Analysis of Fuzzy Decision Trees on Expert Fuzzified Heart Failure Data

Jan Bohacik, C. Kambhampati, Darryl N. Davis, JG Cleland

Department of Computer Science, Department of Informatics, Department of Cardiology
University of Hull, University of Zilina
Hull, UK, Zilina, Slovakia

Abstract—The prevalence of heart failure is 2-3% of the adult population and it is expected to grow. Half of all patients diagnosed with it die within four years. To minimize life-threatening situations and to minimize costs, it is interesting to predict mortality rates for a patient with heart failure. In this paper, a fuzzy decision tree based on classification ambiguity and a fuzzy decision tree based on cumulative information estimations are presented. They are employed on a heart failure data fuzzified on the basis of medical expert knowledge. After a transformation of fuzzy decision trees, the use of medical expert knowledge allows us to create a group of fuzzy rules that is easily interpretable by medical experts. Our study shows that different types of fuzzy decision trees can have significantly different accuracy results and interpretability.

Keywords—fuzzy decision tree; fuzzy rules; fuzzification; cardiology; heart failure

I. INTRODUCTION

Heart failure makes up an important medical, social, and economic problem [11]. Patients with heart failure suffer disabling symptoms, the most common of which are fatigue and dyspnea. The end stage of the disease is comparable to the end stage of terminal cancer. The prevalence of heart failure is estimated to 2%-3% of the adult population and it increases with age. In addition to it, the long-term prognosis for heart failure is poor. Half of all patients diagnosed with heart failure die within four years and about 40% of people admitted to hospital with heart failure die or are readmitted within one year. The cost of medical care for heart failure is measured in billions of dollars per year. Within the costs, the most powerful contributing factor is repeated hospitalization. The prevalence of heart failure increased from the 1950s increasingly. Most likely, a new increase will be observed in the future mainly because of the aging of the population and the trend showing an increasing prevalence of major heart risk factors such as obesity and diabetes.

Once a patient is known to suffer from heart failure, it is important to predict mortality rates so that they are used for an effective prevention. However, there is a lack of methods that accurately predict them. Existing clinical methods include EFFECT Risk Scoring system [8], Emergency Heart Failure Mortality Risk Grade (EHMRG) [9] or Seattle Heart Failure Model (SHFM) [7]. Since hospitals own more and more data about patients, data mining methods are also an option now.

Data mining methods are an integral part of knowledge discovery in databases - a non-trivial extraction of implicit, previously unknown and potentially useful information from data [2]. Data mining methods such as decision trees, nearest neighbor methods, and neural networks are employed in [1][12][13][14][20]. Overall, data mining methods outperform multiple logistic regression and epidemiological methods.

The research reported in the paper considers predicting the death of a heart failure patient within six months with data mining methods using fuzzy decision trees. A fuzzy decision tree method makes use of decision trees with incorporated notions of fuzzy logic. The incorporation allows us to take cognitive uncertainties in heart failure data such as ambiguity and vagueness into consideration. Ambiguity is associated with two or more alternatives where the choice among them remains unspecified [6]. Vagueness is associated with the difficulty to make clear or precise distinctions in the real world. Simultaneously, the notions of fuzzy logic allow us to express the expert knowledge about heart failure patients within the decision trees so that their interpretation is easily understandable and natural to medical experts. The goal is to produce fuzzy decision trees whose use leads to minimal life-threatening situations and minimal costs. Also, their interpretability related to their size or the interpretability of the fuzzy rules after their transformation is important. However, there are several types of fuzzy decision trees and their suitability for fulfilling the goals need to be investigated.

The paper is organized as follows. In Section II, our heart failure data and its expert fuzzification are described. The fuzzy decision tree methods employed are discussed in Section III. Their performance and their interpretability are analyzed in Section IV. Section V concludes the paper.

II. DATA AND FUZZIFICATION

A group of 2032 instances (heart failure patients) classified into two levels of patient status and described by 9 attributes B as queries about clinical findings and physiological measurements is used. Instances are derived from Hull LifeLab which is large, epidemiologically representative, information-rich clinical data [15]. Its purpose is studying patients with heart failure so that its definition and diagnosis, its natural history, its mechanisms and markers of progression, the associated costs to health services and society, and the delivery
of proven treatment to patients are improved. The attributes and their summary are presented in Tab. 1. Describing attributes B are defined as \( B = \{ B_1; \ldots; B_k; \ldots; B_p \} \). If \( B_k \) is a categorical attribute, \( B_k = \{ b_{k,1}; \ldots; b_{k,i}; \ldots; b_{k,k} \} \) where \( b_{k,1}; \ldots; b_{k,i}; \ldots; b_{k,k} \) are possible categorical values. Class attribute \( D \) is used to classify instances into two possible categorical values \( d_1 \) and \( d_2 \) (alive and dead). It is denoted by \( D = \{ d_1; d_2 \} \).

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Data Type</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blood Uric Acid Level ( B_1 )</td>
<td>Numerical</td>
<td>0.11 - 1.06</td>
</tr>
<tr>
<td>Blood Sodium Level ( B_2 )</td>
<td>Numerical</td>
<td>123 - 148</td>
</tr>
<tr>
<td>Blood Creatinine Level ( B_3 )</td>
<td>Numerical</td>
<td>37 - 1262</td>
</tr>
<tr>
<td>NT-proBNP Level ( B_4 )</td>
<td>Numerical</td>
<td>0.89 - 18236</td>
</tr>
<tr>
<td>Age ( B_5 )</td>
<td>Numerical</td>
<td>27 - 96</td>
</tr>
<tr>
<td>Sex ( B_6 )</td>
<td>Categorical</td>
<td>female ( b_{6,1} ), male ( b_{6,2} )</td>
</tr>
<tr>
<td>Pulse Rate ( B_7 )</td>
<td>Numerical</td>
<td>38 - 150</td>
</tr>
<tr>
<td>Weight ( B_8 )</td>
<td>Numerical</td>
<td>29.80 – 193.80</td>
</tr>
<tr>
<td>Height ( B_9 )</td>
<td>Numerical</td>
<td>1.2 – 1.96</td>
</tr>
<tr>
<td>Patient Status ( D )</td>
<td>Categorical</td>
<td>alive ( d_1 ), dead ( d_2 )</td>
</tr>
</tbody>
</table>

The attributes in \( B \) are fuzzified into linguistic variables \( A = \{ A_1; \ldots; A_k; \ldots; A_q \} \). Respective linguistic terms \( a_{k,1}; \ldots; a_{k,j} \) are defined for each \( A_k \in A \). The class attribute \( D \) is fuzzified into a class linguistic variable \( C = \{ c_1; c_2 \} = \{ alive; dead \} \). Membership degrees \( a_{k,i}(p) \) for all \( a_{k,i} \in A_k \) are defined for all \( p \in P \). membership degree \( c_j(p) \) for all \( c_j \in C \) and all \( p \in P \) are defined on the basis of our expert knowledge. \( b_k(p) \) is the particular value for \( B_k \in B \) and \( p \in P \) in our data.

**Blood Uric Acid Level** \( B_1 \) denotes the amount of uric acid in millimoles per liter of the patient's blood. Normal levels of uric acid are influenced in millimoles per liter of the patient's blood. Normal levels of uric acid are considered normal [3]. Levels less than 35 are decreased and levels greater than 145 are increased. Membership degree \( a_{3,1}(p) = decreased(p) \) equals 1 for \( B_1(p) \leq 132, 1.96 \) for \( B_1(p) = \{132;138\} \), and 0 for \( B_1(p) \geq 183 ). Membership degree \( a_{3,2}(p) = decreased(e) \) equals 0 for \( B_1(p) \leq 132 \) or \( B_1(p) \geq 148 \); 1 for \( B_1(p) \in \{138;142\}, \text{ and } \geq 148 \). Membership degree \( a_{3,3}(p) = increased(e) \) equals 0 for \( B_1(p) \leq 142 \), 1 for \( B_1(p) \in \{142;148\}, \text{ and } \geq 148 \).

**Blood Creatinine Level** \( B_3 \) is the amount of creatinine in micrograms per liter of the patient's blood. Normal levels of creatinine depend on the sex of the patient. If \( B_3(p) = \text{ male } \), \( a_{3,1}(p) = low(p) \) equals 1 for \( B_3(p) \leq 58 \), 0.11 for \( B_3(p) = \{53;58\} \), and 0 for \( B_3(p) \geq 75 \). Membership degree \( a_{3,2}(p) = normal(p) \) equals 0 for \( B_3(p) \leq 53 \) or \( B_3(p) \geq 75 \). Membership degree \( a_{3,3}(p) = high(p) \) equals 0 for \( B_3(p) \leq 103 \), 1 for \( B_3(p) \in \{103;133\} \), and 1 for \( B_3(p) \geq 133 \). If \( B_3(p) = \text{ female } \), the values 59, 77, 89, 107 are used instead of 53, 83, 103, 133, respectively.

**NT-proBNP Level** \( B_4 \) represents the amount of the N-terminal prohormone of brain natriuretic peptide (NT-proBNP) in picograms per milliliter of the patient's blood. Normal levels of NT-proBNP depend on the sex and age of the patient [4]. If \( B_4(p) = \text{ male } \) and \( B_4(p) < 45 \), \( a_{4,1}(p) = normal(p) \) equals 1 for \( B_4(p) \leq 81, 0.99 \) for \( B_4(p) \in \{81;99\} \), and 0 for \( B_4(p) \geq 99 \). Membership degree \( a_{4,2}(p) = elevated(p) \) equals 0 if \( B_4(p) \leq 81 \), 0.99 for \( B_4(p) \in \{81;99\} \), and 1 for \( B_4(p) \geq 99 \). If \( B_4(p) = \text{ male } \) and \( B_4(p) \in \{45;55\} \), \( B_4(p) \in \{55;65\} \), \( B_4(p) \in \{65;75\} \), \( B_4(p) \geq 75 \), then values 81 and 99 are replaced with 126 and 154/162 and 198/207 and 253/765 and 935, respectively. If \( B_4(p) = \text{ female } \) and \( B_4(p) < 45 \), \( B_4(p) < 55 \), \( B_4(p) < 65 \), \( B_4(p) > 75 \), then values 162 and 198/171 and 209/207 and 253/315 and 385/558 and 682 are used instead of 81 and 99/126 and 154/162 and 198/207 and 253/765 and 935, respectively.

**Age** \( B_5 \) represents the age of the patient in years. According to [17], the average life expectancy at birth is 78.1 for males and 82.1 for females. And so a four year difference was used in the defined membership degrees. If \( B_5(p) = \text{ male } \), \( a_{5,1}(p) = young(p) \) equals 1 if \( B_5(p) \leq 28 \), 0.48 for \( B_5(p) \in \{28;48\} \), and 0 for \( B_5(p) \geq 48 \). Membership degree \( a_{5,2} = mid\ aged(p) \) equals 0 for \( B_5(p) \leq 28 \), 1 for \( B_5(p) \geq 68 \), 0.48 for \( B_5(p) \in \{28;48\} \), and 0 for \( B_5(p) \geq 68 \). Membership degree \( a_{5,3} = old(p) \) equals 0 for \( B_5(p) \leq 48 \) or \( B_5(p) \geq 88 \).
Sex \((B_6)\) indicates if the patient is female or male. Linguistic terms defined for linguistic variable \(Sex (A_6)\) match possible values of \(B_6\), i.e. values \(female\), \(male\). Their particular membership functions are defined as follows. Membership degree \(a_{6,1}(p) = female(p)\) equals 1 if \(B_6(p) = female\) and 0 if \(B_6(p) = male\). Membership degree \(a_{6,2}(p) = male(p)\) equals 1 if \(B_6(p) = male\) and 0 if \(B_6(p) = female\).

**Pulse Rate** \((B_7)\) is the rate of the patient’s pulse measured by tactile on the outside of an artery in beats per minute. Membership degree \(a_{7,1}(p) = low(p)\) equals 1 for \(B_7(p) \leq 48\), \(\frac{72 - B_7(p)}{72 - 48}\) for \(B_7(p) \in (48; 72)\), and 0 for \(B_7(p) \geq 72\). Membership degree \(a_{7,2}(p) = normal(p)\) equals 0 for \(B_7(p) \leq 48\) or \(B_7(p) \geq 112\), \(\frac{B_7(p) - 48}{112 - 48}\) for \(B_7(p) \in (48; 72)\), 1 for \(B_7(p) \in [72; 88]\), and \(\frac{112 - B_7(p)}{112 - 88}\) for \(B_7(p) \in (88; 112)\). Membership degree \(a_{7,3}(p) = high(p)\) equals 0 for \(B_7(p) \leq 48\), \(\frac{B_7(p) - 48}{112 - 48}\) for \(B_7(p) \in (88; 112)\), and 1 for \(B_7(p) \geq 112\).

**Weight** \((B_8)\) denotes the patient's weight in kilograms. The patient's weight depends on the height and on the sex. The height is incorporated with Body Mass Index (BMI). Suppose \(BMI(p) = \frac{B_8(p)}{B_6(p)^2}\). If \(B_6(p) = male\), membership degree \(a_{8,1}(p) = underweight(p)\) equals 1 for \(BMI(p) \leq 16.55\), \(\frac{20.45 - 16.55}{BMI(p) - 16.55}\) for \(BMI(p) \in (16.55; 20.45)\), and 0 for \(BMI(p) \geq 20.45\). Membership degree \(a_{8,2}(p) = normal(p)\) equals 0 for \(BMI(p) \leq 16.55\) or \(BMI(p) \geq 26.5\), \(\frac{29.45 - 16.55}{BMI(p) - 16.55}\) for \(BMI(p) \in (16.55; 20.45)\), 1 for \(BMI(p) \in [20.45; 23.05]\), \(\frac{26.5 - 23.05}{BMI(p) - 23.05}\) for \(BMI(p) \in (23.05; 26.5)\). Membership degree \(a_{8,3}(p) = overweight(p)\) equals 0 for \(BMI(p) \leq 23.05\) or \(BMI(p) \geq 31.5\), \(\frac{31.5 - BMI(p)}{26.5 - BMI(p)}\) for \(BMI(p) \in (23.05; 26.5)\), 1 for \(BMI(p) \in [26.5; 28.5]\), \(\frac{31.5 - BMI(p)}{31.5 - 28.5}\) for \(BMI(p) \in (28.5; 31.5)\). Membership degree \(a_{8,4}(p) = obese(p)\) equals 0 for \(BMI(p) \leq 28.5\), \(\frac{BMI(p) - 28.5}{BMI(p) - 28.5}\) for \(BMI(p) \in (28.5; 31.5)\), and 1 for \(BMI(p) \geq 31.5\).

**Height** \((B_9)\) is the patient's height in meters. According to [16], in the UK, the average height of men is 1.754 meters and the average height of women is 1.619 meters. If \(B_9(p) = male\), \(a_{9,1}(p) = short(p)\) equals 1 for \(B_9(p) \leq 1.654\), \(\frac{1.754 - B_9(p)}{1.754 - 1.654}\) for \(B_9(p) \in (1.654; 1.754)\), and 0 for \(B_9(p) \geq 1.754\). Membership degree \(a_{9,2}(p) = average(p)\) equals 0 if \(B_9(p) \leq 1.654\) or \(B_9(p) \geq 1.854\), \(\frac{B_9(p) - 1.654}{1.854 - B_9(p)}\) for \(B_9(p) \in (1.654; 1.754)\), and \(\frac{1.854 - B_9(p)}{1.854 - 1.754}\) for \(B_9(p) \in [1.754; 1.854}\). If \(B_9(p) = female\), then values 1.519, 1.619, 1.719 are used instead of 1.654, 1.754, 1.854, respectively.

Class attribute **Patient Status** \((D)\) is a categorical attribute. Linguistic terms defined for class linguistic variable **Patient Status** \((C)\) match possible values of \(D\), i.e. values **alive**, **dead**. Both **alive** and **dead** are used as it allows us to produce more understandable fuzzy rules even though one linguistic term would be enough in this case. Their particular membership functions are defined as follows. Membership degree \(c_1(p) = alive(p)\) equals 1 if \(D(e) = alive\) and 0 if \(D(e) = dead\). Membership degree \(c_2(p) = dead(p)\) equals 1 if \(D(e) = dead\) and 0 if \(D(e) = alive\).

### III. ALGORITHMS FOR BUILDING FUZZY DECISION TREES

Two different algorithms for building of fuzzy decision trees are presented here together with ways how to transform them into fuzzy rules and how to determine the values of \(c_j(p), c_j \in C\), for a patient \(p\). The main difference between them is the measure used for association of a linguistic variable with a decision node. One algorithm, based on [19], uses classification ambiguity and choses the linguistic variable with its lowest value. The other algorithm, based on [10], uses mutual information criterion and choses the linguistic variable with its highest value. The original algorithms in [19] and [10] are adapted to the heart failure domain through expert fuzzification presented in the previous section. They are also described in an easily implementable way here.

#### A. Classification Ambiguity

The algorithm based on classification ambiguity is presented in this subsection. It uses classification ambiguity for association of a linguistic variable with a decision node. It has the following five input parameters: \(A, D, P\), significance level \(\alpha \in [0; 1]\), degree-of-truthfulness threshold \(\beta \in [0; 1]\). Significance level \(\alpha\) serves as a filter of insignificant membership degrees for \(p \in P\). Degree-of-truthfulness threshold controls the minimal truthfulness of fuzzy rules that can be obtained from the build fuzzy decision tree. The decision tree is built as follows:

1. Make the root and associate one of linguistic variables \(A_k \in A\) with the minimal ambiguity to it. Mark the variable as \(\min A_k\). Make a branch for each \(a_{k,l} \in \min A_k\), connect it 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 \(A_k\) and \(a_{k,l}\) from the root to the current branch (included as “\(A_k\) is \(a_{k,l}\)) connected with operator AND;
3. Compute the degree of truthfulness for \(E\) and each \(c_j \in C\). Choose one \(c_j\) for which it has the maximal value. The maximal value is marked as max, and the particular \(c_j\) is marked as \(c_{max}\). If \(\max \geq \beta\) then make a leaf which is associated with linguistic term \(c_{max}\), connect it with the current branch, consider the current branch processed and go to step 2). Otherwise, go to step 4);
4. If there is no \(A_k \in A\), \(A_k\) is not in \(E\), consider the current branch processed and go to step 2). Otherwise, find \(A_k \in A\), \(A_k\) is not in \(E\), with the minimal ambiguity. Mark
The patient status of a patient \( p \) is determined as follows:

1) Make the root and associate one of linguistic variables \( A_k \in A \) with the maximal mutual information to it. Mark the variable as \( \text{max}A_k \). Make a branch for each \( a_{k,i} \in \text{max}A_k \), connect them with the root, and associate them with the particular \( a_{k,i} \) 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 \( A_k \) and \( a_{k,i} \) from the root to the current branch (including as “\( A_k \) is \( a_{k,i} \)” connected with operator AND).

3) Compute the cumulative information of linguistic condition \( E \) for known patients \( P \) and mark it as \( \text{BRANCH} \). Compute the conditional cumulative information of alive/dead for known patients \( P \) provided that linguistic condition \( E \) is used. Choose the minimal value and mark it as \( \text{minCLASS} \). If \( \{ \text{BRANCH} \geq -\log_2(\alpha \cdot \#(P)) \} \) or \( \{ \text{minCLASS} \leq -\log_2(\beta) \} \) or \( E \) contains all linguistic variables in \( A \), go to step 4), otherwise go to step 5);

4) Make a leaf, connect it with the current branch and consider this branch processed. Compute two raised by the exponent of the negative value of the conditional cumulative information of alive/dead for known patients \( P \) provided that linguistic condition \( E \) is used and mark it as \( F(\text{alive})/F(\text{dead}) \). Choose \( F(c_j), c_j \in \{\text{alive}; \text{dead}\} \), with the higher value or choose one of them randomly if they both have the same value. Associate \( c_j \in \{\text{alive}; \text{dead}\} \) which is in the chosen \( F(c_j), F(\text{alive}), F(\text{dead}) \) with the made leaf. Go to step 2);

5) For all \( A_k \in A \) and \( A_k \) which are not in \( E \), compute the mutual information that determines the amount of information which is obtained about \( \text{Patient Status} \) and mark it as \( \text{max}A_k \). Make a node, connect it with the current branch and associate it with linguistic variable \( \text{max}A_k \). Consider the current branch processed. Make a branch for each \( a_{k,i} \in \text{max}A_k \), connect them with the made node, associate them with particular \( a_{k,i} \) and consider them unprocessed. Go to step 2).

The decision tree is transformed into fuzzy rules as follows:

1) Make the root and associate one of linguistic variables \( A_k \in A \) with the maximal mutual information to it. Mark the variable as \( \text{max}A_k \). Make a branch for each \( a_{k,i} \in \text{max}A_k \), connect them with the root, and associate them with the particular \( a_{k,i} \) 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 \( A_k \) and \( a_{k,i} \) from the root to the current branch (including as “\( A_k \) is \( a_{k,i} \)” connected with operator AND).

3) Compute the cumulative information of linguistic condition \( E \) for known patients \( P \) and mark it as \( \text{BRANCH} \). Compute the conditional cumulative information of alive/dead for known patients \( P \) provided that linguistic condition \( E \) is used. Choose the minimal value and mark it as \( \text{minCLASS} \). If \( \{ \text{BRANCH} \geq -\log_2(\alpha \cdot \#(P)) \} \) or \( \{ \text{minCLASS} \leq -\log_2(\beta) \} \) or \( E \) contains all linguistic variables in \( A \), go to step 4), otherwise go to step 5);


IV. EXPERIMENTAL RESULTS

The main purpose of the experiments was to compare the performance of presented fuzzy decision tree methods with each other and with other data mining methods. Experiments were carried out with our software tool written in Java. The core algorithms, other than fuzzy decision tree algorithms, were implemented in Weka [18]. The performance of algorithms was measured with sensitivity $= \frac{tp}{tp+fn}$ and specificity $= \frac{tn}{tn+fp}$ where $tp/fp/fn/tn$ is the number of true positives/false positives/false negatives/true negatives. “C is alive”/”D is alive” was considered negative, “C is dead”/”D is dead” was considered positive. Values $tp$, $fp$, $fn$ and $tn$ were computed during 10-fold cross-validation where the (fuzzified) heart failure data was partitioned randomly into 10 folds of patients. It was made sure that all folds contained roughly the same proportions of alive and dead patients. A patient was considered dead/alive if the value assigned to attribute $D$ was dead/alive. In the fuzzified data, a patient $p$ was considered dead if $\text{dead} \in \arg\max_{E} c_{i}(p)$; otherwise the patient was alive. Of the 10 folds, a single fold was retained as the testing data, and the remaining $9$ folds were used as the learning data. The learning data was analyzed by the method and the cross-validation process was repeated 10 times, with each of the 10 folds used exactly once as the testing data. 

<table>
<thead>
<tr>
<th>Method</th>
<th>Sen</th>
<th>Spec</th>
<th>Sen + Spec</th>
</tr>
</thead>
<tbody>
<tr>
<td>FDT-CA</td>
<td>62.5000</td>
<td>66.7328</td>
<td>129.2328</td>
</tr>
<tr>
<td>FDT-CIE</td>
<td>24.2308</td>
<td>93.1217</td>
<td>117.3525</td>
</tr>
<tr>
<td>Bayes</td>
<td>24.0385</td>
<td>95.3704</td>
<td>119.4084</td>
</tr>
<tr>
<td>C4.5</td>
<td>40.9615</td>
<td>87.8968</td>
<td>128.8583</td>
</tr>
<tr>
<td>MLP</td>
<td>32.1154</td>
<td>93.9815</td>
<td>126.0969</td>
</tr>
<tr>
<td>NNC</td>
<td>24.6154</td>
<td>88.3598</td>
<td>112.9752</td>
</tr>
</tbody>
</table>

The results of 10-fold cross-validation are given in Tab. II. FDT-CA is the fuzzy decision tree method described in Subsection III.A. The parameters which gave the best results were used. Concretely, $\alpha$ was set to 0.6 and $\beta$ was set to 0.8. FDT-CIE is the fuzzy decision tree method described in Subsection III.B. Parameter $\alpha$ was set to 0 and parameter $\beta$ was set to 0.9. Bayes denotes a Bayesian network method which is implemented in Weka as class BayesNet. C4.5 represents a decision tree method which is implemented in Weka as class J48. MLP is a feedforward neural network method using multilayer perception which is implemented in Weka as class MultilayerPerceptron. NNC is a nearest neighbor method using non-tested generalized examples which are implemented in Weka as class NNge. Sen is sensitivity in percentages and Spec is specificity in percentages.

Data in Tab. II are interpreted in the form of a graph in Fig. 1. The results are shown as plots in the ROC space. The distance from the random guess line is an indicator of how well the method predicts if a patient dies within six months. The sum of sensitivity and specificity could also be used as such an indicator. It is very important to avoid classification of dead patients as alive as it would lead to life-threatening situations. On the other hand, many alive patients classified as dead ones would increase the running costs of the treatment considerably. From the results Fig. 1 it can be concluded that a fuzzy decision tree based on classification ambiguity (FDT-CA) achieves considerably better results on our expert fuzzified data than a fuzzy decision tree based on cumulative information estimations (FDT-CIE). It is because FDT-CA takes both ambiguity and vagueness into consideration while FDT-CIE deals only with vagueness in our heart failure data. It is interesting to see that C4.5 decision tree achieves results which are comparable to FDT-CA. However, the interpretability of the C4.5 decision tree is not as favorable as it is in the case of FDT-CA. FDT-CA uses data which is fuzzified according to the wish of medical experts so that it reflects their understanding of the field. Numerical attributes for the C4.5 decision tree are discretized automatically. Also, the C4.5 decision tree does not take cognitive uncertainties such as ambiguity and vagueness into consideration.

![ROC space and plots for all methods.](image)

<table>
<thead>
<tr>
<th>Measure</th>
<th>FDT-CA</th>
<th>FDT-CIE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of fuzzy rules</td>
<td>15.8000</td>
<td>1200.2000</td>
</tr>
<tr>
<td>Length of a fuzzy rule</td>
<td>3.4684</td>
<td>8.0033</td>
</tr>
<tr>
<td>Longest fuzzy rule</td>
<td>5.1000</td>
<td>9.0000</td>
</tr>
<tr>
<td>Shortest fuzzy rule</td>
<td>2.0000</td>
<td>2.0000</td>
</tr>
</tbody>
</table>

The fuzzy decision tree methods also use a transformation of the fuzzy decision trees to fuzzy rules. The interpretability of these fuzzy rules is based on measures derived from [5]. The measures are computed for ten groups of fuzzy rules (learning groups) discovered for particular nine folds in ten-fold cross-validation and the average is taken. Number of fuzzy rules is the average number of fuzzy rules in the ten learning groups. Length of a fuzzy rule is the average number of linguistic
variables in the conditions of all fuzzy rules in the ten learning groups. Longest fuzzy rule is any fuzzy rule in all ten learning groups with the highest number of linguistic variables in its condition. Shortest fuzzy rule is any fuzzy rule in all ten learning groups with the lowest number of linguistic variables in its condition. The values for particular measures are in Tab. III. Fuzzy rules transformed from the fuzzy decision tree based on classification ambiguity (FDT-CA) have a much lower number of fuzzy rules and length of fuzzy rules than fuzzy rules transformed from the fuzzy decision tree based on cumulative information estimations (FDT-CIE).

V. CONCLUSIONS

Two fuzzy decision tree methods for prediction of the death of a patient with heart failure within six months were presented. They consisted of expert fuzzification of attributes from the Hull LifeLab data with 2032 patients, algorithms for creation of the fuzzy decision trees, ways of their transformation into fuzzy rules, and algorithms for prediction. Fuzzification allowed us to create an understandable and natural description of the attributes and also it allowed us to deal with vagueness and ambiguity in the data. The methods should also minimize life-threatening situations and minimize costs. Life-threatening situations appear when patients who are at risk of death within a short time are considered alive, which is measured by sensitivity. These situations should be minimized and so sensitivity should be maximized. Costs are increased when patients with a low risk of death are treated as if they could die, which is measured by specificity. For the costs to be minimized, specificity should be maximized. For the Hull LifeLab data about 2032 patients, a fuzzy decision tree based on classification ambiguity with sensitivity 62.5000% and specificity 66.7328% and a fuzzy decision tree based on cumulative information estimations with sensitivity 24.2308% and specificity 93.1217% were found. The former achieved considerably better results, which was attributed to the incapability of the latter to deal with ambiguity and expert fuzzification. The results of the fuzzy decision trees were competitive to the sensitivity and specificity of other data mining methods such as a Bayesian network, a C4.5 decision tree, a nearest neighbor method, and a neural network classifier. In comparison to the C4.5 decision tree, the fuzzy decision tree based on classification ambiguity achieved a slightly higher sum of sensitivity and specificity. In addition to it, it enjoyed the advantages of expert fuzzification such as understandability and incorporation of cognitive uncertainties. Its interpretability was very good, and after it was transformed into fuzzy rules, the average number of fuzzy rules was 15.8000, the average length of a fuzzy rule was 3.4684 and the average longest length of a fuzzy rule was 5.1000. Overall, two different types of analyzed fuzzy decision trees had significantly different accuracy results and interpretability.

ACKNOWLEDGMENT

The authors would like to thank the University of Hull, UK for its equipment and support. The heart failure data was provided by the Department of Cardiology, Hull York Medical School, University of Hull, UK as a part of Hull LifeLab.

REFERENCES