2014 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)
July 6-11, 2014, Beijing, China

Use of Cumulative Information Estimations for Risk Assessment of Heart Failure Patients

Jan Bohacik
Department of Computer Science
University of Hull
Hull, United Kingdom
j.bohacik@hull.ac.uk
Department of Informatics
University of Zilina
Zilina, Slovakia
jan.bohacik@fri.uniza.sk

C. Kambhampati[*], Darryl N. Davis[*], J. G. F. Cleland[†]
[*]Department of Computer Science
University of Hull
Hull, United Kingdom
{c.kambhampati, d.n.davis}@hull.ac.uk
[†]Department of Cardiology
University of Hull
Hull, United Kingdom
j.g.cleland@hull.ac.uk

***Abstract*— As a consequence of aging population and an increasing prevalence of obesity and diabetes there are more and more patients with heart failure. This leads to a lack of professionals who can treat them and to escalating costs. An interesting solution appears to be home telemonitoring with an intelligent clinical decision support system. In this paper, the use of cumulative information estimations for risk assessment of heart failure patients with such a system is analyzed. These cumulative information estimations are utilized for creation of an algorithmic model using fuzzy decision trees that combine decision trees and notions of fuzzy logic. The algorithmic model employs mutual cumulative information and relative mutual cumulative information for association of an important piece of data about the patients with a decision node. The risk assessment with the presented solution is analyzed from the point of view of minimization of life-threatening situations and minimization of costs. Comparisons with a Bayesian network method, a nearest neighbor method, and a logistic regression method show it is a promising solution.**

Keywords—cumulative information estimation, decision tree, home telemonitoring, e-health, heart failure, cardiology

I. INTRODUCTION

Given the rapidly growing aging population, the increased prevalence of obesity and diabetes, the increased burden of heart failure, and the increasing healthcare costs, there is an urgent need for the development, implementation, and deployment of new models of healthcare services. The burden of heart failure affects 2%-3% of the adult population with disabling symbols, the most common of which are fatigue and dyspnea, while in terms of disability, the end stage of the disease is comparable to the end stage of terminal cancer [13]. 2%-3% of the adult population means over 26 million people with heart failure worldwide and this number is growing rapidly with newly diagnosed people every year. In the UK, heart failure affects about 900,000 people with 60,000 new cases annually [5]. Heart failure is characterized by a poor prognosis: up to 70% of all patients with heart failure die within 5 years after their first hospital admission [18]. In addition to a poor prognosis, a common feature of advanced heart failure is multiple hospital (re-)admissions [5]. A promising strategy that can cope with these challenges seems to be a greater use of home telemonitoring in which physiological data is transferred from the patients' home to the center to monitor them, interpret the data, and make clinical decisions [14]. Home telemonitoring should be integrated with a clinical decision support system that identifies both the nature and optimal response to a problem rather than just its mere existence [4].

Clinical outcomes such as hospital admission and mortality are normally used as regression or classification goals in the analysis of medical data. Traditionally, techniques belonging to the data modeling culture rather than to the algorithmic modeling culture are employed there. Existing methods include Cox regression [6], EFFECT Risk Scoring system [10], Emergency Heart Failure Mortality Risk Grade (EHMRG) [11], logistic regression [1] or Seattle Heart Failure Model (SHFM) [8]. Algorithmic models such as data mining methods do not usually assume there is a causal relationship between home telemonitoring data and clinical outcomes, which is useful as there does not have to be any. Moreover, their goal is predictive accuracy primarily and some of them can store the knowledge about patients in an easily interpretable and understandable way. Data mining methods such as decision tree methods, nearest neighbor methods and neural network methods are employed in [2][15]-[17][21]. In [2], a knowledge based platform of services for more effective and efficient clinical management of heart failure within elderly population is presented. A platform to enhance effectiveness and efficiency of any worsening in a heart failure patient's condition is proposed in [15]. In [16], hospitalization for heart failure is predicted with decision tree methods, nearest neighbor methods and neural network methods. Papers [17] and [21] discuss the issues of preprocessing in heart failure data.

The research reported in the paper considers risk assessment of heart failure patients, i.e. predicting the possibility of death for a heart failure patient within six months, with an algorithmic model creating knowledge represented as a fuzzy decision tree or a set of IF-THEN rules derived from the tree. This set of IF-THEN rules is used by an algorithm that creates the predictions. The fuzzy decision tree and the derived set of IF-THEN rules make use of the notions of fuzzy logic such as fuzzy sets, membership degrees and linguistic variables. The incorporation of fuzzy logic allows us to take cognitive uncertainties such as vagueness into consideration. Vagueness is associated with the difficulty to make clear or precise distinctions in the real world [9]. For example, it is strange to consider a patient's age "young" when the patient is 42 and "mid aged" when the patient is 43. Small changes in numerical values can cause changes in categorical values, which can lead to significant changes in predictions [19]. The knowledge should also be easily readable by the clinicians so that they can make more sophisticated data interpretation and decision-making, and for this reason the number of IF-THEN rules in the set and their lengths are analyzed. The computations in the algorithms within the domain of fuzzy logic are based on cumulative information estimations offering criteria for association of an important piece of data about heart failure patients with the decision node.

The organization of the paper is as follows. Section II describes the data about heart failure patients and its processing. Employed cumulative information estimations are formulated in Section III. The algorithmic model used for risk assessment of heart failure patients is explained in Section IV. Section V contains the experimental results. Section VI concludes the paper.

II. CLINICAL HEART FAILURE DATA

The data used is extracted from Hull LifeLab which is large, epidemiologically representative, information-rich clinical data [3]. It contains information about 2032 heart failure patients which are studied for the purpose of diagnosis, delivery of improved treatment to patients, and estimation of associated costs to health services and society. Mathematically, the patients are represented as set $\boldsymbol{P}$, i.e. each patient $\boldsymbol{p} \in \boldsymbol{P}$. Nine attributes $\boldsymbol{B}$ are identified with queries about clinical findings and physiological measurements in the data. These are referred to as describing attributes in the paper and they are systematically characterized in Table I. Mathematically, these attributes are defined as $\boldsymbol{B} = \{B_1; \ldots; B_k; \ldots; B_9\}$. If B_k is a categorical attribute, $B_k = \{b_{k,1}; \ldots; b_{k,l}; \ldots; b_{k,l_k}\}$ where $b_{k,1}, \ldots, b_{k,l}, \ldots; b_{k,l_k}$ are possible categorical values. The data is further divided into two groups: a) patients who passed away within six months after the data had been obtained and b) patients who were alive six and more months after the data had been obtained. This is reflected in the *Risk Outcome* (D) attribute (referred to as the class attribute as well) with two possible categorical values d_1 (*dead*) and d_2 (*alive*). Mathematically, $D = \{d_1; d_2\}$. The particular value for an attribute B_k (D) and a patient $\boldsymbol{p} \in \boldsymbol{P}$ is marked as $B_k(\boldsymbol{p})$ ($D(\boldsymbol{p})$).

TABLE I. IDENTIFIED ATTRIBUTES IN THE HEART FAILURE DATA

Attribute	Data Type	Values
NT-proBNP Level (B_1)	Numerical	0.89 – 18236
Pulse Rate (B_2)	Numerical	38 - 150
Sex (B_3)	Categorical	*male* ($b_{1,1}$)
		female ($b_{1,2}$)
Age (B_4)	Numerical	27 - 96
Height (B_5)	Numerical	1.20 – 1.96
Weight (B_6)	Numerical	29.80 – 193.80
Blood Creatinine Level (B_7)	Numerical	37 - 1262
Blood Sodium Level (B_8)	Numerical	123 - 148
Blood Uric Acid Level (B_9)	Numerical	0.11 - 1.06
Risk Outcome (D)	Categorical	*dead* (d_1)
		alive (d_2)

Vagueness and interpretability are taken into consideration through fuzzification of attributes in $\boldsymbol{B}$ and attribute D. Each attribute B_k in $\boldsymbol{B}$ is fuzzified into linguistic variable $A_k = \{a_{k,1}; \ldots; a_{k,l}; \ldots; a_{k,l_k}\}$, i.e. describing linguistic variables $\boldsymbol{A} = \{A_1; \ldots; A_k; \ldots; A_9\}$. Attribute D is fuzzified into class linguistic variable $C = \{c_1; c_2\}$. The data is also transformed through definition of membership degrees so that fuzzification is reflected. Membership degrees $a_{k,l}(\boldsymbol{p})$ ($c_j(\boldsymbol{p})$) are defined for all $a_{k,l} \in A_k \in \boldsymbol{A}$ (all $c_j \in C$) and all $\boldsymbol{p} \in \boldsymbol{P}$ using our expert knowledge.

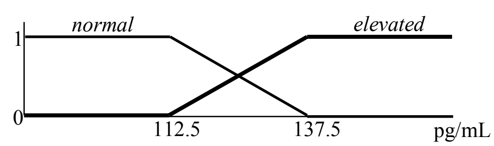

Fig. 1. Fuzzification of *NT-proBNP Level* for patients who are 75 years old or younger.

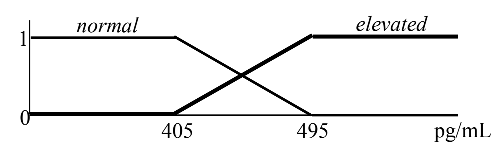

Fig. 2. Fuzzification of *NT-proBNP Level* for patients older than 75.

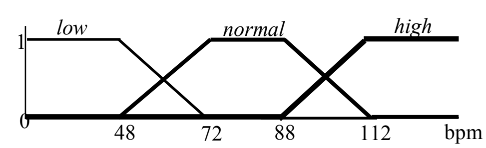

Fig. 3. Fuzzification of *Pulse Rate*.

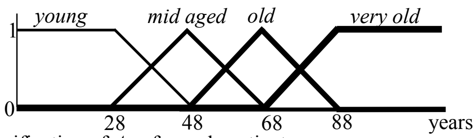

Fig. 4. Fuzzification of *Age* for male patients.

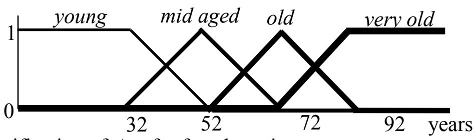

Fig. 5. Fuzzification of *Age* for female patients.

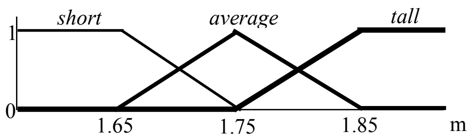

Fig. 6. Fuzzification of *Height* for male patients.

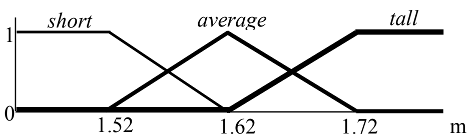

Fig. 7. Fuzzification of *Height* for female patients.

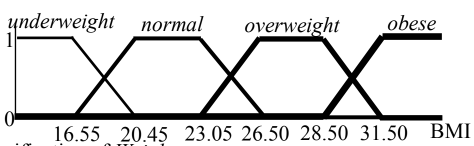

Fig. 8. Fuzzification of *Weight*.

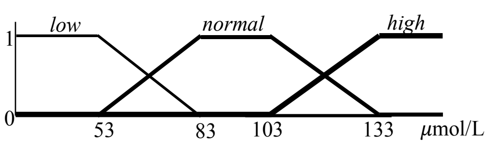

Fig. 9. Fuzzification of *Blood Creatinine Level* for male patients.

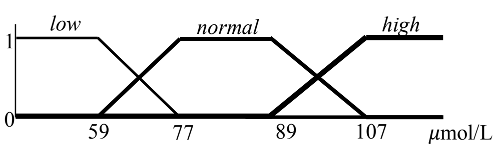

Fig. 10. Fuzzification of *Blood Creatinine Level* for female patients.

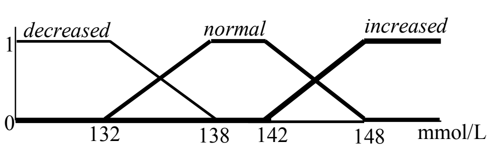

Fig. 11. Fuzzification of *Blood Creatinine Level* for female patients.

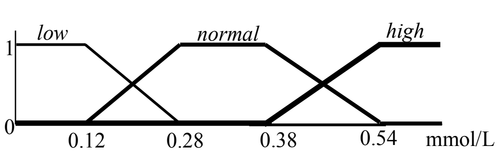

Fig. 12. Fuzzification of *Blood Uric Acid Level* for male patients.

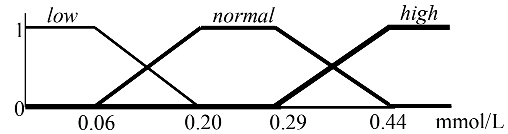

Fig. 13. Fuzzification of *Blood Uric Acid Level* for female patients.

Definition of all membership degrees $a_{k,l} \in A_k \in \boldsymbol{A}$ but any $a_{3,l} \in A_3$ is through membership functions in Fig. 1, Fig. 2, Fig. 3, Fig. 4, Fig. 5, Fig. 6, Fig. 7, Fig. 8, Fig. 9, Fig. 10, Fig. 11, Fig. 12, and Fig. 13. Linguistic terms defined for linguistic variable *Sex* (A_3) match possible values of attribute B_3, i.e. values *male* and *female*. Membership degree $a_{3,1}(\boldsymbol{p}) = male(\boldsymbol{p})$ equals 1 if $B_3(\boldsymbol{p}) = male$ and 0 if $B_3(\boldsymbol{p}) = female$. Membership degree $a_{3,2}(\boldsymbol{p}) = female(\boldsymbol{p})$ equals 1 if $B_3(\boldsymbol{p}) = female$ and 0 if $B_3(\boldsymbol{p}) = male$. Linguistic terms defined for linguistic variable *Risk Outcome* (C) match possible values of attribute D, i.e. values *dead* and *alive*. Membership degree $c_1(\boldsymbol{p}) = dead(\boldsymbol{p})$ equals 1 if $D(\boldsymbol{p}) = dead$ and 0 if $D(\boldsymbol{p}) = alive$. Membership degree $c_2(\boldsymbol{p}) = alive(\boldsymbol{p})$ equals 1 if $D(\boldsymbol{p}) = alive$ and 0 if $D(\boldsymbol{p}) = dead$.

III. Cumulative Information Estimations

Marks, terms, and cumulative information estimations employed in the paper are defined and summarized here. Let *Height* be a linguistic variable defined as *Height* = {*short*; *average*; *tall*}. It is said *short*, *average*, *tall* are associated with (are defined for) *Height*. Membership degree to which a patient $\boldsymbol{p}$ is an element of the fuzzy set associated with *short* is symbolized by *short*($\boldsymbol{p}$). If $\boldsymbol{M}$ is the fuzzy set associated with *short*, #(*short*) is the cardinality of $\boldsymbol{M}$. If *short* is chosen from the linguistic terms predefined for *Height*, it is denoted by '*Height* is *short*'. 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 $\mathrm{T}(\mathrm{a}; \mathrm{b}) = \mathrm{a} \cdot \mathrm{b}$.

3.1 Definition. Let $\boldsymbol{U}$ be the set of all possible patients with heart failure and let $\boldsymbol{p} \in \boldsymbol{U}$ be described by linguistic variables $\boldsymbol{A} = \{A_1; \ldots; A_k; \ldots; A_n\}$. A *linguistic condition* E is a linguistic term associated with a subset terms(E) of linguistic terms defined for variables in $\boldsymbol{A}$. Its lexical name is a connection of terms in terms(E) with conjunction 'AND'. For any possible variable in $\boldsymbol{A}$ 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 $E(\boldsymbol{p})$, $\boldsymbol{p} \in \boldsymbol{U}$, is defined as follows: $E(\boldsymbol{p}) = 1$ if terms(E) = $\emptyset$, otherwise the value of $E(\boldsymbol{p})$ is the result of t-norm applied on all $a_{k,l} \in \mathrm{terms}(E)$.

Linguistic term $a_{k,l}$ in terms(E) of a linguistic condition E, $a_{k,l} \in A_k \in \boldsymbol{A}$, is equally replaced by 'A_k is $a_{k,l}$' and vice versa. $E = \emptyset / E \neq \emptyset$ 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), $A_k \in E / A_k \notin E$. Symbol $E \backslash A_k$ 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 $(E \cup a_{k,l})(\boldsymbol{p}) = \mathrm{T}(E(\boldsymbol{p}); a_{k,l}(\boldsymbol{p}))$, where 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 variable C. Membership degree $(E \cup c_j)(\boldsymbol{p})$ equals $\mathrm{T}(E(\boldsymbol{p}); c_j(\boldsymbol{p}))$.

The cumulative information estimations such as cumulative information, conditional cumulative information, mutual information, cumulative information, and relative mutual information are introduced in [12] and they are formulated with the use of the above-mentioned marks and terms in the following definitions.

3.2 Definition. *Cumulative information of linguistic condition E (linguistic term $E \cup c_j$, $c_j \in C$)* for known patients $\boldsymbol{P}$ is:

$$\mathbf{II}(E; \boldsymbol{P}) = \begin{cases} -log_2\#(E); \text{if } E \neq \emptyset \\ -log_2\#(\boldsymbol{P}); \text{if } E = \emptyset \end{cases} \quad (1)$$

$$(\mathbf{II}(E \cup c_j; \boldsymbol{P}) = -log_2\#(E \cup c_j)). \quad (2)$$

3.3 Definition. *Information of linguistic condition E* (*linguistic term $E \cup c_j$, $c_j \in C$*) for known patients $\boldsymbol{P}$ is:

$$\mathbf{I}(E; \boldsymbol{P}) = log_2\#(\boldsymbol{P}) + \mathbf{II}(E; \boldsymbol{P}), \quad (3)$$

$$(\mathbf{I}(E \cup c_j; \boldsymbol{P}) = log_2\#(\boldsymbol{P}) + \mathbf{II}(E; \boldsymbol{P})). \quad (4)$$

3.4 Definition. *Conditional information* (*conditional cumulative information*) of $c_j \in C$ for known patients $\boldsymbol{P}$ provided that E is known is defined as:

$$\mathbf{I}(c_j / E; \boldsymbol{P})(\mathbf{II}(c_j / E; \boldsymbol{P})) = \mathbf{II}(E \cup c_j; \boldsymbol{P}) - \mathbf{II}(E; \boldsymbol{P}). \quad (5)$$

3.5 Definition. *Mutual information* $\mathbf{I}(C; A_k; E; \boldsymbol{P})$ for determining the amount of information which is obtained about C if values of $E(\boldsymbol{p})$, $\boldsymbol{p} \in \boldsymbol{P}$, $a_{k,l}(\boldsymbol{p})$, $a_{k,l} \in A_k \notin E$, $\boldsymbol{p} \in \boldsymbol{P}$, are known is defined as:

$$\mathbf{I}(C; A_k; E; \boldsymbol{P}) = \textstyle\sum_{c_j \in C} \sum_{a_{k,l} \in A_k} \#(E \cup a_{k,l} \cup c_j) \cdot \mathbf{TMP}(c_j; a_{k,l}; E; \boldsymbol{P}), \text{ where} \quad (6)$$

$$\mathbf{TMP}(c_j; a_{k,l}; E; \boldsymbol{P}) = \mathbf{II}(E \cup c_j; \boldsymbol{P}) + \mathbf{II}(E \cup a_{k,l}; \boldsymbol{P}) - \mathbf{II}(E \cup a_{k,l} \cup c_j; \boldsymbol{P}) - \mathbf{II}(E; \boldsymbol{P}). \quad (7)$$

3.6 Definition. *Cumulative entropy* of linguistic variable $A_k \in \boldsymbol{A}$ on known patients $\boldsymbol{p} \in \boldsymbol{P}$ is defined as:

$$\mathbf{HH}(A_k; \boldsymbol{P}) = \textstyle\sum_{a_{k,l} \in A_k} \#(a_{k,l}) \cdot \mathbf{II}(a_{k,l}; \boldsymbol{P}). \quad (8)$$

3.7 Definition. *Relative mutual information* $\mathbf{I}_{rel}(C; A_k; E; \boldsymbol{P})$ for determining the amount of information obtained about C independently from $\#(A_k)$ if values of $E(\boldsymbol{p})$, $\boldsymbol{p} \in \boldsymbol{P}$, $a_{k,l}(\boldsymbol{p})$, $a_{k,l} \in A_k \notin E$, $\boldsymbol{p} \in \boldsymbol{P}$, are known is defined as:

$$\mathbf{I}_{rel}(C; A_k; E; \boldsymbol{P}) = \frac{\mathbf{I}(C; A_k; E; \boldsymbol{P})}{\mathbf{HH}(A_k; \boldsymbol{P})}. \quad (9)$$

3.8 Definition. *Frequency of class linguistic term $c_j \in C$* for known patients $\boldsymbol{P}$ provided that E is known is defined as:

$$\mathbf{F}(c_j / E; \boldsymbol{P}) = 2^{-\mathbf{I}(c_j / E; \boldsymbol{P})}. \quad (10)$$

IV. ALGORITHMIC MODEL

The algorithmic model supposes known patients $\boldsymbol{P}$ are complex, and, at least, partly unknowable. What is observed is a set of describing linguistic variables $\boldsymbol{A} = \{A_1; \ldots; A_k; \ldots; A_9\}$ and class linguistic variable C. The problem is to find an algorithm such that for all membership degrees $a_{k,l}(\boldsymbol{p})$, $a_{k,l} \in A_k \in \boldsymbol{A}$, $\boldsymbol{p}$ is any heart failure patient, the algorithm will be a good predictor of $c_j(\boldsymbol{p})$, $c_j \in C$. The algorithmic model used in this paper utilizes an algorithm based on [12]. It creates knowledge about known patients and this knowledge is represented as a fuzzy decision tree or a set of IF-THEN rules derived from the tree. This set of IF-THEN rules is used by another algorithm that creates predictions for a patient $\boldsymbol{p} \in \boldsymbol{U}$, $\boldsymbol{U}$ is the set of all possible heart failure patients. The algorithmic model employs mutual cumulative information defined in Definition 3.5 and relative mutual cumulative information defined in Definition 3.7 for association of an important piece of data about known patients in $\boldsymbol{P}$ with a decision node. They are referred to as criterion for association $\mathbf{CA}(C; A_k; E; \boldsymbol{P})$ as one. The growth of the fuzzy decision tree is controlled by two parameters: frequency-of-branch threshold $\omega \in [0; 1]$ and frequency-of-class threshold $\delta \in [0; 1]$. Parameter ω controls its growth on the basis of frequency of branch. The higher the value of this parameter, the lower its height (or equally, the lower the number of assignments in the conditions of the IF-THEN rules) is. Parameter δ controls its growth on the basis of frequency of $c_j \in C$. The lower the value of the parameter, the lower its height (or the lower the number of assignments in the conditions of the IF-THEN rules) is. The algorithm for creation of knowledge represented as a fuzzy decision tree is as follows:

1) Create the root of the tree and associate linguistic variable $maxA_k = \mathrm{argmax}_{A_k \in \boldsymbol{A}}\{\mathbf{CA}(C; A_k; \emptyset; \boldsymbol{P})\}$ with it. Create a branch for each $a_{k,l} \in A_k$, associate each $a_{k,l}$ with the particular branch, connect the branches with the root, and consider the branches unprocessed;
2) END if there is no unprocessed branch. Otherwise, choose one of the unprocessed branches and consider it the current branch. For the current branch, create linguistic term E consisting of all assignments "*Linguistic variable* is *linguistic term*" in the path from the root to the current branch connected with operator "AND";
3) Set branchII = $\mathbf{II}(E; \boldsymbol{P})$ and set minClassI = $\min_{c_j \in C}\{\mathbf{I}(c_j / E; \boldsymbol{P})\}$. If {branchII $\geq -\log_2(\omega \cdot \#(\boldsymbol{P}))$} or {minClassI $\leq -\log_2(\delta)$} or {$\mathbf{A} \backslash E = \emptyset$}, go to step 4, otherwise step 5;
4) Create a leaf, connect this leaf with the current branch, and consider this branch to be processed. Associate frequency$_{c_j} = \mathbf{F}(c_j / E; \boldsymbol{P})$ for both $c_j \in C$ with the leaf. Associate class linguistic term $\mathrm{argmax}_{c_j \in C}\{\mathbf{F}(c_j / E; \boldsymbol{P})\}$ with the leaf. Go to step 2;
5) Create a node, connect it with the current branch, consider this branch processed. Associate $maxA_k = \mathrm{argmax}_{A_k \in \boldsymbol{A}, A_k \notin E}\{\mathbf{CA}(C; A_k; \emptyset; \boldsymbol{P})\}$ with the node. Associate each $a_{k,l} \in maxA_k$ with a newly created branch, connect the branches with the created node, and consider the branches unprocessed. Go to step 2.

A set of IF-THEN rules is derived as follows:

1) For each leaf $leaf_i$, mark the linguistic term associated with it as c^{leaf_i}. For each $leaf_i$, take the branch going to it and formulate linguistic condition E_i for this branch. E_i consists of all assignments "*Linguistic variable* is *linguistic term*" in the path from the root to the branch connected with operator "AND". For each $leaf_i$, set ECR_i = {$\text{frequency}_{c_j}^i$ | $\text{frequency}_{c_j}^i$ associated with $leaf_i$};
2) For each E_i, create a rule in the form of "IF E_i THEN C is c^{leaf_i} (ECR_i)".

The following algorithm uses a set of IF-THEN rules for risk assessment of a patient $\boldsymbol{p} \in \boldsymbol{U}$:

1) For each rule "IF E_i THEN C is c^{leaf_i} (ECR_i)", compute $E_i(\boldsymbol{p})$;
2) For each $c_j \in C$, set linguistic term $c_j(\boldsymbol{p}) = \sum_{\text{all possible } E_i} E_i(\boldsymbol{p}) \cdot \text{frequency}_{c_j}^i$, $\text{frequency}_{c_j}^i \in ECR_i$.

V. EXPERIMENTAL RESULTS

An experimental study aimed at comparison of the algorithmic model using cumulative information estimations with other methods was conducted. Two different criteria for association of a linguistic variable with a node for the algorithmic model were also compared. The core algorithms for the methods, other than the described algorithmic model, are implemented in Weka [20]. The performance of the methods is measured with sensitivity = (tp) / (tp + fn) and specificity = (tn) / (tn + fp). Value tp/fp/fn/tn is the number of true positives/false positives/false negatives/true negatives. "C is $dead$"/"D is $dead$" is considered positive and "C is $alive$"/"D is $alive$" is considered negative. Values tp, fp, fn and tn are computed during 10-fold cross-validation. The (fuzzified) data is partitioned into 10 folds. The partition is random, but all folds contain roughly the same proportions of alive and dead patients. A patient is considered dead/alive in the data if the value assigned to D is dead/alive. A patient is considered dead in the fuzzified data if $dead \in \text{argmax}_{c_j \in C}\{c_j(\boldsymbol{p})\}$, $\boldsymbol{p}$ is a known patient; otherwise the patient is alive. Of the 10 folds, a single fold is retained as the testing data for evaluation, and the remaining 9 folds are used as the learning data. The learning data is analyzed by the method. The validation is repeated 10 times, with each of the 10 folds used exactly once as the testing data.

The accuracy results achieved in 10-fold cross-validation are in Table II. MCI is the algorithmic model described in the previous section with mutual information as the criterion for association. The parameters which gave the best results were used (ω = 0.2 and δ = 0.9). RMCI is the algorithmic model described in the previous section with relative mutual information as the criterion for association (ω = 0.1 and δ = 0.9). BNM denotes a Bayesian network method implemented in Weka as class BayesNet. NNM is a feedforward neural network method using multilayer perception implemented in Weka as class MultilayerPerception. LRM denotes a logistic regression model implemented in Weka as class Logistic. Sen is sensitivity and Spec is specificity, both in percentages.

The sum of sensitivity and specificity is an indicator of how well the method predicts if a patient is alive or dead within six months. Sensitivity is associated with classification of dead patients as alive ones, which leads to life-threatening situations. Specificity is associated with classification of alive patients as dead ones, which leads to increase in costs. The higher the sum of sensitivity and specificity for a method, the better the method predicts. The algorithmic models described in the previous section (MCI and RMCI) achieve better results on the fuzzified heart failure data than the other methods according to Table II. However, the difference between MCI with 125.36% and RMCI with 125.24% is insignificant. It means the choice between mutual cumulative information and relative mutual cumulative information for association of a linguistic variable with a decision node in the algorithmic model is not important for the used fuzzified heart failure data. In general, it is likely unimportant if the number of linguistic terms defined for a linguistic variable is small such as three or four. The worst result with 112.98% is achieved with NNM.

Algorithmic models MCI and RMCI use decision trees for knowledge representation initially. The knowledge can also be interpreted as IF-THEN rules using the transformation described in Section IV. Particular groups of IF-THEN rules for particular algorithmic models can be compared using interpretability measures derived from [7]. The measures are computed for ten groups of IF-THEN rules discovered for particular nine folds in ten-fold cross-validation (learning data) and the average is taken. Number of IF-THEN rules is the average number of IF-THEN rules in the ten groups. Length of a fuzzy rule is the average number of linguistic variables in the conditions of all IF-THEN rules in the ten groups. Longest IF-THEN rule is any IF-THEN rule in all ten groups with the highest number of linguistic variables in its condition. Shortest IF-THEN rule is any IF-THEN rule in all ten groups with the lowest number of linguistic variables in its condition. The values for particular interpretability measures are in Table III. IF-THEN rules associated to the algorithmic model using mutual information as the criterion for association of a linguistic variable with a node (MCI) are much more interpretable than IF-THEN rules associated to the algorithmic model using relative mutual information (RMCI). The average number of IF-THEN rules is 11.00 for MCI in comparison to 24.80 for RMCI. The average length of an IF-THEN rule is 2.94 in comparison to 3.64 for RMCI. The average longest IF-THEN rule is 4.00 for MCI in comparison to 5.00 for RMCI. The average shortest IF-THEN rule is 1.00 for MCI in comparison to 2.00 for RMCI.

TABLE II. ACCURACY RESULTS

Method	Sen	Spec	Sen + Spec
MCI	34.42	90.94	125.36
RMCI	34.23	91.01	125.24
BNM	24.04	95.37	119.41
NNM	24.62	88.36	112.98
LRM	22.69	96.43	119.12

TABLE III. INTERPRETABILITY RESULTS

Interpretability Measure	MCI	RMCI
Number of IF-THEN rules (avg.)	11.00	24.80
Length of an IF-THEN rule (avg.)	2.94	3.64
Longest IF-THEN rule (avg.)	4.00	5.00
Shortest IF-THEN rule (avg.)	1.00	2.00

VI. CONCLUSION

An algorithmic model using a fuzzy decision tree or a set of IF-THEN rules derived from the tree and using cumulative information estimations in computations was employed for prediction of a heart failure patient's death within six months. In the process of learning, the algorithmic model was executed on data about heart failure patients where nine important pieces of data had been identified with queries about clinical findings and physiological measurements. The data had also been preprocessed with fuzzification allowing us to take vagueness in real-world medical situations and interpretability of the obtained knowledge into consideration. The accuracy of the obtained knowledge and the algorithmic model was evaluated using 10-fold cross-validation with the aim of minimization of life-threatening situations and minimization of costs. Interpretability of the knowledge was evaluated with measures such as number of IF-THEN rules, length of an IF-THEN rule, longest IF-THEN rule, and shortest IF-THEN rule. It was found that the difference between the use of mutual cumulative information and relative mutual cumulative information for association of a linguistic variable with a decision node is not significant for accuracy (125.36% vs. 125.24% for the sum of sensitivity and specificity). However, the use of mutual cumulative information led to much more interpretable knowledge (11.00 vs. 24.80 IF-THEN rules and 2.94 vs. 3.64 assignments in the condition of one IF-THEN rule). In the future, the accuracy of the obtained knowledge could be improved through the adoption of modifiers such as "*very*" or "*more or less*".

ACKNOWLEDGMENT

The reported research was financed within HEIF 5. The data describing heart failure patients was provided by the Department of Cardiology, Hull York Medical School, University of Hull, UK as a part of Hull LifeLab.

REFERENCES

[1] A. Bottle, P. Aylin, and A. Majeed, "Identifying patients at high risk of emergency hospital admissions: a logistic regression analysis," *JRSM*, vol. 99, no. 8, pp. 406-414, 2006.

[2] A. Candelieri, D. Conforti, F. Perticone, A. Sciacqua, K. Kawecka-Jaszcz, K. Styczkiewicz, "Early detection of decompensation conditions in heart failure patients by knowledge discovery: the HEARTFAID approaches," in *Proc. of Computers in Cardiology*, 2008, pp. 893-896.

[3] Clinical Effectiveness and Evaluation Unit of the Royal College of Physicians, *Managing Chronic Heart Failure: Learning from Best Practice*. Sudbury, Suffolk, UK: The Lavenham Press Ltd, 2005.

[4] J. G. F. Cleland, C. Lewinter, K. M. Goode, "Telemonitoring for heart failure: the only feasible option for good universal care?," *European Journal of Heart Failure*, vol. 11, no. 3, pp. 227-228, 2009.

[5] M. Connolly, J. Beattie, D. Walker, M. Dancy, A. Hayes, C. Henry, C. Jeffries, S. Machin - NHS Improvement. (2010, July). End of life care in heart failure: A framework for implementation. Available: http://www.endoflifecare.nhs.uk/assets/downloads/End_of_Life_Care_in_heart_failure.pdf

[6] S. R. Giolo, J. E. Krieger, A. J. Mansur, A. C. Pereira. (2012, June). Survival analysis of patients with heart failure: implications of time-varying regression effects in modelling mortality. *PLoS ONE*. 7(6): e37392. Available: http://www.plosone.org/article/info%3Adoi%2F10.1371%2Fjournal.pone.0037392

[7] H. Ishibuchi, T. Nakashima, M. Nii, *Classification and Modeling with Linguistic Information Granules: Advanced Approaches to Linguistic Data Mining* (1st ed.). Berlin, Germany: Springer Verlang, 2004.

[8] E. S. Ketchum, A. F. Jacobson, J. H. Caldwell, R. Senior, M. D. Cerqueira, G. S. Thomas, D. Agostini, J. Narula, and W. c. Levy, "Selective improvement in Seattle Heart Failure Model risk stratification using iodine-123 meta-iodobenzylguanidine imaging," *Journal of Nuclear Cardiology*, vol. 19, no. 5, pp. 1007-1016, 2012.

[9] G. J. Klir, "Where do we stand on measures of uncertainty, ambiguity, fuzziness and the like?," Fuzzy Sets and Systems, vol. 24, no. 2, pp. 141-160, 1987.

[10] D. S. Lee, P. C. Austin, J. L. Rouleau, P. P. Liu, D. Naimark, J. V. Tu, "Predicting Mortality Among Patients Hospitalized for Heart Failure: Derivation and Validation of a Clinical Model," *JAMA*, vol. 290, no. 19, pp. 2581-2587, 2003.

[11] D. S. Lee, A. Stitt, P.C. Austin, T.A. Stukel, M. J. Schull, A. Chong, G. E. Newton, J. S. Lee, J. V. Tu, "Prediction of Heart Failure Mortality in Emergent Care: A Cohort Study," *Annals of Internal Medicine*, vol. 156, no. 11, pp. 767-775, 2012.

[12] V. Levashenko, E. Zaitseva, "Usage of new information estimations for induction of fuzzy decision trees," in *Proc. 3rd International Conference on Intelligent Data Engineering and Automated Learning*, 2002, pp. 493-499.

[13] J. López-Sendón, "The heart failure epidemic," *Medicographia*, vol. 33, no. 4, pp. 363-369, 2011.

[14] G. Pare, K. Moqadem, G. Pineau, C. St-Hilaire. (2010, June). Clinical effects of home telemonitoring in the context of diabetes, asthma, heart failure and hypertension: A systematic review. *Journal of Medical Internet Research*. 12(2): e21. Available: http://www.jmir.org/2010/2/e21/

[15] L. Pecchia, P. Melillo, M. Bracale, "Remote health monitoring of heart failure with data mining via CART method on HRV features," *IEEE Transactions on Biomedical Engineering*, vol. 58, no. 3, pp. 800-804, 2011.

[16] K. T. Phillips, W. N. Street, "Predicting outcomes of hospitalization for heart failure using logistic regression and knowledge discovery methods," in *Proc. of AMIA Annual Symposium*, 2005, pp. 1080.

[17] N. Poolsawad, C. Kambhampati, J. G. F. Cleland, "Future selection approaches with missing values handling for data mining – A case study of heart failure," in *Proc. of the International Conference on Data Mining*, 2011, pp. 828-836.

[18] D. Postmus, D. J. v. Veldhuisen, T. Jaarsma, M. L. Luttik, J. Lassus, A. Mebazaa, M. S. Nieminen, V.-P. Harjola, J. Lewsey, E. Buskens, and H. L. Hillege. (2012, February). The COACH risk engine: a multistate model for predicting survival and hospitalization in patients with heart failure. *European Journal of Heart Failure*. 14(2), 168-175. Available: doi:10.1093/eurjhf/hfr163

[19] J. R. Quinlan, "Decision trees as probabilistic classifier," in *Proc. of the 3rd International Workshop on Machine Learning*, 1987, pp. 31-37.

[20] I. H. Witten, E. Frank, M. A. Hall, *Practical Machine Learning Tools and Techniques (3rd Edition)*. Burlington, MA, USA: Morgan Kaufman Publishers, 2011.

[21] Y. Zhang, C. Kambhampati, D. N. Davis, K. Goode, J. G. F. Cleland, "A comparative study of missing value imputation with multiclass classification for clinical heart failure data," *Proc. of the International Conference on Fuzzy Systems and Knowledge Discovery*, 2012, pp. 2840-2844.