

# Fuzzy rule-based system applied to risk estimation of cardiovascular patients

---

Jan Bohacik, Department of Computer Science, University of Hull, Hull, HU6 7RX, United Kingdom  
and Department of Informatics, University of Zilina, 010 26 Zilina, Slovakia, e-mail:

J.Bohacik@hull.ac.uk, Jan.Bohacik@fri.uniza.sk

Darryl N. Davis, Department of Computer Science, University of Hull, Hull, HU6 7RX, United Kingdom,  
e-mail: D.N.Davis@hull.ac.uk

Cardiovascular decision support is one area of increasing research interest. On-going collaborations between clinicians and computer scientists are looking at the application of knowledge discovery in databases to the area of patient diagnosis, based on clinical records. A fuzzy rule-based system for risk estimation of cardiovascular patients is proposed. It uses a group of fuzzy rules as a knowledge representation about data pertaining to cardiovascular patients. Several algorithms for the discovery of an easily readable and understandable group of fuzzy rules are formalized and analysed. The accuracy of risk estimation and the interpretability of fuzzy rules are discussed. Our study shows, in comparison to other algorithms used in knowledge discovery, that classification with a group of fuzzy rules is a useful technique for risk estimation of cardiovascular patients.

*Key words:* classification, fuzzy rules, linguistic variable elimination, cumulative information estimations, classification ambiguity, medical data mining, cardiology

## 1. Introduction

European health care systems are facing important challenges, such as ageing populations, increase in lifestyle-related health problems and limitations of health care resources. According to [Lieshout](#)

et al., 2008], cardiovascular diseases have been reported as the principal cause of death in most European countries. They account for 43% of mortality among men and for 56% among women. For both men and women coronary heart disease is the most prevalent cause of cardiovascular death; while stroke is relatively more prevalent in women. In cardiovascular risk assessment, diabetes is a very important factor as diabetes patients are at high risk for cardiovascular disease. Its prevalence is still rising due to several factors; overweight being one of these factors. Other important factors are age, gender, genetic factors, clinical factors such as hypertension, and life style factors such as smoking, alcohol consumption, physical exercise and diet.

Monitoring risk factors is important for the prevention of malignant events. Three areas of prevention can be distinguished: a) prevention in the total population; b) prevention in high risk groups; and c) prevention after cardiovascular events. Prevention in the total population includes life style factors and programs targeted at various groups in diverse settings, such as schools, local communities, homes for elderly people, healthcare providers etc. Prevention in high risk groups is targeted at chronic clinical conditions, which mainly affect adults aged 55 years or over, that would otherwise increase the risk for cardiovascular events, such as hypertension and diabetes. These conditions may also have a major negative impact on the patients' functional status, productivity, and quality of life. Acute cardiovascular events such as myocardial infarction and stroke, determine mortality and, if the patient survives, define the quality of life and risk for recurrent events. Prevention in high risk groups and prevention after cardiovascular events includes both life style changes and medication.

As active participants of the BraveHealth project, we are focused on continuous and remote monitoring and real time prevention of malignant events for people already diagnosed as subjects at risk of further cardiological or cardiovascular events. In the project, our patients are required to use a wearable unit with sensors and other devices such as scales and blood pressure cuffs so that we can obtain regular data about them. The data is analyzed in real time by several techniques in the

developing BraveHealth system. These techniques independently decide if a patient is high risk and their results are combined into a final decision about the patient. If the patient is considered high risk, all necessary steps are carried out so that malignant events can be prevented. Cardiovascular risk prediction tools such as the Systematic Coronary Risk Evaluation system (SCORE) [Conroy et al., 2003], the Framingham Risk Score [Wilson et al., 1998] and the Prospective Cardiovascular Munster Heart Study (PROCAM) [Assmann et al., 2002] are not so useful in this situation. These tools are optimized for a 10-year risk prediction of developing fatal cardiovascular disease. However, patients considered here are already at risk. Techniques used include: a) monitoring if some important measures, such as systolic blood pressure, diastolic blood pressure, heart rate, etc., are within limits set by clinicians; and b) data mining techniques such as classifiers. Data mining techniques are used in the data mining step of knowledge discovery in databases, which is a process of nontrivial extraction of implicit, previously unknown and potentially useful information from the data stored in a database [Fayyad et al., 1996]. Data mining techniques are useful since our system is expected to contain a lot of data about the patient and this data is accessible for significant time periods in the history of the patient's treatment. If we took data about patients at some point, after a period of time we could divide it into a group with data about dead patients and a group with data about live patients. These two groups could be used as high and low risk patients for making the classifier. This classifier would be able to give risk predictions for current data about patients. Data mining techniques are also suitable for dealing with the nonlinear and complex data that are often present in cardiovascular domains [Grossi, 2006].

There have been several publications which deal with risk assessment or prediction of cardiovascular disease with data mining techniques [Fidele et al., 2009], [Nicholson et al., 2008], [Palaniappan and Awang, 2008], [Patil and Kumaraswamy, 2009], [Tsipouras et al., 2007], [Yan et al., 2006]. An artificial neural network based classifier is used in [Fidele et al., 2007], [Nicholson et al., 2008], [Palaniappan and Awang, 2008], [Patil and Kumaraswamy, 2009]. In [Fidele et al., 2009], a network with the Levenberg-Marquardt algorithm and the Resilient Back-propagation is employed. It is used

for prediction to determine if a new patient has or does not have a heart disease and then how severe the disease is. Different risk levels of heart attack can be predicted with a Multi-layer Perceptron neural network with Back-propagation as the training algorithm according to [Patil and Kumaraswamy, 2009]. A concrete heart disease (i.e. coronary heart disease, rheumatic valvular heart disease, hypertension, chronic cor pulmonale, and congenital heart disease) is predicted in [Yan et al., 2006]. It also relies on a Multi-layer Perceptron neural network, but with an improved Back-propagation algorithm. Prediction of 10-year risk of event with Bayesian networks is well described in [Nicholson et al., 2008]. Several previously known Bayesian networks are analysed: the Busselton Bayesian Network, the PROCAM-German Bayesian Network, and the PROCAM-adapted Bayesian Network. A clinical support tool, TakeHeartII, is also suggested there. In this tool, the clinician can ask for a risk assessment of cardiovascular disease after providing information about the patient. It shows a graph for 10-year risk of event and the patient's current risk. In [Palaniappan and Awang, 2008], a prototype Intelligent Heart Disease Prediction System (IHDP) and its Internet user interface are introduced. IHDP uses three data mining modelling techniques, namely, a decision tree, a Naive Bayes classifier and a neural network. It employs Data Mining Extensions (DMX) query language and functions for building and accession of the data mining techniques. The concrete algorithms used to make the data mining models (knowledge representations) are not specified. According to the analysis in the paper, the most effective model to predict patients with heart disease appears to be the Naive Bayes classifier followed by the neural network and the decision tree. A method used for automated arrhythmic beat classification and automated ischemic beat classification is introduced in [Tsipouras et al., 2007]. This method relies on rules provided by expert cardiologists and their transformation into fuzzy rules and defined membership functions for attributes on the basis of cardiovascular data. The presented results in the paper indicate an escalation of the performance in accuracy when the initial rules are transformed to more sophisticated fuzzy rules.

The research reported in this paper considers assessing the risk of individual cardiovascular patients with the use of fuzzy rules discovered by data mining techniques. The aim of our study is to investigate and develop techniques capable of helping to decide if a patient is high risk in the BraveHealth system. In the knowledge discovery in databases process, including its data mining step, various found dependences in data are called knowledge. One of more effective knowledge representations, and immediately understandable to clinical partners, is a group of rules in the form “**IF Condition THEN Conclusion**”. *Conditions* contain expressions “*Attribute is possible categorical value*” connected with operator “AND”. For example, “*Age is mid aged AND Respiratory problem is moderate COAD*” where “*Age*” is the age of the patient, “*Respiratory problem*” indicates the patient’s problems with breathing and COAD is an abbreviation for chronic obstructive airway disease. *Conclusions* contain expression “*Risk is possible level of risk*”, e.g. “*Risk is high*”. Notice that attributes can be assigned only to categorical values, i.e. numerical values have to be transformed into categorical values. This contrasts to the use of neural networks where categorical values are transformed to numerical values. However, a group of rules is easily understandable while a neural network is considered a black box knowledge representation.

In cardiovascular data, cognitive uncertainties such as vagueness and ambiguity are often present. Vagueness is associated with the difficulty to make clear or precise distinctions in the real world [Klir, 1987]. For example, it is strange to consider a patient’s age “*mid aged*” when the patient is 55 and “*old*” when the patient is 56. Small changes in numerical values can cause changes in categorical values, which can lead to significant changes in predictions [Quinlan, 1987]. On the other hand, ambiguity is associated with two or more alternatives where the choice among them remains unspecified [Klir, 1987]. For example, the clinician can think that the patient’s respiratory problem can be both mild COAD and moderate COAD and (s)he cannot decide. Vagueness and ambiguity have been solved successfully with notions of fuzzy logic such as fuzzy sets, membership functions and membership degrees for several years. A group of rules which makes use of the notions associated with fuzzy logic is called a group of fuzzy rules. Our goal is to produce a group of fuzzy

rules that identifies high risk individual cardiovascular patients correctly. It is particularly important to address the minimization of high risk patients considered as low risk, which leads to life-threatening situations, and the minimization of low risk patients considered as high risk, which leads to high costs. The rules should also be easily readable by the clinicians in order to make more sophisticated data interpretation and decision-making, and for this reason the number of fuzzy rules in the group and their lengths are analyzed. Clinicians also view the decision about the level of risk for a particular patient as consistent with their knowledge. These critical objectives are reflected in the choice of the algorithms developed, especially those which calculate the ambiguity of the perceived risk. Indeed the use of the quantitative and declarative knowledge representations associated with Fuzzy Logic algorithms mean that a number of the challenges identified by Sittig et al. [Sittig et al., 2008] are addressed. In particular such schemes enable the summarization of patient-level information; their deployment enables the prioritization and filtering of recommendations to the clinical user; and the use of fuzzy logic will ultimately allow us to combine recommendations for patients with co-morbidities.

This paper is organized as follows. Definitions of terms and marks used throughout the paper are in Section 2. In Section 3, cardiovascular data and its transformation (fuzzification) are described. The fuzzy rule discovery itself and the use of a group of fuzzy rules are analyzed in Section 4. The performance of our approach is discussed in Section 5. Section 6 concludes this paper.

## 2. Definitions

Definitions of terms, marks and measures employed in the paper are given and summarized here.

Definition 1:

Let some universe  $U$  be given, a *fuzzy subset*  $M$  of the set  $U$  is a map  $M: U \rightarrow [0; 1]$ , where the value of  $M(x)$  for each  $x \in U$  is interpreted as the degree to which  $x$  is an element of fuzzy

subset  $M$  (i.e. *membership degree*), or equally, as the truthfulness of the statement “ $x$  is an element of fuzzy subset  $M$ ”.

Definition 2:

Let  $M$  be a fuzzy set defined on the universe  $U$ . Fuzzy set  $M$  at significance level  $\alpha$  (marked  $M^\alpha$ )

is defined as follows:  $M^\alpha = \begin{cases} M(x); & \text{if } M(x) \geq \alpha \\ 0; & \text{if } M(x) < \alpha \end{cases}$ .

Definition 3:

Cardinality of fuzzy set  $M$  defined on the universe  $U$  is specified as follows:  $\#(M) = \sum_{x \in U} M(x)$ .

Definition 4:

A *linguistic term* is a (lexical) name associated with a fuzzy set  $M$  which is defined on a universe  $U$ . A *linguistic variable* is a set of linguistic terms. The fuzzy sets, with which these terms are associated, are defined on one universe  $U$ .

When  $M(x)$  is replaced with  $M^\alpha(x)$  in Definition 3, symbol  $\#_\alpha(M)$  is used for denoting the value of the sum. Let linguistic terms “*none*”, “*mild COAD*”, “*moderate COAD*”, “*severe COAD*” be associated with fuzzy subsets of universe  $U$ . Let “*Respiratory problem*” be a linguistic variable defined as *Respiratory problem* = {*none*; *mild COAD*; *moderate COAD*; *severe COAD*}. It is said “*none*”, “*mild COAD*”, “*moderate COAD*”, “*severe COAD*” are associated with (are defined for) “*Respiratory problem*”. Membership degree to which  $x$  is an element of the fuzzy set associated with “*none*” is symbolized by  $none(x)$ . Similarly, if  $M$  is the fuzzy set associated with “*none*”,  $\#(none)$  (resp.  $\#_\alpha(none)/none^\alpha$ ) is often used instead of  $\#(M)$  (resp.  $\#_\alpha(M)/M^\alpha$ ). If “*none*” is chosen from the terms predefined for “*Respiratory problem*”, it is denoted by “*Respiratory problem is none*”. A new linguistic term can be derived from linguistic terms defined for linguistic variables when conjunction “AND” is used and its membership degree is computed with t-norm. Symbol  $\operatorname{argmax}_{x \in X} \{f(x)\}$ , where  $x$  is a linguistic term and  $X$  is a linguistic variable, stands for a set of all

$x \in X$  where  $f(x)$  is the maximal value. Symbol  $\text{rnd } X$ , where  $X$  is a set, stands for one  $x$  chosen from  $X$  randomly.

Definition 5:

Let  $U$  be the set of all possible instances and let all  $e \in U$  be described by linguistic variables  $A = \{A_1; \dots; A_k; \dots; A_n\}$ . A *linguistic condition*  $E$  is a linguistic term associated with a subset  $\text{terms}(E)$  of linguistic terms defined for variables in  $A$ . Its lexical name is a connection of terms in  $\text{terms}(E)$  with conjunction "AND". For any possible variable in  $A$  there is at most one linguistic term from the linguistic terms defined for this variable.  $E$  is associated with a fuzzy set whose membership degree  $E(e)$ ,  $e \in U$ , is defined as follows:  $E(e) = 1$  if  $\text{terms}(E) = \emptyset$ , otherwise the value of  $E(e)$  is the result of t-norm applied on all  $a_{k,l} \in \text{terms}(E)$ .

Definition 6:

Let  $U$  be the set of all possible instances of the task. Let all  $e \in U$  be described by linguistic variables  $A = \{A_1; \dots; A_k; \dots; A_n\}$ . Let them be associated with  $U$  and let linguistic terms  $a_{k,1}, \dots, a_{k,l}, \dots, a_{k,n_k}$ , where  $n_k$  is a natural number, be defined for all  $A_k \in A$ . That is  $A_k = \{a_{k,1}; \dots; a_{k,l}; \dots; a_{k,n_k}\}$ . Let  $V \subseteq U$  be a learning set of instances, i.e. a set of instances  $e$  which the values of  $a_{k,l}(e)$ , any  $a_{k,l} \in A_k$  and any  $A_k \in A$ , are known. Let  $e \in V$  be classified by known values of  $c_j(e)$ ,  $c_j \in C$  where  $C = \{c_1; \dots; c_j; \dots; c_L\}$  is a linguistic variable associated with  $U$  and  $L$  is the defined number of linguistic terms we classify to.  $C$  is the class linguistic variable. *The task of making fuzzy rules* is to make rules: **IF**  $E_i^{\text{Condition}}$  **THEN**  $C$  is  $c_j$  ( $ECR_i$ ) for all  $i = 0, 1, 2, \dots, m$ , where  $E_i^{\text{Condition}} = A_{i_1}$  is  $a_{i_1,j_1}$  AND  $A_{i_2}$  is  $a_{i_2,j_2}$  AND ... AND  $A_{i_{n_i}}$  is  $a_{i_{n_i},j_{n_i}}$ ,  $m$  is the number of rules and  $ECR_i$  is a set of extra criteria for the rule (e.g. its weight).  $E_i^{\text{Condition}}$  contains at least one variable and none of them is there more than once. The rules are used to determine the values of  $c_j(e)$ ,  $c_j \in C$ , for an instance  $e \in U$ , i.e. to classify some  $e \in U$ .



Linguistic term “ $a_{k,l}$ ” in  $terms(E)$  of a linguistic condition  $E$ ,  $a_{k,l} \in A_k \in \mathbf{A}$ , is equally replaced by “ $A_k$  is  $a_{k,l}$ ” and vice versa. For example, “ $a_{2,2}$  AND  $a_{5,3}$ ” equals “ $A_2$  is  $a_{2,2}$  AND  $A_5$  is  $a_{5,3}$ ”. The following symbols are also used.  $E = \emptyset / E \neq \emptyset$  means  $terms(E) = \emptyset / terms(E) \neq \emptyset$ . If there is a linguistic term/no linguistic term defined for linguistic variable  $A_k$  in  $terms(E)$ , we write  $A_k \in E / A_k \notin E$ . Symbol  $E \setminus A_k$  means removing the linguistic term  $a_{k,l} \in A_k$  from  $terms(E)$  if present.  $E \cup a_{k,l}, A_k \notin E$  is a linguistic condition, where “ $E$  AND  $A_k$  is  $a_{k,l}$ ” if  $E \neq \emptyset$  and “ $A_k$  is  $a_{k,l}$ ” if  $E = \emptyset$ . Membership degree  $(E \cup a_{k,l})(e) = \mathbf{T}(E(e); a_{k,l}(e))$ , where  $\mathbf{T}$  is t-norm.  $E \cup c_j$  is a linguistic term “ $E$  AND  $C$  is  $c_j$ ” if  $E \neq \emptyset$  and “ $C$  is  $c_j$ ” if  $E = \emptyset$ , where  $E$  is a linguistic condition,  $c_j$  is a linguistic term defined for class linguistic variable  $C$ . Membership degree  $(E \cup c_j)(e) = \mathbf{T}(E(e); c_j(e))$ , where  $\mathbf{T}$  is t-norm.

Definition 7:

*Degree of truthfulness*  $\mathbf{DTF}_\alpha(E; c_j; \mathbf{V})$  for fixed linguistic condition  $E$ , class linguistic term  $c_j \in C$ , known instances  $\mathbf{V}$  and significance level  $\alpha$  is defined as:

$$\mathbf{DTF}_\alpha(E; c_j; \mathbf{V}) = \frac{\sum_{e \in \mathbf{V}} \mathbf{T}(c_j^\alpha(e); E^\alpha(e))}{\sum_{e \in \mathbf{V}} E^\alpha(e)} \text{ for all } c_j \in C. \quad (1)$$

Definition 8:

*The possibility of classifying an instance*  $e \in \mathbf{U}$  to linguistic term  $c_j \in C$  for given linguistic condition  $E$ , known instance  $\mathbf{V}$ , and significance level  $\alpha$ , is defined as follows:

$$\pi_\alpha(E; c_j; \mathbf{V}) = \frac{\mathbf{DTF}_\alpha(E; c_j; \mathbf{V})}{\max_{c_k \in C} \{\mathbf{DTF}_\alpha(E; c_k; \mathbf{V})\}} \text{ for all } c_j \in C. \quad (2)$$

*Possibility distribution on*  $C$  marked  $\pi_E$  is defined as follows. It consists of values  $\pi_\alpha(E; c_j; \mathbf{V})$  for all  $c_j \in C$  ordered in non-increasing order. The highest value is represented by  $\pi_E(1)$ , the second highest value by  $\pi_E(2)$ , ..., and the last lowest value by  $\pi_E(\#(C))$ .  $\pi_E(\#(C) + 1) = 0$ .

Definition 9:

Classification ambiguity  $\mathbf{AMB}_\alpha(E; C; \mathbf{V})$  of instances in  $\mathbf{V}$  classified into  $c_j \in C$  if a linguistic condition  $E$  is known is defined as follows. If  $\#(C) = 1$ ,  $\mathbf{AMB}_\alpha(E; C; \mathbf{V}) = 0$ . If  $\#(C) > 1$ , then:

$$\mathbf{AMB}_\alpha(E; C; \mathbf{V}) = \sum_{i=2}^{\#(C)} (\pi_E(i) - \pi_E(i+1)) \cdot \ln i. \quad (3)$$

Definition 10:

Classification ambiguity  $\mathbf{AMB}_\alpha(A_k; E; C; \mathbf{V})$  in  $e \in \mathbf{V}$  classified into  $c_j \in C$  when  $A_k \in \mathbf{A}$  is considered and a linguistic condition  $E$  is known is defined for  $A_k \notin E$  as follows:

$$\mathbf{AMB}_\alpha(A_k; E; C; \mathbf{V}) = \sum_{a_{k,l} \in A_k} \frac{\#(E \cup a_{k,l})}{\sum_{a_{k,m} \in A_k} \#(E \cup a_{k,m})} \cdot \mathbf{AMB}_\alpha(E \cup a_{k,l}; C; \mathbf{V}). \quad (4)$$

Definition 11:

Cumulative information of linguistic condition  $E$  (linguistic term  $E \cup c_j$ ,  $c_j \in C$ ) is defined as:

$$\mathbf{II}(E; \mathbf{V}) = \begin{cases} -\log_2 \#(E); & \text{if } E \neq \emptyset \\ -\log_2 \#(\mathbf{V}); & \text{if } E = \emptyset \end{cases} \quad (5)$$

$$(\mathbf{II}(E \cup c_j; \mathbf{V}) = -\log_2 \#(E \cup c_j)). \quad (6)$$

Definition 12:

Information of linguistic condition  $E$  (linguistic term  $E \cup c_j$ ,  $c_j \in C$ ) is defined as:

$$\mathbf{I}(E; \mathbf{V}) = \log_2 \#(\mathbf{V}) + \mathbf{II}(E; \mathbf{V}) \quad (7)$$

$$(\mathbf{I}(E \cup c_j; \mathbf{V}) = \log_2 \#(\mathbf{V}) + \mathbf{II}(E \cup c_j; \mathbf{V})). \quad (8)$$

Definition 13:

Conditional information (conditional cumulative information) of  $c_j \in C$  for known instances  $\mathbf{V}$  provided that  $E$  is known is defined as:

$$\mathbf{I}(c_j / E; \mathbf{V}) (\mathbf{II}(c_j / E; \mathbf{V})) = \mathbf{II}(E \cup c_j; \mathbf{V}) - \mathbf{II}(E; \mathbf{V}). \quad (9)$$

Definition 14:

Mutual information  $\mathbf{I}(C; A_k; E; \mathbf{V})$  for determining the amount of information which is obtained about  $C$  if values of  $E(\mathbf{e})$ ,  $\mathbf{e} \in \mathbf{V}$ , and  $a_{k,l}(\mathbf{e})$ ,  $a_{k,l} \in A_k \notin E$ ,  $\mathbf{e} \in \mathbf{V}$ , are known is defined as:

$$\mathbf{I}(C; A_k; E; \mathbf{V}) = \sum_{c_j \in C} \sum_{a_{k,l} \in A_k} \#(E \cup a_{k,l} \cup c_j) \cdot \mathbf{TMP}(c_j; a_{k,l}; E; \mathbf{V}), \text{ where} \quad (10)$$

$$\mathbf{TMP}(c_j; a_{k,l}; E; \mathbf{V}) = \mathbf{II}(E \cup c_j; \mathbf{V}) + \mathbf{II}(E \cup a_{k,l}; \mathbf{V}) - \mathbf{II}(E \cup a_{k,l} \cup c_j; \mathbf{V}) - \mathbf{II}(E; \mathbf{V}). \quad (11)$$

Definition 15:

Cumulative entropy of linguistic variable  $A_k \in \mathbf{A}$  on known instances  $\mathbf{e} \in \mathbf{V}$  is defined as:

$$\mathbf{HH}(A_k; \mathbf{V}) = \sum_{a_{k,l} \in A_k} \#(a_{k,l}) \cdot \mathbf{II}(a_{k,l}; \mathbf{V}). \quad (12)$$

### 3. Fuzzification of data

As a dataset, a group of 839 instances (cardiovascular patients) classified into two levels of risk and described by 17 attributes  $\mathbf{B}$  as queries about patients' symptoms, medical history, clinical findings and results of physiological measurements is used. Instances are derived from clinical data collected at two clinical sites (the Hull site of 498 instances and the Dundee site of 341 instances) [Davis and Nguyen, 2009]. The description of instances and their summary is given in Table 1.

Describing attributes  $\mathbf{B}$  are defined as  $\mathbf{B} = \{B_1; \dots; B_k; \dots; B_n\}$ . If  $B_k$  is a categorical attribute,  $B_k = \{b_{k,1}; \dots; b_{k,l}; \dots; b_{k,n_k}\}$  where  $b_{k,1}; \dots; b_{k,l}; \dots; b_{k,n_k}$  are possible categorical values. Age ( $B_1$ ) and Gender ( $B_2$ ) represent the age and the gender of the patient. Heart disease ( $B_3$ ), Diabetes ( $B_4$ ), and Stroke ( $B_5$ ) respectively indicate if any heart disease, diabetes or a stroke are present. Similarly, attributes Renal failure ( $B_8$ ), Hypertension ( $B_{10}$ ), Shunt ( $B_{14}$ ), and Coronary artery bypass surgery ( $B_{16}$ ) indicate if renal insufficiency, a high blood pressure, a shunt, or coronary artery bypass surgery are present. Attribute Side ( $B_6$ ) holds the side of surgery. It is either *left* or *right*. Attribute Respiratory problem ( $B_7$ ) indicates problems with breathing. If there are no problems, value *none* is applied. The other possible values are *mild COAD*, *moderate COAD* and *severe COAD*. COAD is an abbreviation for chronic obstructive airway disease. ASA grade ( $B_9$ ) is used to classify the patient into categorical values *one*, *two*, *three* or *four* according to the American Society of Anesthesiologists

classification. Value *one* means the patient is fit and well for her/his age. Value *two* means the patient's cardiovascular disease is mild, i.e. it does not hamper enjoyment of daily activities. Value *three* means the patient's cardiovascular disease is severe, i.e. it restricts the patient's daily activities. Value *four* means the patient's cardiovascular disease is life-threatening. Attribute *ECG* ( $B_{11}$ ) describes electrocardiography, i.e. a transthoracic (across the thorax or chest) interpretation of the electrical activity of the heart over a period of time. Several categorical values are used: *normal*, *q waves*, *st-t waves*, *afib 60 to 90*, *afib  $\leq 90$* , *five ectopic*, *other*. Value *normal* means there are no abnormalities in electrocardiography. Value *q waves* means Q wave abnormalities are present. Value *st-t waves* means ST-T wave abnormalities are present. Values *afib 60 to 90* and *afib  $\leq 90$*  are related to atrial fibrillation. Value *five ectopic* means the patient has five or more ectopic heartbeats per minute. Value *other* represents all other abnormalities. *Duration* ( $B_{12}$ ) is the duration of surgery in hours. *Blood loss* ( $B_{13}$ ) represents the blood loss in surgery in milliliters. *Patch* ( $B_{15}$ ) indicates which material is used for by-pass patching in the patient's surgery. The values *arm vein/leg vein/other vein* means arm veins/leg veins/veins indicate the different patient body part source used; while the values *dacron* and *ptfe* express the use of synthetic material, either Dacron or polytetrafluoroethylene. Value *stent* means a stent is inserted into the patient's body. Value *none* shows there has not been any bypass patching for the patient. Attribute *Consultant* ( $B_{17}$ ) describes the particular consultant employed for the patient's treatment. The real names of consultants are anonymised and replaced with *a*, *b*, *c*, *d* and *e* in this paper. Class attribute *Risk* ( $D$ ) is used to classify instances into two possible categorical values  $d_1$  and  $d_2$  meaning risk levels ("*low*" and "*high*", respectively). It is denoted by  $D = \{d_1; d_2\}$ . The values of class attribute  $D$  are generated according to the following heuristic clinical model [Davis and Nguyen, 2009]: an instance (cardiovascular patient) is classified into "*high*" if the patient's death or severe cardiovascular event (e.g. stroke, myocardial relapse or cardiovascular arrest) appears within 30 days after an operation.

TABLE 1 Dataset of cardiovascular patients.

Attribute	Data Type	Value Range	Frequency
Age ( $B_1$ )	Numerical	38-93	N/A
Gender ( $B_2$ )	Categorical	female ( $b_{2,1}$ )	332
		male ( $b_{2,2}$ )	507
Heart disease ( $B_3$ )	Categorical	yes ( $b_{3,1}$ )	351
		no ( $b_{3,2}$ )	488
Diabetes ( $B_4$ )	Categorical	yes ( $b_{4,1}$ )	90
		no ( $b_{4,2}$ )	749
Stroke ( $B_5$ )	Categorical	yes ( $b_{5,1}$ )	272
		no ( $b_{5,2}$ )	567
Side ( $B_6$ )	Categorical	left ( $b_{6,1}$ )	458
		right ( $b_{6,2}$ )	381
Respiratory problem ( $B_7$ )	Categorical	none ( $b_{7,1}$ )	727
		mild COAD ( $b_{7,2}$ )	92
		moderate COAD ( $b_{7,3}$ )	18
		severe COAD ( $b_{7,4}$ )	2
Renal failure ( $B_8$ )	Categorical	yes ( $b_{8,1}$ )	12
		no ( $b_{8,2}$ )	827
ASA grade ( $B_9$ )	Categorical	one ( $b_{9,1}$ )	4
		two ( $b_{9,2}$ )	645
		three ( $b_{9,3}$ )	182
		four ( $b_{9,4}$ )	8
Hypertension ( $B_{10}$ )	Categorical	yes ( $b_{10,1}$ )	455
		no ( $b_{10,2}$ )	384
ECG ( $B_{11}$ )	Categorical	normal ( $b_{10,1}$ )	604
		q waves ( $b_{10,2}$ )	74
		st-t waves ( $b_{10,3}$ )	35
		afib 60 to 90 ( $b_{10,4}$ )	16
		afib $\leq$ 90 ( $b_{10,5}$ )	7
		five ectopic ( $b_{10,6}$ )	2
		other ( $b_{10,7}$ )	101
Duration ( $B_{12}$ )	Numerical	0.7-5	N/A
Blood loss ( $B_{13}$ )	Numerical	0-2000	N/A
Shunt ( $B_{14}$ )	Categorical	yes ( $b_{14,1}$ )	501
		no ( $b_{14,2}$ )	338
Patch ( $B_{15}$ )	Categorical	arm vein ( $b_{15,1}$ )	3
		leg vein ( $b_{15,2}$ )	4
		other vein ( $b_{15,3}$ )	150
		dacron ( $b_{15,4}$ )	185
		ptfe ( $b_{15,5}$ )	171
		stent ( $b_{15,6}$ )	1
		none ( $b_{15,7}$ )	325
Coronary artery bypass surgery ( $B_{16}$ )	Categorical	yes ( $b_{16,1}$ )	52
		no ( $b_{16,2}$ )	787
Consultant ( $B_{17}$ )	Categorical	a ( $b_{17,1}$ )	237
		b ( $b_{17,2}$ )	114
		c ( $b_{17,3}$ )	102
		d ( $b_{17,4}$ )	383

		$e(b_{17,5})$	3
<i>Risk (D)</i>	Categorical	<i>low</i> ( $d_1$ )	713
		<i>high</i> ( $d_2$ )	126

For several years our data about cardiovascular patients have been collected with respect to crisp classification where only one disease is considered fully possible and all the others are considered fully impossible, which does not always correspond to reality. Further investigation into our data collection is being pondered in the BraveHealth project so that opinions and clinical knowledge of clinicians can be represented more realistically and directly with notions of fuzzy logic. For the purpose of our study, the data containing 839 instances described by attributes  $\mathbf{B}$  and classified into the categorical values of  $D$  has to be fuzzified so that it can be used in the task of making fuzzy rules. The attributes in  $\mathbf{B}$  are transformed into linguistic variables  $\mathbf{A} = \{A_1; \dots; A_k; \dots; A_n\}$  and respective linguistic terms  $a_{k,1}; \dots; a_{k,l}; \dots; a_{k,n_k}$  for each  $A_k \in \mathbf{A}$  are defined. The class attribute  $D$  is transformed into a class linguistic variable  $C = \{c_1; c_2\} = \{high; low\}$ . Membership degrees  $a_{k,l}(e)$  for all  $a_{k,l} \in A_k \in \mathbf{A}$  and all  $e \in \mathbf{V}$ , where  $\mathbf{V}$  is the group of our 839 instances, and membership degrees  $c_j(e)$  for all  $c_j \in C$  and all  $e \in \mathbf{V}$  are defined. Fuzzification of categorical attributes is trivial. A linguistic variable is defined for each categorical attribute and a linguistic term is predefined for each possible categorical value of a categorical attribute. Then, for each instance in  $\mathbf{V}$  and each predefined linguistic term, the membership degree is set to one/zero if the categorical value corresponds/does not correspond to the instance and the linguistic term. In the case of numerical attributes, fuzzification is a special kind of discretization where sharp boundaries are softened with membership functions. The following triangular membership functions, where the numerical values of numerical attribute  $B_k \in \mathbf{B}$  for all  $e \in \mathbf{V}$  are represented by  $X = \{x(e), e \in \mathbf{V}\}$  and  $m_l$ ,  $l = 1, \dots, n_k$ , are centres, are used:

$$\mu_{a_{k,1}}(x) = \begin{cases} 1; & x \leq m_1 \\ \frac{m_2 - x}{m_2 - m_1}; & m_1 \leq x \leq m_2 \\ 0; & x \geq m_2 \end{cases} \quad (12)$$

$$\mu_{a_k, n_k}(x) = \begin{cases} 1; & x \geq m_{n_k} \\ \frac{x - m_{n_k - 1}}{m_l - m_{n_k - 1}}; & m_{n_k - 1} < x < m_{n_k} \\ 0; & x < m_{n_k - 1} \end{cases} \quad (13)$$

$$\mu_{a_k, l}(x) = \begin{cases} 0; & x \geq m_{l+1} \\ \frac{m_{l+1} - x}{m_{l+1} - m_l}; & m_l < x < m_{l+1}, 1 < l < n_k \\ \frac{x - m_{l-1}}{m_l - m_{l-1}}, & m_{l-1} < x < m_l \\ 0; & x \leq m_{l-1}, \end{cases} \quad (14)$$

The concrete values of centres  $m_l$ ,  $l = 1, \dots, n_k$ , are computed according to an iterative algorithm that is described in [Yuan and Shaw, 1995]. At time 0, centres  $m_l[0]$ ,  $l = 1, \dots, n_k$ , are initially set to be evenly distributed on the range of  $X$ , such as  $m_l[0] = \min\{x, x \in X\} + (\max\{x, x \in X\} - \min\{x, x \in X\}) \cdot \frac{l-1}{n_k-1}$ . The centres are then adjusted iteratively in order to reduce the total distance of  $X$  to  $M = \{m_l, l = 1, \dots, n_k\}$ , defined as  $D(X; M) = \sum_{x \in X} \min_l \|x - m_l\|$ . Each iteration at time  $t$  consists of three steps:

- (1) Randomly draw one sample  $x$  from  $X$ , denoted as  $x[t]$ ;
- (2) Find the closest center to  $x[t]$ , i.e. find  $y$  such that  $\|x - m_y\| = \min_l \|x - m_l\|$ ;
- (3) Adjust  $m_y[t+1] = m_y[t] + \eta[t](x[t] - m_y[t])$  and keep  $m_l[t+1] = m_l[t]$  for all  $l \neq y$ , where  $t$  is iteration time,  $\eta[t]$  is a monotonic decreasing scalar learning rate.

The iteration continues until  $D(X; M)$  converges. Function  $\eta[t] = \frac{1}{t}$  is used. Numerical attributes  $B_1 = \text{Age}$ ,  $B_{12} = \text{Duration}$ ,  $B_{13} = \text{Blood loss}$  are transformed into linguistic variables  $A_1 = \text{Age} = \{a_{1,1}; a_{1,2}; a_{1,3}\} = \{\text{mid aged}; \text{old}; \text{very old}\}$ ,  $A_{12} = \text{Duration} = \{a_{12,1}; a_{12,2}\} = \{\text{short}; \text{long}\}$ ,  $A_{13} = \text{Blood loss} = \{a_{13,1}; a_{13,2}; a_{13,3}\} = \{\text{insignificant}; \text{significant}; \text{serious}\}$ .

#### 4. Fuzzy rule algorithms

Three different algorithms for extraction of fuzzy rules in the task of making fuzzy rules are presented here together with ways how to determine the values of  $c_j(e)$ ,  $c_j \in C$ , for an instance  $e \in U$  with these fuzzy rules (i.e. how to classify an instance  $e \in U$  with these fuzzy rules). One algorithm based on [Bohacik, 2000] extracts fuzzy rules on the basis of linguistic variable elimination.

It eliminates the least important  $A_k \in \mathbf{A}$  in a way that leads to dividing  $\mathbf{e} \in \mathbf{V}$  into two groups with subsets of  $\mathbf{A}$  and with minimal inconsistencies between the membership degrees for variables in  $\mathbf{A}$  and  $\mathbf{C}$ . It continues in the groups until no further elimination is considered important and a set of fuzzy rules is formed. The other two algorithms first make a fuzzy decision tree and this is then transformed into fuzzy rules. The main difference between them is the measure used for association of a linguistic variable with a decision node. One algorithm, based on [Levashenko and Zaitseva, 2002], uses mutual information criterion and chooses the linguistic variable with its highest value. The other algorithm, based on [Yuan and Shaw, 1995], uses classification ambiguity and chooses the linguistic variable with its lowest value.

#### 4.1 Algorithm based on linguistic variable elimination

The algorithm based on linguistic variable elimination (marked LVE hereafter) has five input parameters:  $\mathbf{A}$ ,  $\mathbf{C}$ ,  $\mathbf{V}$ , significance level  $\alpha \in [0,1]$  and degree-of-truthfulness threshold  $\beta \in [0,1]$ . Significance level serves as a filter of insignificant membership degrees for  $\mathbf{e} \in \mathbf{V}$ . By this, ambiguity can be eliminated and the importance of higher values of  $a_{k,l}(\mathbf{e})$  and  $c_j(\mathbf{e})$ ,  $\mathbf{e} \in \mathbf{V}$ , can be increased. Degree-of-truthfulness threshold controls the minimal truthfulness of obtained fuzzy rules. A lower value leads to an increase in the number of fuzzy rules, but with some of them covering local dependences in data. Such fuzzy rules usually have lower accuracy of determining the values of  $c_j(\mathbf{e})$ ,  $c_j \in \mathbf{C}$ ,  $\mathbf{e} \in \mathbf{V}$ . If the value of the degree-of-truthfulness threshold increases to a certain level, the increase in accuracy of determining  $c_j(\mathbf{e})$ ,  $c_j \in \mathbf{C}$ ,  $\mathbf{e} \in \mathbf{V}$ , stops. It sometimes happens that there are not any rules with a high value of  $\beta$ . T-norm is defined as  $T(a; b) = \min\{a; b\}$ .

{fuzzy rule} = makeFuzzyRules( $\alpha$ ;  $\beta$ ;  $\mathbf{A}$ ;  $\mathbf{C}$ ;  $\mathbf{V}$ ):

- (1) Set time  $t = 0$ , temporarily made rules  $\mathbf{R} = \emptyset$ , instances of the decision table at time  $t$   $\mathbf{I}[t] = \mathbf{V}$ , considered linguistic variables of decision table at time  $t$   $\mathbf{P}[t] = \mathbf{A}$ , available linguistic variables for elimination at time  $t$   $\mathbf{L}[t] = \mathbf{A}$ , the maximal number of allowable linguistic variables in the



condition of a rule at time  $t$   $maxLength[t] = \#(\mathbf{A})$ ,  $length[t] = 1$ ,  $retainedVariable[t] = false$ ,  $\alpha$  and  $\beta$  are fixed;

- (2) Set the eliminated linguistic variable at time  $t$   $A_{eliminated}[t] = \text{rnd argmax}_{A_k \in L[t]} \{ \mathbf{AMB}_\alpha(A_k; \emptyset; C; I[t]) \}$ . Also set  $I_1[t] = \emptyset$ ,  $P_1[t] = P[t] \setminus A_{eliminated}[t]$ ,  $P_2[t] = P[t]$ ,  $L_{decreased}[t] = L[t] \setminus A_{eliminated}[t]$ . If  $retainedVariable[t] = false$ , then  $Z[t] = V$ . Otherwise,  $Z[t] = I[t]$ ;
- (3) For each  $e \in I[t]$  do: if there is not any  $f \in Z[t]$ ,  $f \neq e$  such that for all  $A_k \in P_1[t]$   $\{ \text{argmax}_{a_{k,l} \in A_k} \{ a_{k,l}(f) \} \} \cap \{ \text{argmax}_{a_{k,l} \in A_k} \{ a_{k,l}(e) \} \} \neq \emptyset$  and  $\{ \text{argmax}_{c_j \in C} \{ c_j(f) \} \} \cap \{ \text{argmax}_{c_j \in C} \{ c_j(e) \} \} \neq \emptyset$ , then set  $I_1[t] = I_1[t] \cup \{e\}$ , otherwise set  $I_2[t] = I_2[t] \cup \{e\}$ ;
- (4) If  $length[t] + 1 < maxLength[t]$ , go to step (5). If not, make one fuzzy rule for all  $e \in I[t]$  and put them without duplication into  $\mathbf{R}$ . The made rules have the following form: **IF**  $A_{k_1}$  is  $\text{rnd argmax}_{a_{k_1,l} \in A_{k_1}} \{ a_{k_1,l}(e) \}$  **AND**  $A_{k_2}$  is  $\text{rnd argmax}_{a_{k_2,l} \in A_{k_2}} \{ a_{k_2,l}(e) \}$  **AND** ... **AND**  $A_{k_q}$  is  $\text{rnd argmax}_{a_{k_q,l} \in A_{k_q}} \{ a_{k_q,l}(e) \}$  **THEN**  $C$  is  $\text{rnd argmax}_{c_j \in C} \{ c_j(e) \}$  ( $\emptyset$ ) where  $\{A_{k_1}; A_{k_2}; \dots, A_{k_q}\} = P_1[t]$  if  $e \in I_1[t]$  and  $\{A_{k_1}; A_{k_2}; \dots, A_{k_q}\} = P_2[t]$  if  $e \in I_2[t]$ ;
- (5) Perform steps (2)-(5) once for  $I[t+1] = I_1[t]$ ,  $P[t+1] = P_1[t]$ ,  $L[t+1] = L_{decreased}[t]$ ,  $length[t+1] = length[t] + 1$ ,  $retainedVariable[t+1] = false$ ,  $t = t + 1$  and once for  $I[t+1] = I_2[t]$ ,  $P[t+1] = P_2[t]$ ,  $L[t+1] = L_{decreased}[t]$ ,  $length[t+1] = length[t] + 1$ ,  $retainedVariable[t+1] = true$ ,  $t = t + 1$ ;
- (6) Compute measure  $\mathbf{DTF}_0(A_k; c_j; V)$  for each fuzzy rule “**IF**  $E_i$  **THEN**  $C$  is  $c_j$  ( $\emptyset$ )” in  $\mathbf{R}$ . Return the fuzzy rules which have  $\mathbf{DTF}_0(A_k; c_j; V) \geq \beta$  as the result of the algorithm.

Classification of  $e \in U$ , whose  $a_{k,l}(e)$  for all  $a_{k,l} \in A_k \in \mathbf{A}$  are known, on the basis of fuzzy rules made with LVE,  $\{c_j(e)\} = \text{classify}(\{\text{fuzzy rule}\}; e; C)$ :

- (1) If the made fuzzy rules do not contain at least one fuzzy rule “**IF**  $E_i$  **THEN**  $C$  is  $c_j$  ( $\emptyset$ )” for each  $c_j \in C$ , it is better to repeat the process of making fuzzy rules. E.g.,  $V$  with greater  $\#(V)$  is used. If there is not such a  $V$  available, it is possible to set  $c_j(e) = 0$  for  $c_j$  without any fuzzy rule;
- (2) For each fuzzy rule “**IF**  $E_i$  **THEN**  $C$  is  $c_j$  ( $\emptyset$ )”, compute the membership degree of  $E_i$ . These membership degrees have the following mark  $E_i^{c_j}(e)$ ;
- (3) Divide fuzzy rules into  $\#(C)$  groups marked  $g^{c_j}$  on the basis of  $c_j \in C$  in their conclusions;
- (4) For each  $g^{c_j}$ , compute the maximum of values of all  $E_i^{c_j}(e)$ . The result of this computation is the value of  $c_j(e)$  (maximum can also be replaced by any other definition of s-norm).

#### 4.2 Algorithm based on maximization of mutual information criterion

The algorithm has six input parameters:  $A, C, V$ , frequency-of-branch threshold  $\alpha \in [0,1]$ , frequency-of-class threshold  $\beta \in [0,1]$ , and criterion for association of a linguistic variable with a node **CRIT**. Parameter  $\alpha$  controls the growth of the decision tree on the basis of the frequency of branch. The higher its value, the lower the height of the decision tree is (or the lower the number of “Linguistic variable is linguistic term” in conditions of made fuzzy rules). Parameter  $\beta$  controls the growth of the decision tree on the basis of the frequency of  $c_j \in C$ . The lower its value, the lower the height of the decision tree (or equally, the lower the number of linguistic variable conditions in made fuzzy rules). Increasing the value of  $\alpha$  and decreasing the value of  $\beta$  can lead to a potentially better classification of  $e \in U \setminus V$ . However, it can lead to a less accurate classification of  $e \in V$ . **CRIT** is a criterion for association of a linguistic variable  $A_k \in A$  with a node when the decision tree is build. There are several criteria of this kind, e.g.  $\mathbf{CRIT}(C; A_k; E; V) = \mathbf{I}(C; A_k; E; V)$  (mutual information),  $\mathbf{CRIT}(C; A_k; E; V) = \frac{\mathbf{I}(C; A_k; E; V)}{\mathbf{H}(A_k; V)}$  (relative mutual information). If the former criterion is used, the algorithm is called MMI hereafter. If the latter criterion is used, the algorithm is called MRMI hereafter. T-norm is defined as  $T(a; b) = a \cdot b$ .

Decision tree = makeTree( $\alpha; \beta; \mathbf{CRIT}; A; C; V$ ):

- (1) Make the root and its associate linguistic variable  $maxA_k = \operatorname{argmax}_{A_k \in \mathbf{A}} \{\mathbf{CRIT}(C; A_k; \emptyset; \mathbf{V})\}$ .  
Make a branch for each  $a_{k,l} \in maxA_k$ , connect them with the root, associate them with the particular  $a_{k,l}$  and consider them unprocessed;
- (2) If there is no unprocessed branch, END. Otherwise, choose one of the unprocessed branches and consider it the current branch. Make linguistic term  $E$  for the current branch.  $E$  consists of all linguistic variable conditions from the root to the current branch connected with operator AND;
- (3) Set  $\text{branchII} = \mathbf{II}(E; \mathbf{V})$  and  $\text{minClassI} = \min_{c_j \in C} \{\mathbf{I}(c_j / E; \mathbf{V})\}$ . If  $\{\text{branchII} \geq -\log_2(\alpha \cdot \#(\mathbf{V}))\}$  or  $\{\text{minClassI} \leq -\log_2(\beta)\}$  or  $\{(A \setminus E) = \emptyset\}$ , go to step (4), otherwise go to step (5);
- (4) Make a leaf, connect it with the current branch and consider this branch processed. Associate  $F(c_j) = \mathbf{F}(c_j / E; \mathbf{V})$  for all  $c_j \in C$  and  $\operatorname{argmax}_{c_j \in C} \{F(c_j)\}$  with the made leaf. Go to step (2);
- (5) Make a node, connect it with the current branch and associate linguistic variable  $maxA_k = \operatorname{argmax}_{A_k \in \mathbf{A}, A_k \notin E} \{\mathbf{CRIT}(C; A_k; \emptyset; \mathbf{V})\}$ . Consider the current branch processed. Make a branch for each  $a_{k,l} \in maxA_k$ , connect them with the made node, associate them with particular  $a_{k,l}$  and consider them unprocessed. Go to step (2).

Transformation of the made decision tree to fuzzy rules,  $\{\text{fuzzy rule}\} = \text{makeFuzzyRules}(\text{decision tree})$ :

- (1) For each leaf  $i$  of the decision tree, mark the linguistic term associated with it as  $c^i$ . For each leaf  $i$ , take the branch going to it and make linguistic condition  $E_i$  for this branch.  $E_i$  consists of all linguistic variable conditions from the root to the branch and they are connected with operator AND. Set  $ECR_i = \{F_i(c_j) \mid F_i(c_j) \text{ which were associated with leaf } i\}$  for each leaf  $i$ ;
- (2) Make a fuzzy rule in the form of “**IF**  $E_i$  **THEN**  $C$  is  $c^i$  ( $ECR_i$ )” for each  $E_i$ .

Classification of  $\mathbf{e} \in \mathbf{U}$ , whose  $a_{k,l}(\mathbf{e})$  for all  $a_{k,l} \in A_k \in \mathbf{A}$  are known, on the basis of made fuzzy rules  $\{c_j(\mathbf{e})\} = \text{classify}(\{\text{fuzzy rule}\}; \mathbf{e}; C)$ :

- (1) For each fuzzy rule “**IF**  $E_i$  **THEN**  $C$  is  $c^i$  ( $ECR_i$ )”, compute  $E_i(\mathbf{e})$ ;
- (2) Set  $c_j(\mathbf{e}) = \sum_{v_i} E_i(\mathbf{e}) \cdot F_i(c_j)$  where  $F_i(c_j) \in ECR_i$ ,  $c_j \in C$ .

### 4.3 Algorithm based on minimization of classification ambiguity

The algorithm has seven input parameters:  $A, C, V$ , significance level  $\alpha \in [0,1]$ , degree-of-truthfulness threshold  $\beta \in [0,1]$ , simplification of fuzzy rules  $\nu \in \{\text{true}, \text{false}\}$ , and keeping the current degree of truthfulness in simplification  $\omega \in \{\text{true}, \text{false}\}$ . Parameters  $\alpha$  and  $\beta$  are the same as in the LVE. If  $\nu = \text{false}/\nu = \text{true}$  and  $\omega = \text{false}/\nu = \text{true}$  and  $\omega = \text{true}$ , the algorithm is marked MCA-F-F/MCA-T-F/MCA-T-T. T-norm is defined as  $T(a; b) = \min\{a; b\}$ .

Decision tree = makeTree( $\alpha; \beta; A; C; V$ ):

- (1) Make the root and its associate linguistic variable  $minA_k = \operatorname{argmin}_{A_k \in A} \{\mathbf{AMB}_\alpha(A_k; \emptyset; C; V)\}$ .  
Make a branch for each  $a_{k,l} \in minA_k$ , connect them with the root, associate them with the particular  $a_{k,l}$  and consider them unprocessed;
- (2) If there is no unprocessed branch, END. Otherwise, choose one of the unprocessed branches and consider it the current branch. Make linguistic condition  $E$  for the current branch. Linguistic condition  $E$  consists of all linguistic variable conditions from the root to the current branch connected with operator AND;
- (3) Set  $\max = \max_{c_j \in C} \{\mathbf{DTF}_\alpha(E; c_j; V)\}$  and  $c_{max} = \operatorname{argmax}_{c_j \in C} \{\mathbf{DTF}_\alpha(E; c_j; V)\}$ . If  $\max \geq \beta$  then make the leaf which is associated with linguistic term  $c_{max}$ , connect it with the current branch, consider the current branch processed and go to the step (2). Otherwise, go to step (4);
- (4) If there is no  $A_k \in V, A_k \notin E$ , consider the current branch processed and go to (2). Otherwise, set the value of  $E\_AMB = \mathbf{AMB}_\alpha(E; C; V)$ ,  $E\_AMB = \min_{A_k \in V, A_k \notin E} \{\mathbf{AMB}_\alpha(A_k; E; C; V)\}$ ,  $minA_k = \operatorname{argmin}_{A_k \in V, A_k \notin E} \{\mathbf{AMB}_\alpha(A_k; E; C; V)\}$ . If  $minAMB < E\_AMB$ , go to step (5). Otherwise, consider the current branch processed and go to step (2);
- (5) Make a node with which linguistic variable  $minA_k$  is associated. Make a branch coming from this node for each  $a_{k,l} \in A_k$  and consider them unprocessed. Connect this node with the current branch and consider this branch processed. Go to step (2);

Transformation of the made decision to fuzzy rules, {fuzzy rule} = makeFuzzyRules(decision tree):

- (1) For each leaf  $i$  of the decision tree, mark the term associated with it as  $c^i$ . For each leaf  $i$ , take the branch going to it and make linguistic condition  $E_i$  for this branch.  $E_i$  contains linguistic variable conditions from the root to the branch and they are connected with operator AND;
- (2) For each leaf  $i$ , make fuzzy rule “**IF**  $E_i$  **THEN**  $C$  is  $c_j$  ( $\emptyset$ )”.

Simplification of the fuzzy rules {simplified fuzzy rule}=simplifyFuzzyRules({fuzzy rule};  $\beta$ ;  $\omega$ ;  $\mathbf{A}$ ;  $C$ ;  $\mathbf{V}$ ):

- (1) For each rule “**IF**  $E_i$  **THEN**  $C$  is  $c_j$  ( $\emptyset$ )” that has more than one linguistic variable  $A_k \in \mathbf{A}$  in its condition do: set  $\max = \max_{A_k \in E_i} \{\mathbf{DTF}_0(E_i \setminus A_k; c_j; \mathbf{V})\}$ ,  $\max A_k = \operatorname{argmax}_{A_k \in E_i} \{\mathbf{DTF}_0(E_i \setminus A_k; c_j; \mathbf{V})\}$ ,  $\text{dtf} = \mathbf{DTF}_0(E_i; c_j; \mathbf{V})$ . If  $\{\omega = \text{true and } \max \geq \text{dtf}\}$  or  $\{\omega = \text{false and } \omega \geq \beta\}$  then replace “**IF**  $E_i \setminus \max A_k$  **THEN**  $C$  is  $c_j$  ( $\emptyset$ )”;
- (2) If there was at least one replaced rule in step (1), go to step (1). Otherwise, if a fuzzy rule is there more than once, keep one fuzzy rule and remove all the other (duplicated) fuzzy rules.

Classification of  $e \in \mathbf{U}$ , whose  $a_{k,l}(e)$  for all  $a_{k,l} \in A_k \in \mathbf{A}$  are known, on the basis of made (simplified) fuzzy rules is done in the same way as it is for fuzzy rules made with LVE.

## 5. Experimental results

The main purpose of the experimental study is to compare the performance of the different fuzzy rule algorithms and with other algorithms on our cardiovascular data. Experiments were carried out with our Java software tool which is being developed with the intention of its integration into the medical decision-making support facilities of the BraveHealth system. The core algorithms, other than the fuzzy rule algorithms, are implemented in Weka [Witten et al., 2011]. The performance of algorithms is measured with sensitivity =  $\frac{\text{tp}}{\text{tp}+\text{fn}}$ , specificity =  $\frac{\text{tn}}{\text{tn}+\text{fp}}$ , positive predictive value =  $\frac{\text{tp}}{\text{tp}+\text{fp}}$ , negative predictive value =  $\frac{\text{tn}}{\text{tn}+\text{fn}}$  and accuracy =  $\frac{\text{tp}+\text{tn}}{\text{tp}+\text{fp}+\text{fn}+\text{tn}}$ . In the formulas, tp/ fp/ fn/ tn is the number of true positives/false positives/false negatives/true negatives. “ $C$  is low”/“ $D$  is low” is considered negative and “ $C$  is high”/“ $D$  is high” is considered positive. Values tp, fp, fn and tn are computed during 10-fold cross-validation. In 10-fold cross-validation, the (fuzzified) dataset is

partitioned into 10 folds of patients. The partition is random, but all folds contain roughly the same proportions of low risk and high risk patients. A patient is considered high risk/low risk in the dataset if the value assigned to attribute  $D$  is *high/low*. A patient is considered high risk in the fuzzified dataset if  $high \in \operatorname{argmax}_{c_j \in C} \{c_j(patient)\}$ ; otherwise, the patient is low risk. Of the 10 folds, a single fold is retained as the testing dataset for evaluation of discovered knowledge, and the remaining 9 folds are used as the learning dataset. The learning dataset is analyzed by the algorithm for the purpose of discovering the knowledge. The cross-validation process is repeated 10 times, with each of the 10 folds used exactly once as the testing dataset.

TABLE 2

Experimental results regarding accuracy.

Algorithm	SEN (%)	SPEC (%)	PPV (%)	NPV (%)	ACC (%)
Bayes	7.94	97.48	35.71	85.70	84.03
C4.5	7.76	98.60	37.50	85.42	84.51
NNC	15.08	90.18	21.35	85.73	78.90
MLP	15.08	89.62	20.43	85.66	78.43
LVE	65.08	75.88	32.28	92.48	74.26
MMI	77.78	89.62	56.98	95.80	87.84
MRMI	72.22	90.18	56.52	94.84	87.49
MCA-F-F	68.25	81.07	38.91	93.53	79.14
MCA-T-F	65.87	87.52	48.26	93.55	84.27
MCA-T-T	60.32	96.07	73.08	93.20	90.70

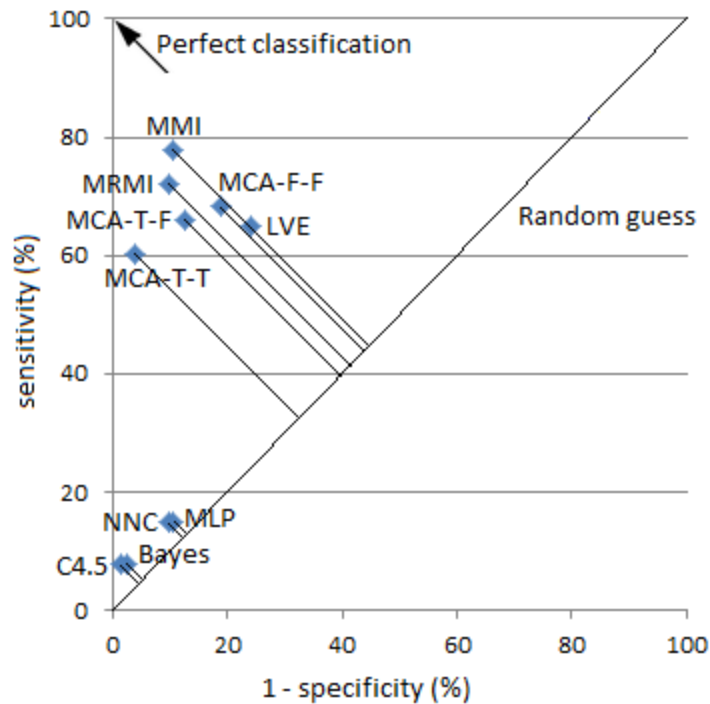


FIGURE 1

ROC graph for particular algorithms.

The accuracy results of our experiments are given in Table 2. Bayes denotes a Bayesian network implemented in Weka as class BayesNet. C4.5 is a decision tree classifier implemented in Weka as class J48. NNC is a nearest neighbor classifier using non-tested generalized examples [Brent, 1995] implemented in Weka as class NNge. MLP is a neural network classifier using multilayer perception implemented in Weka as class MultilayerPerception. LVE is the algorithm based on linguistic variable elimination. MMI is the algorithm based on maximization of mutual information; and MRMI is the algorithm based on maximization of relative mutual information. MCA-F-F is the algorithm based on minimization of classification ambiguity where there is no simplification of fuzzy rules acquired from the decision tree. MCA-T-F is the algorithm based on minimization of classification ambiguity where fuzzy rules acquired from the decision tree are simplified and the degree of truthfulness of fuzzy rules is not kept in simplification. MCA-T-T is the algorithm based on minimization of classification ambiguity where fuzzy rules acquired from the decision tree are simplified and the degree of truthfulness of fuzzy rules is kept in simplification. Bayes, C4.5, NNC and MLP use instances

described with attributes in  $B$  and classified to attribute  $D$ . LVE, MMI, MRMI, MCA-F-F, MCA-T-F, MCA-T-T use instances described with linguistic variables in  $A$  and classified to linguistic variable  $C$ . SEN is sensitivity, SPEC is specificity, PPV is positive predictive value, NPV is negative predictive value and ACC is accuracy.

Our results in Table 2 are interpreted in the form of a graph in Figure 1. The results from the tested algorithms are shown as plots in the ROC space. The distance from the random guess line is an indicator of how well the algorithm classifies a patient as low risk or as high risk. It is especially important to avoid classification of high risk patients as low risk as it would lead to life-threatening situations. Also, many low risk patients classified as high risk would increase the running costs of the BraveHealth system considerably. All the fuzzy rule algorithms described in this paper outperform the comparison (standard) algorithms by a considerable margin. When minimization of life-threatening situations and minimization of costs are considered, the best results are achieved by MMI.

TABLE 3

Experimental results regarding interpretability.

Algorithm	Number of fuzzy rules (avg.)	Length of a fuzzy rule (avg.)	Longest fuzzy rule (avg.)	Shortest fuzzy rule (avg.)
LVE	116.80	6.01	17.00	1.40
MMI	262.80	8.01	17.00	1.00
MRMI	248.90	8.62	17.00	1.00
MCA-F-F	23.30	3.58	6.50	1.50
MCA-T-F	16.90	2.15	3.70	1.00
MCA-T-T	22.80	2.46	4.70	1.00

The interpretability of discovered fuzzy rules is described in Table 3 with measures derived from [Ishibuchi et al., 2004]. The measures are computed for ten groups of fuzzy rules (learning groups) discovered for particular nine folds in 10-fold cross-validation and the average is taken. Number of fuzzy rules is the number of fuzzy rules in a learning group. Length of a fuzzy rule is the number of



linguistic variables in the condition of the fuzzy rule. The average for all lengths of fuzzy rules in all ten learning groups is used. Longest fuzzy rule is any fuzzy rule in a learning group with the highest number of linguistic variables in its condition. The average length of the longest fuzzy rules from all learning groups is in Table 3. Shortest fuzzy rule is any fuzzy rule in a learning group with the lowest number of linguistic variables in its condition. The average length of shortest fuzzy rules from all learning groups is in Table 3. MMI has the best results when minimization of life-threatening situations and minimization of costs are considered, however, the interpretability of its fuzzy rules is a lot worse than the interpretability of the algorithm based on minimization of classification ambiguity (MCA-F-F, MCA-T-F, MCA-T-T). When minimization of life-threatening situations, minimization of costs and interpretability are considered, MCA-T-T gives the best results.

## **6. Conclusions**

A fuzzy rule-based system for risk estimation of cardiovascular patients is presented. It consists of fuzzification of existing categorical and numerical attributes, algorithms for fuzzy rule discovery and algorithms for the use of discovered fuzzy rules in risk estimation. Through the adoption of soft boundaries, fuzzification allows us to minimize the information loss that appears when numerical attributes are discretized. Fuzzified categorical attributes also give us a possibility to assign more values with a categorical attribute, which is applicable when a clinician is not sure which categorical value should be assigned to a patient's particular feature. Fuzzy rules are discovered with linguistic variable elimination or a decision tree is built first and it is then transformed into fuzzy rules. A group of fuzzy rules is evaluated so that a group of fuzzy rules which minimize life-threatening situations, minimize costs and maximize interpretability is preferred. Life-threatening situations appear when patients at high risk are considered low risk, which is measured by sensitivity. This risk should be minimized and so sensitivity should be maximized. Costs are increased when low risk patients are treated as if they were high risk, which is measured by specificity. For the costs to be minimized specificity should be maximized. Interpretability is measured by the average number of fuzzy rules in

the group of fuzzy rules, average length of a fuzzy rule in the group, average length of the longest fuzzy rule and average length of the shortest fuzzy rule.

For our collected data about 839 patients, a group of fuzzy rules with sensitivity 60.32% and specificity 96.07% was found. On average, when 10-fold cross-validation is used, it has 22.80 fuzzy rules, 2.46 assignments in the condition of one fuzzy rule, 4.70/1.0 assignments in the condition of the longest/shortest fuzzy rule. When interpretability is not considered important, a group of fuzzy rules with sensitivity 77.78% and specificity 89.62% can be used. The results show the presented fuzzy-rule based system is a useful technique for risk estimation of cardiovascular patients. The results are competitive to the sensitivity and specificity of other data mining methods such as a Bayesian network, a C4.5-like decision tree, a nearest neighbor classifier using non-tested generalized examples and a neural network classifier using multilayer perception. As indicated earlier, the adoption of these fuzzy rule algorithms, together with other human readable outputs from Bayesian networks and decision trees will further the dissemination of best practices in Clinical Decision Support system design, development, and implementation; and address the many of the grand challenges [Sittig et al., 2008] for such systems, based on a sound theoretical basis.

## Acknowledgements

This work is funded by the European Commission's 7<sup>th</sup> Framework Program: BRAVEHEALTH FP7-ICT-2009-4, Objective ICT-2009.5.1: Personal Health Systems: a) Minimally invasive systems and ICT-enabled artificial organs: a1) Cardiovascular diseases.

## References

- [1] [Assmann et al., 2002] Assmann, G., Gullen, P., Schulte, H. (2002). Simple scoring scheme for calculating the risk of acute coronary events based on the 10-year follow-up of the prospective cardiovascular Munster (PROCAM) study. *Circulation*, Vol. 105, No. 3, pp. 310-315, ISSN: 1524-4539.

- [2] [Bohacik, 2000] Bohacik, J. (2010). Discovering fuzzy rules in databases with linguistic variable elimination. *Neural Network World*, Vol. 20, No. 1, pp. 45-61, ISSN: 1210-0552.
- [3] [Brent, 1995] Brent, M. (1995). *Instance-Based Learning: Nearest Neighbor with Generalization*. (Master's thesis). Retrieved from CiteSeerX (citeseerx.ist.psu.edu).
- [4] [Conroy et al., 2003] Conroy, R. M., Pyorala, K., Fitzgerald, A. P., Sans, S., Menotti, A., De Backer, G., Ducimetiere, P., Jousilahti, P., Keil, U., Njølstad, I., Oganov, R. G., Thomsen, T., Tunstall-Pedoe, H., Tverdal, A., Wedel, H., Whincup, P., Wilhelmsen, L., Graham, I. M., SCORE project group (2003). Estimation of ten-year risk of fatal cardiovascular disease in Europe: the SCORE project. *European Heart Journal*, Vol. 24, No. 11, pp. 987-1003, ISSN 1522-9645.
- [5] [Davis and Nguyen, 2009] Davis, D. N., Nguyen, T. T. (2009). Generating and Verifying Risk Prediction Models Using Data Mining: A Case Study from Cardiovascular Medicine. Chapter of *Data Mining and Medical Knowledge Management: Cases and Applications* (1<sup>st</sup> edition). :IGI Global Inc.
- [6] [Fayyad et al., 1996] Fayyad, U., Piatetsky-Shapiro, G., Smyth, P. (1996). From data mining to knowledge discovery in data mining. *AI Magazine*, Vol. 17, No. 3, pp. 37-54, 1996, ISSN: 0738-4602.
- [7] [Fidele et al., 2009] Fidele, B., Cheeneebash, J., Gopaul, A., Goorah, S. D. (2009). Artificial neural network as a clinical decision-supporting tool to predict cardiovascular disease. *Trends in Applied Sciences Research*, Vol. 4, No. 1, pp. 36-46, ISSN: 1819-3579.
- [8] [Grossi, 2006] Grossi, E. (2006). How artificial intelligence tools can be used to assess individual patient risk in cardiovascular disease: problems with the current methods. *BMC Cardiovascular Disorders*, Vol. 6, No. 1, pp. 20-25, ISSN: 1471-2261.
- [9] [Ishibuchi et al., 2004] Ishibuchi, H., Nakashima, T., Nii, M. (2004). *Classification and Modeling with Linguistic Information Granules: Advanced Approaches to Linguistic Data Mining* (1<sup>st</sup> edition). :Springer Verlag.

- [10] [Klir, 1987] Klir, G. J. (1987). Where do we stand on measures of uncertainty, ambiguity, fuzziness and the like? *Fuzzy Sets and Systems*, Vol. 24, No. 2, pp. 141-160.
- [11] [Levashenko and Zaitseva, 2002] Levashenko, V., Zaitseva, E. (2002). Usage of new information estimations for induction of fuzzy decision trees. *IDEAL '02 Proceedings of the Third International Conference on Intelligent Data Engineering and Automated Learning*, pp. 493-499, ISBN: 3-540-44025-9.
- [12] [Lieshout et al., 2008] Lieshout, J. v., Wensing, M., Grol, R. (2008). *Prevention of cardiovascular diseases: The role of primary care in Europe* (1<sup>st</sup> edition). (Electronic book). Retrieved from the EPA Cardio project ([www.epa-cardio.eu](http://www.epa-cardio.eu)).
- [13] [Nicholson et al., 2008] Nicholson, A. E., Twardy, C. R., Hope, L. R. (2008). Decision Support for Clinical Cardiovascular Risk Assessment. Chapter of *Bayesian Networks: A Practical Guide to Applications* (1<sup>st</sup> edition). :John Wiley & Sons, Ltd.
- [14] [Palaniappan and Awang, 2008] Palaniappan, S., Awang, R. (2008). Intelligent heart diseases prediction system using data mining techniques. *International Journal of Computer Science and Network Security*, Vol. 8, No. 8, pp. 343-350, ISSN: 1738-7906.
- [15] [Patil and Kumaraswamy, 2009] Patil, S. B., Kumaraswamy, Y. S. (2009). Intelligent and effective attack prediction system using data mining and artificial neural network. *European Journal of Scientific Research*, Vol. 31, No. 4, pp. 642-656, ISSN: 1450-216X.
- [16] [Quinlan, 1987] Quinlan, J. R. (1987). Decision trees as probabilistic classifier. *Proceedings of the Fourth International Workshop on Machine Learning*, pp. 31-37.
- [17] [Sittig et al, 2008] Sittig, D. F., Wright, A., Osheroff, J. A., Middleton, B., Teich, J. M., Ash, J. S., Campbell, E., Bates, D. W. (2008). Grand challenges in clinical decision support, *Journal of Biomedical Informatics*, Vol. 41, No.2, pp.387-392, ISSN: 1532-0464.
- [18] [Tsipouras et al., 2007] Tsipouras, M. G., Voglis, C., Fotiadis, D. I. (2007). A framework for expert system creation – application to cardiovascular diseases. *IEEE Transactions on Biomedical Engineering*, Vol. 54, No. 11, pp. 2089-2105, ISSN: 0018-9294.

- [19] [Wilson et al., 1998] Wilson, P. W., D'Agostino, R. B., Levy, D., Belanger, A. M., Silbershatz, H., Kannel, W. B. (1998). Prediction of coronary heart disease using risk factor categories. *Circulation*, Vol. 97, No. 18, pp. 1837-1847, ISSN: 1524-4539.
- [20] [Witten et al., 2011] Witten, I. H., Frank, E., Hall, M. A. (2011). *Data Mining: Practical Machine Learning Tools and Techniques* (3<sup>rd</sup> edition). :Morgan Kaufman.
- [21] [Yan et al., 2006] Yan, H., Jiangb, Y., Zheng, J., Peng, C., Li, Q. (2006). A multilayer perceptron-based medical decision support system for heart disease diagnosis. *Expert Systems with Applications*, Vol. 30, No. 2, pp. 272-281, ISSN: 0957-4174.
- [22] [Yuan and Shaw, 1995] Yuan, Y., Shaw, M.J. (1995). Induction of fuzzy decision trees. *Fuzzy sets and systems*, Vol. 69, No. 2, pp. 125-139, ISSN: 1041-4347.