Title
Machine Learning Outperforms ACC/AHA CVD Risk Calculator in MESA

Journal
Journal of the American Heart Association, 7(22)

ISSN
2047-9980

Authors
Kakadiaris, Ioannis A
Vrigkas, Michalis
Yen, Albert A
et al.

Publication Date
2018-11-20

DOI
10.1161/jaha.118.009476

Peer reviewed

ORIGINAL RESEARCH

Machine Learning Outperforms ACC/AHA CVD Risk Calculator in MESA

Ioannis A. Kakadiaris, PhD; Michalis Vrigkas, PhD; Albert A. Yen, MD; Tatiana Kuznetsova, MD; Matthew Budoff, MD; Morteza Naghavi, MD

***Background*—**Studies have demonstrated that the current US guidelines based on American College of Cardiology/American Heart Association (ACC/AHA) Pooled Cohort Equations Risk Calculator may underestimate risk of atherosclerotic cardiovascular disease (CVD) in certain high-risk individuals, therefore missing opportunities for intensive therapy and preventing CVD events. Similarly, the guidelines may overestimate risk in low risk populations resulting in unnecessary statin therapy. We used Machine Learning (ML) to tackle this problem.

***Methods and Results*—**We developed a ML Risk Calculator based on Support Vector Machines (SVMs) using a 13-year follow up data set from MESA (the Multi-Ethnic Study of Atherosclerosis) of 6459 participants who were atherosclerotic CVD-free at baseline. We provided identical input to both risk calculators and compared their performance. We then used the FLEMENGHO study (the Flemish Study of Environment, Genes and Health Outcomes) to validate the model in an external cohort. ACC/AHA Risk Calculator, based on 7.5% 10-year risk threshold, recommended statin to 46.0%. Despite this high proportion, 23.8% of the 480 "Hard CVD" events occurred in those not recommended statin, resulting in sensitivity 0.76, specificity 0.56, and AUC 0.71. In contrast, ML Risk Calculator recommended only 11.4% to take statin, and only 14.4% of "Hard CVD" events occurred in those not recommended statin, resulting in sensitivity 0.86, specificity 0.95, and AUC 0.92. Similar results were found for prediction of "All CVD" events.

***Conclusions*—**The ML Risk Calculator outperformed the ACC/AHA Risk Calculator by recommending less drug therapy, yet missing fewer events. Additional studies are underway to validate the ML model in other cohorts and to explore its ability in short-term CVD risk prediction. **(*J Am Heart Assoc*. 2018;7:e009476. DOI: 10.1161/JAHA.118.009476.)**

Key Words: Artificial intelligence • Machine learning • clinical decision support • cardiovascular risk • cardiovascular disease risk factors • cardiovascular disease prevention • atherosclerosis • prediction statistics • statin

Approximately every 20 seconds an American will have a heart attack or stroke. Of 790 000 heart attacks each year, 580 000 are new attacks in asymptomatic individuals. Similarly, of 610 000 strokes each year, 425 000 events are first-time. The economic cost of these unpredicted cardiovascular events is tens of billions of dollars annually.[1] Despite the grave nature of the problem, many of these events could be prevented if a more accurate tool for early detection of high-risk individuals became available.

The traditional method of cardiovascular disease (CVD) risk assessment is based on measuring traditional risk factors and predicting events over 10 years or a lifetime. Numerous studies have shown that current 10-year risk calculators, including the 2013 American College of Cardiology/American Heart Association (ACC/AHA) Pooled Cohort Equations Risk Calculator,[2] often overestimate cardiovascular events and, in women and certain ethnic groups, may underestimate risk.[3–7] The existing approach to CVD risk assessment desperately needs an overhaul. A consensus report from the Society for Heart Attack Prevention and Eradication (SHAPE) Task Force concluded that a comprehensive assessment of plaque, blood, and myocardial vulnerability factors is needed for an accurate prediction of CVD events. The task force further noted that despite major advances in the treatment of coronary heart

From the Computational Biomedicine Lab, University of Houston, TX (I.A.K., M.V.); Society for Heart Attack Prevention and Eradication, Palo Alto, CA (A.A.Y., M.N.); Research Unit Hypertension and Cardiovascular Epidemiology, KU Leuven Department of Cardiovascular Sciences, University of Leuven, Belgium (T.K.); Division of Cardiology, Los Angeles Biomedical Research at Harbor-UCLA Medical Center, Torrance, CA (M.B.).

Accompanying Datas S1 through S3, Tables S1 through S6, and Figures S1 through S4 are available at https://www.ahajournals.org/doi/suppl/10.1161/JAHA.118.009476

Correspondence to: Morteza Naghavi, MD, Society for Heart Attack Prevention and Eradication (SHAPE), Palo Alto, CA 94306. E-mail: mn@vp.org

Received April 17, 2018; accepted June 20, 2018.

Clinical Perspective

What Is New?

- The 2013 American College of Cardiology/American Heart Association (ACC/AHA) Pooled Cohort Equations risk calculator has been shown to be inaccurate in certain populations.
- Using the same risk variables, we developed a Machine Learning-based risk calculator in the MESA (Multi-Ethnic Study of Atherosclerosis) cohort and validated in the FLEMENGHO study (Flemish Study on Environment, Genes and Health Outcomes).

What Are the Clinical Implications?

- The Machine Learning Risk Calculator outperformed the ACC/AHA Risk Calculator by recommending less drug therapy, yet missing fewer cardiovascular disease events.
- These findings demonstrate the potential of Machine Learning to improve cardiovascular risk prediction and assist medical decision-making.

disease patients, a large number of victims of the disease who are apparently healthy die suddenly without prior symptoms.[8] Clearly the available screening and diagnostic methods are insufficient to identify the victims before the event occurs; therefore, short-term risk prediction is much needed. To reach the goal, a stepwise multi-phase approach is warranted that includes maximizing the long-term predictive value of traditional risk factors using Machine Learning (ML), gathering unique data on asymptomatic subjects who, shortly after an exam with blood testing, experience an ASCVD event, and applying ML to all available clinical data, including genomic. proteomic, and others to detect the vulnerable patient. In this paper, we report on our initial effort at advancing the field by using the same risk factors used by existing risk calculators but using ML as a new tool instead of the traditional statistical tools used for the existing risk calculator.[2] Rapid growth in information technology and computing power in recent years has spurred the emergence of ML and its applications in our day to day life from automated personal assistants to self-driving cars. The medical community has begun taking advantage of these new possibilities to improve medical care.

ML is generally categorized into 2 types, supervised and unsupervised.[9] Here, we report the use of a supervised ML model developed for predicting CVD risk and guiding the decision of whom should be recommended CVD preventive statin therapy. Unlike traditional statistical prediction methods, which operate with certain assumptions of linearity and force the predictive models to behave accordingly, the ML algorithm we used (Support Vector Machine—SVM) does not follow such assumptions and, instead, relies on learning the intrinsic properties or patterns of a given data set. The use of ML in medicine is clearly lagging other fields and remains experimental. Although several ML-based predictive models related to medical and cardiovascular fields have been published,[10–17] we are not aware of any approved by the FDA (Food and Drug Administration) for CVD prevention.

Our study serves as a step towards addressing this unmet need. Using ML and the same risk factors used by ACC/AHA Risk Calculator, we aimed to improve CVD risk stratification. We tested this approach in MESA (the Multi-Ethnic Study of Atherosclerosis)[18] and also used FLEMENGHO (the Flemish Study on Environment, Genes and Health Outcomes)[19] for external validation of our findings.

Methods

The data, analytic methods, and study materials will not be made available to other researchers for purposes of reproducing the results or replicating the procedure.

Study Participants

Initiated in July 2000 to investigate the prevalence, correlation, and progression of subclinical CVD in individuals who did not exhibit known cardiovascular issues, MESA is a US community-based prospective cohort study of 6814 white, black, Hispanic, or Chinese American men and women aged 45 to 84 years free of clinically apparent CVD at baseline (2000–2002). Study details have been previously published.[18] All participants gave their written informed consent and the institutional review boards at all participating centers approved the study. The MESA study group followed the cohort yearly for up to 13 years from baseline (median, 11.1 years) and monitored for incidence of cardiovascular events. Thus, the MESA data set used in this study includes baseline characteristics data and which participants have experienced CVD events during the follow-up period.

For this study, we excluded 69 (1.0%) subjects with missing risk factor data. Next, we excluded 286 (4.2%) who were >79 years at baseline; this step was performed because the ACC/AHA Risk Calculator was designed for those between 40 and 79 years of age. Thus, our study population was comprised of 6459 participants.

ACC/AHA Pooled Cohort Equations Risk Calculator

The 2013 ACC/AHA Pooled Cohort Equations calculator was designed to estimate 10-year risk of atherosclerotic cardiovascular disease, defined as heart attack, CHD death, or stroke. ACC/AHA risk estimates were computed using the subjects' baseline characteristics and available published

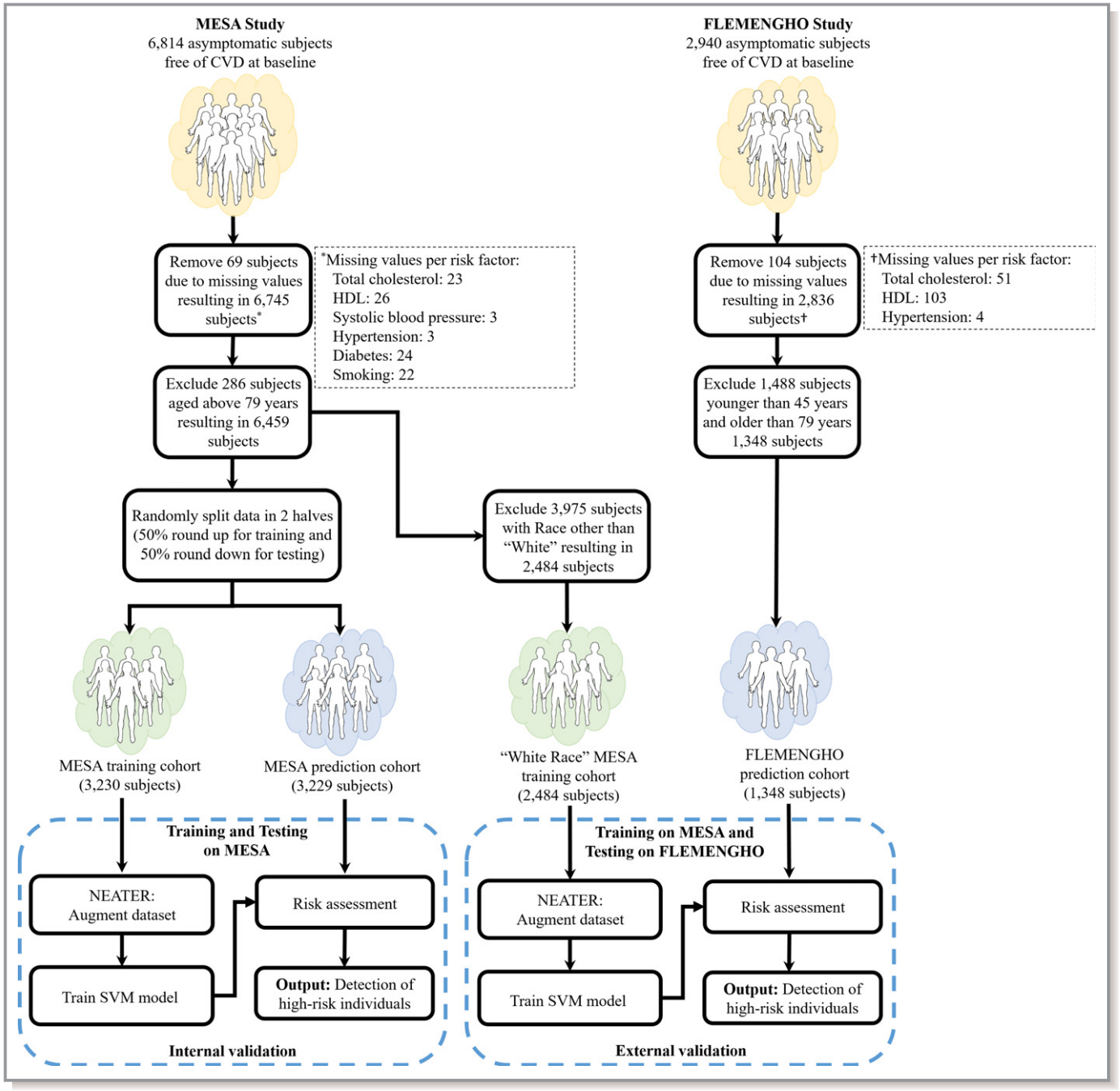


Figure 1. Overview of ML approach. For each ML model, we divided the study population 50/50 into training and prediction subset cohorts. Next, we augmented the training subset using NEATER and trained the SVM prediction model. During prediction, each sample in the prediction cohort was analyzed and classified. Then, the cohorts switched places (ie, prediction becomes training, and vice versa) and the process was repeated. The overall iterative process was repeated 10 times for each ML model, and the results were averaged. CVD indicates cardiovascular disease; FLEMENGHO study, the Flemish Study of Environment, Genes and Health Outcomes; HDL, high-density lipoprotein; MESA, the Multi-Ethnic Study of Atherosclerosis; NEATER, a method for the filtering of oversampled data using non-cooperative game theory; SVM, Support Vector Machine.

equations.[2] We grouped the scores into 2 preselected 10-year risk prediction categories: (1) low risk (<7.5%) and (2) high risk (≥7.5%) categories, for determining which subjects the risk calculator would have recommended statin treatment, based on the 2013 ACC/AHA guideline on the treatment of blood cholesterol to reduce atherosclerotic cardiovascular risk.[20] Because the follow-up for the MESA data is 13 years, we linearly transformed the 10-year risk of the base models into a 13-year risk. Thus, the risk threshold for statin eligibility becomes 9.75% for 13 years.

Machine Learning Risk Calculator

An overview of the ML approach is shown in Figure 1. The algorithm begins with the initial cohort, which is then split into training and prediction sets. Because of the imbalanced nature of data (many more samples without an event than samples with an event), during training, data are augmented using NEATER (a method for the filtering of oversampled data using non-cooperative game theory),[21] and then an SVM[22] classifier is trained. When a new unseen sample appears during prediction, the ML model determines in which category it belongs.

We built 8 ML-based models for "Hard CVD" events and 8 models for "All CVD" events. For each type of events we built 2 models per sex (ie, 1 for males and 1 for females) and for each one of them we built 4 models per ethnicity (eg, white, Chinese, black, and Hispanic). Thus, there were 16 models in total.

The SVM method is one of the most powerful learning algorithms for binary classification problems such as the problem stated in this article.[23–27] SVM is given a training set of examples (or inputs), belonging to 2 classes (eg, positive and negative, event and no event), with associated labels (or output values) and it finds the optimal maximum-margin dividing the 2 classes. This dividing interface is called a hyperplane and achieves maximum discrimination.

During training, the ML model was provided with the baseline values of the same 9 risk factors ("predictor variables," in ML parlance) used by the ACC/AHA Risk Calculator, namely the following: age, sex, ethnicity, total cholesterol, high-density lipoprotein cholesterol, systolic blood pressure, treatment for hypertension, history of diabetes mellitus, and smoking status. "Hard CVD" events included myocardial infarction, fatal CHD, stroke, and stroke death. "All CVD" events included the above list, plus congestive heart failure, transient ischemic attack, peripheral vascular disease, resuscitated cardiac arrest, percutaneous transluminal coronary angioplasties, probable angina, other revascularization, other CVD death, other atherosclerotic death, and cardiac bypass graft surgery.

The investigation of the performance was evaluated in real data samples that the model had not seen before. Moreover, 2-fold cross validation was used to validate the method and the results were averaged for all possible configurations. By using 2-fold cross validation, we ensure that we obtain an unbiased estimation of the model performance, while no significant loss of modeling or prediction capability may be apparent because of the imbalanced nature of the data. More details on the 2-fold cross validation that was performed can be found in the following sections.

For visualization purposes, we projected the high-dimensional feature space into a 3D feature space using the Principal Component Analysis. Because of the high dimensionality of the input training data, the decision hyperplane between the class samples with an event and the class samples without an event is transformed into a hyper-surface. An example of the 3D hyper-surface for the MESA male group for classifying "Hard CVD" and "All CVD" events can be seen in Figure S1.

Imbalanced Data

We cast the problem of event prediction as a binary classification problem. We denote the positive class "+1" (ie, subjects with an event) as the minority class (ie, category of samples with events), and the negative class "−1" (ie, subjects without an event) as the majority class (ie, category of samples without events). The MESA data are severely imbalanced in terms of outcomes, that is, the size of the minority class is much smaller than the size of the majority class, and, as a result, the decision boundary for ML methods would be severely biased and could result in poor performance. To cope with this skewed class distribution issue, we selected the NEATER algorithm,[21] a data augmentation algorithm that is based on filtering oversampled data using cooperative game theory. We elected to use this algorithm based on its ability to effectively increase the performance of the classifier as well as its unique tendency to avoid overfitting, which can be inevitable when using other oversampling techniques. NEATER is able to handle data sets of an imbalanced nature and generate new data. A detailed description of how NEATER works can be found in Data S1.

The main advantage of NEATER is that it makes no prior assumptions about the data, while it reaches high accuracy for both the minority and the majority classes. It is also important to note that NEATER is used for data augmentation only for training purposes and never during prediction.

Two-Fold Cross Validation

To ensure and increase the model's robustness and ability to generalize under unknown samples, we employed 2-fold cross validation to randomly split the original data set into 2 equally sized halves, a training set to train the model, and a test set to evaluate it. This type of cross validation has been widely used in the machine learning literature for predicting high-risk individuals (more details about the 2-fold cross validation can be found in Data S2).[28–31]

External Validation

To test the generalizability of the ML models, and also to check for potential overfitting, we tested the ML risk

calculator on an external data set drawn from the FLEMENGHO study.

FLEMENGHO recruited 2940 white participants between ages 20 and 90 years from August 1985 to December 2005 who were free of clinical CVD at baseline. FLEMENGHO studies a random population sample stratified by sex and age from a geographically defined area in Northern Belgium. All participants provided their written informed consent and local institutional review board approved the study protocol. FLEMENGHO study details have been previously published.[19]

For the FLEMENGHO study, we excluded 104 (3.5%) subjects with missing risk factor data. Then, 1488 (50.6%) subjects who were <45 years and >79 years at baseline were excluded; this step was performed because the ML Risk Calculator was trained in MESA cohort for those between 45 and 79 years of age. The final study population for the external validation was comprised of 1348 subjects. In FLEMENGHO, 621 (21.2%) subjects were followed-up for >13 years. Of them 180 (6.1%) subjects had a CVD event. We treated these samples as "no events" and included them in the study population as such; the reason for this step is that the ML Risk Calculator was trained on a 13-year follow up period while FLEMENGHO's follow-up was >13 years. Thus, to have a fair comparison, this step was necessary for our analysis.

Software

The code was implemented in MATLAB R2017a and C++. For the implementation of SVM, the LIBSVM library[32] was used (details can be found in Data S3).

Statistical Analysis

We performed an analysis to determine the sensitivity, specificity, accuracy, and C-statistic of the ACC/AHA Risk Calculator based on the prediction equations and the 7.5% 10-year risk threshold described previously. Next, we analyzed the performance of the ML Risk Calculator, compared its performance metrics to those of the ACC/AHA Risk Calculator, and calculated categorical net reclassification improvement (NRI) values for paired models. A paired t test was used to compare the population means and compute the *P* values. All statistical tests were 2- tailed, and $P<0.05$ was considered significant.

Results

Descriptive Statistics

Overall, 480 (7.4%) "Hard CVD" and 976 (15.1%) "All CVD" events occurred in the MESA study population (N=6459) during the 13-year follow up period. Baseline characteristics of the study population and subgroups of interest are reported in Table 1. "Hard CVD" events included 221 myocardial infarction, 71 CHD deaths, 178 strokes, and 10 stroke deaths. "All CVD" events included 221 myocardial infarction, 71 CHD deaths, 178 strokes, 10 stroke deaths, 176 angina-driven revascularizations, 11 resuscitated cardiac arrests, 8 other atherosclerotic deaths, 38 other CVD deaths, 111 congestive heart failures (CHF), 53 peripheral vascular diseases (PVD), 22 percutaneous transluminal coronary angioplasties, 9 coronary bypass grafts , 61 transient ischemic attacks, and 7 other revascularizations.

Risk Calculator Performance Comparison

Table 2 presents the sensitivity, specificity, and accuracy of the risk calculators for the prediction of "Hard CVD" and "All CVD" events, respectively. The ACC/AHA Risk Calculator achieved 0.76 sensitivity, 0.56 specificity, and 0.58 accuracy for predicting "Hard CVD" events and 0.75 sensitivity, 0.59 specificity, and 0.62 accuracy for predicting "All CVD" events. In comparison, the ML Risk Calculator for the prediction of "Hard CVD" events had higher sensitivity (0.86), specificity (0.95), and accuracy (0.94). For the prediction of "All CVD" events, ML Risk Calculator sensitivity was increased to 0.96 but with a slight decrease in specificity (0.87) and accuracy (0.89).

The number of false negatives (ie, subjects who are classified by the Risk Calculator as "low risk" but do experience a CVD event) and false positives (ie, subjects who are classified as "high risk" but do not experience a CVD event), and the categorical net reclassification improvement (NRI) between each base ACC/AHA model and its corresponding ML model using the same risk factors are also shown in Table 2. For the ML Risk Calculator, an NRI improvement of 0.49 for "Hard CVD" events and 0.50 for "All CVD" events, when compared with the ACC/AHA Risk Calculator, for all subjects, was achieved.

To investigate the potential impact of removing the statin users, we performed a sensitivity analysis after excluding statin users from the data set and found similar results for ML Risk Calculator performance. For example, for "Hard CVD" events, the ML Risk Calculator had AUC of 0.92 and NRI of 0.46 when statin users were excluded, as compared with AUC of 0.92 and NRI of 0.49 when statin users were included. The baseline characteristics of the study population and subgroups of interest, when statin users were excluded from the analysis, and the performance metrics of the risk calculators are reported in Tables S1 and S2, respectively.

Figure 2 depicts the receiver operating characteristic curves of the different models. It can be observed that all ML models attain high discrimination ability between events and no events

Table 1. Baseline Characteristics of Study Population and Subgroups of Interest

	All (N=6459)	Hard CVD (n=480)	All CVD (n=976)	ACC/AHA <9.75% 13-y risk (n=3487)	ACC/AHA ≥9.75% 13-y risk (n=2972)	ML: Low Risk (13-y) (n=5724)	ML: High Risk (13-y) (n=735)
Age, y	61.3±9.6	65.8±8.9	65.7±8.7	55.3±6.9	68.4±7.1	60.6±9.6	66.4±8.2
Male, n%	3060 (47.4%)	282 (58.7%)	590 (60.4%)	1254 (36.0%)	1806 (60.8%)	2601 (45.4%)	459 (62.4%)
Female, n%	3399 (52.6%)	198 (41.3%)	386 (39.6%)	2233 (64.0%)	1166 (39.2%)	3123 (54.6%)	276 (37.6%)
Ethnicity, n%							
White	2484 (38.5%)	187 (39.0%)	413 (42.3%)	1439 (41.2%)	1045 (35.2%)	2197 (38.4%)	287 (39.0%)
Asian	767 (11.9%)	35 (7.3%)	67 (6.9%)	446 (12.8%)	321 (10.8%)	697 (12.2%)	70 (9.5%)
Black	1780 (27.5%)	138 (28.7%)	282 (28.9%)	794 (22.8%)	986 (33.2%)	1573 (27.5%)	207 (28.2%)
Hispanic	1428 (22.1%)	120 (25.0%)	214 (21.9%)	808 (23.2%)	620 (20.8%)	1257 (21.9%)	171 (23.3%)
Total cholesterol, mg/dL	194.4±35.8	194.6±34.2	193.2±37.3	194.9±34.9	193.8±36.8	194.7±36.6	192.3±30.9
High-density lipoprotein cholesterol, mg/dL	50.9±14.8	47.8±13.9	48.1±13.6	52.6±15.0	48.8±14.3	51.5±15.0	46.5±12.3
Systolic blood pressure, mm Hg	125.9±21.1	136.3±22.2	134.5±21.7	116.9±16.4	136.6±21.1	124.7±20.9	135.8±20.0
Hypertension, n%	2351 (36.4%)	243 (50.6%)	510 (52.2%)	735 (21.1%)	1616 (54.4%)	1974 (34.5%)	377 (51.3%)
Diabetes mellitus, n%	729 (11.3%)	107 (22.3%)	217 (22.2%)	127 (3.6%)	602 (20.3%)	653 (11.4%)	76 (10.3%)
Smoking, n%							
Current smoking	869 (13.5%)	92 (19.2%)	169 (17.3%)	387 (11.1%)	482 (16.2%)	762 (13.3%)	107 (14.6%)
Prior smoking	2365 (36.6%)	180 (37.5%)	419 (42.9%)	1192 (34.2%)	1173 (39.5%)	2073 (36.2%)	292 (39.7%)
Never	3225 (49.9%)	208 (43.3%)	388 (39.8%)	1908 (54.7%)	1317 (44.3%)	2889 (50.5%)	336 (45.7%)
Family history heart attack, n%*	2593 (40.1%)	239 (49.8%)	482 (49.4%)	1364 (39.1%)	1229 (41.3%)	2231 (39.0%)	362 (49.2%)
Coronary artery calcification, Agatston*	138.8±408.6	316.1±577.2	389.9±759.0	41.0±160.8	253.4±555.2	121.0±380.1	274.0±563.7
hsCRP, mg/L*,†	3.8±5.8 1.92 (1.86–1.99)	4.3±5.9 2.23 (1.98–2.47)	4.5±6.8 2.33 (2.16–2.50)	3.6±5.2 1.80 (1.71–1.90)	4.0±6.4 2.07 (1.98–2.16)	3.8±5.8 1.92 (1.85–1.99)	3.7±5.6 1.97 (1.78–2.16)

Continuous variables are expressed as mean±SD. Categorical variables are presented as absolute numbers and frequencies.

*The American College of Cardiology/American Heart Association (ACC/AHA) Risk Calculator does not use these variables; therefore, they were not included in the Machine Learning CVD predictive models.

†High sensitivity C-reactive protein (hsCRP) is also expressed as a geometric mean with 90% confidence interval since this variable is not normally distributed.

with respect to the base models. In particular, the ML Risk Calculator, for "Hard CVD" events, achieved an average AUC of 0.92 and for "All CVD" events, the AUC was 0.94. The ACC/AHA Risk Calculator for "Hard CVD" events achieved an average AUC of 0.71, and for "All CVD" events, the AUC was 0.72. The corresponding receiver operating characteristic curves, when statin users were excluded from the analysis, are depicted in Figure S2.

Statin Eligibility and Missed Treatment Opportunities

Figure 3 shows the performance characteristics of the risk calculators for addressing 2 clinically relevant issues, ie, determining statin eligibility and avoiding missed treatment opportunities. Almost half of the study population (46.0%) were determined by the ACC/AHA calculator to be statin eligible. In contrast, the ML calculator deemed only 11.4% to be at high risk and statin eligible. For "All CVD" events, the ML calculator determined 25.1% to be statin eligible.

Regarding missed treatment opportunities (false negatives), the ACC/AHA calculator also performed poorly, as 23.8% of "Hard CVD" events occurred in individuals that ACC/AHA calculator would not have recommended statin. The ML Risk Calculator fared better, with only 14.4% of "Hard CVD" events and only 4.4% of "All CVD" events occurring in individuals; the ML calculator would not have recommended statin. The breakdown of the missed "Hard CVD" and "All CVD" events comparing the ML Risk calculator with the ACC/AHA Risk Calculator is shown in Figure S3.

Table 2. Risk Calculator Comparison: Sensitivity-Specificity-Other Performance Metrics

Event	Model	Sn (95% CI)	*P* Value	Sp (95% CI)	*P* Value	FN	FP	TP	TN	Acc (95% CI)	*P* Value	NRI (95% CI)	*P* Value
Male													
Hard CVD	ACC/AHA Risk Calculator	0.86±0.1 (0.81–0.90)	–	0.44±0.1 (0.42–0.46)	–	40	1564	242	1214	0.48±0.1 (0.46–0.49)	–	–	–
	ML Risk Calculator	0.90±0.1 (0.86–0.94)	≤0.001	0.93±0.1 (0.92–0.94)	≤0.001	27	204	255	2574	0.92±0.1 (0.91–0.93)	≤0.001	0.53 (0.51–0.55)	≤0.001
All CVD	ACC/AHA Risk Calculator	0.84±0.1 (0.81–0.87)	–	0.47±0.1 (0.45–0.49)	–	96	1312	494	1158	0.54±0.1 (0.52–0.56)	–	–	–
	ML Risk Calculator	0.97±0.1 (0.96–0.99)	≤0.001	0.82±0.1 (0.80–0.84)	≤0.001	15	443	575	2027	0.85±0.1 (0.84–0.86)	≤0.001	0.48 (0.46–0.50)	≤0.001
Female													
Hard CVD	ACC/AHA Risk Calculator	0.63±0.1 (0.56–0.69)	–	0.67±0.1 (0.66–0.69)	–	74	1042	124	2159	0.67±0.1 (0.66–0.69)	–	–	–
	ML Risk Calculator	0.79±0.1 (0.72–0.84)	≤0.001	0.96±0.1 (0.95–0.97)	≤0.001	42	120	156	3081	0.95±0.1 (0.94–0.96)	≤0.001	0.45 (0.43–0.47)	≤0.001
All CVD	ACC/AHA Risk Calculator	0.62±0.1 (0.57–0.67)	–	0.69±0.1 (0.68–0.71)	–	146	926	240	2087	0.68±0.1 (0.67–0.70)	–	–	–
	ML Risk Calculator	0.93±0.1 (0.90–0.95)	≤0.001	0.92±0.1 (0.91–0.93)	≤0.001	28	247	358	2766	0.92±0.1 (0.91–0.93)	≤0.001	0.54 (0.52–0.55)	≤0.001
All													
Hard CVD	ACC/AHA Risk Calculator	0.76±0.1 (0.72–0.80)	–	0.56±0.1 (0.55–0.58)	–	114	2606	366	3373	0.58±0.1 (0.57–0.59)	–	–	–
	ML Risk Calculator	0.86±0.1 (0.82–0.89)	≤0.001	0.95±0.1 (0.94–0.96)	≤0.001	69	324	411	5655	0.94±0.1 (0.93–0.95)	≤0.001	0.49 (0.48–0.50)	≤0.001
All CVD	ACC/AHA Risk Calculator	0.75±0.1 (0.72–0.78)	–	0.59±0.1 (0.56–0.61)	–	242	2238	734	3245	0.62±0.1 (0.60–0.63)	–	–	–
	ML Risk Calculator	0.96±0.1 (0.94–0.97)	≤0.001	0.87±0.1 (0.86–0.88)	≤0.001	43	690	933	4793	0.89±0.1 (0.88–0.89)	≤0.001	0.50 (0.48–0.51)	≤0.001

ACC/AHA indicates American College of Cardiology/American Heart Association; CI, confidence interval; CVD, cardiovascular disease; FN, false negatives; FP, false positives; ML, Machine Learning; NRI, net reclassification improvement; Sn, sensitivity; Sp, specificity; TN, true negatives; TP, true positives.

External Validation

The baseline characteristics of the FLEMENGHO external validation study population are reported in Table S3.

Table 3 provides the sensitivity, specificity, and accuracy of the risk calculators using the MESA data set for training purposes only and the FLEMENGHO as an external validation set. The ML Risk calculator tested on "White Race" FLEMENGHO data set achieved a sensitivity of 0.74, specificity of 0.87, and accuracy of 0.84, much higher than ACC/AHA Risk Calculator (sensitivity 0.63, specificity 0.69, and accuracy 0.68). Also, for comparison purposes, we used the "White Race" sub-cohort of MESA for both training and testing. In this setup, the ML Risk Calculator tested on the "White Race" MESA cohort had a sensitivity of 0.84, specificity of 0.96, and accuracy of

0.95, while the ACC/AHA Risk Calculator achieved 0.73 sensitivity, 0.59 specificity, and 0.60 accuracy. The NRI improvement of ML Risk Calculator tested on FLEMENGHO data set over the ACC/AHA Risk Calculator was 0.29 and for the ML Risk Calculator tested on the "White Race" MESA cohort the NRI was 0.48, respectively. Also, Tables S4 and S5 show the performance metrics of the ML Risk Calculator trained and tested on FLEMENGHO cohort and trained on "White Race" FLEMENGHO cohort and tested on "White Race" MESA Cohort, respectively.

Figure S4 illustrates the discrimination properties of the ML Risk Calculator compared with the ACC/AHA Risk Calculator for the "White Race" MESA and the FLEMENGHO cohorts. The ML Risk Calculator tested on FLEMENGHO data set model achieved an AUC of 0.81 and the AUC of the ACC/AHA Risk Calculator was 0.70. For the testing on the "White Race" MESA cohort, the ML Risk Calculator trained on the "White Race" MESA cohort achieved an average AUC of 0.91 and the ACC/AHA Risk Calculator achieved an AUC of 0.71, indicating that the ML models can more accurately classify those with and without event.

Discussion

In this report, we present a new ML-based risk calculator which uses the same 9 traditional risk factors (ie, age, sex, ethnicity, total cholesterol, high-density lipoprotein cholesterol, systolic blood pressure, treatment for hypertension, diabetes mellitus, and smoking) used by the ACC/AHA Risk Calculator. Despite using identical input, our ML Risk Calculator attained a significantly higher accuracy than the ACC/AHA Risk Calculator. It detected 13% more high-risk individuals and recommended 25% less unnecessary statin therapy in low-risk individuals. Unlike the ACC/AHA Risk Calculator, which is only designed for predicting "Hard CVD" events, our ML Risk Calculator performed well for predicting both Hard and All CVD events. Furthermore, the ML Risk Calculator performed well both for males and females, with NRI improvement values of 0.53 and 0.48 for males, and 0.45 and 0.54 for females, for "Hard CVD" and "All CVD" events, respectively.

To prevent methodological biases and overfitting, we applied a 2-fold internal cross validation and subsequently tested the model an independent external data set, FLEMENGHO. Moreover, to address the inherent problem of class imbalance, which is a common problem in many cohort-based studies, we used the NEATER algorithm. Detailed technical analyses of our validation methodologies and the treatment of class imbalance problem can be found in Data S1. Additionally, characteristics of the synthetic data generated by NEATER for the "Male White Race" MESA subgroup can be seen in Table S6.

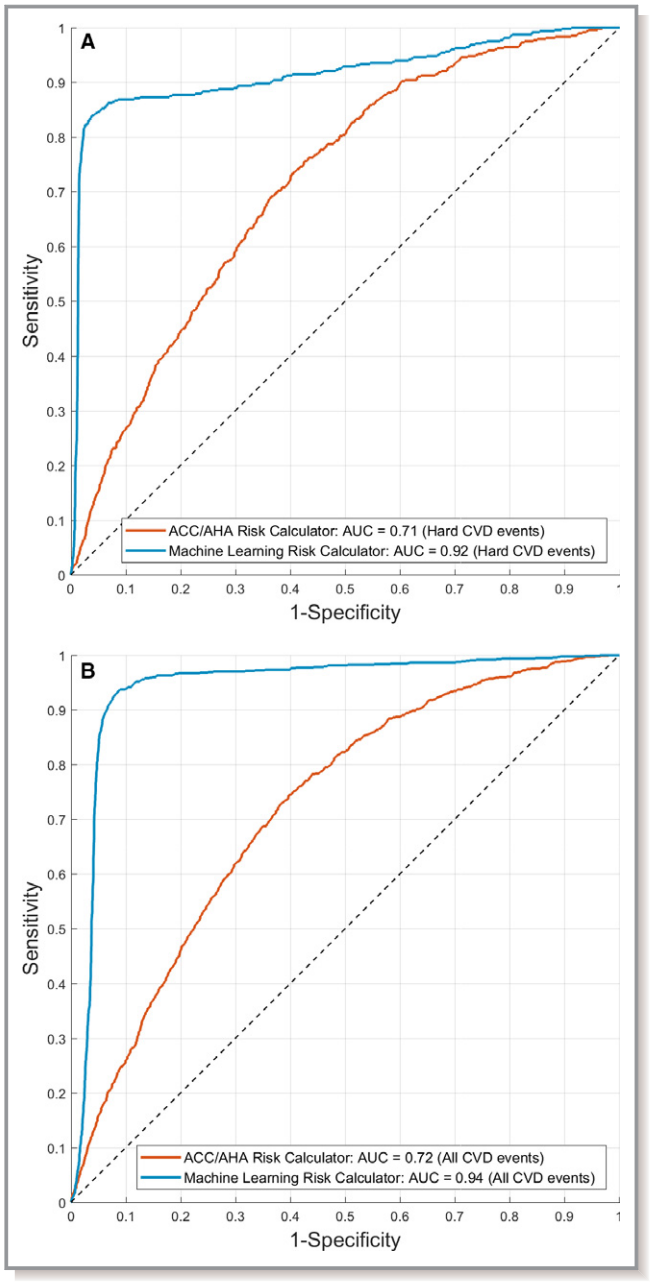


Figure 2. Receiver operating characteristic (ROC) curves for prediction of (A) "Hard CVD" events and (B) "All CVD" events comparing the ML Risk Calculator (blue) with the American College of Cardiology/American Heart Association (ACC/AHA) Risk Calculator (red). AUC indicates area under the curve; CVD, cardiovascular disease; ML, Machine Learning.

We trained our ML model with and without statin users. The results of the sensitivity analysis with and without statin users were not significantly different. Refer to the Table S1 for details.

We attribute the superior performance of our ML model to its flexibility and non-linear function. ML maps the data into a multidimensional space where various separating planes are evaluated and ultimately a "hyperplane" is found. Additionally,

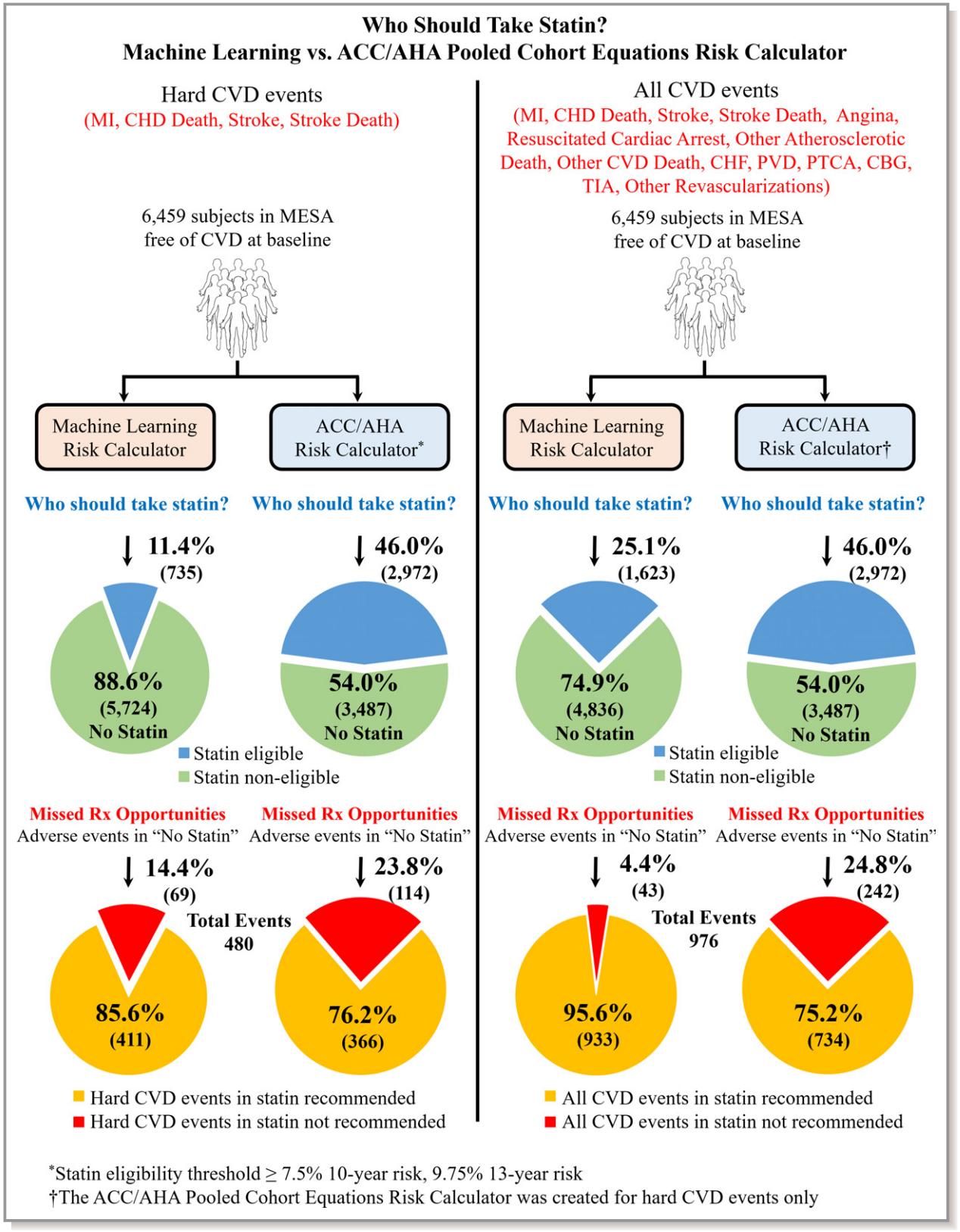


Figure 3. Risk calculator comparison: statin eligibility and missed treatment opportunities. Pie graphs illustrate the performance comparison between ML Risk Calculator and ACC/AHA Risk Calculator for predicting "Hard CVD" events (left) and "All CVD" events (right). ACC/AHA indicates American College of Cardiology/American Heart Association; CBG, coronary bypass grafts; CHD, coronary heart disease; CHF, congestive heart failures; CVD, cardiovascular disease; MI, myocardial infarction; PTCA, percutaneous transluminal coronary angioplasties; PVD, peripheral vascular diseases; TIA, transient ischemic attacks.

DOI: 10.1161/JAHA.118.009476

Table 3. External Validation: Risk Calculator Comparison between Models Trained and Tested on "White Race" MESA Cohort and Models Trained on "White Race" MESA Cohort and Tested on "White Race" FLEMENGHO Cohort

Cohort	Model	Sn (95% CI)	*P* Value	Sp (95% CI)	*P* Value	FN	FP	TP	TN	Acc (95% CI)	*P* Value	NRI (95% CI)	*P* Value
Male													
Train and test on White Race MESA	ACC/AHA Risk Calculator	0.85±0.1 (0.77–0.91)	–	0.45±0.1 (0.42–0.48)	–	16	602	91	488	0.48±0.1 (0.46–0.51)	–	–	–
	ML Risk Calculator	0.90±0.1 (0.82–0.95)	≤0.001	0.98±0.1 (0.97–0.99)	≤0.001	11	22	96	1068	0.97±0.1 (0.96–0.98)	0.002	0.58 (0.55–0.60)	≤0.001
Train on White Race MESA and test on FLEMENGHO	ACC/AHA Risk Calculator	0.74±0.1 (0.66–0.80)	–	0.55±0.1 (0.5–0.59)	–	41	234	114	283	0.59±0.1 (0.55–0.63)	–	–	–
	ML Risk Calculator	0.86±0.1 (0.80–0.91)	≤0.001	0.82±0.1 (0.79–0.85)	≤0.001	21	92	134	425	0.83±0.1 (0.80–0.86)	0.001	0.39 (0.36–0.43)	0.004
Female													
Train and test on White Race MESA	ACC/AHA Risk Calculator	0.58±0.1 (0.49–0.68)	–	0.72±0.1 (0.70–0.75)	–	34	336	46	871	0.71±0.1 (0.69–0.74)	–	–	–
	ML Risk Calculator	0.77±0.1 (0.65–0.85)	≤0.001	0.94±0.1 (0.93–0.96)	≤0.001	19	68	61	1139	0.93±0.1 (0.92–0.95)	0.002	0.41 (0.40–0.46)	≤0.001
Train on White Race MESA and test on FLEMENGHO	ACC/AHA Risk Calculator	0.48±0.1 (0.39–0.57)	–	0.82±0.1 (0.78–0.85)	–	57	103	53	463	0.76±0.1 (0.73–0.79)	–	–	–
	ML Risk Calculator	0.56±0.1 (0.47–0.66)	≤0.001	0.91±0.1 (0.88–0.93)	≤0.001	48	52	62	514	0.85±0.1 (0.82–0.88)	0.002	0.17 (0.14–0.20)	0.004
All													
Train and test on White Race MESA	ACC/AHA Risk Calculator	0.73±0.1 (0.66–0.79)	–	0.59±0.1 (0.57–0.61)	–	50	938	137	1359	0.60±0.1 (0.58–0.62)	–	–	–
	ML Risk Calculator	0.84±0.1 (0.78–0.89)	≤0.001	0.96±0.1 (0.95–0.97)	≤0.001	30	90	157	2207	0.95±0.1 (0.94–0.96)	0.001	0.48 (0.46–0.50)	≤0.001
Train on White Race MESA and test on FLEMENGHO	ACC/AHA Risk Calculator	0.63±0.1 (0.57–0.69)	–	0.69±0.1 (0.66–0.72)	–	98	337	167	746	0.68±0.1 (0.65–0.70)	–	–	–
	ML Risk Calculator	0.74±0.1 (0.68–0.79)	≤0.001	0.87±0.1 (0.85–0.89)	≤0.001	69	144	196	939	0.84±0.1 (0.82–0.86)	0.001	0.29 (0.27–0.31)	0.004

ACC/AHA, American College of Cardiology/American Heart Association; CI, confidence interval; FLEMENGHO, Flemish Study of Environment, Genes and Health Outcomes; FN, false negatives; FP, false positives; MESA, Multi-Ethnic Study of Atherosclerosis; NRI, net reclassification improvement; Sn, sensitivity; Sp, specificity; TN, true negatives; TP, true positives.

the ability to train the ML model with artificially created events using data augmentation techniques such as NEATER can further empower ML over the traditional statistical methods. Our 2-fold cross validation technique assured the independence of testing samples from the training samples. Overall, ML-based prediction models are more versatile and capable than statistical models. As our ML model is exposed to more longitudinal data, including those from which ACC/AHA Risk Calculator was derived, we anticipate a more robust risk calculator. We also plan to introduce new predictor variables such as coronary calcium score and other biomarkers to our model that are expected to further improve its predictive power. Although 10-year risk is the status quo for risk prediction, the ability to predict events in a shorter term (eg, 1-year) is highly desired. Such a short-term risk predictor can open doors for new prophylactic therapies. This development is the focus of the SHAPE initiative titled "Machine Learning Vulnerable Patient: Developing an Artificial Intelligence-based Forecast System for Prediction of Heart Attacks within 12 Months".

Study Strengths and Limitations

A major strength of our study is that we created our ML model based on a robust 13-year follow up data set from MESA, which ranks as the best multiethnic study of atherosclerosis in the world. Unlike data in national registries, population surveys, or other healthcare management databases, MESA meets the highest standards of research quality data, which is key for developing reliable machine learning models. Another strength of our study is that we used human expertise with advanced knowledge of the field to supervise and fine-tune the machine learning models. Yet another strength of our study is the use of oversampling techniques to maximize ML training. Finally, we validated our ML model both internally (2-fold cross validation) and externally (testing on FLEMENGHO cohort).

As for limitations, although the MESA cohort is comprised of a large population of different ethnicities, it suffers from a low event rate in subgroups, which in turn limits the predictive power and generalizability of our ML Risk Calculator. Because of such low number of events, our ML models may not be reliable in other populations or other countries. Although the external validation results were promising, FLEMENGHO is not a multiethnic cohort, therefore our ML model needs to be validated in other multiethnic data sets. Without validating our model across a large number of US and international cohorts we are unable to claim an equal performance as we have seen in MESA and FLEMENGHO. Another limitation is that MESA's age range is 45 to 84 years, which limits the applicability of our ML Risk Calculator for prediction of events in individuals who fall outside this age range. Moreover, the impact of other risk factors or biomarkers in the prediction of the cardiovascular events was not considered in this study. Furthermore, we did not refit the ACC/AHA Risk Calculator to the MESA data set only, which could likely have increased its accuracy in MESA. We decided not to refit because we aimed to compare the ML model with the exact same model as recommended by ACC/AHA Pooled Cohort Equation.

Finally, we would like to reiterate the major limitation of our ML method is that it was created and validated based on 2 data sets only (MESA and FLEMENGHO), while the ACC/AHA Risk Calculator was derived from several data sets. To this end, further studies are underway to validate these findings in other large multiethnic and multinational cohorts.

Summary

In conclusion, we developed a new ML Risk Calculator based on MESA, a multiethnic, community-based cohort of men and women studied for incident atherosclerotic cardiovascular disease. We used the same variables used by the ACC/AHA Risk Calculator yet achieved a much higher predictive accuracy. Further studies are underway to validate this new ML Risk Calculator in other cohorts. As we introduce more data to our ML Risk Calculator, particularly to cases in which events occurred weeks or months following data collection instead of years or decades, the "holy grail" of short-term CVD risk prediction may be within reach.

Acknowledgments

This study was conducted as part of an international initiative led by the Society for Heart Attack Prevention and Eradication (SHAPE) to advance CVD risk assessment. The authors wish to acknowledge the scientific advice of SHAPE's Scientific Advisory Board. Authors would like to thank Ahmed Gul for his administrative assistance with references and formatting. The authors thank the other investigators, the staff, and the participants of the MESA study for their valuable contributions. A full list of participating MESA investigators and institutions can be found at http://www.mesa-nhlbi.org.

Sources of Funding

Dr Kakadiaris and Dr Vrigkas's work has been funded in part by the UH Hugh Roy and Lillie Cranz Cullen Endowment Fund. MESA study was supported by contracts N01-HC-95159, N01-HC-95160, N01-HC-95161, N01-HC-95162, N01-HC-95163, N01-HC-95164, N01-HC-95165, N01-HC-95166, N01-HC-95167, N01-HC-95168 and N01-HC-95169 from the National Heart, Lung, and Blood Institute, and by grants UL1-TR-000040, UL1 TR 001079, and UL1-RR-025005 from National Center for Research Resources.

Disclosures

None.

References

1. Benjamin EJ, Blaha MJ, Chiuve SE, Cushman M, Das SR, Deo R, de Ferranti SD, Floyd J, Fornage M, Gillespie C, Isasi CR, Jiménez MC, Jordan LC, Judd SE, Lackland D, Lichtman JH, Lisabeth L, Liu S, Longenecker CT, Mackey RH, Matsushita K, Mozaffarian D, Mussolino ME, Nasir K, Neumar RW, Palaniappan L, Pandey DK, Thiagarajan RR, Reeves MJ, Ritchey M, Rodriguez CJ, Roth GA, Rosamond WD, Sasson C, Towfighi A, Tsao CW, Turner MB, Virani SS, Voeks JH, Willey JZ, Wilkins JT, Wu JH, Alger HM, Wong SS, Muntner P; American Heart Association Statistics Committee and Stroke Statistics Subcommittee. Heart disease and stroke statistics-2017 update: a report from the American Heart Association. *Circulation*. 2017;135:e146–e603.
2. Goff DC, Lloyd-Jones DM, Bennett G, Coady S, D'Agostino RB, Gibbons R, Greenland P, Lackland DT, Levy D, O'Donnell CJ, Robinson JG, Schwartz JS, Shero ST, Smith SC, Sorlie P, Stone NJ, Wilson PW, Jordan HS, Nevo L, Wnek J, Anderson JL, Halperin JL, Albert NM, Bozkurt B, Brindis RG, Curtis LH, DeMets D, Hochman JS, Kovacs RJ, Ohman EM, Pressler SJ, Sellke FW, Shen WK, Tomaselli GF; American College of Cardiology/American Heart Association Task Force on Practice Guidelines. 2013 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. *Circulation*. 2014;129:S49–S73.
3. Ridker PM, Cook NR. Statins: new American guidelines for prevention of cardiovascular disease. *Lancet*. 2013;382:1762–1765.
4. Muntner P, Colantonio LD, Cushman M, Goff DC, Howard G, Howard VJ, Kissela B, Levitan EB, Lloyd-Jones DM, Safford MM. Validation of the atherosclerotic cardiovascular disease pooled cohort risk equations. *JAMA*. 2014;311:1406–1415.
5. Kavousi M, Leening MJ, Nanchen D, Greenland P, Graham IM, Steyerberg EW, Ikram MA, Stricker BH, Hofman A, Franco OH. Comparison of application of the ACC/AHA guidelines, Adult Treatment Panel III guidelines, and European Society of Cardiology guidelines for cardiovascular disease prevention in a European Cohort. *JAMA*. 2014;311:1416–1423.
6. DeFilippis AP, Young R, Carrubba CJ, McEvoy JW, Budoff MJ, Blumenthal RS, Kronmal RA, McClelland RL, Nasir K, Blaha MJ. An analysis of calibration and discrimination among multiple cardiovascular risk scores in a modern multiethnic cohort. *Ann Intern Med*. 2015;162:266–275.
7. DeFilippis AP, Young R, McEvoy JW, Michos ED, Sandfort V, Kronmal RA, McClelland RL, Blaha MJ. Risk score overestimation: the impact of individual cardiovascular risk factors and preventive therapies on the performance of the American Heart Association-American College of Cardiology-Atherosclerotic Cardiovascular Disease risk score in a modern multi-ethnic cohort. *Eur Heart J*. 2017;38:598–608.
8. Naghavi M, Falk E, Hecht HS, Jamieson MJ, Kaul S, Berman D, Fayad Z, Budoff MJ, Rumberger J, Naqvi TZ, Shaw LJ, Faergeman O, Cohn J, Bahr R, Koenig W, Demirovic J, Arking D, Herrera VLM, Badimon J, Goldstein JA, Rudy Y, Airaksinen J, Schwartz RS, Riley WA, Mendes RA, Douglas P, Shah PK. From vulnerable plaque to vulnerable patient—Part III: executive summary of the screening for heart attack prevention and education (SHAPE) task force report. *Am J Cardiol*. 2006;98:2–15.
9. Sajda P. Machine learning for detection and diagnosis of disease. *Annu Rev Biomed Eng*. 2006;8:537–565.
10. Parthiban L, Subramanian R. Intelligent heart disease prediction system using CANFIS and genetic algorithm. *Int J Biol Life Sci*. 2007;3:157–160.
11. Kourou K, Exarchos TP, Exarchos KP, Karamouzis MV, Fotiadis DI. Machine learning applications in cancer prognosis and prediction. *Comput Struct Biotechnol J*. 2015;13:8–17.
12. Vidyasagar M. Identifying predictive features in drug response using machine learning: opportunities and challenges. *Annu Rev Pharmacol Toxicol*. 2015;55:15–34.
13. Vock DM, Wolfson J, Bandyopadhyay S, Adomavicius G, Johnson PE, Vazquez-Benitez G, O'Connor PJ. Adapting machine learning techniques to censored time-to-event health record data: a general-purpose approach using inverse probability of censoring weighting. *J Biomed Inform*. 2016;61:119–131.
14. Araki T, Ikeda N, Shukla D, Jain PK, Londhe ND, Shrivastava VK, Banchhor SK, Saba L, Nicolaides A, Shafique S, Laird JR, Suri JS. PCA-based polling strategy in machine learning framework for coronary artery disease risk assessment in intravascular ultrasound: a link between carotid and coronary grayscale plaque morphology. *Comput Methods Programs Biomed*. 2016;128:137–158.
15. Deo RC. Machine learning in medicine. *Circulation*. 2015;132:1920–1930.
16. Motwani M, Dey D, Berman DS, Germano G, Achenbach S, Al-Mallah MH, Andreini D, Budoff MJ, Cademartiri F, Callister TQ, Chang HJ, Chinnaiyan K, Chow BJ, Cury RC, Delago A, Gomez M, Gransar H, Hadamitzky M, Hausleiter J, Hindoyan N, Feuchtner G, Kaufmann PA, Kim YJ, Leipsic J, Lin FY, Maffei E, Marques H, Pontone G, Raff G, Rubinshtein R, Shaw LJ, Stehli J, Villines TC, Dunning A, Min JK, Slomka PJ. Machine learning for prediction of all-cause mortality in patients with suspected coronary artery disease: a 5-year multicentre prospective registry analysis. *Eur Heart J*. 2017;38:500–507.
17. Ambale-Venkatesh B, Yang X, Wu CO, Liu K, Hundley WG, McClelland RL, Gomes AS, Folsom AR, Shea S, Guallar E, Bluemke DA, Lima JA. Cardiovascular event prediction by machine learning: the multi-ethnic study of atherosclerosis. *Circ Res*. 2017;121:1092–1101.
18. Bild DE, Bluemke DA, Burke GL, Detrano R, Diez Roux AV, Folsom AR, Greenland P, Jacob DR, Kronmal R, Liu K, Nelson JC, O'Leary D, Saad MF, Shea S, Szklo M, Tracy RP. Multi-ethnic study of atherosclerosis: objectives and design. *Am J Epidemiol*. 2002;156:871–881.
19. Stolarz-Skrzypek K, Kuznetsova T, Thijs L, Tikhonoff V, Seidlerová J, Richart T, Jin Y, Olszanecka A, Malyutina S, Casiglia E, Filipovský J, Kawecka-Jaszcz K, Nikitin Y, Staessen JA. Fatal and nonfatal outcomes, incidence of hypertension, and blood pressure changes in relation to urinary sodium excretion. *JAMA*. 2011;305:1777–1785.
20. Stone NJ, Robinson JG, Lichtenstein AH, Bairey Merz CN, Blum CB, Eckel RH, Goldberg AC, Gordon D, Levy D, Lloyd-Jones DM, McBride P, Schwartz JS, Shero ST, Smith SC, Watson K, Wilson PWF. 2013 ACC/AHA guideline on the treatment of blood cholesterol to reduce atherosclerotic cardiovascular risk in adults: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines. *J Am Coll Cardiol*. 2014;63:2889–2934.
21. Almogahed BA, Kakadiaris IA. Neater: filtering of over-sampled data using non-cooperative game theory. *Soft Comput*. 2015;19:3301–3322.
22. Cristianini N, Shawe-Taylor J. *An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods*. Cambridge, United Kingdom: Cambridge University Press; 2000.
23. Jabeen A, Ahmad N, Raza K. Machine learning-based state-of-the-art methods for the classification of rna-seq data. In: Dey N, Ashour AS, Borra S, eds. *Classification in BioApps: Automation of Decision Making*. Cham: Springer International Publishing; 2018:133–172.
24. Melacci S, Belkin M. Laplacian support vector machines trained in the primal. *J Mach Learn Res*. 2011;12:1149–1184.
25. Baesens B, Van Gestel T, Viaene S, Stepanova M, Suykens J, Vanthienen J. Benchmarking state-of-the-art classification algorithms for credit scoring. *J Oper Res Soc*. 2003;54:627–635.
26. Xiangying W, Yixin Z. Statistical learning theory and state of the art in SVM. Proceedings of IEEE International Conference on Cognitive Informatics. 2003:55–59.
27. Boser BE, Guyon IM, Vapnik VN. A training algorithm for optimal margin classifiers. Proceedings of 5th Annual Workshop on Computational Learning Theory. 1992:144–152.
28. Weng SF, Reps J, Kai J, Garibaldi JM, Qureshi N. Can machine-learning improve cardiovascular risk prediction using routine clinical data? *PLoS One*. 2017;12:e0174944.
29. Isler Y, Narin A, Ozer M. Comparison of the effects of cross-validation methods on determining performances of classifiers used in diagnosing congestive heart failure. *Measurement Science Review*. 2015;15:196–201.
30. Wiens AD, Inan OT. Accelerometer body sensor network improves systolic time interval assessment with wearable ballistocardiography. Proceedings of Annual International Conference of the IEEE Engineering in Medicine and Biology Society. 2015:1833–1836.
31. Alcaraz R, Martínez A, Rieta JJ. Role of the P-wave high frequency energy and duration as noninvasive cardiovascular predictors of paroxysmal atrial fibrillation. *Comput Methods Programs Biomed*. 2015;119:110–119.
32. Chang C-C, Lin C-J. LIBSVM: a library for support vector machines. *ACM Trans Intell Syst Technol*. 2011;2:1–27.

Supplemental Material

DATA S1: IMBALANCED DATA: NEATER

The MESA data are severely imbalanced in terms of outcomes, that is, the size of the class with events (i.e., minority class) is much smaller than the size of the class without events (i.e., majority class), and thus the decision boundary for ML methods would be severely biased and could result in poor performance. To cope with this skewed class distribution issue, we selected the filtering of over-sampled data using non-cooperative game theory (NEATER) algorithm[2], which is an oversampling data augmentation algorithm that employs cooperative game theory to generate artificial data of the minority class. Non-cooperative game theory[3] addresses the interaction between individual rational decision makers, where all the data are players and the goal is to uniformly and consistently label all of the synthetic data created by any oversampling technique. Unlike other over-sampling approaches, NEATER does not automatically consider synthetic data as part of the minority class. Instead, it keeps synthetic samples unlabeled, at first. These samples then participate in a non-cooperative game to determine their most likely class membership, minority or majority. All the synthetic data that end up belonging to the minority class are kept, and the rest are eliminated.

A detailed description of the main steps of the NEATER implementation in this work can be summarized as follows. First, an oversampled method is used, such as the Synthetic Minority Over-Sampling Technique (SMOTE)[4] to generate synthetic data. The use of SMOTE is justified by the fact that it creates samples that are closely related to the minority class, which causes the classifier to create larger decision regions. Then, both the original and synthetically generated data are considered as players, and the possible class memberships are considered strategies available to all game players. Note that, only the synthetic data play to determine their class membership. There are two types of players: I_c, which denotes players that already belong to a

class, and I_u, which denotes unlabeled players or synthetic samples. Each I_u player interacts with a number of its neighbors I_ϕ, one neighbor at a time. Also, each player can choose among two available strategies $S_i = \{m, M\}$ with a probability of 0.5, where m stands for minority and M for majority. A mixed strategy x_i (i.e., combination of strategies from which one is randomly chosen with specified probability) for player i is the probability distribution over his set of strategies S_i. Then, for each player i, its k, where $k = 5$, nearest neighbors are computed and for each player interacting with each of its k neighbors, the utility functions are computed as follows:

$$u_i(x) = \sum_{j \in I_\phi \cap I_u} (x_i^T A_{ij} x_j) + \sum_{d=1}^{2} \sum_{j \in I_\phi \cap I_{c|d}} (x_i^T A_{ij} e_j^d) ,$$

where $d = 1$ is playing the minority class and $d = 2$ is the majority class, $e_j^d \in S_i$ is an extreme mixed strategy with $e_j^1 = (1,0)$ and $e_j^2 = (0,1)$, and A_{ij} is the partial payoff matrix between two players i and j. The set $I_{c|d}$ is the set of players who always play their d^{th} strategy. After that, the average payoff in the whole population is computed:

$$u(x) = x_i^T A_{ij} x_j .$$

Then, iteratively, discrete-time replicator dynamic is applied to study the evolution of the minority strategy probability:

$$x_i^m(t+1) = \frac{\alpha + u_i(e_i^m)}{\alpha + u_i\big(x(t)\big)} x_i^m,$$

if a maximum number of iterations is reached, the process stops, otherwise, t is increased by one and the average payoff for the next player is computed. Finally, for each player in I_u, the class membership with the highest probability is assigned.

An example of the number of the synthetic data of the minority class generated by NEATER and their characteristics for the "Male White Race" MESA subgroup can be seen in **TABLE S6**.

DATA S2: TWO-FOLD CROSS VALIDATION

To ensure and increase the model's robustness and ability to generalize under unknown samples, we employed two-fold cross validation to randomly split the original dataset into two equally sized halves, a training set to train the model, and a test set to evaluate it. This type of cross validation has been widely used in the machine learning literature for predicting high-risk individuals.[5-8] To ensure that the random split of the dataset will always result in having positive and negative examples in both training and testing sets, we employed the following procedure. First, we randomly shuffle the sub-cohort of samples with CVD events into two parts (50% of the positive samples for training and the remaining 50% for testing). Then, the remaining sub-cohort of negative samples is also randomly split into two halves, and the corresponding training and testing subsets are fused so that each of them will contain positive and negative examples. The training and testing sets are independent and do not overlap with each other. At this point, we train our model on subset A and evaluate on subset B, and next we reverse the order (i.e., train on subset B and evaluate on subset A). This process is repeated 10 times, so that statistical reliability of the evaluation process may be ensured[9-11], with each of the different subsets used exactly once as the validation data, and the results are averaged over all the examined configurations. Note that at each iteration the training and evaluation processes start from scratch so that there is no memory of any the previous learned model, and thus biased results are avoided. One of the main reasons for using two-fold cross validation is that the MESA data are extremely imbalanced and there is not enough data of the positive class; furthermore, by repeating the random split multiple times, we are able to train on more positive examples. A fair way to evaluate the model is to split the dataset into two halves and train on as many positive examples as possible, since it is a powerful general technique, when the data are sparse.

DATA S3: SUPPORT VECTOR MACHINE

Support Vector Machine (SVM)[1] is a discriminative classifier, which is designed for supervised learning. The learning model is given a training set of examples (or inputs), belonging to two classes, with associated class labels (or output values). The examples are in form of attribute vectors and the SVM finds the optimal maximum-margin hyper-plane, which separates the input data. Although there exist multiple hyper-planes that offer a solution to the problem, a hyper-plane may be a bad solution if it lies too close to the points, as it is noise-sensitive and may not generalize well. Thus, SVMs aim at finding the hyper-plane that gives the largest minimum distance to the training examples. In other words, given a set of N training examples that consists of pairs of feature vectors x_i with $i = 1, \dots, N$, that denote the pattern to be classified, along with their corresponding class labels y_i, where $x \in \mathbb{R}^d$, with d being the number of features for each sample (i.e., age, sex, ethnicity, total cholesterol, HDL cholesterol, systolic blood pressure, hypertension, diabetes, and smoking status) and $y \in \{-1, +1\}$, where label "-1" corresponds to subjects without an event and label "+1" corresponds to subjects with an event.

The problem is defined as constructing the decision function that correctly classifies an input pattern that is not the training set. The SVM determines the decision hyper-plane between the two classes, the positive class y_1 (i.e., subjects with an event) and the negative class y_2 (i.e., subjects without an event), which is obtained by the solution of the following optimization problem:

$$\underset{w,b,\xi}{\text{minimize}} \left\{ \frac{1}{2} \left|\left|w\right|\right|^2 + C \sum_{i=1}^{N} \xi_i \right\},$$

$$\text{subject to: } y_i(w^T \phi(x_i) + b) \geq 1 - \xi_i, \qquad \xi_i \geq 0, \qquad i = 1, \dots, N$$

where w is a is a normal vector perpendicular to the hyper-plane, $\left|\left|w\right|\right|^2$ indicates the size of the

margin, C is a positive constant that reflects the influence of margin errors, b determines the offset of the hyper-plane from the origin along the normal vector w, and ξ_i are the slack variables, which measure the degree of misclassification of the datum x_i. In our implementation, the kernel "trick" is used with a function $\phi(x_i)$ that maps the data into a higher dimensional space, where various separating planes would be evaluated and ultimately a hyper-plane can be found.

The minimization process is a problem of Lagrangian optimization that can be solved by transforming to the dual form and using Lagrange multipliers to obtain the weight vector w and the bias b of the optimal hyper-plane as follows:

$$\underset{a}{\text{minimize}}\ R(a) = \frac{1}{2}\sum_{i,j=1}^{N} a_i a_j y_i y_j K(x_i, x_j) - \sum_{i=1}^{N} a_i\,,$$

$$\text{subject to: } \sum_{i-1}^{N} y_i a_i = 0, \qquad 0 \le a_i \le C$$

For each testing sample, the kernel matrix K between each of the training samples and the respective testing sample is computed. Thus, the decision function $f(x)$ is given by:

$$f(x) = \ \text{sgn}\left(\sum_{i=1}^{N} y_i a_i K(x_i, x) + b\right),$$

where the terms a_i, with $i = 1, \ldots, N$ constitute a dual representation for the weight vector w in terms of the training set, such as:

$$w = \sum_{i=1}^{N} a_i y_i x_i\,.$$

Moreover, in our experiments, we used as kernel function the radial basis function (RFB) kernel, which is defined as:

$$K(x_i, x_j) = \exp\left(-\gamma \left|\left|x_i - x_j\right|\right|^2\right), \qquad \gamma > 0$$

To estimate the value of the training parameters, for each of the 16 ML-based models (i.e., eight ML-based models for "Hard CVD" events and eight models for "All CVD" events), we used two-fold cross-validation by setting the values of parameter C to 2^k, with $k \in \{-5, \ldots, 15\}$ and the values of the kernel coefficient γ were set to 2^k, with $k \in \{-15, \ldots, 3\}$.

For visualization purposes, we projected the high-dimensional feature space into a 3D feature space using the Principal Component Analysis (PCA). Because of the high dimensionality of the input training data, the decision hyper-plane between the class samples with an event and the class samples without an event is transformed into a hyper-surface. An example of the 3D hyper-surface for the MESA male group for classifying "Hard CVD" and "All CVD" events can be seen in the **FIGURE S1**.

TABLE S1. MESA Cohort Baseline Characteristics of Study Population and Subgroups of Interest. Continuous variables are expressed as mean ± standard deviation. Categorical variables are presented as absolute numbers and frequencies. *The ACC/AHA Risk Calculator does not use these variables; therefore, they were not included in the Machine Learning CVD predictive models. †High sensitivity C-reactive protein (hsCRP) is also expressed as a geometric mean with 90% confidence interval since this variable is not normally distributed.

	Non-Statin Users							**Statin Users**
	All (N = 5,415)	**Hard CVD (N = 381)**	**All CVD (N = 775)**	**ACC/AHA < 9.75% 13yr risk (N = 3,092)**	**ACC/AHA ≥ 9.75% 13yr risk (N = 2,323)**	**ML: Low Risk (13yr) (N = 4,844)**	**ML: High Risk (13yr) (N = 571)**	**Lipid Lowering Medication (N = 1,044)**
Age, y	60.6 ± 9.7	65.5 ± 9.2	65.5 ± 9.0	54.9 ± 6.9	68.2 ± 7.2	59.9 ± 9.6	66.0 ± 8.6	65.0 ± 8.3
Male, n%	2,563 (47.3%)	222 (58.3%)	477 (61.6%)	1,119 (36.2%)	1,445 (62.2%)	2,204 (45.5%)	359 (62.9%)	497 (47.6%)
Female, n%	2,852 (52.7%)	159 (41.7%)	298 (38.5%)	1,973 (63.8%)	878 (37.8%)	2,640 (54.5%)	212 (37.1%)	547 (52.4%)
Ethnicity, n%								
White	2,028 (37.5%)	145 (38.0%)	322 (41.5%)	1,224 (39.6%)	804 (34.6%)	1,806 (37.3%)	222 (38.9%)	456 (43.7%)
Asian	663 (12.2%)	27 (7.1%)	52 (6.7%)	405 (13.1%)	258 (11.1%)	602 (12.4%)	61 (10.7%)	104 (10.0%)
African American	1,484 (27.4%)	107 (28.1%)	223 (28.8%)	717 (23.2%)	767 (33.0%)	1,334 (27.5%)	150 (26.2%)	296 (28.3%)
Hispanic	1,240 (22.9%)	102 (26.8%)	178 (23.0%)	746 (24.1%)	494 (21.3%)	1,102 (22.8%)	138 (24.2%)	188 (18.0%)
Total Cholesterol, mg/dL	196.6 ± 35.5	197.6 ± 33.8	195.9 ± 36.1	196.2 ± 34.7	197.1 ± 36.5	196.7 ± 35.9	195.8 ± 31.3	182.9 ± 35.3

HDL Cholesterol, mg/dL	51.0 ± 14.9	47.7 ± 14.0	47.8 ± 13.7	52.7 ± 15.0	48.7 ± 14.5	51.5 ± 15.1	46.8 ± 12.9	50.3 ± 13.9
Systolic Blood Pressure, mm Hg	125.3 ± 21.0	135.7 ± 22.2	134.5 ± 21.9	116.7 ± 16.5	136.8 ± 21.0	124.2 ± 20.8	134.9 ± 20.3	129.2 ± 21.5
Hypertension, n%	1,724 (31.8%)	173 (45.4%)	364 (47.0%)	578 (18.7%)	1,146 (49.3%)	1,468 (30.3%)	256 (44.8%)	627 (60.1%)
Diabetes, n%	505 (9.3%)	69 (8.1%)	147 (19.0%)	99 (3.2%)	406 (17.5%)	451 (9.3%)	54 (9.5%)	224 (21.5%)
Smoking, n%								
Current Smoking	765 (14.1%)	79 (20.7%)	145 (18.7%)	352 (11.4%)	413 (17.8%)	663 (13.7%)	102 (17.9%)	104 (10.0%)
Prior Smoking	1,938 (35.8%)	134 (35.2%)	314 (40.5%)	1,036 (33.5%)	902 (38.8%)	1,724 (35.6%)	214 (37.4%)	427 (40.9%)
Never	2,712 (50.1%)	168 (44.1%)	316 (40.8%)	1,704 (55.1%)	1,008 (43.4%)	2,457 (50.7%)	255 (44.7%)	513 (49.1%)
*Family History Heart Attack, n%	2,082 (38.5%)	184 (48.3%)	370 (47.7%)	1,158 (37.5%)	923 (39.7%)	1,830 (37.8%)	252 (44.1%)	511 (48.9%)
*Coronary Artery Calcification, Agatston	118.1 ± 370.0	284.8 ± 557.3	355.4 ± 713.2	36.3 ± 155.7	227 ±515.9	103.6 ± 344.6	242.0 ± 524.7	246.0 ±556.3
*†hsCRP, mg/L	3.9 ± 6.0 1.96 (1.88 - 2.03)	4.4 ± 6.3 2.29 (2.01 - 2.56)	4.6 ± 7.0 2.37 (2.17 - 2 .56)	3.7 ± 5.4 1.82 (1.72 - 1.92)	4.2 ± 6.6 2.16 (2.05 - 2.26)	3.9 ± 5.9 1.98 (1.90 - 2.05)	3.6 ± 6.0 1.79 (1.57 - 2.02)	3.3 ± 5.0 1.76 (1.61 - 1.92)

TABLE S2. Risk Calculator Comparison, when Excluding Statin Users from the Analysis: Sensitivity-Specificity-Other Performance Metrics.

Event	Model	Sn (95% CI)	p-value	Sp (95% CI)	p-value	FN	FP	TP	TN	Acc (95% CI)	p-value	NRI (95% CI)	p-value
Male													
Hard CVD	ACC/AHA Risk Calculator	0.84 ± 0.1 (0.78 - 0.88)	--	0.46 ± 0.1 (0.44 - 0.48)	--	36	1,259	186	1,082	0.50 ± 0.1 (0.48 - 0.51)	--	--	--
	ML Risk Calculator	0.89 ± 0.1 (0.84 0.93)	≤0.001	0.93 ± 0.1 (0.92 - 0.94)	≤0.001	24	161	198	2,180	0.93 ± 0.1 (0.92 - 0.94)	≤0.001	0.52 (0.50 - 0.54)	≤0.001
All CVD	ACC/AHA Risk Calculator	0.77 ± 0.1 (0.72 - 0.80)	--	0.53 ± 0.1 (0.50 - 0.55)	--	112	988	365	1,098	0.57 ± 0.1 (0.55 - 0.59)	--	--	--
	ML Risk Calculator	0.97 ± 0.1 (0.95 - 0.99)	≤0.001	0.83 ± 0.1 (0.81 - 0.84)	≤0.001	13	358	464	1,728	0.86 ± 0.1 (0.84 - 0.87)	≤0.001	0.50 (0.48 - 0.52)	≤0.001
Female													
Hard CVD	ACC/AHA Risk Calculator	0.61 ± 0.1 (0.53 - 0.67)	--	0.71 ± 0.1 (0.69 - 0.73)	--	62	781	97	1,912	0.70 ± 0.1 (0.69 - 0.72)	--	--	--
	ML Risk Calculator	0.79 ± 0.1 (0.72 - 0.85)	≤0.001	0.97 ± 0.1 (0.96 - 0.98)	≤0.001	33	86	126	2,607	0.96 ± 0.1 (0.95 - 0.97)	≤0.001	0.44 (0.42 - 0.46)	≤0.001
All CVD	ACC/AHA Risk Calculator	0.54 ± 0.1 (0.48 - 0.60)	--	0.76 ± 0.1 (0.74 - 0.78)	--	137	608	161	1,946	0.74 ± 0.1 (0.72 - 0.75)	--	--	--
	ML Risk Calculator	0.92 ± 0.1 (0.88 - 0.94)	≤0.001	0.92 ± 0.1 (0.90 - 0.93)	≤0.001	25	217	273	2,337	0.92 ± 0.1 (0.90 - 0.93)	≤0.001	0.54 (0.52 - 0.56)	≤0.001
All													
Hard CVD	ACC/AHA Risk Calculator	0.74 ± 0.1 (0.70 - 0.79)	--	0.60 ± 0.1 (0.58 - 0.61)	--	98	2,040	283	2,994	0.60 ± 0.1 (0.59 - 0.62)	--	--	--
	ML Risk Calculator	0.85 ± 0.1 (0.81 - 0.88)	≤0.001	0.95 ± 0.1 (0.94 - 0.96)	≤0.001	57	247	324	4,787	0.94 ± 0.1 (0.93- 0.95)	≤0.001	0.46 (0.45 - 0.47)	≤0.001
All CVD	ACC/AHA Risk Calculator	0.73 ± 0.1 (0.70 - 0.76)	--	0.62 ± 0.1 (0.61 - 0.64)	--	204	1,752	571	2,888	0.64 ± 0.1 (0.63 - 0.65)	--	--	--

	ML Risk Calculator	0.95 ± 0.1 (0.93 - 0.97)	≤0.001	0.88 ± 0.1 (0.86 - 0.89)	≤0.001	38	575	737	4,065	0.89 ± 0.1 (0.88 - 0.90)	≤0.001	0.48 (0.47 - 0.49)	≤0.001

TABLE S3. FLEMENGHO Cohort Baseline Characteristics of Study Population and Subgroups of Interest Including the Statin Users in the Study Population. Continuous variables are expressed as mean ± standard deviation. Categorical variables are presented as absolute numbers and frequencies.

	All (N = 1,348)	**Hard CVD (N = 265)**	**ACC/AHA < 9.75% 13yr risk (N = 844)**	**ACC/AHA ≥ 9.75% 13yr risk (N = 504)**	**ML: Low Risk (13yr) (N = 1,008)**	**ML: High Risk (13yr) (N = 340)**
Male, n%	672 (49.9%)	155 (58.5%)	324 (38.4%)	348 (69.1%)	446 (44.2%)	226 (66.5%)
Female, n%	676 (50.1%)	110 (41.5%)	520 (61.6%)	156 (30.9%)	562 (55.8%)	114 (33.5%)
Age, y	56.9 ± 9.5	61.3 ± 9.4	52.1 ± 6.5	65.0 ± 8.1	55.3 ± 9.1	61.6 ±8.8
Total Cholesterol, mg/dL	232.4 ± 46.8	238.5 ± 48.4	227.2 ± 41.3	237.6 ± 45.3	230.4 ± 45.9	238.4 ± 49.1
HDL Cholesterol, mg/dL	54.5 ± 16.9	51.9 ± 17.2	58.0 ± 15.6	48.2 ± 15.0	56.4 ± 17.1	48.9 ± 15.1
Systolic Blood Pressure, mm Hg	132.2 ± 18.0	137.5 ± 20.2	126.4 ± 14.7	141.8 ± 18.8	129.8 ± 16.4	139.2 ± 20.6
Hypertension, n%	305 (22.6%)	78 (29.4%)	116 (13.7%)	189 (37.5%)	201 (19.9%)	104 (30.6%)
Diabetes, n%	56 (4.2%)	16 (6.0%)	16 (1.9%)	40 (7.9%)	38 (3.8%)	18 (5.3%)
Smoking, n%						
Current Smoking	357 (26.5%)	90 (34.0%)	168 (19.9%)	189 (37.5%)	268 (26.6%)	89 (26.2%)
Prior Smoking	440 (32.6%)	80 (31.2%)	285 (33.8%)	155 (30.8%)	319 (31.6%)	121 (35.6%)
Never	551 (40.9%)	95 (35.8%)	391 (46.3%)	160 (31.7%)	421 (41.8%)	130 (38.2%)

TABLE S4. Risk Calculator Comparison between Models Trained and Tested on FLEMENGHO Cohort: Sensitivity-Specificity-Other Performance Metrics.

Model	Sn (95% CI)	p-value	Sp (95% CI)	p-value	FN	FP	TP	TN	Acc (95% CI)	p-value	NRI (95% CI)	p-value
Male												
ACC/AHA Risk Calculator	0.74 ± 0.1 (0.66 - 0.80)	--	0.55 ± 0.1 (0.50 - 0.59)	--	41	234	114	283	0.59 ± 0.1 (0.55 - 0.63)	--	--	--
ML Risk Calculator	0.85 ± 0.1 (0.79 - 0.90)	≤0.001	0.99 ± 0.1 (0.98 - 1.00)	≤0.001	23	5	132	512	0.96 ± 0.1 (0.94 - 0.97)	≤0.001	0.55 (0.51 - 0.59)	≤0.001
Female												
ACC/AHA Risk Calculator	0.48 ± 0.1 (0.39 - 0.58)	--	0.82 ± 0.1 (0.78 - 0.85)	--	57	103	53	463	0.76 ± 0.1 (0.73 - 0.79)	--	--	--
ML Risk Calculator	0.71 ± 0.1 (0.61 - 0.79)	≤0.001	0.97 ± 0.1 (0.95 - 0.98)	≤0.001	32	16	78	550	0.93 ± 0.1 (0.91 - 0.95)	≤0.001	0.38 (0.34 - 0.41)	≤0.001
All												
ACC/AHA Risk Calculator	0.63 ± 0.1 (0.57 - 0.69)	--	0.69 ± 0.1 (0.66 - 0.72)	--	98	337	167	746	0.68 ± 0.1 (0.65 - 0.70)	--	--	--
ML Risk Calculator	0.79 ± 0.1 (0.74 - 0.84)	≤0.001	0.98 ± 0.1 (0.97 - 0.99)	≤0.001	55	21	210	1,062	0.94 ± 0.1 (0.93 - 0.95)	≤0.001	0.45 (0.42 - 0.48)	≤0.001

TABLE S5. Risk Calculator Comparison between Models Trained on "White Race" FLEMENGHO Cohort and Tested on "White Race" MESA Cohort.

Model	Sn (95% CI)	p-value	Sp (95% CI)	p-value	FN	FP	TP	TN	Acc (95% CI)	p-value	NRI (95% CI)	p-value
Male												
ACC/AHA Risk Calculator	0.85 ± 0.1 (0.77 - 0.91)	--	0.45 ± 0.1 (0.42 - 0.48)	--	16	602	91	488	0.48 ± 0.1 (0.46 - 0.51)	--	--	--
ML Risk Calculator	0.86 ± 0.1 (0.78 - 0.92)	≤0.001	0.78 ± 0.1 (0.76 - 0.81)	≤0.001	15	236	92	854	0.79 ± 0.1 (0.77 - 0.81)	≤0.001	0.34 (0.31 - 0.37)	≤0.001
Female												
ACC/AHA Risk Calculator	0.58 ± 0.1 (0.49 - 0.68)	--	0.72 ± 0.1 (0.70 - 0.75)	--	34	336	46	871	0.71 ± 0.1 (0.69 - 0.74)	--	--	--
ML Risk Calculator	0.78 ± 0.1 (0.67 - 0.86)	≤0.001	0.79 ± 0.1 (0.77 - 0.81)	≤0.001	18	253	62	954	0.79 ± 0.1 (0.77 - 0.81)	≤0.001	0.27 (0.25 - 0.30)	≤0.001
All												
ACC/AHA Risk Calculator	0.73 ± 0.1 (0.66 - 0.79)	--	0.59 ± 0.1 (0.57 - 0.61)	--	50	938	137	1,359	0.60 ± 0.1 (0.58 - 0.62)	--	--	--
ML Risk Calculator	0.82 ± 0.1 (0.76 - 0.87)	≤0.001	0.79 ± 0.1 (0.77 - 0.80)	≤0.001	33	489	154	1,808	0.79 ± 0.1 (0.77 - 0.81)	≤0.001	0.29 (0.27 - 0.31)	≤0.001

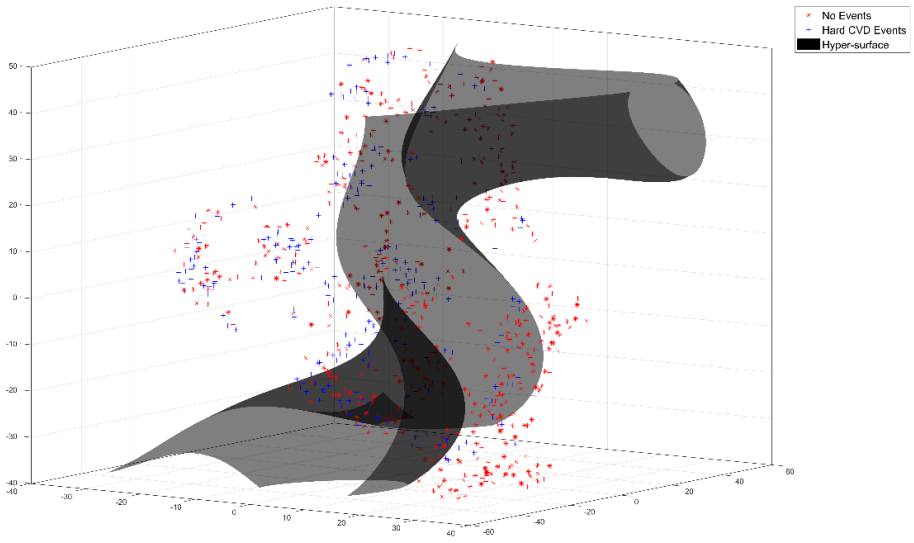
TABLE S6. Characteristics of Synthetic Data Generated by NEATER for "Male White Race" MESA Cohort and Subgroups of Interest. Continuous variables are expressed as mean ± standard deviation. Categorical variables are presented as absolute numbers and frequencies.

	Synthetic Data Generated by NEATER (N = 824)	**Synthetic Data Kept by NEATER (N = 467)**	**p-value***	**Synthetic Data Discarded by NEATER (N = 357)**	**p-value†**	**Majority Data (N = 1,090)**
Age, y	65.5 ± 8.1	66.2 ± 7.7	0.015	64.7 ± 8.5	0.019	61.6 ± 9.6
Total Cholesterol, n%	191.6 ± 30.9	192.6 ± 29.5	0.378	190.5 ± 32.8	0.369	189.4 ± 34.9
HDL, mg/dL	42.1 ± 10.7	41.3 ± 10.9	0.012	43.1 ± 10.2	0.010	45.4 ± 12.0
SBP, mg/dL	132.8 ± 19.2	135.7 ± 19.9	0.002	129.9 ± 17.7	0.001	122.8 ± 17.9
Hypertension, n%	333 (40.4%)	190 (40.7%)	0.486	143 (40.1%)	0.464	347 (31.8%)
Diabetes, n%	159 (19.3%)	86 (18.4%)	0.923	73 (20.4%)	0.855	62 (5.7%)
Smoking, n%			0.834		0.738	
Current Smoking	120 (14.6%)	68 (14.6%)		52 (14.6%)		117 (10.7%)
Prior Smoking	416 (50.5%)	239 (51.1%)		177 (49.6%)		536 (49.2%)
Never Smoking	288 (34.9%)	160 (34.3%)		128 (35.8%)		437 (40.1%)

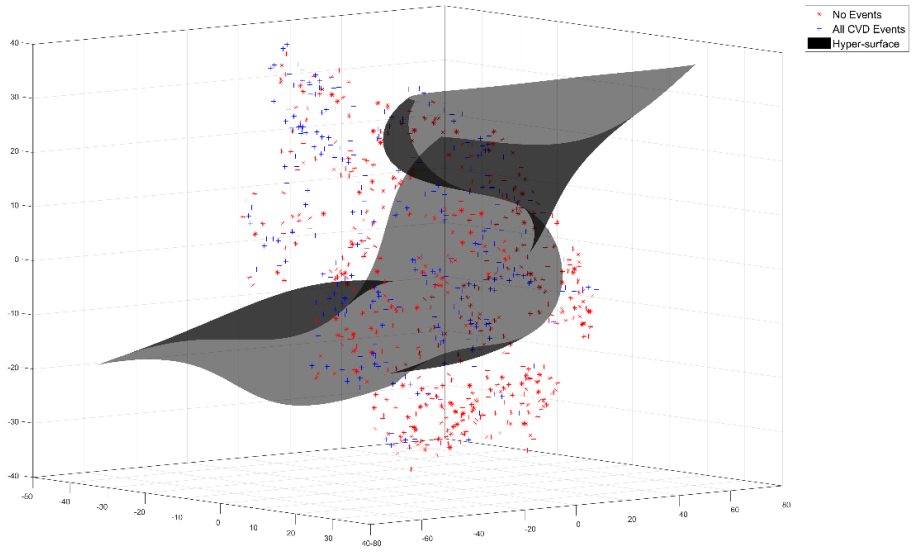
* Interaction between all synthetic data and synthetic data kept by NEATER using multivariate ANOVA

† Interaction between all synthetic data and synthetic data discarded by NEATER using multivariate ANOVA

FIGURE S1. SVM separating hyper-surface for male-group in MESA cohort for classifying (a) Hard CVD events and (b) All CVD events.

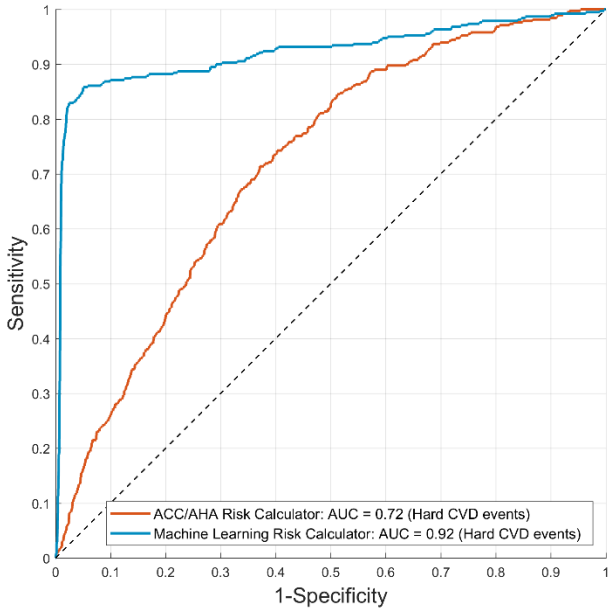


(a)

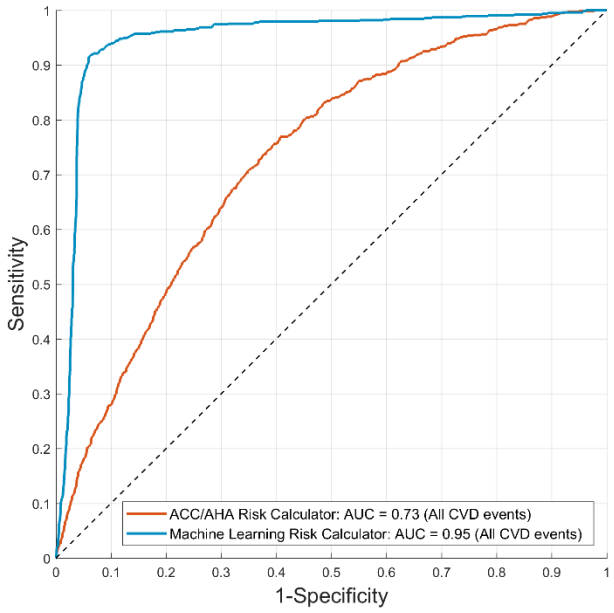


(b)

FIGURE S2. ROC curves for prediction of (a) Hard CVD events and (b) All CVD events, excluding the statin users, comparing the ML Risk Calculator (blue) with the ACC/AHA Risk Calculator (red). AUC: Area under the curve.

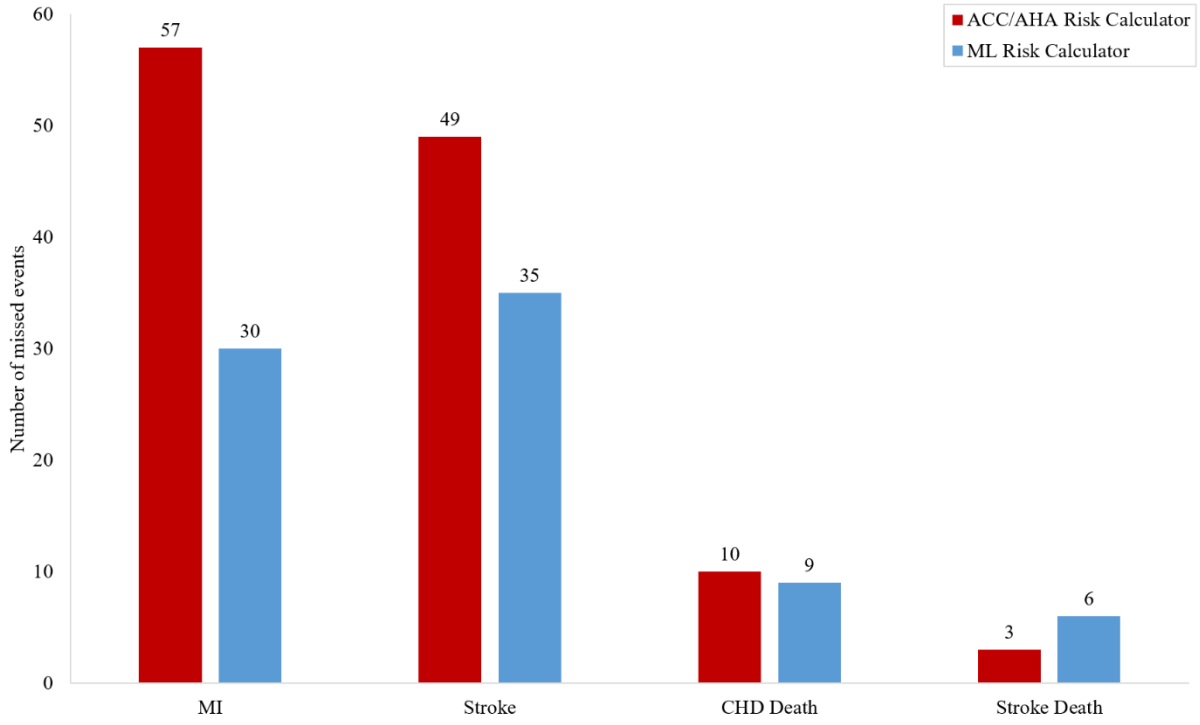


(a)

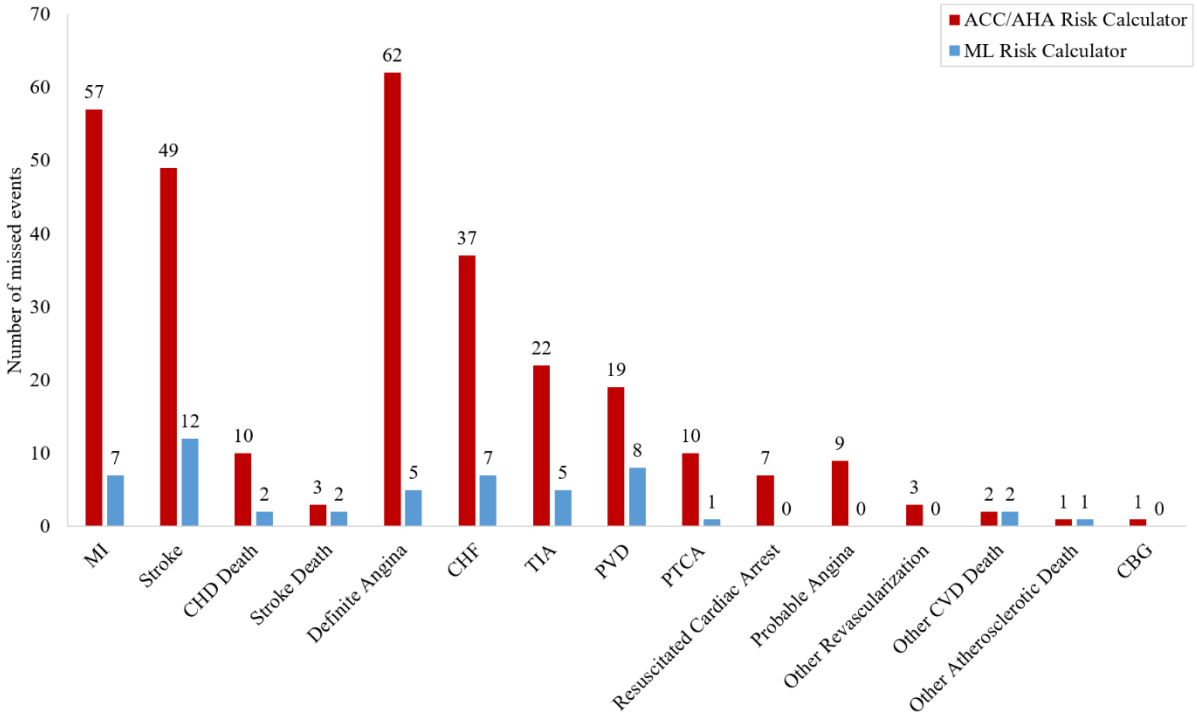

Sensitivity
1-Specificity
ACC/AHA Risk Calculator: AUC = 0.73 (All CVD events)
Machine Learning Risk Calculator: AUC = 0.95 (All CVD events)

(b)

FIGURE S3. Breakdown of the missed (a) Hard CVD events and (b) All CVD events comparing the ML Risk Calculator (blue) with the ACC/AHA Risk Calculator (red). MI: myocardial infarction; CHD: coronary heart disease; CVD: cardiovascular disease; CHF: congestive heart failures; PVD: peripheral vascular diseases; PTCA: percutaneous transluminal coronary angioplasties; CBG: coronary bypass grafts; TIA: transient ischemic attacks.

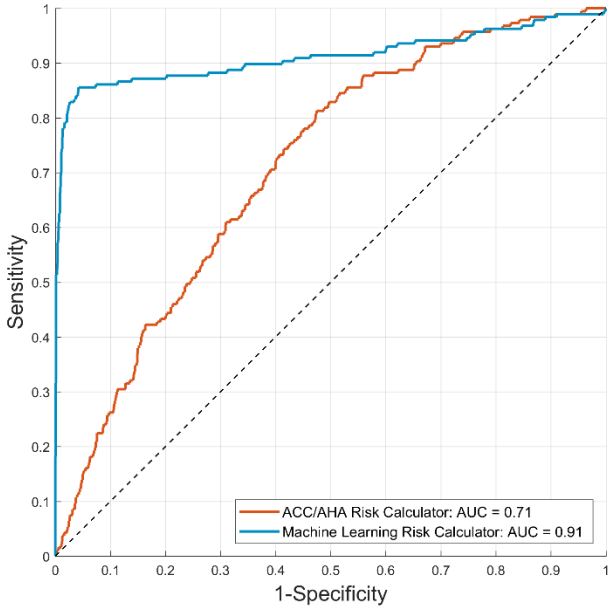


(a)

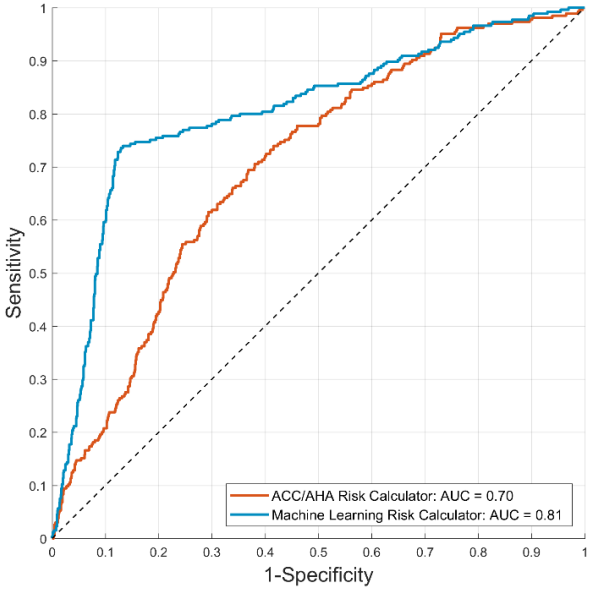

ACC/AHA Risk Calculator
ML Risk Calculator
Number of missed events
70
60
50
40
30
20
10
0
57
7
49
12
10
2
3
2
62
5
37
7
22
5
19
8
10
1
7
0
9
0
3
0
2
2
1
1
1
0
MI
Stroke
CHD Death
Stroke Death
Definite Angina
CHF
TIA
PVD
PTCA
Resuscitated Cardiac Arrest
Probable Angina
Other Revascularization
Other CVD Death
Other Atherosclerotic Death
CBG

(b)

FIGURE S4. ROC curves for prediction of Hard CVD events (a) when training and testing on "White Race" MESA cohort, and (b) when training on "White Race" MESA cohort and testing on FLEMENGHO cohort comparing the ML Risk Calculator (blue) with the ACC/AHA Risk Calculator (red). AUC: Area under the curve.



(a)



(b)

Supplemental References:

1. Chang C-C, Lin C-J. Libsvm: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*. 2011;2:1-27.

2. Almogahed BA, Kakadiaris IA. Neater: Filtering of over-sampled data using non-cooperative game theory. *Proc. International Conference of Pattern Recognition*. 2014:1371--1376.

3. Nash J. Non-cooperative games. *Annals of Mathematics*. 1951;54:286-295.

4. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. Smote: Synthetic minority over-sampling technique. *J. Artif. Int. Res.* 2002;16:321-357.

5. Weng SF, Reps J, Kai J, Garibaldi JM, Qureshi N. Can machine-learning improve cardiovascular risk prediction using routine clinical data? *PLoS One*. 2017;12:e0174944.

6. Isler Y, Narin A, Ozer M. Comparison of the effects of cross-validation methods on determining performances of classifiers used in diagnosing congestive heart failure. *Meas Sci Rev*. 2015;15:196-201.

7. Wiens AD, Inan OT. Accelerometer body sensor network improves systolic time interval assessment with wearable ballistocardiography. *Proc IEEE Eng Med Biol Soc*. 2015;2015:1833-1836.

8. Alcaraz R, Martínez A, Rieta JJ. Role of the p-wave high frequency energy and duration as noninvasive cardiovascular predictors of paroxysmal atrial fibrillation. *Comput Methods Programs Biomed*. 2015;119:110-119.

9. Dietterich TG. Approximate statistical tests for comparing supervised classification learning algorithms. *Neural Comput*. 1998;10:1895-1923.

10. Arlot S, Celisse A. Approximate statistical tests for comparing supervised classification learning algorithms. *Statist Surv*. 2010;4:40-79.

11. Burman P. A comparative study of ordinary cross-validation, v-fold cross-validation and the repeated learning-testing methods. *Biometrika*. 1989;76:503-514.