Knowledge Discovery Approaches for Early Detection of Decompensation Conditions in Heart Failure Patients

A Candelieri, D Conforti
Laboratory of Decision Engineering for Health Care Delivery, Department of Electronics, Informatics and Systems, University of Calabria, Italy
acandelieri@deis.unical.it
mimmo.conforti@unical.it

A Sciacqua, F Perticone
Cardiovascular Diseases Division, Department of Experimental and Clinical Medicine, University "Magna Graecia" of Catanzaro, Italy
sciacqua@unicz.it
perticone@unicz.it

Abstract

A crucial mid-long term goal for the clinical management of chronic heart failure (CHF) patients is to detect in advance new decompensation events, for improving quality of outcomes while reducing costs on the healthcare system. Within the relevant clinical protocols and guidelines, a general consensus has not been reached on how further decompensations could be predicted, even though many different evidence-based indications are known.

In this paper we present the Knowledge Discovery (KD) task which has been implemented and developed into the EU FP6 Project HEARTFAID (www.heartfaid.org), proposing an innovative knowledge based platform of services for effective and efficient clinical management of heart failure within elderly population. KD approaches have represented a practical and effective tool for analyzing data about 49 CHF patients who have been recurrently visited by cardiologist, measuring clinical parameters taken from clinical guidelines and evidence-based knowledge and that are also easy to be acquired at home setting.

Several KD algorithms have been applied on collected data, obtaining different binary classifiers performing a plausible early detection of new decompensations, showing high accuracy on internal validation and independent test.

1. Introduction

Chronic Heart Failure (CHF) is a complex clinical syndrome which impairs the ability of the heart to pump sufficient blood to cover the body's metabolic needs. This can severely affect people's ability in their normal daily activities, and it is a leading cause of hospital admissions and deaths, as well as a strong impact in terms of social and economic effects [1,2].

CHF prevalence increases rapidly with age [3], with a mean age of 74 years, and the increase in the proportion of elderly population concurs in rising prevalence of heart failure. In particular, according to the European Society of Cardiology, there are at least 10 millions of patients with heart failure in the European countries.

Many physicians regard HF primarily as a disease of males, because coronary risk factors are common in males and primarily they are most frequently enrolled in clinical trials of treatments for HF. However, the majority of patients with HF in the general population are females (particularly elderly), who frequently have HF associated with a normal left ventricular ejection fraction (LVEF) [1].

Even though heart failure is a chronic syndrome, it does not evolve gradually. Periods of relative stability alternate with acute destabilizations. During a stable phase the crucial mid-long term goal should be to predict and, hopefully prevent, destabilizations and death of the CHF patient, reducing the burden of heart failure on the healthcare system while improving the quality of life of affected patients.

Although a system of frequent monitoring could be useful for clinicians to recurrently evaluate patient conditions and eventually provide intervention before CHF patient becomes as severely ill as require re-

978-0-7695-3872-3/09 $26.00 © 2009 IEEE
DOI 10.1109/ISDA.2009.204

hospitalization [4], symptoms and signs of HF remain often difficult to identify. Moreover, only using sophisticated methods is possible to predict, define and assess new decompensation events, while the general interest is to find simple methods and criteria on how to predict when a patient will further decompensate.

For such reasons, Knowledge Discovery techniques may be a practical and effective solution for generating and proving new hypothesis, mining and generalizing new medical knowledge directly from pertinent real examples.

Under this respect, we developed and tested a KD task for the extraction of new decision models able to early detect new decompensation events. The KD task was performed within the EU FP6 Project HEARTFAID (www.heartfaid.org), aiming at developing an innovative knowledge based platform for more effective and efficient clinical management of heart failure within elderly population. Decision models extracted through KD task were embedded into the Clinical Decision Support System (CDSS) of HEARTFAID, integrating the platform's knowledge base. We used different Data Mining approaches, such as Decision Trees, Decision Lists, Support Vector Machines and Radial Basis Function Networks. As main result, in this paper we present Decision Tree and Support Vector Machine classifiers able to predict in a plausibly way a CHF patient destabilization.

2. Methods

2.1. Patients recruitment

A group of 49 patients with established CHF diagnosis has been recruited and frequently visited (every two weeks) at the Cardiovascular Diseases Division, Department of Experimental and Clinical Medicine, Faculty of Medicine, University "Magna Graecia" of Catanzaro, Italy.

During each visit, medical doctor has measured and stored values of a set of different clinical parameters, which have been selected from guidelines and clinical evidence-based knowledge: Systolic Blood Pressure (SBP), Heart Rate (HR), Respiratory Rate (RR), Body Weight (weight), Body Temperature (BT) and Total Body Water (TBW). Furthermore, the patient condition evaluated by the cardiologist during the visit, has been reported too.

Patients recruitment started on February 2007 and subjects have been recruited in different periods during the survey campaign. For each patient general information has been stored too, such as gender, age, NYHA class, alcohol use and smoking. Finally, it is worthwhile to remark that monitored parameters are easy to be acquired at home setting by suitable devices.

2.2. Dataset construction

We have defined the KD analysis as a "prediction task", in which every instance is represented by the general patient information and by the value of the aforementioned clinical parameters which have been measured during a visit. The class label for each instance has been defined on the base of the clinical condition assessed by the medical doctor "at the successive visit" (two weeks later). Starting from the available data we built a suitable dataset on which we have applied several classification learning techniques.

With respect to our formulation, only visits in which the patient was in a stable condition have been useful to carry out the designed prediction task. In this way we had to perform a preliminary case-selection so that the final dataset was related to 43 patients, for 301 valuable visits (instances). General description about the dataset is reported in Table 1.

Table 1 – Dataset description

	All	Male	Female
Patients	43	9	34
NYHA class (I, II, III, IV)	0, 28, 15, 0	0, 4, 5, 0	0, 24, 10, 0
Age	71.19±10.03	69.44±11.08	71.65±9.86
Smoking	7	0	7
Alcohol use			
No	18	8	10
Mild	12	1	11
Moderate	8	0	8
Severe	5	0	5
At least 1 new decompensation	9	4	5

Furthermore, such dataset is related to data acquired in the period from February 2007 to July 2008, and has been adopted as training set for the KD task. Further data, which has been acquired from July to October 2008 and consisting of 38 instances (2 cases of decompensations), has been adopted as independent test set. Such dataset contains also visits related to 4 patients who have not passed the preliminary case-selection for training set building.

2.3. KD methodologies and KD task design

In order to realize the KD task we selected and tuned up several classification approaches, either able to represent the extracted knowledge in a form easy to be understood by human or adopting a computational representation for the extracted relations between attribute values and class. Some approaches, such as Decision Trees [5-7] and Decision List [8] (a set of rules to be strictly interpreted in sequence) have been useful for providing simple and comprehensible criteria on how to predict if a patient in a stable condition will further decompensate in 2 weeks. For such cases, it has been possible to initially evaluate their performance in terms of accuracy (ability to correctly predict a new decompensation) and then, for the most performing ones, their plausibility (how much the criteria extracted from available data could be far away from the consolidated medical knowledge).

In a further step, we have adopted "computational" KD approaches, such as Support Vector Machines [9] and a particular Artificial Neural Network architecture known as Radial Basis Function Networks [10]. For such techniques, only performance evaluation has been possible because the extracted models are difficult to be interpreted by the medical doctors. For such reason, only evaluation on the built independent test set may provide an objective indication about plausibility of the function used to predict patient condition in 2 weeks.

We observe that our training set showed a highly skewed class distribution, a common situation in many medical real-world data mining problems [11]. The dataset has only 11 new decompensation events (positive class) and 290 cases of stability persistence (negative class), but in our work the class of primary interest is the minority one, so it has a much higher misclassification cost than the majority class. To this end, we have combined classical classification learning algorithms to a cost sensitive classification approach.

Selection of the most reliable predictive modes have been realized through a 2-phases validation. In the first phase, we have implemented a validation technique similar to leave-one-out, we have named "leave-patient-out". This approach works leaving apart (as independent validation set) all the instances into the training set which are related to the same patient (and not only one instance as in "pure" leave-one-out technique), while all the remaining instances are used as training set. This procedure is repeated for each patient and the classification performances are averaged on all the iterations. Such approach ensures the model that is learned from training instances does not know anything about the patient on whom it will be validated, making each validation phase independent by the learning process. At the end of the first phase we have been able to select an initial set consisting of the most promising predictive models.

During the second validation phase we have used the test set mentioned in the previous section. However, this test set is not really "independent" because data are mainly related to patients who are in training set, even though data refers to two different time periods.

3. Results

A first indication was obtained by a classical statistical approach: Respiratory Rate (RR) measured at current visit is potentially useful for preventing a new decompensation event. In fact, two weeks before a new assessed decompensation, value of RR is higher than during a visit relative to a stability persistence (Mann-Whiteny test: U=740 p-value= 0.002). This difference is showed in Figure 1 (stability persistence: RR=20.21±3.48 min=12 Max=32; successive decompensation: RR=24.27±4.34 min=16 Max=29).

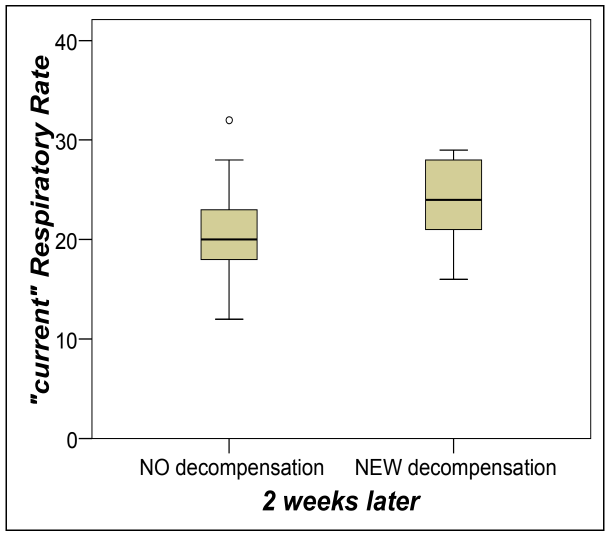

Figure 1 – Current Respiratory Rate and condition at 2 weeks later

The difference on RR is significant for females (Mann-Whitney test: U=128.5 p-value<0.001) but not for males (Mann-Whitney test: U=106.5 p-value=0.497). It is showed in figure 2 and in table 2.

No other significant differences were founded for the other clinical parameters.

Table 2 – Current RR and condition at 2 weeks later in males and females

	"current" Respiratory Rate	
	Male	Female
Successive Decompensation	21.60 ± 4.56 (16-28)	26.50 ± 2.81 (22-29)
Stability	20.31 ± 3.20 (16-28)	20.19 ± 3.55 (12-32)

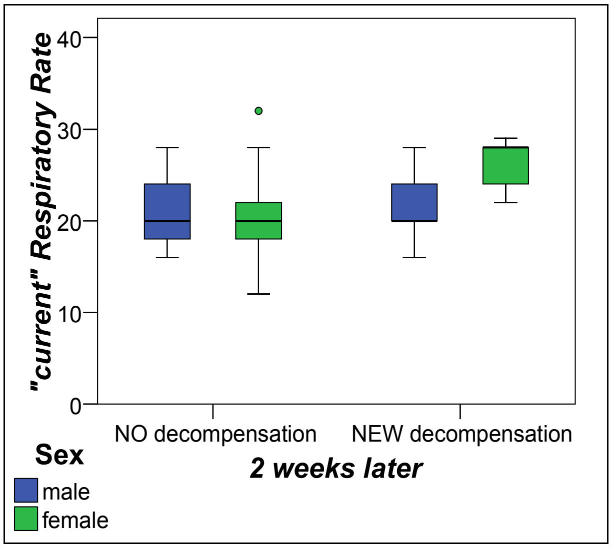

Figure 2 – Current RR and condition at 2 weeks in males and females

For KD analysis all the available information was initially taken into account: value of clinical parameters at current and previous visit, simple and relative variations of such values, and all general information about the patient. However, the best results are relative to predictive models able to correctly assess patient condition in 2 weeks, only using a subset of the clinical parameters measured at current visit and, at the most the, gender of patient.

The first really promising predictive model that we have obtained has been the Decision Tree shown in figure 3.

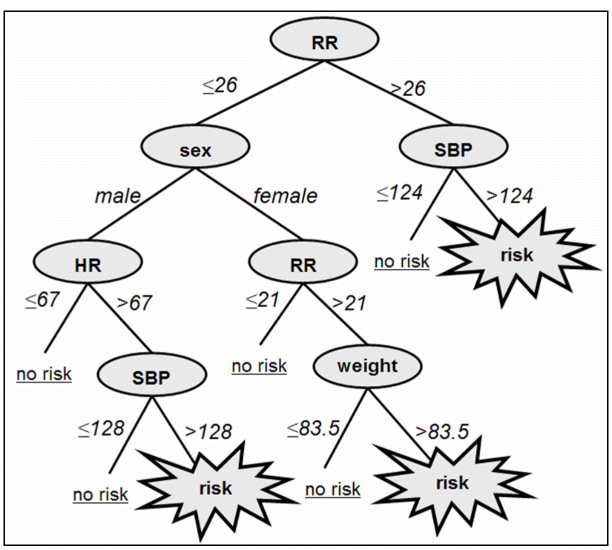

Figure 3 – Decision Tree form KD analysis

The tree presents a really simple structure and it is very easy to understand. The root node of the tree is a test on RR, identifying respiratory rate as the most relevant parameter for predicting a new decompensation (according to the results from statistical analysis). In particular, a CHF patient with RR greater than 26 acts-per-minute (apm) and SBP greater than 124 mmHg, may decompensate in 2 weeks. This rule is associated to 5 of the 11 positive cases in our dataset but it covers also 5 false positive examples.

Another rule that suggests a new decompensation in 2 weeks is related to male patients with RR less or equal than 26 apm, HR garter than 67 bpm and SBP greater than 128 mmHg. This rule is associated to 4 of the 11 positive cases into dataset but it covers also 10 false positive cases. Finally, female patients with RR between 21 and 26 apm, and weight greater than 83.5 Kg may decompensate in 2 weeks. This rule covers only 2 positive cases in our dataset and 14 false positive examples. Female patients with RR less than 21 apm present the condition with higher probability to maintain stability: 172 negative cases in our dataset are covered by this rule and no false negative ones.

As already specified, predictive reliability has been evaluated through a suitable validation technique (leave-patient-out), obtaining the best result for the learned tree (Accuracy: 92.03%, Sensitivity: 63.64%, False Positive Rate: 6.90%). Although the sensitivity is relative low, this is the best value obtained over all the adopted KD approaches.

Another really promising predictive model has been a Support Vector Machine with no kernel (linear-SVM) which uses the same 4 clinical parameters taken into account by the tree (SBP, HR, RR and weight) to provide a prediction about the condition in 2 weeks for a CHF stable patient, without any further information (while the tree needs to know gender). However, the SVM model has shown a greater false positive rate (16.90%) respect to the tree, during internal validation.

Nevertheless, evaluation on independent test set has demonstrated that SVM model is more reliable in predicting new decompensation events. In particular, both all 2 new cases of decompensations into test set has been correctly predicted by the SVM model, making only 1 false positive assessment (on 36 stability persistence instances). On the other hand, the tree has been unable to predict both 2 new decompensations and provided 1 false positive (which is not the one provided by the SVM model). In table 3 we summarize our results.

Table 3 – KD Results

	Leave One Patient Out	Independent Test set
Decision Tree		
Accuracy	92.03%	92,10%
Sensitivity	63.64%	0.00%
FP rate	6.90%	2.78%
SVM		
Accuracy	82.06%	97.37%
Sensitivity	63.64%	100.00%
FP rate	16.90%	2.78%

Moreover, we have to point out that only 1 of the 4 patients belonging to test set but not to training set, have presented a decompensation and that only SVM model has been able to correctly predicted such event. For the other 3 patients (5 instances into test set), stability has been correctly assessed both by SVM and Decision Tree.

10. Conclusions and discussion

We presented a KD task which has been implemented in HEARTFAID project with the goal to obtain predictive models able to early detect if a CHF patient in stable phase will further decompensate. The study was performed on a group of 49 CHF patients recurrently visited by cardiologists, every two weeks. We have presented a Decision Tree and a SVM model which have been evaluated in terms of predictive performance (accuracy and sensitivity) through a suitable internal validation technique and by applying them on an independent test set. Moreover, knowledge extracted from available data and codified into the tree has been evaluated by medical experts in terms of plausibility.

Clinical experts acknowledge that the extracted models may improve the CHF management. According to the tree learned from data and to results from statistical approach, they have confirmed that RR may be a predictive factor for a new destabilization. From a clinical point of view, the possible physic-pathological explanation could be that when the heart becomes unable to pump, an increase in left ventricular telediastolic pressure could occur. This shows an increase in pulmonary pressure and a consequent reduction of gaseous exchanges, which causes an increase of RR as compensation. These changes could occur before than an evident impaired cardiac function, which is characterized by a decrease of SBP and an increase of HR, and it is the cause of the acute and severe symptoms afterwards referred by the patient.

Finally, computational SVM model is difficult to be evaluated by medical doctors in terms of adherence to consolidated medical knowledge, however it has demonstrated to be more effective and reliable than decision tree, respect to an independent set of clinical data.

Such predictive models represent one of the several functionalities provided by the CDSS of the HEARTFAID platform: patient acquires clinical parameters in an automatic or semi-automatic way, they are analyzed by the predictive models and, if a decompensation is predicted into 2 weeks, the platform alerts medical doctors through SMS or e-mail, in order to hopefully prevent destabilizations or, at the least, re-hospitalization.

Acknowledgements

This research work was supported by the EU FP6 Project “HEARTFAID: a Knowledge based Platform of Services for supporting Medical-Clinical Management of the Heart Failure within the Elderly Population”, 2006-2009.

References

[1] Hunt SA, Abraham WT, Chin MH et al. ACC/AHA 2005 Guideline Update for the Diagnosis and Management of Chronic Heart Failure in the Adult: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines. Circulation. 2005 Sep 20; 112(12):e154-235

[2] Swedberg K, Cleland J, Dargie H et al. Guidelines for the diagnosis and treatment of chronic heart failure: executive summary (update 2005): The Task Force for the Diagnosis and Treatment of Chronic Heart Failure of the European Society of Cardiology. European Heart Journal. 2005 Jun; 26(11):1115-40.

[3] McKee PA, Castelli WP, McNamara PM et al. The natural history of congestive heart failure: the Framingham study. New England Journal of Medicine 1971; 285:1441-1446.

[4] Chaudhry SI, Barton B, Mattera J, Spertus J, Krumholz HM. Randomized trial of Telemonitoring to Improve Heart Failure Outcomes (Tele-HF): study design. Journal of Cardiac Failure. 2007 Nov; 13(9):709-14.

[5] Quinlan JR. Induction of decision trees. MachineLearning, 1986, 1(1):81-106.

[6] Quinlan JR. C4.5: Programs for Machine Learning. San Francisco: Morgan Kaufmann. 1993.

[7] Kohavi, R., Quinlan, J.R. Decision tree discovery. Handbook of Data Mining and Knowledge Discovery. 2002. pp.267-276,

[8] Cohen WW. Fast effective rule induction. Proceedings Twelfth Internationl Conference on Machine Learning, Tahoe Cit, CA. San Francisco: Morgan Kaufmann. 1995. 115-123.

[9] Cristianini, N., Schoelkopf, B. Support vector machines

and kernel methods. AI Magazine. 2002. 23(3): pp.31–41.
[10] S. Haykin. Neural Networks: A comprehensive Foundation. Upper Saddle River, NJ:Prentice Hall 1994.
[11] N. Japkowicz and S. Stephen, The class imbalance problem: a systematic study. Intelligent Data Analysis Journal, 6(5), 2002.