Evaluating the Risk of a Rescue Percutaneous Coronary Intervention after Thrombolysis Therapy: A Decision Tree Approach

V Lagani[1], R Ceravolo[2], M Vatrano[2], VA Ciconte[2], D Conforti[3]

[1]Supercomputing Center for the Computational Engineering (CESIC), NEC Italia, Cosenza, Italy
[2]Coronary Intervention Unit, Pugliese Ciaccio Hospital, Catanzaro, Italy
[3]Department of Electronics, Informatics and Systems, University of Calabria, Cosenza, Italy

Abstract

Thrombolysis intervention is a common therapeutic practice used in order to dissolve the coronary atherosclerotic plate whenever Percoutaneous Coronary Intervention (PCI) can not be performed. In case the thrombolytic drug fails, a Rescue PCI is needed in order to restore the normal coronary blood flow. Thus, assessing the individual risk of Thrombolysis failure is a crucial step before deciding to perform a thrombolytic intervention.

Aim of the present study is to develop a simple set of rules able to support physicians in discriminating patients eligible for a Thrombolysis intervention.

The proposed models pointed out some interesting interactions between blood pressure values and clinical parameters related to patient's metabolism, suggesting new interesting mutual influences among Thrombolysis failure predictors.

1. Introduction

This paper addresses a specific medical decision making problem: the evaluation of risk of performing a rescue percutaneous coronary intervention after the subministration of a thrombolytic drug, for patient undergone an Acute Myocardial Infarction (AMI).

In almost all cases AMI are caused by the rupture of atherosclerotic plaques within coronary arterials: wrong alimentary uses, genetic predisposition and other risk factors (e.g. smoke, dyslipidaemia, hypertension) can lead to the inflammation of the vessel internal tissues, with the consequent shrinkage of the arterial inner diameter (stenosis) and the rupture of the plaque. Percutaneous Coronary Intervention (PCI) is the state of the art medical procedure for the effective treatment of the coronary atherosclerotic plaques: during PCI, different devices are introduced by peripheral arterials (e.g. femoral) and pushed until coronaries mouth; thus, utilizing x-rays techniques and appropriate micro devices, the place of stenosis can be localized and reported to its normal diameter.

Despite its effectiveness, PCI can be performed only in specialized centres with qualified medical personnel and appropriate equipment. In small health care structures, without the possibility of performing a PCI, physicians can be forced to administer a thrombolytic drug.

In some cases the drug is not effective, and then a "rescue" PCI must be performed. The decisional problem consists in evaluating the failure probability of the thrombolytic agent, that is, the risk of performing a rescue PCI.

Current medical guidelines provided a certain amount of recommendations for such cases, but the medical knowledge is far to be complete in this field. Moreover, for AMI patients, therapeutic choices must be rapidly performed, and a wrong decision can seriously affects the patient's health.

Thus, physicians would need a set of easily applicable rules, capable to ensure an high level of accuracy in guessing whether the thrombolytic agent will be effective or not. The approach we chose for attempting to satisfy physicians needs consists in the analysis of an historical set of data through Decision Trees techniques. Decision Trees algorithms are statistical methods able to extract useful information from data under the form of decision trees or rules.

Models provided by Decision Trees algorithms can be easily applied in order to resolve decisional problems, and moreover the simple structure of such models allows the physicians to have a strict control over the evaluation process.

The next section will provide a more detailed description of Decision Trees techniques and other methods used in the present work. Section 3 will present the results of our work, next discussed in Section 4.

2. Methods

2.1. Study sample

The dataset comprised retrospective data from 102 patients (80 males and 22 females), subsequently hospitalized after an acute myocardial infarction (AMI) at the Coronary Intervention Unit of the Pugliese Ciaccio Hospital (Catanzaro, Italy). All the patients underwent a fibrinolysis intervention, followed by a Rescue PCI in 44 cases. The set of predictors variables was selected according to the current medical knowledge, and with the aim of avoiding redundant information: sex, age, presence of metabolic syndrome (MS), heart ratio, systolic blood pressure (SBP), diastolic blood pressure (DBP), ejection fraction, total cholesterol, LDL cholesterol, fibrinogens, creatinine, electrolytes (sodium, potassium) , blood count values (hemoglobin, hematocrit, platelets, neutrophils, lymphocytes), and international normalized ratio (INR). The class attribute was Rescue PCI, codified as "1" (Rescue PCI after Thrombolysis) and "0" (No Rescue PCI).

No data pre – processing procedures were applied to the data. Missing values were codified as a separate values for all predictors.

2.2. Decision trees algorithms

Decision Trees algorithms are one of the first non parametric methodology for the modelling of implicit information hidden in the data [1], able to face both classification and regression tasks.

For classification trees, the data are supposed to be a set of m tuples in the form <*Xi,Yi*>, where *Xi* is a collection of predictor variables x_{ij}, j = 1…n, and *Yi* is the class which the instance belong to.

Generally, Decision Trees algorithms recursively partition the set of tuples, until the subgroups are homogenous in terms of the class values. The output of these procedure is a decision tree model, that is, a hierarchical model composed by leaf and branch nodes that reproduces the partition of the data.

Starting from the first branching node (the root), each branching node indicates how to subdivide the instances; when the leaf nodes are reached, the class values of instances are homogeneous and no subdivisions are necessary. Depending by the specific algorithm, a certain level of non homogeneity can be allowed inside the leaf nodes.

Once the decision tree is formed, it is also possible to apply a pruning methodology: one or more branching node can be eliminated and the corresponding sub trees converted in a leaf node, in order to achieve better generalization performance.

One of the most interesting characteristics of decision tree models is their interpretability; decision tree rules can be easily compared with the previous knowledge detained over the specific problem, in order to check concordances, contradictions or the presence of new information suggesting previously unknown relations among the variables.

Several variants of decision tree algorithms exist, differing among them for the partitioning strategy and pruning methodology.

For the present study, we selected one of the most popular decision tree algorithm, the C4.5, developed by J. Quinlan.

The main characteristics of this algorithm are:

- partition rule based on the Information Gain Ratio, a particular score metrics indicating how much a variable is useful in order to predict the class value;
- deals with both numeric and categorical predictor variables;
- implements an intelligent strategy in order to treat missing values;
- pruning procedure based on a statistical estimation of the generalization error.

Further details about Quinlan' algorithm can be found in [2]. Our data contain numerical variables, categorical variables, and moreover some predictors is affected by missing values; thus, choosing the C4.5 algorithm, we were able to analyze our data without applying complex data pre – processing procedures.

2.3. Experimentation protocol

As well as other data mining algorithms, C4.5 presents numerous parameters to be set.

During our experimentations, we decided to optimize the value of two parameters: the minimum number of elements present in a leaf necessary in order to partition the node, and the width of the confidence intervals used for the statistical estimates in the pruning procedure.

All the other parameters were set to their default values. For each combination of parameters values, a ten folds cross validation procedure was performed.

Cross validation is a statistical methodology, for the estimation of the generalization performance of a given model. The instances are subdivided in k folds, and each fold is alternately used as test set, while the remaining ones are used for training the model. At the end of the procedure the k test performances are averaged, and the final model trained over the whole dataset.

The best model was selected using the Area Under the Curve performance measure (AUC) [3]. For models having statistically equivalent AUCs (equivalence tested through a t test for median comparison), we chose the model minimizing false negative error ("Rescue" instances classified as "No Rescue").

3. Results

The best decision tree model, in terms of generalization performance, is reported in Figure 1. The overall accuracy of the model is 80.39%, with an AUC value of 0.816.

Even if the performance of the model does not seem to be excellent, it should be noted that no other model exists able to predict the risk of a Rescue PCI; that is, our model represent a possible innovation for this specific medical problem. The most influencing parameter is the Systolic blood Pressure (SBP); high values of SBP (more then 135 mmHg) reduce the risk of a Rescue PCI after the Thrombolysis intervention.

For the subjects with a low SBP, the presence of Metabolic Syndrome (MS) further influences the risk of thrombolytic agent failure: patients with MS present a higher risk of undergoing a Rescue PCI.

For evaluating the risk of patients without MS, the sex and the level of fibrinogens should be evaluated: female sex and high level of fibrinogens are protective factors against the probability of a Thrombolysis failure.

The C4.5 algorithm automatically find the cut off values for numerical variables; it is worthwhile to note that the threshold chosen for SBP variable (135 mmHg) is the same value indicated by the American Heart Association & National Heart, Lung and Blood Institute (AHA/ NHLBI) for the diagnosis of Metabolic Syndrome [4]. The cut off values or the fibrinogens (435 mg/dl) did not reflect any previous known rule in the medical field; the pato – physiological meaning of such threshold should be further investigate.

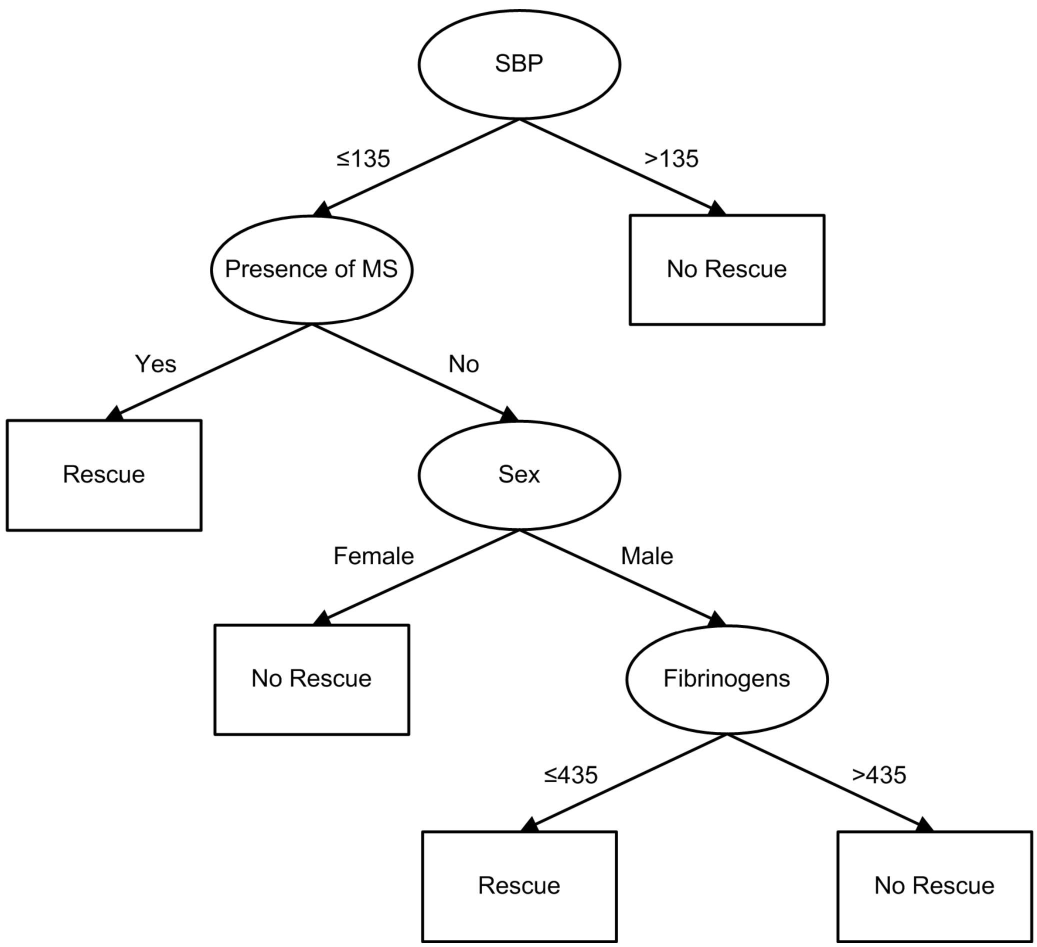

Figure 1: the best Decision Tree Model, as provided by the model selection procedures. Ovals represent branch nodes, rectangles leaf nodes.

4. Discussion and conclusions

As far as we know, this is the first study about the evaluation of risk of a Rescue PCI by exploiting machine learning approaches.

The obtained results have interesting interpretation from a medical point of view. First of all, the positive impact of a high blood systolic pressure is a really new information. It could be stated that a SBP higher than 135 mmHg indicates a good level of heart contractile capacity, that is, an heart still robust and functioning.

A never previously investigated relation is the negative impact of metabolic syndrome (MS). This syndrome consists of a constellation of highly interrelated vascular risk factors and metabolic abnormalities comprising centrally distributed obesity, atherogenic dyslipidemia, high blood pressure, and hyperglycemia [5]. This cluster of risk factors appears to increase the individual's risk of vascular disease by promoting the development of atherosclerosis and type II diabetes mellitus. [6-7] In this context, the MS is characterized by a proinflammatory and a prothrombotic state, mediated by insulin resistance, which might lead to different metabolic abnormalities and might confer a higher resistance to arterial recanalization after thrombolytic therapy. The MS was diagnosed following the criteria established by the American Heart Association & National Heart, Lung and Blood Institute (AHA/ NHLBI) in 2005 [4]. As suggested by our model, the MS was associated with a higher resistance to clotlysis after thrombolytic therapy in patients with acute coronary occlusions. After these results, the MS may not be only associated with an increased risk for cardiovascular morbidity, as previously known, but also with a poor response to thrombolytic therapy in patients with acute myocardial infarction (AMI). MS may make difficult the process of thrombolysis induced arterial recanalization in acute MI through a chronic inhibition of the endogenous fibrinolytic system [8]. Our main finding is in agreement with the results of two recent studies performed in patients with acute coronary syndromes [9] and in acute ischemic stroke [10]. Finally, as a practical consequence of this observation, diagnosis of MS may allow a priori identification of a subgroup of patients with lower probabilities of achieving early successful reperfusion after IV thrombolytic therapy, who may benefit from more aggressive reperfusion approaches.

Nevertheless, it should be noted that our study suffer of an important limitation: the sample size. Even if the results we found are supported by valid scientific arguments, further confirmation should be necessary in order to assess their validity, especially considering the innovative nature of some of our results.

Thus, successive researches will be focused on the strengthen of our founds, with a particular emphasis on the relation about metabolic syndromes and thrombolysis effectiveness.

Acknowledgements

This research work was supported by funds provided by "Supercomputing Center for the Computational Engineering – CESIC", NEC Italy and University of Calabria joint research project, 2008.

References

[1] Breiman L, Friedman J, Ohlsen R, Stone C. Classification and regression trees. Belmont: Wadsworth; 1984.

[2] Quinlan JR. Improved use of continuous attributes in C4.5. Journal of Artificial Intelligence Research 1996;4:77-90.

[3] Lasko, TA, Bhagwat JG, Zou KH, Ohno-Machado L. The use of receiver operating characteristic curves in biomedical informatics. Journal of Biomedical Informatics 2005;38:404-415.

[4] Grundy SM, Cleeman JI, Daniels SR, Donato KA, Eckel RH, Franklin BA, et al. Diagnosis and management of the metabolic syndrome: a statement for health care professionals: an American Heart Association/National Heart, Lung, and Blood Institute Scientific Statement. Circulation 2005;112:2735–2752.

[5] Eckel RH, Grundy SM, Zimmet PZ. The metabolic syndrome. Lancet 2005;365:1415–1428.

[6] Wilson PWF, D'Agostino RB, Parise H, Sullivan L, Meigs JB. Metabolic syndrome as a precursor of cardiovascular disease and type 2 diabetes mellitus. Circulation 2005;112:3066–3072.

[7] McNeill AM, Rosamond WD, Girman CJ, Golden SH, Schmidt MI, East HE, et al. The metabolic syndrome and 11-year risk of incident cardiovascular disease in the Atherosclerosis Risk in Communities Study. Diabetes Care 2005;28:385–390.

[8] Kendall DM, Harmel AP. The metabolic syndrome, type 2 diabetes, and cardiovascular disease: understanding the role of insulin resistance. Am J Manag Care 2002;8:S635–653.

[9] Celik T, Turhan H, Kursaklioglu H, Iyisoy A, Yuksel UC, Ozmen N, Isik E. Impact of metabolic syndrome on myocardial perfusion grade after primary percutaneous coronary intervention in patients with acute ST elevation myocardial infarction. Coronary Artery Dis 2006;17:339–343.

[10] Arenillas JF, Ispierto L, Millán M, Escudero D, Pérez de la Ossa N, Dorado L, et al. Metabolic syndrome and resistance to IV thrombolysis in middle cerebral artery ischemic stroke. Neurology 2008;71:190–195

Address for correspondence

Vincenzo Lagani
CESIC –NEC, C/O cubo 22 b, ponte Pietro Bucci,
Rende (CS) 87036, Italy.
vincenzo.lagani@eu.nec.com