The use of machine learning for the identification of peripheral artery disease and future mortality risk

Elsie Gyang Ross, MD, MSc,[a] **Nigam H. Shah, MBBS, PhD,**[b] **Ronald L. Dalman, MD,**[a] **Kevin T. Nead, MD, MPhil,**[c] **John P. Cooke, MD, PhD,**[d,e] and **Nicholas J. Leeper, MD,**[a] *Stanford, Calif; Philadelphia, Pa; and Houston, Tex*

Objective: **A key aspect of the precision medicine effort is the development of informatics tools that can analyze and interpret "big data" sets in an automated and adaptive fashion while providing accurate and actionable clinical information. The aims of this study were to develop machine learning algorithms for the identification of disease and the prognostication of mortality risk and to determine whether such models perform better than classical statistical analyses.**
Methods: **Focusing on peripheral artery disease (PAD), patient data were derived from a prospective, observational study of 1755 patients who presented for elective coronary angiography. We employed multiple supervised machine learning algorithms and used diverse clinical, demographic, imaging, and genomic information in a hypothesis-free manner to build models that could identify patients with PAD and predict future mortality. Comparison was made to standard stepwise linear regression models.**
Results: **Our machine-learned models outperformed stepwise logistic regression models both for the identification of patients with PAD (area under the curve, 0.87 vs 0.76, respectively; $P = .03$) and for the prediction of future mortality (area under the curve, 0.76 vs 0.65, respectively; $P = .10$). Both machine-learned models were markedly better calibrated than the stepwise logistic regression models, thus providing more accurate disease and mortality risk estimates.**
Conclusions: **Machine learning approaches can produce more accurate disease classification and prediction models. These tools may prove clinically useful for the automated identification of patients with highly morbid diseases for which aggressive risk factor management can improve outcomes. (J Vasc Surg 2016;64:1515-22.)**

The promise of precision medicine is beginning to be realized in some areas of medicine. In Oncology, for example, genetic profiling is now being used to identify patients for whom tailored chemotherapy regimens—directed against the individual's personal cancer mutation—can be used to significantly improve outcomes relative to traditional therapy.[1] Rather than the current empirical approach to treatment, there is hope that with a deeper understanding of biology and pharmacogenomics, we may one day be able to guarantee that every patient receives the right dose of the right medicine at the right time.[2]

In addition to efforts directed toward personalizing therapy, the precision medicine effort is also focused on refining our diagnostic and predictive capabilities. Indeed, it is becoming increasingly clear that traditional risk prediction models (commonly based on a handful of epidemiologic factors) fail to capture important information about the individual's behavior, environment, comorbidities, and personal biologic makeup. Now, with an influx of detailed patient data residing in large, prospective data sets[3,4] and the rapid adoption of electronic health records, there is an increasing pool of information that can potentially be analyzed to refine our predictive capabilities to deliver better care.[5]

Here, we evaluated the utility of using machine learning algorithms and their ability to integrate disparate data inputs to make predictions about presence or absence of disease and the risk of future events. Specifically, we built several models to identify peripheral artery disease (PAD), a highly prevalent[6] and morbid disease[7] that frequently goes undiagnosed,[8,9] and to predict risk of future mortality. We compared our models with standard logistic regression models to evaluate whether there were significant improvements in predictive ability from using modern machine learning techniques.

METHODS

Study population. Patient data were derived from the Genetic Determinants of Peripheral Arterial Disease

From the Division of Vascular Surgery, Stanford Health Care,[a] and the Center for Biomedical Informatics Research, Stanford University,[b] Stanford; the Department of Radiation Oncology, University of Pennsylvania, Philadelphia[c]; and the Department of Cardiovascular Sciences, Houston Methodist Research Institute,[d] and the Center for Cardiovascular Regeneration, Houston Methodist DeBakey Heart and Vascular Center,[e] Houston.

This work was supported by 1T32HL098049-06 (E.G.R.) and 5R01HL12522402 (N.J.L.).

Author conflict of interest: none.

Additional material for this article may be found online at www.jvascsurg.org.

Correspondence: Nicholas J. Leeper, MD, Division of Vascular Surgery, 300 Pasteur Dr, Ste H3638, Stanford, CA 94305 (e-mail: nleeper@stanford.edu).

The editors and reviewers of this article have no relevant financial relationships to disclose per the JVS policy that requires reviewers to decline review of any manuscript for which they may have a conflict of interest.

0741-5214

http://dx.doi.org/10.1016/j.jvs.2016.04.026

Table I. Variables used in predictive models for identification of peripheral artery disease (*PAD*) and future mortality

Categories	*Variables*
Demographics	Age, gender, self-reported race/ethnicity
Anthropomorphic measures	Height, weight, ABI[a]
Physical examination	Blood pressure, heart rate
Laboratory results	Total cholesterol, LDL, HDL, non-HDL, total cholesterol to HDL ratio, LDL to HDL ratio, serum creatinine, serum glucose
Imaging[a]	Coronary angiography results
Medical history	Coronary artery disease, PAD,[a] congestive heart failure, cerebrovascular disease, diabetes, cardiac arrhythmias, menopause
Clinical events during follow-up[a]	Stroke, myocardial infarction, coronary revascularization, new heart failure
Reported symptoms	Joint pain, claudication
Family history	Presence or absence of cardiovascular diseases in parents
Medications[b]	Aspirin, clopidogrel, statins, antihypertensives, diuretics, beta blockers, antiglycemic agents
Physical activity assessment	Walking Impairment Questionnaire,[1,2] rigorous activity engagement
Genomic markers[c]	rs290481, rs819750, rs7100623, rs7003385, rs94855286, rs46599965, rs3745274, rs2171209, rs16824978, rs107572696
Social factors	Ever married, living situation, total education, current income, employment status, alcohol consumption
Smoking behavior	Ever smoked, current smoker, cumulative pack-years

ABI, Ankle-brachial index; *HDL,* high-density lipoprotein; *LDL,* low-density lipoprotein.

[a] Not used in building PAD classification model.

[b] Also assessed medication compliance (How many times in a month do you forget to take your medications?) and total number of medications.

[c] Selected single-nucleotide polymorphisms found to be associated with PAD.

[1] McDermott MM, Liu K, Guralnik JM, Martin GJ, Criqui MH, Greenland P. Measurement of walking endurance and walking velocity with questionnaire: validation of the walking impairment questionnaire in men and women with peripheral arterial disease. J Vasc Surg 1998; 28:1072-81.

[2] Chang P, Nead KT, Olin JW, Myers J, Cooke JP, Leeper NJ. Effect of physical activity assessment on prognostication for peripheral artery disease and mortality. Mayo Clinic Proc 2015; 90:339-45.

(GenePAD) study, a prospective observational study funded by the National Heart, Lung, and Blood Institute with the goal of identifying key demographic, clinical, and genomic factors that differentiate patients with PAD from those without disease.[10,11] Whereas PAD was the disease of interest in this observational study, it was not used as a factor in enrolling patients. Thus, we used this data set because of the depth and breadth of patient variables collected, its longitudinal nature, and the potential impact on improving care for this subset of patients.

Data from the study include a cohort of 1755 patients who were enrolled at presentation to Stanford University Medical Center or Mount Sinai Medical Center for elective coronary angiography between January 2004 and March 2008. Patients included those referred for complaints of angina or dyspnea or who had abnormal stress test results. Clinical data were collected at the time of enrollment, and patients were prospectively observed for any adverse events. PAD status was not known before enrollment, nor was it used as an inclusion or exclusion criterion. Patients were followed up until the study completion in 2012.

Trained research assistants and nurse practitioners conducted extensive participant interviews and performed careful review of patient history. Variables collected included demographic variables, clinical comorbidities, medications, laboratory tests, physical examination variables, physical activity and smoking behaviors, socioeconomic variables, selected genomic markers associated with PAD, and results of coronary angiography (Table I). Ankle-brachial indices (ABIs) were measured for all patients, and PAD was defined as an ABI of <0.9. All patients provided written consent to participate in the GenePAD study, which was approved by the Stanford and Mount Sinai Institutional Review Boards.

Machine learning. Machine learning refers to methods developed within the fields of statistics, computer science, and artificial intelligence that allow the creation of algorithms that can learn from and make predictions using data. Some commonly used algorithms that we used in this study include elastic net[12] and random forest.[13]

Elastic net is a linear modeling technique similar to linear or logistic regression. The advantage of an elastic net is that in addition to fitting an optimized linear model, a penalty is applied to independent variables in the model such that variables that have little influence on the dependent variable are minimized or dropped from the final model. This has the effect of reducing model complexity while improving the generalizability of the model, which can improve predictive accuracy. Random forest is a "tree-based" algorithm whereby multiple decision trees are built using a random assortment of independent variables that are used to predict an outcome label for a random subset of samples. Using a "majority vote" system, a new sample is predicted by the multiple decision trees in the random forest model, and the ultimate classification of this new sample is based on the classification predicted by a majority of the decision trees.

Model building. The goals of our predictive models were to predict which patients had PAD based on baseline demographic, clinical, and genomic factors and to predict their future risk of mortality. In building our machine learning algorithms, we used a "hypothesis-free" approach to identifying which variables (eg, clinical, demographic, or social) should be included in our models and elected to include any variable for which the majority of patients had a data value. That is, we did not include variables based on any a priori hypotheses of their ability to predict disease presence or absence or risk of future mortality. Variables were simply included if they were available. We then included patients with complete data across variables in the model, then randomly split them into a 70% training set and a 30% test set.

For model training, data from 70% of patients were used to approximate model parameters. We used 10-fold cross-validation for training our model to decrease risk of model overfitting.[14] Multiple machine learning algorithms were used to build predictive models including elastic net, a penalized regression model, and random forest using R version 3.2.1.[15]

Model performance. Each model's ability to discriminate between high- and low-risk patients was determined using the area under the receiver operating characteristic curve (AUC) metric.[16] We also evaluated model calibration (ie, the model's ability to accurately predict observed absolute risk) using the Hosmer-Lemeshow test for goodness of fit, where a P value $< .05$ would indicate poor calibration.[17] From our learned models, we selected the model with the best discrimination and calibration. AUC and calibration are based on model performance on the test set of patients.

Identifying PAD. The ultimate goal of our learned model was to establish an accurate classification algorithm that could identify patients with PAD. To this end, cases of PAD were defined as patients with an ABI <0.9 identified at the time of the GenePAD study, whether or not they had a diagnosis before enrollment. Controls were defined as patients with ABI >0.9. We also had a secondary aim of evaluating whether we could build a model that could identify patients with undiagnosed PAD, given how frequently PAD goes undiagnosed. In the GenePAD data set, for instance, of patients identified as having PAD by ABI at the time of enrollment, 68% had no prior diagnosis.[9] Thus, for our "undiagnosed PAD" model, we included patients with an ABI <0.9 and excluded those who had a prior PAD diagnosis. Table I lists variables included in this model.

Predicting mortality. The goal of this predictive model was to most accurately predict all-cause mortality by the end of the GenePAD study, which lasted a total of 8 years. Follow-up consisted of regular reviews of electronic health records and calls to patients or their families for regular updates. All patients included in our PAD prediction model were also included in the mortality prediction model.

Model comparison. To compare whether our models provided better predictive accuracy than standard methodology such as logistic regression, we also built stepwise logistic regression models for identification of PAD patients and prediction of mortality. In the patient training set, we used univariate analysis to identify which variables were significantly associated with presence of PAD or mortality by χ^2 test for categorical variables and analysis of variance for continuous variables. We then performed forward stepwise logistic regression, including variables with a significance level of at least 10%. The final stepwise logistic regression models were selected on the basis of optimizing the Akaike information content. These final models were then applied to the test patient set, and AUC and calibration were calculated accordingly. To compare differences in discriminatory performance of the machine-learned models to the stepwise logistic regression models, we used the bootstrap test for unpaired receiver operating characteristic curves[18] with 2000 bootstrapped samples. P values $< .05$ were considered significant.

Variable importance. An important aspect of model building is identifying which variables may be of greatest importance in contributing to the accuracy of a model. We thus identified the most heavily weighted variables for each machine-learned model and compared differences in algorithm discrimination performance with and without these variables using integrated discrimination improvement index (IDI)[16] and compared calibration performance using the category-free net reclassification improvement index (NRI),[19] which assessed whether patient risk reclassification was appropriate (ie, higher risk assignment for cases and lower risk assignment for noncases).

RESULTS

Characteristics for those patients included in our predictive models are described in Table II. After paring down variables that were incomplete or redundant, we were left with 130 different variables that included genomic markers, physical activity assessments, demographics, clinical assessments, socioeconomic status, and family history. A total of 1047 patients had complete data across these domains and were included in our PAD and mortality prediction models. For the PAD model, this included 183 patients with PAD and 864 controls. For mortality prediction, there were 129 patients with a mortality event and 918 controls. Median follow-up time for included patients was 5.3 years (interquartile range, 4.3-6.2 years).

Identifying PAD. Balancing model accuracy and calibration, the best model that could identify patients with PAD was a penalized linear regression model, which achieved an AUC of 0.87 (95% confidence interval [CI], 0.81-0.92; Fig 1; Supplementary Table I, online only). The stepwise linear regression model did not perform as well (AUC, 0.76; 95% CI, 0.68-0.85; $P = .03$). The penalized regression model also demonstrated good calibration, whereas the stepwise linear regression model had very poor calibration performance using Hosmer-Lemeshow goodness-of-fit tests ($P = .7$, $P < .0001$, respectively; Supplementary Figs 1 and 2, online only). Fig 2 illustrates the most important variables used in the penalized regression model. Removal of these variables from our penalized regression model resulted in significant changes in net reclassification and discriminatory power of the model (NRI, −0.52 [95% CI, −0.23 to −0.80;

Table II. Characteristics for all patients included in predictive models

Patient characteristics	*All*	*PAD*	*No PAD*
No.	1047	183	864
Age, years	65.6 (10.7)	69.4 (9.4)	64.8 (10.8)
Male gender	692 (66)	104 (57)	588 (68)
Race/ethnicity			
White	581 (55)	102 (56)	479 (55)
African American	141 (13)	41 (22)	100 (11)
Asian	72 (7)	7 (4)	65 (7.5)
Hispanic	106 (10)	19 (10)	87 (10)
Body mass index, kg/m^2	29 (6)	28 (5)	29 (6)
ABI	1.0 (0.2)	0.7 (0.2)	1.1 (0.1)
Comorbidities			
Cardiac arrhythmias	215 (20)	42 (23)	173 (20)
Coronary artery disease	795 (76)	165 (90)	630 (73)
Congestive heart failure	63 (6)	17 (9)	46 (5)
Cerebrovascular disease	49 (5)	21 (11)	28 (3)
Major adverse cardiovascular events[a]	250 (24)	61 (33)	189 (22)
Mortality	129 (11)	34 (18)	95 (11)
Medications			
Aspirin	787 (75)	138 (75)	649 (75)
Antihypertensives	919 (88)	173 (94)	746 (86)
Beta blockers	617 (59)	112 (61)	505 (58)
Insulin	82 (8)	30 (16)	52 (6)
Clopidogrel	390 (37)	88 (48)	302 (35)
Statins	741 (71)	127 (69)	641 (74)
Smoking			
Current smoker	104 (10)	34 (18)	70 (8)
Previous smoker	420 (40)	61 (33)	359 (41)

ABI, Ankle-brachial index; *PAD,* peripheral artery disease.

Categorical variables are presented as number (%). Continuous variables are presented as mean (standard deviation).

[a]Major adverse cardiovascular events include stroke, myocardial infarction, coronary revascularization, and new heart failure.

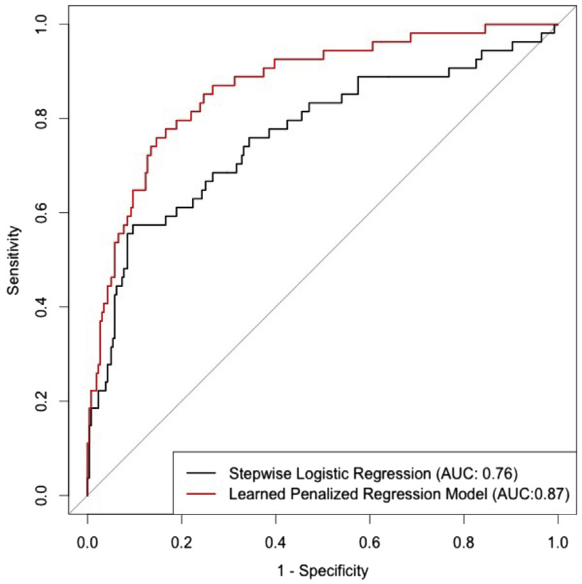


Fig 1. Area under the curve (*AUC*) for stepwise logistic regression and machine-learned penalized regression model for identification of patients with peripheral artery disease (PAD).

P = .0003]; IDI, −0.05 [95% CI, −0.08 to −0.03; P < .0001]).

Identifying undiagnosed PAD. Given how frequently PAD goes undiagnosed, we were also interested in building a predictive model that could potentially identify patients with PAD who did not yet carry a diagnosis. Our training set included a total sample of 993 patients, of whom 103 were found to have an ABI <0.9 and no prior diagnosis of PAD. Our best performing classification model for undiagnosed PAD was a random forest model, which achieved an AUC of 0.84 (95% CI, 0.77-0.91; Supplementary Fig 3, online only). The classical stepwise logistic regression model performed particularly poorly in attempting to identify cases of undiagnosed disease in comparison (AUC, 0.60; 95% CI, 0.5-0.70; P = .0001). The random forest model also demonstrated good calibration, whereas the linear regression model had poor calibration performance using Hosmer-Lemeshow goodness-of-fit tests (P = .06, P < .0001, respectively). The most important variables in the identification of patients with undiagnosed PAD used in the random forest model are illustrated in Supplementary Fig 4 (online only).

Predicting mortality. The best learned model for predicting mortality was a random forest model. The model AUC is 0.76 (95% CI, 0.68-0.84; Fig 3), whereas the stepwise logistic regression model had poorer performance by AUC (0.65; 95% CI, 0.61-0.78; Supplementary Table II, online only). This difference approached significance (P = .10). The learned model demonstrated much better calibration than the stepwise logistic regression model (P = .62, P < .0001, respectively; Supplementary Figs 5 and 6, online only). Fig 4 illustrates the most important variables used in the random forest model. Removal of these variables from the predictive model significantly reduced model discrimination (IDI, −0.07; 95% CI, −0.13 to −0.01; P= .01) but did not significantly affect NRI (−0.28; 95% CI, −0.61 to 0.05; P = .1).

DISCUSSION

We have described an approach leveraging machine learning to created predictive models to identify patients with PAD and to predict mortality in a high-risk population for which good "off-the-shelf" risk prediction models are lacking. Our predictive models, trained in a hypothesis-free fashion on a variety of patient data including genomic, imaging, and socioeconomic variables, outperform standard logistic regression models. Ultimately, such tools may enable the automated detection of individuals with vascular disease and allow personalized, preventive interventions before development of end-stage atherothrombosis.

In general, identification of "at-risk" patients and subsequent risk factor management are integral parts of preventive medical care. Risk stratification methods such as risk prediction scores that are highly used in medicine are often derived from epidemiologic studies and typically describe a handful of risk factors that predict future disease

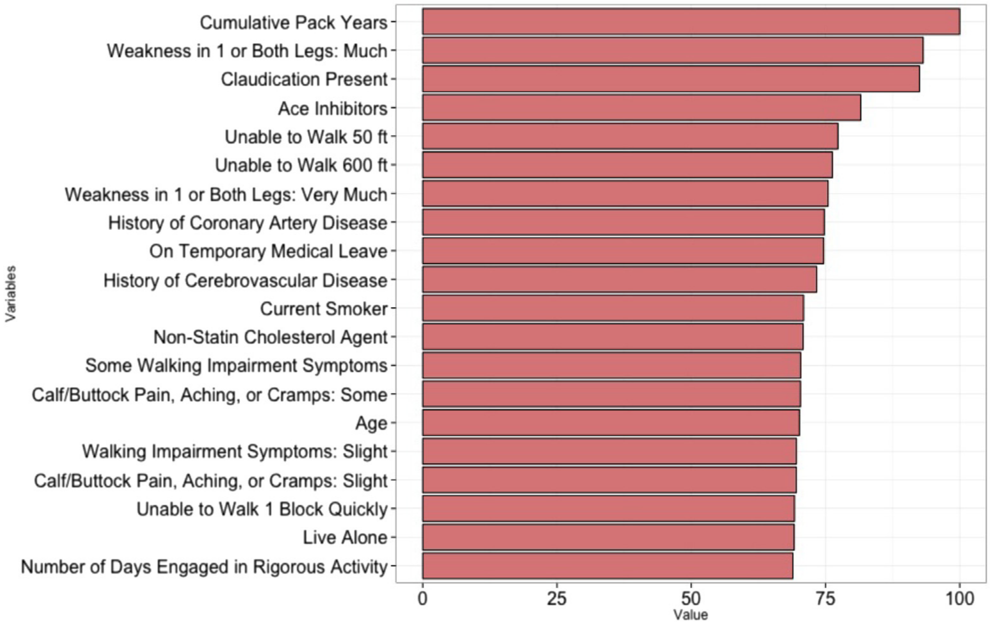


Fig 2. Top 20 weighted variables used in the machine-learned penalized regression model for identification of patients with peripheral artery disease (PAD). *Ace,* Angiotensin-converting enzyme.

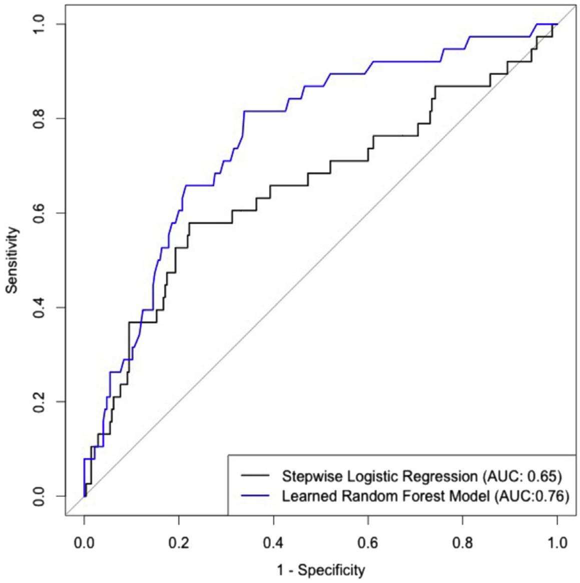


Fig 3. Area under the curve (*AUC*) for stepwise logistic regression and machine-learned random forest model for prediction of mortality.

risk. Such risk scores can be of great clinical utility as they help clinicians identify patients for whom further screening or intervention may change health trajectory.[20-22] There are limitations to such classical risk prediction scores based exclusively on linear modeling, however. In cardiology, for example, the Framingham Risk Score is widely used to identify patients at risk of future major adverse cardiovascular events. However, the Framingham Risk Score often fails to adequately risk stratify patients in more ethnically, socially, and medically diverse populations.[23-25] Investigators have attempted to address these limitations during the last decade by providing "improved" risk scores that incorporate new laboratory, biomarker, or imaging data[26-28]; however, such model adjustments typically have been incremental and confer relatively small or negligible improvements in risk prediction performance.[29] In the specific case of peripheral vascular disease, Duval et al have developed a PAD risk score based on a small number of factors derived from population studies[30]; however, the performance of this tool has been modest.

Our PAD prediction model had excellent discrimination and calibration performance (AUC, 0.87; Hosmer-Lemeshow *P* value = .7) and significantly outperformed a standard stepwise logistic regression model, especially in attempting to identify patients with undiagnosed PAD. Whereas model AUC is an oft-cited metric, calibration is probably a more important aspect of model performance. In addition to providing more accurate classification (ie, disease present or not present), well-calibrated models can also provide meaningful risk estimates for the classification task at hand. For instance, with a well-calibrated PAD model, physicians can be provided with a reliable estimate of the probability of PAD in a patient who might otherwise be missed and potentially prioritize subjects for further screening.

In this study, our models were generated from information captured in a relatively small clinical database. However, we are entering an era in which patient data are becoming ever more ubiquitous, and we will soon be

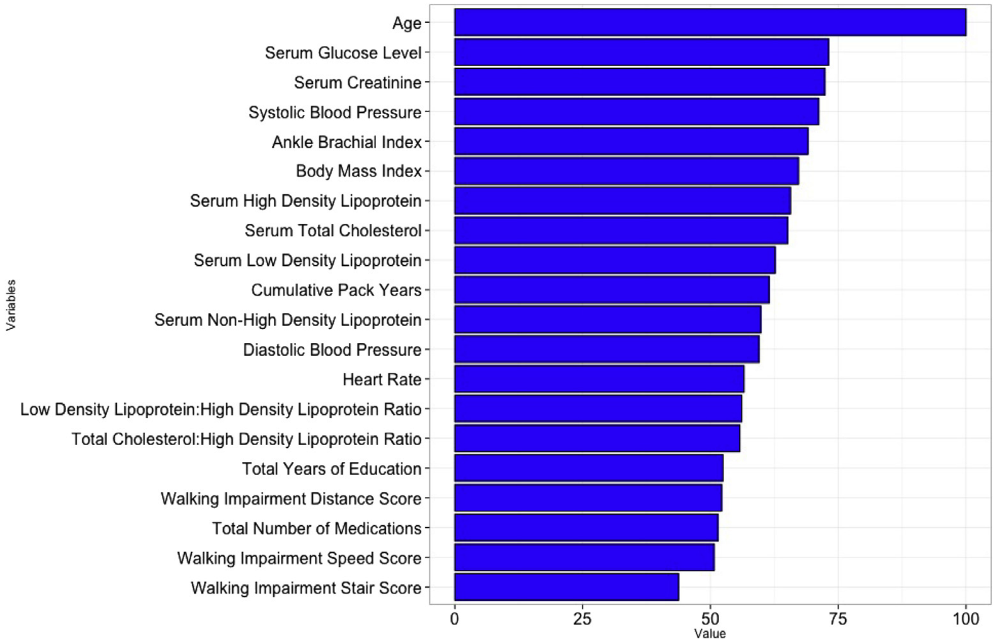


Fig 4. Top 20 weighted variables used in the machine-learned random forest model for mortality prediction.

able to apply predictive analytics to exponentially larger data sets, which should generate even more meaningful and actionable insights.[5] For example, machine learning algorithms are now being designed to continuously surveil the electronic medical record in an automated fashion—*providing real-time predictive analytics for patient care*. Indeed, it is increasingly possible to extract even more nuanced patient data from the electronic health record with rule-based and natural language processing techniques[31,32] and to use these data to accurately predict patient outcomes across a broad range of conditions.[33-35] Using newer analytic approaches, these models will not be static and will one day learn to better predict an individual's clinical trajectory over time.[36,37] Furthermore, such models can be "retrained" locally at different institutions to maximize accuracy in different populations of patients with different clinical and demographic profiles.[38]

We chose to develop algorithms for the detection of PAD for several reasons. First, PAD is a disease that is both highly prevalent and notoriously difficult to diagnose, with >50% of patients in general medical practice going undiagnosed in the United States.[39] Indeed, several studies have shown that PAD frequently goes unidentified, indicating that currently available diagnostic algorithms and screening efforts are insufficient.[8,9,40] Second, PAD is a highly morbid condition known to have worse outcomes than coronary or cerebrovascular disease, but it is undertreated compared with other conditions.[39,40] Third, underlying PAD increases the risk of complications from procedures that these patients frequently undergo, including percutaneous coronary intervention. PAD patients undergoing percutaneous coronary intervention have higher risk of vascular access failure, groin and retroperitoneal hematomas, limb ischemia, and need for blood transfusions.[41,42] Knowledge of a PAD diagnosis before these procedures can help reduce complication rates. Fourth, because intervention for PAD can preserve life and limb,[43,44] it could be useful to have an automated algorithm that can both identify latent disease and single out those individuals at higher risk for adverse outcomes. Last, although there are currently no guidelines recommending routine ABI assessment, predictive algorithms have the potential benefit of providing a cost-effective way of identifying high-risk patients for whom screening and early intervention can significantly affect disease trajectory and decrease health system costs.[45]

A limitation of our study is that we used only patients who had complete data to build our models. In clinical practice, patient data are frequently missing, which may reduce predictive accuracy. One way to address this limitation involves the imputation of missing data, which we found can actually improve predictive accuracy (data not shown). Such methods can be used for real-world data, balancing the limitations of different imputation techniques. Another limitation is that our models may not generalize well to other populations, given that they were trained on data from high-risk patients enrolled at the time of coronary angiography. Although patients within the GenePAD database were not selectively enrolled for their likelihood of having PAD, there may be inherent bias in our data set. As previously mentioned, however, retraining of machine learning algorithms for specific populations is recommended and can greatly improve predictive accuracy. Furthermore, our mortality prediction model had an AUC of 0.76, lower than our PAD classification model of 0.87. Even so, compared with the stepwise regression model for mortality risk, our machine-learned model performed 11 points higher.

CONCLUSIONS

Machine learning algorithms can produce accurate disease classification and prediction models that outperform standard logistic regression models. Such methodology can be employed as potentially more accurate and easily automated ways of identifying at-risk patients for conditions for which good risk prediction methods do not exist or are found to have relatively poor performance. Future studies should attempt to test, automate, and prospectively validate local models, with the aim of reducing the prevalence of undiagnosed diseases and the burden of adverse clinical outcomes related to delays in preventive interventions.

AUTHOR CONTRIBUTIONS

Conception and design: ER, NS, NL
Analysis and interpretation: ER, NS, RD
Data collection: KN, JC, NL
Writing the article: ER
Critical revision of the article: ER, NS, RD, KN, JC, NL
Final approval of the article: ER, NS, RD, KN, JC, NL
Statistical analysis: ER, NS
Obtained funding: ER, RD
Overall responsibility: NL
ER and NS contributed equally to this article and share co-first authorship.

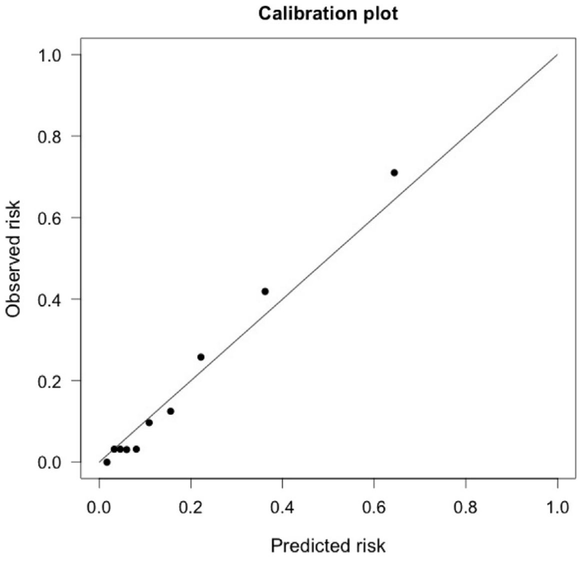
REFERENCES

1. Garraway LA, Verweij J, Ballman KV. Precision oncology: an overview. J Clin Oncol 2013;31:1803-5.
2. Bielinski SJ, Olson JE, Pathak J, Weinshilboum RM, Wang L, Lyke KJ, et al. Preemptive genotyping for personalized medicine: design of the right drug, right dose, right time—using genomic data to individualize treatment protocol. Mayo Clin Proc 2014;89:25-33.
3. Calculate your Ubble age. UK Longevity Explorer (UbbLE); 2016. Available at: http://www.ubble.co.uk/. Accessed February 1, 2016.
4. Ganna A, Ingelsson E. 5 year mortality predictors in 498,103 UK biobank participants: a prospective population-based study. Lancet 2015;386:533-40.
5. Parikh RB, Kakad M, Bates DW. Integrating predictive analytics into high-value care: the dawn of precision delivery. JAMA 2016;315: 651-2.
6. Fowkes FG, Rudan D, Rudan I, Aboyans V, Denenberg JO, McDermott MM, et al. Comparison of global estimates of prevalence and risk factors for peripheral artery disease in 2000 and 2010: a systematic review and analysis. Lancet 2013;382:1329-40.
7. Alberts MJ, Bhatt DL, Mas JL, Ohman EM, Hirsch AT, Röther J, et al. Three-year follow-up and event rates in the international REduction of Atherothrombosis for Continued Health Registry. Eur Heart J 2009;30:2318-26.
8. Hirsch AT, Criqui MH, Treat-Jacobson D, Regensteiner JG, Creager MA, Olin JW, et al. Peripheral arterial disease detection, awareness, and treatment in primary care. JAMA 2001;286:1317-24.
9. Chang P, Nead KT, Olin JW, Cooke JP, Leeper NJ. Clinical and socioeconomic factors associated with unrecognized peripheral artery disease. Vasc Med 2014;19:289-96.
10. Sadrzadeh Rafie AH, Stefanick ML, Sims ST, Phan T, Higgins M, Gabriel A, et al. Sex differences in the prevalence of peripheral artery disease in patients undergoing coronary catheterization. Vasc Med 2010;15:443-50.
11. Nead KT, Cooke JP, Olin JW, Leeper NJ. Alternative ankle-brachial index method identifies additional at-risk individuals. J Am Coll Cardiol 2013;62:553-9.
12. Zou H, Hastie T. Regularization and variable selection via the elastic net. J R Stat Soc Series B Stat Methodol 2005;67:301-20.
13. Breiman L. Random forests. Machine Learning 2001;45:5-32.
14. Kuhn M, Johnson K. Applied predictive modeling. New York: Springer; 2013.
15. R Core Team. R: a language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2015.
16. Steyerberg EW, Vickers AJ, Cook NR, Gerds T, Gonen M, Obuchowski N, et al. Assessing the performance of prediction models: a framework for some traditional and novel measures. Epidemiology 2010;21:128-38.
17. Hosmer DW, Lemeshow S. Goodness of fit tests for the multiple logistic regression model. Commun Stat Theory Methods 1980;9: 1043-69.
18. Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez JC, et al. pROC: an open-source package for R and S+ to analyze and compare roc curves. BMC Bioinformatics 2011;12:77.
19. Pencina MJ, Steyerberg EW, D'Agostino RB. Extensions of net reclassification improvement calculations to measure usefulness of new biomarkers. Stat Med 2011;30:11-21.
20. Lindstrom J, Tuomilehto J. The diabetes risk score: a practical tool to predict type 2 diabetes risk. Diabetes Care 2003;26:725-31.
21. Wilson PW, D'Agostino RB, Levy D, Belanger AM, Silbershatz H, Kannel WB. Prediction of coronary heart disease using risk factor categories. Circulation 1998;97:1837-47.
22. Gail MH, Brinton LA, Byar DP, Corle DK, Green SB, Schairer C, et al. Projecting individualized probabilities of developing breast cancer for white females who are being examined annually. J Natl Cancer Inst 1989;81:1879-86.
23. Ramsay SE, Morris RW, Whincup PH, Papacosta AO, Thomas MC, Wannamethee SG. Prediction of coronary heart disease risk by Framingham and SCORE risk assessments varies by socioeconomic position: results from a study in British men. Eur J Cardiovasc Prev Rehabil 2011;18:186-93.
24. Arts EE, Popa C, Den Broeder AA, Semb AG, Toms T, Kitas GD, et al. Performance of four current risk algorithms in predicting cardiovascular events in patients with early rheumatoid arthritis. Ann Rheum Dis 2015;74:668-74.
25. Tillin T, Hughes AD, Whincup P, Mayet J, Sattar N, McKeigue PM, et al. Ethnicity and prediction of cardiovascular disease: performance of QRISK2 and Framingham scores in a U.K. tri-ethnic prospective cohort study (SABRE—Southall And Brent REvisited). Heart 2014;100:60-7.
26. Murphy TP, Dhangana R, Pencina MJ, Sr D'Agostino RB. Ankle-brachial index and cardiovascular risk prediction: an analysis of 11, 594 individuals with 10-year follow-up. Atherosclerosis 2012;220: 160-7.
27. Liabeuf S, Desjardins L, Diouf M, Temmar M, Renard C, Choukroun G, et al. The addition of vascular calcification scores to traditional risk factors improves cardiovascular risk assessment in patients with chronic kidney disease. PLoS One 2015;10:e0131707.
28. Kadowaki S, Shishido T, Honda Y, Narumi T, Otaki Y, Kinoshita D, et al. Additive clinical value of serum brain-derived neurotrophic factor for prediction of chronic heart failure outcome. Heart Vessels 2016;31: 535-44.
29. Wang TJ. Assessing the role of circulating, genetic, and imaging biomarkers in cardiovascular risk prediction. Circulation 2011;123: 551-65.

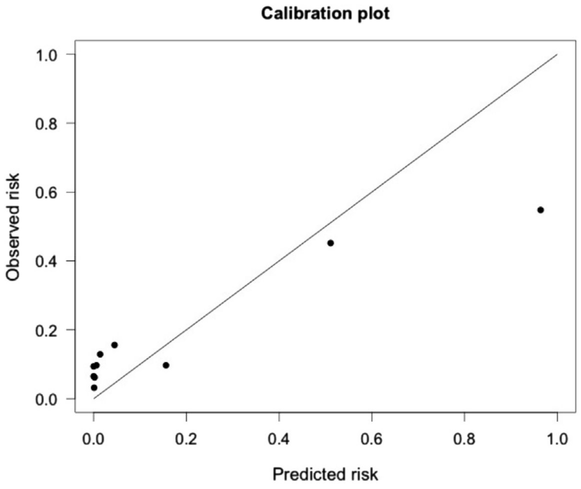
30. Duval S, Massaro JM, Jaff MR, Boden WE, Alberts MJ, Califf RM, et al. An evidence-based score to detect prevalent peripheral artery disease (PAD). Vasc Med 2012;17:342-51.
31. Jonnagaddala J, Liaw ST, Ray P, Kumar M, Chang NW, Dai HJ. Coronary artery disease risk assessment from unstructured electronic health records using text mining. J Biomed Inform 2015;58(Suppl): S203-10.
32. Leeper NJ, Bauer-Mehren A, Iyer SV, Lependu P, Olson C, Shah NH. Practice-based evidence: profiling the safety of cilostazol by text-mining of clinical notes. PLoS One 2013;8:e63499.
33. Finlay GD, Rothman MJ, Smith RA. Measuring the modified early warning score and the Rothman index: advantages of utilizing the electronic medical record in an early warning system. J Hosp Med 2014;9:116-9.
34. Henry KE, Hager DN, Pronovost PJ, Saria S. A targeted real-time early warning score (TREWScore) for septic shock. Sci Transl Med 2015;7: 299ra122.
35. Jung K, Covington S, Sen CK, Januszyk M, Kirsner RS, Gurtner GC, et al. Rapid identification of slow healing wounds. Wound Repair Regen 2016;24:181-8.
36. Weiss JC, Natarajan S, Peissig PL, McCarty CA, Page D. Machine learning for personalized medicine: predicting primary myocardial infarction from electronic health records. AI Magazine 2012;33:33.
37. Jung K, Shah NH. Implications of non-stationarity on predictive modeling using EHRs. J Biomed Inform 2015;58:168-74.
38. Celi LA, Tang RJ, Villarroel MC, Davidzon GA, Lester WT, Chueh HC. A clinical database-driven approach to decision support: predicting mortality among patients with acute kidney injury. J Healthc Eng 2011;2:97-110.
39. Criqui MH, Aboyans V. Epidemiology of peripheral artery disease. Circ Res 2015;116:1509-26.
40. McDermott MM, Kerwin DR, Liu K, Martin GJ, O'Brien E, Kaplan H, et al. Prevalence and significance of unrecognized lower extremity peripheral arterial disease in general medicine practice*. J Gen Intern Med 2001;16:384-90.
41. Nikolsky E, Mehran R, Mintz GS, Dangas GD, Lansky AJ, Aymong ED, et al. Impact of symptomatic peripheral arterial disease on 1-year mortality in patients undergoing percutaneous coronary interventions. J Endovasc Ther 2004;11:60-70.
42. Hildick-Smith DJ, Walsh JT, Lowe MD, Stone DL, Schofield PM, Shapiro LM, et al. Coronary angiography in the presence of peripheral vascular disease: femoral or brachial/radial approach? Catheter Cardiovasc Interv 2000;49:32-7.
43. Olin JW, Sealove BA. Peripheral artery disease: current insight into the disease and its diagnosis and management. Mayo Clin Proc 2010;85: 678-92.
44. Bonaca MP, Creager MA. Pharmacological treatment and current management of peripheral artery disease. Circ Res 2015;116:1579-98.
45. Kent KC, Zwolak RM, Egorova NN, Riles TS, Manganaro A, Moskowitz AJ, et al. Analysis of risk factors for abdominal aortic aneurysm in a cohort of more than 3 million individuals. J Vasc Surg 2010;52:539-48.

Submitted Feb 10, 2016; accepted Apr 4, 2016.

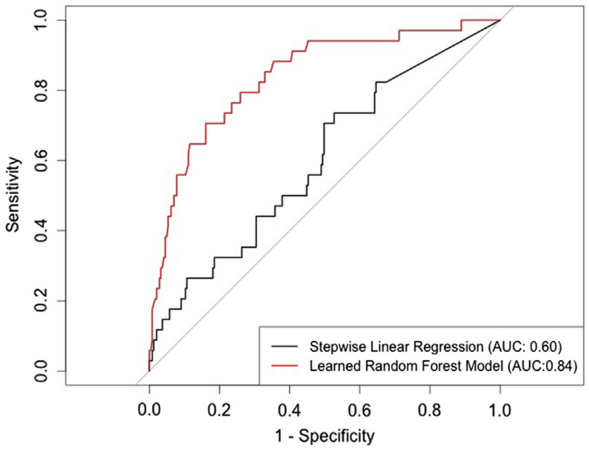
Additional material for this article may be found online at www.jvascsurg.org.



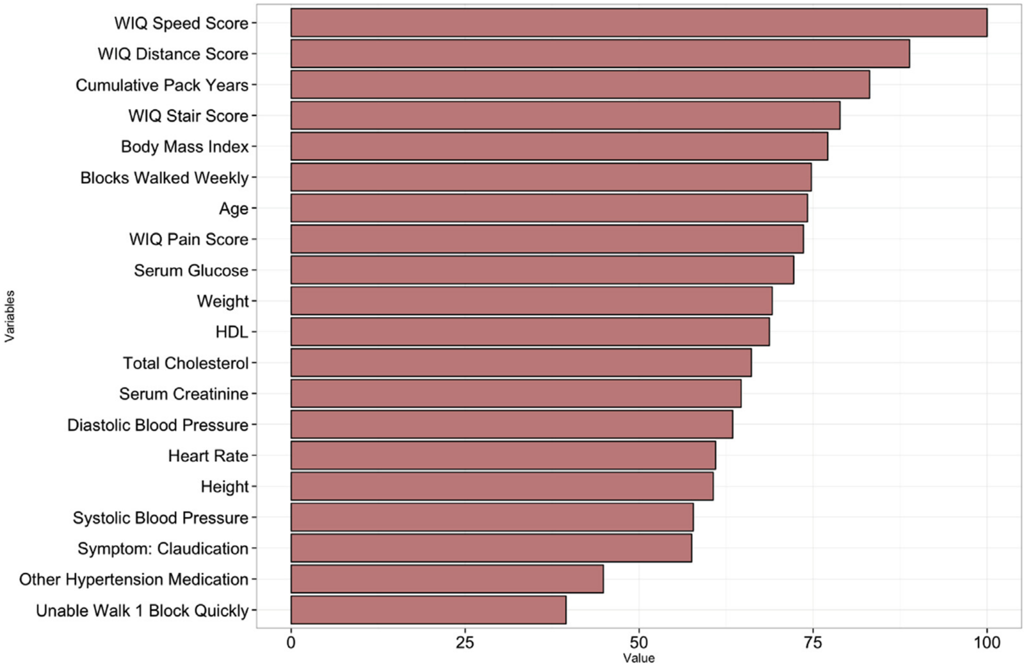
Supplementary Fig 1 (online only). Calibration plot for machine-learned penalized regression model for peripheral artery disease (PAD) classification.



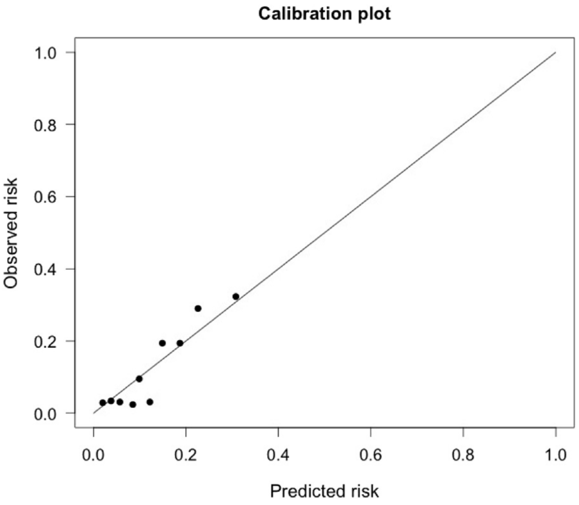
Supplementary Fig 2 (online only). Calibration plot for stepwise logistic regression model for peripheral artery disease (PAD) classification.



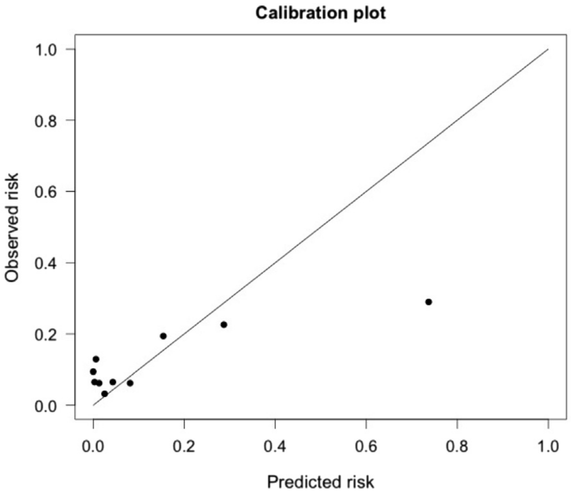
Supplementary Fig 3 (online only). Area under the curve (*AUC*) for stepwise logistic regression and machine-learned random forest model for identification of patients with undiagnosed peripheral artery disease (PAD).



Supplementary Fig 4 (online only). Top 20 weighted variables used in machine-learned random forest model for identification of patients with undiagnosed peripheral artery disease (PAD). *HDL*, High-density lipoprotein; *WIQ*, Walking Impairment Questionnaire.



Supplementary Fig 5 (online only). Calibration plot for machine-learned random forest model for mortality prediction.



Supplementary Fig 6 (online only). Calibration plot for stepwise logistic regression model for mortality prediction.

Supplementary Table I (online only). Comparison of performance[a] of standard stepwise linear regression model to machine-learned penalized regression model for the identification of patients with peripheral artery disease (PAD)

	Machine-learned model[b]	*Stepwise linear regression model*	P *value*
AUC (95% CI)	0.87 (0.81-0.92)	0.76 (0.68-0.84)	.03
Sensitivity	0.76	0.57	
Specificity	0.85	0.9	
Positive predictive value	0.51	0.55	
Negative predictive value	0.94	0.91	

AUC, Area under the curve; *CI,* confidence interval.

[a] Sensitivity and specificity selected on the basis of balance of highest sensitivity for lowest false-positive rate on receiver operating characteristic curve.

[b] Penalized regression model.

Supplementary Table II (online only). Comparison of performance[a] of standard stepwise linear regression model to machine-learned random forest model for the prediction of mortality

	Machine-learned model[b]	*Stepwise linear regression model*	P *value*
AUC (95% CI)	0.76 (0.68-0.84)	0.65 (0.55-0.76)	.1
Sensitivity	0.8	0.6	
Specificity	0.7	0.8	
Positive predictive value	0.25	0.3	
Negative predictive value	0.96	0.93	

AUC, Area under the curve; *CI,* confidence interval.

[a] Sensitivity and specificity selected on the basis of balance of highest sensitivity for lowest false-positive rate on receiver operating characteristic curve.

[b] Random forest model.