Cardiovascular risk prediction: from classical statistical methods to machine learning approaches

Manuscript title: Cardiovascular risk prediction: from classical statistical methods to machine learning approaches

Running title: Computational approaches in cardiovascular risk prediction

AUTHORS: Michela Sperti[1*], Marta Malavolta[1*], Federica Staunovo Polacco[1], Annalisa Dellavalle[1], Rossella Ruggieri[1], Sara Bergia[1], Alice Fazio[1], Carmine Santoro[1], Marco A. Deriu[1].

[1] *Polito[BIO]MedLab, Department of Mechanical and Aerospace Engineering, Politecnico di Torino, Turin Italy;*

* *These authors contributed equally.*

Corresponding Author (marco.deriu@polito.it).

Abstract

Nowadays, cardiovascular risk prediction scores are commonly used in primary prevention settings. Estimating the cardiovascular individual risk is of crucial importance for effective patient management and optimal therapy identification, with relevant consequences on secondary prevention settings. To reach this goal, a plethora of risk scores have been developed in the past, most of them assuming that each cardiovascular risk factor is linearly dependent on the outcome. However, the overall accuracy of these methods often remains insufficient to solve the problem at hand. In this scenario, machine learning techniques have repeatedly proved successful in improving cardiovascular risk predictions, being able to capture the non-linearity present in the data. In this concern, we present a detailed discussion concerning the application of classical versus machine learning-based cardiovascular risk scores in the clinical setting. This review aims to give an overview of the current risk scores based on classical statistical approaches and machine learning techniques applied to predict the risk of several cardiovascular diseases, comparing them, discussing their similarities and differences, and highlighting their main drawbacks to aid the physician having a more critical understanding of these tools.

Keywords: Risk Prediction, Machine Learning, Traditional Risk, Artificial Intelligence, Cardiovascular Diseases.

Introduction

Cardiovascular diseases (CVDs) are a group of disorders and conditions that affect the structures and/or the function of the heart and blood vessels. They are the world's leading cause of mortality. Due to the enlarged globalization and urbanization, it is noticed that the risk factors for CVDs are rapidly increasing. About 17.7 million people die every year due to CVDs and this number is estimated to be 31% of all global deaths. Nearly 75% of CVD deaths occur in low-income middle-income countries. The main risk factors can be classified into two categories: non-modifiable risk and modifiable risk factors; only the latter can be modified with a healthier lifestyle with the aim of practising physical activity, quitting smoking and focusing on treatments to lower blood pressure and change cholesterol levels[1]. The above-mentioned risks are shown in **Errore. L'origine riferimento non è stata trovata.**. The most relevant CVDs affecting the *heart* involve atrial fibrillation (AF), myocardial infarction (MI), valvular heart disease, cardiomyopathies, and hypertensive heart disease, which are related to hypertension. The most important CVDs involving the *blood vessels*, which are known as vascular diseases, include aortic aneurysm, atherosclerosis, and diseases that limit blood affluence to the brain (stroke) or the legs and the arms.

In the 1950s, given the enormous incidence of CVDs, there was a need to develop predictive models that could predict the onset of fatal events and be able to apply the most appropriate therapy. The risk assessment and prevention at the primary stage is quicker to adopt and implement than at the secondary stage of prevention and can potentially lead to a reduced long-term cost upon the sanitary system. However, in clinical practice, it is easier to classify a patient which presented a previous event (secondary prevention), and who is at high risk rather than globally evaluate healthy patients based on identifiable risk factors (primary prevention). In this scenario, it is of crucial importance to develop computational methods able to detect as early as possible the presence of a disease and, more than that, the probability of a patient to develop a certain pathological condition. The prevention of CVD at the primary stage focuses on the detection of high-risk factors that are the major causes of CVD. The key query to be resolved is how to define different risk levels and assess the CVD risk.

The above-mentioned predictive models determine the probability of an individual experiencing a CV event within a predefined period by assessing the entire risk factors profile. Therefore, it is important to choose the most appropriate predictive model to avoid

misdiagnosis and expose patients to unnecessary treatments which would possibly lead to side effects. The so-called global cardiovascular risk profile is therefore an indicator that, given a set of risk factors, uses them to evaluate the probability of experiencing an acute cardiovascular event. The idea of identifying people at risk based on factors considered individually has been abandoned, but their overall assessment is taken into consideration. For this reason, the absolute global risk is not the simple sum of the risk due to individual factors.

To date, with the rapid development and widespread adoption of artificial intelligence (AI), machine learning (ML), and deep learning (DL) techniques, they have become popular methodologies also in the healthcare domain and have repeatedly proven successful in many disease risk prediction scenarios[2–4]. The objective of employing ML to predict the cardiovascular risk in the population is the improvement of current risk scores performances and the exploitation of the hidden information present in the big amount of data available. Traditional methods cannot manage large datasets, since they are limited to a certain number of variables[5].

AI is the computer science subfield involved in systems' creation that performs tasks that would require human intelligence. It also includes ML and DL. ML focuses on approaches that allow machines to learn from data without being explicitly programmed; it tends to minimize errors and maximize the probability that their predictions are reliable. The three main ML classes are supervised learning, unsupervised learning, and reinforcement learning. DL incorporates computational models and algorithms that mimic biological neural networks in the brain[6].

ML methods may be much more valuable than traditional methods to predict the probability that an event will occur without being associated with specific risk factors. It is not always possible to make a diagnosis with only the data of standard clinical trials; ML helps to make a diagnosis quickly without wasting time for additional research and at the same time serves to expand and increase the efficiency of treatments of patients with CVD. Thanks to the evolution of technology, patients can monitor their state of health through mobile devices, and in case they can realize in real-time a worsening of their condition. The main objective of ML is to incorporate data from different sources (clinical measurements, biological data, experimental results, environmental information).

The present review aims at summarizing current risk scores based on classical statistical approaches and ML techniques and discussing their main clinical applications in the context of cardiovascular risk prediction (Figure 2).

1 Classical Statistical Methods

In the following sections, a detailed description of the most famous risk scores based on traditional statistical methods is provided with the most relevant clinical applications: Framingham Risk Score (FRS), Systematic COronary Risk Evaluation (SCORE), QRISK, PROspective CArdiovascular Münster (PROCAM), Reynolds Risk Score (RRS) and CUORE.

1.1 Framingham Risk Score (FRS)

The Framingham Heart Study (FHS) is the longest-running cardiovascular epidemiological cohort study in the USA, which established that age, sex, lifestyle, hypertension, total cholesterol, high-density lipoprotein (HDL) cholesterol, systolic blood pressure, antihypertensive medication use, dyslipidemia, smoking, and diabetes are risk factors for CVDs, and urged the development of efficient treatments for these conditions. The result was a greatly diminished incidence of CVDs[7]. The FHS defines CVD as a composite of coronary heart diseases (CHD: angina, coronary death, coronary insufficiency, MI), cerebrovascular events, heart failure, and peripheral artery disease[8]. The FRS is a multivariable risk function used to point out the possibility of developing CVDs. FRS predicts risks of CVD by using Cox proportional risks regression[9], which is a technique that allows examining the correlation between the incidence of a clinical outcome (such as MI) and a risk factor (such as smoking). The risk factors taken into account were age, systolic blood pressure, antihypertensive medication use, total cholesterol, HDL cholesterol, smoking, and diabetes. Other variables (diastolic blood pressure, body mass index, triglycerides) also were considered, but they were not statistically relevant. Employing Cox proportional risks regression, a CVD risk prediction function was estimated, to evaluate the 10-year absolute CVD risk. All the continuous variables were naturally logarithmically converted to improve the discrimination and calibration of the models. The developed function is the following:

$$FRS = 1 - S^0(t)^{\exp\left(\sum_{i=1}^{p} b_i X_i - \sum_{i=1}^{p} b_i \bar{X}_i\right)} \quad [1]$$

where $S_0(t)$ is the baseline survival at follow-up time *t* (here *t=10 years*), b_i the estimated regression coefficient (log hazard ratio, the values of which are shown in Table I), X_i the

log-transformed value of the ith risk factor, $\bar{X}_i$ the corresponding mean, and p is the number of risk factors. However, this study shows some limitations: FRS is not applicable in Europe because it is based on parameters referred to the North-American population. For this reason, another method for risk calculation was developed in the past (namely SCORE algorithm), which is applicable in Europe. Furthermore, FRS doesn't consider some important risk factors, like CVD family history and lifestyle.

Table I. *Regression Coefficients and Hazard Ratios12. SBP = systolic blood pressure.*

Variable	*Women*		*Men*	
	β	**Hazard Ratio**	**β**	**Hazard Ratio**
Log of age	2,33	10,27	3,06	21,35
Log of total cholesterol	1,21	3,35	1,12	3,08
Log of HDL cholesterol	-0,71	0,49	-0,93	0,39
Log of SBP if not treated	2,76	15,82	1,93	6,91
Log of SBP if treated	2,82	16,82	2,00	7,38
Smoking	0,53	1,70	0,65	1,92
Diabetes	0,69	2,00	0,57	1,78

The FHS was helpful to *evaluate the impact of AF* on the risk of death, after correcting for coexisting cardiac diseases and risk factors in a population-based sample, using the pooled logistic regression analysis[10]. The clinical risk factors for AF are advancing age, diabetes, hypertension, congestive heart failure, rheumatic and nonrheumatic valve disease, and MI. The pooled logistic regression analysis is a Cox regression analysis that depends on time, so it was possible to see how the risk factors changed over time with the clinical variables. About two subjects without AF were matched to each AF subject by gender, age, and date of diagnosis of AF. The differences in survival between AF and matched non-AF participants were analysed to estimate survival and produce mortality curves. It has been noticed that cardiovascular disease risk factors were more frequent in subjects with AF than subjects without AF. In addition, the mortality of men and women with AF was much higher than for the non-AF subjects. At 10 years of follow-up, in people aged between 55 and 74 years, 57.6% of women with AF had died compared with 20.9% of women without AF, and 61.5% of men with AF had died, whereas 30.0% of men without AF had died. Furthermore, in men with AF and aged between 55 and 64 years, 12.6 years of survival have been estimated as against 18.1 years in men without AF, and in women 12.1 years of survival were estimated compared with 21.3 years in women without AF. So, AF is significantly associated with an increased risk of death.

A recent application of FRS is related to the *COVID-19*. It was noticed that patients admitted to hospitals with existing CVDs, experienced worse outcomes compared to CVD-free patients. In fact, in a sample of 700 hospitalizations for COVID-19, unfavourable levels of all individual CVD risk factors were related to an almost doubled risk of subsequent COVID-19[11]. For people in the highest risk groups, the FRS offered some predictive utility

helpful for the identification of at-risk groups, which now are the first groups to be given the vaccine.

1.2 Systematic COronary Risk Evaluation (SCORE)

The SCORE is a method developed in 2003 from the Framingham Heart Study[12]; this method aims to evaluate the 10-year risk to run into a first fatal cardiovascular event. As the Framingham method was not suitable for the European population, a researchers' group developed the so-called SCORE; it also allows the division between high CV risk countries and low CV risk counties and gives more reliable results in people aged between 40 and 65 years. SCORE allows to calculate (through the use of specific charts) the vascular age of a person, that is to say, the age of the vascular system concerning exposure to certain risk factors, including total cholesterol, systolic blood pressure, gender, age, and smoking status; without making any distinction between diabetic patients or non-diabetic patients.

The vascular age may not coincide with the biological age of the person, thus exceeding the limit for the calculation of cardiovascular risk based only on the age of the person. It also demonstrates that it's not always true that elderly people are more exposed to cardiovascular risks: it depends on how many years of exposure to risk factors they have. The concept of vascular age also improves communication with patients, making them more aware of the risk they run into if they do not change their lifestyle. For example, a 50-year-old woman who suffers from high blood pressure (180 mmHg), has a high level of cholesterol in her blood, and smokes risks as much as a 65-year-old woman who has normal blood pressure (140 mmHg), does not smoke, and has a lower cholesterol level. These charts differ if applied to high CV risk countries (Finland, Russia, Norway, British region, Scotland, Denmark, Sweden, and Germany) or low CV risk countries (Belgium, Italy, France, and Spain)[13].

The most important application using the SCORE is a c*omparison between FRS and SCORE in the Australian population.* A specific study was conducted in Australia in which the results obtained with the FRS method were compared with the SCORE method for both high-risk and low-risk countries, as they have similar endpoints, both consider the same risk factors and variables, and both models predict the 10-year CVD death risk[14]. The participants were 4487 women aged 30-74 years, from Australia and free of cardiovascular events. For each patient, the mortality risk was calculated using the FRS and the two SCORE risk models, and these were compared. The accuracy of the CVD risk score models was evaluated using both discrimination and calibration. During the 10-year follow up there

were 152 deaths, of which 28 were caused by cardiovascular diseases, 3.4% of the sample (**Errore. L'origine riferimento non è stata trovata.**). All two methods underestimated the death risk in women below 50 years of age and overestimated the risk in those over the age of 50. The FRS model and SCORE risk chart for low-risk regions predicted similar risk levels in the age categories, but at the same time, the SCORE high-risk model predicted higher risk levels compared to the other two models.

Another application that involves an updated version of SCORE, is called *SCORE OP*. It was developed in 2015 to predict the risk of fatal CV events specifically for patients over the age of 65 in Germany[15]. The group of people consisted of 1657 participants aged 70 and over who had never had cardiovascular problems before, and during the 5 -year follow-up, 118 cardiovascular deaths occurred due to fatal MI, fatal coronary heart disease, sudden cardiac death, death due to other cardiac diseases such as heart failure. The risk factors considered were age, gender, systolic blood pressure, HDL and total cholesterol, smoking status, diabetes mellitus status, and self-reported use of antihypertensive medication. The results show an overestimation of the risk and specifically, SCORE OP predicted higher risks compared to SCORE in female individuals aged >75 and in male individuals aged > 78 (**Errore. L'origine riferimento non è stata trovata.Errore. L'origine riferimento non è stata trovata.**). Although there has been an overestimation of risk, SCORE OP has given a result more discriminatory than SCORE.

A particular implementation of the SCORE is the *IberScore study* in Spain. The goal of this method is the computation of the cardiovascular risk in young, healthy, and working people[16]. In contrast to the original score method, this model separates women from men in two different categories (it does not consider gender as a mere risk factor). The number of people examined by the study was 624389 workers aged 16 to 65, with no cardiovascular history. Among the risk factors, however, there are smoking status, total cholesterol and HDL, systolic blood pressure (SBP), and in addition blood glucose, obesity, and another predictor that summarised a personal history of dyslipidemia, diabetes, hypertension, and/or adherence to a prescribed heart-healthy diet that SCORE did not consider before. They observed 3741 cardiovascular events during the 10-year follow-up (80.5% non-fatal and 19.5% fatal) and the incidence rate was four times greater for men. The results suggest that under 55 years, SCORE underestimated the risk, while Iberscore showed higher risk predictions. Over 55 years, on the contrary, SCORE overestimated the risk, while Iberscore gave more reliable predictions.

1.3 QRISK

QRISK is a prediction algorithm for CVD used in the United Kingdom according to the National Institute for Health and Care Excellence (NICE)[17]. This model provides an estimate of the 10-year risk of experiencing a CVD event, and it is widely used in the context of primary care. Patients with a QRISK score below 20% are considered to be at low risk of developing CVD, whereas patients with a QRISK score above 20% are considered to be at high risk and therefore should be monitored and possibly initiate supportive therapies. The QRISK model was developed from a research database called QRESEARCH which collects sufficient routine data from the anonymised health records of over 35 million patients derived from general practices using the EMIS clinical computer system (https://www.emishealth.com/). The practices are distributed throughout the UK. Historical records (which also include patients who may have died or left) extend back to 1989 making it one of the richest general practice databases in the world (https://www.qresearch.org/). The research uses two-thirds of the database to derive and model the algorithm and the remaining one-third to control and validate the sample. QRISK uses the following risk factors: age, gender, smoking habit, SBP, body mass index (BMI), the ratio of total serious cholesterol to HDL cholesterol or lipoprotein levels, ethnicity, deprivation measures, family history of CVD, AF, RA, chronic kidney disease, diabetes mellitus and use of at least one blood pressure treatment such as thiazide antihypertensive blocker, calcium channel blocker or angiotensin-converting enzyme inhibitor.

There are several versions in the QRISK model, which differ in some risk factors. For example, QRISK2 includes self-assigned ethnicity, type 2 diabetes, kidney disease, RA, and AF, which are not considered in QRISK1. Furthermore, QRISK2 relates age with other factors such as BMI, family history, or smoking habit[18]. QRISK model uses a logarithmic transformation for the ‘age’ variable and linear terms for the rest of the variables according to the fractional polynomial analysis.

An interesting aspect is represented by the inclusion of deprivation and ethnicity within the risk factors of QRISK. This aspect could help prevent and reduce the widening of health inequalities. This feature distinguishes the QRISK score from other predictive methods such as the Framingham Risk Score and is very useful for correctly estimating cardiovascular risk. For example, a conducted study showed that 9% of the total sample was reclassified from low to high risk or vice versa by using QRISK versus FRS and in this article, it is emphasized that FRS algorithm tends to underestimate cardiovascular risk in disadvantaged

areas and overestimate cardiovascular risk in affluent areas. An external validation study was done by an Oxford University independent team which used a dataset external to the QRESEARCH database showing that QRISK is better calibrated on the population of the United Kingdom compared to other predictive models[19]. QRISK resulted better calibrated than FRS. This is because FRS does not include risk factors such as BMI, family history of cardiovascular disease, antihypertensive treatment, and social deprivation.

In a 2017 study, it is shown a correlation between the percentage of cardiovascular risk recorded by QRISK and the use of lipid-lowering therapy treatments such as statins. Statins are a group of molecules that can lower blood cholesterol levels and consequently reduce the risk of heart attacks and strokes (up to 25%). A higher than 10% QRISK score suggests the use of statin therapies for prevention. The data support this thesis by indicating that the majority of patients treated with statins did not register a QRISK score and the majority of patients at high risk according to the QRISK, were not initiated with statin therapy[20].

Another interesting aspect is the implementation of a model called the *QRISK life model* to estimate the individual risk of cardiovascular disease during life. The algorithm is based on the QRISK2 model and uses data routinely collected by UK general practice. Thanks to this algorithm it is possible to define centile values that can be used as thresholds to classify further high-risk patients. If the analysed patient has a value higher than the 90th centile or lifetime risk of CVD higher than 50%, they will be at high risk during their lifetime[21]. For example, a 54-year-old white woman who does not smoke, obese, coming from an affluent area, with a first degree relative with CHD, SBP of 150 mmHg, and HDL cholesterol ratio of 5.3, has a risk of cardiovascular disease at 10 years of 14%, therefore below the intervention threshold at 10 years of 20%. However, her lifetime risk of cardiovascular disease is 62%, which places her above the 95th centile for life risk. If we compare the same person with better risk factors, namely BMI 28, systolic blood pressure 128 mmHg, total HDL cholesterol ratio 4, they have a risk of 49% throughout life, and risk under 10% at 10 years. With the current risk factors, the expected age for the occurrence of a cardiovascular event is 81 years, but with a reduction in risk factors, it shifts to 95 years.

Overall, patients with a high lifetime risk are more likely to be male, from non-white ethnic groups (particularly South Asia), current smokers, and with a positive family history of CHD compared to the ones with a 10-year high risk. The QRISK life model could be used on an individual level to communicate cardiovascular risk to patients, particularly young patients with positive family histories whose early lifestyle changes or drug interventions

could lead to significant improvements. It remains questionable whether therapies should be based on 10-year risk or lifetime risk. On one hand, early lifestyle interventions may be helpful but on the other hand, medical operations carry risks as soon as they are initiated. Furthermore, the main disadvantage of this algorithm is that the values and results found are only applicable to patients from the UK, since it has been developed on the British population of QRESEARCH database (https://www.qresearch.org/).

1.4 PROspective CArdiovascular Münster (PROCAM)

Among the various models which calculate this risk profile, many of these (e.g., FRS) do not consider many important features, like LDL cholesterol or family history of CHD. This is done by the PROCAM score, based on a study carried out in Germany[22], which involved for its development 5389 men with ages between 35 and 65. By the 10 year follow-up, 325 acute coronary events resulted. To predict the CHD, the "simple" clinical judgment is not sufficient, because it involves too much subjectivity; thus, statistical methods have been implemented by PROCAM. These give a more realistic percentage of risk. In that case, the method consisted of a Cox proportional hazards model[23], which has many advantages, like logistic regression. It used 8 independent risk variables, listed in order of importance: age, LDL cholesterol, smoking (yes or no), HDL cholesterol, SBP, family history of premature MI (yes or no), diabetes mellitus (yes or no), and triglycerides. Of the 57 clinical features measured in the PROCAM study, 8 were used to construct the risk algorithm because they were found to be independently predictive of event risk. These variables, together with the β-coefficients of the Cox model, the hazard ratios, and the 95% confidence intervals, are shown in Table II[22]. The table also shows the mean value of each variable in the total study population.

Table II. *Variables used in the construction of the Cox proportional hazards model. The P-value is <0,001 for all variables.*

Variables	β	Hazard Ratio
Age, y	0,103	1,108
LDL cholesterol, mg/dl	0,013	1,013
Smoking, %	0,658	1,931
HDL cholesterol, mg/dl	-0,032	0,968
SBP, mmHg	0,010	1,010
Family History of MI, %	0,382	1,465

Diabetes mellitus, %	0,399	1,491
Triglycerides, mg/dl	0,317	1,373

With the Cox model, only the relative risk can be calculated. Relative risk is the ratio of the proportion of patients who present the event in the risk group and the proportion of patients experiencing the event in the group not exposed to the risk factor. Consequently, to transform the relative risk into an absolute one, the Kaplan-Meier statistic was used, which permits the construction of the survival curves (i.e., the graph of the relationship between the cumulative probability of surviving and the observation time) and to measure the observed risk. To increase the easiness and usefulness of the risk algorithm in clinical practice, a simple scoring scheme was created: each risk factor was divided into classes, as can be seen in Table III[22].

Table III. *Coefficients of PROCAM Scoring Scheme.*

Age, y		**LDL, mg/dL**		**SBP, mmHg**		**Triglycerides, mg/dL**	
Range	*Score*	*Range*	*Score*	*Range*	*Score*	*Range*	*Score*
35-39	0	<100	0	<120	0	<100	0
40-44	6	100-129	5	120-129	2	100-149	2
45-49	11	130-159	10	130-139	3	150-199	3
50-54	16	160-189	14	140-159	5	≥200	4
55-59	21	≥190	20	≥160	8		
60-65	26						
Diabetes mellitus		**Smoker**		**HDL, mg/dL**		**MI in family history**	
Range	*Score*	*Range*	*Score*	*Range*	*Score*	*Range*	*Score*
No	0	No	0	<35	11	No	0
Yes	6	Yes	8	35-44	8	Yes	4
				45-54	5		
				≥55	0		

Then the global risk was calculated by use of the Cox model in combination with the survival curves. After that, a regression equation was calculated between the logarithm of global risk and the classes of each risk factor. The coefficients obtained in this way were then standardized and rounded to the integer. Table IV shows the risk of a coronary event associated with each score, calculated using the full PROCAM algorithm[22].

Table IV. *10 years risk of acute coronary events associated with each PROCAM score*[22].

Points	Risk (%)	Points	Risk (%)	Points	Risk (%)	Points	Risk (%)
≤20	<1.0	36	4.2	25	1.6	51	16.8
21	1.1	37	4.8	26	1.7	52	17.5
22	1.2	38	5.1	27	1.8	53	19.6
23	1.3	39	5.7	28	1.9	54	21.7
24	1.4	40	6.1	29	2.3	55	22.2
30	2.4	41	7.0	45	10.2	56	23.8
31	2.8	42	7.4	46	10.5	57	25.1
32	2.9	43	8.0	47	10.7	58	28.0
33	3.3	44	8.8	48	12.8	59	29.4
34	3.5	45	10.2	49	13.2	≥60	≥30.0
35	4.0	46	10.5	50	15.5		

The accuracy of the simple point scoring system was similar to coronary event prediction using continuous variables, with the important difference that the use of the scoring scheme gives more accurate targeting of lipid-lowering therapy and strong cost savings. Indeed, considering the so-called ROC (receiver-operating characteristics) curve, that draws the ratio between true positive rate and false positive rate, whose ideal value is 1 (100%), it shows an area under the curve of 82.4%, compared with 82.9% for the Cox model (with continuous variables).

Trying to use the FRS to determine the relative risk of an acute coronary event and evaluating it through ROC curve analysis, the area under the curve was 77.8%, so it was significantly less than that obtained with the PROCAM models. There are, however, important differences between the PROCAM and the FHS data sets. For example, FRS does not include information on family history of CHD, triglycerides, or LDL cholesterol. The Framingham risk prediction function systematically overestimated risk in the PROCAM cohort. The PROCAM score performs better (than FRS) with the data set in which it was developed and optimized. This is a classical characteristic of the classical statistical methods (which will be improved by using ML techniques, described later). In fact, more real validation of both the Framingham and PROCAM scoring systems would need their application to a third independent data set.

An example of PROCAM risk score application is represented by a systematic review of published research in which the score was used together with FRS to test if bariatric surgery reduces long-term CV events[24]. CV risk models were applied in a validation cohort previously published. The validation cohort was a historical cohort of 197 patients who underwent Roux-en-Y gastric bypass and 163 control patients (with nonoperative treatment), identified through the Rochester Epidemiology Project. FRS and PROCAM risk scores were applied to calculate 10-year CV risk. In the validation cohort, the absolute 10-year FRS for CV events was lower at follow-up in the bariatric surgery group (7.0% to 3.5%, $p<0.001$) compared with controls (7.1% to 6.5%, $p=0.13$), with an absolute difference in risk reduction between the two groups of 3% ($p<0.001$). PROCAM risk in the bariatric surgery group decreased from 4.1% to 2.0% ($p<0.001$), whereas the control group exhibited only a modest decrease (4.4% to 3.8%, $p=0.08$). Comparing the average data from the validation study with the Roux-en-Y group, it was noticed that the tendency of CV risk was similar. Thus, bariatric surgery offers significant reductions in predicted CV risk. Furthermore, some studies demonstrated a mortality benefit after bariatric surgery. This is because obesity not only increases cardiovascular risk, but also the risk of developing respiratory diseases, certain cancers, and more. In a study with a 7.1-year mean follow-up period, long-term mortality from any cause was decreased by 40% compared with controls, and mortality from CV disease was decreased by 56%[25]. In conclusion, risk models could be applied to bariatric surgery patients to inform them of their future risk for CV events and all-cause death.

1.5 Reynolds Risk Score (RRS)

RRS for the prevention of CVD has been developed in 2007 for women and men. The difference between RRS and the elder Risk Scores such as FRS concerns the features involved in evaluating the risk classification of the various patients. For example, while in FRS the main features used for the classification are age, smoking, blood pressure, antihypertensive treatment, diabetes mellitus, and total and high-density lipoprotein cholesterol, RRS adds some features as high-sensitivity C-reactive protein, family history of premature MI, and haemoglobin A1c (among diabetics only). It does not predict CVD in particular but it calculates a composite CVD outcome that may refer to MI incident, ischemic stroke, coronary occlusion, and death due to CVD in general[26]. A detailed example of statistical analysis of population, creation, and validation of RRS algorithm has been described by the Women's Health Initiative (WHI). The algorithm has been validated on a

cohort of women of various ethnicity from all around the world. The population included 1722 between black, Hispanic, Asian, and white women, and also a subcohort of 1994 women was created following the same stratification. The WHI registered all patients' data such as age, sex, and smoking habits, also they arranged to make blood analysis on all the patients to register their blood parameters. Patients have been followed for 10 years during the evolution of their cardiovascular clinical picture to register any episode, and this made it possible to build and validate the RRS algorithm by comparing the algorithm's outcomes to real clinical results. At the end of the 10 years, 1722 cardiovascular cases occurred, and 752 of them have been classified as MIs, while 754 were ischemic stroke and 216 were death caused by cardiovascular disease. RRS is based on a mathematical relationship used to weigh the various parameters that can be involved in the manifestation of CVD. Every parameter has a different weight, in fact, in the determination of the percentual of risk, each of them is multiplicated for a coefficient to weigh its incidence in CVD. The following relationship describes the structure and the coefficient of the various features in developing RRS:

$$10 - year\ CVD\ risk\ (\%) = [1 - 0{,}98634(e^{B - 22{,}325})]\ 100\% \quad [2]$$

where

$$B = 0{,}0799\ age + 3{,}137\ ln(SBP) + 0{,}180\ ln(high\ sensitivity\ C - reactive\ protein) + 1{,}382\ ln(total\ cholesterol) - 1{,}172\ ln(high - density\ lipoprotein\ cholesterol) + 0{,}134\ hemoglobin\ A1c(\%)\ (if\ diabetic) + 0{,}818\ (in\ presence\ of\ smoking\ habits) + 0{,}438\ (if\ family\ history\ of\ premature\ MIs).$$

This formula shows the weight of the various features on cardiovascular risk. As explained some features are considered only in presence of certain factors, for example, to calculate RRS, haemoglobin is required only in diabetic patients, and also some events involving the patient's lifestyle are included in the mathematical relationship such as the presence of smoking habits and family history of premature MI. The value of coefficients states that some features are heavier determining the personal RRS, the preponderant coefficients are that of the SBP, indicating a strong relationship between these features and CVD risk. Afterwards, the coefficient of total cholesterol and smoking habits considerably high, confirming that also patient's lifestyle has a great weight on cardiovascular health. The process of RRS validation has been analysed by WHS using a larger population of women[27]. The entire population was divided into three parts: two-thirds were used as a derivation data

set, while one-third had the role of the validation set. The first part of the process consisted of the determination of the model using the Cox reduction method using two-thirds of collected data, and the second part is centred on the validation of the model using one-third of the total population[23]. The extendibility of RRS is based on the determination on the Bayes Information Criterion (BIC), a sort of regularization in which lower values indicate a better fitting and the increasing of variable induces a penalty. BIC usually describes the goodness of the fitting of a statistical method to sample data for given values of unknown parameters. The coefficients for RRS are shown in Table V.

Table V. *This table shows the coefficient β and standards errors (SE) for RRS. The P-value is <0,001 for all variables. Abbreviations: HbA_{1c} = hemoglobin A1c; SBP = Systolic blood pressure; HDL-C = high-density lipoprotein cholesterol; hsCRP = high- sensitivity C-reactive protein; MI = myocardial infarction*[27].

Variables	β (SE)	$\chi 2$
Age	0,080 (0,006)	193,5
HbA_{1c}% with diabetes	0,134 (0,017)	62,3
Current smoking	0,818 (0,109)	55,9
ln(SBP)	3,137 (0,423)	55,1
ln(HDL-C)	-1,172 (0,172)	46,2
ln(Total Cholesterol)	1,382 (0,239)	33,3
ln(hsCRP)	0,180 (0,043)	17,5
Parental history of MI < Age 60 y	0,438 (0,118)	13,7
Age	0,080 (0,006)	193,5

Coefficient β is also called “regression coefficient” and is used to calculate how well the curve that predicts the outcomes is near to the sample that we analyse. The coefficient $\chi 2$ indicates the “standard error” and estimates the accuracy of the algorithm in predicting outcomes. Finally, p-value also called “probability value”, evaluates the probability to obtain outcomes similar to results obtained during the testing of the algorithm. In other words, p-value helps to understand if the difference between obtained and supposed outcomes is due to fortuity, low levels of p-value indicate that can be a very significant difference that cannot be attributed to randomness in sampling. Moreover, the BIC coefficient for RRS reaches a very low value underlining a good predicting ability with clinical simplifications. RRS could be confronted with the most used risk scores, such as

FRS or ATP-III prediction model. FRS and ATP-III present very different values of BIC if reported to RRS that shows a loss of predicting ability with clinical simplification.

RRS has been confronted with other algorithms by WHI in a scientific essay led by Dr. Nancy Cook[26]. The group of researchers analysed the predictive curve of the main used risk scores, the results have been organized in tabs where the main features observed are: calibration and accuracy. The first is the ability of the curve to fit the parameters and discriminate the various subjects basing on their parameters, and the second is directly linked to the ability to predict in the right way the outcome, in other words, it is based on the right correspondence between supposed and obtained outcome. The average risk obtained from the different risk scores was 3.8%, 4.6%, and 10,9% respectively for ATP-III, RRS, and FRS, highlighting an intermediate value of RRS between the two most used risk scores. In general, RRS seems to be more calibrated in the determination of CVD since the other way to calculate CVD risk score tends to overestimate or underestimate the risk according to WHI studies. After 10 years from the beginning of the screening, WHI collected all data from patients controlling the appearance of CVDs among the samples. Data based on the course of cardiological clinic painting were compared with the prediction curves typical of the various risk score to understand the accuracy and the calibration of the most used risk score and value the most suitable on a particular sample.

RRS according to the studies of Cook, Nancy and co-workers, provides a very effective way to calculate the patient's risk to occur in a cardiovascular event in the following 10 years since the determination of the risk score[26]. Using RRS instead of elder risk scores (such as FRS, SCORE, and ATP-III) guarantees a better calibration and accuracy of the prediction curve, and this important improvement makes it possible to create a more accurate diagnosis minimizing the risk to underestimate or overestimate the outcomes. Moreover, RRS includes the quantity of C-reactive protein, haemoglobin A1c and cardiovascular history among the relevant features used to evaluate the risk score. The quantity of reactive C-protein increases during inflammatory events, and in this way RRS becomes useful in considering systemic inflammatory processes in the prediction phase[28]. High levels of reactive C-protein are strongly linked to CVD as ischemic stroke and ictus since blood vessels seem to be very sensitive to PCR and a great amount of it can cause damns or lacerations of arterial and venous tissues. The addition of PCR and CVD history among risk factors makes RRS an effective prediction method in patients with inflammatory diseases and helps physicians evaluating the possibility to begin statin therapy. Statins are a group

of drugs used to reduce the number of lipids in the organism, including cholesterol. Patients with type 2 diabetes mellitus (DM) are at an increased risk for atherosclerosis and the great quantity of glucose in a patient's blood can compromise the endurance of the blood vessels. However, RRS can monitor the level of PCR in blood, which is one of the most important markers to predict the onset of type 2 DM, and metabolic syndromes are often preceded by weak but numerous inflammatory events[28]. Hemoglobin A1c is considered only in diabetics patients, so diabetes results to be another important marker for the prediction of CVD. So this method can be used in presence of patients with diseases as diabetes or systemic inflammatory events and provides an accurate risk prediction involving factors brought by other diseases[28]. Despite the great performances of RRS, Buring & Cook highlights the possibility to generate an even more accurate and calibrated risk score using the Cox reduction method and minimizing BIC[23,27].

1.6 CUORE

In 1998 the CUORE project - Epidemiology and Prevention of Cerebral and CVDs - was launched, under the coordination of the *Istituto Superiore di Sanità* (ISS), to estimate the probability of a first cardiovascular event occurs within 10 years of assessment[29]. The idea behind this is to monitor CV risk factors, lifestyles, and diseases, to picture the situation of the population's health and to identify areas with a priority of intervention, and to develop preventive actions[30]. It is known that the Framingham method[9], which resulted very suitable for the American population, cannot fit with the same accuracy populations registering a lower rate of CHD, such as the European ones[30]. Indeed, in this case, the Framingham method-based risk charts tend to overestimate the absolute risk. More reliable would be applying an algorithm of risk validated on the population of reference. In this scenario, the CUORE risk score, developed during the CUORE project, answers the need for a better fitting tool of evaluation on the Italian population, and others with low CHD rate[31]. It is based on the mathematical equation [3] to do an appraisal of the probability of a first major CV event[32].

$$CUORE = 1 - S(t)^{e^{[\sum_{i=1}^{p} \beta_i X_i - G(\mu)]}} \quad [3]$$

where $S(t)$ is the survival at 10 years, β_i are the risk factor coefficients, performed using the Cox proportional hazards model, separately for men and women[23] (Table VI), x_i is the i^{th} factor of risk, $G(\mu)$ is the linear combination of the mean of the factors, and p is the number of the considered factors of risk.

The project's official website http://www.cuore.iss.it/ offers the possibility to consult collected data, calculated risk, and statistics, to have a global picture of the population's health situation. Moreover, it offers different options to calculate and evaluate the CV risk.

Table VI. *Cox proportional hazard model for the CV assessment in 10 years*[31].

Variable	*Women*		*Men*	
	β	**Hazard Ratio**	**β**	**Hazard Ratio**
Age	0,07	2,05	0,08	2,12
SBP (mmhH)	0,01	1,29	0,02	1,39
Total Cholesterol (mg/dl)	0,01	1,32	0,01	1,24
Hdl Cholesterol (mg/dl)	-0,01	1,17	-0,02	1,28
Smoking Status	0,48	1,62	0,79	2,20
Diabetes	0,40	1,49	0,56	1,74
Antihypertensive Treatment	0,34	1,41	0,47	1,60

The first and most precise one is the calculation of an Individual Score, obtained either from the downloadable executive software, practical and easy to use in a medical environment, or filling in the online survey. This score tends to be more precise since it considers continuous values for some factors of risk and includes more variables. There are eight factors of risk: sex, age, diabetes, smoking habit, SBP, total and HDL cholesterol, and hypertensive treatment. It presents, though, some limitations: measurements must follow a standard methodology; age must be between 35 and 69 for both sexes and subjects must be without previous CV events; it is not suitable for pregnant women; it cannot be applied to outliers of the risk factors: SBP higher than 200 mmHg or lower than 90 mmHg, total cholesterol higher than 320 mg/dl or lower than 130 mg/dl, HDL cholesterol higher than 100mg/dl or lower than 20 mg/dl; clinical exams of glycemia and cholesterol should be not older than three months.

The other possible option is to estimate the risk score with the Cards of Risk. These are absolute global risk classes calculated by categories of risk factors, considering six factors of risk: sex, age, diabetes, smoking habit, SBP, total cholesterol. This score is less precise since it considers fewer factors, therefore it is possible to obtain different results with the two methods. Similar to the previous one, it presents some limitations. All the previous prescriptions are valid, but, in this case, they apply only to people aged 40-69 years old. In any case, it is recommended to periodically assess the CV risk, according to the risk range

of belonging. Furthermore, the project provides periodical surveys on the Italian population of all the Italian territory, to keep the algorithm updated to the actual features of the population, and took part in many projects for CVD prevention, and has an active role to this extent.

Lately, a *comparative study between CUORE and Framingham algorithms* was carried out to study their reliability in an Italian population[32]. There have been considered 996 subjects of both sexes, patients of General Practitioners working in the Province of Rome. R2 value has been used to evaluate the goodness of the regression model. Both the methods aim to estimate the risk of CHD, but they differ in the variable to consider: Framingham considers angina pectoris, recognized and unrecognized MI, coronary insufficiency (unstable angina), and CHD deaths too, while CUORE can be applied to subjects who haven't had any previous cardiovascular event, and the risk factors are obtained from the use of standardized methodologies. Differently from the majority of the published studies, it emerged that FRS is more reliable than CUORE on this specific Italian population (CVD risk is less overestimated by FRS with respect to CUORE). The reason for such results can be maybe found in the fact that all these algorithms tend to fit some populations better than others, and maybe the features of the considered patients differ from the ones from other studies. However, no unambiguous indication of the best algorithm to estimate the individual cardiovascular risk is still available. On the other hand, the CUORE algorithm was revealed to be more objective than the Framingham when it comes to the presence of other medical conditions, such as psoriasis or severe mental illnesses (SMI).

CUORE was also employed in patients affected by rheumatic diseases. Research from 2013[29], aiming to investigate the relation between CV risk and psoriasis features in a real-world setting, highlighted how the Framingham score on the Italian population was not corresponding to the real evidence, while the CUORE resulted closer to it, proving the connection of the two conditions. A few years later, an Italian bicentric study[33] explored the performances of different algorithms in the prediction of cardiovascular events in patients with rheumatoid arthritis (RA), another chronic inflammatory autoimmune disease that electively affects the joints. Two Italian cohorts were used, and the discriminatory ability for CV risk prediction was evaluated by analysing the area under the ROC curves, while the calibration between predicted and observed events were assessed by Hosmer-Lemeshow (HL) tests and sensitivity and specificity calculated for low to intermediate and intermediate to high-risk cut-offs. CV events considered were: heart failure, sudden cardiac death, TIA,

CAD (stable and unstable angina pectoris), MI, PAD, and CVA. In this case, FRS, SCORE, CUORE, Reynold's Risk Score (RRS), and QRISK2 tended to assess the CV risk lower than what was possible to observe in reality in patients with RA. Since this category of patients seems to be more exposed to CV problems, recently the EULAR (European League Against Rheumatism) recommended adding a multiplier factor of 1.5 in the risk algorithms used for patients with RA. It has been tested during the study, but it did not lead to a significant improvement (**Errore. L'origine riferimento non è stata trovata.**). Overall, despite an increased CV risk in case of psoriatic arthritis has been demonstrated, all the five algorithms underestimated the CV risk in this study. In particular, data showed a poor model fit with a significant different distribution of predicted and observed events for CUORE, SCORE and RRS, but not for FRS and QRISK2. Indeed, QRISK2 showed the best discriminative ability and calibration on the considered population.

This score was applied also to *patients affected by severe mental illness (SMIs)* in 2020, which was obtained interesting result[30]. In the last 10 years, many studies highlighted that patients with SMIs show higher rates of metabolic syndrome and its connection with the risk of developing CVDs[30]. These patients are twice times more affected by diabetes, obesity, and metabolic syndrome than the general population, and consequently more prone to develop CV diseases. Moreover, they are normally prescribed complex medication regimens and tend to be more inactive and keep unhealthy dietary habits, no matter the socioeconomic status. The study, on a sample of Northern Italian patients aged 35-69 years, employed the CUORE algorithm. It considered the risk in these patients compared with the data from the general population, defining SMI as schizophrenia or bipolar disorder. The analysis showed CV risk raise with ageing in both sexes, comparable risk in both patients with low or high educational level, and it transpired those women with SMI were consistently more at risk of experience a CV event than the general population counterpart, even at a younger age (about 40% increased risk). Such results are supported by the higher number of smoked cigarettes, the higher triglycerides, and lower HDL cholesterol levels displayed in females with SMIs compared to the control group of healthy individuals. Men with SMIs showed higher triglycerides, lower HDL cholesterol, and more cigarette smoking than healthy subjects. However, the increased risk in males is likely offset by lower SBP in patients with SMI. The CUORE have been validated on the general Italian population, but it would be appropriate to incorporate specific mental health measures.

It emerges that the CUORE algorithm keeps an overall well fit to studies based on the Italian population, and good accuracy in presence of a bunch of comorbidities such as psoriasis or severe mental illnesses, but keeps the limitations related to a fixed function on a specific population with a limited number of variable and needs to be reviewed time by time.

2 Machine Learning Approaches

In the following sections, a general description and the main clinical applications of the most used ML and DL methods to predict cardiovascular risk is provided. A short description of these approaches is shown in Table VII.

Table VII. *Different ML algorithms.*

Logistic regression	One of the most commonly used linear methods. The result is the probability that an object belongs to a specific class. Very common in medical applications. It is usually not prone to overfitting.
K-nearest neighbours (KNN)	Supervised learning algorithm used for objects' classification based on the characteristics of nearby objects; the input consists of k examples of training items closest in space. When the new test data is provided, KNN calculates the distance between the query data and the training samples. The Euclidean distance metric is often used. It is simple to use and implement, and it is applicable to both classification and regression problems.
Probabilistic models (Naïve Bayes)	A non-structural method used to demonstrate the existence of a prescribed type of mathematical object. It works by showing that if one randomly chooses objects from a specified class, the probability that the result is of the prescribed type is much greater than zero. Even if the demonstration uses probability, the conclusion is determined for certain, without any possible error.
Adaptive boosting (AdaBoost)	Statistical classification meta-algorithm that can be used combined with other learning algorithms; it considers as training data only samples that can improve the predictive power of the model, improving executing time avoiding the calculation of irrelevant subsets of the original dataset.
Random forest	An algorithm that uses a set of decision trees and the bootstrap method, which allows you to create many examples of the same size based on some randomly selected initial data and repeat; each decision tree is

	based on one of these examples. It usually presents high accuracy, but it is also often prone to overfitting.
Deep learning (neural networks)	The statistical model is composed of artificial neurons, inspired by a simplification of biological neural networks.

2.1 Logistic regression

Logistic regression was used for the prediction of cardiovascular events in *hypertensive Indonesian patients*. A major health problem in Indonesia is hypertension, due to its incidence at 25,8% among the population[34]. In this study, the risk variables of Surabaya Hajj hospital, such as age, heart rate, stress levels, body mass index were considered influential on hypertension through a regression curve without a specific pattern. In this analysis, an additive nonparametric logistic regression approach based on local linear estimators was implemented to model the case of hypertension from the above-mentioned variables[35]. The non-parametric regression model was used to determine the relationship between the response variable (y) and the predictive variable (t). The results of the study highlighted that hypertension was most predominant among people over 65 years with BMI around 25-30 kg/m^2 (condition of obesity) and normal heart rate around 60-100 bpm and most of them were living under stressful conditions. The model obtained an accuracy of in-sample data of 95% and 89,47% in the case of out-sample data. As a consequence, the additive nonparametric logistic regression resulted the best estimator for determining hypertension incidence. Furthermore, previous research showed a comparison between the classification accuracy of hypertension using binary logistic regression and the C4.5 algorithm with case studies in UPT Puskesmas Ponjong I, Gunungkidul[34]. Binary logistic regression is a method that describes the relationship between the response variable and multiple predictor variables with the variable equal to 1 to declare the existence of a feature and the value 0 to declare the absence of a feature. In this ML method, the factors influencing the onset of hypertension were sex, systolic blood pressure, and other diseases. The C4.5 algorithm is a data mining classification method used to build a decision tree. Factors influencing the onset of hypertension in this other technique were systolic blood pressure, medical history, other diseases, diastolic blood pressure, and sex, while the age factor did not affect it. Based on this analysis, the classification of hypertension by the binary logistic regression method obtained an accuracy of 72.53% and the algorithm C4.5 obtained an accuracy of 64.08%. In conclusion, the binary logistic regression was better than the C4.5 algorithm in classifying the type of hypertension.

The logistic regression was also used to predict the cardiovascular events in *hemodialysis patients*, where the incidence rate is about 30 times higher than in the general population. Several research groups have recently made several attempts to predict the cardiovascular outcome of dialysis patients (http://hdl.handle.net/10589/83083). For this purpose, a dataset of real data extracted from the EuCliD system, owned by the Fresenius Medical Care company, was used. The data of more than 4500 patients undergoing hemodialysis treatment three times a week for 18 months were analysed, in particular patients treated for the first time in their life with hemodialysis treatment. This category of patients was selected for the high onset of cardiovascular disease during the initial period of dialysis. The ML techniques chosen for the development of these models were logistic regression based on the Lasso algorithm, random forests, support vector machines, and self-organizing maps (a type of side-connected neural network). The best predictive performance was obtained with the use of the random forests model: a ROC AUC equal to 73% with sensitivity greater than 70% was obtained, demonstrating the ability of random forests to exploit non-linear relationships identified in the feature space. Two different feature selection methods both based on a wrapper strategy were integrated for the construction of the models to identify subgroups of features effective in prediction[36]. From the analysis of the implemented models and the analysis of the subgroups of features identified, it was possible to note that the presence of an inflammatory state, malnutrition and incorrect ultrafiltration of the patients were significant predictors of the onset of cardiovascular events in patients with hemodialysis treatment. These factors highlighted a greater risk of a sudden deterioration in the conditions of the cardiovascular system: therefore, personalized treatment strategies can be devised to decrease the cardiovascular risk in patients who have suffered dialysis treatment and are identified as being at greater cardiovascular risk.

2.2 K-nearest neighbours (KNN)

The KNN was used for the *detection and localization of MI*[37]. The presence or absence of MI is characterized by specific waves or segments in the ECG beats. The main indicators are the amplitude of the T wave, the amplitude of the Q wave, and the elevation or depression of the ST level. The location of the MI was predicted by analysing 12 ECG leads. Using the KNN method, a subdivision into 11 classes was made for localization: 1 normal class and 10 classes with different MI positions, taken from the ECG leads. The results showed good sensitivity and specificity values for MI detection, 99,97% and 99,9% respectively. To reduce the computational cost of the KNN classifier, a pruning algorithm

was employed. This method decreased the accuracy of both MI detection and localization. An important aspect to consider is the control of the robustness of the feature extraction because the detection and location of the MI strongly depend on the correct extraction of the time-domain characteristics.

The KNN was employed also for the *prognosis of cardiac ischemia*, which can be detected from the morphological characteristics derived from the observation of the ECG, in particular from the deviation of the ST segment and the changes in the T wave[38]. Good values of sensitivity, specificity, and accuracy were achieved. In this study other approaches were tried: Artificial Neural Network and Support vector machine. The experimental results confirmed that the ANN model outperformed with an accuracy of 96,62%, higher compared to SVM and KNN classifiers[38].

2.3 Probabilistic methods (Naïve Bayes)

Probabilistic models are used in many fields. Naïve Bayes (NB) methods are a set of supervised learning algorithms based on applying Bayes' theorem with the "naive" assumption of conditional independence between every pair of features given the value of the class variable. This approach does not depend on the assumption that hazards are proportional. For this reason, NB might be better suited to model data that do not follow a proportional hazards model, or for which the functional relationship between covariates and the hazard is not specified[39]. The problem with standard NB lies in the poorly calibration ability in some settings, and that is why it was necessary to extend it, developing other better methods based on NB. Among the various methods of this type, one has been developed called Censored Naive Bayes (CNB), which is nonparametric concerning the underlying distribution of event times and models the marginal covariate distributions in a flexible way. In many studies, some patients have incomplete data due to loss of follow-up or censorship and ignoring or assigning a particular event status to these subjects can lead to biased probability estimates. CNB is a method that considers this problem. However, CNB is poorly calibrated when the covariates are not mutually independent. For this reason, it has been developed another method called Censored Naive Bayes-Principal Components (CNB-PC), which is a modification of CNB which results in a much better-calibrated technique. In fact, CNB and CNB-PC methods were applied to *predict the risk of CVDs from electronic health record data*[40]. The data, which came from a healthcare system in the Midwestern United States and were extracted from the HMO Research Network Virtual Data Warehouse (HMORN VDW) associated with that system, were as follows: insurance

enrollment, demographics, pharmaceutical dispensing, utilization, vital signs, laboratory, census, and death records. The analysis was restricted to those subjects with two medical encounters in the in-network clinic with blood pressure information at least 30 days but at most 1.5 years apart, with drug coverage at the end of the baseline period. Patients under the age of 40 were excluded. Subjects with pre-existing serious comorbidities other than diabetes were excluded too. The final analysis dataset contained 87363 individuals. The cohort was randomly split into a training dataset (65522 subjects), and a test dataset (21841 subjects). From the results of this study, it was evident that CNB was rather poorly calibrated, overpredicting the risk, whereas CNB-PC had better performances[40]. In this study, Cox proportional hazards (CPH) was also used to compare the performances. The model emulates a situation where the effect of age on risk is nonlinear: subjects at middle age are at highest risk, and those at low and high age are at lower risk. This pattern is observed modelling the probability of nonfatal CV events. The additional flexibility of the CNB models allowed obtaining more accurate predictions when relevant covariates were omitted from the CPH model[41].

Another method based on NB is Gaussian Naïve Bayes (GNB), which was used for a classification problem and the results seemed to be more promising in terms of heart failure (HF) risk prediction accuracy than other techniques in the literature. This study was conducted on a sample of 297 people, and the following parameters were considered: age, sex, chest pain type, resting blood pressure, serum cholesterol, fasting blood sugar, resting electrocardiographic results, maximum heart rate achieved, exercise-induced angina, old peak, peak exercise slope, number of major vessels coloured by fluoroscopy and thallium scan. The robustness of the proposed model was evaluated using accuracy, specificity, and sensitivity as evaluation metrics and a forward feature selection. The best model was achieved with a subset of 9 features (including sex, chest pain type, resting electrocardiographic results, maximum heart rate achieved, exercise-induced angina, old peak, peak exercise slope, number of major vessels coloured by fluoroscopy, thallium scan) with an accuracy on test of 93%, a training accuracy of 84%, a sensitivity of 88% and a specificity of 98%. The same study was conducted using the RF method, to make a comparison between RF and GNB and the latter performed better. It is possible to verify the greater effectiveness of GNB compared to RF also by observing the respective ROC AUC of 96% and 94%, respectively. The GNB approach outperformed the RF approach[42].

2.4 Tree-based models

2.4.1 Adaptive boosting

In 2021 D'Ascenzo et al. conducted a study to compare the performance of 4 ML methods and identify the best model to predict the *risk of death, MI, and risk of bleeding 1 year after discharge*[43]. A sample of 23270 adult patients (derivation cohort) and an additional sample of 3444 patients (external validation cohort) were analysed. The derivation cohort was randomly split into a training dataset (80%) and a validation dataset (20%). The analysed methods were AB, NB, RF, and KNN. AB, as known as the PRAISE model, was found to be the most performing model for each of the three cases. The PRAISE model considers 25 variables (clinical, anatomical, and procedural characteristics). A subset of the 25 features was found to be important for the three types of risk prediction separately. The most significant variables for the prediction of death are LVEF, age, hemoglobin level, statin (at discharge), EGFR, ACE inhibitor or ARB, previous bleeding, and malignancy. The main variables for the prediction of acute MI are hemoglobin level, LVEF, age, EGFR, peripheral artery disease, diabetes, complete revascularisation, multivessel disease. The main features for the prediction of risk of bleeding are hemoglobin level, age, LVEF, EGFR, diabetes, drug-eluting stent, complete revascularisation, malignancy. The PRAISE model classified 60% of acute coronary syndrome (ACS) patients as at low risk of ischemic and hemorrhagic events (1% probability) and 10% of ACS patients as at high risk of these events (>19%). The above-mentioned percentages are similar to those observed in the clinical setting. Among the considered models, the PRAISE model is the one with the most performing ROC curve in all of the three cases and accuracy, sensitivity, specificity, and F2 score values of 89%, 55%, 91%, and 43% on the evaluation set and 78%, 88%, 77% and 60% on the validation set, respectively[43]. The limitations of this approach are the slight underestimation of the adaptive enhancement classifier among high-risk patients and the no inclusion of the time as a variable since it has been created to be used with a fixed time frame (e.g. 1-year follow-up)[43].

2.4.2 Random forest

RF methods represent one of the most important families of ML methods based on the use of decision trees. This ML method is used in the CVD prediction setting due to its enhanced predictive ability and the ability to find linear and non-linear interactions between parameters and risk factors. This could improve the prediction in presence of previous

diseases. Finally, RF can be used in all applications involving a large group of data, in fact, the reliability of the algorithm is not compromised by the great number of data[44,45]. ML's unique feature is that algorithms do not require any assumptions on the underlying data; therefore, non-linearities can be caught more easily. Since considered risk factors often result insufficient to analyse complex relationships between heterogeneous data, this explains the necessity to implement ML methods in risk scores to underline the weight of specific illnesses on cardiovascular risk. Some of these methods were implemented in the study of inflammatory diseases using C-reactive protein as a marker of an inflammatory state. The use of the RF method was validated in the prediction of CVD in patients affected by Psoriatic Arthritis (PsA) and the results showed a higher sensitivity in presence of this kind of disease[46]. The results obtained with PsA generated an AUC of 83% which is significantly higher than the value of AUC obtained with the use of different ML methods as the KNN. The use of the RF was studied also in presence of other diseases as Ankylosing Spondylitis (AS), an inflammatory disease that can be localized in vertebral articulations. In this case, the RF method was implemented to predict the cardiovascular risk in AS patients and was compared with other ML methods and traditional risk scores[47]. The results showed better accuracy and specificity, of 65% and 66% respectively, and a lower sensitivity of 61% of the RF method if compared to the KNN method. However, the best algorithms that permit the prediction of CV risk resulted to be Reynolds Risk Score (RRS) and SCORE which have a higher specificity, accuracy, and sensitivity than ML methods. Finally, the most important limitations of this study are: first, the fact that the cohort of people suffering from AS involves a small number of subjects and therefore a large validation set for ML methods like RF, that need a significant number of data to be correctly balanced, is not available; second, the entire cohort sample is composed only by Italian Caucasian people.

2.5 Deep learning (neural networks)

Artificial neural networks (ANN) were used in *HF prediction using cardiopulmonary exercise testing (CPX)*, which was employed to quantify the degree of exercise intolerance patients with chronic HF. Recently many studies demonstrated how CPX can be useful in determining the risk of mortality, hospitalization, and other adverse outcomes in subject with HF[48], and how peak VO_2, ventilatory abnormalities, heart rate recovery (HRR), and other responses, could provide clinically significant and independent information for estimating prognosis in these patients. ANNs overcome the limitation of the traditional

statistical methods, providing a more accurate value for the risk estimation[48]. The study aimed to evaluate the predictive accuracy of the chosen CPX using an ANN and compare it to conventional methods to estimate the risk of CV mortality. The dataset included 2635 patients, which were split into training and testing subsets, therefore the 2635 recruited patients, 550 were not included in the training set since they have been used for testing the ANN. The subjects were followed for a major cardiac event for three years. ANN analyses were performed using a feed-forward multilayer learning strategy with a back-propagation training method. ROC curves were evaluated to compare the predicting value of each method. In the follow-up, it emerged that 291 subjects died from cardiovascular causes, and their data have been compared with the ones from the survivors. Each component of the CPX resulted significantly connect to CV mortality. Through logistic regression and proportional hazard, it was possible to determine the predictive strength of the observed variables (VE/VCO_2 slope revealed to be the strongest predictor of risk), while ANN showed higher predictive accuracy using 3 clinical variables together with the 5 CPX variables. ANN exhibited a better performance compared to the other two more traditional methods with an AUC of 72%, a sensitivity of 79% and a specificity of 63%. Moreover, ANN reached its ROC area peak at 5 hidden neurons, after which it decreased due to overfitting. However, no significant differences in the ROC areas in the three methods were observed. This study was applied to a cohort of 75% male, so it may not represent a female population, and the performance was evaluated using few variables, which could be the limit of this study.

Myers and co-workers applied the sparse autoencoder (SAE) approach to perform unsupervised learning for heart disease prediction[49]. SAE is an unsupervised neural network, which gives in output a model reconstructed through backpropagation of the input data and the reconstruction error, difference of the original input and its preconstruction. The dataset consisted of 4238 samples and 16 features, refers to a cardiovascular study on residents of Framingham, in Massachusetts, predicting their 10-years risk of CHD. Due to missing data, only 3656 have been considered, 3099 negative and 557 positive. The authors did a comparison between the ANN and SAE and they noticed that the combination of the two methods performed better. Then, comparative experiments were carried out using some of the algorithms proposed in this review to verify the effectiveness of the model. In particular, there was a comparison between the above-mentioned method and the performance of KNN, Classification and Regression Trees (CART), Logistic Regression

(LR), Naïve Bayes. The above-mentioned method reported the best performances with 90% of accuracy, 89 % of precision, 91% of recall and 90% of F1-score[49].

Traditionally, congestive heart failure (CHF) detection with ML on ECG relied on the extraction of heart rate variability features to build classifiers able to discriminate between healthy and CHF subjects. Porumb et al. proposed a new framework of congestive heart failure detection relying on raw ECG signals, through a 1-D Convolutional Neural Network, which is another type of neural network in DL[50]. Working on raw physiological data improves the interpretability of the model and ensures the transparency of the method[50]. A retrospective analysis was performed on two public datasets: the MIT-BIH Normal Sinus Rhythm Database from PhysioNet for the control group, including 18 healthy non-arrhythmic subjects for a total of 275974 beats considered, and the BIDMC Congestive Heart Failure Database from PhysioNet for the CHF group including 15 pathological subjects, for a total of 214531 beats considered. About 20 hours of recording were available for each subject, consequently, due to the very large amount of data collected, it was randomly selected one single beat every 5 seconds. All data are from the same laboratory. Total data was randomly split in smaller subsets as follows: 50% for training, 25% for validation and 25% for testing. Random splitting was repeated 10 times to reduce variance. Results showed the number of false positive is about 1% of the mean number of normal heartbeats, and the false negative are around 3% of the total number of heartbeats in CHF class. 478 from the control group and 854 CHF beats were misclassified, in both cases mainly belonging to the same subject. Since more than the 95% of the heartbeats were properly classified, patients could still be correctly categorized. Results have been confirmed by an additional simulation averaging the heartbeats in 5 s interval. No changes was detected in the performance. Employing the majority voting scheme for heartbeats within 5 minutes of ECG, it was possible to evaluate the accuracy of the method in the discrimination of the patients from the control or CHF group. Results suggest that it was possible to build a highly accurate model for CHF diagnosis using an about 5 minutes long raw ECG recording. Authors showed, indeed, that 99% of the 5 minutes ECG extracts could be employed for providing a correct CHF diagnosis. The limitation of this study is the limited dataset and it is needed to extend the obtained results to larger datasets.

López-Martínez et al. presented an alternative approach using ANN to classify hypertensive patients[51]. The dataset contains patients from the National Health and Nutrition Examination Survey (NHANES) from 2007 to 2016. Data are public and collected under survey form,

combining also social determinants of health data such as smoking, alcohol, and diet. This dataset is very unbalanced since the survey involved a generic population, not a specific one. Despite an acceptable model accuracy of 73%, the disbalance of the dataset affects the precision and sensitivity of the model which reveals a high number of false negatives. Some other ML algorithms were employed with the same NHANES data to compare their AUC and to validate the performance and sensitivity of this method in disease diagnostic. All the methods show very similar accuracy working on the same inaccurate dataset, whereas AUC and F1-score result was slightly higher in the proposed model. ANN with back-propagation was used to classify hypertensive patients. Out of 7330 patients included in the dataset, 69,9% were non-hypertensive and 30,1% were hypertensive. Therefore, the model shows a sensitivity of 40% and a specificity of 87%. The predictive power of positives and negatives was respectively 57% and 77%, with 12% of false alarm cases. Such results suggest that the proposed ANN model might not fit positive detection in healthcare diagnosis but are conversely effective in finding non-hypertensive patients with an accuracy of 73%. In conclusion, this model cannot be used for giving a final diagnosis but can be used to make clinicians aware of the probability that the patients could develop hypertension[51].

CONCLUSION

In the present review, two different approaches to the prevention of cardiovascular risk were described. According to the considered risk factors and the specific cardiovascular event, it could be convenient to move towards either traditional methods or more innovative methods of ML, as one predictive model can be more effective than another.

The main difference between ML and traditional methods is that ML learns from data and is dependent on training data, considering a greater number of both linear and non-linear variables, while statistical models use a limited number of variables only with a linear relationship between events and risk factors. ML algorithms are often highly flexible algorithms that require a penalty to avoid overfitting. Therefore, ML algorithms can take into account much more information[52]. In studies where the main objective is to predict the event without considering the association between specific risk factors and a clinically understandable event, traditional methods are replaced by the use of ML, since traditional methods are often relevant and necessary in association analyses but do not fit in prediction analyses, where the focus is on the output and not on the initial variables. ML and traditional methods often have the same objective, but are built differently: ML does not need pre-specified basic structures, but rather looks for the best fit: at the end, a better predictive

model is generated, at the expense of understanding how the risk factors are related to the event of interest. In particular, the final goal is always to perform accurate and precise predictions, being robust to future applications (e.g., validation on external datasets) and to elucidate the unknown relationships between data. Traditional statistical methods are based on theoretical and a priori hypotheses and assumptions and they draw inference from predictors. ML methodologies are instead data-driven models empirically optimized based on agnostic approaches which enable exploratory data analysis and insights. In addition, traditional methods are based on regression models that suppose a linear relationship between the event and risk factors, while ML can consider the non-linearity of some risk factors giving a more precise result (e.g., the risk of death increases with age, but the increase seen in the 40-50 age group is different from that in the 70-80 age group). At the same time, however, an algorithm could have reduced performance if applied, for example, to patients belonging to ethnic minorities if these subjects were not included inside the training phase of the algorithm.

Thanks to its flexibility, it is believed that ML in the future will have better performance than traditional statistical modelling. The prediction of cardiovascular risk is increasingly approaching an automated sphere controlled by AI. Despite all the difficulties of using ML approaches, this will greatly facilitate the work of doctors, increase the effectiveness of diagnostics and contribute to the choice of effective treatment measures.

KEY MESSAGES

- *Traditional Risk Score methods* are based on theoretical and a priori hypotheses and assumptions. They can catch the linear relationship between the event and risk factors, as they are regressor models.
- *Machine Learning methods* are data-driven models based on agnostic approaches that allow for explorative data analysis. They can catch the non-linearity between risk factors and events giving a more accurate risk prediction. The training of the model can be affected by the data.
- Machine Learning methods can improve the CVD risk prediction supporting clinical decision and early diagnosis.

REFERENCES

1. Abidov A, Chehab O. Cardiovascular risk assessment models: Have we found the

perfect solution yet? J Nucl Cardiol [Internet]. 2020 Dec 21;27(6):2375–85. Available from: https://doi.org/10.1007/s12350-019-01642-x

2. Ming C, Viassolo V, Probst-Hensch N, Chappuis PO, Dinov ID, Katapodi MC. Machine learning techniques for personalized breast cancer risk prediction: comparison with the BCRAT and BOADICEA models. Breast Cancer Res [Internet]. 2019 Dec 20;21(1):75. Available from: https://breast-cancer-research.biomedcentral.com/articles/10.1186/s13058-019-1158-4

3. Churpek MM, Carey KA, Edelson DP, Singh T, Astor BC, Gilbert ER, et al. Internal and External Validation of a Machine Learning Risk Score for Acute Kidney Injury. JAMA Netw Open [Internet]. 2020 Aug 11;3(8):e2012892. Available from: https://jamanetwork.com/journals/jamanetworkopen/fullarticle/2769146

4. Halasz G, Sperti M, Villani M, Michelucci U, Agostoni P, Biagi A, et al. A Machine Learning Approach for Mortality Prediction in COVID-19 Pneumonia: Development and Evaluation of the Piacenza Score. J Med Internet Res [Internet]. 2021 May 31;23(5):e29058. Available from: https://www.jmir.org/2021/5/e29058

5. Firyulina MA, Kashirina IL. Using of machine learning methods in cardiology. J Phys Conf Ser [Internet]. 2021 May 1;1902(1):012135. Available from: https://iopscience.iop.org/article/10.1088/1742-6596/1902/1/012135

6. Colpani O, Preventiva F, Farmacologiche S. Machine learning : the ability to predict applied to research and clinical practice. 2019;11(4):5–11.

7. Andersson C, Johnson AD, Benjamin EJ, Levy D, Vasan RS. 70-year legacy of the Framingham Heart Study. Nat Rev Cardiol [Internet]. 2019;16(11):687–98. Available from: http://dx.doi.org/10.1038/s41569-019-0202-5

8. D'Agostino RB, Vasan RS, Pencina MJ, Wolf PA, Cobain M, Massaro JM, et al. General cardiovascular risk profile for use in primary care: The Framingham heart study. Circulation. 2008;117(6):743–53.

9. Orfanoudaki A, Chesley E, Cadisch C, Stein B, Nouh A, Alberts MJ, et al. Machine learning provides evidence that stroke risk is not linear: The non-linear Framingham stroke risk score. PLoS One [Internet]. 2020;15(5):1–20. Available from: http://dx.doi.org/10.1371/journal.pone.0232414

10. Benjamin EJ, Wolf PA, D'Agostino RB, Silbershatz H, Kannel WB, Levy D. Impact of atrial fibrillation on the risk of death: The Framingham Heart Study.

Circulation. 1998;98(10):946–52.
11. Batty GD, Hamer M. Vascular risk factors, framinghamrisk score, and covid-19: Community-based cohort study. Cardiovasc Res. 2020;116(10):1664–5.
12. DAWBER TR, MEADORS GF, MOORE FE. Epidemiological approaches to heart disease: the Framingham Study. Am J Public Health. 1951;41(3):279–81.
13. Cuende JI, Cuende N, Calaveras-Lagartos J. How to calculate vascular age with the SCORE project scales: A new method of cardiovascular risk evaluation. Eur Heart J. 2010;31(19):2351–8.
14. Goh LGH, Welborn TA, Dhaliwal SS. Independent external validation of cardiovascular disease mortality in women utilising Framingham and SCORE risk models: A mortality follow-up study. BMC Womens Health. 2014;14(1):1–11.
15. Piccininni M, Rohmann JL, Huscher D, Mielke N, Ebert N, Logroscino G, et al. Performance of risk prediction scores for cardiovascular mortality in older persons: External validation of the SCORE OP and appraisal. PLoS One. 2020;15(4):1–18.
16. Fernández-Labandera C, Calvo-Bonacho E, Valdivielso P, Quevedo-Aguado L, Martínez-Munoz P, Catalina-Romero C, et al. Prediction of fatal and non-fatal cardiovascular events in young and middle-aged healthy workers: The IberScore model. Eur J Prev Cardiol. 2019;
17. Hippisley-Cox J, Coupland C, Vinogradova Y, Robson J, Brindle P. Performance of the QRISK cardiovascular risk prediction algorithm in an independent UK sample of patients from general practice: A validation study. Heart. 2008;94(1):34–9.
18. Collins GS, Altman DG. An independent and external validation of QRISK2 cardiovascular disease risk score: A prospective open cohort study. BMJ. 2010;340(7758):1231.
19. Collins GS, Altman DG. Predicting the 10 year risk of cardiovascular disease in the United Kingdom: Independent and external validation of an updated version of QRISK2. BMJ. 2012;345(7867):1–12.
20. Finnikin S, Ryan R, Marshall T. Statin initiations and QRISK2 scoring in general practice. 2017;(December). Available from: https://doi.org/10.3399/bjgp17X693485
21. Hippisley-Cox J, Coupland C, Robson J, Brindle P. Derivation, validation, and evaluation of a new QRISK model to estimate lifetime risk of cardiovascular

disease: Cohort study using QResearch database. Bmj. 2011;342(7788):93.
22. Assmann G, Cullen P, Schulte H. Simple scoring scheme for calculating the risk of acute coronary events based on the 10-year follow-up of the Prospective Cardiovascular Münster (PROCAM) study. Circulation. 2002;105(3):310–5.
23. Cox DR. Regression Models and Life-Tables. J R Stat Soc Ser B. 1972;34(2):187–202.
24. Batsis JA, Sarr MG, Collazo-Clavell ML, Thomas RJ, Romero-Corral A, Somers VK, et al. Cardiovascular Risk After Bariatric Surgery for Obesity. Am J Cardiol. 2008;102(7):930–7.
25. Adams TD, Gress RE, Smith SC, Halverson RC, Simper SC, Rosamond WD, et al. Long-Term Mortality After Gastric Bypass Surgery. Surv Anesthesiol [Internet]. 2008 Apr;52(2):94. Available from: http://journals.lww.com/00132586-200804000-00040
26. Cook NR, Paynter NP, Eaton CB, Manson JE, Martin LW, Robinson JG, et al. Comparison of the framingham and reynolds risk scores for global cardiovascular risk prediction in the multiethnic women's health initiative. Circulation. 2012;125(14):1748–56.
27. Buring JE, Cook NR. Algorithms for the Assessment of Global Cardiovascular Risk in Women - The Reynolds Risk Score. Am Med Assoc. 2016;297(6):611–20.
28. Ridker PM, Silvertown JD. Inflammation, C-Reactive Protein, and Atherothrombosis. J Periodontol. 2008;79(8s):1544–51.
29. Doukaki S, Caputo V, Bongiorno MR. Psoriasis and cardiovascular risk: Assessment by cuore project risk score in Italian patients. Dermatol Res Pract. 2013;2013.
30. Salvi V, Aguglia A, Barone-Adesi F, Bianchi D, Donfrancesco C, Dragogna F, et al. Cardiovascular risk in patients with severe mental illness in Italy. Eur Psychiatry. 2020;63(1):4–9.
31. Zanchè A, Lapi F, Parretti D. Cardiovascular risk chart : past , present and future. Rivista SIMG. 2021;28(1):32–7.
32. La Torre G, Palmeri V, Pagano L, Nati G, Sernia S, Mannocci A. Evaluation of cardiovascular risk profile: A comparative analysis between CUORE algorithm and the Framingham risk scores. Eur Rev Med Pharmacol Sci. 2018;22(8):2398–404.
33. Navarini L, Margiotta DPE, Caso F, Currado D, Tasso M, Angeletti S, et al.

Performances of five risk algorithms in predicting cardiovascular events in patients with Psoriatic Arthritis: An Italian bicentric study. PLoS One. 2018;13(10):1–11.

34. Rumaenda W, Wilandari Y, Safitri D. Perbandingan Klasifikasi Penyakit Hipertensi Menggunakan Regresi Logistik Biner Dan Algoritma C4.5 (Studi Kasus Upt Puskesmas Ponjong I, Gunungkidul). J GAUSSIAN [Internet]. 2016;5(2):299–309. Available from: http://ejournal-s1.undip.ac.id/index.php/gaussian

35. Ana E, Chamidah N, Andriani P, Lestari B. Modeling of hypertension risk factors using local linear of additive nonparametric logistic regression. J Phys Conf Ser. 2019;1397(1).

36. Dimensioni P, Di C, Di T. FEATURE SELECTION PER Introduzione e contesto. 2017;(April).

37. Arif M, Malagore IA, Afsar FA. Detection and localization of myocardial infarction using K-nearest neighbor classifier. J Med Syst. 2012;36(1):279–89.

38. Murthy HSN, Meenakshi DM. ANN, SVM and KNN Classifiers for Prognosis of Cardiac Ischemia- A Comparison. Bonfring Int J Res Commun Eng. 2015;5(2):07–11.

39. Webb GI, Keogh E, Miikkulainen R, Miikkulainen R, Sebag M. Naïve Bayes. In: Encyclopedia of Machine Learning [Internet]. Boston, MA: Springer US; 2011. p. 713–4. Available from: http://link.springer.com/10.1007/978-0-387-30164-8_576

40. Wolfson J, Bandyopadhyay S, Elidrisi M, Vazquez-Benitez G, Vock DM, Musgrove D, et al. A Naive Bayes machine learning approach to risk prediction using censored, time-to-event data. Stat Med [Internet]. 2015 Sep 20;34(21):2941–57. Available from: http://doi.wiley.com/10.1002/sim.6526

41. Wolfson J, Bandyopadhyay S, Elidrisi M, Vazquez-Benitez G, Vock DM, Musgrove D, et al. A Naive Bayes machine learning approach to risk prediction using censored, time-to-event data. Stat Med. 2015;34(21):2941–57.

42. Ali L, Khan SU, Golilarz NA, Yakubu I, Qasim I, Noor A, et al. A Feature-Driven Decision Support System for Heart Failure Prediction Based on $\chi 2$ Statistical Model and Gaussian Naive Bayes. Comput Math Methods Med. 2019;2019.

43. D'Ascenzo F, De Filippo O, Gallone G, Mittone G, Deriu MA, Iannaccone M, et al. Machine learning-based prediction of adverse events following an acute coronary syndrome (PRAISE): a modelling study of pooled datasets. Lancet. 2021;397(10270):199–207.

44. Svetnik V, Liaw A, Tong C, Christopher Culberson J, Sheridan RP, Feuston BP. Random Forest: A Classification and Regression Tool for Compound Classification and QSAR Modeling. J Chem Inf Comput Sci. 2003;43(6):1947–58.
45. Yang L, Wu H, Jin X, Zheng P, Hu S, Xu X, et al. Study of cardiovascular disease prediction model based on random forest in eastern China. Sci Rep. 2020;10(1):1–8.
46. Navarini L, Sperti M, Currado D, Costa L, Deriu MA, Margiotta DPE, et al. A machine-learning approach to cardiovascular risk prediction in psoriatic arthritis. Rheumatol (United Kingdom). 2020;59(7):1767–9.
47. Navarini L, Caso F, Costa L, Currado D, Stola L, Perrotta F, et al. Cardiovascular Risk Prediction in Ankylosing Spondylitis: From Traditional Scores to Machine Learning Assessment. Rheumatol Ther [Internet]. 2020;7(4):867–82. Available from: https://doi.org/10.1007/s40744-020-00233-4
48. Myers J, De Souza CR, Borghi-Silva A, Guazzi M, Chase P, Bensimhon D, et al. A neural network approach to predicting outcomes in heart failure using cardiopulmonary exercise testing. Int J Cardiol [Internet]. 2014;171(2):265–9. Available from: http://dx.doi.org/10.1016/j.ijcard.2013.12.031
49. Mienye ID, Sun Y, Wang Z. Improved sparse autoencoder based artificial neural network approach for prediction of heart disease. Informatics Med Unlocked [Internet]. 2020;18:100307. Available from: https://doi.org/10.1016/j.imu.2020.100307
50. Porumb M, Iadanza E, Massaro S, Pecchia L. A convolutional neural network approach to detect congestive heart failure. Biomed Signal Process Control [Internet]. 2020;55:101597. Available from: https://doi.org/10.1016/j.bspc.2019.101597
51. López-Martínez F, Núñez-Valdez ER, Crespo RG, García-Díaz V. An artificial neural network approach for predicting hypertension using NHANES data. Sci Rep [Internet]. 2020;10(1):1–14. Available from: https://doi.org/10.1038/s41598-020-67640-z
52. Goldstein BA, Navar AM, Carter RE. Moving beyond regression techniques in cardiovascular risk prediction: Applying machine learning to address analytic challenges. Eur Heart J. 2017;38(23):1805–14.

Conflicts of interest. — The authors certify that there is no conflict of interest with any financial organization regarding the material discussed in the manuscript.

Authors' contributions. — MS, MM and MAD have given substantial contributions to the conception of the manuscript, FSP, AD, SB, AF, RR and CS to acquisition, analysis and interpretation of the data. All authors have participated to drafting the manuscript, MAD revised it critically. All authors read and approved the final version of the manuscript.

TABLES

Table I. Regression Coefficients and Hazard Ratios12. SBP = systolic blood pressure.

Table II. Variables used in the construction of the Cox proportional hazards model. The P-value is <0,001 for all variables.

Table III. Coefficients of PROCAM Scoring Scheme.

Table IV. 10 years risk of acute coronary events associated with each PROCAM score22.

Table V. This table shows the coefficient β and standards errors (SE) for RRS. The P-value is <0,001 for all variables. Abbreviations: HbA1c = hemoglobin A1c; SBP = Systolic blood pressure; HDL-C = high-density lipoprotein cholesterol; hsCRP = high- sensitivity C-reactive protein; MI = myocardial infarction27.

Table VI. Cox proportional hazard model for the CV assessment in 10 years31.

Table VII. Different ML algorithms.

TITLES OF FIGURES

Figure 1. The two main categories of risk factors: non-modifiable risk and modifiable risk. Modifiable risk factors also include lifestyle factors (diet, physical exercise, stress, and smoking) and partly social factors (social deprivation and environment).

Figure 2. A summary of the main pros and cons of traditional risk scores vs machine learning-based approaches.