer mutually exclusive since the crisp thresholds have been converted to fuzzy ones using the sigmoid function. A test record may be covered by more than one fuzzy rule. Therefore, we define a rule evaluation metric, which is required in order to decide which rule will classify a test record, in the case when two or more rules fire for a record. For this reason, the likelihood ratio was used to measure how "strong" a rule is:

$$p_i = 2\sum_{i=1}^{n_y} fr_i \log(fr_i / e_i), \quad (6)$$

where n_y is the number of classes, fr_i is the observed frequency of class i records that are covered by the rule, and e_i is the expected frequency of a rule that makes random predictions. A large p_i suggests that the number of correct predictions made by the rule is significantly larger than that expected by random guessing. Other metrics for rule evaluation could be considered, however this was preferred since it takes into account both the accuracy and the coverage of the rules. This metric is applied to each $Cond_i^f$. The general crisp rules are transformed to fuzzy ones as:

$$R_y^f\left(A,\Theta^f\right)=max\begin{Bmatrix} p_{j_1}\times Cond_{j_1}^f\left(A,\Theta^f\right)\ , \\ p_{j_2}\times Cond_{j_2}^f\left(A,\Theta^f\right)\ ,\ \ldots\ , \\ p_{j_n}\times Cond_{j_n}^f\left(A,\Theta^f\right) \end{Bmatrix}. \quad (7)$$

These fuzzy general rules comprise the fuzzy model:

$$M^f\left(A,\Theta^f\right)=arg\max_{y=1,\ldots,ny}\left(R_y^f\left(A,\Theta^f\right)\right), \quad (8)$$

where n_y is the number of classes. As shown in Eq (8), for each feature vector A, the fuzzy general rule with the higher value defines its class.

2.3. Fuzzy model's parameters optimization

The fuzzy model $M^f\left(A,\Theta^f\right)$ is optimized with respect to its parameters Θ^f, using a training dataset (D_{train}). Having X the confusion matrix:

$$X_{M^f\left(A,\Theta^f\right),y}= \# \ of\ patterns\ in\ y\ classified\ to\ M^f\left(A,\Theta^f\right), \quad (9)$$

the cost function, used for this purpose, is defined as:

$$F\left(\Theta,D_{train}\right)=\frac{1}{\left|D_{train}\right|}\sum_{i=1}^{n_y}X_{i,i}\ , \quad (10)$$

The optimization method used is the Healed Topographical Multilevel Single Linkage (HTMLSL) [20], a stochastic algorithm based on MLSL. The algorithm attempts to find all local minima of an objective function $F(x)$ inside a bounded set $S\subset\mathbb{R}^n$, which are potentially global. These local minima are obtained by a local-search procedure, starting from suitably chosen points in a properly maintained sample. At the k^{th} iteration: (i) Construct a sample selecting at random $N\psi$ points from S and evaluate the objective function for each point; (ii) Choose from the sample a subset of points to be used as starting points for local searches; (iii) Perform a local search from each starting point. If a new minimum is discovered store it; (iv) Determine whether to stop or not. If not, repeat starting from step i. From the stored local minima the one with the lowest value might be the global minimum.

3. Results

The study was conducted in the invasive cardiology department in the University Hospital of Ioannina. We included 199 subjects suspected for CAD undergoing their first CA for the diagnosis or exclusion of CAD. Patients with known coronary disease were excluded from the study. Patients with more than mild valvular heart disease were also excluded. 89 of the subjects were normal, and for the rest 110 the presence of CAD was confirmed by two experts. In order to characterize the subjects, 19 features were used (Table 1).

Hypertension was defined as SBP more than 140mmHg and/or DBP more than 90mmHg or use of antihypertensive agents. Diabetes mellitus was defined as a fasting blood glucose concentration more than 126mg/dl or antihyperglycemic drug treatment. Current smoking was defined as having smoked the last cigarette less than a week before CA. Hyperlipidemia was defined as total cholesterol over 220mg/dl or use of lipid-lowering agents (statins or fibrates). Body mass index (BMI) was calculated as weight (kg) divided by the square of height (m^2). Carotid-Femoral Pulse Wave Velocity (PWVcf) and Augmentation index (AIx) were measured non-invasively using applanation tonometry, as indices of vascular stiffness.

Table 1. Feature that characterize each subject

#	Feature	Units
1	Age	years
2	Sex	male(1), female(0)
3	Family History	yes(1), no(0)
4	Smoking	smoker (2), ex-smoker (1), non-smoker (0)
5	Diabetes	FBGC ≥ 126mg/dl (1) else (0)
6	Hypertension	DBP>90mmHg and/or SBP>140mmHg (1) else (0)
7	Hyperlipidemia	total cholesterol over 220mg/dl (1) else (0)
8	Creatinine	mg/dL
9	Glucose	mg/dL
10	Total Cholesterol	mg/dL
11	HDL	mg/dL
12	TRG	mg/dL
13	BMI	kg/ m^2
14	Waist	cm
15	HR	bpm
16	SBP	mmHg
17	DBP	mmHg
18	PWVcf	m/sec
19	AIx	%

Table 2. Results

	Crisp rule-based classifier							Fuzzy model						
	TP[1]	TN[2]	FP[3]	FN[4]	Se[5] (%)	Sp[6] (%)	Acc[7] (%)	TP[1]	TN[2]	FP[3]	FN[4]	Se[5] (%)	Sp[6] (%)	Acc[7] (%)
Run 1	10	6	3	1	90.91	66.67	80	9	8	1	2	81.82	88.89	85
Run 2	6	5	4	5	54.55	55.56	55	8	5	4	3	72.73	55.56	65
Run 3	8	4	5	3	72.73	44.44	60	11	5	4	0	100	55.56	80
Run 4	5	7	2	6	45.45	77.78	60	6	8	1	5	54.55	88.89	70
Run 5	7	4	5	4	63.64	44.44	55	10	7	2	1	90.91	77.78	85
Run 6	7	5	4	4	63.64	55.56	60	9	6	3	2	81.82	66.67	75
Run 7	3	4	5	8	27.27	44.44	35	9	5	4	2	81.82	55.56	70
Run 8	10	3	6	1	90.91	33.33	65	10	4	5	1	90.91	44.44	70
Run 9	5	5	4	6	45.45	55.56	50	8	5	4	3	72.73	55.56	65
Run 10	7	5	3	4	63.64	62.5	63.16	8	5	3	3	72.73	62.5	68.42
Overall	**68**	**48**	**41**	**42**	**61.82**	**53.93**	**58.29**	**88**	**58**	**31**	**22**	**80**	**65.17**	**73.37**

[1]TP: True Positive, [2]TN:True Negative, [3]FP:False Positive, [4]FN:False Negative, [5]Se:Sensitivity, [6]Sp:Specificity, [7]Acc:Accuracy

In order to assess the subjects status, concerning the existence or not of CAD, CA was performed by the Judkins technique. All coronary angiograms were visually assessed by two experienced angiographers and a consensus was reached. Significant CAD was defined as at least one 50% or greater diameter stenosis in at least one coronary artery vessel. The absence of CAD was defined as completely smooth epicardial coronary arteries without any narrowing visible in CA.

The ten fold stratified cross validation method, which is considered as the most reliable evaluation method [17], was used for evaluation. The stratification was implemented by dividing the subjects into two subsets: those with CAD and those with no CAD. The procedure was applied to each fold, generating ten different crisp set of rules and fuzzy models. Both, have been evaluated in our dataset. Table 2 presents the true positive, true negative, false positive and false negative results, for each fold. In addition, the sensitivity, specificity and accuracy of each fold and in overall (average of all folds) are provided. The overall accuracy of the crisp set of rules is 58%, while the average accuracy for the fuzzy DSS is 73%. The difference, 15%, is statistically significant at 95% confidence interval. Due to the fact that ten different crisp set of rules were generated (one for each fold), there is no final crisp set of rules. Therefore, in Fig. 2, we present indicative crisp rules, generated using the whole dataset.

4. Discussion and conclusions

In the current study we introduced a novel methodology for the automated generation of a DSS. The methodology is data-driven and is implemented in three steps. Initially a set of rules is constructed from a decision tree, induced by the training records. Then, the resulted crisp set of rules is transformed to a fuzzy model and, finally, the parameters of this fuzzy model are optimized. The above methodology was used for the creation of a DSS for the diagnosis of CAD.

The proposed DSS is advantageous since it requires easily obtained data, i.e. the patient's history, routine blood tests and non-invasive assessment of vascular stiffness. Furthermore, the rule-based nature of the DSS makes the decision making process transparent. The use of CA for the initial annotation is an advantage, concerning the quality of our dataset due to the fact that CA is the most reliable examination regarding CAD diagnosis. It should be mentioned that the classification results of the crisp rule-based classifier are significantly improved when it is transformed to a fuzzy model and the fuzzy parameters are optimized. As we can see from Table 2, the improvement in accuracy ranges from 5-35% when considering each fold separately. The overall improvement in accuracy is 15%.

In Table 3 we present a comparison of several methods for CAD diagnosis. A direct comparison

$\mathbf{r_1}$: **IF** (a_2=0 **AND** a_4=0 **AND** $a_{15} \leq 65$ **AND** $a_{18} \leq 10.5$ **AND** $a_{19} \leq 48$ **AND** a_3=0)	**THEN** y=0
$\mathbf{r_2}$: **IF** (a_2=1 **AND** a_{15}>49 **AND** $a_1 \leq 69$ **AND** a_3=1 **AND** a_6=1 **AND** $a_9 \leq 103$)	**THEN** y=0
$\mathbf{r_3}$: **IF** (a_2=1 **AND** a_{15}>49 **AND** a_1>69 **AND** a_5=1 **AND** $a_8 \leq 1$)	**THEN** y=0
$\mathbf{r_4}$: **IF** (a_2=1 **AND** a_{15}>49 **AND** $a_1 \leq 69$ **AND** a_3=0 **AND** a_5=0 **AND** $a_{13} \leq 27.99$ **AND** a_{12}>191)	**THEN** y=1
$\mathbf{r_5}$: **IF** (a_2=1 **AND** a_{15}>49 **AND** $a_1 \leq 69$ **AND** a_3=0 **AND** a_5=0 **AND** a_{13}>27.99 **AND** $a_{15} \leq 53$)	**THEN** y=1
$\mathbf{r_6}$: **IF** (a_2=1 **AND** a_{15}>49 **AND** a_1>69 **AND** a_5=1 **AND** a_8>1)	**THEN** y=1

Figure 2. Indicative crisp rules

cannot be derived since different features and datasets have been used. In [13] a subset of our dataset has been used and comparable performance has been reported. It should be mentioned that most of the methods reported in Table 3 are based on neural networks [6,9,11-13,15]. These methods are not able to provide interpretation for their decisions. On the other hand our DSS is able to provide the desired comprehensibility since it is based on a set of rules. Also, in most of the reported works, CA was not performed to all subjects thus, an uncertainty concerning the presence or absence of CAD is introduced to the used datasets.

Table 3. Comparison with other methods.

Method	Features	No[1]	Se[2] (%)	Sp[3] (%)	Acc[4] (%)
Neural networks [11]	subject's data	2004	100	26	
Neural networks [12]	laboratory examinations	162			66
Signal processing [10]	arterio-scillography	51	73	90	
Case based reasoning [8]	scintigraphy	100	98	70	70
Neural networks [9]	scintigraphy	115	76	77	
Neural networks [6]	heart sounds	55	83	87	
Signal processing [1]	ECG	345	80	90	
Neural networks [15]	subject's data, ECG and scintigraphy	327	96	84	92
Neural networks [13]	subject's data	139	78	75	78
Fuzzy rules	subject's data	**199**	**80**	**65**	**73**

[1]No: # of subjects, [2]Se: Sensitivity, [3]Sp: Specificity, [4]Acc: Accuracy

A limitation of our study is that our patient population was of high risk for CAD and, thus not representative of the general population. Further studies need to be performed in order to create a DSS that could accurately diagnose CAD in the general population. Also, the extension of the DSS to the determination of CAD's severity is of great interest.

Acknowledgements

This research is part funded by the program "Heraklitos" of the Hellenic Ministry of Education and by the EC as part of the project NOESIS (IST-2002-507960).

5. References

[1] J.W. Deckers, B.J. Rensing, R.V.H. Vinke, and M.L. Simoons, "Comparison of exercise algorithms for diagnosis of coronary artery disease", in *proc. CinC*, 1988, pp. 231-4.

[2] F. Ng, S. Wong,, F. Mora, G. Passariello, and D. Almeida, "Diagnostic value of QT interval hysteresis loop in coronary artery disease", in *proc. CinC*, 1998, pp. 417-20.

[3] M. Sabry-Rizk, S. El-Khafif, E. Carson, W. Zgallai, K. Grattan, C. Morgan, and P. Hardiman, "Suspicious polyphase patterns of normal looking ECGs provide fast early diagnoses of a coronary artery disease", in *proc. IEEE Eng. in Med. and Biol.*, 1999, vol. 2, pp. 1391-1393.

[4] S.M. Szildgyi, L. Szildgyi, and L. David, "Assessment of R wave amplitude changes in coronary artery disease", in *proc. IEEE Eng. in Med. Biol.*, 1997, vol. 1, pp. 272-274.

[5] E.J. Tkacz, and P. Kostka, "An application of wavelet neural network for classification of patients with coronary artery disease based on HRV analysis", in *proc. IEEE Eng. in Med. Biol.*, 2000, vol. 2, pp. 1391-1393.

[6] M. Akay, Y.M. Akay, and W. Welkowitz, "Neural networks for the diagnosis of coronary artery disease", in *proc. Int. Joint Conf. on Neural Net.s*, 1992, pp. 419-424.

[7] Y.M. Akay, M. Akay, W. Welkowitz, and J. Kostis, "Noninvasive detection of coronary artery disease", *IEEE Eng. in Med. and Biol. Mag.,* 1994, vol. 13, pp. 761-764.

[8] M. Haddad, K.P. Adlassnig, and G. Porenta, "Feasibility analysis of a case-based reasoning system for automated detection of coronary heart disease from myocardial scintigrams", *Artif. Intel. Med*, 1997, vol. 9, pp. 61-78.

[9] H.B. Ebadian, H.S. Zadeh, S. Setayeshi, and S.T. Smith, "Neural network and fuzzy clustering approach for automatic diagnosis of coronary artery disease in nuclear medicine", *IEEE Trans. Nuclear Science*, 2004, vol. 51, pp. 184-192.

[10] M. Pouladian, M.R. H.Golpayegani, A.A. Tehrani-Fard, and M. Bubvay-Nejad, "Noninvasive detection of coronary artery disease by arterio-oscillography", *IEEE Trans. Biomed. Eng*, 2005, vol. 52, pp. 743-747.

[11] B.A. Mobley, E. Schechter, W.E. Moore, P.A. McKee, and J.E. Eichner, "Neural network predictions of significant coronary artery stenosis in men", *Artif. Intel. Med*, 2005, vol. 34, pp. 151-161.

[12] P. Lapuerta, S.P. Azen, and L. Labree, "Use of neural networks in predicting the risk of coronary artery disease", *Comp. and Biomed. Research*, 1995, vol. 28, pp. 38-52.

[13] D. Frossyniotis et. al., "An Intelligent system for the early diagnosis of Coronary artery disease", in *proc. NNESMED*, 2001.

[14] M. Kukar, C. Groselj, "Reliable diagnostics for coronary artery disease", in *proc. IEEE CBMS*, 2002, pp. 7-12.

[15] M. Kukar, et. al., "Analysing and improving the diagnosis of ischaemic heart disease with machine learning", *Artif. Intel. Med*, 1999, vol. 16, pp. 25-50.

[16] N.B. Merz, "Assessment of patients at intermediate cardiac risk", *Am. J. Cardiol.*, 2005, vol. 96(supl), pp.2J-10J.

[17] P.N. Tan, M. Steinbach, V. Kumar. *Introduction to data mining*, Addison Wesley, 2005.

[18] J.R. Quinlan, *C4.5: Programs for Machine Learning* Morgan Kauffman, 1993.

[19] L.X. Wang, *A course in fuzzy systems and control*, Prentice-Hall, 1986.

[20] F.V. Theos, I.E. Lagaris, and D.G. Papageorgiou, "PANMIN: Sequential and Parallel Global Optimization Procedures with a variety of options for the local search strategy", *Comp. Phys. Commun*, 2004, vol. 159, pp. 63-69.