A Comparison of Linear and Non-linear Classifiers for the Detection of Coronary Artery Disease in Stress-ECG

Georg Dorffner\textsuperscript{1}, Ernst Leitgeb\textsuperscript{2}, and Heinz Koller\textsuperscript{3}

\textsuperscript{1} Austrian Research Institute for Artificial Intelligence
Schottengasse 3, A-1010 Vienna, Austria
and Dept. of Medical Cybernetics and AI, Univ. of Vienna
georg@ai.univie.ac.at

\textsuperscript{2} Leitgeb Company, Strasshof, Austria

\textsuperscript{3} LKH Klagenfurt, Austria

Abstract. In this paper we report about a retrospective comparative study on three classifiers (multilayer perceptron, logistic classifier, and nearest neighbor classifier) applied to the task of detecting coronary artery disease in variables obtained from stress-ECG (treadmill exercise). A 10-fold cross-validation on all three methods was applied and the results were compared to expert performance. The results indicate that the multilayer perceptron had significantly higher specificity (correctly classified normals) than both the other classifiers and experts. In addition, they perform with lower standard deviation than experts, pointing to a more reliable, objective measure for diagnosis.

1 Stress-ECG and Coronary Artery Disease

The electrocardiogram (ECG) provides direct evidence of cardiac rhythm and conduction and indirect evidence of certain aspects of myocardial anatomy, blood supply and function. Electrocardiography has been used for many years as a key non-invasive method in the diagnosis and early detection of ischemic heart disease (coronary artery disease, or CAD), which is the leading cause of mortality in Western countries \[40\].

To improve the accuracy of the electrocardiogram and obtain more information on the dynamic state of the heart, exercise testing was introduced \[2\]. During stress testing not only the electrocardiogram is continuously registered but also other physiological parameters are monitored (blood pressure, physical symptoms and angina pectoris). According to different established protocols, the workload is increased step by step and the changes of parameters during stress and recovery are recorded and analysed. Skilled cardiologists achieve around 66 \% specificity (correctly classified normals) and 81 \% sensitivity (correctly classified CAD cases) in detecting CAD based on the resulting data \[3\].

The design of optimized automatic classifiers is an important contribution to the diagnosis and treatment of this wide-spread disease. In this study, we evaluated three linear and non-linear classifiers applied to this task.
2 DATA

For the retrospective comparison study, a data set of 550 subjects, including 318 patients with CAD and 132 normals were available. From this, a set of 200 subjects (175 CAD and 25 normals) came from a different recording setting and was used as an independent test set. The remaining 350 subjects (“cross-validation set”) were used in a 10-fold cross-validation using the different classifiers. Among the 107 normals in that group, 31 were athletes with no suspicion of CAD whatsoever. Subjects in the cross-validation set were aged from 18 to 89 years and included 283 males and 67 females.

Stress-ECG was done on a standard treadmill setting. The stress program consisted of 11 steps with increasing power, starting with no stress, and increasing by 25 Watts at each step, up to 250 Watts. After this, four more recording steps at rest (immediately after stress, 1, 3, and 5 minutes after stress) followed. During each step several psychological variables were registered and recorded, such as systolic and diastolic blood pressure, and heart rate; symptoms (like fatigue, sweating, etc.), and angina pectoris; different types of arrhythmia; ST-depressions in the ECG signal.

These variables were observed and judged by a cardiologist, the latter two based on the ongoing ECG. In each case, the stress part of the program was either completed (up to 250 Watts) or interrupted at severe contra-indications (e.g., severe arrhythmia). In addition, for each subject a number of demographic data was recorded, such as age, height, weight, sex, and an indication of prior infarction. Coronary angiography was used in all cases as the gold-standard reference method for deciding whether CAD was present or not.

3 Preprocessing

The raw data consisting of the above variables was preprocessed by following as closely as possible expert knowledge about the most informative parts in the observed data. For instance, instead of using the heart rate directly, an allowable interval (based on the age and weight of the subject) was computed, and the heart rate increase or decrease was compared to this interval at each time step. This corresponds to the analysis routinely applied by expert cardiologists. Symptoms and types of arrhythmia were categorized into two classes (according to severity), since each symptom occurs too rarely to be encoded separately.

With respect to the time structure of the variables, we again abided by the general routine, separating between stress and rest periods, and summing the contributions of each time step for each of the two parts (see, for instance, [4]).

This preprocessing scheme resulted in 29 numerical values (10 for the stress part, 10 for the rest part, and 9 for the time-independent patient-demographic data) used as input for the classifiers.