Proceedings of the World Congress on Electrical Engineering and Computer Systems and Science (EECSS 2015)
Barcelona, Spain – July 13 - 14, 2015
Paper No. 302

Matlab Tools Developed Under CardioRisk Project

Simão Paredes, Teresa Rocha
Instituto Politécnico de Coimbra, Instituto Superior de Engenharia de Coimbra, Portugal
Coimbra, Portugal
sparedes@isec.pt; teresa@isec.pt

Jorge Henriques, Paulo Carvalho, Diana Mendes
CISUC, Departamento de Engenharia Informática, Universidade de Coimbra, Portugal
Coimbra, Portugal
jh@dei.uc.pt, carvalho@dei.uc.pt; diana.sxm@gmail.com

Ramona Cabiddu
Cardiopulmonary Physiotherapy Laboratory, Federal University São Carlos, São Carlos, Brazil
ramona.cabiddu@gmail.com

João Morais
Centro Hospitalar de Leiria
joaomorais@chleiria.min-saude.pt

Abstract- The development of computational tools for cardiovascular disease (CVD) risk assessment can be a valuable element for aiding the clinical decision. Actually, clinical guidelines recommend the use of risk assessment tools (scores) to identify the CVD risk of each patient as the correct stratification of patients may significantly contribute to the optimization of the health care strategies.

CardioRisk project addresses this issue, namely the management of myocardial infarction (MI) patients. The main goal is the development of personalized clinical models for CVD risk assessment of acute events (death and new hospitalization), in order to improve the stratification of patients according to their care needs.
Three main hypotheses guided the developed work: *i)* it is possible to extract and combine data from several sources of information; *ii)* it is possible to identify the best classifier for each patient/group of patients; *iii)* it is possible to incorporate heart rate variability parameters in order to improve the risk assessment.

A Matlab framework was implemented to support the development and validation of several algorithms created within the CardioRisk project. This paper, intends to describe these software tools giving an insight of the respective algorithms as well as the expected interactions with the user (physician that intends to record data and assess the CVD risk assessment of a given patient).

The CardioRisk project allowed the development of methodologies that can be potentially relevant to apply in the CVD risk assessment as they provide some important achievements: *i)* performance improvement; *ii)* ability to deal with missing information; *iii)* incorporation of new risk factors/clinical expertise.

***Keywords*:** Cardiovascular Risk Assessment; Clinical Decision-Support Systems; Models Combination; Personalization.

1. Introduction

The cardiovascular disease (CVD) which includes coronary heart disease (e.g. myocardial infarction), cerebrovascular disease (stroke), heart failure, hypertension, is the world's primary cause of death. According to the World Health Organization (WHO) estimates, the number of people who die from CVD will increase to reach 23.3 million by 2030 (WHO, 2013). As a result, there is a vast research activity

directly related with cardiovascular disease trying to minimize the social and economic cost of this disease (EU, 2012)

In this context, the CardioRisk project addresses the CVD risk assessment, i.e. the evaluation of the probability of occurrence of an event (death, myocardial infarction/hospitalization) directly originated by CVD. The importance of this assessment is critical, as it can aid the physicians in patients stratification identifying and adapting the health care plan to the real needs of the different patients (Perk, 2012) as well as to increase the patients responsibility on their own health. In fact, the information and communication technology research lines reflect this paradigm, e.g. the Information and Communications Technology (ICT) research directions in disease prevention project (PREVE) states that the main goal should be: *"having the individual as a co-producer of health and empowering individuals to take responsibility of their health with personalized ICT"* (VTT , 2010).

The main goal of the CardioRisk project is the development of personalized clinical models for cardiovascular (CVD) risk assessment of acute events (e.g. death and new hospitalization). The proposed methodologies were developed based on three main hypotheses: *i)* it is possible to extract and combine data from several sources of information (current CVD risk assessment tools; clinical expertise; clinical literature); *ii)* it is possible to identify the best classifier for each patient/group of patients; *iii)* it is possible to incorporate heart rate variability data in order to improve the risk assessment.

The project specifically addresses the secondary prevention, namely, the management of myocardial infarction (MI) patients; short term (30 days after the event) and combined endpoint: death/new hospitalization.

The paper is organized as follows: section 2 briefly describes the CardioRisk project, namely the clinical platform concept. Section 3 presents the developed software and briefly describes the respective algorithms. Some final considerations are drawn in section 4.

2. CardioRisk Project

The CardioRisk project addresses the coronary artery disease (CAD), namely, the management of myocardial infarction (MI) patients in order to improve their stratification. Three main algorithms were developed: *i)* combination of available CVD risk assessment tools (information fusion); *ii)* grouping of patients; *iii)* incorporation of parameters resulting from Heart rate variability (HRV) analysis.

From the technical perspective, the development of a clinical platform (Fig. 1) integrating the patient data (Hospital Information System), the ECG acquisition (Holter device) as well as the developed algorithms was the main achievement.

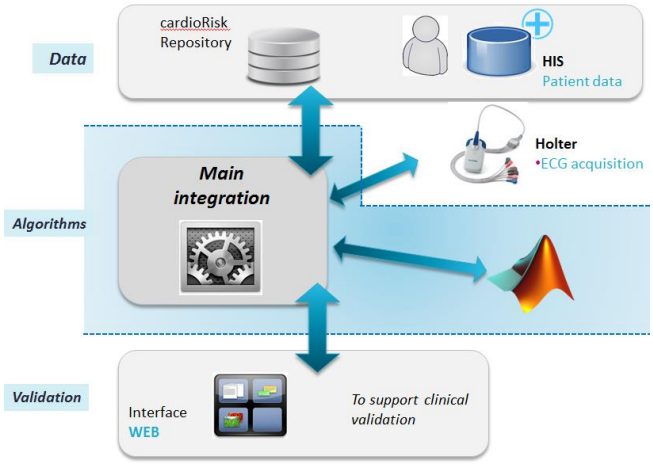


Fig. 1. CardioRisk Architecture.

As depicted in Fig. 1, the integrated clinical application comprises three main levels: *i)* data acquisition; *ii)* data analysis algorithms and *iii)* physician interface and clinical validation.

The first level involves the development of an application to collect the necessary information, namely the relevant information from Hospital Information System (HIS) and ECG collected by means of the Holter devices.

The second is devoted to the development of models for short-term risk assessment, incorporating the developed algorithms for fusion of CVD risk assessment tools, personalization of patients, and HRV analysis. This module uses as input the data from the first module and generates the required outcomes to be used by module 3. This third module presents the algorithms' results to the physicians and supports all the necessary functionalities for the clinical validation.

This clinical application also has to support an observational study focused on patients (approximately 100 patients) admitted in the intensive cardiac care unit (ICCU) with a first episode of acute MI, managed according to the current European guidelines.

The CardioRisk consortium involves three partners from two different countries (Portugal, Italy). The team is composed of two research institutions, the University of Coimbra and the Politecnico di Milano, and a public hospital, Leiria Hospital Centre. The project started in July 2013 and has a total duration of 24 months (until June 2015).

3. Matlab Tools

A Matlab framework was created to support the development and validation of the several algorithms as well as to integrate all this software according to the main goals of the project.

3. 1. Combination of CVD Risk Assessment Tools (information fusion)

This approach aims to combine CVD risk assessment tools (risk scores) and it is based on two main hypotheses: *i)* it is possible to create a common representation to the individual CVD risk assessment tools; *ii)* it is possible to combine in a common framework the resulting individual models.

In effect, these tools (currently applied in the clinical daily activity) are diversely represented (charts, equations, etc.) which does not facilitate their integration/combination. Therefore a common representation will decisively facilitate their combination. Additionally, these different representations are not suitable to deal with missing risk factors nor can they incorporate additional clinical knowledge. A unified framework would be very valuable, since it would create a flexible scheme that will allow the incorporation of clinical knowledge, as well as to deal with missing information. Furthermore, this combination structure also explores the implementation of optimization methodologies that are essential to increase the CVD risk prediction performance.

Therefore, the proposed methodology comprises two main steps. Initially, a common representation based on a naïve-Bayes classifier was applied to each individual risk score tool. Then a proper combination of the individual model's parameters followed by an optimization base on genetic algorithms was implemented, enabling the integration of the information provided by the individual tools[1] (Paredes, 2012).

CardioRisk specifically addresses the secondary prevention (CAD patients), thus this approach was applied to the combination of three well-known risk scores tools, GRACE, TIMI (no ST-elevation) and PURSUIT (Tang, 2007), (Antman, 2000), (Boersma, 2000). These models are employed in secondary prevention on CAD patients, for short-term risk assessment (1 month) of death/myocardial infarction.

Fig. 2 presents the software interface to assess the CVD risk assessment of a given patient.

The software allows the input of risk factors of a specific patient, calculates the respective CVD risk, and shows it according to the two defined categories: {"low risk", "high risk"} (Fig. 3). The physician can easily assess the risk of a specific patient or he can use the tool as a predictor, in the sense that can try different values for several parameters and foreseen the result of its adjustment.

[1] The development of this algorithm was partially supported by the results achieved by this research team in the context of the European project HeartCycle (FP7-216695).

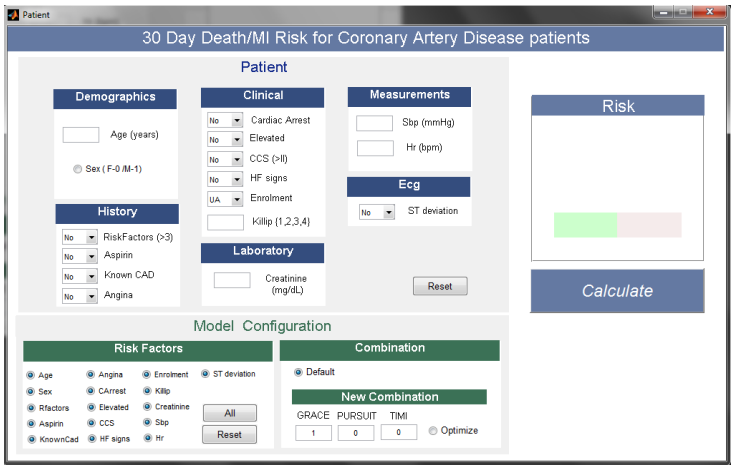


Fig. 2. Software to Assess the Individual Risk of a Patient.

The software allows the input of risk factors of a specific patient, calculates the respective CVD risk, and shows it according to the two defined categories: {"low risk", "high risk"} (Fig. 3). The physician can easily assess the risk of a specific patient or he can use the tool as a predictor, in the sense that can try different values for several parameters and foreseen the result of its adjustment.

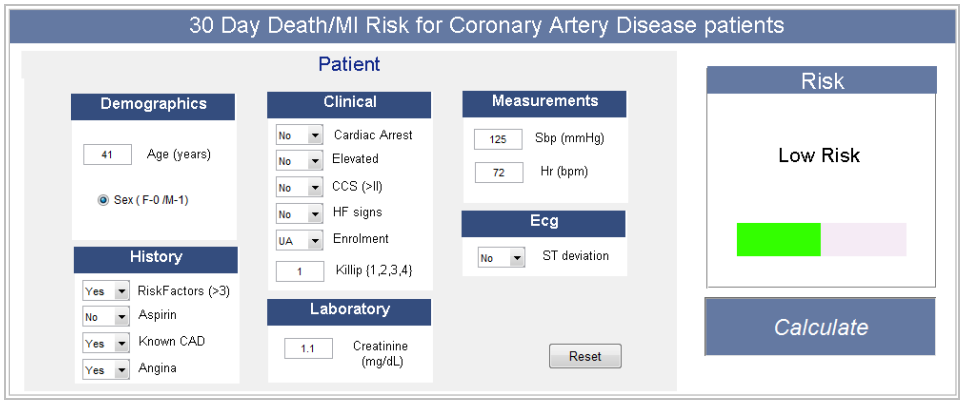


Fig. 3. Individual patient risk (example).

The physician can assess the risk of an individual patient without any additional configuration of the model (default parametrization) or, on the contrary, configure the Bayesian global model (Fig. 4). Actually, the physician may load a previously optimized parameters for the global model or alternatively perform a set of new configurations that are going to originate new values for the parameters of the global model. The physician has the possibility of selecting the risk factors that are going to integrate the global model as well as adjust the weights of the individual Bayesian models. Additionally, this configuration includes an optimization procedure (based on a Genetic Algorithms approach) to calibrate the global model to a specific population.

Some promising results were obtained (Paredes, 2015), as it was possible to improve the performance when compared to the one achieved by the individual risk assessment tools. Table. 1 presents the results obtained with a real patient dataset made available by the Santa Cruz hospital, Portugal. This dataset contains data from N=460 consecutive patients that were admitted in the Hospital, with ACS-NSTEMI between March 1999 and July 2001. The event rate of combined endpoint (death/myocardial infarction) is 7.2% (33 events) (Paredes, 2012).

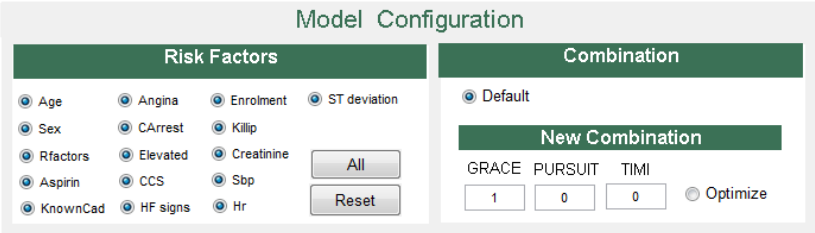


Fig. 4. Model Configuration.

Table. 1. Performance of Combination approach.

%	GRACE	PURSUIT	TIMI	Combination Scenario 1
SE *(%)*	81.82	69.70	48.58	75.76
SP *(%)*	53.40	43.80	72.60	74.71
Gmean (%)	66.10	55.24	59.33	75.23

(SE-sensitivity / SP-specificity)

Furthermore, the proposed methodology can deal with missing risk factors as well as to allow the incorporation of new risk factors (such as rules based on clinical knowledge). Other important advantage that results from the Bayesian nature of the global model is the clinical interpretability of the model. This software was integrated in the global clinical platform (data analysis algorithms module) (Fig. 1).

3. 2. Grouping of Patients

This methodology addresses exclusively the performance of CVD risk assessment. Here, as already mentioned, is explored the hypothesis that it is possible to identify the best classifier for each patient/group of patients. In fact, this hypothesis results directly from the evidence that risk assessment tools perform differently among different populations, which originates that if the patients are properly grouped it is possible to find the best model (classifier) for each group. Two different approaches were developed: *i)* Clustering patients; *ii)* Similarity measures.

Fig. 5 presents the clustering method. Initially the data is pre-processed and then a clustering algorithm is applied (subtractive clustering algorithm) (Paredes, 2014). So, patients are grouped based on the values of respective risk factors, which require the adoption of a distance metric that allows the quantification of the distance between patients.

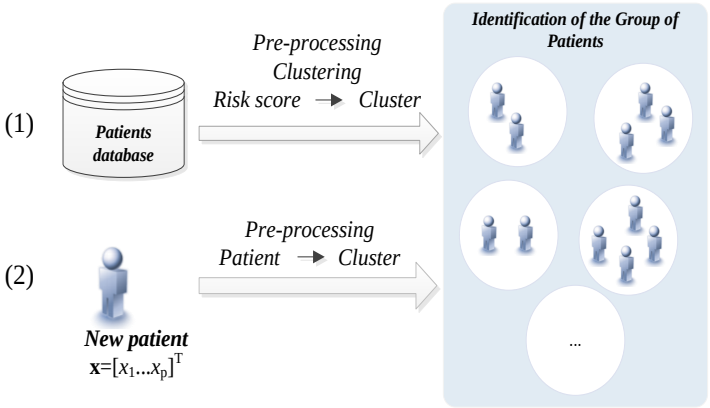


Fig. 5. Clustering Patients Approach.

The classification of a new patient can be simply described in two steps: *i)* the patient is assigned to a specific cluster (the closest one); *ii)* the patient is classified by the CVD risk assessment tool with the best performance in that cluster (Paredes, 2014). However, the clusters creation is not trivial and when validated with a real patient dataset originated some problems in the clear identification of groups of patients (Fig. 6). This difficulty was originated by the reduced dimension of the dataset (number of patients). Therefore, this method was discarded and an alternative approach was considered to create groups of patients.

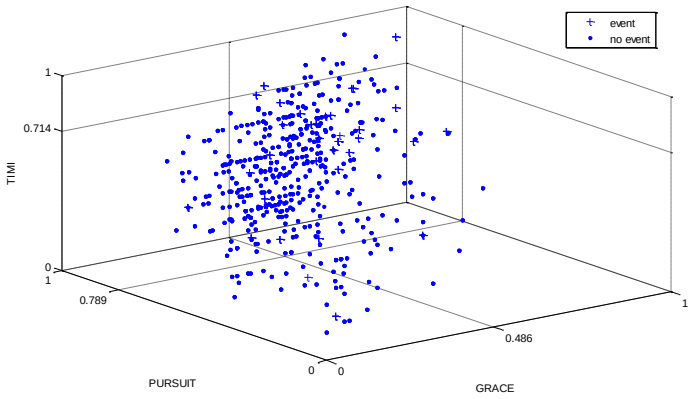


Fig. 6. Clustering applied to Santa Cruz dataset.

The similarity measures methodology proposes a simpler strategy to form groups of patients (Fig. 7). The classification of a new patient is based on a similarity measure, assuming that if a new patient is closest to one that is correctly classified by a CVD risk assessment tool, it is probable that the same tool will also be able to classify it accurately.

In this way, the groups of patients are formed by the patients correctly classified by each CVD risk tool which is different from the clustering algorithm where groups of patients are created exclusively based on the values of the risk factors (Paredes, 2014).

However, the identification of the closest patient is not obvious, which imposes a comparison among several distance metrics (e.g. Euclidean, Hamming). Additionally, with the goal of improving the identification of the closest patient, a specific weight was assigned to each risk factor. An optimization procedure, based on genetic algorithms, was carried out to adjust those weights.

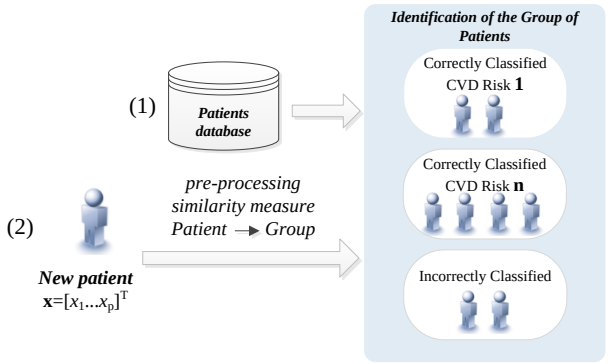


Fig. 7. Similarity Measures Approach.

Here, the best results were obtained through similarity measures approach as presented in Table 2.

Table 2. Performance of Grouping approach.

%	GRACE	PURSUIT	TIMI	Grouping Scenario 1
SE *(%)*	81.82	69.70	48.58	75.76
SP *(%)*	53.40	43.80	72.60	69.79
Gmean (%)	66.10	55.24	59.33	72.71

(SE-sensitivity / SP-specificity)

These two algorithms (clustering; similarity measures) were implemented in Matlab. The grouping approach based on the similarity measures was integrated in the clinical platform.

3. 3. HRV Analysis

The Heart Rate Variability (HRV) signal can be easily derived from the ECG and consists in the oscillation in the interval between consecutive heart beats. Cardiac rhythmicity is controlled by the autonomic nervous system (ANS). In fact, depressed HRV has been reported in several cardiovascular diseases, including coronary heart disease and heart failure (Taylor, 2010).

Both spectral and non-linear HRV derived parameters are important to this assessment. In the frequency domain, three main spectral components can be identified on the HRV signal spectrum: the very low frequency (VLF: 0.01-0.04 Hz), the low frequency (LF: 0.04-0.15 Hz) and the high frequency (HF: 0.15-0.4 Hz) components. In healthy subjects LF and HF can increase under different conditions. In normal subjects LF and HF exhibit a circadian pattern, with higher values of LF during the day and of HF at night. Moreover, an increased LF is observed during standing, mental stress and moderate exercise, while an increase in HF is induced by controlled respiration. Characteristic changes in the VLF, LF and HF bands were found in MI patients. The power spectral density (PSD) of the signal is calculated and decomposed into single spectral components, according to the method described in (Baselli G, 1997). The frequency and power values associated to each rhythmic component can subsequently be calculated. Parameters that can be derived include the normalized power of the LF and the HF components along with the LF/HF ratio.

Additionally, a non-linear analysis was carried out, as cardiac activity is determined and modulated by non-linear mechanisms. (Bianchi AM, 2010). The Detrended Fluctuation Analysis (DFA) algorithm has been applied in this project. The 1/f slope and the Lempel-Ziv Complexity (LZC) were also considered in the HRV analysis.

The algorithms to analyse the ECG as well as to perform the HRV analysis, were implemented in Matlab and integrated in the clinical platform.

3. 4. Clinical Platform

As represented in Fig. 1, a clinical platform was implemented. This software, had the following requirements: *i)* incorporation of the developed algorithms for combination of CVD risk assessment tools, grouping of patients and HRV analysis; ii) support the ECG collection performed during the observational study (mainly to perform HRV analysis); iii) support the integration with the HIS, Hospital Information System, to access the patient data. This platform was implemented in Matlab.

This platform was implemented in Matlab and developed based on three-tier architecture: *i)* presentation tier, *ii)* logic tier and *iii)* data tier. The data tier uses a proper database engine (SQLite engine) as it was considered adequate for the storage of relatively small sets of data. Furthermore, SQLite allows the application to be easily portable as it is a server less database engine.

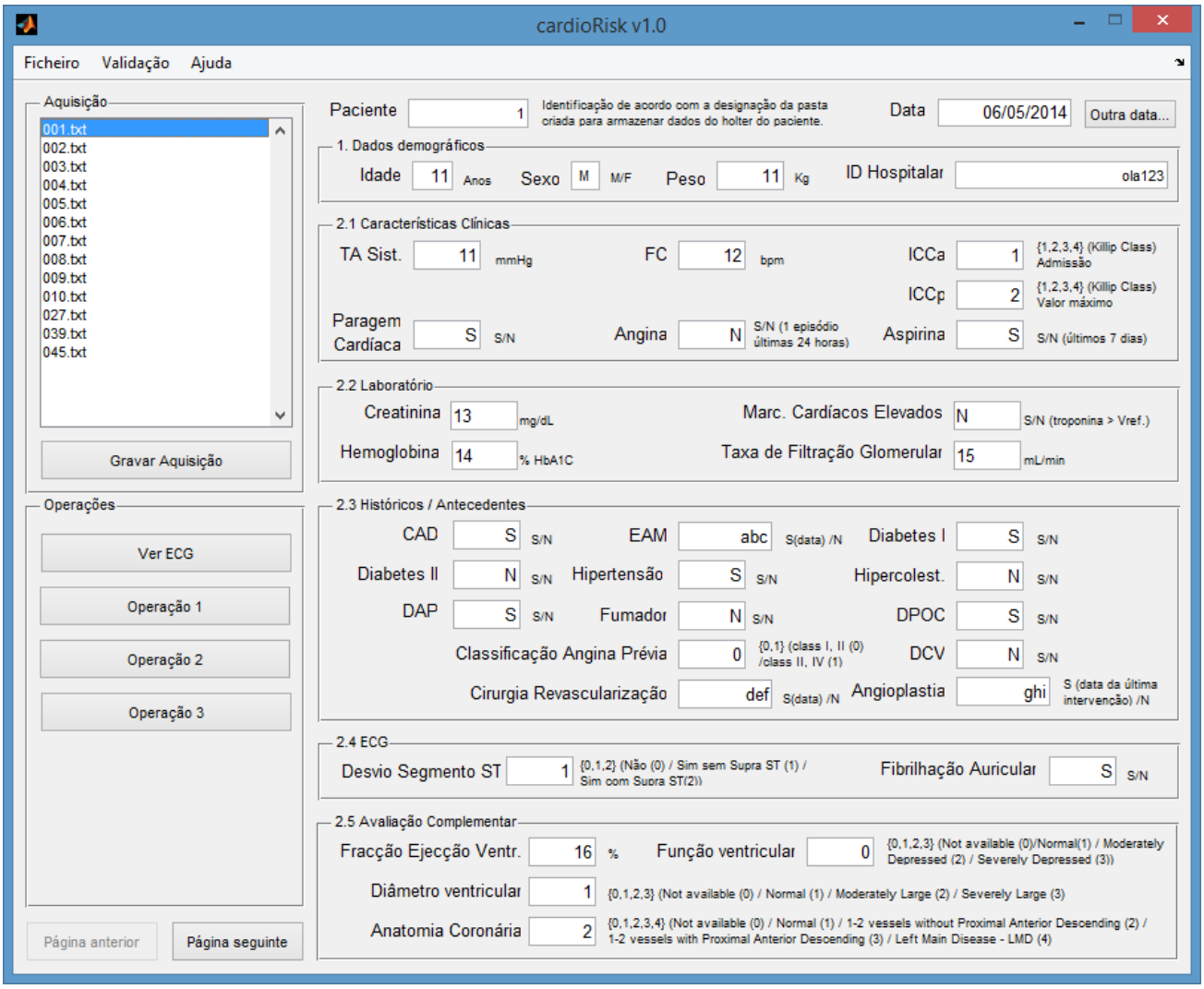
The resulting interface was designed to be used intuitively by the physician, as showed in Fig. 8.

4. Conclusion

This paper intended to provide an overview of the software developed in the CardioRisk project and simultaneously to give a brief overview of the developed methodologies as well as to present some results obtained in the preliminary validation phase. Some references to previous publications of this research team were indicated as it would be completely impossible to detail all the methodologies in this limit of pages.

The CardioRisk project allowed the development of methodologies that can be potentially relevant to apply in the CVD risk assessment as they provide some important achievements: *i)* performance improvement; *ii)* ability to deal with missing information; *iii)* incorporation of new risk factors/clinical expertise.

The ongoing research is focused on the integration of the outputs of the three developed approaches, namely in the integration of the HRV parameters.



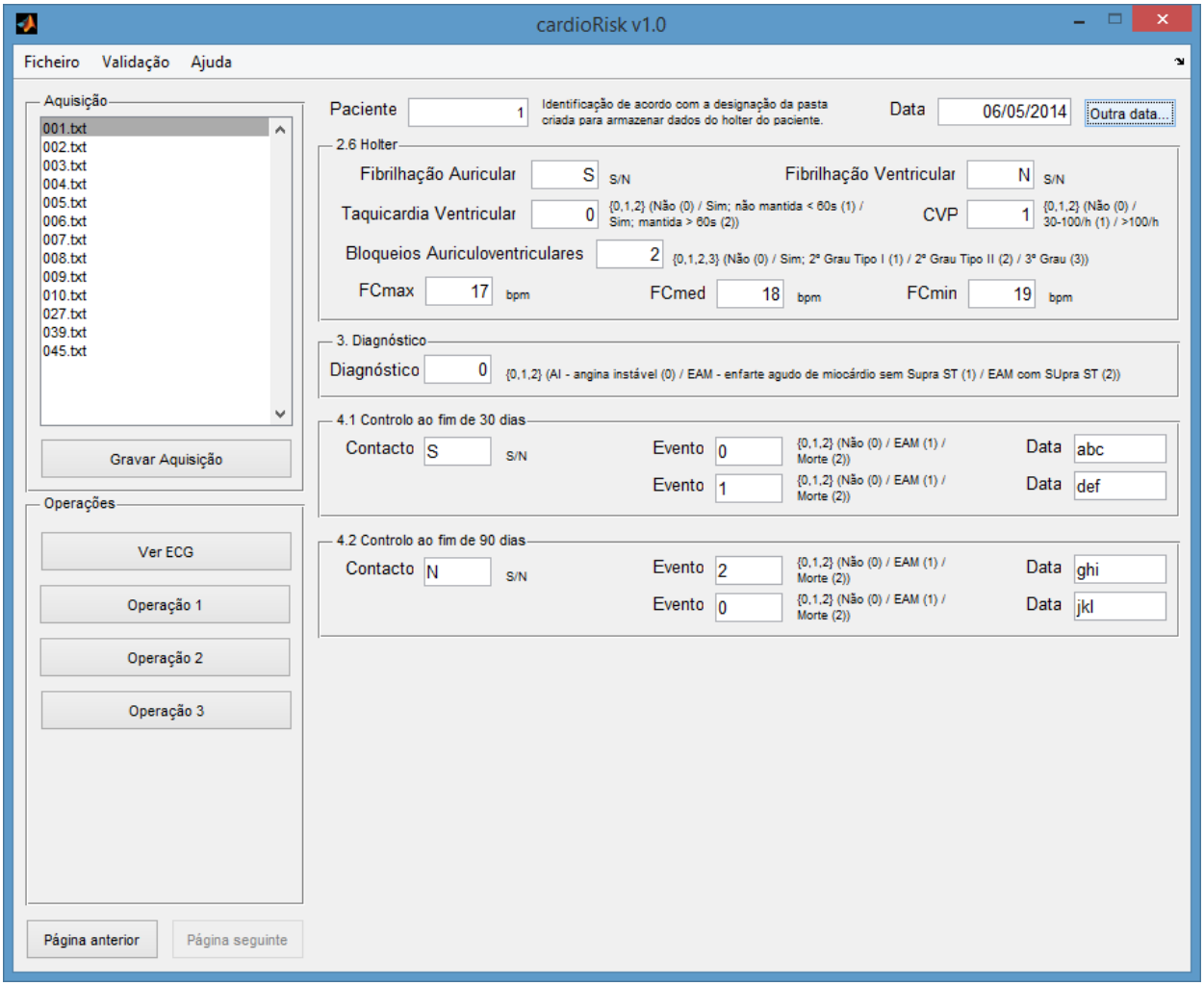


Fig. 8. Clinical Platform's GUI (application launch).

Acknowledgements

This work was supported by CardioRisk (PTDC /EEI-SII/2002/2012).

References

Antman, E. (2000). The TIMI risk score for Unstable Angina / Non-St Elevation MI – A Method for Prognostication and Therapeutic Decision Making. *Journal of American Medical Association- JAMA, 284*, 835-842.

Baselli, G. (1997). Spectral Decomposition in Multichannel Recordings Based on Multivariate Parametric Identification. *IEEE Trans. Biomed. Eng., 44*, 1092-101.

Bianchi, A. (2010). Long-Term Correlations and Complexity Analysis of the Heart Rate Variability Signal During Sleep. Comparing Normal and Pathologic Subjects. *Methods of Information in Medicine, 49*, 479-*83.*

Boersma, E. (2000). Predictors of Outcome in Patients with Acute Coronary Syndromes without Persistent ST-Segment Elevation; Results from an International Trial of 9461 Patients. *Circulation, American Heart Association - AHA, 101*, 2557-2657.

Nolan, J. (1998). Prospective Study Of Heart Rate Variability And Mortality In Chronic Heart Failure. Results of the *United Kingdom Heart Failure Evaluation and Assessment of Risk Trial (UK-Heart). 98*, 1510-6.

Paredes, S. (2012). Cardiovascular Event Risk Assessment - Fusion of Individual Risk Assessment Tools Applied to the Portuguese Population. *Proc. of the 15th Int. Conference on Inform. Fusion,* 925-932.

Paredes, S. (2014). Personalization Algorithms Applied to Cardiovascular Disease Risk Assessment. *36th Annu. Int. IEEE EMBS Conference.*

Paredes, S. (2015). The CardioRisk Project: Improvement of Cardiovascular Risk Assessment. *The Int. Conference on Computational Sci.*

Perk, J. (2012). European Guidelines on Cardiovascular Disease Prevention in Clinical Practice. *European Heart Journal, 33*, 1635–1701.

Tang, E. W. (2007). Global Registry of Acute Coronary Events (GRACE) Hospital Discharge Risk Scores Accurately Predicts Long Term Mortality Post-Acute Coronary Syndrome. *American Heart Journal*, *153*(1), 30-35.

Taylor, C. B. (2010). *Depression, Heart Rate Related Variables and Cardiovascular Disease.*

VTT . (2010). *ICT Research Directions in Disease Prevention, FP7-248197.*

Web sites:

Web-1: http://ec.europa.eu/research/health/medical-research/cardiovasculardiseases/projectsfp7_en.html consulted 18 March 2015.

Web-2: http://www.who.int/mediacentre/factsheets/fs317/en/index.html consulted in December 2014