Computers in Biology and Medicine 142 (2022) 105204

Contents lists available at ScienceDirect

Computers in Biology and Medicine

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

Understanding the bias in machine learning systems for cardiovascular disease risk assessment: The first of its kind review

Jasjit S. Suri [a,*], Mrinalini Bhagawati [b], Sudip Paul [b], Athanasios Protogeron [c], Petros P. Sfikakis [d], George D. Kitas [e], Narendra N. Khanna [f], Zoltan Ruzsa [g], Aditya M. Sharma [h], Sanjay Saxena [i], Gavino Faa [j], Kosmas I. Paraskevas [k], John R. Laird [l], Amer M. Johri [m], Luca Saba [n], Manudeep Kalra [o]

[a] Stroke Diagnostic and Monitoring Division, AtheroPoint™, Roseville, CA, USA
[b] Department of Biomedical Engineering, North-Eastern Hill University, Shillong, India
[c] Department of Cardiovascular Prevention & Research Unit Clinic & Laboratory of Pathophysiology, National and Kapodistrian University of Athens, Greece
[d] Rheumatology Unit, National Kapodistrian University of Athens, Greece
[e] Arthritis Research UK Centre for Epidemiology, Manchester University, Manchester, UK
[f] Department of Cardiology, Indraprastha APOLLO Hospitals, New Delhi, India
[g] Semmelweis University, Budapest, Hungary
[h] Division of Cardiovascular Medicine, University of Virginia, Charlottesville, VA, USA
[i] Department of CSE, International Institute of Information Technology, Bhubaneswar, India
[j] Department of Pathology, A.O.U., di Cagliari -Polo di Monserrato s.s, Cagliari, Italy
[k] Department of Vascular Surgery, Central Clinic of Athens, N. Iraklio, Athens, 14122, Greece
[l] Cardiology Department, St. Helena Hospital, St. Helena, CA, USA
[m] Department of Medicine, Division of Cardiology, Queen's University, Kingston, Canada
[n] Department of Radiology, University of Cagliari, Italy
[o] Department of Radiology, Massachusetts General Hospital, Boston, MA, USA

ARTICLE INFO

Keywords:
Risk prediction
Coronary artery disease
Carotid ultrasound
Artificial intelligence
Bias
Gold standard
And unseen data

ABSTRACT

Background: Artificial Intelligence (AI), in particular, machine learning (ML) has shown promising results in coronary artery disease (CAD) or cardiovascular disease (CVD) risk prediction. Bias in ML systems is of great interest due to its over-performance and poor clinical delivery. The main objective is to understand the nature of risk-of-bias (RoB) in ML and non-ML studies for CVD risk prediction.

Methods: PRISMA model was used to shortlisting 117 studies, which were analyzed to understand the RoB in ML and non-ML using 46 and 32 attributes, respectively. The mean score for each study was computed and then ranked into three ML and non-ML bias categories, namely *low-bias* (LB), *moderate-bias* (MB), and high-bias (HB), derived using two cutoffs. Further, bias computation was validated using the analytical slope method.

Results: Five types of the gold standard were identified in the ML design for CAD/CVD risk prediction. The low-moderate and moderate-high bias cutoffs for 24 ML studies **(5, 10**, and **9** studies for each LB, MB, and HB) and 14 non-ML (**3, 4,** and **7** studies for each LB, MB, and HB) were in the range of **1.5** to **1.95**. $\mathbf{Bias}_{ML} < \mathbf{Bias}_{non\text{-}ML}$ by **~43%**. A set of recommendations were proposed for lowering RoB.

Conclusion: ML showed a lower bias compared to non-ML. For a robust ML-based CAD/CVD prediction design, it is vital to have **(i)** stronger outcomes like death or CAC score or coronary artery stenosis; **(ii)** ensuring scientific/clinical validation; **(iii)** adaptation of multiethnic groups while practicing unseen AI; **(iv)** amalgamation of conventional, laboratory, image-based and medication-based biomarkers.

1. Introduction

Cardiovascular disease (CVD) is one of the leading causes of mortality with 18 million deaths worldwide [1]. The ultimate cause of CVD is atherosclerotic vascular disease and its formation in the heart's coronary arteries [2]. The CVD risk worsens due to comorbidities like diabetes [3], chronic kidney disease [4,5], rheumatoid arthritis [6,7],

* Corresponding author. Stroke Monitoring and Diagnostic Division, AtheroPoint™, Roseville, CA, 95661, USA.
E-mail address: jasjit.suri@atheropoint.com (J.S. Suri).

https://doi.org/10.1016/j.compbiomed.2021.105204
Received 28 November 2021; Received in revised form 29 December 2021; Accepted 29 December 2021
Available online 4 January 2022

Acronym table

*Abb*Definition*	
ACC	American College of Cardiology
AHA	American Heart Association
ANOVA	Analysis of variance
ASCVD	Atherosclerotic cardiovascular disease
AUC	Area-under-the-curve
BMI	Body mass index
CAC	Coronary artery calcification
CAD	Coronary artery disease
CAS	Coronary artery syndrome
CCVRC	Conventional cardiovascular risk cal#
CHD	Coronary Heart Disease
cIMT	Carotid Intima-Media Thickness
CKD	Chronic kidney disease
CT	Computed Tomography
CUSIP	Carotid ultrasound image phenotype
CV	Cross-validation
CVD	Cardiovascular disease
CVE	Cardiovascular events
DL	Deep learning
DM	Diabetes mellitus
EEGS	Event-equivalent gold standard
ESC	European Society of Cardiology
FH	Family history
FNR	False-negative rate
FPR	False-positive rate
FRS	Framingham risk score
GUI	Graphical user interface
HTN	Hypertension
IMTV	Intima-media thickness variability
IPN	Intraplaque neovascularization
LBBM	Laboratory-based biomarker
LSTM	Long short-term memory network
MedUSE	Medication use
ML	Machine learning
MPH	Maximum plaque height
MRI	Magnetic Resonance Imaging
NPV	Negative predictive value
Non-ML	Non-machine learning
OBBM	Office-based biomarker
PE	Performance evaluation matrices
PPV	Positive predictive value
PTC	Plaque tissue characterization
QRISK3	QResearch cardiovascular risk algorithm
RA	Rheumatoid arthritis
RF	Random forest
RoB	Risk of bias
ROC	Receiver operating-characteristics
RRS	Reynolds risk score
SCORE	Systematic coronary risk evaluation
SMOTE	Synthetic minority over-sampling technique
SVM	Support vector machine
TPA	Total plaque area
US	Ultrasound
WHO	World Health Organization
Abb*	Abbreviation
#	Calculator

hypertension [8], and high-lipids [9]. This puts patients with heart disease and stroke at greater risk, and therefore there is a need for early CVD risk detection.

There are popular calculators used for CVD risk assessment such as the QRISK3 calculator [10], Framingham risk score (FRS) [11], the systematic coronary risk evaluation score (SCORE) [12], the Reynolds risk score (RRS) [13], and the atherosclerosis cardiovascular disease (ASCVD) [14]. Typically, these calculators use conventional and indirect risk factors (or office-based biomarkers (OBBM)) and laboratory-based biomarkers (LBBM) [15,16]. These risk prediction algorithms follow specific guidelines such as the European Society of Cardiology (ESC) and the American College of Cardiology/American Heart Association (ACC/AHA) [17,18]. However, these CVD calculators offer challenges such as (i) being unable to handle the non-linear relationship between these risk factors (or covariates) and outcomes and (ii) lacking the CVD risk of granularity [19–21]. Thus, there is a clear need for a more accurate CVD risk assessment solution that can measure true atherosclerotic non-invasive plaque burdens along with OBBM and LBBM, rather than indirect risk factors.

Recent research suggests that coronary artery disease (CAD) risk prediction can be determined via the carotid artery framework. Because of the same genetic makeup in these two sets of arteries, this gives an added advantage of both CVD and stroke risk assessments [22–24]. Because of this, carotid artery imaging has been quite regularly adapted as a surrogate biomarker for CVD risk assessment [25–30]. Generally, the three major medical imaging modalities, such as magnetic resonance imaging (MRI) [31], computed tomography (CT) [32], and ultrasound (US) [33,34] have been adapted for screening the carotid artery [35]. Among these, B-mode carotid US offers several advantages such as small footprint, economic, superior ergonomics, easy access via neck window, and ability to produce high-resolution images (due to compound and harmonic image reconstruction) [36,37]. These imaging modalities can be used for extracting carotid ultrasound image-based phenotypes (CUSIP), such as carotid intima-media thickness (cIMT), intima-media thickness variability (IMTV), maximum plaque height (MPH) [38], total plaque area (TPA) derived from common carotid artery or carotid bulb [39–41], which can be used for CVD risk computation [42–46]. These image-based CUSIP biomarkers, when fused with OBBM and LBBM further improve the reliability of the CVD risk assessment system as shown by AtheroEdge 2.0 (Roseville, CA, USA) [6,7,20,44,47–49]; however, these biomarkers do not address the non-linearity between the biomarkers and outcomes. Thus, we need a more advanced solution that can handle this challenge.

Artificial Intelligence (AI) has opened the door in all walks of life, particularly in the healthcare sector, and more precisely in medical imaging [50,51]. Machine learning (ML), the subcategory of AI has further dominated the field of the medical imaging industry and recently in diagnosing diabetes [52,53], and several types of cancers namely, liver [54–56], thyroid [57,58], coronary [48,59,60], prostate [61], ovarian [62], and skin [63,64]. Since the covariates and outcomes can handle the non-linearity, these ML models are more suited for CVD risk computations. Other applications of ML include plaque tissue characterization (PTC) for stroke risk, particularly for the classification of symptomatic *vs*. asymptomatic plaque [65–68]. These ML methods have shown higher classification accuracy or higher performance; however, they lack (a) robust outcome design (choice of the gold standard) and (b) the clinical delivery component while over-performing the accuracy of the ML system, which in turn leads to bias in ML.

This study analyzes the bias in these ML-based and non-ML-based studies for CVD risk assessment. We discovered five types of outcomes (gold standard) for ML design such as (i) death [69–72], (ii) coronary artery calcification (CAC) or coronary artery stenosis [73–79], (iii) different heart conditions like myocardial infarction, stroke, angina, and coronary artery syndrome (CAS) [80–86], (iv) chronic disease like

diabetes, hypertension, and chronic kidney disease (CKD) [87–89], and (v) event-equivalent gold standard (EEGS) [90–92], that acts as an indirect measurement of outcomes. As part of the comparative study, we also compute the bias in non-ML studies [5,11,15,19,26,93–101], which computes CVD risk using (a) CUSIP such as AECRS 1.0 score [101–104] and (b) the 10-year risk assessment strategies [97,98,105–107]. The bias cutoff is then determined, and the studies are ranked based on mean aggregate score, analyzed and finally, recommendations are presented to lower the risk of bias (RoB) in ML and non-ML systems. While we hypothesize that bias in ML is lower than non-ML, we further demonstrate this using the analytical slope method. We finally present a set of recommendations for lowering the RoB.

The layout of this study is as follows. Section 2 presents the PRISMA model for the selection of the citations. Along with this, the section discusses the statistical distributions of the AI parameters used in this study. Section 3 shows the ML architectures used for CVD risk assessment. The central component, of this study, involves the computation of bias in ML where the aggregate score is computed for all the ML-based studies, which is then used for determining the bias cutoff. This is shown in section 4. CVD risk analysis and comparison between ML *vs*. non-ML is discussed in section 5. We demonstrate in section 6, the hypothesis that ML-based techniques are less biased compared to non-ML using the analytical slope method. Section 7 shows the distribution of bias by AI attributes in ML and recommendations for reduction in RoB. The critical discussion is presented in section 8, and finally, the study concludes in section 9.

2. Search strategy and statistical distributions

2.1. PRISMA model

A PRISMA search was performed by using IEEE Xplore, PubMed, Google Scholar, and Science Direct. The main keywords used for selecting studies were "CVD risk prediction using Artificial Intelligence", "CVD risk assessment in AI framework", "CVD risk prediction using carotid", "CVD risk prediction using ML", "CVD risk prediction". Fig. 1 shows the PRISMA model with all of the AI/ML studies used in this study. After an intensive search, a total of 18,561 studies were identified, and duplicates studies were removed. After this systematic process, 15,000 studies were retained. There were three exclusion criteria (i) non-relevant articles, (ii) records excluded after screening of articles, and (iii) studies with insufficient data were excluded. After the application of the exclusion criteria, 14,815, 62, and 6 studies were excluded which are marked as E1, E2 (non-AI), and E3 in Fig. 1. The crucial and relevant information from these selected studies was extracted, and their statistical distributions were plotted and further linked to bias.

2.2. Statistical distributions

Fig. 2 shows the statistical distribution of (a) publication trend-by-year (b) ML *vs*. non-ML (c) types of outcomes or gold standard (d) risk granularity types (binary, tertiary, and quaternary). Fig. 2 (a) displays the trend of ML publications for CVD risk over the last 10-years. From 2016 it shows an increasing trend of publications till 2021 except for the year 2019, but there is no narrative review of the studies on the bias in ML for CVD risk prediction. Of all the selected 117 CVD-related articles, 24 were pure ML, and 14 were pure non-ML. Fig. 2 (b) shows the 60% studies used ML [69–92] and 40% studies used non-ML studies [5,11,15,19,26,93–101]. ML-based CVD calculators were analyzed based on the outcome (ground truth) design, which was divided into five clusters (Fig. 2 (c)) such as death 17% [69–72], coronary artery calcification (CAC), which was 24% [73–79], myocardial infarction (MI) along with stroke and angina which was 29% [80–86], diabetes along with hypertension which was 12% [87–89] and the event-equivalent gold

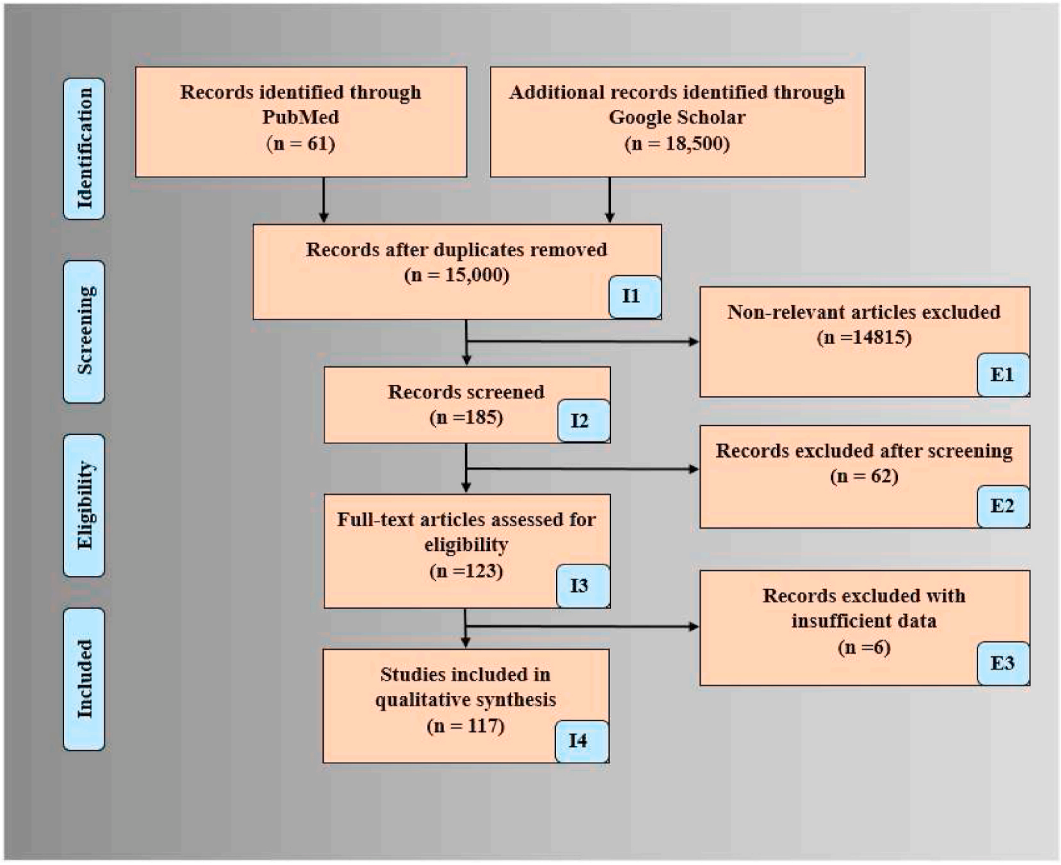


Fig. 1. PRISMA Analysis for selected studies.

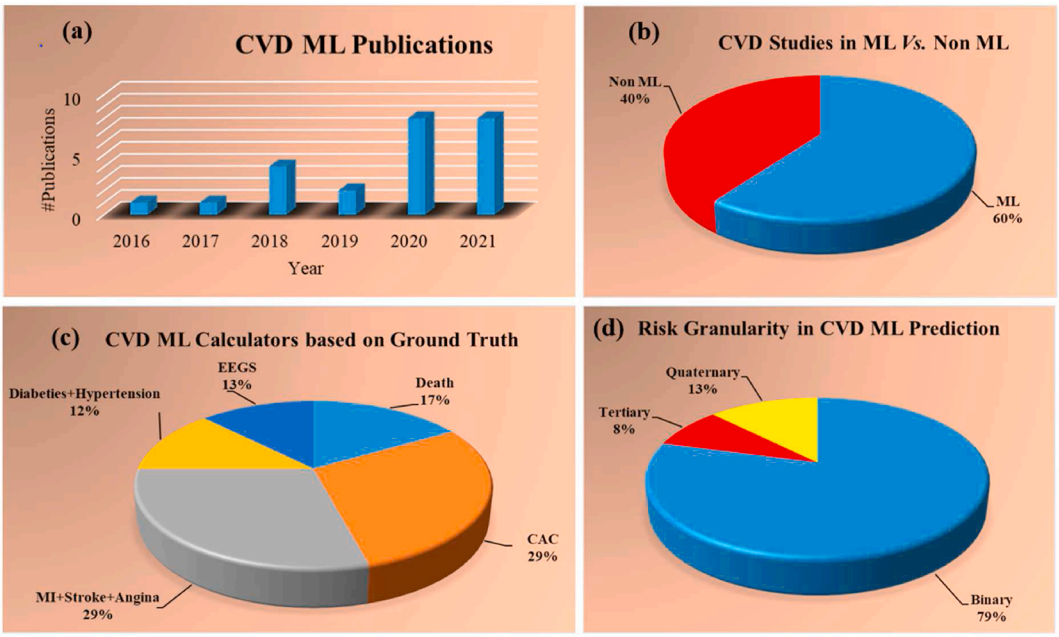


Fig. 2. Statistical distribution (a) publication trend by year (b) ML *vs.* non-ML (c) types of outcomes or gold standard (d) risk granularity types (binary, tertiary, and quaternary). (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

standard (EEGS) which was 13% [90–92]. Since the outcomes of the CVD ML calculators play a crucial role in determining the reliability and stability of the ML design, it is therefore important to choose the outcome designs (gold standard) to have minimal impact on the bias of ML systems. Risk granularity (or type of classes) was also analyzed in the CVD ML prediction. Three classes were identified such as binary, tertiary, and quaternary. Fig. 2 (d) shows the contribution consisting of 77% binary [70–72,76,78–90,92], 9% tertiary [75,91] and 14% quaternary [69,74,77]. Even though granularity does not hamper the performance of the ML system directly, it could introduce a bias during the monitoring of statins on CVD patients. This is because multiclass CVD risk assessments are more effective since the drug delivery is more refined. It is, therefore, necessary to have risk granularity as one of the attributes for overall bias in ML analysis.

The type of classifier determines the performance of the ML model. Fig. 3 shows the distribution of (i) different classifiers, (ii) type of features, (iii) types of covariate clusters, and (iv) types of optimizations used in AI framework. Fig. 3 (a) shows the percentage distribution of the classifiers used in the ML framework consisting of 5% support vector machine (SVM) + synthetic minority oversampling technique (SMOTE) [72], 20% random forest (RF) [81,85,87,92], 5% K-nearest neighbour (KNN) [80], 10% decision tree [83], 5% neural network (NN) [92], 5% principal component analysis (PCA) [91], 5% RF along with Adaboost [86], 10% ML [69,85], 10% deep learning (DL) [73,78], 5% stochastic

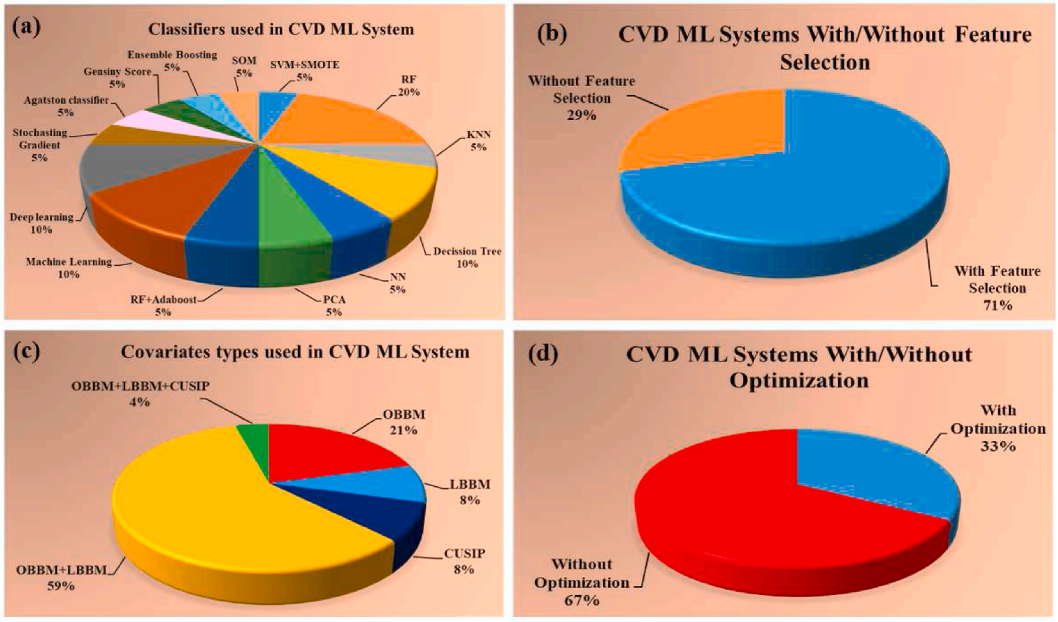


Fig. 3. Statistical distribution (a) types of classifiers; (b) adaption of feature selection during ML; (c) types of risk factors (covariates); and (d) optimization applied to ML.

gradient [70], 5% Agatston classifier [74], 5% Gensiny score [76], 5% ensemble boosting [75], and 5% self-organization map (SOM) [89]. The ML systems were also analyzed without and with feature selection. Fig. 3 (b) shows the percentage distribution, where 29% was without features [72,74,77,79,83,85,89,90] and 71% was with selected features [69–71, 75,76,80–82,84,86–88]. The statistical distribution of the covariates such as OBBM, LBBM, and CUSIP is shown in Fig. 3 (c). OBBM took 21% [71,78,86,90,92], LBBM took 8% [82,83], CUSIP took 8% [70,74], OBBM along with LBBM took 59% [72,73,75–77,79,81,84,85,87–89, 91], OBBM, LBBM, and CUSIP all combined took 4% [69]. Fig. 3 (d) shows the characterization, where 33% of the ML studies were optimized [71,72,74,75,81,83,91,92] and 67% of the ML studies were not optimized [69,70,73,76–80,82,84–90,92]. Three out of four attributes (such as classifier types, selected features, and optimizations of ML systems) were the three legs of the stool that stabilizes the ML system design to prevent it from being biased in ML designs. It is therefore important to study these ML attributes for optimal performance. The clinical covariates such as OBBM, LBBM, CUSIP, and medications are equally important clusters for training the ML systems, therefore these covariates must be chosen appropriately to have minimal impact on the ML bias.

3. Machine learning architecture for CVD risk assessment

The fundamental design in the ML system consists primarily of (a) model generation using the CVD risk factors on training cases and (b) prediction of risk labels (risk probabilities of the classes) on test cases. So these training models are based on (i) the choice of outcome (gold standard) designs, (ii) the type of the classifiers, (iii) the clusters of training covariates (risk factors), (iv) the objective of risk prediction for short term or long term (10-years), (v) the type of the cross-validation protocol (such as 10-fold or 5-fold), (vi) the scientific validation, and (vii) clinical evaluation. These factors play a crucial role in determining the overall performance and pitfalls leading to bias. Therefore, the objective of this section is to understand the different kinds of ML architectures keeping the above framework in mind.

Three architectures are presented in this section covering different aspects of the ML design and clinical delivery. Fig. 4 shows a typical ML architecture used for CVD risk prediction with the concept of data partition after pre-processing of the data using SMOTE algorithm. This was taken from AtheroEdge™ 3.0 system (AtheroPoint, CA, USA) [69]. It basically has two sections: (i) the offline training block and (ii) the online prediction of CVD risk. The block showing ten-fold cross--validation divides the balanced datasets into two subsets (a) training

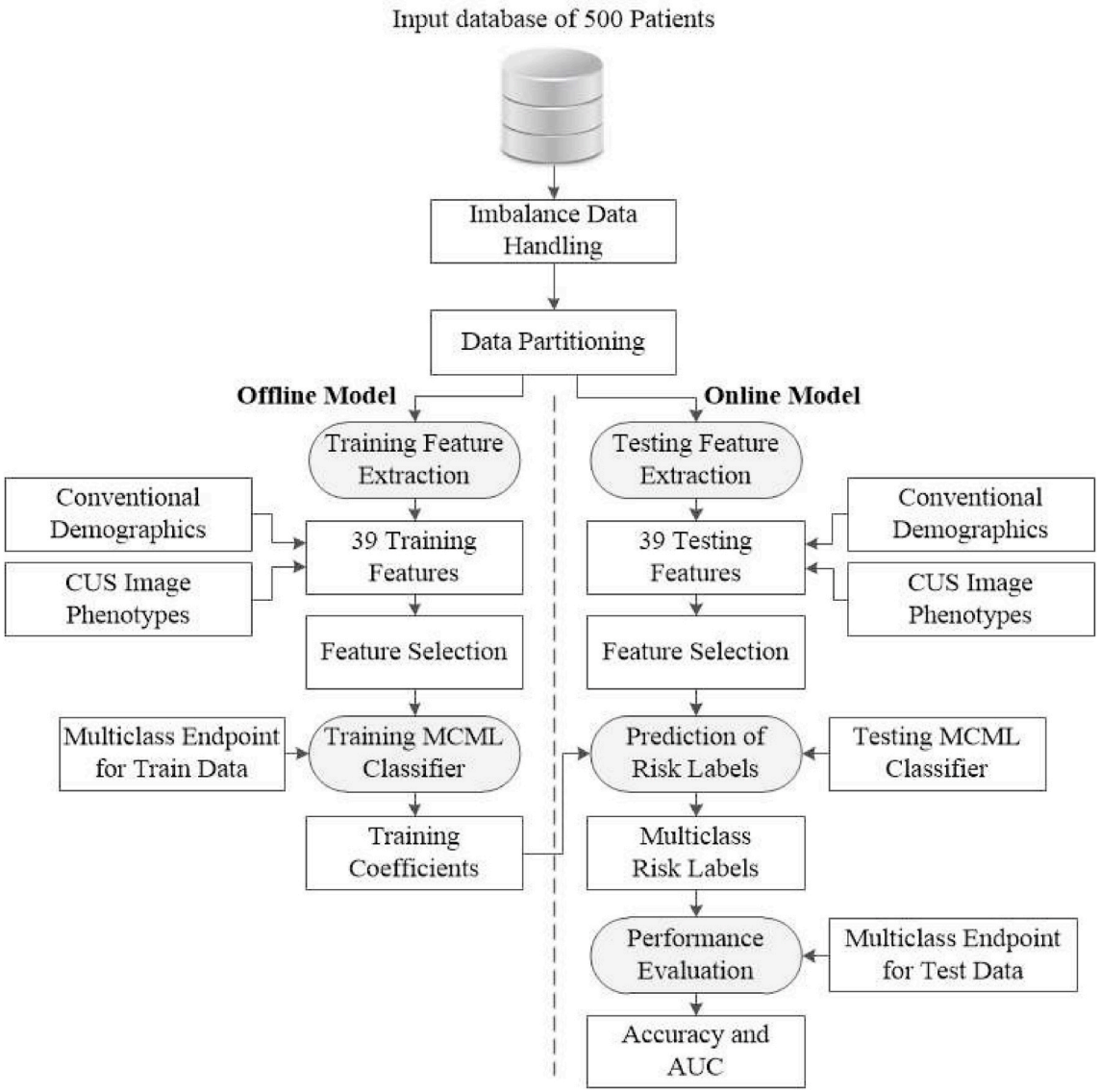


Fig. 4. The AtheroEdge 3.0 ML-system (Courtesy of AtheroPoint, Roseville, CA, USA) [69].

dataset and (b) testing dataset. The offline system consists of training model generation using the combination of training covariates (risk factors) and the gold standard (which could be a binary class or multiple classes). An example of a binary class is low- and high-risk labels, while in the case of the multiclass framework, one can take coronary angiographic score (CAS) consisting of four grades (low, moderate, mild, and high-risk). The model generation is accomplished using standardized classifiers such as XGBoost, support vector machine (SVM), random forest (RF), and neural networks (NN). The CVD risk is predicted during the online system by transforming the test risk factors (on test patients) by the training model coefficients (right side of Fig. 4). One can also adapt the benchmarking protocol where the CVD predicted risk from the ML system could be compared against the conventional calculators such as FRS, the SCORE, and ASCVD. Thus, we conclude that the classifiers' design outcomes and choice play a vital role in the ML-based CVD systems. Thus, these can be considered as ML attributes for computing the bias in ML.

Another category of architecture is when the experimental data and validation data are different. An example would be when the training is on data type A and testing on data type B, and vice-versa. A similar paradigm was adopted in Ref. [72]. The architecture of a similar approach is shown in Fig. 5. The added benefit of this paradigm is the role of follow-up to predict the CVD risk at the 10th year, given the baseline of the current year. In this design, the authors use a dataset having 13-year follow-up information from the Multi-Ethnic Study of Atherosclerosis (MESA) data. This dataset consists of 6459 atherosclerotic CVD-free participants at baseline. Moreover, for validation, the Flemish Study of Environment, Genes, and Health Outcomes (FLEMENGHO) cohort was used. Note that the system was data augmented by the use of NEATER, a technique for filtering oversampled data by the use of non-cooperative game theory as the system had unbalanced classes. The training system used an SVM classifier along with the outcome labels for model generation. The same classifier was used during the prediction model. Since this is a cross-validation paradigm, databases were swapped for the training and testing sets for unbiased results, followed by the mean computations of the predicted values. Since the partition was randomly done, therefore one can use ten trials for better mean statistics. Thus, we conclude that (a) the design outcomes and (b) the choice of the classifiers play a vital role in ML-based CVD systems. One can therefore consider these as ML attributes for

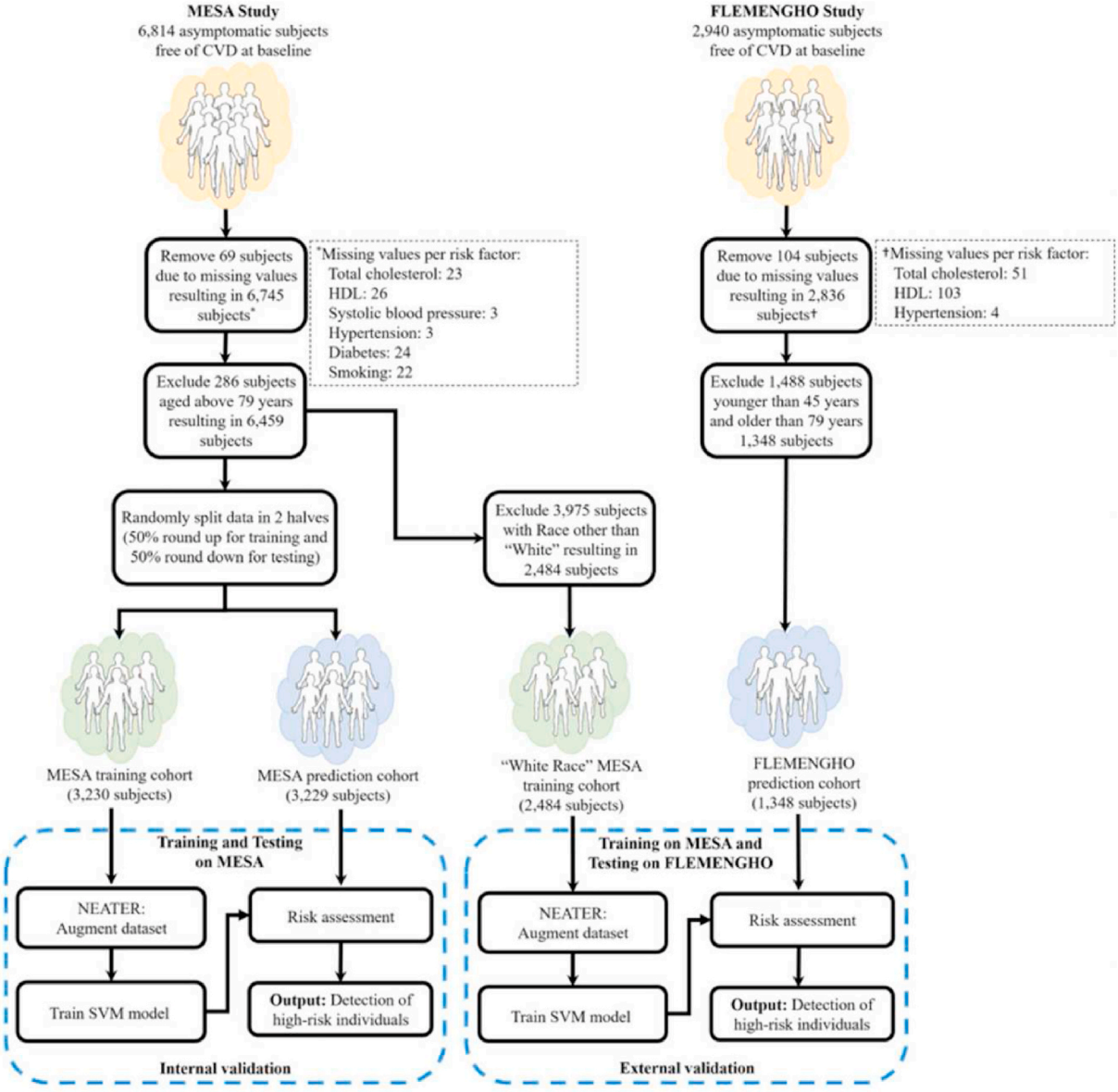


Fig. 5. ML-system for 13 years follow-up CVD risk prediction (Courtesy of Journal of American Heart Association, Open Access) [72].

computing the bias in ML.

The next set of architecture (Fig. 6) predicts the 10-year CVD risk outcomes given the baseline covariates and benchmarking against the conventional CVD risk calculator such as HelenicSCORE. A set of three models was developed to distinguish a patient into high or low CVD risk groups by the use of different types of classifiers, namely k-NN, decision tree (DT), and RF [80]. Generally, the ML algorithms operate in two stages: (a) training of the model through a random set of data (training set), and (b) testing or evaluation of the model on the remaining data (testing set). The model should be trained by using a training set to estimate all the features in the training set. Therefore, selecting the features is very crucial and should portray the original dataset which means the selection or sampling should be random and not be biased in any manner. Stratified k-fold cross-validation (CV) is a commonly used mechanism for sampling. In this, the whole dataset is divided into "k" segments with equal size (folds), that obtain the ratio between classes, and then "k" independent iteration of validation/training was performed. Even though the system was designed using ML-based for risk prediction, the system lacks risk granularity. Thus, we conclude that the design outcomes, choice of the classifiers play a vital role in ML-based CVD systems. From the above discussion, we firmly believe there is a need to estimate the RoB in ML systems, and the ML attributes can be utilized for framing the design for the CVD bias risk estimation.

4. ML bias computation: scoring attributes, determining bias cutoffs, and ranking

The main objective of this study is to automatically detect the RoB in the ML-based studies that compute the CVD risk assessment. Thus, the first step is to estimate in which "bias bin" the study lies. These bias bins are estimated by determining the cutoffs which divide the "mean AI scores" in the ML studies. We broadly create three bins, such as low-, moderate- and high-bias, using two kinds of cutoffs based on the AP (**ai**) Bias 1.0 (AtheroPoint, Roseville, CA, USA). Thus, the first step needs the mean AI score value for each study, which is computed by aggregating the "individual AI scores for each attribute", divided by the "total AI attributes". Note that the individual AI score per attribute is decided using the grading scheme based on an AI expert having a minimum of ten to fifteen years of experience.

The first cutoff is determined by computing the intersection of the "mean score plot" (in decreasing order for all the participating ML studies) with a "cumulative frequency plot" (CFP) curve. The first intersection point is the *low-moderate* (LM) cutoff, while the *moderate-high* (MH) cutoff was empirically computed based on the slope transitions of the mean score plot of the participating studies. These two cutoffs then divide the total number of studies into low-, moderate-, and high-bias bins. Using this strategy, one can rank the ML-based studies and color code them demarcating the *low-*, *moderate-* and *high-bias* studies from green, yellow-to-orange, and red. These partitioned studies can then be analyzed for improving the RoB in ML-based CVD risk assessment systems.

The grading scheme used for scoring the AI study is presented in Table A.1 (Appendix A). This table shows 46 AI attributes clubbed broadly into nine clusters namely (i) demographics, (ii) AI architecture design, (iii) optimization, (iv) performance parameter, (v) scientific validation, (vi) statistical test, (vii) clinical validation, (viii) clinical evaluation, and (ix) survival and hazard analysis. Each of these attributes has independent criteria for grading and scoring 24 ML studies that are self-explanatory and presented in Table A.1 [67–90]. Thus, one can compute the LM and MH cutoffs as shown in Fig. 7. As seen in this figure, the LM and MH cutoffs are **1.83** and **1.59**, respectively. Accordingly, the *low-bias* group has **5** ML studies [69,72,75,76,87], *moderate-bias* group has **10** ML studies [70,71,78,80,81,85,86,88,89,92] and *high-bias* group contains **9** ML studies [71,72,75,77,80–82,88,89]. Table 1 shows the ranking of the 24 ML-based CVD studies. The ranking strategy is a powerful paradigm for automatically isolating and

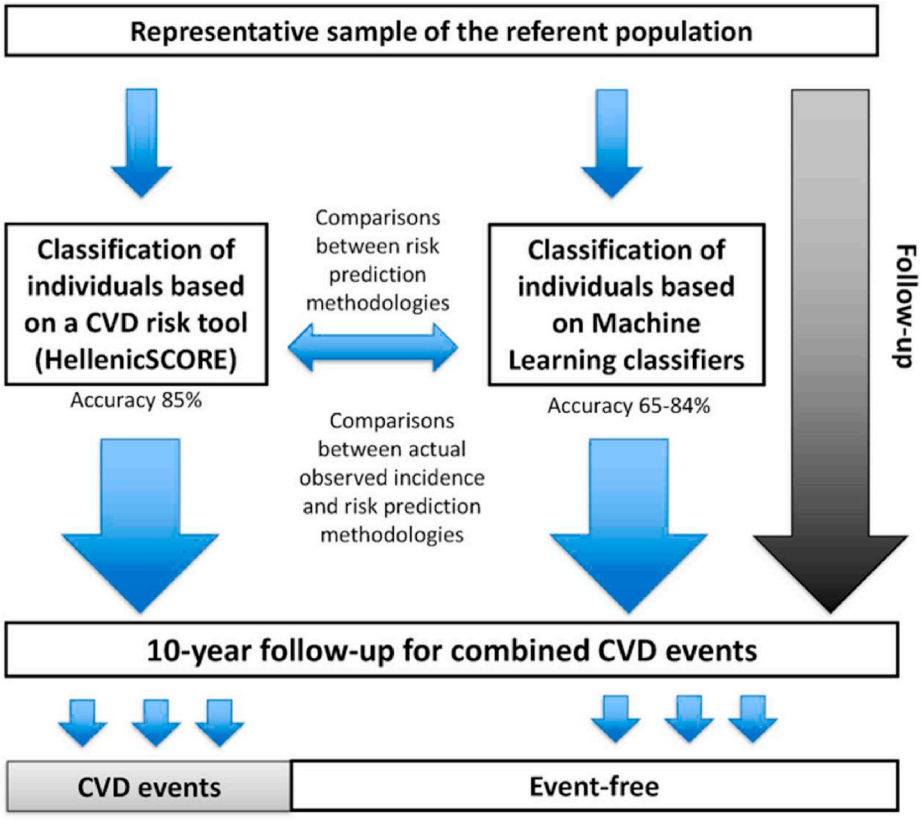


Fig. 6. The ML-system for 10-year follow-up CVD risk prediction (Courtesy of Journal of BMC Med Res Methodol., Open Access) [80].

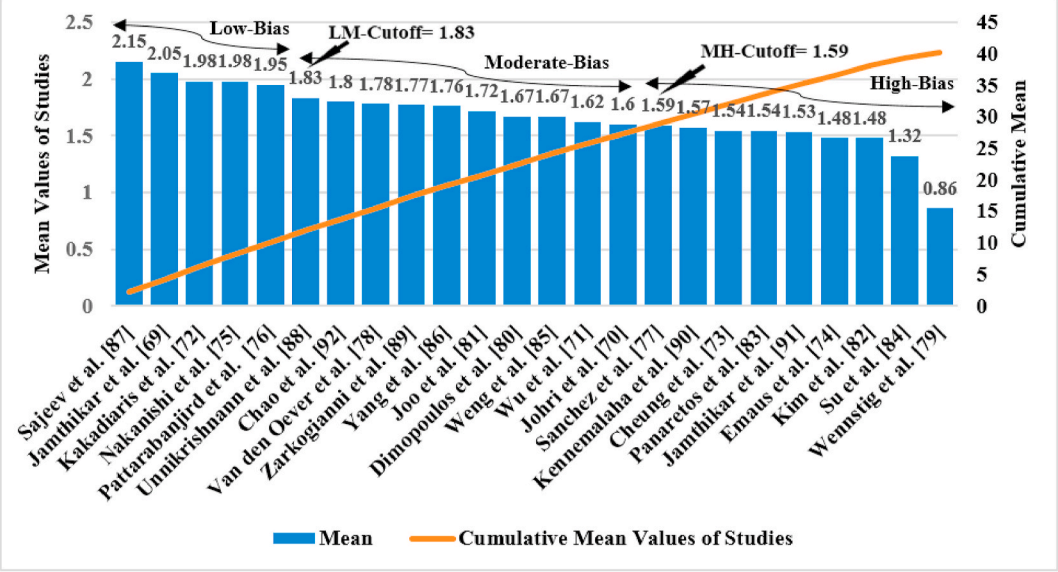


Fig. 7. Ranking for the 24 ML studies. **LM**: *Low-moderate* cutoff; **MH**: *Moderate-high* cutoff.

identifying the RoB in ML studies once the ML scores to the attributes are assigned. The same ranking algorithm paradigm was applied to non-ML-based studies, computing the cutoffs, bias bins, and detecting the non-ML studies in the three-bin classes, to be presented in the next section.

5. Understanding AI attribute behaviour: ML *vs.* non-ML

The objective of this section is to understand in a deeper way how the AI attributes are involved in affecting the bias in the ML algorithms.

Table 1
Ranking table derived using Tables A.1 and Table A.2.

		A47	A48	A49	A50
SN	**Studies**	**Sum**	**Mean**	**Ranking**	**STDDEV**
1	Sajeev *et al.* [87]	99	2.15	1	2.03
2	Jamthikar *et al.* [69]	94.5	2.05	2	2.15
3	Kakadiaris *et al.* [72]	91	1.98	3	2.04
4	Nakanishi *et al.* [75]	91	1.98	4	2.07
5	Pattarabanjird *et al.* [76]	89.5	1.95	5	2.09
6	Unnikrishnan *et al.* [88]	84	1.83	6	1.9
7	Chao *et al.* [92]	83	1.8	7	1.94
8	Van den Oever *et al.* [78]	82	1.78	8	1.94
9	Zarkogianni *et al.* [89]	81.5	1.77	9	1.95
10	Yang *et al.* [86]	81	1.76	10	1.98
11	Joo *et al.* [81]	79	1.72	11	1.87
12	Dimopoulos *et al.* [80]	77	1.67	12	1.86
13	Weng *et al.* [85]	77	1.67	13	1.87
14	Wu *et al.* [71]	74.5	1.62	14	1.92
15	Johri *et al.* [70]	73.5	1.6	15	1.92
16	Sanchez *et al.* [77]	73	1.59	16	1.86
17	Kennemalaha *et al.* [90]	72	1.57	17	1.9
18	Cheung *et al.* [73]	71	1.54	18	1.88
19	Panaretos *et al.* [83]	71	1.54	19	1.88
20	Jamthikar *et al.* [91]	70.5	1.53	20	1.8
21	Emaus *et al.* [74]	68	1.48	21	1.8
22	Kim *et al.* [82]	68	1.48	22	1.79
23	Su *et al.* [84]	60.5	1.32	23	1.61
24	Wennstig *et al.* [79]	39.5	0.859	24	1.24

Thus, this section concentrates on the statistical distribution of the attributes in cluster formation. We actually split the clusters and zero-in into each attribute and correlate all the 24 ML studies for each ML attribute. There are two components to this section. In part A, we simply study the characteristics of each of the three bins once the ranking has been determined. In part B, we understand the behaviour of these attributes.

5.1. Characteristics of three cluster bias bins

Low-Bias Cluster Studies: This cluster had **5** ML studies [69,72,75,76,87]. The general characteristics had more sophisticated use of AI architecture, and supervised techniques for developing training ML models. ML pipeline consisted of feature extraction [69,87] and application of top-to-the-notch classification techniques.

Moderate-Bias Cluster Studies: This cluster consisted of **10** ML studies [70,71,78,80,81,85,86,88,89,92]. The general characteristics of this cluster primarily consisted of moderate sophisticated use of AI architecture, the unsupervised paradigm for training the ML model, partially adopting feature extraction methods, and usage and limited of popular classification models. The cross-validation performance parameters were not as high as low-bias cluster studies. Thus, the mean score values of this cluster were relatively lower compared to the *low-biased* cluster.

High-Bias Cluster Studies: This cluster had **9** ML studies [73,74,77,79,82–84,90,91]. The general characteristics of the studies in this cluster include non-robust AI architecture, lack of feature extraction process, and further only statistical classification techniques were applied [79,82]. These studies were also short of cross-validation protocol [74,84]. Thus, these studies secured low ranks among all the ML participating studies.

5.2. Understanding mean AI attribute behaviour over all the ML studies

The AI attributes were compared between them to see which contributed to the *low-bias*, *moderate-bias*, and high-bias groups. The aggregate mean of each attribute was computed among all the 24 ML studies and for 14 non-ML studies. Then the percentage contribution for each AI attribute was based on the mean AI score for each attribute. These were plotted in bar charts to better understand, as shown in Fig. 8 and Fig. 9.

Fig. 8 shows the attributes of four cluster types (in %) such as (a) demographical risk factors, (b) attributes of AI architecture (c)

prevalence of optimization paradigm, and (d) attributes for the performance parameters. The corresponding clusters using absolute mean scores are exhibited in Figure A.1 (Appendix A). In demographical attributes (Fig. 8 (a)), the smoking attribute was used a maximum number of times (**86.6%**) in the ML studies. The same figure shows the decreasing order of the remaining demographical risk factors in the following order such as hypertension (76.6%), non-image data (76.6%), body mass index (BMI) (70%), family history (FH) (63.4%), sample size > 5K (K~1000) (48.4%), covariate type >2 (48.4%), ethnicity (either mono or multi-ethnic) (38.4%), and data collection from multiple institutes (26.6%). This can be seen in the following studies [69,72,75,76,87,88].

In the AI architecture attribute cluster (Fig. 8 (b)), the feature extraction attribute obtained the highest percentage among all the attributes (**76.6%**) which means most of the studies performed feature extraction for the ML system [69,75,76,87,88,92]. Other attributes in this cluster follow the decreasing order as five types of GT (**65%**), five classifier types (**58.4%**), number of architecture used (53.4%) in each study, number of classifiers used (**46.4%**), usage of seen data in AI for CV (**40%**), training technique (supervised or unsupervised) (**34.2%**), number of classes used (binary, tertiary, quaternary) (**33.4%**), use of feature selection process (**30%**), type data used (seen or unseen) (**20%**), number of GT used (single label, multilabel) (**20%**), application of hyperparameters (4%). In the optimization paradigm (Fig. 8 (c)), the percentage of the studies performing optimization is **50%,** while the number of studies using optimization techniques is **43.4%**. This can be seen in Refs. [71,72,75,76,92]. Maximum number of studies (**79.2%**) have used P-value as performance parameter (Fig. 8 (d)) among all the attributes in performance parameter cluster [69,72,76,87,88,92]. The remaining attributes of this cluster are a number of performance parameters used (**65.8%**), cross-validation process applied (**50.8%**), and sensitivity (**45.8%**), and accuracy (**44.2%**), specificity (30.8%) of the ML systems, precision of the ML system (20.8%), F1 score calculation (**8.2%**), and application data augmentation (**4.2%**).

Fig. 9 describes the percentage contributions of the clusters (a) scientific validation of ML system (b) performance of the statistical test (c) clinical evaluation of the ML systems, and (d) survival and hazard analysis of the ML studies. The corresponding clusters using absolute mean scores are exhibited in Figure A.2 (Appendix). The percentage of ML studies performing scientific validation by testing the ML system on other datasets is **25%** (Fig. 9 (a)). The statistical test carried out are the chi-square test and Cohen & Kappa test having equal percentage contribution (**8.4%**) (Fig. 9 (b)). Clinical evaluation cluster (Fig. 9 (c)) shows that **4.2%** of ML studies do a clinical evaluation of the ML system on a multicentre. Fig. 9 (d) shows the survival and hazard analysis cluster, the more ML studies have carried out the survival test (**20.8%**) while only **12.6%** of ML studies had performed the hazard analysis test. It is already observed that the ML system was very accurate compared to the non-ML system in predicting the CVD risk from the comparison of the attribute mean scores (Fig. 10 and Fig. 11).

The unique features of ML systems and non-ML systems were compared which are displayed in Table 2. It shows some common factors contributing to *high-bias* in both ML and non-ML studies such as no clinical validation, verification, statistical tests were done for both types of studies. The variability operation was also not performed for both the ML and non-ML studies. The hypothesis that the ML-based approaches for CVD risk assessment are less biased and more accurate is being discussed and proved in a further section.

6. Bias computation using analytical slope method: A novel paradigm

The section aims to design an analytical solution for computing the bias in ML and non-ML studies based on the concept of the "proportion of scores (POS)" obtained during the evaluation of AI attributes. Our definition of POS is derived by the rationale that it is reciprocal of the probability of each category (type) of the score. If p_i represents the probability of score "i", then the POS for score "i" is mathematically represented as $\left(\frac{1}{p_i}\right)$. Using this p_i notation the bias of a technique can be measured as the sum of reciprocal of the probabilities for all the scores "i", ranging from 0 to 5. Note that 0 to 5 are the scores obtained for each AI during the scoring paradigm. This concept can now be applied to ML

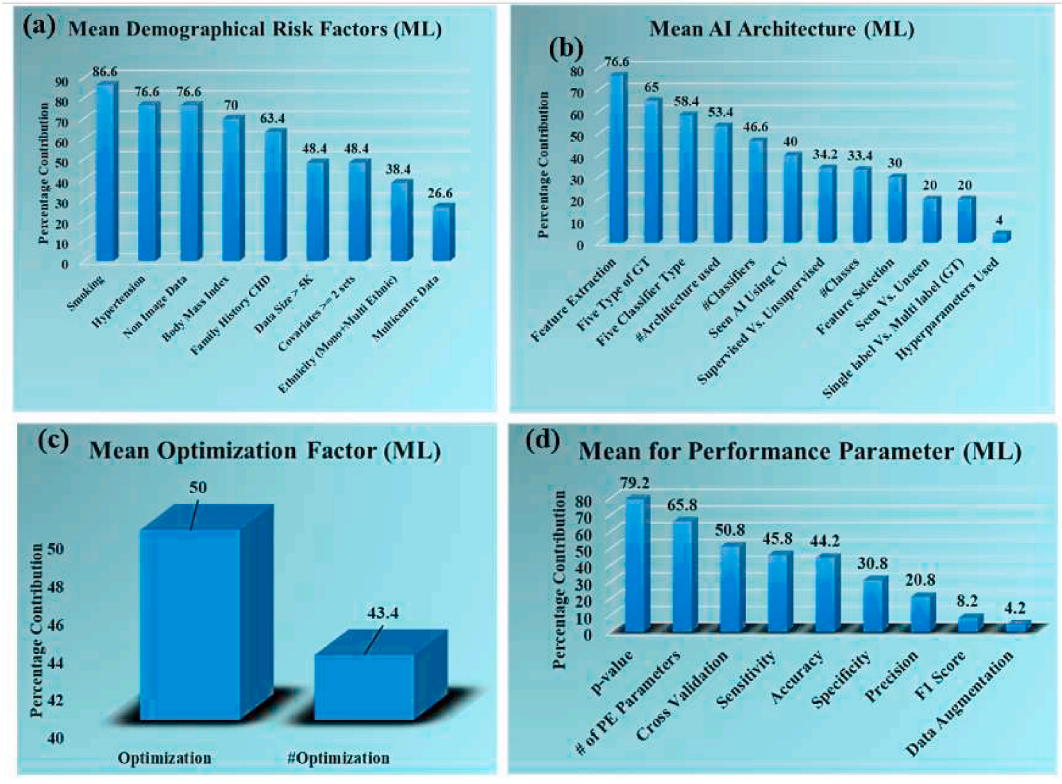


Fig. 8. Percentage distribution for first-four bias clusters (a) demographics (b) AI architecture (c) optimization factor, and (d) performance parameters.

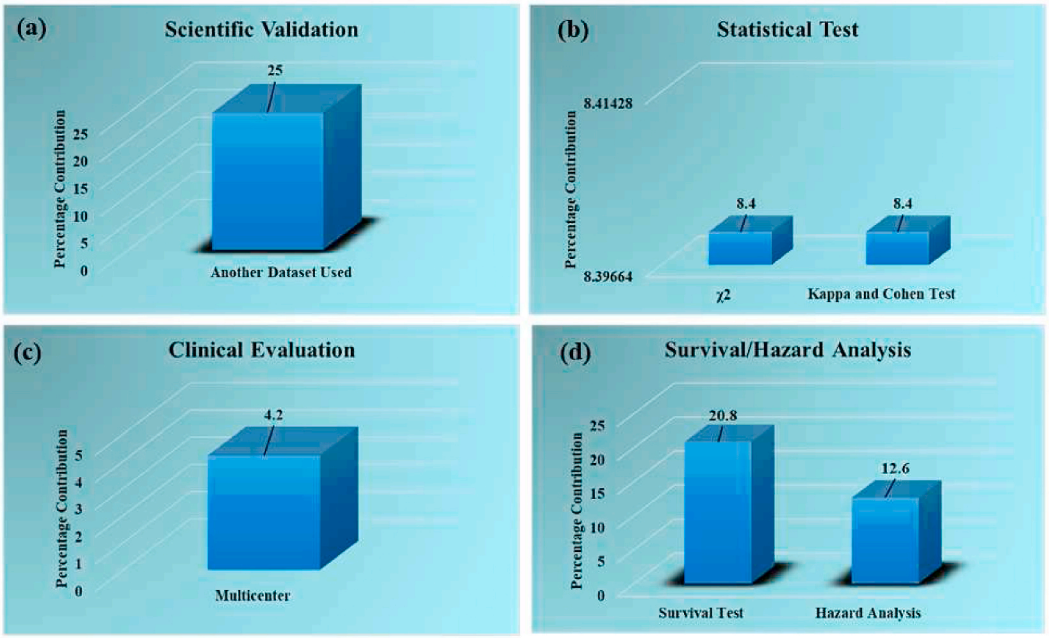


Fig. 9. Percentage distribution for clusters (a) scientific validation (b) statistical test (c) clinical evaluation, and (d) survival/hazard analysis.

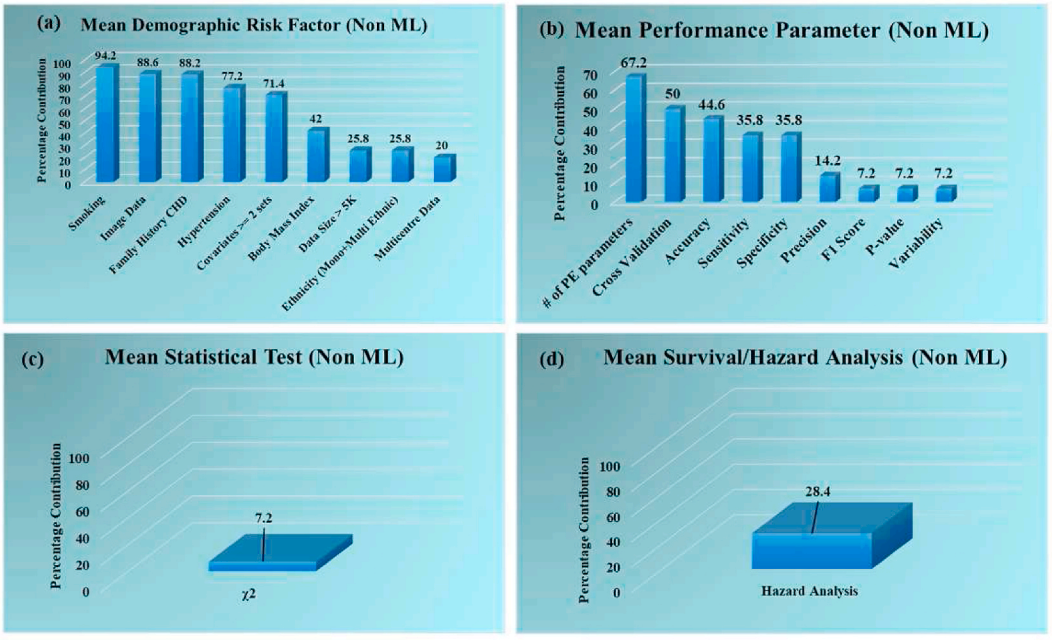


Fig. 10. The percentage for Clusters (a) demographics (non-ML) (b) performance parameters (non-ML) (c) statistical test (non-ML) (d) survival/hazard analysis (non-ML).

and non-ML setups.

Since there are 24 ML studies and 46 AI attributes, this corresponds to "24 times 46" leading to 1104 number of AI scores. On the contrary, there were "14 non-ML studies and 32 attributes", leading to 448 non-ML scores. The plot of these scores is shown in Fig. 12. The distribution of ML studies is on the top, and on the bottom is for the non-ML studies. Note that the length of each segment corresponding to each score leads to the motivation for the design of bias-effect in ML and non-ML. This needs to be now analyzed in terms of POS and eventually bias effect. The POS values for ML and non-ML are shown in Table 3.

Using the POS values, we can now compute the bias effect for ML and non-ML, shown in Table 4. The POS for the ML are **0.38**, **0.5**, **0.156**, **0.288**, **0.21**, and **1.08** corresponding to scores **0**, **1**, **2**, **3**, **4**, and **5** respectively. And the POS for the non-ML is **0.5**, **0.3**, **0.066**, **0.116**, **0.175**, and **1.44** for score **0**, **1**, **2**, **3**, **4**, and **5** respectively. Therefore, bias effect for ML ($Bias_{ML}$) is **20.20** and bias effect for non-ML ($\mathbf{Bias}_{non\text{-}ML}$) is **35.51**. So, it is evident that $\mathbf{Bias}_{ML} < \mathbf{Bias}_{non\text{-}ML}$. Thus, we conclude that the bias effect in ML for CVD risk is less than in non-ML for CVD risk by [(35.51-20.20)/20.20]~**43%**.

Note that the above technique offers several advantages **(i)** the mathematical representation is sound and straight-forward, **(ii)** it handles different proportions of data such as ML *vs.* non-ML, **(iii)** the technique is generalized, and **(iv)** it proves the hypothesis that the ML techniques are strong (less biased) than the non-ML (more biased). Even

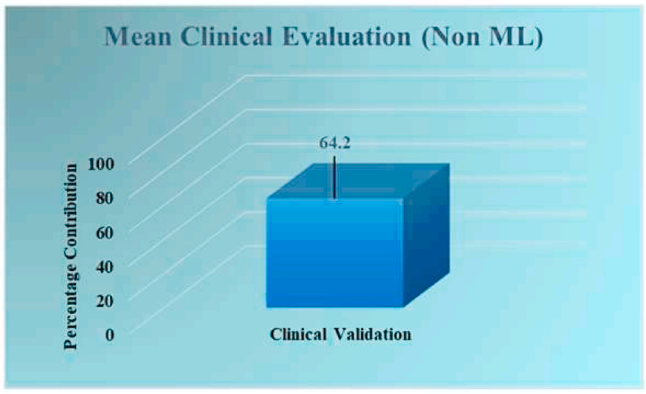


Fig. 11. The percentage for Cluster clinical evaluation (non-ML).

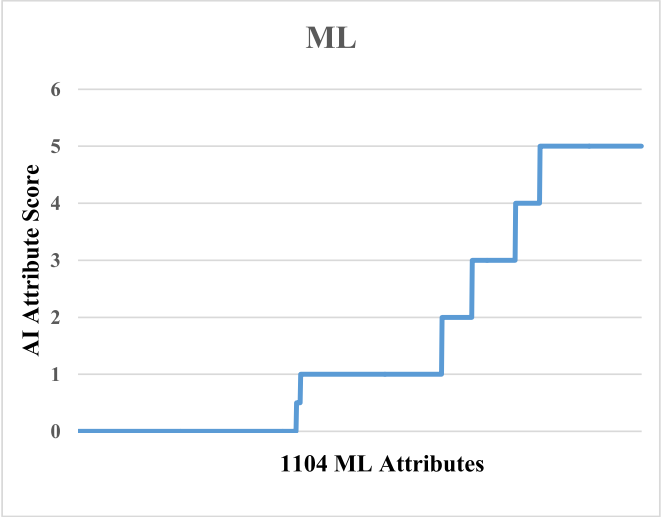
Table 2
Comparison of ML *vs.* non-ML systems for CVD risk assessment.

SN	Attributes	Machine Learning	Non-Machine Learning
1	Total Studies	24	14
2	Cutoffs	*Low-Moderate* = 1.83 *Moderate-High* = 1.59	*Low-Moderate* = 1.95 *Moderate-High* = 1.56
3	# Studies in each bias bin	High:9; Moderate:10; Low:5	High:3; Moderate:4; Low:7
4	Reason of studies in high-bias	No clinical validation, verification, statistical tests were not done, variability operation was not performed	No clinical validation, verification, statistical tests were not done, variability operation was not performed, data augmentation was not done, survival analysis was not performed
5	ML Training	Supervised/Unsupervised	No training
6	Ability to handle non-linearity	✓	×
7	ML hyper-parameter-based optimization	✓	×
8	Technique's nomenclature	Machine Learning Deep Learning	Logistic regression Statistical calculation
9	Data size	Relatively Large	Relatively Small, Moderate
10	Classifier Flexibility	✓	×
11	Risk granularity	>2	2
12	Risk Threshold	Flexible or can change	Ad-hoc
13	Gold standard labels	Multiple	Binary
14	Ability to handle different ethnicity	✓	×
15	Risk Stratification control	✓	×
16	Accuracy	Improved by changing the Neural network depth	No iteration can be performed
17	Loss Function	✓	×
18	Ground truth	Considered	No concept of ground truth
19	Class of GT	Multiple GT	Not possible
20	Ensemble model (fusion)	✓	×
21	Ability to handle #attributes	Large	Moderate

though the analytical paradigm offers several advantages, it is bound by the total number of AI attributes during the design of the ranking paradigm and can change, if the design of the AI attributes changes.

7. Bias distribution: ML vs. non-ML and recommendations

Our main objective is to understand the bias in the ML and non-ML



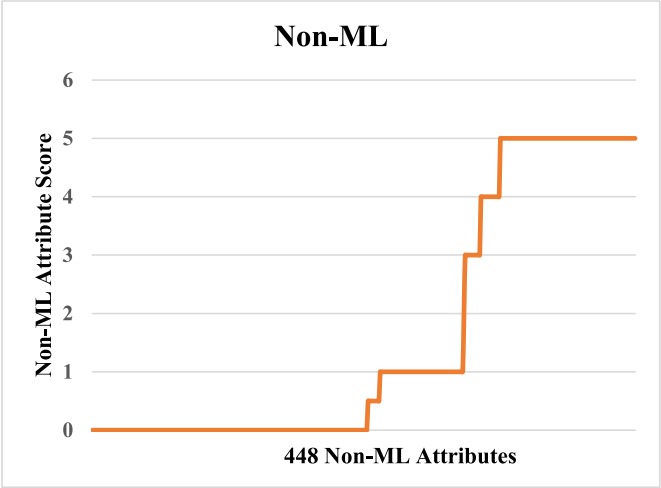


Fig. 12. Distribution of AI scores. **Top**: ML framework. **Bottom**: Non-ML framework.

Table 3
Proportion computation for the ML *vs.* non-ML.

POS for ML			POS for non-ML		
Score Type	Score Frequency	POS Computation	Score Type	Score Frequency	POS Computation
0	426	(426/1100) x100~38%	0	226	(226/446) x100~50%
1	276	(276/1100) x100~25%	1	68	(68/446) x100~15%
2	58	(58/1100) x100–5.2%	2	1	(1/446) x100–0.22%
3	84	(84/1100) x100–7.2%	3	13	(13/446) x100–2.91%
4	47	(47/1100) x100–4.2%	4	16	(16/446) x100–3.5%
5	198	(198/1100) x100~18%	5	111	(111/446) x100~24%

studies by scoring each attribute. Since there are 46 attributes and 24 ML studies, thus there are 1104 attribute scores, therefore it can help us understand the bias in AI, and lead to recommendations for improving the RoB. Therefore, we categorize the mean score for each attribute corresponding to *low-bias*, *moderate-bias*, and *high-bias* categories of studies. This is shown in Table A.4. The spirit of this computation was derived from our previous publications [108,109], but extended to ML

Table 4
Bias effect computation for ML *vs.* non-ML.

Bias for ML				Bias for non-ML			
Score Cat#	POS for ML	PROD 1 (P1)	Bias (1/P1)	Score Cat.	POS for Non-ML	PROD 2 (P2)	Bias (1/P2)
1	0.380	0.380	2.631	1	0.500	0.500	2.000
2	0.250	0.500	2.00	2	0.150	0.300	3.333
3	0.052	0.156	6.410	3	0.022	0.066	15.151
4	0.072	0.288	3.472	4	0.029	0.116	8.629
5	0.042	0.210	4.761	5	0.035	0.175	5.714
6	0.180	1.080	0.925	6	0.24	1.440	0.694
Total **Bias**$_{ML}$			**20.20**	Total **Bias**$_{non-ML}$			**35.51**

#Cat: Category; **Bias**$_{ML}$: Bias due to ML; **Bias**$_{non-ML}$: Bias due to non-ML.

(24 studies) and non-ML systems (14 studies) for CVD risk assessment. Similar to ML, since the non-ML has 32 attributes and 14 studies, resulting in 448 scores, it can correspondingly help understand the non-ML paradigm's bias. This section consists of two subsections. In the first section, we discussed the bias in the ML studies, and in the second section, we focused on the bias in the non-ML studies.

7.1. Bias distribution in ML attributes

The distribution of attributes is shown by three colors, red (*high-bias*), cream (*moderate-bias*), and green (*low-bias*). These three colors represent each attribute and are arranged from *high-bias* to *low-bias* values in decreasing order. This can be seen in Fig. 13. Our observations showed that number of studies in *low-bias*, *moderate-bias*, and *high-bias* groups were **5, 10,** and **9**, respectively. The *high-bias* attributes are non-image data, non-consideration of smoking data, p-value, and family history. The *moderate-bias* is due to BMI, feature extraction, feature selection, number of parameters used. The attributes that contribute to the *low-bias* in the ML studies are smoking, supervised training in ML system, high p-value, a greater number of optimization techniques used, and high specificity value. The attributes with the lowest mean scores are variability, verification, diagnostic order of ratio (DOR), Analysis of variance (ANOVA), clinical validation, FDA approval, ML systems were clinically passed or not. Therefore, these attributes are commonly causing bias in ML studies.

Based on the observation, we can say that the overall performance of the ML systems can be improved by considering the missing attributes, such as verification and validation. It can be achieved by including family history, smoking information, and hypertension data or attributes into the demographic cluster. Using a higher number of classifiers, optimization techniques, supervised training can lead to obtaining a *low-bias* AI system. The use of a higher number of performance parameters can lead to *low-bias* in ML systems. The clinically approved ML system always delivers better results in estimating CVD risk. Thus, making the ML systems *low-biased*. An unusual behaviour has been observed for a few ML attributes such as smoking with high mean scores for both *low-bias* and high-bias groups.

7.2. Bias distribution for the non-ML attributes

In this section, we classified the 14 non-ML studies among low- (green), moderate- (cream), and high-bias (red) groups (Fig. 14). Using the cutoff for the low-moderate group as **1.95** and for moderate-high cutoff as **1.56**, our observations showed **3**, **4,** and **7** studies were grouped into *low-bias*, *moderate-bias*, and high-bias groups respectively. The clinical setting was not performed in non-ML systems which was the main reason behind the high-bias for such a system. The other non-ML attributes resulting in high-bias were due to no consideration of family history and smoking information. The cross-validation that was not performed further leads to the high-bias in the non-ML system. Lack of performance evaluation for the system also caused high-bias in the non-ML systems. Statistical tests like the chi-square test, DOR, ANOVA were not performed to check the system's robustness leading to high-bias. The clinical verification for the non-ML systems was not performed, which had severely affected the bias and resulted in the high-bias of the system. The *moderate-bias* attributes were from image data or the consideration of the image phenotypes as a risk factor. From the observation, we can say that the main factors for improving the performance quality and reducing the bias in the non-ML systems were to perform clinical validation, verification, pass the clinical test, and get approval for the clinical settings. More demographic attributes should be introduced for

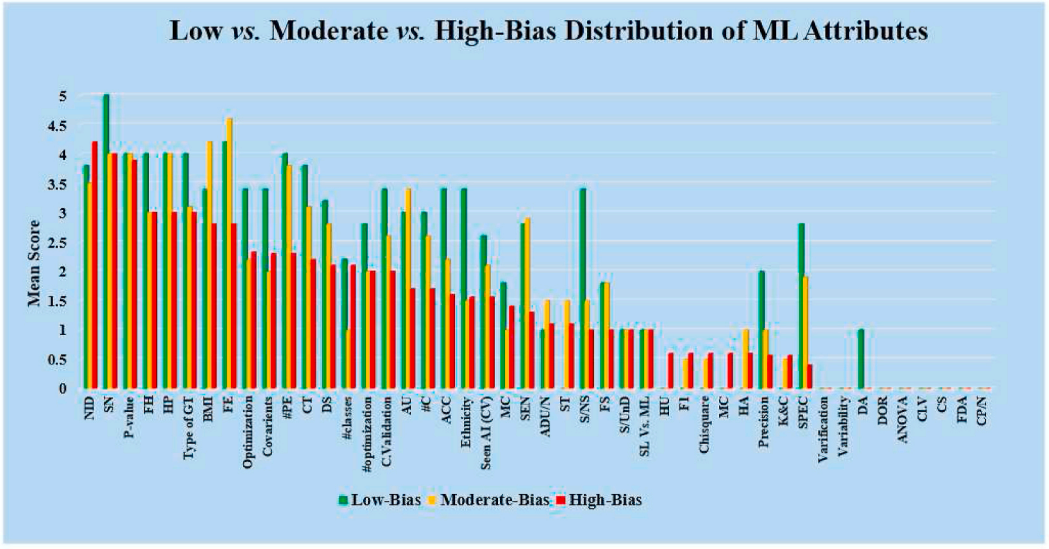


Fig. 13. Low *vs.* moderate *vs.* high-bias distribution of 46 AI attributes. **AN**: Attribute name; **NID**: Non-Image Data; **SN**: Smoking; **FH**: Family History; **HP**: Hypertension; **BMI**: Body mass index; **GT**: Ground truth; **FE**: Feature extraction; **#PE**: Number of performance parameter; **CT**: Classifier type, **DS**: Data size; **C. Validation**: Cross-validation; **AU**: Architecture used; **#C**: Number of classifiers; **ACC**: Accuracy; **MC**: Multicentre; **SEN**: Sensitivity; **AU/N**: Another dataset used or not; **ST**: Survival test; **S/NS**: Supervised/Non-supervised; **FS**: Feature selection; **S/UnD**: Seen *Vs.* Unseen Data; **SL *vs.* ML**: Single-label *vs.* Multilabel; **HU**: Hyper-parameters used; **F1**: F1 score; **MC**: Multicentre (clinical evaluation); **HA**: Hazard analysis; **K&C**: Kappa-Cohen test; **SPEC**: Specificity; **DA**: Data augmentation; **DOR**: Diagnostic Odds Ratio test; **ANOVA**: ANOVA test; **Cl.V**: Clinical validation; **CS**: Clinical setting; **FDA**: Food and Drug Administration approved; **CP/N**: Clinically passed or not.

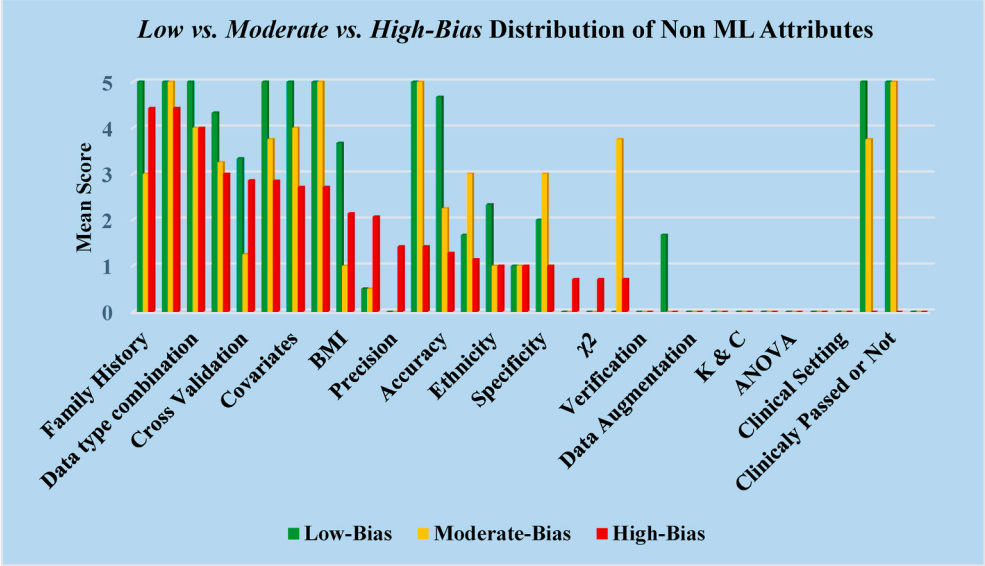


Fig. 14. *Low vs. moderate vs. high-bias* distribution of non-ML attributes. **χ2:** Chi-square test; **K&C**: Kappa & Cohen Test; **BMI**: Body mass index; **ANOVA**: ANOVA Test.

the reduction of bias in AI systems.

7.3. Primary and secondary recommendations

The proposed study presented the bias in ML for CVD risk assessment. We here summarize the crucial factors by which the RoB can be reduced. This can be categorized into **(a)** primary and **(b)** secondary sets of recommendations. The primary recommendations include:

(i) *Outcome design:* The choice of the outcome plays a very important role in ML design for CVD risk assessment. The order of outcomes was EEGS < Chronic Diseases < Heart Conditions < CAC < Death, with death being the best, followed by CAC score, heart conditions, chronic diseases like hypertension, diabetes, and EEGS which is a function of carotid arterial morphological phenotypes such as lumen diameter. Stronger outcome designs are recommended for robust ML design. **(ii)** *Morphological image-based phenotypes risk factors:* Due to the same genetic makeup of morphological plaque between the coronary artery and carotid artery, the CUSIP is a robust non-invasive biomarker for CVD/CAD risk prediction when used in a reduced speckle framework [110,111]. This CUSIP can be fused with OBBM and LBBM. Thus, the CUSIP risk factor can be recommended for removing the RoB. **(iii)** *Clinical Evaluation:* The clinical evaluation piece gets missed due to the lack of 360 coverage. Therefore, it is vital to have the ML system try to evaluate the unseen patients whose results are known *a priori.* Thus, the ML system can be evaluated clinically. **(iv)** *Risk Granularity*: Risk prediction generally provides binary results for several fields in medicine. This strategy makes it difficult or challenging for drug prescription and proper patient care. Because of this, new techniques came into practice with more than one class (multiclass) framework that gives a multi-label granularity of risk prediction. These will help in proper control of medication and monitoring. We, therefore, recommend the outcome design to be multiclass. **(v)** *Validation using Cross Modality*: Typically, scientific validation is ignored using cross-modality due to inherent reasons such as cost. However, in scientific validation, it is vital to validate the ML systems using cross-modality imaging. **(vi)** *Sample Size*: In AI, the sample size is vital; therefore, the performance and sample size constraints are vital, keeping the power analysis in mind. **(vii)** *Inter- and Intra-observer Variability*: Our findings displays that none of the ML-based studies had carried out variability operation (inter-and intra-observer) analysis. Thus, compromising the robustness of the ML results or outcomes. Therefore, we recommend that inter-and intra-observer variability operations can be performed, further ensuring the reliability of the ML systems in clinical settings. An example can be seen in Refs. [112,113]. The secondary recommendations are **(i)** strong model design, segregating cross-validation, prediction, graphical user interface (GUI), validation, and training. **(ii)** using multiple datasets for reproducibility, **(iii)** multi-ethnic data collection for studies, **(iv)** relaxation in ground-truth events for clinical validation, **(v)** high-quality documentation, and **(vi)** opting for the process of peer-review.

8. Discussions

Typically, CVD or CAD risk prediction is conducted using conventional calculators such as FRS, SCORE, or ASCVD, but these calculators presume a linear relationship between the covariates (risk factors or variables) and outcomes (so-called gold standard or ground truth). As a result, CVD is either underestimated or overestimated. Further, the CVD estimated risk is limited to binary categories (or classes). This helps in improve drug delivery (statin) to prevent coronary heart disease (CHD).

The proposed review was a first of its kind to understand the bias in ML-based framework for CVD risk prediction. The second aspect of this study was focused on machine learning literature for CVD risk prediction by categorizing the outcomes into five different classes such as (i) death, (ii) CAC score, (iii) CVE/CVE, (iv) chronic diseases, and (v) EEGS. The novelty of the design outcome provided a deeper look into the AI framework for CVD estimation towards robust design. The third novelty of this system is focused on understanding the ML *vs.* non-ML paradigms since these risk prediction methods use different kinds of covariates.

This study involved scoring attributes of the ML studies with the help of experts who have more than 15 years of experience in AI. Further, the objective was to compute the aggregate score followed by the design of a ranking scheme and determine the two cutoffs (LM and MH) for classifying studies into *low-*, *moderate-*, and *high-bias*. This allowed in understanding the *low-, moderate-*, and *high-biases* for every AI attribute, thereby giving solid recommendations for a reduction in RoB. A novel analytical slope method was designed to compute the bias effect in ML and non-ML paradigm. This validated the hypothesis that the bias effect in ML systems is lower compared to non-ML systems, suggesting that ML

Table 5
Benchmarking Table for ML-based CVD risk prediction studies.

C1	C2	C3	C4	C5	C6	C7	C8	C9	C10
SN	**First Author (Year)**	**Cohort**	**#RF**	**Patient Cohort/ Focus**	**Covariates (Biomarkers)**	**Prediction (Outcome)**	**Risk Calculator**	**Performance Attributes**	**Summary of the Study**
R1	Jamthikar *et. al.* [4] (2021)	-	120	CKD	HT, BMI, Age, Lipids, eGFR, HbA1c,5 image phenotype	Composite CVD & CKD	FRS, ML	Contribution of eGFR, image-based phenotypes in CVD Risk assessment	Indicates link between CVD and CKD especially by reducing eGFR, albumin-to-creatinine ratio and albuminuria should also be considered in CVD risk assessment models
R2	Jamthikar *et. al.* [20] (2020)	UK, Germany, Scotland, USA, Japan, and Europe	208	-	ESR, BMI, Age, Lipids, eGFR, HbA1c,5 Image Phenotype	CVD	ML	Risk Contribution, AUC	By using powerful AI-based techniques, it examines the ways for the inclusion of more non-invasive image-based phenotypes in CVD risk prediction.
R3	Jamthikar *et. al.* [47] (2019)	-	-	-	cIMTave, cIMTmax, cIMTmin, IMTV, mTPA	Event, Non-event, Stroke	ML & Deep Learning	cIMT, mTPA	The combination of image-based risk factors and conventional CVD risk factors leads to more accurate CVD/stroke risk determination, handling multiple sources of information in the big data by the use of AI-based framework (ML & DL) will be the future of CVD/stroke risk prediction.
R4	Jamthikar *et. al.* [6] (2020)	-	120	Rheumatoid Arthritis	cIMTave, cIMTmax, cIMTmin	CVD	ML	AUC, Specificity, sensitivity	The image-based biomarkers for carotid atherosclerotic such as cIMT & plaque are significantly associated with RA-specific inflammatory markers.
R5	Jamthikar *et.al.* [114] (2020)	-	202		Age, Gender, DBP, SBP, HT, HbA1c, LDL-C, HDL-C, TC, TG, TC/HDL, FH, Smoking, eGFR	CVD	ML	cIMT, AUC, specificity, sensitivity	CVD/stroke risk calculators (ML-based) show a higher ability to predict 10-year CVD/stroke risk when compared to the 13 other types of risk calculators (statistical-based) including AECRS 2.0 model.
R6	Khanna et al. [7] (2019)	-	10	Rheumatoid Arthritis	BMI, Smoking, Lipid, HT, Diabetes, Insulin Resistance	CVD	ML & DL	cIMT, AUC	It provides an understanding of RA pathogenesis and its relation with carotid atherosclerosis by B-mode US methods. Lacunas is a type of conventional risk score. It also shows the role of ML-based algorithms by using tissue characterization and could facilitate CVD risk prediction in patients with RA.
R7	Viswanathan et al. [49] (2021)	-	-	Diabetes, Coronavirus	Obesity, Oxidative stress, Coagulation activity, Hypoxia, Cytokine, RASS dysregulation	CVD	-	Linking of the risk factors with CVD	This review showed a bidirectional link between diabetes mellitus (DM) and coronavirus disease 2019 (COVID-19). The other link is from DM to COVID-19 because of drug-induced or decreased furin levels in DM or impaired immunity is from DM to COVID-19 (upper part of DKA) One is being from COVID-19 to DM (lower part of DKA) because of pancreatic damage or renin-angiotensin-aldosterone system dysregulation. This reduces the insulin release and increase in insulin resistance. Both these processes trigger the CVD.
R8	Proposed Study	Unseen	117	Heart Conditions, Chronic diseases	OBBM, LBBM, CUSIP, MedUSE	CVD	ML	Outcomes (ground truth) & Risk factors linking	Five types of the gold standard were identified in the ML design for CAD/CVD risk prediction. The *low-moderate* and *moderate-high* cutoffs for ML and non-ML studies were determined. Risk granularity was used in the analysis. A set of five recommendations were presented for a reduction in RoB.

#RF: number of references used in the study.

studies are superior solutions for CVD risk assessments. We presented a set of five key recommendations for reduction in RoB such as (i) choosing better outcome design in the order of death > CAC > CVE > Chronic disease > EEGS; (ii) fusion of image-based phenotypes (CUSIP) with conventional risk factors (OBBM + LBBM); (iii) dire need for scientific validation by using either other data sets or adopting the Unseen-AI paradigm; (iv) designing the CVD systems with multiple CVD risk granularity; and (v) designing system which can be adapted in routine clinical settings.

8.1. Benchmarking: Comparative studies on cardiovascular disease

Table 5 shows a comparison between seven review studies that focused on CVD risk prediction. This table shows nine attributes (column **C2** to column **C10**) for each of the seven studies ([4,6,7,20,47,49,114]) corresponding to the rows **R1** to **R8**. These nine attributes presented were the year of the study (**C2**), cohort (**C3**), a number of studies used in the study (**C4**), type of disease in the cohort (**C5**), type of covariates (**C6**), type of outcome (**C7**), type of risk calculator adopted (**C8**), performance parameters (**C9**), and finally the summary/conclusion of the study (**C10**). Column **C4** shows the # of references in the study, ranging from 10 to 208, while our study had used 117 studies. The description of the type of disease is shown in **C5**, such as CKD (row R1), RA (row R4 and R6), COVID (row R7), Diabetes (row R7). Our proposed study used heart conditions and chronic diseases as part of the disease type. Five out of seven studies used conventional biomarkers (rows **R1, R2**, **R5**, **R6**, **and R7** in **C6**) such as smoking status, gender, obesity, BMI, age, hypertension, lipids (LDL, HDL), eGFR, HbA1c, oxidative stress, coagulation activity, hypoxia, cytokine, RASS dysregulation. Two studies used only image-based phenotypes as risk factors (row **R3** and **R4**) such as cIMT (ave., max., min), IMTV, and TPA. No studies have used both OBBM and CUSIP biomarkers in conjunction for estimation of the CVD risk. Column **C7** shows the type of the outcome, which was typically CVD in all the studies (R1 to R8). The type of the risk calculator is shown in column **C8** and most of them adapted ML, however, some studies used the combination of ML and CCVRC, such as FRS, SCORE, ASCVD. The summary of each study is presented in column **C9**. Only the proposed study (row **R8**) discussed the "Unseen AI" compared to other studies. The proposed study discusses granularity in CVD risk prediction and is not restricted to the binary system of risk prediction (**R8, C10**).

8.2. A short note on bias in machine learning for CVD risk assessment

Cardiovascular disease is causing the death of approximately 18 million people globally [1]. For reducing these death rates early and accurate detection of the CVD risk is vital. Therefore, for the improvement in the prediction of CVD risk, the use of AI systems came into the application as compared to the conventional tool. But there are some challenges with AI systems since it sometimes focuses only on accuracy, forgetting about the validation of the AI systems in the clinical setting. This overemphasizes the accuracy and underemphasizes the validation of AI systems, which causes bias in the AI system [115]. So, it is important to understand the bias in each AI system for its improvement in CVD risk prediction [116]. In this review, we have done the same by ranking the AI systems. The ranking is done by scoring each AI attribute based on the performance of the attributes. The mean score and the cumulative mean were calculated and based on that two cutoffs (low--moderate cutoff, moderate-high cutoff) were determined. And the AI studies were classified as *low-bias*, *moderate-bias,* and *high-bias* groups. This whole system of ranking and classifying the AI systems for CVD risk prediction was done by experts having more than 15-years of experience in the field of CVD. Further, the correlation of AI attributes and non-ML attributes with the bias has been done.

8.3. Strength, weakness, and extensions

The main strength of this study was the establishment of a ranking-based solution for binning the ML studies into three bins (LB, MB, and HB). Further, we analytically proved that bias in ML is lower compared to non-ML-based studies. Lastly, we clearly established the five-set of recommendations for improving the RoB. Overall, our review also characterized the AI attribute behaviour among the 24 ML and 14 non-ML studies, establishing the relationship between LB, MB, HB distributions.

While the results were encouraging and promising, we noticed a lack of solid focus and data missing in these studies. Due to a lack of funding in research and data sharing, there is a lack of participation by several advanced AI groups. Not all the studies presented all the kinds of attributes as discussed in the benchmarking table. This study can be extended by exploring exhaustively other attributes for stronger bias measurements and classification methods [117].

9. Conclusions

This was the first pilot study of its kind which computed bias in ML for CVD risk assessment. PRISMA model was used to select the best 24 ML studies and 14 non-ML studies. The study showed two powerful schemes for bias estimation such as (a) *ranking solution* clubbed with an estimation of three bias bins and clustering these 24 ML studies in these three bins *(low-bias*, *moderate-bias*, and *high-bias)*, and (b) *analytical slope method* for comparison between ML *vs*. non-ML for bias estimation for validating the hypothesis. Besides the above, the study offered (i) discovery of five types of the gold standard during the design of ML-based CVD risk assessment solution; and (ii) understanding the bias in ML *vs*. non-ML using the ranking method. Further, we presented five types of recommendations for the reduction of RoB such as (i) superior outcome (gold standard) design; (ii) amalgamation of covariates such as OBBM, LBBM, and CUSIP in machine learning framework; (iii) need for scientific validation; (iv) CVD granularity for better drug delivery and monitoring the atherosclerotic disease, and (v) need for clinical evaluation thereby adapting in clinical settings. We demonstrated that the bias-effect in ML was less compared to non-ML by **43%** using the analytical slope method, thus validating and establishing the hypothesis. As AI further evolves such as ensemble techniques and deep learning solutions, we will see high confidence in the risk estimation for CVD.

Declaration of competing interest

Dr. Jasjit S. Suri is with AtheroPoint™ (Roseville, CA, USA), focused on cardiovascular disease and stroke.

Appendix A

All the ML and non-ML studies were ranked based on the scoring of the AI attributes. There was a total of 46 attributes as shown in Table A.1. As seen in the table there were 9 attributes (A2, A4, A6, A7, A9, A14, A15, A18, A20, A21, and A22) which were binary nature i.e. having two states (higher & lower). These were graded 5 for the higher state i.e., when the attribute was considered and 1 for the lower state. Since the authors have made an effort to design the system, we have chosen to grade 1 instead of 0. There were a few attributes (A29-A46) that were also binary in nature where we gave graded 5 for the higher state and 0 for the lower state. It is because there was no option to choose any intermediate state. Also, these were performance and test attributes that were crucial to determining the bias of the ML system. The next set of attributes A1, A3, A5, A8, A10-A13,

A16, A17, A19, and A23-A28) were multiclass in nature. These attributes were graded 5 for the highest state then gradually decreasing to 0 for the lowest state. We opt for multiple classes in these attributes as there were options to choose an intermediate state based on the distribution.

Table A.1
Grading Scheme.

SN	AN	Attributes	Descriptions (grade value)
1	A1	Data Size	5 (>5000); 4 (4000–5000); 3 (3000–4000); 2 (2000–3000); 1 (1000–2000); 0.5 (<1000)
2	A2	Family History	5 (when the FH was used); 1 (when FH was not used)
3	A3	Covariates	4 (OBBM, LBBM, CUSIP, Medications); 3 (OBBM, LBBM, CUSIP); 2 (OBBM, LBBM); 1 (OBBM)
4	A4	BMI	5 (when considered) 1 (when not considered)
5	A5	Ethnicity	5 (Multi-ethnic); 2 (Mono ethnic); 1 (Not mentioned)
6	A6	Hypertension	5 (if Hypertension was considered); 1 (if not considered)
7	A7	Smoking	5 (if Smoking was considered); 1 (If not considered)
8	A8	Data Type Combination	5 (Point and Image data combined); 4 (Image data alone); 3 (Point data alone)
9	A9	Data collection	5 (Multiple institute); 1 (Single institute)
10	A10	#Architecture	5 (# Architecture>5); 4 (# Architectures 4); 3 (# Architectures 3); 2 (# Architectures 2); 1 (# Architectures 1)
11	A11	Supervised/Unsupervised	5 (if supervised); 2 (if non supervised); 1 (if not mentioned)
12	A12	Five Classifier Type	5 (Ensemble learning); 4 (Deep learning); 3 (Machine learning); 2 (Statistical); 1 (visual)
13	A13	# Classifiers	5 (# Classifiers 4); 4 (# Classifiers 3); 3 (# Classifiers 2); 2 (# Classifiers 1); 1 (if not used)
14	A14	Feature Extraction	5 (when the FE was used); 1 (when FE was not used)
15	A15	Feature Selection	5 (when the FS was used); 1 (when FS was not used)
16	A16	# Classes	5 (# Classes 4); 3 (# Classes 3); 1 (# Classes 2)
17	A17	Five Type of GT	5 (Death); 4 (CAC); 3 (CVE/Cerebrovascular event); 2 (Chronic Disease); 1 (EEGS)
18	A18	Seen *Vs.* Unseen Data	5 (Unseen Data); 1 (Seen Data)
19	A19	Seen AI using CV	5 (K > 10); 4 (K = 10); 3 (K = 5); 2 (K < 4); 1 (K not used)
20	A20	Single-label/Multilabels (GT)	5 (Multilabels); 1 (Single-label)
21	A21	Hyper-parameters used	5 (when the hyper-parameters were used); 1 (when hyper-parameters were not used)
22	A22	Optimization	5 (when the optimization was done); 1 (when optimization was not done)
23	A23	#Optimizers	5 (#Optimizers>2); 4 (#Optimizers 1); 1 (Not used)
24	A24	Cross-Validation	5 (k fold); 4 (10-fold); 3 (4 ≤ fold); 2 (Cohort); 1 (Validation not done)
25	A25	# of PE Parameters	5 (# PE Parameters>5); 4 (# PE Parameters 4); 3 (# PE Parameters 3); 2 (# PE Parameters 2); 1 (# PE Parameters 1)
26	A26	Accuracy	5 (95–100); 4 (90–95); 3 (90–80); 2 (80–70); 1 (<70); 0 (Not mentioned)
27	A27	Sensitivity	5 (95–100); 4 (90–95); 3 (90–80); 2 (80–70); 1 (<70); 0 (Not mentioned)
28	A28	Specificity	5 (95–100); 4 (90–95); 3 (90–80); 2 (80–70); 1 (<70); 0 (Not mentioned)
29	A29	F1 score	5 (when the F1 score was used); 0 (when F1 score was not used)
30	A30	Precision	5 (when the precision was used); 0 (when precision was not used)
31	A31	P-value	5 (when the P-value was used); 0 (when P-value was not used)
32	A32	Verification	5 (when the verification was done); 0 (when verification was not done)
33	A33	Variability	5 (when the variability was done); 0 (when the variability was not done)
34	A34	Data Augmentation	5 (when the data augmentation was done); 0 (when data augmentation was not done)
35	A35	$\chi 2$	5 (when the $\chi 2$ test was done); 0 (when $\chi 2$ was not done)
36	A36	Kappa-Cohen	5 (when the Kappa-Cohen test was done); 0 (when Kappa-Cohen was not done)
37	A37	DOR	5 (when the DOR test was done); 0 (when DOR test was not done)
38	A38	ANOVA	5 (when the ANOVA test was done); 0 (when ANOVA was not done)
39	A39	Clinical Validation	5 (when the Cl. V was done); 0 (when Cl. V was not done)
40	A40	Multicentre	5 (when the MC was done); 0 (when MC was not done)
41	A41	Clinical Setting	5 (when the CS was done); 0 (when CS was not done)
42	A42	Another dataset used or not	5 (when another dataset was used); 0 (when another dataset was not used)
43	A43	FDA Approved	5 (when the system is FDA approved); 0 (when the system was not approved)
44	A44	Clinically Passed or not	5 (when the system was clinically passed); 0 (when it was not passed)
45	A45	Survival Test	5 (when the ST was done); 0 (when the ST was not done)
46	A46	Hazard Analysis	5 (when the HA was done); 0 (when HA was not done)

Table A.2
Scoring for each attribute and ranking of studies.

		Demographics									AI Architecture Design												Optimization	
		A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	A12	A13	A14	A15	A16	A17	A18	A19	A20	A21	A22	A23
SN	**Studies**	**DS**	**FH**	**Covariates**	**BMI**	**Ethnicity**	**HP**	**SN**	**NID**	**MC**	**#AU**	**S/NS**	**CT**	**#C**	**FE**	**FS**	**#Classes**	**Type of GT**	**S/UnS**	**Seen AI (CV)**	**SL Vs. ML**	**HU**	**Optimization**	**#Optimizer**
1	Sajeev et al. [87]	5	1	3	1	5	5	5	3	5	4	5	3	3	5	1	1	2	1	2	1	0	1	1
2	Jamthikar et al. [69]	0.5	5	5	5	1	5	5	5	1	3	5	3	2	5	1	5	5	1	4	1	0	1	1
3	Kakadiaris et al. [72]	5	5	3	1	5	5	5	3	1	2	5	3	3	1	1	1	5	1	2	1	0	5	4
4	Nakanishi et al. [75]	5	5	3	5	5	5	5	5	1	1	1	5	2	5	1	3	4	1	4	1	0	5	4
5	Pattarabanjird et al. [76]	0.5	5	3	5	1	1	5	3	1	5	1	5	5	5	5	1	4	1	1	1	0	5	4
6	Unnikrishnann et al. [88]	2	1	3	5	1	5	5	3	1	3	5	3	2	5	5	1	2	1	1	1	0	1	1
7	Chao et al. [92]	2	5	1	1	1	5	5	4	1	4	1	4	3	5	1	1	1	1	1	1	0	5	4
8	Van den Oever et al. [78]	5	5	1	5	1	1	1	5	1	2	1	4	2	5	1	1	4	1	1	1	0	1	1
9	Zarkogianni et al. [89]	0.5	1	3	5	1	5	5	3	1	5	2	5	2	5	1	1	2	1	4	1	0	1	1
10	Yang et al. [86]	5	1	1	5	1	5	5	3	1	5	1	3	3	5	1	1	3	1	1	1	0	1	1
11	Joo et al. [81]	5	1	3	5	2	1	5	3	1	4	1	2	2	5	1	1	3	1	1	1	0	5	4
12	Dimopoulos et al. [80]	2	1	3	5	1	5	5	3	1	3	1	3	4	5	1	1	3	1	5	1	0	1	1
13	Weng et al. [85]	5	5	3	5	5	5	5	3	1	4	1	3	5	1	1	1	3	1	2	1	0	1	1
14	Wu et al. [71]	0.5	1	1	1	1	5	1	3	1	3	1	1	1	5	5	1	5	1	1	1	0	5	5
15	Johri et al. [70]	0.5	5	1	5	1	5	5	5	1	1	1	3	2	5	1	1	5	1	4	1	0	1	1
16	Sanchez et al. [77]	3	5	3	1	1	1	5	5	1	1	1	1	1	1	1	5	4	1	1	1	5	1	1
17	Kennemalaha et al. [90]	1	5	1	5	1	5	5	5	1	1	1	0	1	1	1	1	1	1	1	1	0	1	1
18	Cheung et al. [73]	5	1	3	5	5	1	5	5	1	2	1	4	2	5	1	1	4	1	1	1	0	1	1
19	Panaretos et al. [83]	2	5	1	5	1	5	5	3	1	2	1	3	2	1	1	1	3	1	5	1	0	5	4
20	Jamthikar et al. [91]	0.5	1	3	1	1	5	5	5	1	2	1	3	2	5	1	3	1	1	1	1	0	5	4
21	Emaus et al. [74]	1	1	3	1	2	1	1	4	5	1	1	4	2	1	1	5	4	1	1	1	0	5	4
22	Kim et al. [82]	5	5	1	1	1	5	5	3	1	1	1	2	2	5	1	1	3	1	2	1	0	1	1
23	Su et al. [84]	0.5	5	3	5	1	5	5	3	1	2	1	2	2	5	1	1	3	1	1	1	0	1	1
24	Wennstig et al. [79]	0.5	1	3	1	1	1	1	5	1	3	1	1	1	1	1	1	4	1	1	1	0	1	1

Table A.3
Scoring for attributes (A24 to A46).

		Performance Parameters											Scientific Validation	Stastistical Test				Clinical Evaluation/ Effectiveness			Clinical Validation		Survival/ Hazard Analysis	
		A24	A25	A26	A27	A28	A29	A30	A31	A32	A33	A34	A35	A36	A37	A38	A39	A40	A41	A42	A43	A44	A45	A46
SN	**Studies**	**Validation**	**#PE**	**Acc**	**SEN**	**SPEC**	**F1**	**Precision**	**p-value**	**Verification**	**Variability**	**DA**	**ADU/N**	**χ2**	**K&C**	**DOR**	**ANOVA**	**Cl.V**	**MC**	**CS**	**FDA**	**CP/N**	**ST**	**HA**
1	Sajeev et al. [87]	3	5	3	3	2	0	5	5	0	0	5	5	0	0	0	0	0	0	0	0	0	0	0
2	Jamthikar et al. [69]	4	4	4	3	5	0	0	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	Kakadiaris et al. [72]	3	4	4	3	5	0	0	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	Nakanishi et al. [75]	4	3	3	3	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
5	Pattarabanjird et al. [76]	3	4	3	2	0	0	5	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6	Unnikrishnann et al. [88]	2	4	2	1	3	0	0	5	0	0	0	0	0	0	0	0	0	0	0	0	0	5	5
7	Chao et al. [92]	2	3	3	3	0	0	0	5	0	0	0	5	0	0	0	0	0	0	0	0	0	5	0
8	Van den Oever et al. [78]	2	4	3	5	3	0	5	0	0	0	0	5	0	5	0	0	0	0	0	0	0	0	0
9	Zarkogianni et al. [89]	4	4	1	1	1	0	0	5	0	0	0	0	0	0	0	0	0	0	0	0	0	5	5
10	Yang et al. [86]	1	4	2	5	5	0	0	5	0	0	0	0	5	0	0	0	0	0	0	0	0	0	0
11	Joo et al. [81]	1	5	2	2	2	5	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
12	Dimopoulos et al. [80]	5	4	3	3	1	0	0	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
13	Weng et al. [85]	3	2	2	1	2	0	0	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
14	Wu et al. [71]	2	4	3	5	1	0	0	5	0	0	0	5	0	0	0	0	0	0	0	0	0	0	0
15	Johri et al. [70]	4	4	1	3	1	0	0	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
16	Sanchez et al. [77]	2	3	3	0	0	0	5	5	0	0	0	5	0	0	0	0	0	0	0	0	0	0	0
17	Kennemalaha et al. [90]	1	4	1	4	1	0	0	5	0	0	0	0	5	0	0	0	0	0	0	0	0	5	5
18	Cheung et al. [73]	2	2	1	0	0	0	0	5	0	0	0	5	0	0	0	0	0	0	0	0	0	0	0
19	Panaretos et al. [83]	5	2	1	0	0	0	0	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
20	Jamthikar et al. [91]	2	2	3	5	1	0	0	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
21	Emaus et al. [74]	1	1	0	1	0	0	0	0	0	0	0	0	0	5	0	0	0	5	0	0	0	5	0
22	Kim et al. [82]	3	3	3	0	0	5	0	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
23	Su et al. [84]	1	3	2	2	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
24	Wennstig et al. [79]	1	1	0	0	0	0	0	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

AN: Attribute name; NID: Non-Image Data; SN: Smoking; FH: Family History; HP: Hypertension; BMI: Body mass index; GT: Ground truth; FE: Feature extraction; #PE: Number of performance parameter; CT: Classifier type, DS: Data size; C. Validation: Cross-validation; AU: Architecture used; #C: Number of classifiers; ACC: Accuracy; MC: Multicentre; SEN: Sensitivity; AU/N: Another dataset used or not; ST: Survival test; S/NS: Supervised/Non-supervised; FS: Feature selection; S/UnS: Seen Vs. Unseen Data; SL Vs. ML: Single-label vs. Multilabel; HU: Hyper-parameters used; F1: F1 score; MC: Multicentre (Clinical evaluation); HA: Hazard analysis; K&C: Kappa-Cohen test; SPEC: Specificity; DA: Data augmentation; DOR: DOR test; ANOVA: ANOVA test; Cl.V: Clinical validation; CS: Clinical setting; FDA: Food and Drug Administration approved; CP/N: Clinically passed or not.

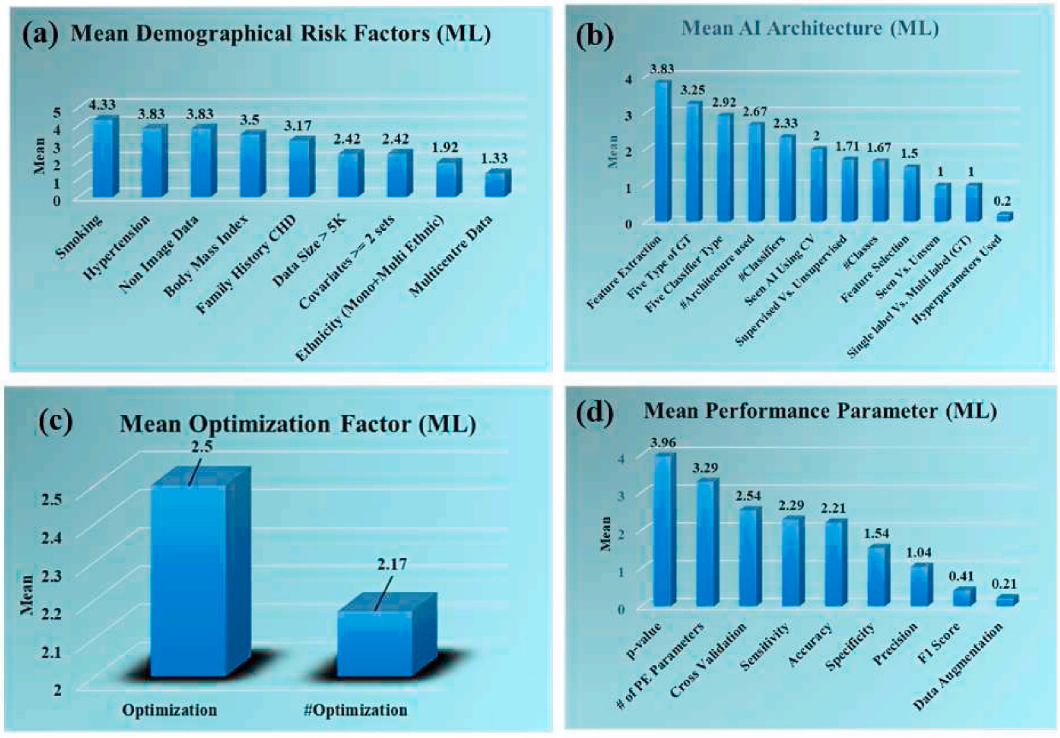


Figure A.1. Mean for Clusters (a) demographics (b) AI architecture (c) optimization factor (d) performance parameters.

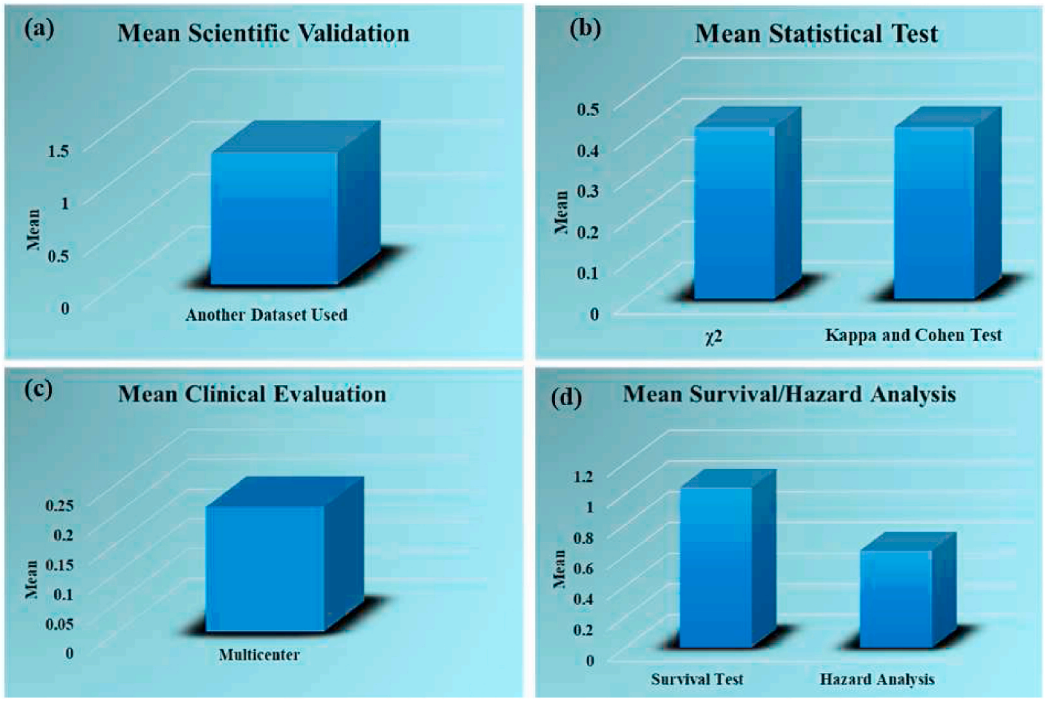


Figure A.2. Mean for Clusters (a) scientific validation (b) statistical test (c) clinical evaluation (d) survival/hazard analysis.

Table A.4
Mean computation for each attribute by three bias categories (LB, MB, HB)-link section 6.

SN	AN	Attributes	*Low-Bias*	*Moderate-Bias*	*High-Bias*
0	-	-	[69,72,75,76,87]	[70,71,78,80,81,85,86, 88,89,92]	[73,74,77,79,82,83,84,90,91]
1	A8	NID	3.8	3.5	4.2
2	A7	SN	5	4	4
3	A31	P-value	4	4	3.89
4	A2	FH	4	3	3
5	A6	HP	4	4	3
6	A17	Type of GT	4	3.1	3
7	A4	BMI	3.4	4.2	2.8
8	A14	FE	4.2	4.6	2.8
9	A22	Optimization	3.4	2.2	2.33
10	A3	Covariates	3.4	2	2.3
11	A25	#PE	4	3.8	2.3
12	A12	CT	3.8	3.1	2.2
13	A1	DS	3.2	2.8	2.1
14	A16	#classes	2.2	1	2.1
15	A23	#optimization	2.8	2	2
16	A24	C. Validation	3.4	2.6	2
17	A10	AU	3	3.4	1.7
18	A13	#C	3	2.6	1.7
19	A26	ACC	3.4	2.2	1.6
20	A5	Ethnicity	3.4	1.5	1.56
21	A19	Seen AI (CV)	2.6	2.1	1.56
22	A9	MC	1.8	1	1.4
23	A27	SEN	2.8	2.9	1.3
24	A10	ADU/N	1	1.5	1.1
25	A45	ST	0	1.5	1.1
26	A11	S/NS	3.4	1.5	1
27	A15	FS	1.8	1.8	1
28	A18	S/UnS	1	1	1
29	A20	SL *Vs.* ML	1	1	1
30	A21	HU	0	0	0.6
31	A29	F1	0	0.5	0.6
32	A36	Chi-square	0	0.5	0.6
33	A41	MC	0	0	0.6
34	A46	HA	0	1	0.6
35	A30	Precision	2	1	0.56
36	A37	K&C	0	0.5	0.56
37	A28	SPEC	2.8	1.9	0.4
38	A32	Verification	0	0	0
39	A33	Variability	0	0	0
40	A34	DA	1	0	0
41	A38	DOR	0	0	0
42	A39	ANOVA	0	0	0
43	A40	Cl.V	0	0	0
44	A42	CS	0	0	0
45	A43	FDA	0	0	0
46	A44	CP/N	0	0	0

AN: Attribute name; NID: Non-Image Data; SN: Smoking; FH: Family History; HP: Hypertension; BMI: Body mass index; GT: Ground truth; FE: Feature extraction; #PE: Number of performance parameter; CT: Classifier type, DS: Data size; C. Validation: Cross-validation; AU: Architecture used; #C: Number of classifiers; ACC: Accuracy; MC: Multicentre; SEN: Sensitivity; AU/N: Another dataset used or not; ST: Survival test; S/NS: Supervised/Non-supervised; FS: Feature selection; S/UnS: Seen *Vs.* Unseen Data; SL *Vs.* ML: Single-label *vs.* Multilabel; HU: Hyper-parameters used; F1: F1 score; MC: Multicentre (Clinical evaluation); HA: Hazard analysis; K&C: Kappa-Cohen test; SPEC: Specificity; DA: Data augmentation; DOR: DOR test; ANOVA: ANOVA test; Cl.V: Clinical validation; CS: Clinical setting; FDA: Food and Drug Administration approved; CP/N: Clinically passed or not.

References

[1] S. Kaptoge, L. Pennells, D. De Bacquer, M.T. Cooney, M. Kavousi, G. Stevens, L. M. Riley, S. Savin, T. Khan, S. Altay, World Health Organization cardiovascular disease risk charts: revised models to estimate risk in 21 global regions, Lancet Global Health 7 (10) (2019) e1332–e1345.

[2] S.K. Banchhor, N.D. Londhe, T. Araki, L. Saba, P. Radeva, N.N. Khanna, J.S. Suri, Calcium detection, its quantification, and grayscale morphology-based risk stratification using machine learning in multimodality big data coronary and carotid scans: a review, Comput. Biol. Med. 101 (2018) 184–198.

[3] V. Viswanathan, A.D. Jamthikar, D. Gupta, N. Shanu, A. Puvvula, N.N. Khanna, L. Saba, T. Omerzum, K. Viskovic, S. Mavrogeni, M. Turk, J.R. Laird, G. Pareek, M. Miner, P.P. Sfikakis, A. Protogerou, G.D. Kitas, C. S, S. Joshi, H. Fiscian, A. A. Folson, D.H. Wu, Z. Ruzsa, A. Nicolaides, A. Sharma, D.L. Bhatt, J.S. Suri, Low-cost preventive screening using carotid ultrasound in patients with diabetes, Front. Biosci. 25 (2020) 1132–1171, https://doi.org/10.2741/4850.

[4] A.D. Jamthikar, A. Puvvula, D. Gupta, A.M. Johri, V. Nambi, N.N. Khanna, L. Saba, S. Mavrogeni, J.R. Laird, G. Pareek, Cardiovascular disease and stroke risk assessment in patients with chronic kidney disease using integration of estimated glomerular filtration rate, ultrasonic image phenotypes, and artificial intelligence: a narrative review, Int. Angiol.: a Journal of the International Union of Angiology 40(2) (2021) 150–164, https://doi.org/10.23736/S0392-9590.20.04538-1.

[5] V. Viswanathan, A.D. Jamthikar, D. Gupta, A. Puvvula, N.N. Khanna, L. Saba, K. Viskovic, S. Mavrogeni, M. Turk, J.R. Laird, Integration of estimated glomerular filtration rate biomarker in image-based cardiovascular disease/stroke risk calculator: a south Asian-Indian diabetes cohort with moderate chronic kidney disease, Int. Angiol.: a journal of the International Union of Angiology 39 (4) (2020) 290–306.

[6] A.D. Jamthikar, D. Gupta, A. Puvvula, A.M. Johri, N.N. Khanna, L. Saba, S. Mavrogeni, J.R. Laird, G. Pareek, M. Miner, P.P. Sfikakis, A. Protogerou, G. D. Kitas, R. Kolluri, A.M. Sharma, V. Viswanathan, V.S. Rathore, J.S. Suri, Cardiovascular risk assessment in patients with rheumatoid arthritis using carotid ultrasound B-mode imaging, Rheumatol. Int. 40 (12) (2020) 1921–1939.

[7] N.N. Khanna, A.D. Jamthikar, D. Gupta, M. Piga, L. Saba, C. Carcassi, A. A. Giannopoulos, A. Nicolaides, J.R. Laird, H.S. Suri, S. Mavrogeni, A. D. Protogerou, P. Sfikakis, G.D. Kitas, J.S. Suri, Rheumatoid arthritis: atherosclerosis imaging and cardiovascular risk assessment using machine and deep learning-based tissue characterization, Curr. Atherosclerosis Rep. 21 (2) (2019) 7.

[8] M. Porcu, L. Mannelli, M. Melis, J.S. Suri, C. Gerosa, G. Cerrone, G. Defazio, G. Faa, L. Saba, Carotid plaque imaging profiling in subjects with risk factors (diabetes and hypertension), Cardiovasc. Diagn. Ther. 10 (4) (2020) 1005.
[9] L. Saba, G. Micheletti, W. Brinjikji, P. Garofalo, R. Montisci, A. Balestrieri, J. Suri, J. DeMarco, G. Lanzino, R. Sanfilippo, Carotid intraplaque-hemorrhage volume and its association with cerebrovascular events, Am. J. Neuroradiol. 40 (10) (2019) 1731–1737.
[10] J. Hippisley-Cox, C. Coupland, P. Brindle, Development and validation of QRISK3 risk prediction algorithms to estimate future risk of cardiovascular disease: prospective cohort study, BMJ (2017) 357.
[11] Sr D'Agostino, R. B, R.S. Vasan, M.J. Pencina, P.A. Wolf, M. Cobain, J. M. Massaro, W.B. Kannel, General cardiovascular risk profile for use in primary care: the Framingham Heart Study, Circulation 117 (6) (2008) 743–753.
[12] R.M. Conroy, K. Pyörälä, A.e. Fitzgerald, S. Sans, A. Menotti, G. De Backer, D. De Bacquer, P. Ducimetiere, P. Jousilahti, U. Keil, Estimation of ten-year risk of fatal cardiovascular disease in Europe: the SCORE project, Eur. Heart J. 24 (11) (2003) 987–1003.
[13] P.M. Ridker, J.E. Buring, N. Rifai, N.R. Cook, Development and validation of improved algorithms for the assessment of global cardiovascular risk in women: the Reynolds Risk Score, JAMA 297 (6) (2007) 611–619.
[14] D.C. Goff, D.M. Lloyd-Jones, G. Bennett, S. Coady, R.B. D'agostino, R. Gibbons, P. Greenland, D.T. Lackland, D. Levy, C.J. O'donnell, ACC/AHA guideline on the assessment of cardiovascular risk: a report of the American College of Cardiology/American heart association task force on practice guidelines, J. Am. Coll. Cardiol. 63 (25 Part B) (2013) 2935–2959, 2014.
[15] E. Cuadrado-Godia, A.D. Jamthikar, D. Gupta, N.N. Khanna, T. Araki, M. Maniruzzaman, L. Saba, A. Nicolaides, A. Sharma, T. Omerzu, Ranking of stroke and cardiovascular risk factors for an optimal risk calculator design: logistic regression approach, Comput. Biol. Med. 108 (2019) 182–195.
[16] N.N. Khanna, A.D. Jamthikar, D. Gupta, A. Nicolaides, T. Araki, L. Saba, E. Cuadrado-Godia, A. Sharma, T. Omerzu, H.S. Suri, 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.
[17] M. Roffi, C. Patrono, J.P. Collet, C. Mueller, M. Valgimigli, F. Andreotti, J.J. Bax, M.A. Borger, C. Brotons, D.P. Chew, B. Gencer, G. Hasenfuss, K. Kjeldsen, P. Lancellotti, U. Landmesser, J. Mehilli, D. Mukherjee, R.F. Storey, S. Windecker, E.S.C.S.D. Group, 2015 ESC guidelines for the management of acute coronary syndromes in patients presenting without persistent ST-segment elevation: task force for the management of acute coronary syndromes in patients presenting without persistent ST-segment elevation of the European society of Cardiology (ESC), Eur. Heart J. 37 (3) (2016) 267–315.
[18] M. Task Force, G. Montalescot, U. Sechtem, S. Achenbach, F. Andreotti, C. Arden, A. Budaj, R. Bugiardini, F. Crea, T. Cuisset, C. Di Mario, J.R. Ferreira, B.J. Gersh, A.K. Gitt, J.S. Hulot, N. Marx, L.H. Opie, M. Pfisterer, E. Prescott, F. Ruschitzka, M. Sabate, R. Senior, D.P. Taggart, E.E. van der Wall, C.J. Vrints, , ESC Committee for Practice Guidelines (CPG), J.L. Zamorano, S. Achenbach, H. Baumgartner, J. J. Bax, H. Bueno, V. Dean, C. Deaton, C. Erol, R. Fagard, R. Ferrari, D. Hasdai, A. W. Hoes, P. Kirchhof, J. Knuuti, P. Kolh, P. Lancellotti, A. Linhart, P. Nihoyannopoulos, M.F. Piepoli, P. Ponikowski, P.A. Sirnes, J.L. Tamargo, M. Tendera, A. Torbicki, W. Wijns, S. Windecker, R. Document, J. Knuuti, M. Valgimigli, H. Bueno, M.J. Claeys, N. Donner-Banzhoff, C. Erol, H. Frank, C. Funck-Brentano, O. Gaemperli, J.R. Gonzalez-Juanatey, M. Hamilos, D. Hasdai, S. Husted, S.K. James, K. Kervinen, P. Kolh, S.D. Kristensen, P. Lancellotti, A.P. Maggioni, M.F. Piepoli, A.R. Pries, F. Romeo, L. Ryden, M. L. Simoons, P.A. Sirnes, P.G. Steg, A. Timmis, W. Wijns, S. Windecker, A. Yildirir, J.L. Zamorano, 2013 ESC guidelines on the management of stable coronary artery disease: the Task Force on the management of stable coronary artery disease of the European Society of Cardiology, Eur. Heart J. 34 (38) (2013) 2949–3003.
[19] N.N. Khanna, A.D. Jamthikar, T. Araki, D. Gupta, M. Piga, L. Saba, C. Carcassi, A. Nicolaides, J.R. Laird, H.S. Suri, Nonlinear model for the carotid artery disease 10-year risk prediction by fusing conventional cardiovascular factors to carotid ultrasound image phenotypes: a Japanese diabetes cohort study, Echocardiography 36 (2) (2019) 345–361.
[20] A.D. Jamthikar, D. Gupta, L. Saba, N.N. Khanna, K. Viskovic, S. Mavrogeni, J. R. Laird, N. Sattar, A.M. Johri, G. Pareek, M. Miner, P.P. Sfikakis, A. Protogerou, V. Viswanathan, A. Sharma, G.D. Kitas, A. Nicolaides, R. Kolluri, J.S. Suri, Artificial intelligence framework for predictive cardiovascular and stroke risk assessment models: a narrative review of integrated approaches using carotid ultrasound, Comput. Biol. Med. 126 (2020) 104043.
[21] U.R. Acharya, O. Faust, V. Sree, G. Swapna, R.J. Martis, N.A. Kadri, J.S. Suri, Linear and nonlinear analysis of normal and CAD-affected heart rate signals, Comput. Methods Progr. Biomed. 113 (1) (2014) 55–68.
[22] L. Cui, Y. Xing, Y. Zhou, L. Wang, K. Liu, D. Zhang, Y. Chen, Carotid intraplaque neovascularisation as a predictive factor for future vascular events in patients with mild and moderate carotid stenosis: an observational prospective study, Ther Adv Neurol Disord 14 (2021), 17562864211023992.
[23] J.E. Herr, M.-F. Hétu, T.Y. Li, P. Ewart, A.M. Johri, Presence of calcium-like tissue composition in carotid plaque is indicative of significant coronary artery disease in high-risk patients, J. Am. Soc. Echocardiogr. 32 (5) (2019) 633–642.
[24] A.M. Johri, K.A. Lajkosz, N. Grubic, S. Islam, T.Y. Li, C.S. Simpson, P. Ewart, J. S. Suri, M.F. Hetu, Maximum plaque height in carotid ultrasound predicts cardiovascular disease outcomes: a population-based validation study of the American society of echocardiography's grade II-III plaque characterization and protocol, Int. J. Cardiovasc. Imag. 37 (5) (2021) 1601–1610.
[25] S. Delsanto, F. Molinari, P. Giustetto, W. Liboni, S. Badalamenti, J.S. Suri, Characterization of a completely user-independent algorithm for carotid artery segmentation in 2-D ultrasound images, IEEE Trans. Instrum. Meas. 56 (4) (2007) 1265–1274.
[26] N.N. Khanna, A.D. Jamthikar, D. Gupta, T. Araki, M. Piga, L. Saba, C. Carcassi, A. Nicolaides, J.R. Laird, H.S. Suri, A. Gupta, S. Mavrogeni, A. Protogerou, P. Sfikakis, G.D. Kitas, J.S. Suri, Effect of carotid image-based phenotypes on cardiovascular risk calculator: AECRS1.0, Med. Biol. Eng. Comput. 57 (7) (2019) 1553–1566.
[27] F. Molinari, C.S. Pattichis, G. Zeng, L. Saba, U.R. Acharya, R. Sanfilippo, A. Nicolaides, J.S. Suri, Completely automated multiresolution edge snapper–a new technique for an accurate carotid ultrasound IMT measurement: clinical validation and benchmarking on a multi-institutional database, IEEE Trans. Image Process. 21 (3) (2012) 1211–1222.
[28] L. Saba, A. Jamthikar, D. Gupta, N.N. Khanna, K. Viskovic, H.S. Suri, A. Gupta, S. Mavrogeni, M. Turk, J.R. Laird, G. Pareek, M. Miner, P.P. Sfikakis, A. Protogerou, G.D. Kitas, V. Viswanathan, A. Nicolaides, D.L. Bhatt, J.S. Suri, Global perspective on carotid intima-media thickness and plaque: should the current measurement guidelines be revisited? Int. Angiol. 38 (6) (2019) 451–465.
[29] L. Saba, K.M. Meiburger, F. Molinari, G. Ledda, M. Anzidei, U.R. Acharya, G. Zeng, S. Shafique, A. Nicolaides, J.S. Suri, Carotid IMT variability (IMTV) and its validation in symptomatic versus asymptomatic Italian population: can this be a useful index for studying symptomaticity? Echocardiography 29 (9) (2012) 1111–1119.
[30] P.J. Touboul, M.G. Hennerici, S. Meairs, H. Adams, P. Amarenco, N. Bornstein, L. Csiba, M. Desvarieux, S. Ebrahim, M. Fatar, R. Hernandez Hernandez, M. Jaff, S. Kownator, P. Prati, T. Rundek, M. Sitzer, U. Schminke, J.C. Tardif, A. Taylor, E. Vicaut, K.S. Woo, F. Zannad, M. Zureik, Mannheim carotid intima-media thickness consensus (2004-2006). An update on behalf of the advisory board of the 3rd and 4th watching the risk symposium, 13th and 15th European stroke conferences, Mannheim, Germany, 2004, and Brussels, Belgium, 2006, Cerebrovasc. Dis. 23 (1) (2007) 75–80.
[31] L. Saba, N. Agarwal, R. Cau, C. Gerosa, R. Sanfilippo, M. Porcu, R. Montisci, G. Cerrone, Y. Qi, A. Balestrieri, Review of Imaging Biomarkers for the Vulnerable Carotid Plaque, JVS: Vascular Science, 2021.
[32] L. Saba, J.S. Suri, Multi-Detector CT Imaging: Principles, Head, Neck, and Vascular Systems, vol. 1, CRC Press, 2013.
[33] J. Seabra, J. Sanches, Ultrasound Imaging: Advances and Applications, Springer, New York, 2012.
[34] P. Radeva, J.S. Suri, Vascular and intravascular imaging trends, analysis, and challenges, volume 2; plaque characterization, Vascular Intravascular Imaging Trends (2019).
[35] K. Liu, J.S. Suri, Automatic Vessel Indentification for Angiographic Screening, Google Patents, 2005.
[36] N.D. Londhe, J.S. Suri, Superharmonic imaging for medical ultrasound: a review, J. Med. Syst. 40 (12) (2016) 1–16.
[37] J.M. Sanches, A.F. Laine, J.S. Suri, Ultrasound Imaging, Springer, 2012.
[38] A.M. Johri, K.A. Lajkosz, N. Grubic, S. Islam, T.Y. Li, C.S. Simpson, P. Ewart, J. S. Suri, M.-F. Hétu, Maximum plaque height in carotid ultrasound predicts cardiovascular disease outcomes: a population-based validation study of the American society of echocardiography's grade II–III plaque characterization and protocol, Int. J. Cardiovasc. Imag. 37 (5) (2021) 1601–1610.
[39] N. Ikeda, A. Gupta, N. Dey, S. Bose, S. Shafique, T. Arak, E.C. Godia, L. Saba, J. R. Laird, A. Nicolaides, biology, Improved correlation between carotid and coronary atherosclerosis SYNTAX score using automated ultrasound carotid bulb plaque IMT measurement, Ultrasound in medicine 41 (5) (2015) 1247–1262.
[40] N. Ikeda, N. Dey, A. Sharma, A. Gupta, S. Bose, S. Acharjee, S. Shafique, E. Cuadrado-Godia, T. Araki, L. Saba, Automated segmental-IMT measurement in thin/thick plaque with bulb presence in carotid ultrasound from multiple scanners: stroke risk assessment, Comput. Methods Progr. Biomed. 141 (2017) 73–81.
[41] N. Ikeda, T. Araki, N. Dey, S. Bose, S. Shafique, A. El-Baz, J. Suri, Automated and accurate carotid bulb detection, its verification and validation in low quality frozen frames and motion video, Int. Angiol. 33 (6) (2014) 573–589.
[42] M. Biswas, L. Saba, S. Chakrabartty, N.N. Khanna, H. Song, H.S. Suri, P. P. Sfikakis, S. Mavrogeni, K. Viskovic, J.R. Laird, E. Cuadrado-Godia, A. Nicolaides, A. Sharma, V. Viswanathan, A. Protogerou, G. Kitas, G. Pareek, M. Miner, J.S. Suri, Two-stage artificial intelligence model for jointly measurement of atherosclerotic wall thickness and plaque burden in carotid ultrasound: a screening tool for cardiovascular/stroke risk assessment, Comput. Biol. Med. 123 (2020) 103847.
[43] V. Viswanathan, A.D. Jamthikar, D. Gupta, N. Shanu, A. Puvvula, N.N. Khanna, L. Saba, T. Omerzum, K. Viskovic, S. Mavrogeni, Low-cost preventive screening using carotid ultrasound in patients with diabetes, Front. Biosci. 25 (2020) 1132–1171.
[44] L. Saba, F. Molinari, K. Meiburger, M. Piga, G. Zeng, R. Achraya, A. Nicolaides, J. Suri, What is the correct distance measurement metric when measuring carotid ultrasound intima-media thickness automatically? Int. Angiol.: a journal of the International Union of Angiology 31 (5) (2012) 483–489.
[45] F. Molinari, K.M. Meiburger, L. Saba, U.R. Acharya, L. Famiglietti, N. Georgiou, A. Nicolaides, R.S. Mamidi, H. Kuper, J.S. Suri, Automated carotid IMT measurement and its validation in low contrast ultrasound database of 885 patient Indian population epidemiological study: results of AtheroEdge® software, in: Multi-modality Atherosclerosis Imaging and Diagnosis, Springer, 2014, pp. 209–219.

[46] F. Molinari, K.M. Meiburger, J. Suri, Automated high-performance cIMT measurement techniques using patented AtheroEdge™: a screening and home monitoring system, in: 2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, IEEE, 2011.

[47] A. Jamthikar, D. Gupta, N.N. Khanna, T. Araki, L. Saba, A. Nicolaides, A. Sharma, T. Omerzu, H.S. Suri, A. Gupta, S. Mavrogeni, M. Turk, J.R. Laird, A. Protogerou, P.P. Sfikakis, G.D. Kitas, V. Viswanathan, G. Pareek, M. Miner, J.S. Suri, A special report on changing trends in preventive stroke/cardiovascular risk assessment via B-mode ultrasonography, Curr. Atherosclerosis Rep. 21 (7) (2019) 25.

[48] A.D. Jamthikar, A. Puvvula, D. Gupta, A.M. Johri, V. Nambi, N.N. Khanna, L. Saba, S. Mavrogeni, J.R. Laird, G. Pareek, M. Miner, P.P. Sfikakis, A. Protogerou, G.D. Kitas, A. Nicolaides, A.M. Sharma, V. Viswanathan, V. S. Rathore, R. Kolluri, D.L. Bhatt, J.S. Suri, Cardiovascular disease and stroke risk assessment in patients with chronic kidney disease using integration of estimated glomerular filtration rate, ultrasonic image phenotypes, and artificial intelligence: a narrative review, Int. Angiol. 40 (2) (2021) 150–164.

[49] V. Viswanathan, A. Puvvula, A.D. Jamthikar, L. Saba, A.M. Johri, V. Kotsis, N. N. Khanna, S.K. Dhanjil, M. Majhail, D.P. Misra, Bidirectional link between diabetes mellitus and coronavirus disease 2019 leading to cardiovascular disease: a narrative review, World J. Diabetes 12 (3) (2021) 215.

[50] L. Saba, M. Biswas, V. Kuppili, E.C. Godia, H.S. Suri, D.R. Edla, T. Omerzu, J. R. Laird, N.N. Khanna, S. Mavrogeni, The present and future of deep learning in radiology, Eur. J. Radiol. 114 (2019) 14–24.

[51] M. Biswas, V. Kuppili, L. Saba, D.R. Edla, H.S. Suri, E. Cuadrado-Godia, J.R. Laird, R.T. Marinhoe, J.M. Sanches, A. Nicolaides, State-of-the-art review on deep learning in medical imaging, Front. Biosci. 24 (2019) 392–426.

[52] M. Maniruzzaman, N. Kumar, M.M. Abedin, M.S. Islam, H.S. Suri, A.S. El-Baz, J. S. Suri, Comparative approaches for classification of diabetes mellitus data: machine learning paradigm, Comput. Methods Progr. Biomed. 152 (2017) 23–34.

[53] M. Maniruzzaman, M.J. Rahman, M. Al-MehediHasan, H.S. Suri, M.M. Abedin, A. El-Baz, J.S. Suri, Accurate diabetes risk stratification using machine learning: role of missing value and outliers, J. Med. Syst. 42 (5) (2018) 1–17.

[54] V. Kuppili, M. Biswas, A. Sreekumar, H.S. Suri, L. Saba, D.R. Edla, R.T. Marinhoe, J.M. Sanches, J.S. Suri, Extreme learning machine framework for risk stratification of fatty liver disease using ultrasound tissue characterization, J. Med. Syst. 41 (10) (2017) 1–20.

[55] L. Saba, N. Dey, A.S. Ashour, S. Samanta, S.S. Nath, S. Chakraborty, J. Sanches, D. Kumar, R. Marinho, J.S. Suri, Automated stratification of liver disease in ultrasound: an online accurate feature classification paradigm, Comput. Methods Progr. Biomed. 130 (2016) 118–134.

[56] U.R. Acharya, S.V. Sree, R. Ribeiro, G. Krishnamurthi, R.T. Marinho, J. Sanches, J.S. Suri, Data mining framework for fatty liver disease classification in ultrasound: a hybrid feature extraction paradigm, Med. Phys. 39 (7) (2012) 4255–4264.

[57] U.R. Acharya, S.V. Sree, M.M.R. Krishnan, F. Molinari, R. Garberoglio, J.S. Suri, Non-invasive automated 3D thyroid lesion classification in ultrasound: a class of ThyroScan™ systems, Ultrasonics 52 (4) (2012) 508–520.

[58] U.R. Acharya, O. Faust, S.V. Sree, F. Molinari, R. Garberoglio, J. Suri, Cost-effective and non-invasive automated benign & malignant thyroid lesion classification in 3D contrast-enhanced ultrasound using combination of wavelets and textures: a class of ThyroScan™ algorithms, Technol. Cancer Res. Treat. 10 (4) (2011) 371–380.

[59] A. Murgia, A. Balestrieri, P. Crivelli, J.S. Suri, M. Conti, F. Cademartiri, L. Saba, Therapy, Cardiac computed tomography radiomics: an emerging tool for the non-invasive assessment of coronary atherosclerosis, Cardiovasc. Diagn. Ther. 10 (6) (2020) 2005.

[60] U.R. Acharya, S.V. Sree, M.M.R. Krishnan, N. Krishnananda, S. Ranjan, P. Umesh, J.S. Suri, Automated classification of patients with coronary artery disease using grayscale features from left ventricle echocardiographic images, Comput. Methods Progr. Biomed. 112 (3) (2013) 624–632.

[61] G. Pareek, U.R. Acharya, S.V. Sree, G. Swapna, R. Yantri, R.J. Martis, L. Saba, G. Krishnamurthi, G. Mallarini, A. El-Baz, Prostate tissue characterization/classification in 144 patient population using wavelet and higher order spectra features from transrectal ultrasound images, Technol. Cancer Res. Treat. 12 (6) (2013) 545–557.

[62] U.R. Acharya, S.V. Sree, S. Kulshreshtha, F. Molinari, J.E.W. Koh, L. Saba, J. S. Suri, GyneScan: an improved online paradigm for screening of ovarian cancer via tissue characterization, Technol. Cancer Res. Treat. 13 (6) (2014) 529–539.

[63] V.K. Shrivastava, N.D. Londhe, R.S. Sonawane, J.S. Suri, Exploring the color feature power for psoriasis risk stratification and classification: a data mining paradigm, Comput. Biol. Med. 65 (2015) 54–68.

[64] V.K. Shrivastava, N.D. Londhe, R.S. Sonawane, J.S. Suri, Computer-aided diagnosis of psoriasis skin images with HOS, texture and color features: a first comparative study of its kind, Comput. Methods Progr. Biomed. 126 (2016) 98–109.

[65] U.R. Acharya, O. Faust, S.V. Sree, F. Molinari, L. Saba, A. Nicolaides, J.S. Suri, measurement, An accurate and generalized approach to plaque characterization in 346 carotid ultrasound scans, IEEE Trans. Instrum. Meas. 61 (4) (2011) 1045–1053.

[66] 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, Stroke risk stratification and its validation using ultrasonic echolucent carotid wall plaque morphology: a machine learning paradigm, Comput. Biol. Med. 80 (2017) 77–96.

[67] U.R. Acharya, O. Faust, A. Alvin, G. Krishnamurthi, J.C. Seabra, J. Sanches, J. S. Suri, Understanding symptomatology of atherosclerotic plaque by image-based tissue characterization, Comput. Methods Progr. Biomed. 110 (1) (2013) 66–75.

[68] J.S. Suri, Imaging Based Symptomatic Classification and Cardiovascular Stroke Risk Score Estimation, Google Patents, 2011.

[69] A.D. Jamthikar, D. Gupta, L.E. Mantella, L. Saba, J.R. Laird, A.M. Johri, J.S. Suri, Multiclass machine learning vs. conventional calculators for stroke/CVD risk assessment using carotid plaque predictors with coronary angiography scores as gold standard: a 500 participants study, Int. J. Cardiovasc. Imag. 37 (4) (2021) 1171–1187.

[70] A.M. Johri, L.E. Mantella, A.D. Jamthikar, L. Saba, J.R. Laird, J.S. Suri, Role of artificial intelligence in cardiovascular risk prediction and outcomes: comparison of machine-learning and conventional statistical approaches for the analysis of carotid ultrasound features and intra-plaque neovascularization, Int. J. Cardiovasc. Imag. 37(11): (2021) 3145–3156, https://doi.org/10.1007/s10554-021-02294-0.

[71] C.-C. Wu, W.-D. Hsu, Y.-C. Wang, W.-M. Kung, I.S. Tzeng, C.-W. Huang, C.-Y. Huang, Y.-C. Li, An innovative scoring system for predicting major adverse cardiac events in patients with chest pain based on machine learning, IEEE Access 8 (2020) 124076–124083.

[72] I.A. Kakadiaris, M. Vrigkas, A.A. Yen, T. Kuznetsova, M. Budoff, M. Naghavi, Machine learning outperforms ACC/AHA CVD risk calculator in MESA, J. Am. Heart Assoc. 7 (22) (2018) e009476.

[73] C.Y. Cheung, D. Xu, C.Y. Cheng, C. Sabanayagam, Y.C. Tham, M. Yu, T.H. Rim, C. Y. Chai, B. Gopinath, P. Mitchell, R. Poulton, T.E. Moffitt, A. Caspi, J.C. Yam, C. C. Tham, J.B. Jonas, Y.X. Wang, S.J. Song, L.M. Burrell, O. Farouque, L.J. Li, G. Tan, D.S.W. Ting, W. Hsu, M.L. Lee, T.Y. Wong, A deep-learning system for the assessment of cardiovascular disease risk via the measurement of retinal-vessel calibre, Nat Biomed Eng 5 (6) (2021) 498–508.

[74] M.J. Emaus, I. Isgum, S.G.M. van Velzen, H. van den Bongard, S.A.M. Gernaat, N. Lessmann, M.G.A. Sattler, A.J. Teske, J. Penninkhof, H. Meijer, J.P. Pignol, H. M. Verkooijen, g. Bragatston study, Bragatston study protocol: a multicentre cohort study on automated quantification of cardiovascular calcifications on radiotherapy planning CT scans for cardiovascular risk prediction in patients with breast cancer, BMJ Open 9 (7) (2019) e028752.

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

[76] T. Pattarabanjird, C. Cress, A. Nguyen, A. Taylor, S. Bekiranov, C. McNamara, A machine learning model utilizing a novel SNP shows enhanced prediction of coronary artery disease severity, Genes 11 (12) (2020) 1446.

[77] F. Sanchez-Cabo, X. Rossello, V. Fuster, F. Benito, J.P. Manzano, J.C. Silla, J. M. Fernandez-Alvira, B. Oliva, L. Fernandez-Friera, B. Lopez-Melgar, J. M. Mendiguren, J. Sanz, J.M. Ordovas, V. Andres, A. Fernandez-Ortiz, H. Bueno, B. Ibanez, J.M. Garcia-Ruiz, E. Lara-Pezzi, Machine learning improves cardiovascular risk definition for young, asymptomatic individuals, J. Am. Coll. Cardiol. 76 (14) (2020) 1674–1685.

[78] L.B. van den Oever, L. Cornelissen, M. Vonder, C. Xia, J.N. van Bolhuis, R. Vliegenthart, R.N. Veldhuis, G.H. de Bock, M. Oudkerk, P.M. van Ooijen, Deep learning for automated exclusion of cardiac CT examinations negative for coronary artery calcium, Eur. J. Radiol. 129 (2020) 109114.

[79] A.K. Wennstig, H. Garmo, U. Isacsson, G. Gagliardi, N. Rintela, B. Lagerqvist, L. Holmberg, C. Blomqvist, M. Sund, G. Nilsson, The relationship between radiation doses to coronary arteries and location of coronary stenosis requiring intervention in breast cancer survivors, Radiat. Oncol. 14 (1) (2019) 40.

[80] A.C. Dimopoulos, M. Nikolaidou, F.F. Caballero, W. Engchuan, A. Sanchez-Niubo, H. Arndt, J.L. Ayuso-Mateos, J.M. Haro, S. Chatterji, E.N. Georgousopoulou, C. Pitsavos, D.B. Panagiotakos, Machine learning methodologies versus cardiovascular risk scores, in predicting disease risk, BMC Med. Res. Methodol. 18 (1) (2018) 179.

[81] G. Joo, Y. Song, H. Im, J. Park, Clinical implication of machine learning in predicting the occurrence of cardiovascular disease using big data (Nationwide Cohort Data in Korea), IEEE Access 8 (2020) 157643–157653.

[82] J.O.R. Kim, Y.S. Jeong, J.H. Kim, J.W. Lee, D. Park, H.S. Kim, Machine learning-based cardiovascular disease prediction model: a cohort study on the Korean National health insurance service health screening database, Diagnostics 11 (6) (2021).

[83] D. Panaretos, E. Koloverou, A.C. Dimopoulos, G.M. Kouli, M. Vamvakari, G. Tzavelas, C. Pitsavos, D.B. Panagiotakos, A comparison of statistical and machine-learning techniques in evaluating the association between dietary patterns and 10-year cardiometabolic risk (2002-2012): the ATTICA study, Br. J. Nutr. 120 (3) (2018) 326–334.

[84] X. Su, Y. Xu, Z. Tan, X. Wang, P. Yang, Y. Su, Y. Jiang, S. Qin, L. Shang, Prediction for cardiovascular diseases based on laboratory data: an analysis of random forest model, J. Clin. Lab. Anal. 34 (9) (2020) e23421.

[85] 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) e0174944.

[86] L. Yang, H. Wu, X. Jin, P. Zheng, S. Hu, X. Xu, W. Yu, J. Yan, Study of cardiovascular disease prediction model based on random forest in eastern China, Sci. Rep. 10 (1) (2020) 1–8.

[87] S. Sajeev, S. Champion, A. Beleigoli, D. Chew, R.L. Reed, D.J. Magliano, J. E. Shaw, R.L. Milne, S. Appleton, T.K. Gill, A. Maeder, Predicting Australian adults at high risk of cardiovascular disease mortality using standard risk factors and machine learning, Int. J. Environ. Res. Publ. Health 18 (6) (2021).

[88] P. Unnikrishnan, D.K. Kumar, S. Poosapadi Arjunan, H. Kumar, P. Mitchell, R. Kawasaki, Development of health parameter model for risk prediction of CVD using SVM, Comput Math Methods Med, 2016 (2016) 3016245.

[89] K. Zarkogianni, M. Athanasiou, A.C. Thanopoulou, Comparison of machine learning approaches toward assessing the risk of developing cardiovascular disease as a long-term diabetes complication, IEEE J Biomed Health Inform 22 (5) (2018) 1637–1647.

[90] A. Kenne Malaha, J. Magne, L. Jarlan, K. Mansour, M. Ait-Ouatet, S. Galinat, M. P. Teissier, P. Lacroix, I. Desormais, V. Aboyans, Vascular ultrasound for cardiovascular risk stratification in asymptomatic patients with type-2 diabetes, Prim Care Diabetes 15 (4) (2021) 726–732.

[91] A.D. Jamthikar, D. Gupta, A.M. Johri, L.E. Mantella, L. Saba, R. Kolluri, A. M. Sharma, V. Viswanathan, A. Nicolaides, J.S. Suri, Low-cost office-based cardiovascular risk stratification using machine learning and focused carotid ultrasound in an asian-Indian cohort, J. Med. Syst. 44 (12) (2020) 208.

[92] H. Chao, H. Shan, F. Homayounieh, R. Singh, R.D. Khera, H. Guo, T. Su, G. Wang, M.K. Kalra, P. Yan, Deep learning predicts cardiovascular disease risks from lung cancer screening low dose computed tomography, Nat. Commun. 12 (1) (2021) 2963.

[93] Z. Huang, W. Dong, H. Duan, J. Liu, A regularized deep learning approach for clinical risk prediction of acute coronary syndrome using electronic health records, IEEE (Inst. Electr. Electron. Eng.) Trans. Biomed. Eng. 65 (5) (2017) 956–968.

[94] A. Jamthikar, D. Gupta, E. Cuadrado-Godia, A. Puvvula, N.N. Khanna, L. Saba, K. Viskovic, S. Mavrogeni, M. Turk, J.R. Laird, Ultrasound-based stroke/cardiovascular risk stratification using Framingham risk score and ASCVD risk score based on "integrated vascular age" instead of "chronological age": a multi-ethnic study of asian Indian, Caucasian, and Japanese cohorts, Cardiovasc. Diagn. Ther. 10 (4) (2020) 939.

[95] A.M. Johri, C.M. Calnan, M.F. Matangi, J. MacHaalany, M.-F. Hétu, Focused vascular ultrasound for the assessment of atherosclerosis: a proof-of-concept study, J. Am. Soc. Echocardiogr. 29 (9) (2016) 842–849.

[96] H.T. Jørstad, E.B. Colkesen, S.M. Boekholdt, J.G. Tijssen, N.J. Wareham, K.-T. Khaw, R.J. Peters, Estimated 10-year cardiovascular mortality seriously underestimates overall cardiovascular risk, Heart 102 (1) (2016) 63–68.

[97] N.N. Khanna, A.D. Jamthikar, D. Gupta, A. Nicolaides, T. Araki, L. Saba, E. Cuadrado-Godia, A. Sharma, T. Omerzu, H.S. Suri, A. Gupta, S. Mavrogeni, M. Turk, J.R. Laird, A. Protogerou, P.P. Sfikakis, G.D. Kitas, V. Viswanathan, J. S. Suri, 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.

[98] V. Kotsis, A.D. Jamthikar, T. Araki, D. Gupta, J.R. Laird, A.A. Giannopoulos, L. Saba, H.S. Suri, S. Mavrogeni, G.D. Kitas, K. Viskovic, N.N. Khanna, A. Gupta, A. Nicolaides, J.S. Suri, Echolucency-based phenotype in carotid atherosclerosis disease for risk stratification of diabetes patients, Diabetes Res. Clin. Pract. 143 (2018) 322–331.

[99] A. Puvvula, A.D. Jamthikar, D. Gupta, N.N. Khanna, M. Porcu, L. Saba, K. Viskovic, J.N. Ajuluchukwu, A. Gupta, S. Mavrogeni, Morphological carotid plaque area is associated with glomerular filtration rate: a study of south asian indian patients with diabetes and chronic kidney disease, Angiology 71 (6) (2020) 520–535.

[100] J.H. Stein, C.E. Korcarz, R.T. Hurst, E. Lonn, C.B. Kendall, E.R. Mohler, S. S. Najjar, C.M. Rembold, W.S. Post, Use of carotid ultrasound to identify subclinical vascular disease and evaluate cardiovascular disease risk: a consensus statement from the American Society of Echocardiography Carotid Intima-Media Thickness Task Force endorsed by the Society for Vascular Medicine, J. Am. Soc. Echocardiogr. 21 (2) (2008) 93–111.

[101] V. Viswanathan, A.D. Jamthikar, D. Gupta, A. Puvvula, N.N. Khanna, L. Saba, K. Viskovic, S. Mavrogeni, J.R. Laird, G. Pareek, M. Miner, P.P. Sfikakis, A. Protogerou, A. Sharma, P. Kancharana, D.P. Misra, V. Agarwal, G.D. Kitas, A. Nicolaides, J.S. Suri, Does the carotid bulb offer a better 10-year CVD/stroke risk assessment compared to the common carotid artery? A 1516 ultrasound scan study, Angiology 71 (10) (2020) 920–933.

[102] E. Cuadrado-Godia, M. Maniruzzaman, T. Araki, A. Puvvula, M. Jahanur Rahman, L. Saba, H.S. Suri, A. Gupta, S.K. Banchhor, J.S. Teji, T. Omerzu, N.N. Khanna, J. R. Laird, A. Nicolaides, S. Mavrogeni, G.D. Kitas, J.S. Suri, A. Fellow, Morphologic TPA (mTPA) and composite risk score for moderate carotid atherosclerotic plaque is strongly associated with HbA1c in diabetes cohort, Comput. Biol. Med. 101 (2018) 128–145.

[103] Z. Huang, W. Dong, H. Duan, J. Liu, A regularized deep learning approach for clinical risk prediction of acute coronary syndrome using electronic health records, IEEE Trans. Biomed. Eng. 65 (5) (2018) 956–968.

[104] J.H. Stein, C.E. Korcarz, R.T. Hurst, E. Lonn, C.B. Kendall, E.R. Mohler, S. S. Najjar, C.M. Rembold, W.S. Post, F. American, Society of echocardiography carotid intima-media thickness task, Use of carotid ultrasound to identify subclinical vascular disease and evaluate cardiovascular disease risk: a consensus statement from the American Society of Echocardiography Carotid Intima-Media Thickness Task Force. Endorsed by the Society for Vascular Medicine. J Am Soc Echocardiogr 21 (2) (2008) 93–111, quiz 189-90.

[105] E. Cuadrado-Godia, A.D. Jamthikar, D. Gupta, N.N. Khanna, T. Araki, M. Maniruzzaman, L. Saba, A. Nicolaides, A. Sharma, T. Omerzu, H.S. Suri, A. Gupta, S. Mavrogeni, M. Turk, J.R. Laird, A. Protogerou, P. Sfikakis, G.D. Kitas, V. Viswanathan, J.S. Suri, Ranking of stroke and cardiovascular risk factors for an optimal risk calculator design: logistic regression approach, Comput. Biol. Med. 108 (2019) 182–195.

[106] A.M. Johri, C.M. Calnan, M.F. Matangi, J. MacHaalany, M.F. Hetu, Focused vascular ultrasound for the assessment of atherosclerosis: a proof-of-concept study, J. Am. Soc. Echocardiogr. 29 (9) (2016) 842–849.

[107] H.T. Jorstad, E.B. Colkesen, S.M. Boekholdt, J.G. Tijssen, N.J. Wareham, K. T. Khaw, R.J. Peters, Estimated 10-year cardiovascular mortality seriously underestimates overall cardiovascular risk, Heart 102 (1) (2016) 63–68.

[108] J. Suri, S. Agarwal, S.K. Gupta, A. Puvvula, K. Viskovic, N. Suri, A. Alizad, A. El-Baz, L. Saba, M. Fatemi, Systematic review of artificial intelligence in acute respiratory distress syndrome for COVID-19 lung patients: a Biomedical imaging perspective, IEEE Journal of Biomedical Health Informatics 25(11): (2021 Nov) 4128–4139, https://doi.org/10.1109/JBHI.2021.3103839.

[109] J.S. Suri, A. Puvvula, M. Majhail, M. Biswas, A.D. Jamthikar, L. Saba, G. Faa, I. M. Singh, R. Oberleitner, M. Turk, S. Srivastava, P.S. Chadha, H.S. Suri, A. M. Johri, V. Nambi, J.M. Sanches, N.N. Khanna, K. Viskovic, S. Mavrogeni, J. R. Laird, A. Bit, G. Pareek, M. Miner, A. Balestrieri, P.P. Sfikakis, G. Tsoulfas, A. Protogerou, D.P. Misra, V. Agarwal, G.D. Kitas, R. Kolluri, J. Teji, M. Porcu, M. Al-Maini, A. Agbakoba, M. Sockalingam, A. Sexena, A. Nicolaides, A. Sharma, V. Rathore, V. Viswanathan, S. Naidu, D.L. Bhatt, Integration of cardiovascular risk assessment with COVID-19 using artificial intelligence, Rev. Cardiovasc. Med. 21 (4) (2020) 541–560.

[110] P. Sudeep, P. Palanisamy, J. Rajan, H. Baradaran, L. Saba, A. Gupta, J.S. Suri, Speckle reduction in medical ultrasound images using an unbiased non-local means method. Biomedical Signal Processing and Control 28, Elsevier, 2016, pp. 1–8.

[111] J.S. Suri, D. Kumar, Medical Image Enhancement System, Google Patents, 2008.

[112] L. Saba, S.K. Banchhor, T. Araki, H.S. Suri, N.D. Londhe, J.R. Laird, K. Viskovic, J. S. Suri, Intra-and inter-operator reproducibility analysis of automated cloud-based carotid intima media thickness ultrasound measurement, J. Clin. Diagn. Res. 12 (2) (2018).

[113] L. Saba, J.C. Than, N.M. Noor, O.M. Rijal, R.M. Kassim, A. Yunus, C.R. Ng, J. S. Suri, Inter-observer variability analysis of automatic lung delineation in normal and disease patients, J. Med. Syst. 40 (6) (2016) 142.

[114] A. Jamthikar, D. Gupta, L. Saba, N.N. Khanna, T. Araki, K. Viskovic, S. Mavrogeni, J.R. Laird, G. Pareek, M. Miner, P.P. Sfikakis, A. Protogerou, V. Viswanathan, A. Sharma, A. Nicolaides, G.D. Kitas, J.S. Suri, Cardiovascular/stroke risk predictive calculators: a comparison between statistical and machine learning models, Cardiovasc. Diagn. Ther. 10 (4) (2020) 919–938.

[115] S. Benjamens, P. Dhunnoo, B. Meskó, The state of artificial intelligence-based FDA-approved medical devices and algorithms: an online database, NPJ digital medicine 3 (1) (2020) 1–8.

[116] P.F. Whiting, A.W. Rutjes, M.E. Westwood, S. Mallett, J.J. Deeks, J.B. Reitsma, M. M. Leeflang, J.A. Sterne, P.M. Bossuyt, Q.- Group*, QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies, Ann. Intern. Med. 155 (8) (2011) 529–536.

[117] L. Saba, R. Sanfilippo, S. Sannia, M. Anzidei, R. Montisci, G. Mallarini, J.S. Suri, Association between carotid artery plaque volume, composition, and ulceration: a retrospective assessment with MDCT, Am. J. Roentgenol. 199 (1) (2012) 151–156.