Predictive Models for Risk Assessment of Worsening Events in Chronic Heart Failure Patients

Maria Carmela Groccia[1], Danilo Lofaro[1], Rosita Guido[1], Domenico Conforti[1], Angela Sciacqua[2]

[1]de-Health Lab, DIMEG, Università della Calabria, Rende, Italy
[2]Dipartimento di Scienze Mediche e Chirurgiche, Università Magna Graecia, Catanzaro, Italy

Abstract

This work aims at developing and assessing a machine learning based Knowledge Discovery task for risk prediction of major cardiovascular worsening events in chronic heart failure patients.

Clinical data from 50 patients with chronic heart failure was analyzed. For each patient, personal data, different vital and clinical parameters and the presence of cardiovascular worsening events have been stored every three months per two years.

We defined the Knowledge Discovery analysis as a predictive task stated as supervised binary classification problem. The class label was defined based on the occurrence or not of cardiovascular worsening events between two consecutive visits. To take into account the temporality of the worsening events, six different temporal weighting strategies, applied to the vital parameters, were tested.

Several machine learning algorithms were applied for each strategy obtaining different predictive models. Models performance have been evaluated mainly in term of area under the ROC curve (AUC), and Linear Support Vector Machine got the best performing predictive model.

The implemented Knowledge Discovery task have shown to be a reliable tool for support cardiologists for risk predictions of major cardiovascular worsening events.

1. Introduction

Chronic Heart Failure (CHF) is a complex syndrome in which the heart weakens and its activity becomes insufficient for the body. Patients with CHF are predisposed to major cardiovascular worsening events that may compromise the residual function of the heart and cause hospital admissions and deaths [1, 2].

Under this respect, it may be helpful for clinicians to assess the patient's condition and provide a more appropriate clinical management before worsening.

Knowledge Discovery (KD) techniques can be appropriately taken into account due to their effective impact in cardiovascular domain for prediction tasks [3, 4].

When using clinical events, it is important to take into account temporality of events, so that the Machine Learning (ML) algorithms can exploit this information. In previous studies, different strategies have been proposed to take in account the temporality of clinical events, showing that temporality strategies lead to better predictive performance [5-7].

In this work, we defined a KD analysis as a predictive task to extract new decision models that can help clinicians to early detect major cardiovascular worsening events in patients with CHF. We analysed clinical data from 50 patients with CHF collected at the Cardiovascular Diseases Division, Department of Experimental and Clinical Medicine, University "Magna Graecia" of Catanzaro, Italy.

Different temporal weighting strategies, applied to the vital parameters, were analysed to take into account the temporality of the worsening events. To evaluate the impact of each strategy on predictive performance, several ML algorithms have been applied obtaining different binary classifiers. Models performance have been evaluated mainly in term of area under the ROC curve (AUC) due to the unbalanced class distribution of the collected data. Sensitivity and specificity are also been taken in account.

2. Methods

A group of 50 patients with an established diagnosis of CHF were followed for two years. At the first visit, personal data and clinical parameters such as gender, age, NYHA class, aetiology, cardiovascular history and other diseases were collected. General characteristics of the collected data are show in Table *1*. In each subsequent visit, different vital parameters were measured and recorded: Weight, Heart Rate, Body Temperature, Systolic Blood Pressure, Diastolic Blood Pressure, and Respiratory Rate. Furthermore, the date of the visit, the date and type

of major cardiovascular worsening events were reported.

Table 1. General characteristics of the collected data.

Feature	
Age (years)	72.55±14.19
NYHA Class	
I	3 (6%)
II	38 (76%)
III	9 (18%)
Aetiology	
Ischemic	23(46%)
Idiopathic dilatation	9 (18%)
Hypertension	8 (16%)
Valvular	12 (24%)
Alcoholic	2 (4%)
Cardiovascular history	
Instable angina	1 (2%)
PTCA	1 (2%)
BY-PASS	7 (14%)
Atrial flutter	13 (26%)
Pace Maker	3 (6%)
Cardiac resynchronization	1 (2%)
ICD	3 (6%)
Mitral insufficiency	22 (44%)
Aortic insufficiency	5 (10%)
Hypertension	29 (58%)
TIA	3 (6%)
Other diseases	
Diabetes	15 (30%)
Hypothyroidism	4 (8%)
Renal failure	4 (8%)
BPCO	5 (10%)
Asthma	1 (2%)
Sleep apnea	4 (8%)
Pulmonary fibrosis	1 (2%)
Gastrointestinal diseases	4 (8%)
Hepatic diseases	3 (6%)
Major cardiovascular worsening events	13 (26%)

We defined the KD analysis as a predictive task stated as supervised binary classification problem. From the available data, we have constructed a data set suitable for the prediction task; every instance is represented by the general patients' information and the value of the vital parameters measured during a visit. The class label was defined based on the occurrence or not of cardiovascular worsening events between two consecutive visits.

Based on the research in [6], we explored various temporal weighting strategies and analysed their impact on the performance of predictive models.

Each temporal weighting strategy are based on the following form: for the vital parameters collected n days before the current visit, a weight w is assigned according to the value of a specific function $f(n)$.

In detail, six temporal weighting functions were tested:

- $w_1 = e^{(-n)}$
- $w_2 = \frac{1}{n+1}$
- $w_3 = e^{(-0.1 \times n)}$
- $w_4 = \frac{(n-1211)^2}{1211^2}$
- $w_5 = 1 - \frac{n}{1211}$
- $w_6 = 1 - \frac{n^2}{1211^2}$

where n is the number of days between two consecutive visits and 1211 is the total number of days in which the patients were followed. One day was added to make sure that parameters measured on the 1210th day receive low but non-zero weights.

The vital parameters recorded in the current visit have a maximum weight equal to one. The weight decreases monotonically with increasing number of days between visits. All assigned weights are between 0 and 1. The temporal weighting strategies differ in the speed by which weights decrease along the patient history.

Therefore, the values of vital parameters used by ML algorithms at each visit are weighted with the parameters measured in all previous visits.

In order to realize the KD task, the original dataset was divided into training (70%) and test (30%) set. The dataset turns out to be an unbalanced distribution of the class of interest, 10 patients that present major cardiovascular worsening events and 40 patients without the presence of events. Since in our case the class of interest is that of the minority, an oversampling approach [8] has been adopted to balance the classes in the training set.

Several ML algorithms were applied to the training set obtaining different predictive models: Linear/Radial Support Vector Machine (SVM), Decision Trees, Naïve Bayes and Neural Networks [9]. Parameters tuning has been optimized basing of performance on a 5-fold cross-validation repeated five times.

The performance of the models was evaluated on the test set in term of AUC (classifier's ability to avoid false classification), sensitivity (the proportion of instances with major cardiovascular worsening event that were correctly predicted) and specificity (the proportion of instances without major cardiovascular worsening events that were correctly predicted) [10].

At first, in order to evaluate the impact of each temporal weighting strategy on predictive performances of ML algorithms, we considered as predictive features only the features varying over time, that is the vital parameters. Then, we selected the best temporal weighting strategy and reapplied the ML algorithms considering all features.

3.　　Results

Table *2* shows the performances obtained by each ML algorithm with the different temporal weighting strategies on the test set. The results refer to the performance obtained by the ML algorithms that use only the vital parameters as predictive features.

Table 2. AUC, sensitivity (Sens) and specificity (Spec) of the models evaluated measured on the test set with the six weighted strategies and vital parameters as predictive features.

	Temporal Weighting Strategy					
	w_1	w_2	w_3	w_4	w_5	w_6
SVM Lin.						
AUC	0.90	0.91	0.90	0.87	0.88	0.88
Sens	0.72	0.71	0.72	0.72	0.72	0.81
Spec	1.00	1.00	1.00	1.00	0.75	0.75
SVM Rad.						
AUC	0.75	0.75	0.75	0.69	0.71	0.70
Sens	0.98	0.98	0.98	0.97	0.97	0.97
Spec	0.00	0.00	0.00	0.00	0.00	0.00
Decision Tree						
AUC	0.74	0.84	0.71	0.48	0.46	0.47
Sens	0.98	0.92	0.93	0.95	0.92	0.94
Spec	0.50	0.75	0.50	0.00	0.00	0.00
Neural Network						
AUC	0.70	0.47	0.84	0.65	0.43	0.71
Sens	0.99	0.92	0.91	0.90	0.94	0.92
Spec	0.00	0.00	0.25	0.25	0.25	0.00
Naïve Bayes						
AUC	0.46	0.47	0.46	0.62	0.52	0.51
Sens	0.86	0.86	0.86	0.78	0.77	0.77
Spec	0.25	0.25	0.25	0.50	0.50	0.25

Evaluation on test set has demonstrated that the temporal weighting strategy w_2 allows achieving the best value of AUC and that, in general, the Linear SVM has the highest performance. Specifically, Linear SVM achieves the best performance with the temporal weighting strategy w_2: AUC=0.91, Sens=0.71 and Spec=1.00. This indicate that the Linear SVM with the temporal weighting strategy w_2, has an high ability to avoid false classification, correctly predicts the 71% of instances with major cardiovascular worsening event, and correctly predicts all instances without major cardiovascular worsening event. With weighted strategy w_2, the decision tree also showed good performance. On the contrary, the other tested models achieved very low or zero specificity values with the same strategy and this indicates that they are not able to classify major cardiovascular worsening events.

On these bases, in the second experiment we considered the vital parameters weighed with the best strategy w_2 and all general information about the patient (age, etiology, other diseases and information regarding cardiovascular history) as predictive features. Then, all ML algorithms were reapplied. In Table *3* the performance of the tested ML algorithms on test set are presented.

Table 3. AUC, sensitivity (Sens) and specificity (Spec) of the tested ML algorithms with all parameters and patients' general information as predictive features.

Model	AUC	Sens	Spec
SVM Lin.	0.73	0.83	0.50
SVM Rad.	0.78	0.99	0.00
Decision Tree	0.61	0.96	0.25
Neural Network	0.82	0.94	0.25
Naïve Bayes	0.57	0.99	0.00

In general, with more information the performance of Linear SVM and Decision tree decreases. Nevertheless, Linear SVM is the more reliable in predicting major cardiovascular worsening events also in this experiment. The other tested methods even if have an acceptable AUC (> 70%), have all lower specificity.

4.　　Conclusions

In this work, we defined a KD analysis as a predictive task to extract new decision models that can help clinicians to early detect major cardiovascular worsening events in patients with CHF.

Different temporal weighting strategies were tested in order to take into account temporality of the worsening events. The impact of each strategy on the performance of different predictive models was analysed.

Linear SVM with temporal weighting strategy w_2 showed a very good prediction performance on the test set. Prediction performance with only vital parameters (AUC=0.91, Sens=0.71 and Spec=1.00) demonstrate the effectiveness of the linear SVM as predictive model for major cardiovascular worsening events in patients with CHF.

As future work, we plan to embed the selected predictive model in a clinical decision support system in order to test it in the cardiologists' daily practise.

References

[1] P. Ponikowski, A. A. Voors, S. D. Anker, H. Bueno, J. G. F. Cleland, A. J. S. Coats, V. Falk, J. R. González-Juanatey, V. P. Harjola, E. A. Jankowska, M. Jessup, C. Linde, P. Nihoyannopoulos, J. T. Parissis, B. Pieske, J. P. Riley, G. M. C. Rosano, L. M. Ruilope, F. Ruschitzka, F. H. Rutten, and P. van der Meer, "2016 ESC Guidelines for the Diagnosis and Treatment of Acute and Chronic Heart Failure," *Rev Esp Cardiol (Engl Ed),* vol. 69, no. 12, pp. 1167, Dec, 2016.

[2] R. P. Morrissey, L. Czer, and P. K. Shah, "Chronic heart failure: current evidence, challenges to therapy, and future directions," *Am J Cardiovasc Drugs,* vol. 11, no. 3, pp. 153-71, Jun, 2011.

[3] B. Jin, C. Che, Z. Liu, S. Zhang, X. Yin, and X. Wei, "Predicting the Risk of Heart Failure With EHR Sequential Data Modeling," *Ieee Access,* vol. 6, pp. 9256-9261, 2018, 2018.

[4] D. Lofaro, M. C. Groccia, R. Guido, D. Conforti, S. Caroleo, G. Fragomeni, and A. Murray, "Machine Learning Approaches for Supporting Patient-Specific Cardiac Rehabilitation Programs," *2016 Computing in Cardiology Conference (Cinc), Vol 43,* vol. 43, pp. 149-152, 2016, 2016.

[5] J. Zhao, A. Henriksson, M. Kvist, L. Asker, and H. Boström, "Handling Temporality of Clinical Events for Drug Safety Surveillance," *AMIA Annu Symp Proc,* vol. 2015, pp. 1371-80, 2015.

[6] J. Zhao, J. Huan, S. Miyano, A. Shehu, H. X, B. Ma, S. Rajasekaran, V. Gombar, I. Schapranow, I. Yoo, J. Zhou, B. Chen, P. V, and B. Pierce, "Temporal Weighting of Clinical Events In Electronic Health Records for Pharmacovigilance," *Proceedings 2015 Ieee International Conference on Bioinformatics and Biomedicine*, pp. 375-381, 2015, 2015.

[7] J. Zhao, and A. Henriksson, "Learning temporal weights of clinical events using variable importance," *Bmc Medical Informatics and Decision Making,* vol. 16, JUL 21 2016, 2016.

[8] I. Yoo, P. Alafaireet, M. Marinov, K. Pena-Hernandez, R. Gopidi, J. Chang, and L. Hua, "Data Mining in Healthcare and Biomedicine: A Survey of the Literature," *Journal of Medical Systems,* vol. 36, no. 4, pp. 2431-2448, AUG 2012, 2012.

[9] S. Kotsiantis, D. Kanellopoulos, and P. Pintelas, "Handling imbalanced datasets: A review," in *GESTS International Transactions on Computer Science and Engineering*, 2006, pp. 25-36.

[10] M. Sokolova, and G. Lapalme, "A systematic analysis of performance measures for classification tasks," *Information Processing & Management,* vol. 45, no. 4, pp. 427-437, JUL 2009, 2009.

Address for correspondence.

Prof. Domenico Conforti
de-Health Lab, DIMEG, Università della Calabria. Via P. Bucci 87100, Rende (CS), Italy.
domenico.conforti@unical.it