Biomedical Signal Processing and Control 91 (2024) 106035

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

journal homepage: www.elsevier.com/locate/bspc

Deep-stratification of the cardiovascular risk by ultrasound carotid artery images

Maria del Mar Vila [a,b], Lucas Gago [c], Pablo Pérez-Sánchez [a,d], Maria Grau [e,f,g], Beatriz Remeseiro [h,1], Laura Igual [c,*,1]

[a] CIBER Enfermedades Cardiovasculares, Instituto de Salud Carlos III, Monforte de Lemos 3-5, Pabellón 11, 28029 Madrid, Spain
[b] IMIM, Institut Hospital del Mar d'Investigacions Mèdiques, Dr. Aiguader 88, 08003 Barcelona, Spain
[c] Departament de Matemàtiques i Informàtica, Universitat de Barcelona, Gran Via de les Corts Catalanes 585, 08007 Barcelona, Spain
[d] Institute for Biomedical Research of Salamanca (IBSAL), P. de San Vicente, 182, 37007 Salamanca, Spain
[e] Serra Húnter Fellow, Department of Medicine, Universitat de Barcelona, Carrer Casanova 143, 08036 Barcelona, Spain
[f] August Pi i Sunyer Biomedical Research Institute (IDIBAPS), C/ del Rosselló, 149, 08036 Barcelona, Spain
[g] CIBER Epidemiología y Salud Pública, Instituto de Salud Carlos III, Monforte de Lemos 3-5, Pabellón 11, 28029 Madrid, Spain
[h] Department of Computer Science, Universidad de Oviedo, Campus de Gijón s/n, 33203 Gijón, Spain

ARTICLE INFO

Keywords:
Cardiovascular event
Survival model
Reclassification
Convolutional neural networks
Machine learning

ABSTRACT

Cardiovascular risk estimation functions predict the risk of cardiovascular events with clinical data and survival models. These functions accurately stratify individuals into low, moderate, and high-risk categories. However, they tend to classify a considerable number of individuals into the middle-risk category, and often, a subsequent reclassification into high-risk groups is required. Atherosclerosis is the leading cause of cardiovascular events, and ultrasound images of the Carotid Artery (CA) can detect its burden by measuring the carotid intima-media thickness and identifying atherosclerotic plaques. Current risk estimation functions do not consider ultrasound imaging. This paper proposes the use of *deep* ultrasound CA image features in survival models to improve risk stratification. In particular, we define new deep CA image features, extracting information from a convolutional neural network, and add them to an existing risk function. The experiments carried out show that using deep image features improves the AUC of the risk function to 0.842, and these features are enough to replace the information provided by blood biomarkers. Furthermore, the use of these features resulted in a 20% improvement in the reclassification of risk categories, specifically for individuals who suffered an event, as shown by the net reclassification improvement metric.

1. Introduction

Cardiovascular Diseases (CVDs) are the leading cause of death in developed countries. Most at-risk individuals of cardiovascular events suffer from atherosclerosis, a chronic inflammatory process morphologically characterized by an asymmetric focal thickening of the innermost layer of the artery. Ultrasound Carotid Artery (CA) images are used to detect the burden of atherosclerosis since they provide the possibility to measure the Carotid Intima-Media Thickness (CIMT) (see Fig. 1) of the artery and identify the presence of atherosclerotic plaques (Mannheim Consensus [1]). Monitoring the CIMT and detecting the atherosclerotic plaque, as well as its characteristics or changes, may have significant clinical relevance for CVD risk assessment.

In the field of cardiovascular epidemiology, practitioners use risk prediction functions [2–6] to estimate the risk of suffering an event in a period of time. These functions are based on survival models and estimate the risk using a set of clinical variables of each individual. Classically, the predictive results of risk functions are divided into different risk categories for stratification. Each risk category is defined by an interval probability of suffering a cardiovascular event in the next ten years. The cut-off points that define the risk stratification in *low*-risk group, *moderate*-risk group, and *high*-risk group have practical implications for deciding pharmacological intervention measures. Moreover, according to several clinical trials in primary prevention [7], treating populations is only cost-effective in the *high*-risk group.

One example of these risk prediction functions is REGICOR, which was built based on REGICOR study [2] and has shown accurate predictions of cardiovascular events [2,4,5,8]. REGICOR is a population-based observational cohort study conducted in the province of Girona (Spain).

* Corresponding author.
E-mail address: ligual@ub.edu (L. Igual).
[1] These authors jointly supervised this work.

https://doi.org/10.1016/j.bspc.2024.106035
Received 10 July 2023; Received in revised form 21 December 2023; Accepted 29 January 2024
Available online 3 February 2024

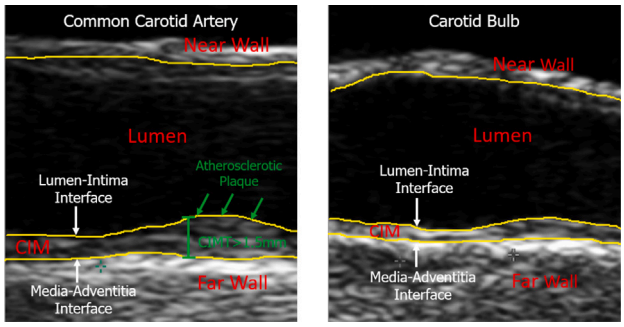

Fig. 1. Ultrasound CA images from two territories: common CA (left) and bulb (right). The different parts of the CA are delimited with lines: Near wall, Far wall, Lumen, Bulb, Carotid Intima-Media (CIM) region. In both cases, the CIMT is estimated in the CIM region (between Lumen-Intima Interface and Media-Adventitia Interface). Atherosclerotic plaque is also shown (portion of the CIM with a CIMT greater than 1.5 mm [1]).

Table 1
REGICOR data [2] on the ten-year incidence rate of Ischemic Heart Disease (IHD).

Risk category and interval probability	Subjects (%)	Subjects with IHD (%)	IHD in the risk group (%)
Low < 5	2449 (65.76%)	31 (25.83%)	1.27%
Moderate $\in [5, 15)$	1139 (30.59%)	72 (60.00%)	6.32%
High ≥ 15	136 (3.65%)	17 (14.17%)	12.50%
Total	3724 (100%)	120 (100%)	3.22%

Table 1 shows data on the ten-year incidence rate of Ischemic Heart Disease (IHD) in a population of 3724 individuals aged between 35 and 74 years who participated in REGICOR study. The risk category for each subject is based on the percentage probability of suffering a cardiovascular event in the next ten years (first column). Although the results shown in [2] are presented in four risk categories, we have summarized the table by combining the two middle categories (*low-moderate* and *high-moderate*) into a single category (*moderate*). The second and third columns indicate the number of subjects in each category and the number of events (IHD), respectively. The last column shows the percentage of events that occurred in a period of ten years for each risk category. As can be seen, the percentage of events falls inside each interval probability. More specifically, the percentage of subjects in the *low*, *moderate*, and *high* categories is 1.2%, 6.32%, and 12.50%, respectively, which is within the interval probability estimated: [0, 5)%, [5, 15)%, and [15,.)%. On the other hand, the values in the third column show that the *moderate* category concentrates the highest percentage of events (60%). This fact makes this stratification strategy ineffective in treating the population of the *moderate* group. Therefore, new pathological information should be considered to reclassify individuals from this group to the *high*-risk group.

In this research, we present a novel approach to improve the survival model for risk stratification proposed by Marrugat et al. [2]. In particular, our approach uses *deep* ultrasound CA image features in the survival model aiming at reclassifying individuals from the *moderate*- to the *high*-risk category. In particular, we consider a deep neural network (DNN) architecture to extract a set of deep features from the CA images, add them to the REGICOR function, and analyze the new survival model in terms of prediction and reclassification. We show that the DNNs are able to learn new feature embeddings useful to improve risk stratification by classifying events from *low* or *moderate* categories to higher categories. To do so, we compare the performance of our model using different sets of deep features and with another model that uses a set of phenotypes manually defined from the CIM region. Additionally, we assess the relevance of these features by comparing the risk function outlined in [2] to a model with the same variables, except for the invasive ones (i.e., those requiring blood extraction), which are substituted with the image features. In this sense, we assess a model that does not include invasive variables but incorporates additional information about atherosclerosis, including details of the atherosclerotic plaque location. Finally, the best sets of image features are compared in terms of reclassification. In particular, we investigate whether the inclusion of deep image features in the survival model results in the reclassification of individuals who have experienced an event from the *moderate*-risk group to the *high*-risk group.

The rest of the manuscript is organized as follows. Section 2 exposes the related work. Section 3 presents the proposed methodology for cardiovascular risk estimation using deep image features. Section 4 describes the REGICOR dataset used in the experiments. Section 5 explains the experimentation setup and the performance measures used. Section 6 describes the experiments carried out and shows the corresponding results. Finally, Section 7 closes with the main conclusions and future lines of research.

2. Related work

The tissue of CA walls provides information about the patients' arteries and cardiovascular health. For this reason, the study of ultrasound CA plaque images has been considered clinically relevant. Moreover, the long induction period of atherosclerosis makes it suitable for the study of subclinical CVD for preventive purposes. Several attempts in the literature tried to assess the cardiovascular risk of subjects using CA image features, as discussed below.

The main purpose of studies in [9,10] was to create an image-based system to characterize the plaque and the carotid artery walls in ultrasound CA images. In these works, authors used the lumen diameter for risk stratification as ground truth to solve a classification problem between two ranges: low and high risk. Their proposal was to estimate the spatial distribution of gray levels to examine the texture for Common CA (CCA) far and near wall, reaching high classification accuracies with two classification methods: Support Vector Machine (SVM) and Principal Component Analysis (PCA).

Other works [11–14] focused on analyzing the CA in longitudinal studies (i.e., repeated observations of the same subjects during a period of time). These works obtained CA image features, either manually or using semi-automatic methods, and used them to create risk prediction functions. Most of the image features considered are related to IMT (mean IMT, maximum IMT, minimum IMT, IMT variability) and atherosclerotic plaque (total plaque area, grayscale median of plaque). Since these features are used repeatedly in the literature, from now on, we will refer to this set of six features as *classical image phenotypes*. In particular, the model presented in [13] used the classical image phenotypes in two instants of time combined with conventional (non-imaging) risk variables achieving better results than classical survival models. Alternatively, Kyriacou et al. [15] used Probabilistic Neural Networks (PNN) and SVM classifiers to combine clinical features, phenotypes, and other CA image features (based on texture and morphology) for plaque classification (event vs. non-event). Despite the promising results for predicting events in individuals with plaque, this classification task is different from the risk stratification problem we address in this study.

Moreover, several works in the literature [16–19] compare classical risk prediction functions with different Machine Learning (ML) approaches. The main difference between both approaches is that ML methods use the event/non-event information for prediction (binary classification task) and survival models use the time until the event to predict the risk of suffering an event. In these studies, they used classical clinical variables, also known as *risk factors* and some additional characteristics from a population initially free of cardiovascular disease in a longitudinal study. The ML techniques outperform existing approaches in cardiovascular event prediction. In this context, Weng et al. [16] compared different ML algorithms: random Forest (RF), logistic regression, gradient boosting, and neural networks, which

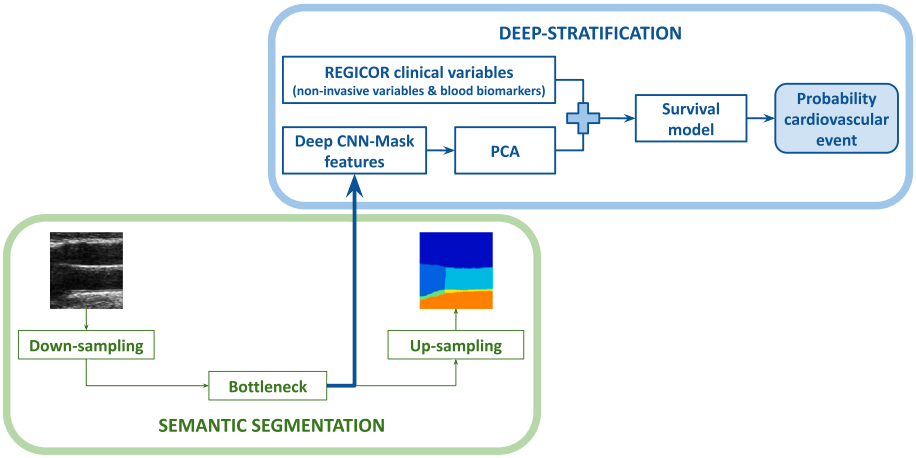

Fig. 2. Proposed methodology for the deep-stratification of the cardiovascular risk. The survival model receives an input vector with 12 features, which include 8 clinical variables used in the REGICOR risk function and 4 deep CNN-Mask features extracted from a semantic segmentation model of the carotid intima-media [24] and transformed by PCA.

reached the best results in cardiovascular prediction. The most recent approaches [17,18] proposed boosted ensemble algorithms followed by automated feature selection using information gain ratio. Both of them reached very high accuracies in cardiovascular death [17] and cardiovascular event [18] prediction. Although they did not use CA images, Ambale-Venkatesh et al. [19] included features from electrocardiography images and reached the highest accuracy using the RF algorithm to predict the events and to select relevant features from a large set of variables (more than 700).

In terms of subject reclassification (i.e., moving subjects who suffered an event to higher risk categories and subjects free of event to lower risk categories), recent research works [20,21] used Net Reclassification Improvement (NRI) to evaluate if the addition of one or more variables to a survival model improves its predictive capacity. In particular, these improvements are based on the reclassification of subjects with event into higher categories. Moreover, Tamarappoo et al. [18] applied this measure to evaluate the improvement in their proposed ML prediction model and reached an improvement of approximately 50% in reclassification.

Finally, it is worth mentioning some studies [22,23], that characterize other types of information that affect cardiovascular health but are not conventional images or clinical variables. These studies deal with biomedical signals, such as PPG [22] or ECG [23]. The techniques in these two studies are, respectively, Hilbert transforms and wavelet transforms. The results achieved in both cases are very competitive in terms of signal characterization and detection of specific patterns. However, this is not the focus of our study.

3. Methodology

We propose to improve the survival model based on the REGICOR risk function [2] by combining its clinical variables with a novel set of deep features extracted from ultrasound CA images. Fig. 2 depicts the different stages of the proposed method, which are subsequently described.

3.1. REGICOR clinical variables

Most of the risk prediction functions in the field of cardiovascular epidemiology are based on survival models, which are Cox Proportional Hazards models (CoxPh) [2–4,6]. The CoxPh analyzes the risk affecting the survival of a population of subjects [25]. In these models, the outcome is the time until an event occurs, and the predictor covariates are the risk factors. In particular, CoxPh survival models, such as the ones used in Framingham [3] or in REGICOR [2], allow us to estimate the probability of suffering a cardiovascular event in the next period of time [25], say ten years, by means of the following formula:

$$prob(event|x) = 1 - S^{exp(\sum_{n=1}^{N} \beta_n x_n - \sum_{n=1}^{N} \beta_n \bar{x}_n)}, \quad (1)$$

where N is the number of risk factor variables, x_n is the value of the nth variable for the individual x, β_n is the Cox coefficient for the nth variable, $\bar{x}_n$ is the average value of the nth variable in the validation population, and S is the average ten-year survival in the validation population. One of the best-known survival models in the literature is Framingham [3]. REGICOR risk function [2] is a model with the same risk factors as Framingham [3] but validated in the Spanish population. In particular, in the REGICOR risk function: $N = 8$, the average values correspond to the Spanish population and S is the average ten-year survival in the Spanish population. The REGICOR risk function uses the following eight clinical variables:

- *Non-invasive variables*: age, sex, diastolic blood pressure, systolic blood pressure, and smoke.
- *Blood biomarkers*: diabetes, total cholesterol, and HDL cholesterol.

3.2. Deep CNN-mask features

The extraction of the deep features is based on the model defined by Gago et al. [24] for the segmentation of the carotid intima-media. Specifically, this work designed a semantic segmentation model that uses the U-Net [26], a widely recognized architecture that incorporates a downsampling path for feature extraction and an upsampling path for region segmentation. The downsampling path of the proposed model employs an EfficientNet B0 [27] as its backbone while maintaining the bottleneck and the upsampling path of the original U-Net.

The deep features used in our proposal are obtained from this semantic segmentation model. In particular, we use the deep features extracted from the bottleneck, which is located between the down-sampling and up-sampling paths (see Fig. 2). As the segmentation model produces a mask of the CIM region, the 1152 features derived in this manner are referred to as *deep CNN-Mask features*.

3.3. Dimensionality reduction using principal component analysis

The *one in ten rule* [28] used in statistics determines that the number of variables for a survival model should be around 1 variable for every 10 events. Taking into account the number of events in our study (3.22%, see Section 1) and the number of deep features (1152) obtained for each territory (CCA and bulb), the dimension of the feature vector must be reduced.

For this purpose, we use Principal Component Analysis (PCA) [29], an algorithm that applies a linear transformation to compress a dataset onto a subspace of lower dimensionality while retaining the majority of the relevant information. The number of principal components to retain is typically determined by looking at the proportion of variance explained by each component and selecting a threshold, such as retaining the top n components that explain a certain percentage of the total variance.

In the current study, there are four CA images available for each subject: left and right sides of CCA and bulb territories. The deep CNN-Mask features are computed from the four images and then PCA is applied. The final vector per territory is calculated as the average of the feature vectors obtained from both sides (left and right). According to the experimentation, a variance percentage of 85% is applied to the two CCA images and 70% in the case of the bulb, thus obtaining two features per territory. That is, four deep features are concatenated with the eight REGICOR variables (which results in a final feature vector with $N = 12$), to feed the survival model, as following described.

3.4. Survival model

At this point, we have obtained a 12-dimensional vector with the following features: eight clinical variables, which are the ones used in the REGICOR risk function; and four deep CNN-mask features, which are obtained after applying PCA to the deep features computed from a semantic segmentation model trained on ultrasound CA images from two territories (CCA and bulb). These 12 features feed the survival model, as shown in Eq. (1), to finally estimate the probability of suffering a cardiovascular event in the next ten years. We call this strategy *deep-stratification* approach. In particular, for this work, we evaluate the respective CoxPh model (see Section 3.1), which estimates the time until the event. This model is obtained with `coxph()` function from Epi package [30] (R software [31]).

4. Dataset

This research work analyzes the participants from REGICOR [2], a population-based observational cohort study. The dataset consists of a sample of 5083 subjects from *Girona's Heart Registry* [32] with ultrasound CA images and other clinical information. All these data were collected from 2007 to 2010 and the subjects represent a general population aged between 35 and 84 who were followed up for approximately ten years. Two trained sonographers performed the CA US scans with an Acuson XP128 US system equipped with L75-10 MHz transducer and a computer program extended frequency (Siemens-Acuson). US longitudinal images were obtained in B-mode with a resolution of 23.5 pixels/mm and a size of 470 × 445 pixels. The original images were saved in DICOM format and then converted to PNG. The set of images collected for each subject was obtained from left and right CA in two different territories (CCA and bulb). During the CA acquisition, some images were discarded if the sonographers considered that the image quality was not sufficient. Due to the poor quality of the bulb images, they collected a total of 10,151 CCA images and 9143 Bulb images. Among them, 4727 CCA images and 3721 bulb images have CIMT reference values, given by the Amsterdam Medical Center with a semi-automatic method [33].

The clinical data include eight classical risk factors used in the REGICOR risk function. We call them *REGICOR variables* and they are: gender, age, smoking, systolic and diastolic blood pressure, total cholesterol, HDL cholesterol, and diabetes. Moreover, these data include the time until the event or the time to the date of the last contact. Given the nature of our study, we only considered subjects with no history of CVD (coronary artery disease, stroke, or intermittent claudication), therefore 314 subjects were discarded. Finally, a total of 4769 subjects were included in the dataset.

Table 2
Summary of clinical data of the REGICOR subjects considered in this study, grouped by 'sex' and with the p-value for the differences between the two groups. Categorical variables are expressed as *n (%)* and continuous variables as *mean (standard deviation)*.

	All N = 4769	Men N = 2620	Women N = 2149	p.value
Age	59.5 (11.8)	59.7 (11.8)	59.4 (11.8)	0.483
Total cholesterol	205 (35.5)	209 (35.2)	201 (35.3)	<0.001
HDL cholesterol	52.7 (11.8)	56.6 (11.6)	48.0 (10.3)	<0.001
Systolic blood pressure	129 (19.7)	126 (20.1)	133 (18.2)	<0.001
Diastolic blood pressure	77.2 (10.1)	75.1 (9.77)	79.7 (9.91)	<0.001
Diabetes	608 (12.7%)	252 (9.62%)	356 (16.6%)	<0.001
Smoke	810 (17.1%)	347 (13.3%)	463 (21.7%)	<0.001
Event	151 (3.17%)	58 (2.21%)	93 (4.33%)	<0.001

Table 2 shows a summary of the clinical data considered in this research. Note that, from the analyzed cohort of 4769 subjects who were free from CVD at baseline, there are 151 incident cases (3.17%) with cardiovascular events (acute myocardial infarction, other ischemic heart diseases, stroke, and the same causes of death) during the ten-year follow-up period.

5. Experimental setup

This section describes the setup used in the experimentation carried out to evaluate our deep-stratification approach. First, we explain the performance measures to evaluate the prediction capability of the survival model and the reclassification. Next, we describe the train–test split applied to the REGICOR dataset for validation purposes.

Note that the code of our proposed framework will be publicly available after paper acceptance.[2]

5.1. Evaluation metrics

In order to evaluate the performance of our deep-stratification approach we used two different metrics: the Area Under the Curve (AUC) and the Net Reclassification Improvement (NRI).

The AUC metric is used to assess the performance of the survival model. In this context, this metric agrees with Harrell's C [34], which is a goodness of fit measure of risk scores models such as statistical survival models.

On the other hand, NRI [35] is used to evaluate the reclassification improvement in risk prediction. NRI is a widely used metric in the recent clinical literature [18,20,21], which intuitively summarizes the improvement in the classification of individuals in risk categories. Therefore, it is used to summarize the incremental improvement obtained with new variables. The total improvement is the sum of the improvements in the reclassification of individuals without event and with event separately as is shown in the following formula:

$$NRI = (N1 - N2) + (N3 - N4) \tag{2}$$

where $N1$ and $N2$ are the percentages of individuals who have suffered an event and those who have been reclassified into higher and lower categories, respectively; and $N3$ and $N4$ are the percentages of individuals free of event who have been reclassified into lower and higher categories, respectively.

Note that when the problem is the probability of suffering an event during a period of time, we use the event variable (*yes/no*) and the time until the event (or the time until the date of the last contact in case of non-event subjects). For this reason, the statistical techniques utilized for NRI, such as the `nricens` function from `nricens`[36] package (R software [31]), which is employed in this study, incorporate survival functions that consider the time to the event in addition to the values shown in the formula. That is, these functions generate an NRI output that cannot be derived directly from the aforementioned formula (2).

[2] https://github.com/mmarvila/CA_deep_stratification.

Table 3
Number of images in the REGICOR dataset: train–validation–test split to evaluate the image-based features (CIMT GT) and test split to evaluate the survival models (CIMT GT + NO CIMT GT).

		Train	Validation	Test	Total
CIMT GT	CCA	2836	946	945	4727
	Bulb	2232	744	745	3721
NO CIMT GT	CCA			5424	5424
	Bulb			5422	5422
Total	CCA	2836	946	6369	10 151
	Bulb	2232	744	6167	9143

Table 4
Kept variance and number of features obtained when applying PCA to the two sets of deep features.

Territory	CNN-Mask		CNN-CIMT	
	Var.	No. feats.	Var.	No. feats.
CCA	99%	>8		
	95%	4	90%	>8
	85%	2	85%	8
	70%	1	70%	5
BULB	95%	>8	95%	>8
	90%	5	90%	6
	85%	4	85%	4
	70%	2	70%	2

5.2. Train–test split

Table 3 shows the number of images used to evaluate our deep-stratification approach. As explained in Section 4, the REGICOR dataset contains a total of 10,151 CCA images and 9143 bulb images. Among them, 4727 CCA images and 3721 bulb images have CIMT reference values that can be used as the Ground Truth (GT) for CIMT estimation. Note that the first two rows of the table (CIMT GT) correspond to the same train–validation–test split used in [24] (60% training set, 20% validation set, and 20% test set), which has been used here to evaluate image-based features.

The images without CIMT values can be used for the evaluation of the survival model. For this reason, we have also considered them and, therefore, the test set used in this case is composed of 6369 CCA and 6167 bulb images.

6. Results

This section details the four experiments conducted to evaluate our deep-stratification approach and analyzes the results obtained. The first two experiments are intended to select the best set of image-based features. In particular, we compare two sets of deep features in Experiment 1 and a set of hand-crafted features in Experiment 2. In both cases, we consider two territories (CCA and bulb) and several sizes of feature vectors. Next, in Experiment 3, we assess the relevance of the image-based features with respect to the blood biomarkers. For this purpose, we compare the survival model of the REGICOR function [2] but with the image-based features instead of the blood biomarkers. Finally, in Experiment 4, we analyze the most competitive configurations of the previous experiments in terms of the reclassification of the survival model.

The computational time and complexity of the proposed methodology mostly depends on the model used to extract deep features. As mentioned in [37], the time needed to process an input image is 0.026 s using a GeForce RTX 2080ti 11 GB GPU also from NVIDIA. On the other hand, the computational time to obtain the proposed CoxPh survival model is less than one millisecond per individual.

6.1. Experiment 1: Analysis of the deep features

In order to evaluate the adequacy of the proposed deep CNN-Mask features, we compare them with another set of deep features. For the extraction of the alternative deep features, we use a CNN model defined for CIMT estimation and plaque detection in [24]. This model was tuned using Bayesian optimization, with a final architecture composed of four blocks of convolutional layers followed by a max-pooling operation and two blocks of fully connected layers followed by batch normalization and dropout. In this case, the deep features were extracted from the second block of fully connected layers. The 32 features derived in this manner are referred to as *deep CNN-CIMT features*.

As we mentioned in Section 4 the CA images of the REGICOR dataset correspond to two territories, CCA and bulb. The quality of the images in CCA is better, whilst there is an increased burden of atherosclerotic plaque in the bulb images. For this reason, we compute the deep features from both territories, individually and jointly for comparison.

For each CA image, the size of the feature vector is 1152 for the proposed deep CNN-Mask features and 32 for the alternative deep CNN-CIMT features. Since the number of events in the dataset is small (151 from 4769 subjects, see Table 2), the size of the feature vectors must be reduced. As previously explained in Section 3.3, the number of events in the dataset should be about 10 per risk factor variable in the survival model [28]. Thus, we set the maximum number of variables to be included in the survival model to 16. Taking into account that our proposed model already includes the 8 REGICOR risk function variables [2] (see Fig. 2), the number of deep features should be less than or equal to 8. To reduce the size of the two feature vectors considered, we apply PCA to obtain a maximum of 8 variables and keep at least 70% of the variance. Table 4 shows the number of features obtained after applying PCA with different variance percentages. For CCA images, a variance of 99% in deep CNN-Mask features and 90% in CNN-CIMT features result in feature vectors with more than 8 variables, so lower variances must be considered for the proposed survival model. Regarding the other territory, bulb, a variance lower than 95% must be considered in both feature sets.

As detailed in Section 3, each subject has two CA images (left and right) per territory (CCA and bulb). Thus, the final feature vector is calculated as the average of the feature vectors extracted from both sides. In case one of the two images is missing, the subject is discarded. For this reason, using the 6369 CCA and 6167 bulb images mentioned in Section 5.2, we obtain the deep features for the following number of subjects: 2796 subjects for CCA, 2760 subjects for bulb, and 2501 for both territories.

Table 5 shows the performance of the survival model using the two sets of deep features (CNN-Mask and CNN-CIMT), in the two territories (CCA and bulb) individually and jointly, and with different kept variances for PCA, based on the results reported in Table 4. In all cases, the deep features are concatenated with the 8 REGICOR variables [2] that are used as the baseline (AUC = 0.825). For each configuration, we report the total number of features (No. feats.), the predictive capacity of the survival model (AUC), the AUC increase with respect to the baseline, and the p-value of the AUC increase, which is considered statistically significant if $p < 0.06$. Note that equal values in the AUC metric have different increments due to rounding precision (three decimal digits), and vice versa.

As can be observed, the AUC increase reaches statistically significant values using either of the two sets of deep features. However, the best AUC results are obtained with the deep CNN-Mask features, with AUC greater than 0.840 in several configurations. With respect to the two territories, their combination provides the highest values when the deep CNN-Mask features are applied. Regarding the number of features, the highest AUC values are obtained with larger feature vectors, mainly because these configurations correspond to the joint use of the two territories. Note that the configurations that achieve statistically significant results ($p < 0.06$) are selected for further analysis (Experiments 3 and 4).

Table 5
Experiment 1. AUC results of the survival model fed with the 8 REGICOR variables and different sets of deep features (CNN-Mask and CNN-CIMT), applied to the input images of two territories (CCA and bulb). The number of features obtained after applying PCA to the deep features is specified between parentheses (F), after the variance percentage. The statistically significant results ($p < 0.06$) are in bold.

Deep	Territory		No.	AUC		
features	CCA	Bulb	feats.	AUC	increase[b]	p
CNN-Mask	PCA 95% (4F)	–	12	0.831	0.007	0.118
CNN-Mask	PCA 85% (2F)	–	10	0.831	0.006	0.230
CNN-Mask	PCA 70% (1F)	–	9	0.827	0.002	0.628
CNN-Mask	–	PCA 90% (5F)	**13**	**0.842**	**0.018**	**0.050**
CNN-Mask	–	PCA 85% (4F)	**12**	**0.842**	**0.017**	**0.052**
CNN-Mask	–	PCA 70% (2F)	10	0.834	0.010	0.296
CNN-Mask	PCA 95% (4F)	PCA 85% (4F)	**16**	**0.847**	**0.023**	**0.008**
CNN-Mask	PCA 85% (2F)	PCA 90% (5F)	**15**	**0.846**	**0.022**	**0.016**
CNN-Mask[a]	PCA 85% (2F)	PCA 70% (2F)	**12**	**0.842**	**0.017**	**0.054**
CNN-CIMT	PCA 85% (8F)	–	16	0.833	0.008	0.076
CNN-CIMT	PCA 70% (5F)	-	**13**	**0.832**	**0.008**	**0.046**
CNN-CIMT	–	PCA 90% (6F)	14	0.835	0.011	0.198
CNN-CIMT	–	PCA 85% (4F)	12	0.836	0.011	0.070
CNN-CIMT	–	PCA 70% (2F)	10	0.834	0.010	0.296
CNN-CIMT	PCA 70% (5F)	PCA 70% (2F)	15	0.836	0.011	0.066

[a] Our proposed method.
[b] AUC increase with respect to the 8 REGICOR variables [2] (baseline).

Table 6
Coefficients for the CoxPh model and p-values of the risk factors used in the survival model: eight factors from the REGICOR risk function and the six hand-crafted phenotypes selected based on the statistical analysis performed.

Risk factor	Territory	Coefficient	p
Age	–	0.06	<0.01
Sex	–	0.38	0.06
Total cholesterol	–	0.01	0.04
HDL cholesterol	–	−0.03	0.00
Systolic blood pressure	–	0.00	0.62
Diastolic blood pressure	–	0.01	0.23
Diabetes	–	0.49	0.02
Smoker	–	0.12	0.66
log(mean IMT)	CCA	2.55	<0.01
minimum IMT	CCA	−1.26	0.26
log(IMT variability)	CCA	−0.43	0.17
Small plaque and echoic	CCA	−0.61	0.32
Huge plaque and non-echoic	CCA	0.36	0.59
Huge plaque and echoic	CCA	−0.70	0.25
log(mean IMT)	Bulb	−1.48	0.30
minimum IMT	Bulb	0.22	0.83
log(IMT variability)	Bulb	0.71	0.14
Small plaque and echoic	Bulb	0.15	0.61
Huge plaque and non-echoic	Bulb	0.13	0.78
Huge plaque and echoic	Bulb	0.26	0.44

6.2. Experiment 2: Analysis of the hand-crafted features

The aim of this experiment is to analyze the use of a new set of *hand-crafted features*, which could replace the deep features in our proposed methodology (see Fig. 2). Based on the classical image phenotypes mentioned in Section 2, we consider six phenotypes manually defined from the CIM region (see Section 1). Thus, for each CA image we obtain the following image features: mean IMT, maximum IMT, minimum IMT, IMT variability, Total Plaque Area (TPA), and Grayscale Median (GSM) of plaque. Notice that TPA is measured in mm^2 and it is estimated in the region where the CIMT reaches more than 1.5 mm, following the Mannheim consensus [1]. GSM refers to the grayscale median value in the same area where TPA is evaluated.

These phenotypes are extracted from the four CA images available for each subject (left and right sides of CCA and bulb territories), thus obtaining a total of 24 phenotypes per subject. As previously explained in Section 6.1, a maximum of 8 image-based features should be added so a dimensionality reduction procedure must be applied. The definition of these 24 hand-crafted features is based on prior knowledge about specific characteristics of the carotid images. Therefore, we can reduce the number of features, from 24 to a maximum of 8, by selecting the most relevant phenotypes according to the results of a statistical analysis performed on the training data. The procedure carried out is summarized as follows:

- **Base-e logarithmic adjustments.** Some phenotypes are not normally distributed and hence they should be normalized using base-e logarithmic adjustment. These phenotypes are mean IMT, maximum IMT, and IMT variability (on both sides).
- **Categorization of variables.** Due to the low percentage of plaques (3.4% in CCA and 25.8% in bulb), TPA and GSM phenotypes are categorized into three classes. The categories for TPA are *non-plaque*, *small plaque*, and *high burden of plaque*, where the threshold between the last two categories is the median of all TPA values from the same side (five for CCA and six for bulb). GSM is categorized into *non-plaque*, *echolucent*, and *non-echolucent*, where the threshold for echolucency detection is the third quartile of all GSM values from the same side (107 in CCA and 95 in bulb). Finally, in order to analyze the interaction between the TPA and the GSM phenotypes, we create a new variable as a combination of both. The categories of this new variable are *non-plaque* (the subject does not have any plaque), *small plaque and non-echoic* (the subject has at least plaque in one side of the arteries, but none is huge or echoic), *small plaque and echoic* (the subject has at least plaque in one side that is echoic, but there is not any huge plaque), *huge plaque and non-echoic* (the subject has at least a huge plaque that is not echoic), and *huge plaque and echoic* (the subject has at least a huge plaque that is echoic). This categorization is done for CCA phenotypes and bulb phenotypes separately. Finally, since there is no event in the *small plaque and non-echoic* category for CCA, we merge it with the *non-plaque* category.
- **Mean between left and right sides.** In order to reduce the number of phenotypes, the mean IMT, maximum IMT, minimum IMT, and IMT variability are defined as the average of right and left side values [32]. If a value is missing on one side, we use the available value for this subject. If the value is missing on both sides, then the phenotype is considered missing.
- **Co-linearity.** We eliminate co-linear variables using the Variance Inflation Factor (VIF) [37]. In particular, we analyze all the variables from the model and discard the ones with VIF > 2. Maximum IMT phenotypes from CCA and bulb are also discarded for the model.
- **Discarding variables.** Ultimately, our approach involves systematically eliminating non-statistically significant phenotypes from the survival model. However, we make a deliberate choice to retain the phenotype variables that are deemed confounders. Note that a variable is considered a confounder if at least one of the coefficients of the variables that remained in the model changed more than 15%. After the analysis, all the selected phenotypes are considered confounders so we keep all of them.

As a result, Table 6, shows the coefficients for the CoxPh model (see Section 3.1) and their corresponding p-value. These coefficients correspond to a survival model including the eight REGICOR variables and the eight phenotypes selected. Note that the variable is statistically significant for the model if $p < 0.06$.

Table 7 shows the performance of the survival model using the hand-crafted features in the two territories (CCA and bulb), individually for the Statistical Analysis (SA) above described and jointly for PCA (maximum 8 variables and at least 70% of variance). In all cases, the hand-crafted features were concatenated with the 8 REGICOR variables [2] that are used as the baseline (AUC = 0.825). As in Experiment 1, we report the total number of features (No. feats.), the predictive capacity of the survival model (AUC), the AUC increase with respect to

Table 7
Experiment 2. AUC results of the survival model fed with the eight REGICOR variables and the hand-crafted features applied to the input images of two territories (CCA and bulb). SA stands for statistical analysis and PCA is followed by the variance percentage applied. In both cases, the number of features obtained is specified between parentheses (F). The statistically significant results ($p < 0.06$) are in bold.

Image features	Territory		No. feats.	AUC	AUC increase[a]	p
	CCA	Bulb				
Hand-crafted	**SA (4F)**	–	12	**0.843**	0.018	**0.010**
Hand-crafted	–	SA (4F)	12	0.839	0.015	0.116
Hand-crafted	**SA (4F)**	**SA (4F)**	16	**0.856**	0.031	**0.004**
Hand-crafted	PCA 99% (5F)		13	0.830	0.006	0.278
Hand-crafted	PCA 95% (4F)		12	0.826	0.001	0.734
Hand-crafted	PCA 90% (3F)		11	0.824	−0.001	1.314
Hand-crafted	PCA 80% (2F)		10	0.824	<0.000	1.280

[a] AUC increase with respect to the 8 REGICOR variables [2] (baseline).

Table 8
Experiment 3. AUC results of the survival model fed with the five non-invasive REGICOR variables and different configurations of image-based features selected in Experiments 1 and 2, applied to the input images of two territories (CCA and bulb). SA stands for statistical analysis and PCA is followed by the variance percentage applied. In both cases, the number of features obtained is specified between parentheses (F).

Image features	Territory		No. feats.	AUC	AUC increase[b]	p
	CCA	Bulb				
CNN-Mask	–	PCA 90% (5F)	10	0.826	0.002	0.894
CNN-Mask	–	PCA 85% (4F)	9	0.826	0.002	0.892
CNN-Mask	PCA 95% (4F)	PCA 85% (4F)	13	0.831	0.007	0.57
CNN-Mask	PCA 85% (2F)	PCA 90% (5F)	12	0.831	0.006	0.63
CNN-Mask[a]	PCA 85% (2F)	PCA 70% (2F)	9	0.827	0.002	0.874
CNN-CIMT	PCA 70% (5F)	–	10	0.813	−0.012	1.716
Hand-crafted	SA (4F)	–	9	0.825	0.000	0.978
Hand-crafted	SA (4F)	SA (4F)	13	0.839	0.014	0.352

[a] Our proposed method.
[b] AUC increase with respect to the 8 REGICOR variables [2] (baseline).

the baseline, and the p-value of the AUC increase, which is considered statistically significant if $p < 0.06$. Note that equal values in the AUC metric have different increments due to rounding precision (three decimal digits), and vice versa.

As can be seen, the AUC increase is statistically significant when using the features selected by the statistical analysis on the CCA territory, both individually or combined with the bulb. In particular, the best performance is achieved when using the two territories jointly. On the contrary, the configurations that used the feature vectors obtained after applying PCA to the hand-crafted features do not have statistical significance. Note that the configurations that achieve statistically significant results ($p < 0.06$) are selected for further analysis (Experiments 3 and 4).

6.3. Experiment 3: Analysis of the REGICOR variables

The aim of this experiment is to analyze the power of image-based features and see if it is possible for them to replace the 3 blood biomarkers used in the REGICOR risk function [2].

Table 8 shows the performance of the survival model using the 5 non-invasive REGICOR variables concatenated with the different configurations of image-based features selected in the previous experiments, according to their statistical significance. Note that the target of the experiment is to analyze if these features can replace the 3 blood biomarkers, so the 8 REGICOR variables are used as the baseline (AUC = 0.825). As in previous experiments, we report the total number of features (No. feats.), the predictive capacity of the survival model (AUC), the AUC increase with respect to the baseline, and the p-value of the AUC increase, which is considered statistically significant if $p < 0.06$. Note that equal values in the AUC metric have different increments due to rounding precision (three decimal digits), and vice versa.

As can be observed in Table 8, there is a positive increment in most of the survival models fed with image-based features with respect to the survival model fed with the three blood biomarkers (REGICOR risk function [2]). Note that only in the case of the deep CNN-CIMT features the AUC increment is negative.

6.4. Experiment 4: Analysis of the reclassification results

This experiment aims at analyzing the models selected in Experiments 1 and 2 in terms of their reclassification results for cardiovascular events. For this purpose, we consider the NRI metric (see Section 5.1) and the following cut-off points, which correspond to the risk categories defined in [2] and discussed in Table 1:

- *low*: Subjects with a probability of suffering an event <0.05.
- *low-moderate*: Subjects with a probability of suffering an event in the range $[0.05, 0.1)$.
- *high-moderate*: Subjects with a probability of suffering an event in the range $[0.1, 0.15)$.
- *high*: Subjects with a probability of suffering an event ≥ 0.15.

Table 9 shows the performance of the survival model using the 8 REGICOR variables [2] concatenated with the different configurations of image-based features selected in Experiments 1 and 2, according to their statistical significance. For each configuration, we report the total NRI value (NRI) and its Confidence Interval (CI), the NRI value for the subjects who suffered an event (NRI events) and its CI, and the NRI value for the subjects free of event (NRI controls) and its CI. Note that the NRI values are statistically significant if their CI does not include 0 and they are shown in bold.

The results presented in Table 9 indicate statistically significant improvements in reclassification for three specific configurations. These improvements were achieved using deep features, resulting in an increase of more than 17%. Only these three configurations show statistically significant results in either the "NRI events" or in the "NRI controls" column. Particularly, with our proposal, we show a significant increment in subjects who suffered an even, with an NRI of 20.02%, while the increment using CNN-CIMT features is lower, 17.14%. Instead, the other configuration using CNN-Mask shows a statistically significant increment in subjects free of event, although it is relatively small (1.8%). In contrast, the results obtained from the hand-crafted features do not demonstrate statistical significance in any case.

Table 10 shows the reclassification of the event group using the previously defined risk categories. The risk categories and the number of subjects are shown using the REGICOR risk function [2] (rows) and our proposal (columns). The values above the diagonal indicate subjects that have been assigned a higher category compared to REGICOR [2]. Conversely, values below the diagonal represent the number of subjects that have been downgraded. Similarly, the values on the diagonal indicate the number of subjects classified with the same category using both our proposal and REGICOR [4]. Although here we cannot show the calculation of the NRI values (as we mentioned in Section 5.1) Table 10 shows that, for the events, there are many more individuals who are reclassified into higher categories ($17 = 2 + 5 + 6 + 4$) than not than lower categories ($4 = 1 + 1 + 1 + 1$).

6.5. Comparison with the literature

The predictive capacity of our proposed method (AUC = 0.842, Table 5) is similar to or higher than the results proposed in the literature (see Section 2). Even without the 3 blood biomarkers used in the REGICOR risk function, we achieved a good cardiovascular risk prediction (AUC = 0.827, Table 8).

Table 9
Experiment 4. NRI results of the survival model fed with the 8 *REGICOR variables* and the different configurations of image-based features selected in Experiments 1 and 2, applied to the input images of two territories (CCA and bulb). SA stands for statistical analysis and PCA is followed by the variance percentage applied. In both cases, the number of features obtained is specified between parentheses (F). The statistically significant results (CI does not include 0) are in bold.

Image	Territory		NRI	NRI events	NRI controls
features	CCA	Bulb	[CI 95%]	[CI 95%]	[CI 95%]
CNN-Mask	–	PCA 90% (5F)	11.54 [−0.05; 0.30]	9.99 [−0.07; 0.28]	1.55 [0.00; 0.03]
CNN-Mask	–	PCA 85% (4F)	11.56 [−0.05; 0.30]	10.05 [−0.07; 0.28]	1.52 [0.00; 0.03]
CNN-Mask	PCA 95% (4F)	PCA 85% (4F)	**17.91 [0.01; 0.34]**	16.1 [−0.01; 0.32]	**1.81 [0.01; 0.03]**
CNN-Mask	PCA 85% (2F)	PCA 90% (5F)	16 [−0.01; 0.32]	14.76 [−0.02; 0.31]	1.24 [<0.00; 0.03]
CNN-Mask[a]	PCA 85% (2F)	PCA 70% (2F)	**20.82 [0.04; 0.38]**	**20.02 [0.03; 0.37]**	0.81 [−0.01; 0.02]
CNN-CIMT	PCA 70% (5F)	–	**17.50 [0.03; 0.33]**	**17.14 [0.03; 0.32]**	0.36 [−0.01; 0.01]
Hand-crafted	SA (4F)	–	9.79 [−0.03; 0.21]	8.85 [−0.04; 0.20]	0.95 [<0.00; 0.02]
Hand-crafted	SA (4F)	SA (4F)	6.47 [−0.11; 0.26]	5.58 [−0.12; 0.25]	0.9 [−0.01; 0.02]

[a] Our proposed method.

Table 10
Comparison between the REGICOR risk function [2] and our proposed method in terms of reclassification results for cardiovascular events.

REGICOR	Our proposed method				
risk function	<0.05	[0.05, 0.1)	[0.1, 0.15)	≥0.15	Total
<0.05	24	2	0	0	26
[0.05; 0.1)	1	4	5	6	16
[0.1; 0.15)	1	1	2	4	8
≥0.15	0	0	1	9	10
Total	26	7	8	19	60

The best result found in the literature is reported in [13] (AUC = 0.93). This study uses risk factors and phenotypes at two different times, thus making data collection more complex. Regarding the ML approaches, we do not reach the results reported in [19] (AUC = 0.86), but their proposed survival model includes data from questionnaires (lifestyles, history, medication, etc.), more biomarkers, electrocardiography, and magnetic resonance imaging features. That is, they use a set of characteristics that make the study more expensive and that are not appropriate to be obtained in primary care centers. Regarding deep features, our best result is with "PCA 95%, CCA & 85%, BULB" features, which reach 0.847 in the AUC metric (Table 5). In the literature, we also found the study conducted by Rine Nakanishi [17] who demonstrates a slightly superior AUC result of 0.85. However, they do use computed tomography images, a more expensive technique than US imaging.

In terms of reclassification, the total NRI reported by Tamaroppoo et al. [18] (53%) is better than the one obtained with our method (20.82%, Table 9). However, our method outperforms it in the event group, which is the objective of the presented proposal due to its clinical relevance. More specifically, the NRI of the event group reported in [18] is 8% versus 20.02% achieved with our proposal.

7. Conclusions

This work presents, for the first time in the literature, a survival model that integrates CA image features extracted from deep neural networks. The new survival model is capable of predicting the risk of suffering a cardiovascular event, which results in a deep-stratification of the cardiovascular risk. The proposal improves the survival model presented in [2] by adding information from CA ultrasound images. For that, we concatenated deep features from a CNN previously defined for CA image semantic segmentation. These features are able to improve the model in terms of prediction (AUC = 0.84, with an increment of 0.017 with respect to REGICOR risk function [2]) and reclassification (NRI = 20.8%, NRI events = 20%). We successfully achieved a reduction in the number of individuals in the middle-risk category and moved them to the *high*-risk category, which is our main goal.

In order to validate our proposal, we performed a comparison of different sets of image features and different configurations of these features. First, we compared our proposal with another set of deep features that reached a statistically significant improvement in prediction and reclassification, but with a smaller increase than with our proposal (AUC = 0.83, NRI = 17.5%, and NRI events = 17.1%). Second, we compared with a set of phenotypes manually defined from the CIM region and selected according to the statistical analysis results. In this case, the improvement reached in prediction is high (AUC = 0.86) and statistically significant, but the findings in reclassification were not statistically conclusive. In addition, our findings demonstrate that CA image features are able to replace invasive variables, such as blood biomarkers, while simultaneously providing localized information concerning atherosclerotic plaque.

The main limitation of our work is the small number of events in the dataset (151 events over 4769). With so few events, we are forced to greatly reduce the number of new image features and we are also exposed to overfitting. In this scenario, the survival models may not have enough statistical power to show all the differences between the different sets of image features. In addition, the proposed method has been tested on a single dataset, so it has not been possible to analyze its generalizability.

For future research, it would be useful to validate our method with an independent cohort that has more events. This would allow us to overcome the aforementioned limitations, increasing the statistical significance of the study and testing the power of generalization of our approach. In addition, it would be interesting to use other deep features obtained with CNNs combined with other deep neural networks trained for another task, such as a binary classification task (to predict event/non-event) or a regression task (to estimate the time until the event). Another line of research could be the interpretability of the specific features extracted from CA images. Understanding the contribution of individual features to the survival model, for example generating saliency maps, can provide information about which regions in CA images are associated with cardiovascular risk. Finally, it would be interesting to perform a longitudinal analysis of the deep features, including changes over time in CA, as it is suggested in [13], which uses risk factors and phenotypes at two different times.

CRediT authorship contribution statement

Maria del Mar Vila: Data curation, Investigation, Methodology, Software, Writing – original draft. **Lucas Gago:** Software. **Pablo Pérez-Sánchez:** Methodology. **Maria Grau:** Conceptualization, Funding acquisition, Investigation, Writing – review & editing. **Beatriz Remeseiro:** Conceptualization, Investigation, Methodology, Software, Supervision, Writing – review & editing. **Laura Igual:** Conceptualization, Funding acquisition, Methodology, Supervision, Writing – original draft, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The data that has been used is confidential.

Acknowledgments

This work was partially supported by the MICINN, Spain Grant RTI2018-095232-B-C21, AGAUR, Spain Grants (2021-SGR-01104, 2017-SGR-222 and 2017-SGR-1742), and the Spanish Ministry of Economy and Competitiveness, Spain through the Instituto de Salud Carlos III-FEDER (CIBERCV and FIS CPII17/00012).

References

[1] P.-J. Touboul, M. Hennerici, S. Meairs, H. Adams, P. Amarenco, et al., Mannheim carotid intima-media thickness and plaque consensus (2004–2006-2011), Cardiovasc. Dis. 34 (4) (2012) 290–296.

[2] J. Marrugat, J. Vila, J.M. Baena-Díez, M. Grau, J. Sala, R. Ramos, I. Subirana, M. Fitó, R. Elosua, Validez relativa de la estimación del riesgo cardiovascular a 10 años en una cohorte poblacional del estudio REGICOR, Rev. Esp. Cardiol. 64 (15) (2011) 385–394.

[3] R.B. D'Agostino Sr., S. Grundy, L.M. Sullivan, P. Wilson, for the CHD Risk Prediction Group, Validation of the framingham coronary heart disease prediction scores: Results of a multiple ethnic groups investigation, JAMA 286 (2) (2001) 180–187.

[4] J. Marrugat, I. Subirana, R. Ramos, J. Vila, A. Marín-Ibañez, M.J. Guembe, F. Rigo, M.J. Tormo Díaz, C. Moreno-Iribas, J.J. Cabré, A. Segura, J.M. Baena-Díez, A.G. de la Cámara, J. Lapetra, M. Grau, M. Quesada, M.J. Medrano, P. González Diego, G. Frontera, D. Gavrila, E.A. Aicua, J. Basora, J.M. García, M. García-Lareo, J.A. Gutierrez, E. Mayoral, J. Sala, R. D'Agostino, R. Elosua, Derivation and validation of a set of 10-year cardiovascular risk predictive functions in Spain: The FRESCO study, Prev. Med. 61 (2014) 66–74.

[5] R. Conroy, K. Pyörälä, A. Fitzgerald, S. Sans, A. Menotti, G. De Backer, D. De Bacquer, P. Ducimetière, P. Jousilahti, U. Keil, I. Njølstad, R. Oganov, T. Thomsen, H. Tunstall-Pedoe, A. Tverdal, H. Wedel, P. Whincup, L. Wilhelmsen, I. Graham, on behalf of the SCORE project group, Estimation of ten-year risk of fatal cardiovascular disease in Europe: the SCORE project, Eur. Heart J. 24 (11) (2003) 987–1003.

[6] J. Hippisley-Cox, C. Coupland, Y. Vinogradova, J. Robson, R. Minhas, A. Sheikh, P. Brindle, Predicting cardiovascular risk in England and Wales: prospective derivation and validation of QRISK2, BMJ 336 (1475) (2006).

[7] M. Grau, J. Marrugat, Funciones de riesgo en la prevención primaria de las enfermedades cardiovasculares, Rev. Esp. Cardiol. 61 (4) (2008) 404–416.

[8] J. Marrugat, R. D'Agostino, L. Sullivan, R. Elosua, P. Wilson, J. Ordovas, P. Solanas, F. Cordón, R. Ramos, J. Sala, R. Masiá, W. Kannel, Comparing the predictive powers of survival models using Harrell's C or Somers' D, J. Epidemiol. Community Health 57 (8) (2003) 634–638.

[9] L. Saba, P.K. Jain, H.S. Suri, N. Ikeda, T. Araki, B.K. Singh, A. Nicolaides, S. Shafique, A. Gupta, J.R. Laird, et al., Plaque tissue morphology-based stroke risk stratification using carotid ultrasound: a polling-based PCA learning paradigm, J. Med. Syst. 41 (6) (2017) 98.

[10] T. Araki, P.K. Jain, H.S. Suri, N.D. Londhe, N. Ikeda, A. El-Baz, V.K. Shrivastava, L. Saba, A. Nicolaides, S. Shafique, et al., Stroke risk stratification and its validation using ultrasonic echolucent carotid wall plaque morphology: a machine learning paradigm, Comput. Biol. Med. 80 (2017) 77–96.

[11] M.-L.M. Gronholdt, B.G. Nordestgaard, T.V. Schroeder, S. Vorstrup, H. Sillesen, Ultrasonic echolucent carotid plaques predict future strokes, Circulation 104 (1) (2001) 68–73.

[12] Y. Irie, N. Katakami, H. Kaneto, M. Takahara, M. Nishio, R. Kasami, K. Sakamoto, Y. Umayahara, S. Sumitsuji, Y. Ueda, K. Kosugi, I. Shimomura, The utility of ultrasonic tissue characterization of carotid plaque in the prediction of cardiovascular events in diabetic patients, Atherosclerosis 230 (2) (2013) 399–405.

[13] N.N. Khanna, A.D. Jamthikar, D. Gupta, A. Nicolaides, T. Araki, L. Saba, E. Cuadrado-Godia, A. Sharma, T. Omerzu, H.S. Suri, et al., Performance evaluation of 10-year ultrasound image-based stroke/cardiovascular (CV) risk calculator by comparing against ten conventional CV risk calculators: a diabetic study, Comput. Biol. Med. 105 (2019) 125–143.

[14] C. Mitchell, C. Korcarz, A. Gepner, J. Kaufman, W. Post, R. Tracy, A. Gassett, N. Ma, R. McClelland, J. Stein, Ultrasound carotid plaque features, cardiovascular disease risk factors and events: The multi-ethnic study of atherosclerosis, Atherosclerosis 276 (2018) 195–202.

[15] E.C. Kyriacou, S. Petroudi, C.S. Pattichis, M.S. Pattichis, M. Griffin, S. Kakkos, A. Nicolaides, Prediction of high-risk asymptomatic carotid plaques based on ultrasonic image features, IEEE Trans. Inf. Technol. Biomed. 16 (5) (2012) 966–973.

[16] S.F. Weng, J. Reps, J. Kai, J.M. Garibaldi, N. Qureshi, Can machine-learning improve cardiovascular risk prediction using routine clinical data? PLOS ONE 12 (4) (2017) 1–14.

[17] R. Nakanishi, P. Slomka, R. Rios, J. Betancur, M. Blaha, K. Nasir, M. Miedema, J. Rumberger, H. Gransar, L. Shaw, A. Rozanski, M. Budoff, D. Berman, Machine learning adds to clinical and CAC assessments in predicting 10-year CHD and CVD deaths, JACC Cardiovasc. Imaging 14 (3) (2021) 615–625.

[18] B. Tamarappoo, A. Lin, F. Commandeur, P. McElhinney, S. Cadet, M. Goeller, A. Razipour, X. Chen, H. Gransar, S. Cantu, R. Miller, S. Achenbach, J. Friedman, S. Hayes, L. Thomson, N. Wong, A. Rozanski, P. Slomka, D. Berman, D. Dey, Machine learning integration of circulating and imaging biomarkers for explainable patient-specific prediction of cardiac events: A prospective study, Atherosclerosis (2021) 76–82.

[19] B. Ambale-Venkatesh, X. Yang, C.O. Wu, K. Liu, W.G. Hundley, R. McClelland, A.S. Gomes, A.R. Folsom, S. Shea, E. Guallar, et al., Cardiovascular event prediction by machine learning: the multi-ethnic study of atherosclerosis, Circ. Res. 121 (9) (2017) 1092–1101.

[20] D. Gonzalo-Calvo, R. Elosua, A. Vea, I. Subirana, S. Sayols-Baixeras, J. Marrugat, V. Llorente-Cortés, Soluble low-density lipoprotein receptor-related protein 1 as a biomarker of coronary risk: Predictive capacity and association with clinical events, Atherosclerosis 287 (2019) 93–99.

[21] S. Winther, S. Schmidt, B. Foldyna, T. Mayrhofer, L. Rasmussen, J. Dahl, U. Hoffmann, P. Douglas, J. Knuuti, M. Bøttcher, Coronary calcium scoring improves risk prediction in patients with suspected obstructive coronary artery disease, J. Am. Coll. Cardiol. 80 (21) (2022) 1965–1977.

[22] G.N.K. Reddy, M.S. Manikandan, R.B. Pachori, Automated Hilbert envelope based respiration rate measurement from ppg signal for wearable vital signs monitoring devices, in: 2022 International Conference on Artificial Intelligence of Things, ICAIoT, 2022, pp. 1–6, http://dx.doi.org/10.1109/ICAIoT57170.2022.10121855.

[23] A. Sharma, S. Patidar, A. Upadhyay, U.R. Acharya, Accurate tunable-Q wavelet transform based method for QRS complex detection, Comput. Electr. Eng. 75 (2019) 101–111, http://dx.doi.org/10.1016/j.compeleceng.2019.01.025, URL https://www.sciencedirect.com/science/article/pii/S0045790618322316.

[24] L. Gago, M. del Mar Vila, M. Grau, B. Remeseiro, L. Igual, An end-to-end framework for intima media measurement and atherosclerotic plaque detection in the carotid artery, Comput. Methods Programs Biomed. 223 (2022) 106954.

[25] N.E. Breslow, Analysis of survival data under the proportional hazards model, Int. Statist. Rev. / Rev. Int. Statist. 43 (1) (1975) 45–57.

[26] O. Ronneberger, P. Fischer, T. Brox, U-net: Convolutional networks for biomedical image segmentation, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, 2015, pp. 234–241.

[27] M. Tan, Q. Le, EfficientNet: Rethinking model scaling for convolutional neural networks, in: K. Chaudhuri, R. Salakhutdinov (Eds.), Proceedings of the 36th International Conference on Machine Learning, in: Proceedings of Machine Learning Research, vol. 97, 2019, pp. 6105–6114.

[28] F.E. Harrell, K.L. Lee, D.B. Mark, Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors, Statist. Med. 15 4 (1996) 361–387.

[29] H. Abdi, L.J. Williams, Principal component analysis, Wiley Interdiscip. Rev. Comput. Stat. 2 (4) (2010) 433–459.

[30] B. Carstensen, M. Plummer, E. Laara, M. Hills, Epi: A package for statistical analysis in epidemiology, 2021, R package version 2.44. URL https://CRAN.R-project.org/package=Epi.

[31] R Core Team, R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, 2021, URL https://www.R-project.org/.

[32] M. Grau, I. Subirana, D. Agis, R. Ramos, et al., Grosor íntima-media carotídeo en población española: valores de referencia y asociación con los factores de riesgo cardiovascular, Rev. Esp. Cardiol. 65 (12) (2012) 1086–1093.

[33] E. de Groot, G.K. Hovingh, A. Wiegman, P. Duriez, A.J. Smit, J.-C. Fruchart, J.J. Kastelein, Measurement of arterial wall thickness as a surrogate marker for atherosclerosis, Circulation 109 (23 suppl 1) (2004) III–33.
[34] F.E. Harrell Jr., R.M. Califf, D.B. Pryor, K.L. Lee, R.A. Rosati, Evaluating the yield of medical tests, JAMA 247 (18) (1982) 2543–2546.
[35] M.J. Pencina, R.B. D'Agostino, R.S. Vasan, Statistical methods for assessment of added usefulness of new biomarkers, Clin. Chem. Lab. Med. 48 (12) (2010) 1703–1711.
[36] E. Inoue, Nricens: NRI for risk prediction models with time to event and binary response data, 2018, R package version 1.6. URL https://CRAN.R-project.org/package=nricens.
[37] G. James, D. Witten, T. Hastie, R. Tibshirani, An Introduction to Statistical Learning with Applications in R, Springer, 2017.