

# Machine Learning Approaches for Supporting Patient-Specific Cardiac Rehabilitation Programs

Danilo Lofaro<sup>1</sup>, Maria Carmela Groccia<sup>1</sup>, Rosita Guido<sup>1</sup>, Domenico Conforti<sup>1</sup>, Sergio Caroleo<sup>2</sup>, Gionata Fragomeni<sup>3</sup>

<sup>1</sup>de-Health Lab, DIMEG, University of Calabria, Rende, Italy

<sup>2</sup>“Villa del Sole” Clinic, Catanzaro, Italy

<sup>3</sup>Dept. Surgical and Clinical Science, Magna Graecia University, Catanzaro, Italy

## Abstract

*Cardiac rehabilitation is a well-recognised non-pharmacological intervention that prevents the recurrence of cardiovascular events. Previous studies investigated the application of data mining techniques for the prediction of the rehabilitation outcome in terms of physical, but fewer reports are focused on using predictive models to support clinicians in the choice of a patient-specific rehabilitative treatment path. Aim of the work was to derive a prediction model for help clinicians in the prescription of the rehabilitation program.*

*We enrolled 129 patients admitted for cardiac rehabilitation after a major cardiovascular event. Data on anthropometric measures, surgical procedure and complications, comorbidities and physical performance scales were collected at admission. The prediction outcome was the rehabilitation program divided in four different paths. Different algorithms were tested to find the best predictive model. Models performance were measured by prediction accuracy.*

*Mean model accuracy was 0.790 (SD 0.118). Best model selected was Lasso regression showing an average classification accuracy on test set of 0.935.*

*Data mining techniques have shown to be a reliable tool for support clinicians in the decision of cardiac rehabilitation treatment path.*

## 1. Introduction

Cardiovascular disease is the leading cause of death in the majority of developing and developed countries [1, 2].

Although life expectancy has been extended and great progresses have been made in detection and treatment of heart diseases, heart failure continues to produce a heavy burden of cardiovascular morbidity and mortality increasing with age [3, 4].

Cardiac physical rehabilitation is a well-recognised non-pharmacological intervention recommended for the

follow up management of cardiovascular diseases [5]. Physical and exercise training programs have been shown to prevent the recurrence of cardiovascular events and increases life expectancy, improve function, exercise capacity, quality of life, psychosocial well-being, morbidity, and mortality [6].

Cardiac rehabilitation conventionally consists of phase I (inpatient cardiac rehabilitation), phase II (outpatient cardiac rehabilitation), and phase III (maintenance). It starts with an inpatient hospital-based program. Due to the short hospital stays and time-consuming examinations, phase I programs are mostly limited to early mobilization and education to life-style changes and do not include an exercise training component. Outpatient hospital-based programs last from two to four months. The content of phase II cardiac rehabilitation varies greatly from hospital to hospital and it usually includes single or group exercises, education and counselling [7, 8].

Numerous previous studies investigated the application of data mining techniques for the prediction of cardiac rehabilitation outcome in terms of physical performance as well as of length of stay after acute cardiovascular events [9-11], but fewer reports are focused on using predictive models to support clinicians in the choice of a patient-specific rehabilitative treatment path.

In this work, we analyzed data relative to two years activity of a rehabilitation clinic in South Italy to derive a prediction model to support clinicians in the prescription of the phase II cardiac rehabilitation program.

## 2. Methods

We collected data relative to 129 patients admitted for cardiac rehabilitation after a major cardiovascular event (Coronary Artery Disease or Myocardial Infarction). Data collected at admission were age, sex, body mass index, secondary death risk calculated by the GISSI-Prevenzione mortality risk chart [12], cardiovascular risk class [13], number of coronary arteries treated, perioperative

complications, infections, neurological comorbidities, diabetes, pleural effusion, pericardial effusion, Barthel index, mMRC breathlessness scale, FEV1/FVC ratio, Borg Rating of Perceived Exertion scale and 6 minutes walking test.

The prediction outcome was the rehabilitation program divided in four different programs. These four programs are assigned to patients by physicians basing on their perceived health status and abilities to perform physical exercises. The rehabilitation programs are based on a mix of respiratory, ambulation, passive (executed by physical therapists) and active (divided in calisthenic and ergometric) exercises. Differences among rehabilitation programs are based on exercise frequency, progression of exercises intensity in terms of Metabolic Equivalents in Task (METS) [14] and range of METS achieved in ergometric exercises.

Looking at original dataset characteristics, having a relative high number of predictor (17) and low number of cases, to derive the predictive model of rehabilitation program, we select a set of classification algorithms having internal procedure for feature selection and problem dimension reduction. In particular, algorithms used to find the best predictive model were L1 norm penalized (Lasso) regression, Support Vector Machine (SVM) with linear kernel, SVM with polynomial kernel, SVM with radial kernel, Random Forest (RF), Bagged Flexible Discriminant Analysis (FDA) [15], C 5.0 Classification Tree with Boosting and Bagged CART.

Study sample was divided in training (nr. 99) and test set (nr. 31) basing on the distribution of the outcome classes. Models parameters tuning has been optimized basing of performance on a 5-fold cross-validation repeated five times on the training set. Models performance were evaluated by calculating a series of measures for multi-class classification metrics based on a generalization of measures derived from binary problems [16]. Given classes  $C_i, i = 1 \dots l$ ,  $tp_i$  are the number of true positive for  $C_i$ , and  $fp_i$  – false positive,  $fn_i$  – false negative, and  $tn_i$  – true negative counts respectively. We defined:

$$\text{Average Accuracy} = \frac{\sum_{i=1}^l \frac{tp_i + tn_i}{tp_i + fn_i + fp_i + tn_i}}{l}$$

$$\text{Error Rate} = \frac{\sum_{i=1}^l \frac{fp_i + fn_i}{tp_i + fn_i + fp_i + tn_i}}{l}$$

$$\text{Precision} = \frac{\sum_{i=1}^l \frac{tp_i}{tp_i + fp_i}}{l}$$

$$\text{Recall} = \frac{\sum_{i=1}^l \frac{tp_i}{tp_i + fn_i}}{l}$$

R 3.2 and caret package 6.0 were used for analyses.

### 3. Results

In Table 1 general characteristics of enrolled patients at admission at the cardiac rehabilitation clinic are shown. The majority of the patients were male. We observed a high proportion of subjects with High cardiovascular risk at admission (as expected), pleural and pericardial effusion, diabetes and perioperative complications.

Table 1. Characteristics of the study sample. Data are presented as mean  $\pm$  sd for normal distributed numerical variables, median (IQR) for nonnormal variables and number (%) for categorical variables.

<b>Variables</b>	
Age (years)	67.98 $\pm$ 9.15
Sex Male	105 (81.4%)
Body Mass Index (Kg/m <sup>2</sup> )	26.32 $\pm$ 3.56
GISSI-Prevenzione Mortality Risk	38.62 $\pm$ 24.92
Cardiovascular Risk Class	
High	94 (74%)
Medium	4 (3.1%)
Low	29 (22.8%)
Nr. of Treated Coronaries	2 (1.00 - 3.00)
Perioperative Complications	67 (51.9%)
Infections	17 (13.2%)
Neurological Comorbidities	11 (8.5%)
Diabetes	53 (41.1%)
Pleural Effusion	93 (72.1%)
Pericardial Effusion	53 (41.1%)
Barthel Index	77 (38.00 - 85.00)
mMRC Breathlessness Scale	1 (0.00 - 2.00)
FEV1/FVC Ratio	61.61 $\pm$ 11.56
Borg RPE Scale	13.22 $\pm$ 1.77
6 Minutes Walking Test (m)	308.92 $\pm$ 101.33
Rehabilitation Program	
A	5 (4%)
B	28 (22.2%)
C	53 (42.1%)
D	40 (31.7%)

In Table 2 the performance of the classification algorithms tested are presented. Mean model accuracy was 0.790 (SD 0.118). The model with the best performance was Lasso Regression with penalization parameter  $\lambda = 0.057$ . The model showed an average accuracy = 0.935 (error rate = 0.063), precision = 0.941 and recall = 0.9. Bagged flexible discriminant analysis showed similar classification abilities as well as C 5.0 and SVM with linear kernel. On the other hand, other tested models showed average accuracy under the 70%, with very low precision and recall values.

Table 2. Accuracy, Error Rate, Precision and Recall of the models evaluated measured on the test set.

Model	Accuracy	Error Rate	Precision	Recall
Lasso	0.935	0.063	0.941	0.901
SVM Lin.	0.887	0.110	0.85	0.837
SVM Poly	0.677	0.316	0.232	0.246
SVM Rad.	0.677	0.316	0.236	0.246
RF	0.694	0.300	0.234	0.238
Bag. FDA	0.919	0.079	0.902	0.881
Boost. C 5.0	0.871	0.126	0.828	0.812
Bag. CART	0.677	0.316	0.204	0.221

#### 4. Discussion and conclusion

In this work, we developed a prediction model to support physicians in the task of assign a cardiac rehabilitation program for patients after a major cardiovascular event. The selected model showed a very good prediction performance on the test set, and it can be considered sufficiently accurate to be tested in a clinical setting.

We tested several algorithms founding different performances even among models within the same family. In particular, L1 norm penalized Lasso regression achieved the best performance on the test set, selection of a parsimonious model results in better prediction in this particular setting. This is confirmed by the performance obtained with the SVM with linear kernel, FDA. Poor performance achieved by SVMs with polynomial and radial kernel could be explained by the distribution of the original data. On the other hand, the differences between C 5.0 and random forest or CART are not so clear. It is possible that the method for the selection of the final model could have influenced the performance as well as the imbalance in the prediction classes.

To date, the assignment of a cardiac rehabilitation program to patients were an exclusive decision of the physicians, principally based on the health status of the patient and his perceived ability to tolerate a certain intensity in physical exercise. There are not international guidelines defining a set of standard procedure for cardiac rehabilitation. It is possible to find many evidences about the efficacy of several approaches, but programs of cardiac rehabilitation are different from hospital to hospital depending on country/regional health policy, budget, available facilities, etc. Starting from previous cases to develop a deductive model to help in the assignment rehabilitation program could help physicians in standardize their prescription, possibly help in the training phase of new medical personnel and in the

engagement of non-medical staff in the decisional process.

Our model is intended to be part of a larger informative system, designed to support outpatient clinic staff in the management of the entire rehabilitation path. Next step of our work will be to test the selected model on new patients, and try to develop a new model basing on personal patient outcome. In particular, this new model will use “real time” data from day-by-day exercise outcome and directly from machines to select the best rehabilitation program for the specific patients.

In conclusion, our work demonstrates that data mining approach can be suitable to support clinicians in the prescription of cardiac rehabilitation program and could represent a solid base for the development of decisional support system in the cardiac rehabilitation field.

#### References

- [1] WHO. Cardiovascular diseases (CVDs), Fact sheet No 317: 2016 [updated June 2016]. Available from: <http://www.who.int/mediacentre/factsheets/fs317/en/>.
- [2] Global, regional, and national age-sex specific all-cause and cause-specific mortality for 240 causes of death, 1990-2013: a systematic analysis for the Global Burden of Disease Study 2013. *Lancet*, 2015;385:117-71.
- [3] Berthold HK, Gouni-Berthold I. Lipid-lowering drug therapy in elderly patients. *Curr Pharm Des*, 2011;17:877-93.
- [4] Lavie CJ, Milani RV, Ventura HO. Obesity and cardiovascular disease: risk factor, paradox, and impact of weight loss. *J Am Coll Cardiol*, 2009;53:1925-32.
- [5] Leon AS, Franklin BA, Costa F, Balady GJ, Berra KA, Stewart KJ, et al. Cardiac rehabilitation and secondary prevention of coronary heart disease: an American Heart Association scientific statement from the Council on Clinical Cardiology (Subcommittee on Exercise, Cardiac Rehabilitation, and Prevention) and the Council on Nutrition, Physical Activity, and Metabolism (Subcommittee on Physical Activity), in collaboration with the American association of Cardiovascular and Pulmonary Rehabilitation. *Circulation*, 2005;111:369-76.
- [6] Taylor RS, Brown A, Ebrahim S, Jolliffe J, Noorani H, Rees K, et al. Exercise-based rehabilitation for patients with coronary heart disease: systematic review and meta-analysis of randomized controlled trials. *Am J Med*, 2004;116:682-92.
- [7] Vanhees L, McGee HM, Dugmore LD, Schepers D, van Daele P. A representative study of cardiac rehabilitation activities in European Union Member States: the Carinex survey. *J Cardiopulm Rehabil*, 2002;22:264-72.
- [8] Balady GJ, Williams MA, Ades PA, Bittner V, Comoss P, Foody JM, et al. Core components of cardiac rehabilitation/secondary prevention programs: 2007 update: a scientific statement from the American Heart Association Exercise, Cardiac Rehabilitation, and Prevention Committee, the Council on Clinical Cardiology; the Councils on Cardiovascular Nursing, Epidemiology and Prevention, and Nutrition, Physical

- Activity, and Metabolism; and the American Association of Cardiovascular and Pulmonary Rehabilitation. *Circulation*, 2007;115:2675-82.
- [9] Hachesu PR, Ahmadi M, Alizadeh S, Sadoughi F. Use of Data Mining Techniques to Determine and Predict Length of Stay of Cardiac Patients. *Healthc Inform Res*, 2013;19:121-9.
- [10] Suh Mk, Rofouei M, Nahapetian A, Kaiser WJ, Sarrafzadeh M, editors. *Optimizing Interval Training Protocols Using Data Mining Decision Trees*. 2009 Sixth International Workshop on Wearable and Implantable Body Sensor Networks; 2009 3-5 June 2009.
- [11] Van A, Gay, V. C., Kennedy, P. J., Barin, E., Leijdekkers P, editors. *Understanding risk factors in cardiac rehabilitation patients with random forests and decision trees*. Australasian Data Mining Conference (AusDM 11); 2011; Ballarat, Australia: ACS.
- [12] Marchioli R, Avanzini F, Barzi F, Chieffo C, Di Castelnuovo A, Franzosi MG, et al. Assessment of absolute risk of death after myocardial infarction by use of multiple-risk-factor assessment equations: GISSI-Prevenzione mortality risk chart. *Eur Heart J*, 2001;22:2085-103.
- [13] Fletcher GF, Balady GJ, Amsterdam EA, Chaitman B, Eckel R, Fleg J, et al. Exercise standards for testing and training: a statement for healthcare professionals from the American Heart Association. *Circulation*, 2001;104:1694-740.
- [14] Jette M, Sidney K, Blumchen G. Metabolic equivalents (METS) in exercise testing, exercise prescription, and evaluation of functional capacity. *Clin Cardiol*, 1990;13:555-65.
- [15] Hastie T, Tibshirani R, Buja A. Flexible Discriminant Analysis by Optimal Scoring. *Journal of the American Statistical Association*, 1994;89:1255-70.
- [16] Sokolova M, Lapalme G. A systematic analysis of performance measures for classification tasks. *Inform Process Manag*, 2009;45:427-37.

Address for correspondence.

Prof. Domenico Conforti  
DIMEG, University of Calabria. Via P. Bucci 87100, Rende  
(CS), Italy.  
domenico.conforti@unical.it