Prediction of 1-Year Mortality from Acute Myocardial Infarction Using Machine Learning

Han Cheol Lee, MD, PhD[a], Jin Sup Park, MD, PhD[a,*], Jeong Cheon Choe, MD[a], Jin Hee Ahn, MD, PhD[a], Hye Won Lee, MD, PhD[a], Jun-Hyok Oh, MD, PhD[a], Jung Hyun Choi, MD, PhD[a], Kwang Soo Cha, MD, PhD[a], Taek Jong Hong, MD, PhD[a], and Myung Ho Jeong, MD, PhD[b], the Korea Acute Myocardial Infarction Registry (KAMIR) and Korea Working Group on Myocardial Infarction (KorMI) Investigators

Risk stratification at hospital discharge could be instrumental in guiding postdischarge care. In this study, the risk models for 1-year mortality using machine learning (ML) were evaluated for guiding management of acute myocardial infarction (AMI) patients. From the Korea Acute Myocardial Infarction Registry (KAMIR) dataset, 22,182 AMI patients were selected. The 1-year all-cause mortality was recorded at 12-month follow-up periods. Anomaly detection was conducted for removing outliers; principal component analysis for dimensionality reduction, recursive feature elimination algorithm for feature selection. Model selection and training were conducted with 70% of the dataset after the creation and cross-validation of hundreds of models with decision trees, ensembles, logistic regressions, and deepnets algorithms. The rest of the dataset (30%) was used for comparison between the ML and KAMIR score-based models. The mean age of the AMI patients was 64 years, 71.8% were male, and 56.7% were eventually diagnosed with ST-elevation myocardial infarction. There were 1,332 patients suffering from all-cause mortality (6%) during a median 338 days of follow-up. The ML models for 1-year mortality were well-calibrated (Hosmer-Lemeshow p >0.05) and showed good discrimination (area under the curve for test cohort: 0.918). Compared with the performance of the KAMIR score model, the ML model had a higher area under the curve, net reclassification improvement, and integrated discrimination improvement. The ML model for 1-year mortality was well-calibrated and had excellent discriminatory ability and higher performance. In a comprehensive clinical evaluation process, this model could support risk stratification and management in postdischarge AMI patients. **(Am J Cardiol 2020;133:23−31)**

Clinical practice guidelines endorse routine use of acute myocardial infarction (AMI) mortality risk models to aid decision-making for ischemic heart disease.[1] In this vein, the Global Registry of Acute Coronary Events (GRACE) 6-month postdischarge model[2] (hereafter, GRACE model) and the Korean Acute Myocardial Infarction Registry (KAMIR) score-based risk model[3] (hereafter, KAMIR model) have been developed and validated. However, the GRACE model uses AMI data obtained before the development of a new antiplatelet agent that results in a decreased mortality rate and an increased rate of percutaneous coronary intervention (PCI) for AMI patients[4]; thus, it might be inadequate for current clinical practice. The KAMIR model has good discrimination (c-index 0.82) with more in-hospital predictors such as hyperglycemia and left ventricular dysfunction; however, it is not externally validated.[3] To overcome the disadvantages of current risk models, machine learning (ML) could be a good alternative.[5] This study utilized a multi-center prospective registry of AMI patients to develop and validate a 1-year mortality risk model using ML, and then compared the performance to an existing risk model.

Methods

This article follows the guidelines in the Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis consensus document.[6]

We used data from the KAMIR for 2006 to 2007 and from the Korea Working Group of Myocardial Infarction registry for 2008 to 2013.[7] The 53 participating centers included university and community hospitals with facilities for PCI and on-site cardiac surgery. A trained study coordinator collected data using a standardized case report form

[a]Department of Cardiology and Medical Research Institute, Pusan National University Hospital, Busan, South Korea; and [b]Department of Cardiology, Chonnam National University Hospital, Gwangju, South Korea. Manuscript received May 25, 2020; revised manuscript received and accepted July 13, 2020.

Funding: This work was supported by a clinical research grant from Pusan National University Hospital (MRI-2020) and was also supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF-2017R1D1A1B03033520). The funding agencies had no role in the study design; in the collection, analysis, and interpretation of data; in writing the report; and in the decision to submit the article for publication.

See page 29 for disclosure information.

*Corresponding author: Tel: +82-10-4541-3657; fax: +82-51-247-5875.

E-mail address: dr.jinsup@gmail.com (J.S. Park).

0002-9149/
https://doi.org/10.1016/j.amjcard.2020.07.048

and protocol. The study protocol conformed to the ethical guidelines of the 1975 Declaration of Helsinki, as reflected by prior approval from the human research committee of each participating institution, and the Institutional Review Board of Pusan National University Hospital approved this study (IRB number: 2008064). We obtained informed consent for the use of data from each patient.

From the 31,149 patients with AMI registered in the database, those with clinical events during index hospitalization (n = 999, 3.8%) and follow-up loss at 1-year (n = 7,968, 25.6%), were excluded. Clinical follow-up was conducted at 12 months after index hospitalization for AMI. During follow-up, data from patients, including clinical status, all interventions, outcome events, and adverse events were recorded. Because our aim was to model risk for postdischarge mortality, we excluded those who died during index hospitalization, keeping a final sample of 22,182 participants for analysis (Figure 1).

All clinical, laboratory variables, and clinical events were prospectively collected. Trained study coordinators sourced clinical and laboratory characteristics and outcomes using a standardized case report form and protocol. All-cause mortality was considered cardiac unless a definite noncardiac cause could be established. All events were identified by the patient's physician and confirmed by the principal investigator of each hospital.

Continuous variables were compared using the Student's *t* test or Wilcoxon rank sum test, whereas categorical variables were analyzed using the chi-squared or Fisher's exact test.

Missing data were handled in 2 ways: any feature with >20% of the data missing was eliminated; for <20% missing data, a random forest (RF) was used to impute the missing data. The RF was trained with 10 models, 512 node-threshold, and smart pruning that considered pruning the nodes with <1% of the instances. RF algorithms did noticeably better than the popular k-nearest neighbors imputation method when there was a high correlation.[8]

We performed anomaly detection using an optimized implementation of the Isolation Forest algorithm (sample size: 1,024, forest size: 128, detection rate: 1%, number of cases: 220).[9] The anomaly detection introduced a metric that measured the depth of each instance of the tree and repeated this process hundreds of times to ascertain whether it had been consistently easy to isolate each instance or probably anomalous. Anomaly detection represented this by transforming the average depth into a number between 0 (similar) and 1 (anomalous), considering anomaly scores >0.60 as highly likely anomalous data points. Finally, the 21,962 participants of the dataset remaining following removal of the anomalies were used for risk prediction model development. We split the dataset such that the class frequencies were equal between the training/validation (n = 15,373; 70% [n = 12,299 for training; 80%, n = 3,074 for validation; 20%]) and test (n = 6,589; 30%) datasets.

Feature engineering was conducted in 2 ways: principal component analysis (PCA) and recursive feature elimination (RFE). PCA was used to transform a training dataset to yield uncorrelated features and reduce dimensionality, which reduces supervised model overfitting that leads to

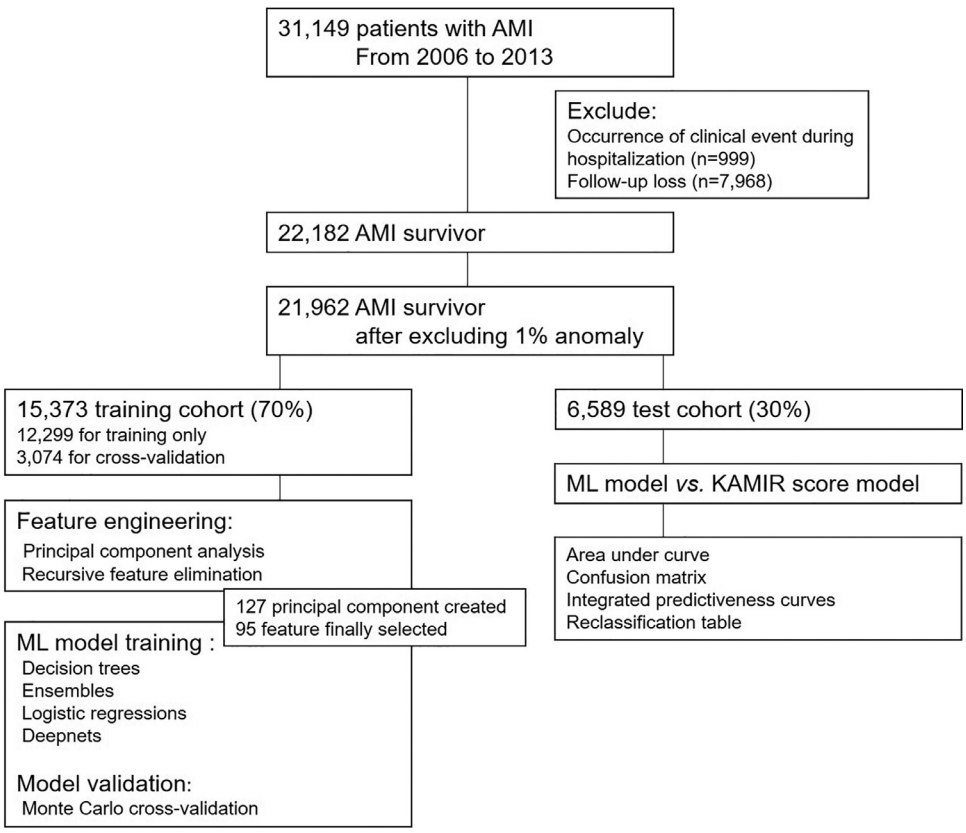

Figure 1. Study flowchart.

poor model performance.[10] It included 77 variables from the original training dataset (n = 12,299), and produced 127 principal components after transformation from a different data type using the factorial analysis of mixed data algorithm.

RFE is a backward selection method. First, we built a linear model for 1-year mortality based on all the variables. After the model calculated variable coefficients and their importance to 1-year mortality, we removed the lowest ranking variables until the RFE had an optimal number of features. Subsequently, 95 features remained for model selection and training.[11]

An automated optimization process provided by BigML was performed for model selection and hyperparametrization to solve classification and regression problems with the training dataset. Different promising parameters were iteratively found through Bayesian parameter optimization based on the sequential model-based algorithm configuration optimization technique.[12] For model selection and training, decision trees, ensembles ($\leq$ 256 trees), logistic regressions (only for classification problems), and deepnets were applied using 95 selected features with 1-year mortality as an outcome. For model validation, Monte Carlo cross-validation was iteratively performed on those models close to the optimum. Finally, 2 bootstrap decision forest models and 1 boosted tree model were selected based on the highest Phi coefficient, and the final ML model was created as a combination of these 3 models.

After the predictive probability for 1-year mortality of the ML model was calculated using batch prediction, discrimination and calibration of the ML model were evaluated in the training and test sets using the c-statistic and Hosmer-Lemeshow goodness-of-fit statistic, respectively. Observed and predicted probabilities of the 1-year mortality were calculated for deciles of risk obtained by applying the ML model to the test data. The KAMIR score, which was developed for 12-month post-AMI mortality, includes the following factors: old age, high Killip class, elevated initial serum creatinine level, elevated initial levels of serum glucose, lower left ventricular ejection fraction, and not undergoing PCI in the hospital.[3] The predictive performance of the ML model and the KAMIR model was compared in several ways. The difference in the area under the curve (AUC) for the ML and KAMIR models (both evaluated in the test dataset) was evaluated using the methods of DeLong et al.[13] We also compared the specificity, positive predictive value, and negative predictive value of the ML and KAMIR models evaluated with the test dataset. Moreover, category-based and category-free (continuous) net reclassification improvement (NRI) indices and integrated discrimination improvement were calculated and compared between the ML and KAMIR models.[14] Finally, integrated predictiveness curves from the application of both models to the test dataset were compared.[15] All the methods from the ML and statistics are summarized in Figure 1.

A p value <0.05 was considered statistically significant. BigML provided all ML processes, including the supervised and unsupervised algorithms. R and Analyse-it validation edition (Analyse-it Software, Ltd., Leeds, United Kingdom) were used to perform the statistical analyses.

Results

Table 1 shows the baseline characteristics of the training cohort. The mean age was 63.3 $\pm$ 12.7 years, 71.6% were male. One-third (36.8%) of the training cohort was experiencing preinfarction angina. The most common past medical conditions were hypertension (50%), current smoker (43%), diabetes (27.7%), prior history of ischemic heart disease (15.5%), and dyslipidemia (11.4%). Almost half of the training cohort was diagnosed with ST-segment elevation AMI (STEMI) (56.7%), and three-quarters (71.5%) presented with Killip class I. Initial thrombolysis therapy was performed on 9.1% of STEMI patients, and the overall rate of those who had undergone PCI was 86.5% (92.9% in the STEMI subgroup, and 79.1% in the non-STEMI subgroup). In patients who underwent invasive coronary procedure, 80.3% received transfemoral approach, 24.6% had 3-vessel disease including left main disease involved, and 43.9% had left anterior descending artery as culprit lesion. A drug-eluting stent was implanted in over 90% of patients, and the procedure success rate was 81.3%.

Overall, 1,332 patients (6%) who were discharged from hospital alive suffered all-cause death during median 338 (25th interquartile, 75th interquartile 50, 400) days of follow-up periods (989 [6.4%] in the training cohort and 343 [5.2%] in the test cohort). The cause of death was adjudicated as cardiovascular in 84.2% of all-cause death, noncardiovascular in 15.8%. Patients who suffered all-cause death were older, more likely to be female, and a similar rate as diagnosed with STEMI, but less likely to have had lower Killip class. They had differences in a burden of past medical conditions, more prior history of ischemic heart disease, diabetes, and hypertension; however, they had less dyslipidemia, fewer current smokers, and less family history of ischemic heart disease. They had less chance of having undergone PCI and were less likely to succeed in the procedure. They were more likely to have had in-hospital complications such as major bleeding, arrhythmic event, new-onset heart failure, and cardiogenic shock during hospitalization. They were less likely to have had postdischarge medications including aspirin, P2Y12 inhibitors, beta-blockers, angiotensin-converting enzyme inhibitors, and statin.

We applied a classifier-dependent feature selection method, as described before, to assess the relation between the models' performance and the number of variables incorporated into them. The model performance plateaued after including 19 variables more. Seventeen variables according to their predictive importance for 1-year mortality could be found, excluding a predictive value for pulse pressure. Among the top predictors across the top 5 classifiers were age, left ventricular ejection fraction during hospitalization, initial Killip class, body mass index (BMI), and initial serum creatinine (Figure 2).

ML model discrimination was excellent (AUC = 0.924 [training cohort]; 0.918 [test cohort]). Compared with the KAMIR model (evaluated in the test cohort), discrimination was improved (AUC [95% CI] = 0.918 [0.902 to 0.935] vs 0.837 [0.817 to 0.857], p <0.001; Figure 3). The model was also well-calibrated in the test cohort (Figure 4).

Table 1
Patients characteristics of training cohort

Variables	Survival status at 12th month		p value
	Alive (n = 14,384) Mean (SD) or number (percentage)	Dead (n = 989) Mean (SD) or number (percentage)	
Age (years)	63±13	73±11	<0.001
Men	10504 (72.6%)	526 (59.6%)	<0.001
Body mass index (kg/m^2)	24±3	23±3	<0.001
Coronary heart disease	2181 (15.1%)	191 (21.7%)	<0.001
Diabetes mellitus	3913 (27.2%)	340 (38.8%)	<0.001
Hypertension	7228 (49.9%)	544 (61.6%)	<0.001
Dyslipidemia[†]	1654 (12.4%)	73 (9.1%)	0.006
Current smoker	6338 (44%)	263 (29.9%)	<0.001
Family history of IHD	1199 (9%)	41 (5.3%)	<0.001
Preinfarct angina pectoris, n (%)	6080 (42.3%)	369 (42%)	0.873
Systolic blood pressure at presentation (mm Hg)	132±25	126±26	<0.001
Heart rate at presentation (beats/min)	78±19	88±25	<0.001
Killip class			<0.001
I	10675 (77.5%)	356 (42%)	
II	1830 (13.3%)	160 (18.9%)	
III	974 (7.1%)	225 (26.6%)	
IV	297 (2.2%)	106 (12.5%)	
ST-segment elevation MI	8295 (57.2%)	501 (56.7%)	0.767
Serum glucose (mg/dl)	166±76	200±102	<0.001
Peak Troponin I (ng/ml)	42.36±138.66	64.37±158.37	<0.001
Serum creatinine (mg/dl)	1.13±1.51	1.77±3.18	<0.001
Serum C-reactive protein (mg/dl)	9.49±53.80	18.90±74.21	<0.001
Low-density lipoprotein cholesterol (mg/dl)	117±41	105±44	<0.001
Ejection fraction (%)	52±14	42±14	<0.001
Initial thrombolysis	765 (5.3%)	25 (2.9%)	<0.001
Percutaneous coronary intervention	12788 (88.3%)	599 (67.8%)	<0.001
3-vessel coronary disease	3486 (26.3%)	311 (45.9%)	<0.001
LAD as Culprit lesion	6381 (48.3%)	338 (50.4%)	0.292
Lesion type B2/C	9443 (78.3%)	513 (84.1%)	0.001
Postprocedural TIMI 3 flow grade	11287 (97.5%)	468 (88.6%)	<0.001
Successful PCI	12051 (94.9%)	508 (82.2%)	<0.001
In-hospital complication			
Major bleeding	28 (0.2%)	7 (0.8%)	<0.001
Arrhythmia	302 (2.1%)	77 (8.7%)	<0.001
Newly-onset heart failure	47 (0.3%)	17 (1.9%)	<0.001
Cardiogenic shock	122 (0.8%)	109 (12.3%)	<0.001
Medications at discharge			
Aspirin	14030 (98.1%)	349 (94.1%)	<0.001
P2Y12 inhibitors	13201 (94.5%)	329 (89.2%)	<0.001
Beta adrenergic blockers	11103 (78%)	260 (70.3%)	<0.001
ACE inhibitors	8935 (63.1%)	210 (56.6%)	0.010
Angiotensin receptor blockers	2971 (21.1%)	85 (23%)	0.380
Spironolactone	1143 (8.2%)	67 (18.2%)	<0.001
Statins	1096 (75.8%)	240 (64.5%)	<0.001

ACE = angiotensin-converting enzyme; IHD = Ischemic heart disease; LAD = left anterior descending; MI = myocardial infarction; PCI = percutaneous coronary intervention; TIMI = thrombolysis in myocardial infarction.

[†] Dyslipidemia is defined as prior history of diagnosis with dyslipidemia or treatment with statin.

Specificity for all-cause death was higher for both the ML and KAMIR models, but sensitivity was significantly higher in the ML model. The proportion of false-positives was similarly low between models, but the proportion of false-negatives was almost double in the KAMIR model (Table 2). The average recall and precision rate for the ML model in the test cohort were 91.35% and 76.57%. Focusing on the all-cause death only, the recall and precision rates decreased by 85.19% and 53.64% at the 41% probability threshold.

Integrated predictiveness curves for the ML and KAMIR models show that for predicted probabilities below the observed all-cause death (5.2%), the ML model consistently yielded lower risk, and for values above the observed all-cause death, it consistently yielded higher risk, reflecting its ability to identify patients at either risk level better (Figure 5).

The ML model significantly improved the prediction of all-cause death when we performed the reclassification table. The ML and KAMIR models had categorical NRI 0.258 (95% CI 0.190 to 0.326, p <0.001), continuous NRI

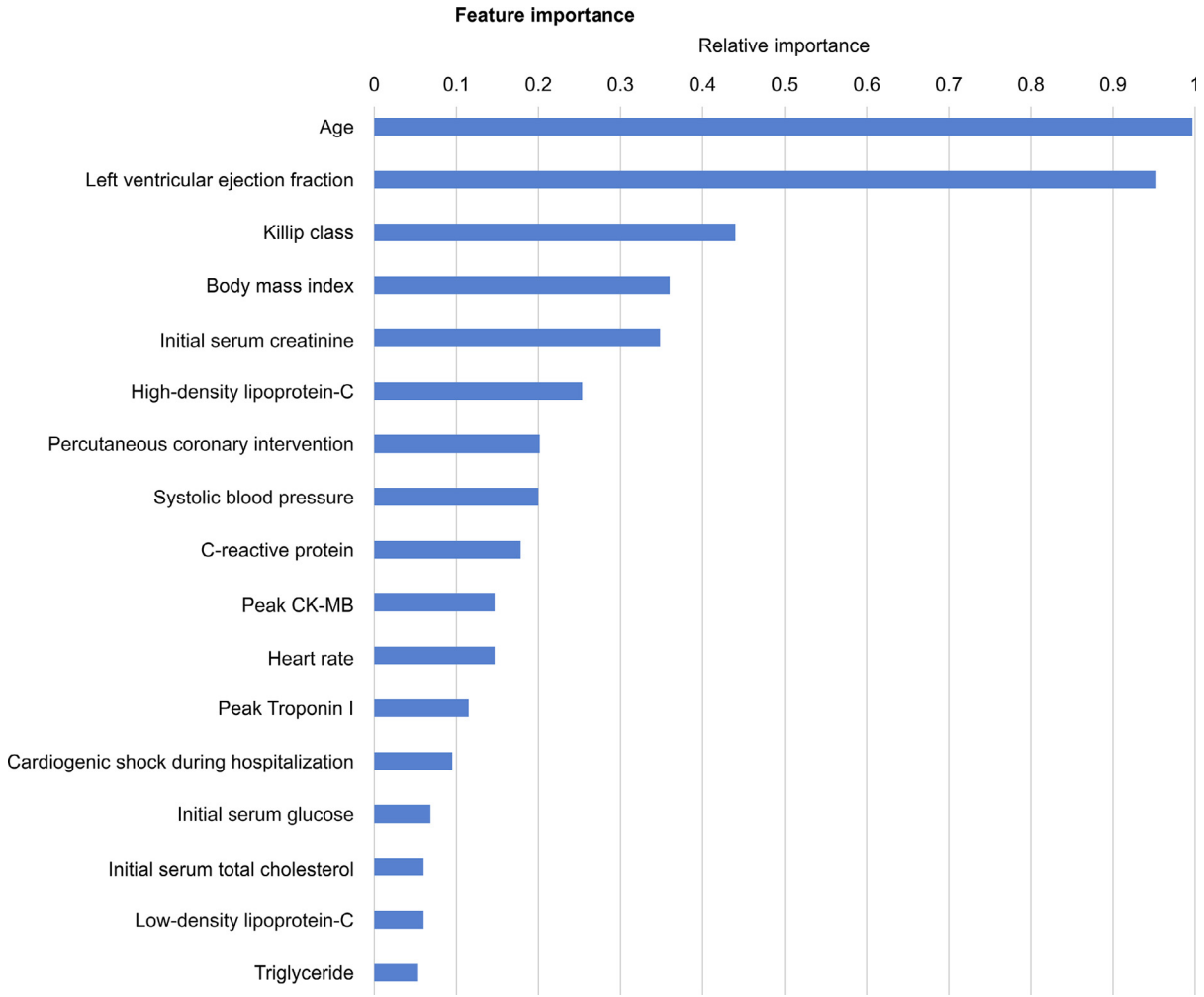

Figure 2. Features of importance based on recursive feature elimination before feature transformations in the train cohort.

0.803 (95% CI 0.699 to 0.907, p <0.001), and integrated discrimination improvement 0.351 (95% CI 0.310 to 0.391, p <0.001; Table 3).

We considered the case of a hypothetical AMI patient with obesity (BMI: 30 kg/m^2), aged 80, who had admission systolic blood pressure of 140 mm Hg, and heart rate of 120 beats/min, Killip class 3, initial serum CK-MB level of 300 mg/dl, troponin I level of 45 pg/L, glucose level of 180 mg/dl, creatinine level of 2.0 mg/dl, some degree of dyslipidemia (total cholesterol: 180 mg/dl, high-density lipoprotein-C 52 mg/dl, low-density lipoprotein-C: 120 mg/dl), and high level of C-reactive protein (8 ug/dl), who underwent PCI. The patient developed cardiogenic shock during hospitalization but recovered. Echocardiography showed a severely reduced ejection fraction of 30%. The probability of 1-year mortality was calculated using the ML model, which predicted that patients with these features will die within 1 year at a 80% probability.

Discussion

Accurate prediction of prognosis is fundamental to patient-centered care. Using data from a nationwide AMI registry, we developed alternative statistical methods to build risk models for 1-year mortality in outpatients with AMI. An ML model comprising a fusion of decision forest and boosted tree models presented proper risk stratification for post-AMI patients with good internal validation and excellent calibration. Furthermore, its performance was significantly better than that of the KAMIR model.[3] These results imply that ML risk models, driven by several important features within the dataset, could outperform risk models developed using conventional statistical methods. Risk prediction using ML could be easily performed based on easily accessible electronic medical records, for which the volume is increasing rapidly.

The GRACE model has been demonstrated to predict mortality for up to 4 years with good accuracy and extensive external validations.[2] However, it has several limitations compared with the KAMIR model. Cohorts enrolled from the GRACE model had a lower PCI rate and the use of P2Y12 inhibitors was approximately 30%. The GRACE model did not include the following risk factors in its development: admission hyperglycemia, presence of stroke or peripheral artery disease, and left ventricular systolic function, which showed significant associations with mortality

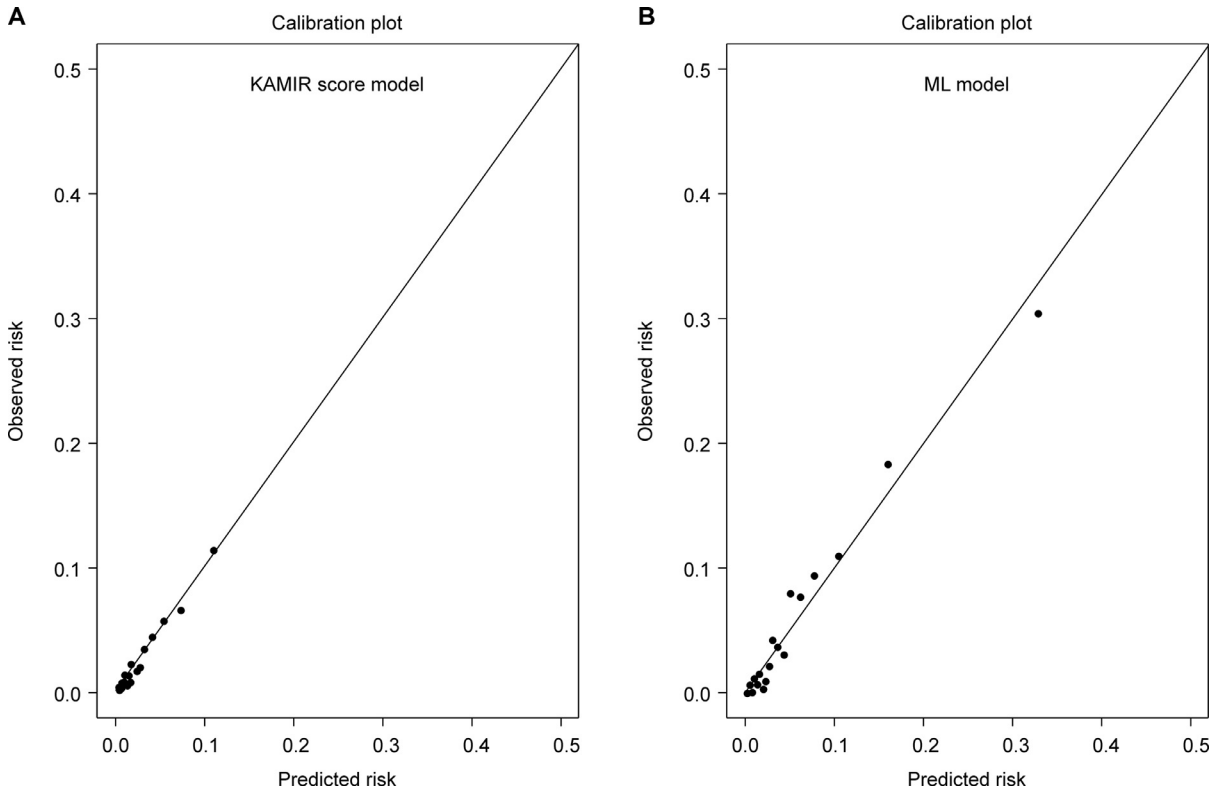

Figure 3. Calibration plot for the KAMIR score-based risk model (*A*) and machine learning risk model (*B*) in the test cohort.

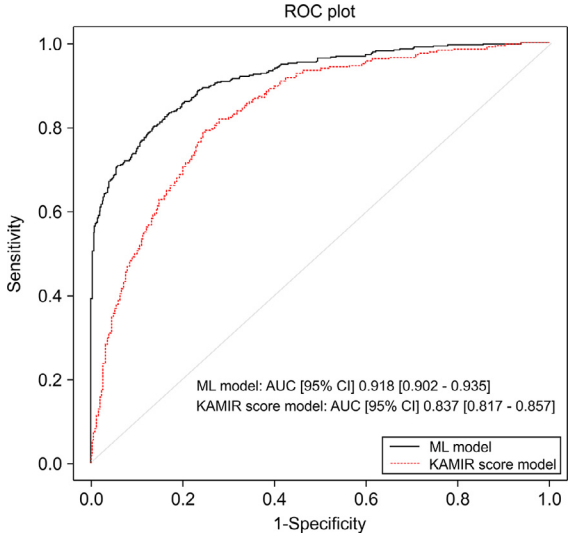

Figure 4. Comparison of receiver operating characteristic curve and area under curve between the KAMIR score-based risk model *(dotted line)* and the machine learning risk model *(solid line)* in the test cohort.

Table 2
Confusion matrix

ML risk model				Threshold: 0.41
	Death	Survival	Actual	Recall
Death	184	159	343	53.64%
Survival	32	6214	6246	99.49%
Predicted	216	6373	6589	
Precision	85.19%	97.51%		
KAMIR score model				Threshold: 0.5
	Death	Survival	ACTUAL	RECALL
Death	16	327	343	4.66%
Survival	31	6214	6246	99.28%
Predicted	47	6542	6589	
Precision	34.04%	94.99%		

in AMI patients.[16,17] Furthermore, the KAMIR model demonstrated significant differences in more predictive accuracy than the GRACE model in post-AMI patients.[3] Comparison of the performance between the models in the test cohort showed that the ML model was better in accuracy, sensitivity, and reclassification of 1-year mortality.

RF and support vector machine algorithms have typically outperformed logistic regression, decision tree, and neural networks in classification problems on large health datasets.[18] This could enable the creation of nonlinear and conditional relations that might be developed by the logistic regression and decision tree models. The neural network exhibited lower performance than the models created by other algorithms. Massive-sized training data might be required to train the neural network properly compared with RF models. In addition to the algorithm, the overall performance of the ML methods in classification problems can be improved by dimensionality reduction using feature engineering.[19] PCA is widely used in feature engineering to recognize the peak disparity characteristics in the applied dataset, thus reducing it to a smaller number of significant features.[20] The RFE method, which removes the weakest features until the desired number of features is reached, is widely used in disease prediction using ML.[21] PCA and RFE were performed in the training cohort, and a fusion of

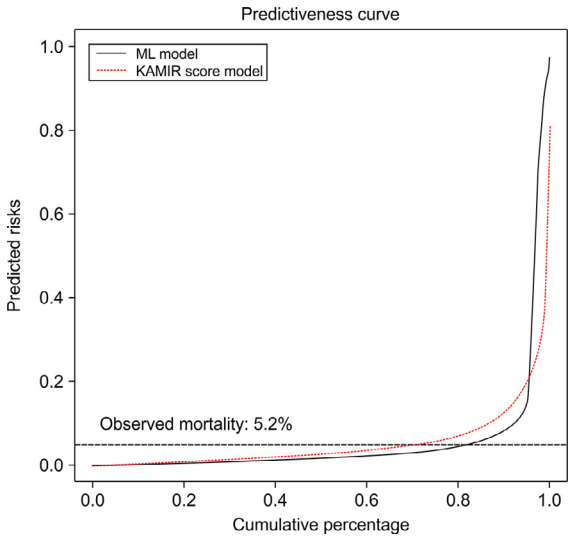

Figure 5. Comparison of predictiveness curves between the KAMIR score-based risk model *(dotted line)* and the machine learning risk model *(solid line)* in the test cohort.

Table 3
Reclassification table

KAMIR model	ML model		
	<5%	5−10%	>10%
Survival (n = 6,246)			
<5%	4521 (72.3%)	86 (1.3%)	58 (0.9%)
5−10%	640 (10.2%)	218 (3.4%)	40 (0.6%)
>10%	154 (2.4%)	330 (5.2%)	199 (3.2%)
Death (n = 343)			
<5%	39 (11.4%)	3 (0.8%)	29 (8.4%)
5−10%	20 (5.8%)	15 (4.4%)	57 (16.6%)
>10%	7 (2.0%)	25 (7.3%)	148 (43.1%)

2 decision forest and 1 tree algorithms was used to develop the ML risk model.

The features of importance provided by the ML algorithm differed somewhat compared with logistic regression in conventional statistical methods. However, it could provide suggestions that give direction. Older age, higher creatinine levels, and Killip class, and lower systolic blood pressure derived from the ML algorithm are all risk factors that could have a substantially detrimental effect for the prognosis of post-AMI patients with the statistical method.[22] Left ventricular systolic function, which ranked as second grade, is a significant predictor of long-term mortality after AMI.[23] Initial glucose levels were also listed, supporting the association between hyperglycemia and early mortality in AMI patients.[24] Renal dysfunction is independently associated with increased risk of death in post-AMI patients,[25] which reflected the level of serum creatinine as the selected feature by the algorithm. Additional high-ranking variables, confirming previous studies, included low-density lipoprotein levels,[26] high-density lipoproteins,[27] BMI,[28] and C-reactive protein.[29] All variables in the KAMIR model ranked as important features; this could be reliable and used to identify critical prognostic factors along with conventional statistical methods.

This study had several limitations. First, the study cohort was not derived from random study data, which could lead to bias in the measurement of outcomes. Approximately 25% of the registered patients were excluded owing to follow-up loss at 1 year. However, the registry data collected prospectively represented real-world data, hence conveying more recent clinical practice. The collecting of registry data began >15 years ago; invariably, there have been changes in the management of AMI patients since then. However, KAMIR data have shown a tendency to reduce the mortality from the start. Second, we focused on 1-year mortality only. Post-AMI patients experience many complications related to AMI, such as recurrent myocardial infarction, repeated coronary artery angiography, decreased quality of life due to depression, and post-MI heart failure. A future model should include these outcomes for better prediction and risk stratification in a post-MI survivor. Third, the current prospective registry differs in aspects such as the low prevalence of dyslipidemia, high rate of invasive treatment, and lower mortality rate compared with the western registries. It is difficult to generalize and adapt the ML model for prediction of outcome in the rest of the population, and it needs external validation. Finally, the models have not been externally validated. However, the use of iterated cross-validation inside the training cohort, then diagnosis with the test cohort increases the results' strength and reduces the risk of overfitting (i.e., an overoptimistic estimation of the model's performance).

Disclosures

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

Authors' Contribution

Han Cheol Lee, Jin Sup Park: Conceptualization, Methodology.

Jeong Cheon Choe, Jin Hee Ahn, Hye Won Lee, Jun-Hyok Oh, Jung Hyun Choi, Kwang Soo Cha, Taek Jong Hong, Myung Ho Jeong: Data collection, Cleansing.

Han Cheol Lee: Writing - Original draft preparation.

Jin Sup Park: Visualization, Investigation, Supervision, Validation, Writing - Reviewing and Editing.

Appendix

*Complete list of the included variables

Age, sex, body mass index, symptom-to-door time, preinfarct angina pectoris, systolic blood pressure, diastolic blood pressure, pulse pressure, heart rate, Killip class, coronary heart disease, hypertension, diabetes, dyslipidemia, current smoker, family history of heart disease, thrombolysis, percutaneous coronary intervention, cath room to balloon time, diseased vessel, culprit lesion, lesion type, pre-TIMI flow, post-TIMI flow, intervention result, complication during hospitalization, left ventricular ejection fraction, serum glucose, serum creatinine, serum CK-MB, serum TnI, total cholesterol, triglyceride, high-density lipoprotein, low-density lipoprotein cholesterol, C-reactive protein,

discharge medications (aspirin, clopidogrel, cilostazol, ticlopidine, triflusal, calcium channel blocker, beta blocker, angiotensin-converting enzyme inhibitor, angiotensin receptor blocker, nitrate, nicorandil, diuretics, spiro0lactone, statin, fibrate, eztrole, vytorin, amiodarone, antidepressant, digoxin, insulin, oral anticoagulation, oral hypoglycemic agent, sedatives, final diagnosis.

Acknowledgment

The authors acknowledge the contribution of the Korea Acute Myocardial Infarction Registry (KAMIR) and Korea Working Group on Myocardial Infarction (KorMI) investigators[#] to this study.

[#]Investigators in Korea Acute Myocardial Infarction Registry (KAMIR) and Korea Working Group on Myocardial Infarction (KorMI):

Myung Ho Jeong, MD; Young Keun Ahn, MD; Sung Chull Chae, MD; Jong Hyun Kim, MD; Seung Ho Hur, MD; Yong Jo Kim, MD; In Whan Seong, MD; Dong Hoon Choi, MD; Jei Keon Chae, MD; Taek Jong Hong, MD; Jae Young Rhew, MD; Doo Il Kim, MD; In Ho Chae, MD; Jung Han Yoon, MD; Bon Kwon Koo, MD; Byung Ok Kim, MD; Myung Young Lee, MD; Kee Sik Kim, MD; Jin Youg Hwang, MD; Myeong Chan Cho, MD; Seok Kyu Oh, MD; Nae Hee Lee, MD; Kyung Tae Jeong, MD; Seung Jea Tahk, MD; Jang Ho Bae, MD; Seung Woon Rha, MD; Keum Soo Park, MD; Chong Jin Kim, MD; Kyoo Rok Ha, MD; Tae Hoon Ahn, MD; Moo Hyun Kim, MD; Ki Bae Seung, MD; Wook Sung Chung, MD; Ju Young Yang, MD; Chong Yun Rhim, MD; Hyeon Cheol Gwon, MD; Seong Wook Park, MD; Young Youp Koh, MD; Seung Jae Joo, MD; Soo Joong Kim, MD; Dong Kyu Jin, MD; Jin Man Cho, MD; Byung Ok Kim, MD; Sang-Wook Kim, MD; Jeong Kyung Kim, MD; Tae Ik Kim, MD; Dung young Nah, MD; Si Hoon Park, MD; Sang Hyun Lee, MD; Seung Uk Lee, MD; Hang-Jae Chung, MD; Jang Hyun Cho, MD; Seung Won Jin, MD; Yang Soo Jang, MD; Jeong Gwan Cho, MD, and Seung Jung Park, MD.

1. Amsterdam EA, Wenger NK, Brindis RG, Casey DE, Ganiats TG, Holmes DR, Jaffe AS, Jneid H, Kelly RF, Kontos MC, Levine GN, Liebson PR, Mukherjee D, Peterson ED, Sabatine MS, Smalling RW, Zieman SJ. 2014 AHA/ACC guideline for the management of patients with non-ST-elevation acute coronary syndromes: executive summary. *J Am Coll Cardiol* 2014;64:2645–2687.
2. Eagle KA, Lim MJ, Dabbous OH, Pieper KS, Goldberg RJ, Van de Werf F, Goodman SG, Granger CB, Steg PG, Gore JM, Budaj A, Avezum Á, Flather MD, Fox KAA. A validated prediction model for all forms of acute coronary syndrome: estimating the risk of 6-month postdischarge death in an international registry. *JAMA* 2004;291:2727–2733.
3. Kim HK, Jeong MH, Ahn Y, Kim JH, Chae SC, Kim YJ, Hur SH, Seong IW, Hong TJ, Choi DH, Cho MC. Hospital discharge risk score system for the assessment of clinical outcomes in patients with acute myocardial infarction (Korea Acute Myocardial Infarction Registry [KAMIR] score). *Am J Cardiol* 2011;107:965–971. e961.
4. Roe MT, Messenger JC, Weintraub WS, Cannon CP, Fonarow GC, Dai D, Chen AY, Klein LW, Masoudi FA, McKay C, Hewitt K, Brindis RG, Peterson ED, Rumsfeld JS. Treatments, trends, and outcomes of acute myocardial infarction and percutaneous coronary intervention. *J Am Coll Cardiol* 2010;56:254–263.
5. Motwani M, Dey D, Berman DS, Germano G, Achenbach S, Al-Mallah MH, Andreini D, Budoff MJ, Cademartiri F, Callister TQ, Chang HJ. 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.
6. Moons KGM, Altman DG, Reitsma JB, Ioannidis JPA, Macaskill P, Steyerberg EW, Vickers AJ, Ransohoff DF, Collins GS. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): explanation and elaboration. *Ann Intern Med* 2015;162:W1–W73.
7. Park JS, Cha KS, Shin D, Lee DS, Lee HW, Oh JH, Choi JH, Lee HC, Hong TJ, Lee SH, Kim JS. Prognostic significance of presenting blood pressure in patients with ST-elevation myocardial infarction undergoing percutaneous coronary intervention. *Am J Hypertens* 2014;28:797–805.
8. Tang F, Ishwaran H. Random forest missing data algorithms. *Stat Anal Data Min* 2017;10:363–377.
9. Liu FT, Ting KM, Zhou Z-H. Isolation forest. In: 2008 Eighth IEEE International Conference on Data Mining; 2008.
10. Ringnér M. What is principal component analysis? *Nat Biotechnol* 2008;26:303–304.
11. Guyon I, Weston J, Barnhill S, Vapnik VJML. Gene selection for cancer classification using support vector machines. *Mach Learn* 2002;46:389–422.
12. Hutter F, Hoos H, Leyton-Brown K. *An Evaluation of Sequential Model-Based Optimization for Expensive Blackbox Functions*. Amsterdam, The Netherlands: Association for Computing Machinery; 2013.
13. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics* 1988;44:837–845.
14. Pencina MJ, D'Agostino RB Sr, D'Agostino RB Jr, Vasan RS. Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond. *Stat Med* 2008;27:157–172. discussion 207-112.
15. Pepe MS, Feng Z, Huang Y, Longton G, Prentice R, Thompson IM, Zheng Y. Integrating the predictiveness of a marker with its performance as a classifier. *Am J Epidemiol* 2008;167:362–368.
16. Kontos MC, Jamal S, Tatum JL, Ornato JP, Jesse RL. Predictive power of systolic function and congestive heart failure in patients with patients admitted for chest pain without ST elevation in the troponin era. *Am Heart J* 2008;156:329–335.
17. Bhadriraju S, Ray KK, DeFranco AC, Barber K, Bhadriraju P, Murphy SA, Morrow DA, McCabe CH, Gibson CM, Cannon CP, Braunwald E. Association between blood glucose and long-term mortality in patients with acute coronary syndromes in the OPUS-TIMI 16 trial. *Am J Cardiol* 2006;97:1573–1577.
18. Oermann EK, Rubinsteyn A, Ding D, Mascitelli J, Starke RM, Bederson JB, Kano H, Lunsford LD, Sheehan JP, Hammerbacher J, Kondziolka D. Using a machine learning approach to predict outcomes after radiosurgery for cerebral arteriovenous malformations. *Sci Rep* 2016;6:21161.
19. Karapinar Senturk Z. Early diagnosis of Parkinson's disease using machine learning algorithms. *Med Hypotheses* 2020;138:109603.
20. Rajaguru H, Chakravarthy SRS. Analysis of decision tree and k-nearest neighbor algorithm in the classification of breast cancer. *Asian Pac J Cancer Prev* 2019;20:3777–3781.
21. Yang X. Identification of risk genes associated with myocardial infarction based on the recursive feature elimination algorithm and support vector machine classifier. *Mol Med Rep* 2018;17:1555–1560.
22. Cheng JM, Helming AM, van Vark LC, Kardys I, Den Uil CA, Jewbali LS, van Geuns RJ, Zijlstra F, van Domburg RT, Boersma E, Akkerhuis KM. A simple risk chart for initial risk assessment of 30-day mortality in patients with cardiogenic shock from ST-elevation myocardial infarction. *Eur Heart J Acute Cardiovascular Care* 2016;5:101–107.
23. Kober L, Torp-Pedersen C, Pedersen OD, Hoiberg S, Camm AJ. Importance of congestive heart failure and interaction of congestive heart failure and left ventricular systolic function on prognosis in patients with acute myocardial infarction. *Am J Cardiol* 1996;78:1124–1128.
24. Zhao CJ, Hao ZX, Liu R, Liu Y. Admission glucose and risk of early death in non-diabetic patients with ST-segment elevation myocardial infarction: a meta-analysis. *Med Sci Monit* 2015;21:1387–1394.
25. Anavekar NS, McMurray JJ, Velazquez EJ, Solomon SD, Kober L, Rouleau JL, White HD, Nordlander R, Maggioni A, Dickstein K, Zelenkofske S. Relation between renal dysfunction and cardiovascular outcomes after myocardial infarction. *N Engl J Med* 2004;351:1285–1295.

26. Oduncu V, Erkol A, Kurt M, Anboğa İH, Karabay CY, Şengül C, Bulut M, Özveren O, Fotbolcu H, Akgün T, Türkyılmaz E. The prognostic value of very low admission LDL-cholesterol levels in ST-segment elevation myocardial infarction compared in statin-pretreated and statin-naive patients undergoing primary percutaneous coronary intervention. *Int J Cardiol* 2013;167:458–463.
27. Martin SS, Khokhar AA, May HT, May HT, Kulkarni KR, Blaha MJ, Joshi PH, Toth PP, Muhlestein JB, Anderson JL, Knight S, Li Y. HDL cholesterol subclasses, myocardial infarction, and mortality in secondary prevention: The Lipoprotein Investigators Collaborative. *Eur Heart J* 2015;36:22–30.
28. Bucholz EM, Rathore SS, Reid KJ, Jones PG, Chan PS, Rich MW, Spertus JA, Krumholz HM. Body mass index and mortality in acute myocardial infarction patients. *Am J Med* 2012;125:796–803.
29. Mani P, Puri R, Schwartz GG, Nissen SE, Shao M, Kastelein JJ, Menon V, Lincoff AM, Nicholls SJ. Association of initial and serial c-reactive protein levels with adverse cardiovascular events and death after acute coronary syndrome: a secondary analysis of the VISTA-16 Trial. *JAMA Cardiol* 2019;4:314–320.