Decision stump for prognosis of mortality rates in heart failure patients

Jan Bohacik*[1], *Department of Informatics at the University of Zilina, Slovakia

Abstract: Demographic developments and the rapid increase in chronic illnesses such as heart failure account for the enormous potential of telemedical solutions, and telemonitoring in particular. In addition to newly trained professionals for provision of telemonitoring service, automatic decision support systems are being pondered. In this paper, prognosis of death for a heart failure patient within a certain period of time as an important part of automatic decision support systems is investigated with a decision stump. Life-threatening situations and costs of treatment are taken into consideration primarily and achieved results are presented using sensitivity, specificity, positive predictive value, negative predictive value, accuracy, and the sum of sensitivity and specificity.

Key words: decision stump, prediction, mortality rates, heart failure

1. Introduction

Nowadays, it is common all around the world that health of patients with heart failure is pondered. Heart failure is a complex clinical syndrome with varying pathophysiological and clinical expression and requires frequent examinations. More effective treatment strategies directed at preventing clinical deterioration and consequent rehospitalization are then urgently needed, and great efforts have been done in the last years to identify and evaluate new home telemonitoring strategies capable of reducing unplanned hospital readmissions and improving quality of life (7). Home telemonitoring helps patients and families become active participants in their care and also provides a patient's clinical team with accurate and timely information to support clinical decisions and interventions.

An important part of home telemonitoring that could potentially improve the efficiency, cost-effectiveness and quality and safety of care is a decision support system (3). A decision support system is able to provide support and decisions and is mainly used to provide software-based healthcare-related advice to assist doctors and nurses in making decisions and

[1] Ján Boháčik with all Slovak diacritics.

developing solutions in complex or non-routine situations. The decision support system elaborated here is for prognosis of a patient's possibility to die within some time. An issue here is a lack of methods that could do this type of prognosis. Since telemonitoring can potentially provide enormous amounts of collectable data about patients with heart failure, data mining and machine learning appear to be very useful methods. There are several examples of data mining and machine learning applied in medical domains (1)(5)(6)(9). In this article, a decision stump is used for prognosis of mortality rates in heart failure patients.

The following organization is chosen for the article. The heart failure patient data which was used in the research is presented in Section 2. The application of decision stump for prognosis of mortality rates in heart failure patients is detailed in Section 3. An experimental study including the heart failure patient data and decision stump is discussed in Section 4. Section 5 concludes the article.

Table. 1: Heart failure patient data.

Attribute	**Type**	**Obtained values**
Blood Sodium Level (A_1)	Continuous	123.0000 – 148.0000 (mmol/L)
Age (A_2)	Continuous	27.0000 – 96.0000 (years)
Weight (A_3)	Continuous	29.8000 – 193.8000 (kg)
Height (A_4)	Continuous	1.2000 – 1.9600 (m)
Sex (A_5)	Discrete	*female* ($a_{5,1}$)
		male ($a_{5,2}$)
Blood Creatinine Level (A_6)	Continuous	37.0000 – 1262.0000 (umol/L)
Pulse Rate (A_7)	Continuous	38.0000 – 150.0000 (bpm)
Blood Uric Acid Level (A_8)	Continuous	0.1100 – 1.0600 (mmol/L)
NT-proBNP Level (A_9)	Continuous	0.8900 – 18236.0000 (pmol/L)
Prognosis (C)	Discrete	*alive* (c_1)
		dead (c_2)

2. Heart failure patient data

The data of Hull LifeLab, which is large, epidemiologically representative, information-rich clinical data (2), is used in the research. There are 2032 heart failure patients $\boldsymbol{p} \in \boldsymbol{P}$, $\boldsymbol{P}$ is the set of patients, which are described by nine attributes $\boldsymbol{A} = \{A_1; ...; A_k; ...; A_9\}$ and each heart failure patient $\boldsymbol{p} \in \boldsymbol{P}$ has one assigned prognosis $c_j \in C = \{c_1; c_2\}$. It is presented in Table. 1 in detail where the type of each attribute and its possible values are mentioned as well. If A_k is a discrete attribute, $A_k = \{a_{k,1}; ...; a_{k,l}; ...; a_{k,l_k}\}$ where $a_{k,1}$, ...; $a_{k,l}$, ..., a_{k,l_k} are possible discrete values. *Blood Sodium Level* (A_1) represents the amount of sodium in millimoles per litre of the patient's blood and its values are in mmol/L. *Age* (A_2) is the patient's age in years. *Weight* (A_3) is the patient's weight in kilograms. *Height* (A_4) shows the patient's height in meters. *Sex* (A_5) indicates if the patient is female or male. *Blood Creatinine Level* (A_6) is the amount of creatinine in micromoles per litre of the patient's blood. *Pulse Rate* (A_7) is the rate of the patient's pulse measured by tactile on the outside of an artery in beats per minute. *Blood Creatinine Level* (A_8) represents the amount of creatinine in micromoles per litre of the patient's blood. *NT-proBNP Level* (A_9) is the amount of the N-terminal prohormone of brain natriuretic peptide (NT-proBNP) in picograms per millilitre of the patient's blood. Value *alive* (c_1) means that the patient is prognosticated to be alive at least more than six months. Value *dead* (c_2) means that the patient is prognosticated to be dead within six months. Both *alive* and *dead* are defined for *Prognosis* (C).

3. Decision stump

Mortality rates for particular heart failure patients are predicted on the basis of a decision stump which is a machine learning model consisting of a one-level decision tree (4). The split at the root level is based on a specific attribute from all possible attributes in $\boldsymbol{A} = \{A_1; ...; A_k; ...; A_9\}$ (Figure 1). In this case, *Blood Uric Acid Level* (A_8) is used at the root level of the decision stump. The decision stump makes prognosis based on the value of the attribute, *Blood Uric Acid Level* in this particular case. If it is less than value 0.8500, then the prognosis is alive and if it is greater than or equal to value 0.8500, then the prognosis is dead. Thus, one attribute value is used when prognosis is being done for a heart failure patient and this may lead to reductions of costs for finding the values of other attributes provided that the decision stump for our problem has high accuracy.

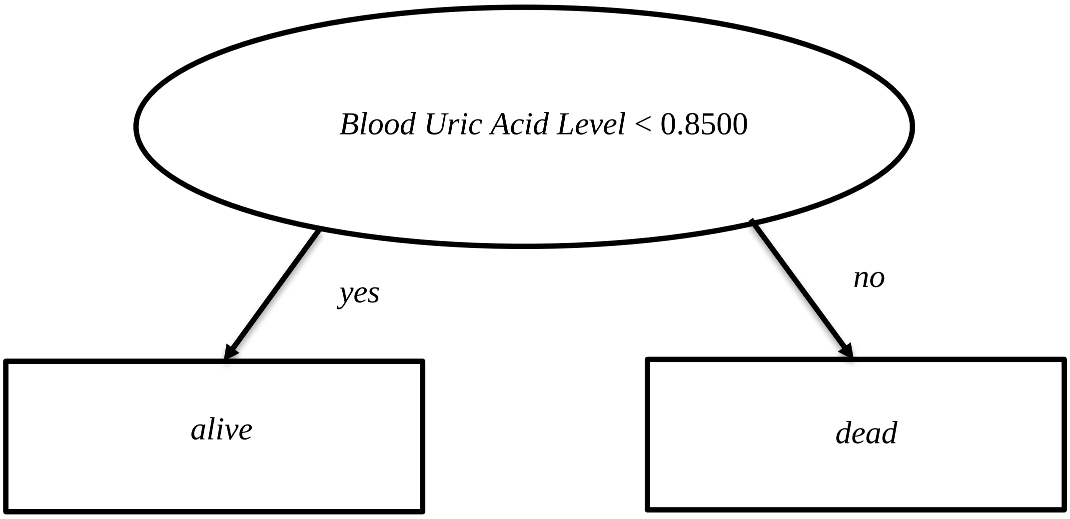

Figure. 1: An example of a decision stump.

4. Experiments and results

The decision stump was analysed on the heart failure patient data with a Java software tool developed by the author. The core algorithm of the decision stump was implemented in Weka (8) as class DecisionStump. The accuracy was measured with sensitivity tn / (tn + fp), specificity tp / (tp + fp), positive predictive value tp / (tp + fp), negative predictive value tn / (tn + fn), accuracy (tp + tn) / (tp + fp + fn + tn), and the sum of sensitivity and specificity. Tp/tn/fp/fn is the number of true positives/true negatives/false positives/false negatives where "*C* is *alive*" is considered negative and "*C* is *dead*" is considered positive. Tp, fp, fn and tn are computed in 10-fold cross-validation where the patient data is partitioned into 10 folds of patients. Of the 10 folds, a single fold is retained as the testing data for evaluation of accuracy, and the remaining 9 folds are used as the learning data. The learning data is used for creation of decision stump. The cross-validation process is repeated 10 times, with each of the 10 folds used exactly once as the testing data. It should not be prognosticated that soon-to-be death patients remain alive as this leads to life-threatening (measured by sensitivity) and it should not be prognosticated that patients that are still alive are going to die as this leads to increased running costs (measured by specificity). Both of these situations are combined as the sum of sensitivity and specificity.

Table. 2: Results of experiments for decision stump.

Accuracy Measure	Value
Sensitivity	0.5384615384615384
Specificity	0.8478835978835979
Positive Predictive Value	0.5490196078431373
Negative Predictive Value	0.8423127463863338
Accuracy	0.7687007874015748
Sum of sensitivity and specificity	1.3863451363451365

Accuracy results for decision stump are displayed in Table. 2. Sensitivity is 0.5384615384615384, specificity is 0.8478835978835979, positive predictive value is 0.5490196078431373, negative predictive value is 0.8423127463863338, and the sum of sensitivity and specificity is 1.3863451363451365.

5. Conclusions

A decision stump was applied for prognosis of heart failure for patients with occurred heart failure situations. There were data about 2032 patients used in the research where some of them were alive and others were dead six months after the data was obtained. The accuracy was evaluated in 10-fold cross-validation where measures sensitivity, specificity, positive predictive value, negative predictive value, accuracy, and the sum of sensitivity and specificity were computed. The sum of sensitivity and specificity measures minimization of life-threatening situations and minimization of costs. The higher the value of the sum is, the better the prognosis is. The obtained sensitivity was 0.5384615384615384, the obtained specificity was 0.8478835978835979, and the sum of sensitivity and specificity was 1.3863451363451365. The results are promising when the simplicity of decision stump is taken into consideration.

6. Bibliography

(1) Cazzolato, M. T., Ribeiro, M. X.: A statistical decision tree algorithm for medical data stream mining, Proceedings of the IEEE 26th International Symposium on Computer-Based Medical Systems, Pages: 389-392, Year: 2013.

(2) Clinical Effectiveness and Evaluation Unit of the Royal College of Physicians: Managing chronic heart failure: learning from best practice (Published by: The Lavenham Press Ltd; In: United Kingdom), Pages: 48, Year: 2005.

(3) de Vries, A. E.: The use of computer decision support systems and telemonitoring in heart failure. PhD thesis (Printed by: Grafimedia Facilitair Bedrijf RUG; In: Netherlands), Pages: 170, Year: 2013.

(4) Iba, W., Langley, P.: Induction of one-level decision trees, Proceedings of the 9th International Conference on Machine Learning, Pages: 233–240, Year: 1992.

(5) Mao, Y., Chen, Y.Hackmann, G., Chen, M., Lu, C., Kollef, M., Bailey, T.C.: Medical data mining for early deterioration warning in general hospital wards, Proceedings of the IEEE 11th International Conference on Data Mining Workshops, Pages: 1042-1049, Year: 2011.

(6) Trnka, A., Kovarova, M., Doci, I.: Occurrence of sexual disorders in Slovakia - a five-year analysis with Data Mining, Forum Staticum Slovacum (Volume: 7; Number: 2), Pages: 185-192, Year: 2011.

(7) Pinna, G. D., Maestri, R., Roma, M., Scanferlato, J.L., Giordano, A., Comazzi, F., Ferrari, M., Andrews, D., Johnson, P., Capomolla, S., Mortara, A.: Home telemonitoring of chronic heart failure patients: novel system architecture of the home or hospital in heart failure study, Computers in Cardiology (Volume: 30), Pages: 105-108, Year: 2003.

(8) Witten I. H., Frank E., Hall M. A.: Practical machine learning tools and techniques (Third Edition) (Published by: Morgan Kaufman; In: United States of America), Pages: 664, Year: 2011.

(9) Zhao, J., Wang, T.: A general framework for medical data mining, Proceedings of the 2010 International Conference on Future Information Technology and Management Engineering, Pages: 163-165, Year: 2010.

7. Address of the authors:

Dr Jan Bohacik
Department of Informatics
Faculty of Management Science and Informatics
University of Zilina
010 26
Zilina
SLOVAKIA
Jan.Bohacik@fri.uniza.sk

Acknowledgments

The heart failure patient data was provided by the Department of Cardiology, Hull York Medical School, University of Hull, UK as a part of Hull LifeLab during the author's stay at the University of Hull, UK.

The article was accepted for publication in November 2014 by the publisher of the Journal of Information Technologies (Vol. 7, No. 2, ISSN: 1337-7467), i.e. by the Faculty of Mass Media Communication, University of Ss. Cyril and Methodius in Trnava, Slovakia.