

Original Paper

# The Impact of Time Horizon on Classification Accuracy: Application of Machine Learning to Prediction of Incident Coronary Heart Disease

Steven Simon<sup>1</sup>, BA, MD; Divneet Mandair<sup>2</sup>, BA, MD; Abdel Albakri<sup>2</sup>, BA, MD; Alison Fohner<sup>3</sup>, PhD; Noah Simon<sup>4</sup>, PhD; Leslie Lange<sup>5</sup>, PhD; Mary Biggs<sup>6</sup>, PhD; Kenneth Mukamal<sup>7</sup>, MD; Bruce Psaty<sup>3,8</sup>, MD, PhD; Michael Rosenberg<sup>1</sup>, MD

<sup>1</sup>Division of Cardiology, University of Colorado School of Medicine, Aurora, CO, United States

<sup>2</sup>University of Colorado School of Medicine, Aurora, CO, United States

<sup>3</sup>Department of Epidemiology, University of Washington, Seattle, WA, United States

<sup>4</sup>Department of Biostatistics, University of Washington, Seattle, WA, United States

<sup>5</sup>Division of Biomedical Informatics and Personalized Medicine, University of Colorado Anschutz Medical Campus, Aurora, CO, United States

<sup>6</sup>Cardiovascular Health Research Unit, University of Washington, Seattle, WA, United States

<sup>7</sup>Department of Medicine, Beth Israel Deaconess Medical Center, Boston, MA, United States

<sup>8</sup>Department of Medicine, University of Washington, Seattle, WA, United States

**Corresponding Author:**

Steven Simon, BA, MD

Division of Cardiology

University of Colorado School of Medicine

13001 E 17th Pl

Aurora, CO, 80045

United States

Phone: 1 303 724 6946

Email: [steven.simon@cuanschutz.edu](mailto:steven.simon@cuanschutz.edu)

## Abstract

**Background:** Many machine learning approaches are limited to classification of outcomes rather than longitudinal prediction. One strategy to use machine learning in clinical risk prediction is to classify outcomes over a given time horizon. However, it is not well-known how to identify the optimal time horizon for risk prediction.

**Objective:** In this study, we aim to identify an optimal time horizon for classification of incident myocardial infarction (MI) using machine learning approaches looped over outcomes with increasing time horizons. Additionally, we sought to compare the performance of these models with the traditional Framingham Heart Study (FHS) coronary heart disease gender-specific Cox proportional hazards regression model.

**Methods:** We analyzed data from a single clinic visit of 5201 participants of a cardiovascular health study. We examined 61 variables collected from this baseline exam, including demographic and biologic data, medical history, medications, serum biomarkers, electrocardiographic, and echocardiographic data. We compared several machine learning methods (eg, random forest, L1 regression, gradient boosted decision tree, support vector machine, and k-nearest neighbor) trained to predict incident MI that occurred within time horizons ranging from 500-10,000 days of follow-up. Models were compared on a 20% held-out testing set using area under the receiver operating characteristic curve (AUROC). Variable importance was performed for random forest and L1 regression models across time points. We compared results with the FHS coronary heart disease gender-specific Cox proportional hazards regression functions.

**Results:** There were 4190 participants included in the analysis, with 2522 (60.2%) female participants and an average age of 72.6 years. Over 10,000 days of follow-up, there were 813 incident MI events. The machine learning models were most predictive over moderate follow-up time horizons (ie, 1500-2500 days). Overall, the L1 (Lasso) logistic regression demonstrated the strongest classification accuracy across all time horizons. This model was most predictive at 1500 days follow-up, with an AUROC of 0.71. The most influential variables differed by follow-up time and model, with gender being the most important feature for the

L1 regression and weight for the random forest model across all time frames. Compared with the Framingham Cox function, the L1 and random forest models performed better across all time frames beyond 1500 days.

**Conclusions:** In a population free of coronary heart disease, machine learning techniques can be used to predict incident MI at varying time horizons with reasonable accuracy, with the strongest prediction accuracy in moderate follow-up periods. Validation across additional populations is needed to confirm the validity of this approach in risk prediction.

(*JMIR Cardio* 2022;6(2):e38040) doi: [10.2196/38040](https://doi.org/10.2196/38040)

## KEYWORDS

coronary heart disease; risk prediction; machine learning; heart; heart disease; clinical; risk; myocardial; gender

## Introduction

Cardiovascular disease (CVD) is the leading cause of morbidity and mortality in the United States and worldwide. The prevalence of CVD in adults within the United States has reached 48% and greater than 130 million adults in the United States are projected to have CVD by 2035, with total costs expected to reach US \$1.1 trillion [1]. The leading cause of deaths attributable to CVD are from coronary heart disease, followed by stroke, hypertension, and heart failure [1]. This year alone, roughly 605,000 Americans will have an incident myocardial infarction (MI) and greater than 110,000 will die from MI [1]. Given the high prevalence of MI, there is significant focus on identifying those most likely to develop incident coronary heart disease [2-5]. If properly identified, primary preventive pharmacologic and lifestyle strategies can be applied to those at the highest risk [6].

Historically, risk prediction models have been developed by applying traditional statistical models (ie, regression-based models and Cox) to cohort data [7-10]. These analyses have provided a breadth of information about the risk of CVD and have been very useful clinically, given their straightforward relationships between a small number of variables and the outcome of interest [11-16]. However, these risk scores often do not achieve high reliability when applied to novel data sets [10,17]. Currently, roughly half of MIs and strokes occur in people who are not predicted to be at an elevated risk for CVD [18].

Machine learning has been introduced as a novel method for processing large amounts of data, focused primarily on accurate prediction rather than understanding the relative effect of risk factors on disease. In some applications, machine learning methods have been found to improve upon traditional regression models for predicting various cardiovascular outcomes [19-22]. A key aspect of applying machine learning methods is the bias-variance trade-off or balancing how accurately a model fits the training data (bias) and how well it can be applied broadly (variance) in out-of-sample testing or validation data [23]. Machine learning models tend to excel when dealing with a large number of covariates and nonlinear or complex relationships of covariates, often at the expense of overfitting a particular training set [24]. However, with an increased ability to model complex interactions between covariables comes a decrease in understanding how risk factors relate to an outcome. Additionally, one key limitation of many machine learning methods is that they are often classification models that do not include well-developed methods to incorporate information

about time-to-event data. Investigators often select a single time horizon for classification, but how varying time horizons affect the relative prediction accuracy is a relatively unexplored aspect of machine learning methods. We hypothesize that there is a trade-off in the selection of the predictive time horizon, in which the use of shorter time horizons offers an increased relevance of predictors to outcomes and greater effect sizes. This is balanced against an increase in the number of events when the time horizon is of longer duration. Based on this trade-off, we would predict that moderate time horizons would have the highest predictive accuracy.

With this investigation, we examined the impact of varying time horizons on the prediction of incident MI. Using data from the Cardiovascular Health Study (CHS) [25], we examined the predictive accuracy of multiple machine learning algorithms over varying time frames of 500 days through 10,000 days of follow-up to identify incident MI. Additionally, we used the Framingham Heart Study (FHS) coronary heart disease gender-specific Cox proportional hazards regression model for comparison to the machine learning models. We aimed to find what time horizon would have the highest predictive accuracy and examine how this compared with the prediction accuracy of the FHS regression model.

## Methods

### Ethical Considerations

Data were approved for use by the Cardiovascular Health Study Policies and Procedures Committee with accompanying data and materials distribution agreement.

### Data Set Creation

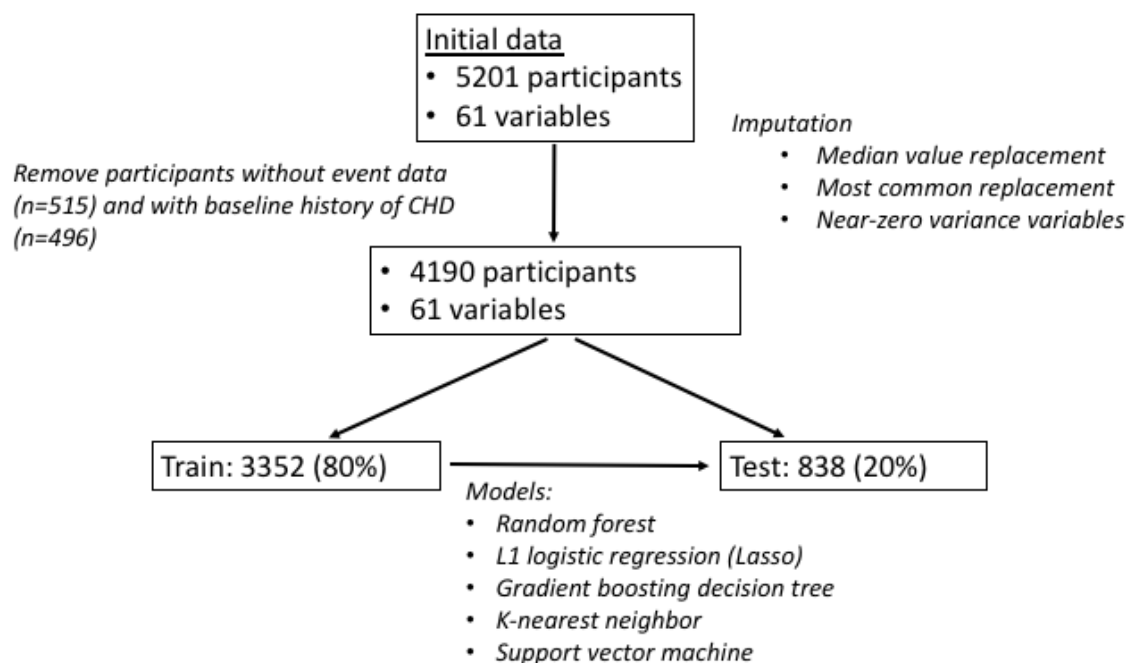
We used anonymized data from the CHS [25], the design and objectives of which have been previously described. Briefly, the CHS is a longitudinal study of men and women aged 65 years or older, recruited from a random sample of Medicare-eligible residents of Pittsburgh, PA, Forsyth County, NC, Sacramento, CA, and Hagerstown, MD. The original cohort of 5201 participants was enrolled in 1989-1990 and serves as the sample for this study. Baseline data were obtained in this cohort, and routine clinic visits and telephone interviews were conducted periodically going forward.

We excluded patients with a baseline history of prior MI from the cohort. We examined 61 variables collected from the baseline exam, including demographic and biologic data (Table S1 in [Multimedia Appendix 1](#)).

Using an end point of incident MI, we applied multiple machine learning methods across varying time horizons to define an optimal risk prediction. Missing variable data was quite uncommon for baseline demographic and laboratory data. Although overall infrequent, missing data was more common

for electrocardiogram variables. In these cases of missing data, imputation was performed on missing variables using median value replacement for continuous variables and most common replacement for categorical variables (Figure 1).

**Figure 1.** Analysis flowchart. CHD: Cardiovascular Health Study.



## Statistical Analysis

The data set was randomly split into a training set (80%) and a testing or validation set (20%). The training data set was used to construct 5 machine learning models: random forest, L1 (LASSO) regression, support vector machine, k-nearest neighbor, and gradient boosted decision tree. Hyperparameter tuning to identify the optimal values for parameters that are not learned during the training process was performed using the validation set. These models were then applied to the test set to examine model performance, which was assessed using an area under the receiver operating characteristic curve (AUROC). Additionally, we used the FHS coronary heart disease Cox proportional hazards regression model as a comparison to the machine learning models (Table S2 in [Multimedia Appendix 1](#)) [7,9,26].

Starting at 500 days, we looped each model over 500-day time horizons in order to identify the optimal predictive horizon up through 10,000 days of follow-up time. For each time horizon, variable importance algorithms were applied to the L1 regression and random forest models. In the L1 regression model, coefficients that are less helpful to the model were shrunk to zero, thereby removing unneeded variables altogether. The remaining coefficients are the variables selected. Because models use normalized inputs, direct comparison of coefficients can be performed based on the absolute value of the average coefficient for each input. In the random forest algorithm, we performed a “permutation” feature selection, which measures the prediction strength of each variable by measuring the

decrease in accuracy when a given variable is essentially voided within the model.

Preliminary analyses identified a high degree of bias related to the cases that were selected within the held-out split sample, and so we performed 50 analyses with different random seeds, with separate results stored for each model, time horizon, and seed number (a total of 1000 separate models for each type of model). Results were compiled based on the average AUROC, coefficient value (L1 regression), and impurity or accuracy (random forest) for each model. Model comparison was performed using linear mixed effects models, with seed number as the random effect and unstructured covariance matrix pattern.

All modeling was performed using publicly available packages on R software (version 1.1.463; The R Foundation for statistical computing). The code used for analysis is provided in [Multimedia Appendix 1](#). Model comparisons (mixed effects models) were performed using Stata IC (version 14; Stata, Inc).

## Results

Baseline characteristics of the study participants are presented in [Table 1](#). There were a total of 4190 participants included. The average age of the cohort was 72.6 years, and 2522 (60.2%) participants were female. At baseline, 2201 (53%) had a history of ever using tobacco, 2300 (55%) had a diagnosis of hypertension, and 389 (9.3%) had a diagnosis of diabetes. Over 30 years of follow-up, there were 813 incident MI events at a median follow-up time of 4725 days.

**Table 1.** Baseline Characteristics of the study participants.

Characteristics	Values (N=4190)
Age (years), mean (SD)	72.6 (5.6)
Gender (male), n (%)	1668 (39.8)
Tobacco consumption, n (%)	2201 (53)
Hypertension, n (%)	2300 (55)
Diabetes, n (%)	389 (9.3)
Total Cholesterol (mg/dL), mean (SD)	211 (38)
BMI, mean (SD)	26.4 (1.9)

### Comparison of Prediction Models Across Time Horizons

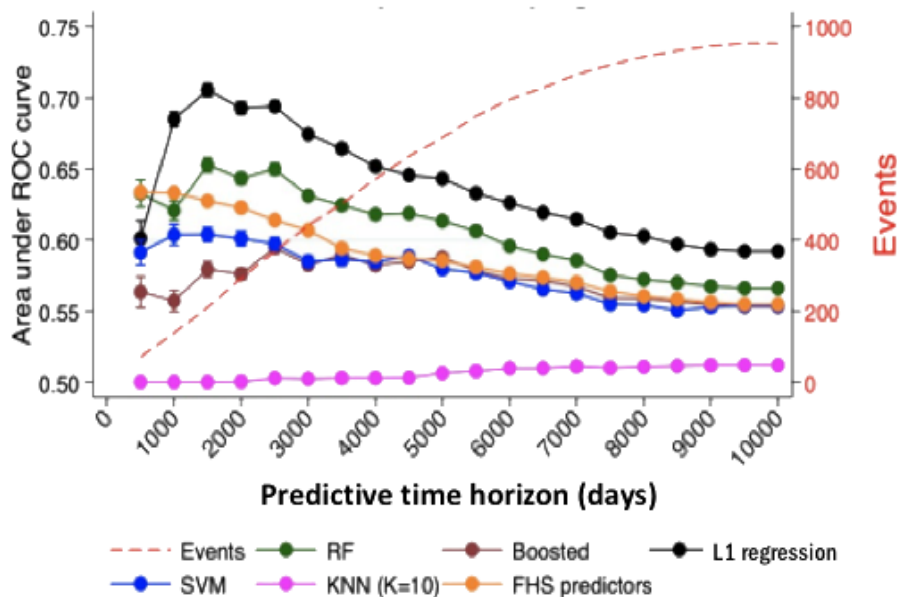
Relative performance of the machine learning methods and FHS model is displayed in Figure 2 as the AUROC across cut points for the time horizon. The machine learning models were generally most predictive over moderate time horizons of 1500-2500 days of follow-up.

In addition to examining AUROC, we also examined the area under the precision-recall curve (Figure 3), which favored later time horizons, but with no change in the order of model performance. The L1 regression model still had the highest performance across time points.

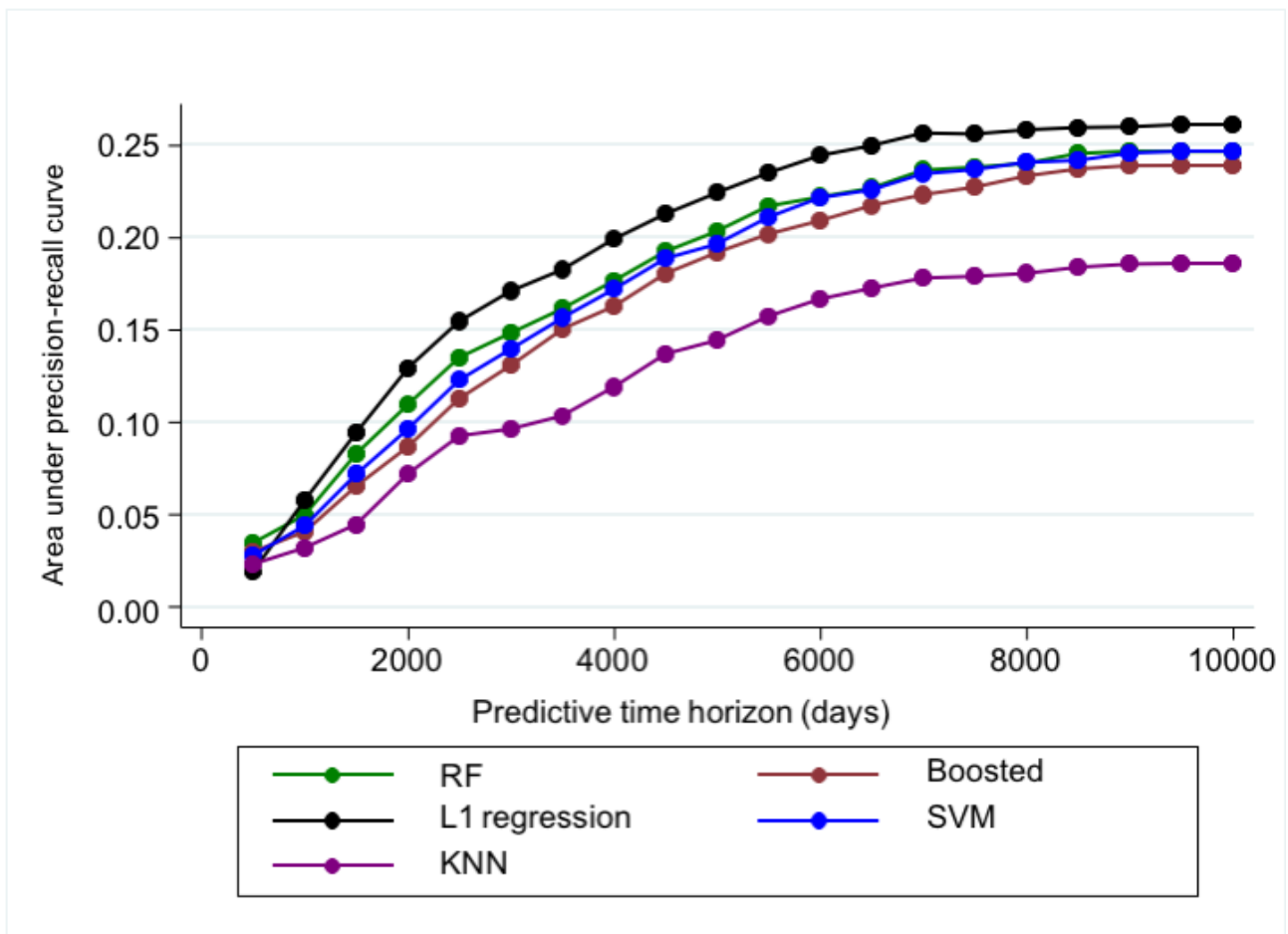
The L1 logistic regression was overall the most predictive across all time points (Figure 4) and displayed the highest prediction accuracy at 1500-day time horizon with an AUROC of 0.71. The k-nearest neighbor model performed relatively poorly across all time points.

When compared with the FHS model, the L1 model performed worse at 500 days of follow-up but had superior prediction accuracy at all subsequent follow-up times. The random forest model performed better than the FHS model starting at 1500 days of follow-up and longer. The remaining machine learning models were less predictive than the FHS model across all time frames (Figure 2).

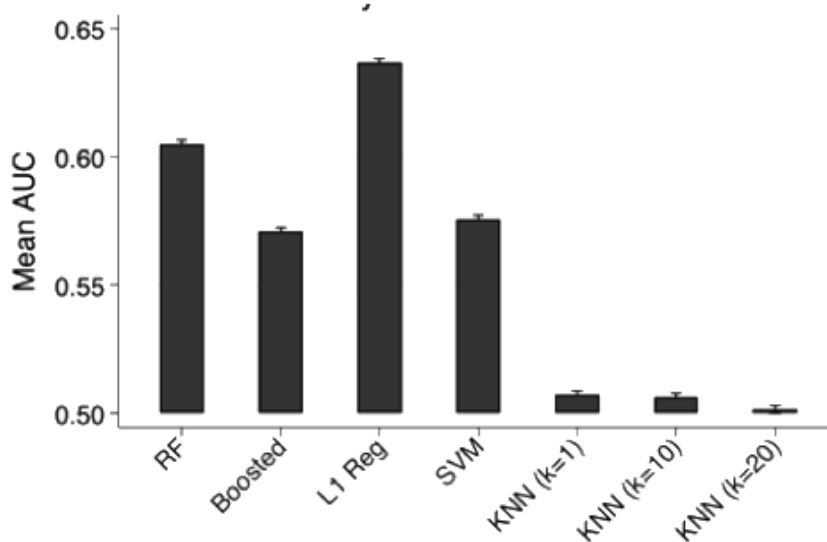
**Figure 2.** Predictive accuracy over varying time horizons. FHS: Framingham Heart Study; KNN: k-nearest neighbor; RF: random forest; ROC: receiver operating characteristics; SVM: support vector machine.



**Figure 3.** Predictive Accuracy using area under precision-recall curve. KNN: k-nearest neighbor; PR: precision-recall; RF: random forest; SVM: support vector machine.



**Figure 4.** Prediction accuracy across all time horizons. AUC: area under the curve; KNN: k-nearest neighbor; RF: random forest; SVM: support vector machine.



## Feature Selection

Some machine learning algorithms allow for analysis of variable

contributions to the model. For this analysis, feature importance was performed across all time points for the L1 regression and random forest models (Table 2).

**Table 2.** Feature selection (top features).

Model	Short-term follow-up (500-1000 days)	Intermediate follow-up (1500-2500 days)	Long-term follow-up (>2500 days)
L1 regression	<ul style="list-style-type: none"> <li>Gender (0.90)</li> <li>Calcium channel blockers (0.47)</li> <li>IVCD<sup>a</sup> by ECG<sup>b</sup> (0.40)</li> <li>Diabetes mellitus (0.32)</li> <li>Smoking (0.22)</li> <li>Systolic blood pressure (0.21)</li> </ul>	<ul style="list-style-type: none"> <li>Gender (1.03)</li> <li>Diabetes mellitus (0.33)</li> <li>Calcium channel blockers (0.42)</li> <li>Hypertension (0.27)</li> <li>Alcohol (per week) (-0.21)</li> </ul>	<ul style="list-style-type: none"> <li>Gender (0.50)</li> <li>Calcium channel blockers (0.33)</li> <li>Diabetes mellitus (0.20)</li> </ul>
Random forest	<ul style="list-style-type: none"> <li>Weight</li> <li>FEV1<sup>c</sup></li> <li>BMI</li> <li>Height</li> <li>LDL-C<sup>d</sup></li> </ul>	<ul style="list-style-type: none"> <li>Weight</li> <li>FEV1</li> <li>BMI</li> <li>Height</li> <li>Gender</li> </ul>	<ul style="list-style-type: none"> <li>Weight</li> <li>Total cholesterol</li> <li>BMI</li> <li>Height</li> <li>LDL-C</li> </ul>

<sup>a</sup>IVCD: intraventricular conduction delay.

<sup>b</sup>ECG: electrocardiogram.

<sup>c</sup>FEV1: forced expiratory volume in one second.

<sup>d</sup>LDL-C: low-density lipoprotein cholesterol.

For the L1 regression, the most important variables (based on the absolute value of coefficients applied to normalized inputs) at short-term follow-up intervals (ie, <1000 days) were gender, history of diabetes, use of calcium channel blockers or  $\beta$ -blockers, and having a ventricular conduction defect by electrocardiogram. At intermediate follow-up interval (ie, 1500-2500 days), the most important variables were gender, use of calcium-channel blocker, history of diabetes, and history of hypertension. At longer follow-up times (ie, >2500 days), the most important variables were gender, use of calcium channel blocker, and history of diabetes.

For the random forest variable selection based on accuracy, the most important variables at short-term follow-up intervals (ie, <1000 days) were weight, forced expiratory volume (FEV) by pulmonary function testing, BMI, height, and low-density lipoprotein (LDL) cholesterol. At intermediate follow-up interval (1500-2500 days), the most important variables were weight, FEV, BMI, height, and gender. At longer follow-up times (ie, >2500 days), the most important variables were weight, height, BMI, LDL cholesterol, and total cholesterol.

## Discussion

### Principal Findings

This study demonstrates the ability to use machine learning methods for the prediction of incident MI over varying time horizons in cohort data. Using AUROC as the primary metric for model performance, prediction across all models was most accurate in the moderate (ie, 1500-2500 day) follow-up horizon. The L1 regularized regression provided the most accurate prediction across all time frames, followed by the random forest algorithms. These two models compared favorably to the FHS coronary heart disease prediction variables, especially at longer follow-up intervals. Applying ranked variable importance

algorithms demonstrated how the variables selected differed over time and in different models.

Prediction was most accurate in the moderate follow-up horizon. We suspect that this was due to the balance of accumulating enough events while still being close in time to the baseline data collected. A predictor that is measured closer in time to the outcome is more likely to be relevant in prediction, and as more events accumulate over time, the power to identify a predictive model increases. Prior studies have looked at machine learning prediction of coronary heart disease at short and intermediate follow-up times; however, to our knowledge, this is the first study to apply models to annual time horizons from short- to long-term follow-up [27].

The L1 regularized regression generally provided the most accurate prediction across all time frames. These regularized regression models expand upon traditional regression models by searching across all variables for the best subset of predictors prior to fitting a regression model. An L1 (Lasso) regression differs from other regularized regression models in that it can shrink the importance of many variables to zero, allowing for feature selection in addition to preventing overfitting. As such, it is very useful when using many variables, like in a cohort or electronic health record data. Prior studies have found these models to be comparable to more advanced machine learning methods for predicting clinical outcomes [28]. The random forest model also performed quite well. Random forest is a regularized form of classification and regression tree model that searches for the covariates that best split the data based on outcome, and then continues to split using additional covariates until many decision “trees” are formed. These models avoid overfitting and can also overcome nonlinearity and handle many variables. The accuracy of the L1 regression and random forest prediction models based on AUROC is reasonable in our study in comparison to prior work [29]. It is worthy of note that we

did not include interaction or polynomial terms in the L1 regression, and as such, this model would not be able to identify nonlinear effects between predictors in the same manner as random forest. Our finding that L1 regression provided superior predictive accuracy despite this limitation suggests that nonlinear effects may be less important with these predictors for coronary artery disease or MI, although further work would be needed to support this claim.

With machine learning models, the relationship between any one variable and the outcome is not as clear as with standard regression models. However, some methods can provide the relative importance of each variable to the model creation. We performed ranked variable analysis for the L1 regression and random forest models. We found that, generally, the models found traditional risk factors to be the most important; however, these most important variables changed over time.

The random forest variable importance found weight, height, LDL-cholesterol, and BMI to be highly important across time frames. FEV was important in short- and medium-term follow-up but less important in longer-term follow-up. For the L1 regression, gender, history of diabetes, and the use of calcium channel blockers were important variables across all time horizons. Although these associations are interesting, causation cannot be applied to these analyses, and it can only suggest further study on the importance of these variables.

### Limitations

This study has some notable limitations. First, the CHS [25] data for incident MI are failure time data, and our model does

not allow for censored observations due to lack of follow-up. Second, both testing and validation were performed only within the CHS cohort. Although on the one hand, this is an important examination of a specific population, it limits the applicability of our findings to the global population. Machine learning models are very sensitive to the training population and have been found to be biased when created in one population and applied in another. Since the CHS cohort is composed of individuals over the age of 65 years, this analysis provides an opportunity to study machine learning models in this group. We used the original cohort of 5201 participants enrolled in the CHS, which leaves out a subsequent, predominantly African American cohort, making the results less applicable to the global population. Given these limitations, this analysis needs to be validated in novel cohorts. Additionally, this model cannot easily be directly applied to clinical practice; however, this study presents a model for performing similar analysis in more clinically applicable data sets, including electronic health record data. We aim to accomplish this with future studies.

### Conclusions

In a population free of coronary heart disease, machine learning techniques can be used to accurately predict development of incident MI at varying time horizons. Moderate follow-up time horizons appear to have the most accurate prediction given the balance between proximity to baseline data and allowing ample number of events to occur. Future studies are needed to validate this technique in additional populations.

### Acknowledgments

This research was supported by contracts 75N92021D00006, HHSN268201200036C, HHSN268200800007C, N01HC55222, N01HC85079, N01HC85080, N01HC85081, N01HC85082, N01HC85083, and N01HC85086, as well as grants U01HL080295 and U01HL130114 from the National Heart, Lung, and Blood Institute (NHLBI), with additional contribution from the National Institute of Neurological Disorders and Stroke (NINDS). Additional support was provided by R01AG023629 from the National Institute on Aging (NIA). A full list of principal CHS investigators and institutions can be found at CHS-NHLBI.org website. This work was also funded by grants from the National Institute of Health/NHLBI (MAR: 5K23 HL127296, R01 HL146824).

### Conflicts of Interest

None declared.

### Multimedia Appendix 1

Tables and code used for model analysis.

[\[DOCX File, 43 KB-Multimedia Appendix 1\]](#)

### References

1. Benjamin E, Muntner P, Alonso A, Bittencourt M, Callaway C, Carson A. Heart disease and stroke statistics-2019 update: a report from the American Heart Association. *Circulation* 2019;137(12):e493.
2. Alonso A, Norby FL. Predicting atrial fibrillation and its complications. *Circ J* 2016;80(5):1061-1066. [doi: [10.1253/circj.cj-16-0239](https://doi.org/10.1253/circj.cj-16-0239)]
3. Schnabel RB, Sullivan LM, Levy D, Pencina MJ, Massaro JM, D'Agostino RB, et al. Development of a risk score for atrial fibrillation (Framingham Heart Study): a community-based cohort study. *The Lancet* 2009 Feb;373(9665):739-745. [doi: [10.1016/s0140-6736\(09\)60443-8](https://doi.org/10.1016/s0140-6736(09)60443-8)]
4. Chamberlain AM, Agarwal SK, Folsom AR, Soliman EZ, Chambless LE, Crow R, et al. A clinical risk score for atrial fibrillation in a biracial prospective cohort (from the Atherosclerosis Risk in Communities [ARIC] study). *Am J Cardiol* 2011 Jan;107(1):85-91 [FREE Full text] [doi: [10.1016/j.amjcard.2010.08.049](https://doi.org/10.1016/j.amjcard.2010.08.049)] [Medline: [21146692](https://pubmed.ncbi.nlm.nih.gov/21146692/)]

5. Alonso A, Krijthe BP, Aspelund T, Stepos KA, Pencina MJ, Moser CB, et al. Simple risk model predicts incidence of atrial fibrillation in a racially and geographically diverse population: the charge - af consortium. *JAHA* 2013 Mar 12;2(2):e000102. [doi: [10.1161/jaha.112.000102](https://doi.org/10.1161/jaha.112.000102)]
6. Chou R, Dana T, Blazina I, Daeges M, Jeanne TL. Statins for prevention of cardiovascular disease in adults: evidence report and systematic review for the us preventive services task force. *JAMA* 2016 Nov 15;316(19):2008-2024. [doi: [10.1001/jama.2015.15629](https://doi.org/10.1001/jama.2015.15629)] [Medline: [27838722](https://pubmed.ncbi.nlm.nih.gov/27838722/)]
7. Wilson PWF, D'Agostino RB, Levy D, Belanger AM, Silbershatz H, Kannel WB. Prediction of coronary heart disease using risk factor categories. *Circulation* 1998 May 12;97(18):1837-1847. [doi: [10.1161/01.cir.97.18.1837](https://doi.org/10.1161/01.cir.97.18.1837)] [Medline: [9603539](https://pubmed.ncbi.nlm.nih.gov/9603539/)]
8. Goff DJ, Lloyd-Jones D, Bennett G, Coady S, D'Agostino R, Gibbons R, et al. 2013 ACC/AHA guideline on the assessment of cardiovascular risk. *Circulation* 2014 Jun 24;129(25\_suppl\_2):S49-S73 [FREE Full text] [doi: [10.1161/01.cir.0000437741.48606.98](https://doi.org/10.1161/01.cir.0000437741.48606.98)] [Medline: [24222018](https://pubmed.ncbi.nlm.nih.gov/24222018/)]
9. D'Agostino RB, Grundy S, Sullivan LM, Wilson P, CHD Risk Prediction Group. Validation of the Framingham coronary heart disease prediction scores: results of a multiple ethnic groups investigation. *JAMA* 2001 Jul 11;286(2):180-187. [doi: [10.1001/jama.286.2.180](https://doi.org/10.1001/jama.286.2.180)] [Medline: [11448281](https://pubmed.ncbi.nlm.nih.gov/11448281/)]
10. DeFilippis AP, Young R, McEvoy JW, Michos ED, Sandfort V, Kronmal RA, et al. 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 Feb 21;38(8):598-608 [FREE Full text] [doi: [10.1093/eurheartj/ehw301](https://doi.org/10.1093/eurheartj/ehw301)] [Medline: [27436865](https://pubmed.ncbi.nlm.nih.gov/27436865/)]
11. The ARIC investigators. The atherosclerosis risk in community (ARIC) study: design and objectives. *Am J Epidemiol* 1989;129(4):687-702. [doi: [10.1093/oxfordjournals.aje.a115184](https://doi.org/10.1093/oxfordjournals.aje.a115184)]
12. Nasir K, Tsai M, Rosen BD, Fernandes V, Bluemke DA, Folsom AR, et al. Elevated homocysteine is associated with reduced regional left ventricular function. *Circulation* 2007 Jan 16;115(2):180-187. [doi: [10.1161/circulationaha.106.633750](https://doi.org/10.1161/circulationaha.106.633750)]
13. Bild D, Bluemke D, Burke G, Detrano R, Diez Roux AV, Folsom A, et al. Multi-ethnic study of atherosclerosis: objectives and design. *Am J Epidemiol* 2002 Nov 01;156(9):871-881. [doi: [10.1093/aje/kwf113](https://doi.org/10.1093/aje/kwf113)] [Medline: [12397006](https://pubmed.ncbi.nlm.nih.gov/12397006/)]
14. Rosenberg M, Gottdiener J, Heckbert S, Mukamal K. Echocardiographic diastolic parameters and risk of atrial fibrillation: the cardiovascular health study. *Eur Heart J* 2012 Apr;33(7):904-912 [FREE Full text] [doi: [10.1093/eurheartj/ehr378](https://doi.org/10.1093/eurheartj/ehr378)] [Medline: [21990265](https://pubmed.ncbi.nlm.nih.gov/21990265/)]
15. Rosenberg MA, Lopez FL, Bůžková P, Adabag S, Chen LY, Sotoodehnia N, et al. Height and risk of sudden cardiac death: the atherosclerosis risk in communities and cardiovascular health studies. *Ann Epidemiol* 2014 Mar;24(3):174-179.e2 [FREE Full text] [doi: [10.1016/j.annepidem.2013.11.008](https://doi.org/10.1016/j.annepidem.2013.11.008)] [Medline: [24360853](https://pubmed.ncbi.nlm.nih.gov/24360853/)]
16. Rosenberg M, Patton K, Sotoodehnia N, Karas M, Kizer J, Zimetbaum P, et al. The impact of height on the risk of atrial fibrillation: the cardiovascular health study. *Eur Heart J* 2012 Nov;33(21):2709-2717 [FREE Full text] [doi: [10.1093/eurheartj/ehs301](https://doi.org/10.1093/eurheartj/ehs301)] [Medline: [22977225](https://pubmed.ncbi.nlm.nih.gov/22977225/)]
17. Christophersen IE, Yin X, Larson MG, Lubitz SA, Magnani JW, McManus DD, et al. A comparison of the CHARGE-AF and the CHA2DS2-VASc risk scores for prediction of atrial fibrillation in the Framingham Heart Study. *Am Heart J* 2016 Aug;178:45-54 [FREE Full text] [doi: [10.1016/j.ahj.2016.05.004](https://doi.org/10.1016/j.ahj.2016.05.004)] [Medline: [27502851](https://pubmed.ncbi.nlm.nih.gov/27502851/)]
18. Ridker PM, Danielson E, Fonseca FA, Genest J, Gotto AM, Kastelein JJ, et al. Rosuvastatin to prevent vascular events in men and women with elevated C-reactive protein. *N Engl J Med* 2008 Nov 20;359(21):2195-2207. [doi: [10.1056/nejmoa0807646](https://doi.org/10.1056/nejmoa0807646)]
19. Shouval R, Hadanny A, Shlomo N, Iakobishvili Z, Unger R, Zahger D, et al. Machine learning for prediction of 30-day mortality after ST elevation myocardial infarction: an acute coronary syndrome Israeli survey data mining study. *Int J Cardiol* 2017 Nov 01;246:7-13. [doi: [10.1016/j.ijcard.2017.05.067](https://doi.org/10.1016/j.ijcard.2017.05.067)] [Medline: [28867023](https://pubmed.ncbi.nlm.nih.gov/28867023/)]
20. Mansoor H, Elgendy IY, Segal R, Bavry AA, Bian J. Risk prediction model for in-hospital mortality in women with ST-elevation myocardial infarction: a machine learning approach. *Heart Lung* 2017 Nov;46(6):405-411. [doi: [10.1016/j.hrtlng.2017.09.003](https://doi.org/10.1016/j.hrtlng.2017.09.003)] [Medline: [28992993](https://pubmed.ncbi.nlm.nih.gov/28992993/)]
21. Li X, Liu H, Yang J, Xie G, Xu M, Yang Y. Using machine learning models to predict in-hospital mortality for ST-elevation myocardial infarction patients. *Stud Health Technol Inform* 2017;245:476-480. [Medline: [29295140](https://pubmed.ncbi.nlm.nih.gov/29295140/)]
22. Kakadiaris IA, Vrigkas M, Yen AA, Kuznetsova T, Budoff M, Naghavi M. Machine learning outperforms ACC/AHA CVD risk calculator in MESA. *JAHA* 2018 Nov 20;7(22):e009476. [doi: [10.1161/jaha.118.009476](https://doi.org/10.1161/jaha.118.009476)]
23. Hastie T, Tibshirani R, Friedman J. *The Elements of Statistical Learning: Data Mining, Inference, Prediction*. 2nd ed. New York, NY: Springer-Verlag; 2009:-793.
24. Gorodeski EZ, Ishwaran H, Kogalur UB, Blackstone EH, Hsieh E, Zhang Z, et al. Use of hundreds of electrocardiographic biomarkers for prediction of mortality in postmenopausal women. *Circ Cardiovasc Qual Outcomes* 2011 Sep;4(5):521-532. [doi: [10.1161/circoutcomes.110.959023](https://doi.org/10.1161/circoutcomes.110.959023)]
25. Fried LP, Borhani NO, Enright P, Furberg CD, Gardin JM, Kronmal RA, et al. The cardiovascular health study: design and rationale. *Ann Epidemiol* 1991 Feb;1(3):263-276. [doi: [10.1016/1047-2797\(91\)90005-w](https://doi.org/10.1016/1047-2797(91)90005-w)]
26. Anderson KM, Wilson PW, Odell PM, Kannel WB. An updated coronary risk profile. A statement for health professionals. *Circulation* 1991 Jan;83(1):356-362. [doi: [10.1161/01.cir.83.1.356](https://doi.org/10.1161/01.cir.83.1.356)] [Medline: [1984895](https://pubmed.ncbi.nlm.nih.gov/1984895/)]



27. Dogan M, Beach S, Simons R, Lendasse A, Penaluna B, Philibert R. Blood-based biomarkers for predicting the risk for five-year incident coronary heart disease in the Framingham heart study via machine learning. *Genes (Basel)* 2018 Dec 18;9(12):641 [FREE Full text] [doi: [10.3390/genes9120641](https://doi.org/10.3390/genes9120641)] [Medline: [30567402](https://pubmed.ncbi.nlm.nih.gov/30567402/)]
28. Tiwari P, Colborn KL, Smith DE, Xing F, Ghosh D, Rosenberg MA. Assessment of a machine learning model applied to harmonized electronic health record data for the prediction of incident atrial fibrillation. *JAMA Netw Open* 2020 Jan 03;3(1):e1919396 [FREE Full text] [doi: [10.1001/jamanetworkopen.2019.19396](https://doi.org/10.1001/jamanetworkopen.2019.19396)] [Medline: [31951272](https://pubmed.ncbi.nlm.nih.gov/31951272/)]
29. Ambale-Venkatesh B, Yang X, Wu CO, Liu K, Hundley WG, McClelland R, et al. Cardiovascular event prediction by machine learning. *Circ Res* 2017 Oct 13;121(9):1092-1101. [doi: [10.1161/circresaha.117.311312](https://doi.org/10.1161/circresaha.117.311312)]

## Abbreviations

**AUROC:** area under the receiver operating characteristic curve

**CHS:** Cardiovascular Health Study

**CVD:** cardiovascular disease

**FEV:** forced expiratory volume

**FHS:** Framingham Heart Study

**MI:** myocardial infarction

*Edited by T Leung; submitted 16.03.22; peer-reviewed by A Andy, K Uludag, M Hofford, D Verran; comments to author 12.05.22; revised version received 28.06.22; accepted 09.08.22; published 02.11.22*

*Please cite as:*

*Simon S, Mandair D, Albakri A, Fohner A, Simon N, Lange L, Biggs M, Mukamal K, Psaty B, Rosenberg M*

*The Impact of Time Horizon on Classification Accuracy: Application of Machine Learning to Prediction of Incident Coronary Heart Disease*

*JMIR Cardio* 2022;6(2):e38040

URL: <https://cardio.jmir.org/2022/2/e38040>

doi: [10.2196/38040](https://doi.org/10.2196/38040)

PMID:

©Steven Simon, Divneet Mandair, Abdel Albakri, Alison Fohner, Noah Simon, Leslie Lange, Mary Biggs, Kenneth Mukamal, Bruce Psaty, Michael Rosenberg. Originally published in *JMIR Cardio* (<https://cardio.jmir.org>), 02.11.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in *JMIR Cardio*, is properly cited. The complete bibliographic information, a link to the original publication on <https://cardio.jmir.org>, as well as this copyright and license information must be included.