Fusion of Risk Assessment Models with application to Coronary Artery Disease Patients

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

Abstract – Several risk score models are available in literature to predict death/myocardial infarction event for coronary artery disease (CAD) patients, within a short period of time. However, the choice of the most adequate model is not straightforward since there might not be a consensus about the best model to use in clinical practice Moreover, individually, these models present some weaknesses, such as the inability to deal with missing information.

This work addresses these problems, proposing a Bayesian classifier strategy enabling the simultaneous use of several models (models' fusion). Thus, a higher number of risk factors can be used in the common model, while it can deal with missing information. The validation of the strategy is carried out through the combination of three current risk score models (GRACE, TIMI, PURSUIT). Results were obtained based on a dataset that comprises 460 consecutive patients admitted to the Cardiology Department of Santa Cruz Hospital, Lisbon, from 1999 to 2001. A comparison with the voting scheme, which considers exclusively the outputs of models to combine (models output combination) is also carried out. The proposed Bayesian approach had very satisfactory results, confirming the potential of its application to the clinical practice.

I. INTRODUCTION

Each year, cardiovascular disease (CVD) causes over 4.3 million deaths in Europe and almost 2.0 million in the European Union. In fact, CVD is the main cause of illness and death in Europe being responsible for 23% of the total disease burden [1]. Moreover, CVD alone represents approximately €192 billion /year to health costs in the European Union. Coronary Artery Disease (CAD), the cause of approximately half of all CVD deaths, is the single most common cause of death in Europe, and results in direct health costs of €23 billion [2].

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}

£ Instituto Politécnico de Coimbra, Departamento de Física e Matemática, Coimbra. {dml@isec.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}

Therefore, the correct diagnosis and prognostication of cardiovascular disease is a key factor in trying to reduce these social and economic costs.

In this context, risk assessment is an important component of preventive health care. Clinical professionals can evaluate the probability of occurrence of an event, based on past and current exposure to risk factors of the patient [3][4]. Therefore, it will help clinical professionals to identify and adapt treatment to individual patients.

To this aim several risk score models are available in literature. These risk scores differ on input risk factors, disease (coronary artery disease, heart failure,...), events prediction (death, myocardial infarction, unstable angina, hospitalization), the time over which the risk is calculated (days/months/years) prevention type (primary/secondary) and patients' specific condition (for example diabetics).

Of particular relevance are the statistical risk scores GRACE, TIMI (no ST-elevation), TIMI (ST-elevation) and PURSUIT [5][6][7][8]. These models are employed for secondary prevention on CAD patients, in particular for assessing the risk of death/myocardial infarction within a short period of time (days/months). Usually, one of these statistical models is commonly in use as the standard model in different institutions. However, the choice of the best adequate model can be difficult since there might not be a consensus about the best model to use. Therefore, combining existing models (current knowledge) is potentially interesting.

This work intends to compare two different approaches to combine different risk assessment models: a) Model´s fusion based on Bayesian classifiers, b) Individual Models Output Combination. Models' fusion is implemented based on an approach developed by the authors and detailed in [9]. In a first phase a common representation based on a Naïve-Bayes classifier is applied to each individual risk score model. Then a proper combination of the individual model's parameters is implemented, enabling the integration of the information provided by the individual models. On the other hand, individual models output combination is implemented through a well known voting scheme.

The paper is organized as follows: in section II an outline of the methodology of the two combination schemes is presented. In section III the results of the two methods are compared and discussed. Section IV includes the conclusions and the main research paths to be followed in the near future.

II. METHODOLOGY

Combination of classifiers can be done based on different techniques that can be organized in two main categories: a) Models' fusion; b) Individual models output combination.

A. *Models' Fusion – Bayesian Approach*

To implement this strategy, two main phases were required: 1) common representation of all individual models; 2) a combination scheme that exploits the probabilistic nature of Naïve-Bayes inference mechanism.

1) Common Representation of Individual Models

Current individual risk score models are described by different equations/scores/charts [5][6][7][8]. So, in order to ease their combination, all the individual risk score models were represented as Naïve-Bayes classifiers. This classifier was selected since it is efficient, simple and can deal with lack of input information (missing risk factors) [10]. Its inference mechanism is given by:

$$P(C \mid \mathbf{x}) = P(C \mid X_1,...,X_n) = \alpha\, P(C) \prod_{i=1}^{n} P(X_i \mid C) \qquad (1)$$

Usually $\mathbf{x} = \{X_1,...,X_k\}$ is a set of observations (clinical examination, laboratory measurements,…) and C a hypothesis (e.g. risk level is "High"). The term $P(C \mid \mathbf{x})$ 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). $P(X \mid C)$ is a likelihood expressing the probability of the observation X being made if the hypothesis is correct (equivalent to the sensitivity of the clinical examination). α is a normalization constant. This particular Bayesian inference mechanism (Naïve Bayes) assumes that observation instances $\mathbf{x} = \{X_1,...,X_k\}$, are conditionally independent, given the value of hypothesis C [11][12]. However, even if this condition is not verified, naïve Bayes often presents a good performance [13].

Conditional probability tables (CPT) of each individual model were derived based on equations/scores available in literature. The training dataset must contain all the risk factors that belong to the different individual models. As a result, conditional probability tables for the *n* observations were constructed based on (2).

$$P(X_i = x_j \mid C = c_k) = \frac{\sum_{1}^{m} (X_i = x_j \wedge C = c_k)}{\sum_{1}^{m} (C = c_k)} \qquad i = 1,....,n \qquad (2)$$

It is assumed that class C has several categories (mutually exclusive), and that variable c_k denotes the *k* class label of variable C. Furthermore, it is assumed that variable x_j denotes a particular value of the attribute X_i and m is the total number of training instances.

2) Combination Strategy

In this phase the individual models ($M_1,\ldots,M_M$) are combined to generate a global model (M_C).

Based on a training set, each classifier is characterized by a specific CPT, $P(X_i^{\,j} \mid C_i)$, and by its respective prior probabilities of output classes, $P(C_i)$, regarding a specific number of mutually exclusive classes. Where i denotes the individual model and j identifies the risk factor.

The global model is formed at each time by the union of the selected active individual models, i.e., models that verify the selection criteria. Moreover, some risk factors (model inputs) may be considered by more than one model, while other inputs belong only to a particular model. In order to perform this combination some conditions have to be verified: *i*) Individual models have the same number of output levels (e.g. low, high); this ensures that individual models share the same risk assessment goal. *ii*) Among individual models, shared variables' CPT $P(X_i^{\,j} \mid C_i)$ present approximately the same values. Parameters of global model $P(C)$ and $P(X_i \mid C)$ were determined based on a frequency calculation considering the individual models as presented in (2).

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.

3) Optimization

If there is an available dataset characterizing a specific population, 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. A more detailed description of this methodology can be found in [9].

B. *Individual Models Output Combination*

This type of combination includes several methods that can be organized in two main groups: static and dynamic. Static methods apply the same method for the entire data space. Examples of such methods are cross validation majority, voting and weighted voting [14][15]. The dynamic combination considers the characteristics of each specific instance to be classified, in order to define the most proper classifiers to combine/select. Examples are dynamic selection, dynamic voting and dynamic voting with selection [14].

The voting method (static) is explored in this work. 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 [14]. This technique has a serious drawback since it has no mechanism to solve the draw situations. In this work it does not happen, since there are two output classes and an odd number of models (and therefore votes) to combine.

III. RESULTS

A. Santa Cruz Hospital Dataset (Testing 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 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 per cent values. This dataset was used as the testing dataset to validate the two combination strategies.

TABLE I
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%)

Table II presents the occurrence of the D-death and MI-myocardial infarction events during two time periods: one month and one year.

TABLE II
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

B. Training Data Set

With respect to the training process the approach proposed by Twardy *et. al.* [11] was followed here. Continuous variables (age, sbp, hr and cr) 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 $(x_{1i},...,x_{ni})$ for all $1 \le i \le N$: training set N=1000. This training dataset was applied to the models described in the following item, to obtain the respective output class $(x_{1i},...,x_{ni},c_i)$ for all $1 \le i \le N$.

C. Individual Models Description

Table III presents the selected individual risk score models to predict death/MI for CAD patients within a short period.

TABLE III
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 models Performance

As mentioned in II a), 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 in "low/intermediate risk" category. Table IV shows the performance of the three individual models when the testing dataset (real patient data described in III a) is considered for a time period of 30 days. As observed the three models present a very different ability to predict the combined endpoint (Death/Myocardial Infarction).

TABLE IV
PERFORMANCE OF SELECTED INDIVIDUAL MODELS

Model	SE	SP	ACC	AUC
TIMI	33.3	73.5	70.7	0.534
GRACE	60.6	74.9	73.9	0.678
PURSUIT	42.4	74.2	71.9	0.583

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

E. Models' Combination

Training data set was used to calculate the parameters of Bayesian individual models (common representation). Then, these models were combined to generate the Bayesian global model according to the methodology explained in II a).
The global voting model was implemented considering the outputs (0/1) of the three individual models.
Table V shows the performance of both models (fusion and voting).

TABLE V
COMBINATION OF THE THREE INDIVIDUAL CLASSIFIERS

Model	SE	SP	ACC	AUC
Global Model Bayesian	60.6	70.7	66.5	0.651
Global Model Voting	48.5	75.6	73.7	0.594

It is possible to confirm that the Bayesian approach (models' fusion) presents a better sensitivity and discrimination capability than the global model obtained through voting. In contrast, its specificity and accuracy are lower than the voting model. However, the Bayesian approach presents additional advantages, since its inference mechanism is able to cope with missing information, as detailed in [10]. It also allows the direct incorporation of clinical expertise through CPT definition by the physician.

F. Conditional Probability Table Optimization

Additionally, contrarily to the voting scheme, the proposed fusion methodology can be adjusted to a specific population. In this case, if a dataset is available, an optimization can be performed improving the behaviour of the Global Bayesian model. Table VI presents such optimization results, by means of a genetic algorithm approach.

TABLE VI
BAYESIAN MODEL ADJUSTMENT TO THE DATASET

Model	SE	SP	ACC	AUC
Global Bayesian BO	60.6	70.7	66.5	0.651
Global Bayesian AO	60.6	76.1	75	0.682

***BO**: Before Optimization; **AO**: After Optimization*

It is possible to conclude that genetic algorithms' optimization improved the performance of Bayesian global model. This multiobjective optimization attempted to maximize simultaneously the specificity and the sensitivity of the global model. The objective functions to minimize were:

$$f_1 = 1 - \frac{TP}{TP + FN}; \quad f_2 = 1 - \frac{TN}{TN + FP} \tag{3}$$

TP: True Positive; **TN**: True negative; **FN**: False negative; **FP**: False Positive

restricted to:

$$-\beta\, p(X \mid C) \le \delta \le \beta\, p(X \mid C) \tag{4}$$

The value of $\beta = 0.7$ was defined after several experiments.

Although this restriction reduces the efficiency of the optimization algorithm, it assures that optimization procedure does not ignore the knowledge provided by the original models [9].

IV. CONCLUSIONS

This work addressed the combination of risk assessment models for CAD patients. As referred, combination of risk assessment models can avoid the difficulty of choosing a standard model as well as increase the number of considered risk factors. Two different approaches were evaluated: a) model's fusion based on Bayesian classifiers, b) combination of individual models outputs.

Bayesian approach presented a very interesting performance when compared with the voting scheme. The inference mechanism of Bayesian methodology makes it possible to cope with missing information and inherently allows the direct incorporation of other relevant models created from clinical expertise (CPT definition by the physician). Additionally, it can be adjusted for a given population. Preliminary results are very promising, suggesting the potential of the Bayesian approach to combine current models in a clinical practice context. Future work will analyze the capability of the strategy to deal with missing information as well as the incorporation of additional clinical knowledge. Moreover, assuming that some score models are more generally accepted than others, the Bayesian approach should cope with the relative importance of such models.

V. REFERENCES

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

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

[3] 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.

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

[5] 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, January, 2007.

[6] Antman, E. et. al., "The TIMI risk score for Unstable Angina / Non-St Elevation MI – A method for Prognostication and Therapeutic Decision Making", V. 284 nº7, American Medical Association, 2000.

[7] Morrow, D. et. al., "TIMI Risk Score for ST-Elevation Myocardial Infarction: A Convenient Bedside, Clinical Score for Risk Assessment at Presentation: An Intravenous nPA for Treatment of Infarcting Myocardium Early II Trial Substudy", Vol 21, pp.2031-2037, Circulation, American Heart Association, 2000.

[8] 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.

[9] 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.

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

[11] 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.

[12] Twardy C., Nicholson A., Korb K., McNeil J., "Knowledge engineering cardiovascular Bayesian networks from the literature", School of Computer Science & Software Engineering, Monash University, Melbourne, technical report, 2005.

[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] Bauer, E., Kohavi R., "An Empirical Comparison of Voting Classification Algorithms: Bagging, Boosting and Variants", Machine Learning, pp. 1-38, 2008.

[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.