Cardiovascular Event Risk Assessment – Fusion of Individual Risk Assessment Tools Applied to the Portuguese Population

S. Paredes†, T. Rocha†, P. de Carvalho‡, J. Henriques‡, J. Morais, J. Ferreira§, M. Mendes§*

Abstract – Cardiovascular disease (CVD) is the world's largest killer, responsible for 17.1 million deaths per year. Thus, the improvement of the prognosis of this disease is an important factor to defeat the current statistics. Although there are several risk tools available to assess the risk of occurrence of a cardiovascular event within a given period of time, these tools present some major drawbacks. In particular, each individual tool considers a reduced number of risk factors, does not permit to incorporate additional clinical knowledge and presents difficulties in coping with missing risk factors.

In order to overcome the identified weaknesses, a flexible framework is proposed here, based on the fusion of a set of distinct risk assessment tools. The methodology is based on two main hypotheses: *i*) it is possible to derive a common representation for the individual risk assessment tools, *ii*) it is possible to combine (fusion) the obtained individual models, in order to achieve the referred goals. Additionally, through the implementation of optimization techniques, an increasing in the global risk prediction performance is also investigated.

The validation of the strategy is carried out through the combination of three current risk assessment tools (GRACE, TIMI, PURSUIT) developed to predict the risk of an event in coronary artery disease (CAD) patients. The combination of these tools is validated with two real patients testing datasets: *i*) Santa Cruz Hospital, Lisbon/Portugal, N=460 ACS-NSTEMI[1] patients; *ii*) Santo André Hospital, Leiria/Portugal, N=99 ACS-NSTEMI patients.

Considering the obtained results with the available datasets it is possible to state that the initial goals of this work were achieved. This evidence makes this work a valid contribution for the improvement of the risk assessment applied to cardiovascular diseases.

I. INTRODUCTION

The cardiovascular disease[2] (CVD) disease is the world's largest killer, responsible for 17.1 million deaths per year [1]. In fact, each year, cardiovascular disease causes over 4.3 million deaths in Europe and almost 2.0 million in the European Union. Consequently, CVD is the main cause of illness and death in Europe, responsible for 23% of the total disease burden [2]. Moreover, CVD alone represents approximately €192 billion /year to health costs in the European Union [3]. Furthermore, the population of the EU and the western world is aging. The number of elderly people aged 65-79 will increase approximately by 37% by 2030 [4]. It is recognized that this demographic change in the population will result in unaffordable health costs.

In this context, the correct diagnosis and prognosis of cardiovascular disease assumes a particular importance in trying to reduce these statistics. The assessment of the risk of occurrence of an event, i.e. the evaluation of the probability of occurrence of an event given the patient's past and current exposure to risk factors, is critical to improve prognosis. This way, it is possible to increase the quality of preventive health care, as this assessment data will help physicians to identify and adapt the treatment/care plan to an individual patient [5][6].

Several risk score tools[3] were developed to assess the probability of occurrence of a CVD event within a certain period of time (months/years). Available risk score tools differ on the assessed period of time (months/years), disease (coronary artery disease, heart failure, ...), predicted events (death/non-fatal), risk factors considered in the model, patient's conditions (ambulatory patients, hospitalized patients, cardiac transplant candidates,…).

These tools are very useful although they present some important weaknesses: *i*) they ignore the information provided by other risk assessment tools that were previously developed, *ii*) each individual tool considers a reduced number of risk factors, *iii*) they have difficulty in coping with missing risk factors, *iv*) they do not allow the incorporation of additional clinical knowledge, *v*) some tools do not assure the clinical interpretability of the respective parameters.

The proposed approach aims to defeat these flaws and simultaneously consider the valuable information provided by these tools. Therefore, rather than to derive a new model, the proposed methodology intends to create a flexible global framework (global model) based on the combination of available risk assessment tools.

The combination of individual tools will also permit to overcome of the additional difficulty of selecting the best tool to use in the daily clinical practice. Actually, one of these statistical tools is typically selected as the standard

This work was supported by HeartCycle EU project (FP7-216695) and CISUC (Center for Informatics and Systems of University of Coimbra).

† Instituto Politécnico de Coimbra, Departamento de Engenharia Informática e de Sistemas, Rua Pedro Nunes, 3030-199 Coimbra. {sparedes@isec.pt, teresa@isec.pt}

‡ CISUC, Departamento de Engenharia Informática, Universidade de Coimbra, Pólo II, 3030-290 Coimbra,{carvalho@dei.uc.pt, jh@dei.uc.pt}

*Serviço de Cardiologia, Hospital Santo André, EPE Rua das Olhalvas, 2410-197 Leiria, {joaomorais@hsaleiria.min-saude.pt}

§Serviço de Cardiologia, Hospital Santa Cruz, Lisboa, {jorge_ferreira@netcabo.pt, miguel.mendes.md@sapo.pt }

[1] ACS-NSTEMI Acute Coronary Syndrome with non-ST segment elevation

[2] Cardiovascular disease is caused by disorders of the heart and blood vessels, including coronary heart disease (heart attacks), cerebrovascular disease (stroke), raised blood pressure (hypertension), peripheral artery disease, rheumatic heart disease, congenital heart disease and heart failure.

[3] In order to clarify, risk assessment models that have been statistically validated and are available in literature are going to be designated through this work as **risk assessment tools**.

model to be applied in a given institution. However, the choice of the risk assessment tool can be difficult since there might not be a consensus about the best tool to use.

The developed methodology [7] is based on two main hypotheses:

i) It is possible to implement a common representation of individual risk assessment tools. In fact, current risk assessment tools are diversely represented (charts, equations, scores,…). This does not facilitate their integration/combination. The ability to deal with missing risk factor along with the flexibility to incorporate additional clinical knowledge (new risk factors) are further aspects that influenced the selection of the classifier to implement this first step. Moreover, its parameters/rules must be clinically interpretable;

ii) It is possible to implement a combination of the obtained individual models. The ability of combining available knowledge from various sources is useful since it creates a flexible global framework which originates the mentioned benefits. Additionally, the parameters of the global model resulting from this combination can be adjusted by means of optimization methodologies (such as genetic algorithms), in order to increase the CVD risk prediction performance.

This approach was validated with current risk assessment tools specific for secondary prevention on coronary artery disease (CAD) patients, in particular for assessing the risk of death/myocardial infarction within a short period of time (days/months). Here, of particular relevance are the statistical risk assessment tools GRACE, TIMI (no ST-elevation) and PURSUIT [8][9][10]. This validation was supported by two real patients testing datasets: *i*) Santa Cruz Hospital, Lisbon/Portugal, N=460 ACS-NSTEMI patients; *ii*) Santo André Hospital, Leiria/Portugal, N=99 ACS-NSTEMI patients.

The paper is organized as follows: in section II an outline of the developed methodology is presented. In section III the results of the validation procedure with the two datasets are discussed. Section IV summarizes the main conclusions and the main research paths to be followed up in the near future.

II. METHODOLOGY

Figure 1 presents the proposed combination methodology:

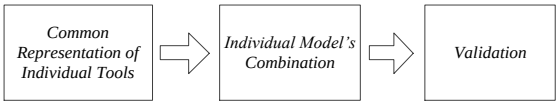

Figure 1 – Proposed Methodology

The first step of this approach consists in the creation of a common representation, based on a machine learning classification methodology that can be applied to all the selected individual tools[4]. The classifier must be selected considering both the combination of these individual models and also dealing with missing risk factors. Moreover, this common representation must assure the clinical interpretability of the model.

The second step of the methodology is the combination of the individual models. In this phase the individual models that were originated from the previous step (common representation of individual risk assessment tools) are combined. The global model that results from this combination scheme must be derived based on the available input risk factors and the individual models' selection criteria. For instance, if one individual model does not have any of its input values available, then that model should not be considered for integration in the combination scheme. This approach allows a very flexible model which is able to consider a variable number of input risk factors, it enables to incorporate empirical clinical knowledge and it avoids the necessity of choosing a particular model as a standard model for the clinical practice. However, the clinical relevance of a CVD risk prediction system depends directly of its performance. Optimization techniques, namely genetic algorithms, are adopted in this stage to increase the global model's performance (maximize sensitivity and specificity).

The third phase is validation that is determinant to evaluate the potential clinical importance of the proposed methodology, namely using sensitivity and specificity metrics. This phase is performed based on real data and it intends to be as inclusive as possible.

A. Common Representation of Individual Tools

Current individual risk score models are described by different equations/scores/charts [8][9][10]. So, in order to ease their combination, all the individual risk score tools were represented based on a similar structure, i.e. the same classifier.

The naïve Bayes classifier was the selected classifier. In fact, the naïve Bayes has a competitive performance with the remaining classifiers, is simple and can deal with missing risk factors. Besides these features, naïve Bayes assures the interpretability of the model which is one of the main goals of this approach. Finally, the structure of naïve Bayes simplifies the incorporation of empirical clinical knowledge [11][12]. Figure 2 depicts the structure of a naïve Bayes classifier.

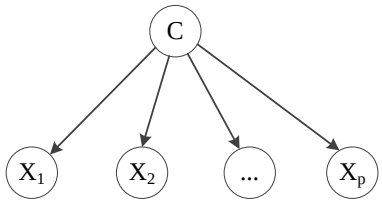

Figure 2 – Naïve Bayes Structure

The variable $X = [X_1 ... X_p]$ is a vector of random variables and C is a random variable that denotes the class of an instance, where c is a particular class label. A vector $\mathbf{x} = [x_1 ... x_p]$ represents a particular instance that contains the observed values of the different p attributes, i.e., $X = \mathbf{x}$ is

[4] The selection of current available risk assessment tools must be done according to the specific CVD risk assessment context, e.g. secondary prevention, coronary artery disease patients, 1 month, ...

the same as $X_1 = x_1 \wedge ... \wedge X_p = x_p$. In the context of this work, $X = [X_1 ... X_p]$ is a set of observations (risk factors) such as clinical examination and laboratory measurements and *C* is the hypothesis (e.g. risk level is "High").
The inference mechanism is described by equation (1)

$$P(C \mid X) = P(C \mid X_1, ..., X_p) = \alpha \, P(C) \prod_{i=1}^{p} P(X_i \mid C) \qquad (1)$$

The term $P(C \mid X)$ is the probability that the hypothesis is correct after observations have occurred (e.g., the probability that risk is "High" given the results of a clinical examination, measurements,…). $P(C)$ is the probability that the hypothesis is correct before seeing any observation (in this example, the prevalence of the risk level). α is a normalization constant. $P(X_i \mid C)$ is a likelihood expressing the probability of the observation X_i being made if the hypothesis is correct (equivalent to the sensitivity of the clinical examination). This particular Bayesian inference mechanism (naïve Bayes) assumes that attributes $X = [X_1 ... X_p]$, are conditionally independent, given the value of hypothesis C [13].
It is recognized that the violation of the assumption of independence may affect the performance of naïve Bayes classifier [12][13][14]. In the present combination scheme the selection of risk factors considered by the individual tools results from a statistical analysis process. This procedure usually starts with a large set of candidate risk factors, where the most relevant, typically not correlated, are selected. Therefore, the eventual violation of the attribute independence is controlled as the attributes' independence is addressed in the statistical derivation of each individual risk assessment tool. Moreover, the present methodology addresses this potential lack of performance through the implementation of an optimization procedure, that is carried out in the models' combination phase, by means of a genetic algorithm approach.
The structure of naïve Bayes classifier is completely defined (Figure 2) as a result the construction of the classifier is restricted to parameters' learning. Thus, the model has to learn from the training data set, the conditional probability $P(X_i \mid C)$ of each attribute X_i given the class C as well as the prior probability $P(C)$ of the class C.
Then, the process of representing an individual risk assessment tool as a naïve Bayes classifier can be systematized as follows:

- A training dataset (*N* instances $\mathbf{x} = [x_1 ... x_p]$ composed of p attributes) is generated.
- This training dataset is applied to a given risk assessment tool in order to obtain a complete labelled dataset $J = \{(\mathbf{x}_1, c_1), ..., (\mathbf{x}_N, c_N)\}$.
- Based on J and through the maximum likelihood estimation (2) it is possible to derivate a naïve Bayes classifier that resembles the behavior of that specific risk assessment tool. The prior probability $P(C)$ results directly from distribution of the class values. The conditional probabilities can be calculated through the following expression:

$$P(X_i = x_i \mid C = c) = \frac{\sum_{1}^{N} (X_i = x_i \wedge C = c)}{\sum_{1}^{N} (C = c)} \qquad (2)$$

It is important to refer that this probabilities' estimation is reliable only when the attributes are qualitative. Hence the discretization of numeric attributes may have a great impact in the construction of the conditional probabilities tables and therefore in the performance of the classifier. The Equal Width Discretization (EWD) was the selected discretization method to allow the application of the maximum likelihood estimation to numeric attributes [15].
This process must be repeated to each one of the individual risk assessment tools that integrate the combination scheme.

B. Individual Model's Combination

1) Combination strategy

The models' combination phase is responsible for the combination of the naïve Bayes classifiers that resemble the behavior of each one of the risk assessment tools that integrate the combination scheme. This procedure, is described in Figure 3.

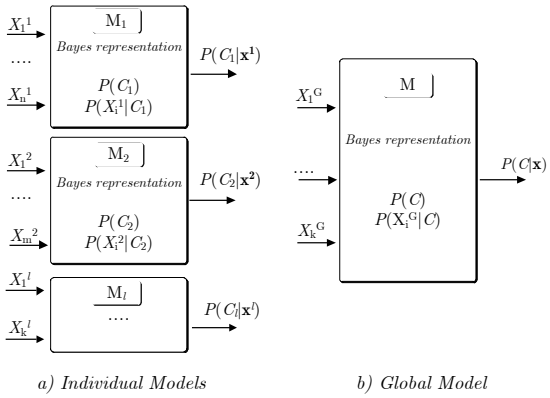

Figure 3 – Combination Scheme

Several individual models $M_i \in M = \{M_1, ..., M_l\}$ are considered to integrate the combination scheme, where each classifier is characterized by a specific conditional probability table $P(X_i^j \mid C)$, and by their respective prior probability of output class $P(C_j)$. $P(X_i^j \mid C_j)$ represents the conditional probability table of attribute i of model j, $P(C_j)$ is the prior probability distribution of model j

regarding a specific number of mutually exclusive classes, $\mathbf{x}^j \subseteq \mathbf{x} = [x_1 ... x_p]$ is the input instance considered by the model j (subset of the p risk factors that are considered by the individual model j).

The combination scheme implements the direct combination of the individual models' parameters (models' fusion) according to the following equation:

$$P(C) = \sum_{j=1}^{l} P(C_j) \times \frac{w_j}{\Phi} \quad ; \quad \Phi = \sum_{j=1}^{l} w_j$$
$$P(X_i \mid C) = \sum_{j=1}^{b} P(X_i^j \mid C_j) \times \frac{w_j}{\vartheta} \quad ; \quad \vartheta = \sum_{j=1}^{b} w_j \tag{3}$$

Where l is the number of individual models, b is the number of individual models that contain the attribute $X_i \in X$, C_j denotes each individual model that contains X_i, w_j is the weight of model j.

This combination scheme is very flexible which permits the implementation of a combination strategy that depends on the characteristics of each specific combination, namely it: *i*) permits to assign to each individual model a different weight that is proportional to the respective performance; *ii*) allows disabling a specific model. In this way, different individual model selection criteria to integrate the combination scheme may be implemented; *iii*) allows the incorporation of additional features to improve risk prediction. For instance, the clinical partners that collaborated in this work identified the possibility of assign different weights to the different attributes/risk factors.

The model selection criterion highly influences the global classification performance. According to the implemented condition, the information of a given model was considered if there was available at least one of its inputs. Moreover, some risk factors may be considered by more than one model, while other inputs belong only to a single model. Therefore, the classification of the global model is dependent on the availability of input risk factors as well as on the selection criteria to define the individual models that should be included in the combination scheme.

Additionally, to allow the combination of different individual models the following condition has to be verified: "*Individual models have the same number of output levels (e.g. "low/intermediate", "high")*". This restriction ensures that models share the same risk assessment goal[5].

This methodology also makes the incorporation of clinical expertise a straightforward operation. In fact, a new model can be directly created by the physician based on a CPT definition, and easily incorporated in the combination scheme. This is an important characteristic of this method.

2) Optimization

An additional optimization step can be performed to improve the performance of the global model. Conditional probability tables $P(X_i \mid C)$ of the global model can be optimized by means of an optimization strategy, such as genetic algorithms (GA). This algorithm focuses on the parameters ($P(X_i \mid C); P(C)$) of the global model that was created through (3).

An evaluation function must be defined in order to assign a quality measure to each candidate solution. As a result from several experiments, the selected evaluation function is composed of two functions (f_1, f_2 *multiobjective optimization*[6]) since the optimization attempts to maximize simultaneously the specificity and the sensitivity of the global model. The objective functions are given by (4) in order to transform the maximization of specificity and sensitivity into a minimization problem.

$$f_1 = 1 - \frac{TP}{TP + FN}; \; f_2 = 1 - \frac{TN}{TN + FP} \tag{4}$$

TP : True positive; FP : False positive; TN : True negative; FN : False negative

The optimization procedure is restricted to the neighbourhood of the initial values in order to assure the clinical interpretability of the final model.

For instance, considering three possible categories for the attribute X_1, $\{x_1^1, x_1^2, x_1^3\}$ and two mutually exclusive risk classes $\{c_1, c_2\}$ for the output C, the respective conditional probability table is defined by a 3×2 matrix, as shown in equation .

$$\begin{bmatrix} P(X_1 = x_1^1 \mid C = c_1) \pm \delta_{11} & P(X_1 = x_1^1 \mid C = c_2) \pm \delta_{12} \\ P(X_1 = x_1^2 \mid C = c_1) \pm \delta_{21} & P(X_1 = x_1^2 \mid C = c_2) \pm \delta_{22} \\ P(X_1 = x_1^3 \mid C = c_1) \pm \delta_{31} & P(X_1 = x_1^3 \mid C = c_2) \pm \delta_{32} \end{bmatrix} \tag{5}$$

Variables δ_{kj} denote the variations on the probability of the category k of attribute X_i given the class j.

C. Validation

The validation procedure focused on the evaluation of the performance of the global model that is originated through the combination of current risk assessment tools.

In this particular case, some current tools suitable to predict risk in coronary artery disease (CAD) patients have been selected.

Two different datasets made available by two Portuguese hospitals were used as testing datasets, while the training

[5] It is important to emphasize that this restriction does not obstructs the clinical application of the proposed methodology. In fact, from the clinical perspective the main goal is the identification of the high risk patients.

[6] Multiobjective optimization is applied when a single objective with several constraints does not adequately represent the optimization problem. In multiobjective optimization there is a vector of objective functions, where a tradeoff between objectives must be found.

dataset required to generate the parameters to represent the individual Bayesian classifiers was derived based on proper values available in literature.

The validation phase was composed of three main steps: *i*) performance assessment of individual tools. Here, individual tools were tested with both populations. This data provided some additional knowledge to adjust the weights of individual tools; *ii*) performance assessment of the global model. The performance of the global model was evaluated when different weights were assigned to the individual models. For all testing datasets, the Bayesian global model has been compared with a voting model[7] as well as with each one of the individual tools. In order to increase the statistical significance of the obtained results, bootstrapping validation was employed which allowed the derivation of confidence intervals of the formulas assessed. Parametric statistical significance tests (Student's t-test, Levene's test) were executed to increase the reliability of the conclusions extracted from this comparison; *iii*) missing risk factors. The ability of the Bayesian global model to deal with missing risk factors was also assessed. Each variable has been successively removed. For each variable, the performance of three different models was evaluated: *i*) Bayesian global model before optimization; *ii*) Bayesian global model after optimization; *iii*) Voting model[8] [14]. The different models' performance were compared based on parametric statistical significance tests (Student' t-test, Levene's test) that were complemented with an analysis of variance. Also in this situation the validation was based on the bootstrapping validation with $n = 1000$ bootstrap samples.

III. Results

A. Testing Datasets

1) Santa Cruz Hospital Dataset

This dataset contains data from N=460 consecutive patients that were admitted in the Santa Cruz Hospital, Lisbon, with Acute Coronary Syndrome with non-ST segment elevation (ACS-NSTEMI) between March 1999 and July 2001.

Table I contains the number of events considering two different endpoints (death/myocardial infarction) and two different periods (one month/one year).

TABLE I
ENDPOINTS OF SANTA CRUZ HOSPITAL DATASET

Time	Event	n	%	Total
30 days	D	13	2.8	33
	MI	24	5.2	7.2%
1 year	D	32	7.0	70
	MI	49	10.7	15.4

D: Death; **MI**: Myocardial Infarction

[7] In voting, the classification produced by a classifier is considered as a vote for a particular class value. The class with the highest number of votes is selected as the final classification.

[8] If the missing variables were continuous the replacement has been done based on the respective mean values, in the case of Boolean variables their value were successively replaced by 0 and 1 values.

Table II presents the main clinical characteristics of such patients (a detailed analysis can be found in Gonçalves *et. al.* [16]). Continuous variables with a normal distribution are expressed as mean value and standard deviation. Discrete variables are presented as frequencies and percent values.

TABLE II
CLINICAL CHARACTERISTICS OF PATIENTS THAT INTEGRATE THE DATASET

Model	Event
Age (years)	63.4 ± 10.8
Sex (Male/Female)	361 (78.5%) / 99 (21.5%)
Risk Factors:	
Diabetes (0/1)	352 (76.5%) / 108 (23.5%)
Hypercholesterolemia (0/1)	180 (39.1%) / 280 (60.9%)
Hypertension (0/1)	176 (38.3%) / 284 (61.7%)
Smoking (0/1)	362 (78.7 %) / 98 (21.3%)
Previous History / Known CAD	
Myocardial Infarction (0/1)	249 (54.0%) / 211 (46.0%)
Myocardial Revascularization (0/1)	239 (51.9%) / 221 (48.1%)
PTCA	146 (31.7%)
CABG	103 (22.4%)
Sbp (mmHg)	142.4 ± 26.9
Hr (bpm)	75.3 ± 18.1
Creatinine (mg/dl)	1.37 ± 1.26
Enrolment [0 UA, 1 MI]	180 (39.1 %) / 280 (60.9%)
Killip 1/2/3/4	395 (85.9%) / 31 (6.8%) / 33 (7.3 %) / 0%
CCS [0 I/II; 1 CSS III/IV]	110 (24.0%) / 350 (76.0%)
ST Segment Deviation (0/1)	216 (47.0%) / 244 (53.0%)
Signs of Heart Failure(0/1)	395 (85.9%) / 65 (14.1%)
Tn I > 0.1 ng/ml (0/1)	313 (68.0%) / 147 (32.0%)
Cardiac Arrest Admission (0/1)	460 (100%) / 0%
Aspirin (0/1)	184 (40.0%) / 276 (60.0%)
Angina (0/1)	19 (4.0%) / 441 (96.0%)

2) Santo André Hospital Dataset

Table III presents the main clinical characteristics of patients' data collected in Santo André Hospital, Leiria, Portugal.

TABLE III
CLINICAL CHARACTERISTICS OF PATIENTS THAT INTEGRATE THE DATASET

Model	Event
Age (years)	68.0 ± 11.8
Sex (Male/Female)	68 (68.7%) / 31 (31.3%)
Risk Factors:	
Diabetes DMIT (0/1)	91(91.9%) / 8 (8.1%)
Diabetes DMNIT (0/1)	70 (70.7%) / 29 (29.3%)
Hypercholesterolemia (0/1)	59 (59.6%)/ 40 (40.4%)
Hypertension (0/1)	26 (26.3%) / 73 (73.7%)
Smoking (0/1)	83 (83.8) / 16 (16.2%)
Previous History / Known CAD	66 (66.7%) / 33 (33.3%)
Sbp (mmHg)	145.7 ± 32.1
Hr (bpm)	83.2 ± 20.2
Creatinine (mg/dl)	1.11 ± 0.42
Enrolment [0 UA, 1 MI]	6 (6.1%) / 93 (93.9%)
Killip 1/2/3/4	70 (70.7%) / 21 (21.2%) / 7 (7.1%) /1 (1%)
CCS [0 I/II; 1 CSS III/IV]	78 (78.8%) / 21 (21.2%)
ST Segment Deviation (0/1)	98 (99%) / 1 (1%)
Signs of Heart Failure(0/1)	70 (70.7%) / 29 (29.3%)
Tn I > 0.1 ng/ml (0/1)	7 (7.1%) / 92 (92.9%)
Cardiac Arrest Admission (0/1)	98 (99%) / 1 (1%)
Aspirin (0/1)	71 (71.7%) / 28 (28.3%)
Angina (0/1)	33 (33.3%) / 66 (66.7%)

The available dataset contains data from N=99 patients that were admitted in the Hospital with Acute Coronary Syndrome (ACS-NSTEMI) during 2007.

B. Training Data Set

The approach proposed by Twardy *et. al.* [17] was followed to the generation of the training data set. Continuous variables were normally distributed. Values for mean and standard deviation were taken from the literature [16]. Discrete variables are binary and were generated through a random process. The training data set was created $\mathbf{x}^i = [x_1^i ... x_p^i]$ for all $1 \le i \le N$: with $N = 1000$. This training dataset was applied to the selected risk assessment tools (Table IV) in order to obtain the respective output class $J = \{(\mathbf{x}_1, c_1), ..., (\mathbf{x}_N, c_N)\}$.

C. Individual Risk Assessment Tools

Table IV presents the selected individual risk assessment tools to predict death/MI for CAD patients within a short period.

TABLE IV
SHORT-TERM RISK ASSESSMENT MODELS

Model	Event	Time	Prev. Type	Risk Factors
GRACE [7]	D MI	6 m	S	Age, SBP, CAA HR, Cr, STD, ECM, CHF
PURSUIT [8]	D MI	30 d	S	Age, Sex, SBP, CCS, HR, STD, ERL, HF
TIMI [9]	D MI UR	14 d	S	Age, STD, ECM, KCAD, AS, AG, RF

D: Death; **MI**: Myocardial Infarction; **UR**: Urgent revascularization
m: months; **d**: days; **S**: Secondary Prevention;
Cr-Creatinine, **HR** – Heart Rate, **CAA** – Cardiac Arrest at Admission, **CHF** – Congestive Heart Failure, **STD** - ST Segment. Depression, **ECE** - Elevated Cardiac Markers/Enzymes, **KCAD**- Known CAD, **ERL** – Enrolment (MI/UA), **HF** –Heart Failure, **CCS** – Angina classification, **AS** - Use of aspirin in the previous 7 days, **AG** - 2 or more angina events in past 24 hrs, **RF** - 3 or more cardiac risk factors

D. Individual Risk Assessment Tools' Performance

The proposed combination scheme requires that individual models have the same number of output levels. This work defines the risk stratification in two categories: {"low/intermediate risk" , "high risk" }.Therefore, the "high risk" category in the original models matches the new "high risk". The remaining original categories were grouped into "low/intermediate risk" category.

Table V shows the performance of the three individual models when a period of 30 days is considered.

As observed the three models present a very different ability to predict the endpoint in the three different testing situations. GRACE was the risk assessment tool with the best performance and discrimination capability in the three test situations. TIMI and PURSUIT presented a poor performance, so they are not as suitable as GRACE to the endpoint prediction in the considered datasets.

TABLE V
PERFORMANCE OF SELECTED INDIVIDUAL RISK ASSESSMENT TOOLS

Model	%	Santa Cruz 30 days/D/MI	Santa Cruz 30 days/D	Santo André 30 days/D
GRACE	SE	60.6	76.9	60.0
	SP	74.9	73.8	60.6
	Acc	73.9	73.9	60.6
	AUC	0.67	0.765	0.600
PURSUIT	SE	42.4	38.5	20.0
	SP	74.2	73.4	72.3
	Acc	72.0	72.4	69.7
	AUC	0.575	0.565	0.5*
TIMI	SE	33.3	23.1	20.0
	SP	73.5	72.9	93.6
	Acc	70.7	71.5	89.9
	AUC	0.525	0.5*	0.575

***SE**: Sensitivity (%); **SP**: Specificity (%); **ACC**: Accuracy (%); **AUC:** area under the Receiver Operating Characteristic*

E. Individual Models' Combination

The Bayesian global model was derived according to the methodology explained in II. The global voting model was implemented considering the votes (0/1) of the three individual models.

TABLE VI
PERFORMANCES COMPARISON – **SANTA CRUZ, (DEATH/MI)**

	%	GRACE	PURSUIT	TIMI	ByG	Vot.
Original	SE	60.6	42.4	33.3	60.6	48.5
	SP	74.9	74.2	73.5	67.0	75.6
	Gmean	67.3	56.0	49.4	63.4	60.6
	AUC	0.675	0.575	0.525	0.635	0.625
Boot Samples n=1000	SE	60.8 *(60.2; 61.3)*	42.4 *(41.9;43.1)*	33.5 *(33.0; 34.0)*	60.6 *(60.1;61.3)*	48.6 *(48.0;49.2)*
	SP	74.9 *(74.8; 75.1)*	74.2 *(74.1;74.3)*	73.6 *(73.5; 73.7)*	67.0 *(66.9;67.2)*	75.6 *(75.5;75.8)*
	Gmean	67.3 *(67.0; 67.6)*	55.8 *(55.5;56.2)*	49.3 *(48.9; 49.7)*	63.6 *(63.3;63.9)*	60.3 *(60.0;60.7)*

TABLE VII
PERFORMANCES COMPARISON – **SANTA CRUZ, DEATH**

	%	GRACE	PURSUIT	TIMI	ByG	Vot.
Original	SE	76.9	38.5	23.1	61.5	53.8
	SP	73.8	73.4	72.9	65.7	74.7
	Gmean	75.3	53.1	40.6	63.5	63.0
	AUC	0.765	0.565	0.5	0.625	0.625
Boot Samples n=1000	SE	77.3 *(76.5;78.0)*	38.2 *(37.4;39.2)*	23.0 *(22.3;23.7)*	61.6 *(60.7;62.5)*	53.7 *(52.9;54.7)*
	SP	73.8 *(73.6;73.9)*	73.3 *(73.1;73.4)*	72.9 *(72.8;73.1)*	65.8 *(65.6;65.9)*	74.6 *(74.5;74.8)*
	Gmean	75.2 *(74.9;75.6)*	51.8 *(51.1;52.5)*	38.8 *(38.0;39.5)*	63.1 *(62.7;63.6)*	62.7 *(62.2;63.3)*

TABLE VIII
PERFORMANCES COMPARISON – **SANTO ANDRÉ, DEATH**

	%	GRACE	PURSUIT	TIMI	ByG	Vot.
Original	SE	60.0	20.0	20.0	80.0	40.0
	SP	60.6	72.3	93.6	67.0	74.5
	Gmean	60.2	38.0	43.2	73.2	54.5
	AUC	0.6	0.5	0.575	0.725	0.575
Boot Samples n=1000	SE	61.2 *(59.8;62.8)*	19.9 *(18.6;21.2)*	21.5 *(20.3;22.9)*	80.3 *(78.9;81.5)*	41.4 *(40.0;43.1)*
	SP	60.4 *(59.9;60.8)*	72.1 *(71.6;72.5)*	93.2 *(92.7;93.5)*	66.8 *(66.4;67.2)*	74.1 *(73.7;74.5)*
	Gmean	58.7 *(57.7;59.7)*	29.0 *(27.4;30.5)*	35.2 *(33.4;36.9)*	72.3 *(71.5;73.1)*	50.6 *(49.3;52.1)*

Based on the previous tables (VI, VII, VIII), it is possible to conclude that in some testing situations (Table VIII) the Bayesian global model presents a better performance than the other models. However, GRACE tool showed higher discrimination capability when applied patients of Santa Cruz dataset.

These results demonstrate that the proposed combination scheme should be complemented with the adjustment of its parameters (optimization procedure) in order to improve its performance.

F. Optimization

The proposed fusion methodology can be adjusted to a specific population. If a dataset is available, an optimization can be performed improving the behavior of the global Bayesian model. Table IX presents the optimization results, obtained through a genetic algorithm approach.

TABLE IX
PERFORMANCES COMPARISON

		Santa Cruz 30 days/D/MI		Santa Cruz 30 days/D		Santo André 30 days/D	
		ByG	ByG AO	ByG	ByG AO	ByG	ByG AO
Original	SE	60.6	72.7	61.5	76.9	80.0	80.0
	SP	67.0	69.1	65.7	70.7	67.0	82.9
	G_{mean}	63.4	70.9	63.5	73.7	73.2	81.5
	AUC	0.635	0.7	0.625	0.725	0.725	0.8
Boot Samples n=1000	SE	60.6 *(60.1;61.3)*	72.9 *(72.4;73.4)*	61.6 *(60.7;62.5)*	77.3 *(76.5;78.0)*	80.3 *(78.9;81.5)*	79.8 *(78.6;81.0)*
	SP	67.0 *(66.9;67.2)*	69.1 *(69.0;69.2)*	65.8 *(65.6;65.9)*	70.6 *(70.5,70.8)*	66.8 *(66.4;67.2)*	83.8 *(83.3;84.2)*
	G_{mean}	63.6 *(63.3;63.9)*	70.9 *(70.6;71.1)*	63.1 *(62.7;63.6)*	73.6 *(73.3;74.0)*	72.3 *(71.5;73.1)*	80.9 *(80.0;81.6)*

SE: Sensitivity; **SP**: Specificity; **D:** Death; **MI:** Myocardial Infarction; ***(;)****=95% Confidence Interval;* **ByG** – Bayesian Global Model; **ByG AO** – Bayesian Global Model After Optimization.

It is possible to conclude that genetic algorithms' optimization improved the performance of the Bayesian global model. The optimization was performed in the neighborhood of the initial values, although this restriction may reduce the efficiency of the optimization algorithm, it assures that the optimization procedure does not ignore the knowledge provided by the original risk assessment tools.

G. Missing Risk Factors

The ability of the different classifiers to deal with missing risk factors was assessed through the comparison of the Bayesian approach (before and after the optimization procedure) with the voting model.

Replacement of missing risk factors in voting model was done according to the variables' type, such as: *i*) binary variables were replaced successively by values 0 and 1; *ii*) as Killip level is ordinal, it was replaced sequentially by values 1, 2 and 3; *iii*) a single imputation method based on the mean value was applied to the remaining variables that are continuous. Three different situations were evaluated: *i*) one missing risk factor; *ii*) two missing risk factors; *iii*) three missing risk factors.

TABLE X
MISSING RISK FACTORS - **SANTA CRUZ, (DEATH/MI)**

	Parameter	Bayesian	Bayesian After Opt.	Voting
SE	mean	57.1 *(55.3;58.8)*	65.4 *(62.9;67.7)*	47.8 *(43.9;52.6)*
	std. dev.	5.1	7.1	11.5
	range	[45.4;72.2]	[48.5;75.7]	[27.3;72.7]
SP	mean	65.5 *(63.1;67.9)*	65.9 *(63.4;68.4)*	74.7 *(71.3;78.2)*
	std. dev.	7.1	7.4	10.3
	range	[47.3;74.4]	[48.0;75.8]	[46.9; 94.8]
G_{mean}	mean	61.0 *(60.2;61.8)*	65.2 *(64.2;66.0)*	59.0 *(57.5;60.6)*
	std. dev.	2.3	2.8	4.7
	range	[57.8; 67.6]	[60.3; 71.2]	[50.5; 69.0]

TABLE XI
MISSING RISK FACTORS - **SANTA CRUZ, (DEATH)**

	Parameter	Bayesian	Bayesian After Opt.	Voting
SE	mean	60.0 *(58.1;61.9)*	63.0 *(61.0;65.1)*	49.6 *(45.2;54.1)*
	std. dev.	5.6	7.0	13.2
	range	[46.1;76.9]	[46.2;84.6]	[23.0;76.9]
SP	mean	64.6 *(62.2,66.9)*	68.8 *(66.8,70.8)*	74.3 *(70.9,77.8)*
	std. dev.	6.9	5.8	10.2
	range	[46.5, 72.9]	[55.0, 78.0]	[46.0, 93.7]
G_{mean}	mean	62.0 *(61.0;63.0)*	65.2 *(63.5;66.8)*	59.5 *(57.5;61.6)*
	std. dev.	2.9	4.9	6.1
	range	[57.9;66.8]	[58.7;75.1]	[46.2;70.6]

TABLE XII
MISSING RISK FACTORS - **SANTO ANDRÉ, (DEATH/MI)**

	Parameter	Bayesian	Bayesian After Opt.	Voting
SE	mean	70.8 *(66.4;75.1)*	75.1 *(71.1;79.1)*	45.4 *(38.4;52.3)*
	std. dev.	12.9	11.9	20.8
	range	[20;80]	[20;80]	[20;100]
SP	mean	65.5 *(64.4;66.6)*	79.4 *(78.1;80.7)*	73.5 *(70.9;76.1)*
	std. dev.	3.3	3.9	7.8
	range	[61.7;81.9]	[75.5;81.9]	[58.5;88.9]
G_{mean}	mean	62.1 *(61.0;63.0)*	65.2 *(63.5;66.8)*	59.5 *(57.5;61.6)*
	std. dev.	2.9	4.9	6.1
	range	[57.9;66.8]	[58.7;75.1]	[46.2;70.6]

It is possible to conclude that in the majority of the test cases the global Bayesian model after optimization presents the best performance (highest sensitivity/highest specificity). However, in some situations (Santa Cruz Dataset) the voting model presented the highest specificity's value. This lack of performance of the Bayesian global model in some testing situations must be further investigated.

IV. CONCLUSIONS

This work addressed the combination (fusion) of CVD risk assessment tools. As referred, the combination of these individual risk assessment tools can overcome the respective weaknesses, namely: *i*) consider simultaneously the

information provided by the selected individual current risk assessment tools, *ii*) increase the number of risk factors to compute the risk, *iii*) improve the capability to deal with missing risk factors, *iv*) allow the incorporation of additional clinical knowledge, *v*) assure the clinical interpretability of the respective parameters. Besides, it eliminates the need of a consensus on the best model to use in the clinical practice. Finally, this global model can be easily adjusted for a given population.

The obtained results are very promising, suggesting the potential of the Bayesian approach to fuse current risk assessment tools in a clinical practice context.

Future work will further investigate the capability of this combination strategy to deal with missing information as well as the incorporation of additional clinical knowledge. Validation considering a significant number of patients as well as its application to other populations will give additional significance to the developed strategy.

V. REFERENCES

[1] World Health Organization, "Cardiovascular Diseases (CVDs)", fact sheet n°317. http://www.who.int/mediacentre/factsheets/fs317/en/index.html. (accessed December 2010)

[2] European Heart Network, "Healthy Hearts for All", Annual Report 2009.Available:http://www.ehnheart.org/publications/annual-reports.html (accessed March 2011).

[3] European Heart Network, "Healthy Hearts for All", Annual Report 2008.Available:http://www.ehnheart.org/publications/annual-reports.html (accessed March 2011).

[4] Commission of the European Communities, "Confronting demographic change: a new solidarity between the generations - Green paper". Available: http://eur-lex.europa.eu/LexUriServ (Accessed in December 2010).

[5] Bertrand, M. et al. "Management of acute coronary syndromes in patients presenting without persistent ST-segment elevation", European Heart Journal, Vol. 23, 1809–1840, 2002.

[6] Graham, I. et. al., "Guidelines on preventing cardiovascular disease in clinical practice: executive summary", European Heart Journal, Vol.28, 2375 – 2414, 2007.

[7] S. Paredes, T. Rocha, P. Carvalho, J. Henriques, M. Harris, J. Morais, "Long Term Cardiovascular Risk Models' Combination", Computer Methods and Programs in Biomedicine Journal, 2011.

[8] Tang. E, et. al., "Global Registry of Acute Coronary Events(GRACE) hospital discharge risk scores accurately predicts long term mortality post-acute coronary syndrome", AHJ, Vol. 154, pp. 29-35 2007.

[9] Antman, E. et. al., "The TIMI risk score for Unstable Angina / Non-St Elevation MI – A method for Prognostication and Therapeutic Decision Making", Journal of American Medical Association ,Vol. 284, pp. 835-842, 2000.

[10] Boersma E., K. Pieper, E. Steyerberg, "Predictors of outcome in patients with acute coronary syndromes without persistent ST-segment elevation. Results from an international trial of 9461 patients", Circulation 101;2557–2567, 2000.

[11] Kotsiantis, S, "Supervised Machine Learning: A Review of Classification Techniques", Informatica Vol.31, 249-268, 2007.

[12] Zheng, "A comparative study of semi-naïve Bayes methods in classification learning." Proceedings of the 4th Australasian Data Mining Conference, (pp. pp. 141-156.).

[13] Friedman N., Geiger D., Goldszmidt M., "Bayesian network classifiers", Machine Learning, Vol.29, 131-163, 1997.

[14] Tsymbal A., S. Puuronen and D. Patterson, "Ensemble feature selection with the simple Bayesian classification", in Information Fusion, Vol.4, Issue 2, pp. 87-100 (Elsevier, 2003).

[15] Yang. (2009). Discretization for naïve- Bayes learning managing discretization bias and variance. Machine Learning, Vol. 74, pp. 39-74.

[16] Gonçalves P., Ferreira J., Aguiar C., Seabra-Gomes R., "TIMI, PURSUIT and GRACE risk scores: sustained prognostic value and interaction with revascularization in NSTE-ACS", European Heart Journal, Vol. 26, pp. 865-872, 2005.

[17] Twardy C., Nicholson A., Korb K., McNeil J., "Data Mining cardiovascular Bayesian networks, School of Computer Science Software Engineering, Monash Univ., Melbourne, Tec. report, 2004.