



OPEN

Machine learning and deep learning to predict mortality in patients with spontaneous coronary artery dissection

Chayakrit Krittanawong^{1,2}✉, Hafeez Ul Hassan Virk³, Anirudh Kumar⁴, Mehmet Aydar⁵, Zhen Wang^{6,7}, Matthew P. Stewart^{8,9} & Jonathan L. Halperin²

Machine learning (ML) and deep learning (DL) can successfully predict high prevalence events in very large databases (big data), but the value of this methodology for risk prediction in smaller cohorts with uncommon diseases and infrequent events is uncertain. The clinical course of spontaneous coronary artery dissection (SCAD) is variable, and no reliable methods are available to predict mortality. Based on the hypothesis that machine learning (ML) and deep learning (DL) techniques could enhance the identification of patients at risk, we applied a deep neural network to information available in electronic health records (EHR) to predict in-hospital mortality in patients with SCAD. We extracted patient data from the EHR of an extensive urban health system and applied several ML and DL models using candidate clinical variables potentially associated with mortality. We partitioned the data into training and evaluation sets with cross-validation. We estimated model performance based on the area under the receiver-operator characteristics curve (AUC) and balanced accuracy. As sensitivity analyses, we examined results limited to cases with complete clinical information available. We identified 375 SCAD patients of which mortality during the index hospitalization was 11.5%. The best-performing DL algorithm identified in-hospital mortality with AUC 0.98 (95% CI 0.97–0.99), compared to other ML models ($P < 0.0001$). For prediction of mortality using ML models in patients with SCAD, the AUC ranged from 0.50 with the random forest method (95% CI 0.41–0.58) to 0.95 with the AdaBoost model (95% CI 0.93–0.96), with intermediate performance using logistic regression, decision tree, support vector machine, K-nearest neighbors, and extreme gradient boosting methods. A deep neural network model was associated with higher predictive accuracy and discriminative power than logistic regression or ML models for identification of patients with ACS due to SCAD prone to early mortality.

Machine learning (ML), a branch of artificial intelligence (AI), is applicable to risk modeling in medicine. Deep learning (DL) is a form of ML typically implemented through multi-layered neural networks to interpret and classify complex datasets and enhance clinical decision-making. When applied to big data in medicine, ML and DL, particularly convolutional neural networks (ResNet, GoogleNet, or VCG families) are well suited to clinical image recognition and estimation of prognosis in large datasets^{1,2}. Information about the performance of this technology in smaller datasets, such as rare diseases or patient cohorts with heterogeneous conditions, infrequent events, or comorbidities that cause competing risks is considerably more limited. ML may perform poorly at predicting low frequency events, such as those with an incidence $< 10\%$, and scant data are available on the performance of DL in heterogeneous populations with a low frequency of events³. Spontaneous coronary artery

¹Section of Cardiology, Baylor College of Medicine, 1 Baylor Plaza, Houston, TX 77030, USA. ²Icahn School of Medicine at Mount Sinai, The the Zena and Michael A. Wiener Cardiovascular Institute, Mount Sinai Heart, New York, NY, USA. ³Department of Cardiovascular Diseases, Case Western Reserve University, University Hospitals Cleveland Medical Center, Cleveland, OH, USA. ⁴Heart and Vascular Institute, Cleveland Clinic, Cleveland, OH, USA. ⁵Department of Computer Science, Kent State University, Kent, OH, USA. ⁶Robert D. and Patricia E. Kern Center for the Science of Health Care Delivery, Mayo Clinic, Rochester, MN, USA. ⁷Division of Health Care Policy and Research, Department of Health Sciences Research, Mayo Clinic, Rochester, MN, USA. ⁸The Institute of Applied and Computational Sciences, Harvard University, Boston, MA, USA. ⁹School of Engineering and Applied Sciences, Harvard University, Boston, MA, USA. ✉email: Chayakrit.Krittanawong@bcm.edu

dissection (SCAD) is a heterogeneous condition (high noise) and uncommon cause of acute coronary syndrome (ACS) associated with low mortality (infrequent events)^{4,5}. Although there have been other ML-based approaches for predicting mortality risk in patients with ACS, predictors of early mortality in patients with SCAD have not been identified. We hypothesized that ML and particularly DL models could predict in-hospital mortality in patients with ACS due to SCAD, based on information extracted from electronic health records (EHR), with greater accuracy than conventional risk classification methods^{6,7}. Accordingly, we compared the performance of conventional logistic regression, ML modeling, and custom-built DL models to predict mortality in patients with SCAD using data from the EHR of a large urban health system.

Methods

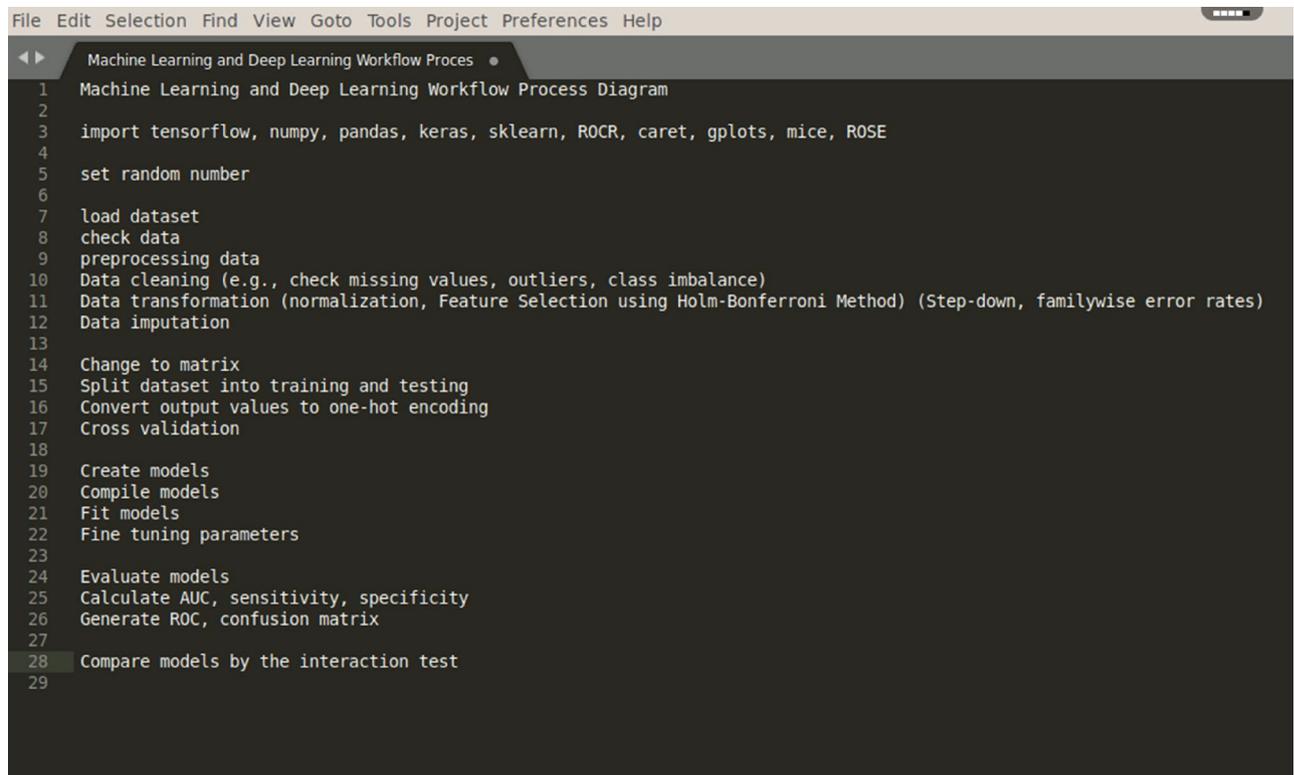
Study population. We identified patients with a principal diagnosis of SCAD by querying the entire Mount Sinai Health System EHR for the period from January 1, 2008, to December 31, 2018, using International Classification of Diseases (ICD) 9 (414.12) and 10 codes (I25.42) for SCAD diagnosis, including only those with procedural (CPT) codes for coronary angiography and/or percutaneous coronary intervention, and excluding those with diagnoses indicating iatrogenic coronary dissection, perforation, or laceration. We excluded patients who have the following: (1) missing critical demographic information (i.e., age); (2) missing data for mortality; (3) age < 18; and/or (4) they had concomitant iatrogenic puncture or laceration of the coronary vessels. All coded co-morbidities were accumulated. The protocol was approved by the Institutional Review Board governing research involving human subjects at the Icahn School of Medicine at Mount Sinai. The MSHS Ethics Committee approved a waiver of documentation of informed consent; de-identified patient data was obtained from the MSHS Data Warehouse (<https://msdw.mountsinai.org/>).

Baseline variables and feature selection. Feature selection to select potential variables for SCAD patients based on 2 steps. First, we conducted stepwise backward regression on all variables using the Holm-Bonferroni Method (Step-down, familywise error rates) and we chose variables for our full model when variables were $P < 0.0004$ ⁸. Second, we identified potential variables using clinical judgment based on previously published findings. From over 400 variables, the candidate features present on admission were selected to develop the prediction models; these demographic, clinical characteristic, comorbidity, medication, vital sign and laboratory value items are listed in Online Supplementary Table 1. The primary outcome was in-hospital mortality.

Machine learning analysis. Algorithms towards diagnosis or forecasting (prognosis) of an event were based on supervised learning to predict mortality using preprocessed data and several ML and DL approaches. The ML algorithms were logistic regression, support vector machine (SVM), decision tree, random forest, K-nearest neighbors, AdaBoost and extreme gradient boosting. The DL model employed a deep neural network running Python version 3.6 (Keras) with Tensorflow backend in the high performance computing (HPC) clusters⁹. Statistical properties of continuous variables (e.g., laboratory measurements) were summarized using histograms or kernel density estimation when necessary. Log-transformations were used to normalize the underlying distribution of variables. Variables with > 10% missing data were excluded and remaining missing data were addressed through imputation techniques on an individual variable basis, using R version 3.3.1 (MICE and missForest packages)¹⁰. Event imbalance was addressed by random over-sampling¹¹. We randomly partitioned the data and repeated multiple times.

Sensitivity analyses were performed by comparing various data partitions, missing imputation vs ignoring missing data (LightGBM) and by data augmentation. In these analyses, we also examined whether results changed when limited to cases with complete unimputed data or treating event imbalance with ROSE package software. Machine-learning algorithms gain functionality from variables in the training dataset. The histogram for each clinical characteristic was normalized and analyzed separately for relationship by linear regression. Hyper-parameters, a specific learned function, was randomly tuned using several values for each parameter to derive optimum values (online Supplementary Table 2). A random forest algorithm, for example, has hyper-parameters specifying the number of branches and maximum width of each branch corresponding to the number of interactions considered in the model, whereas for the neural network, hyper-parameters control for the complexity of the model, size of the network, and how network connections are activated, or “learned”. For other prediction methods, logistic regression models did not prove viable. Hyper-parameters were tuned by cross-validated performance to minimize overfitting, searching the grid by sampling incremental combinations of multiple variables for the model, and including a parameter to control for dropping from the model variables that did not contribute to minimizing loss of functionality to identify the optimal parameters¹². Once optimized based on the training dataset, performance of the final predictive model was assessed using the evaluation dataset. Each of several ML and DL models using tenfold cross-validation was used to estimate performance, minimize biases, and optimize hyperparameters. A summary of the workflow process for ML and DL is presented in Fig. 1. All methods were carried out in accordance with relevant guidelines and regulations. Given this particular dataset and heterogeneity in nature, we used oversampling techniques to control for dataset imbalance as well as reported balance accuracy.

Deep learning model. The full model architecture explanation with mathematical can be found in the onlinesupplementaryMethod. The neural network architecture for this binary classification ($x = [x_1, x_2, \dots, x_n]^T$ adding weight and bias) by binary cross-entropy $[-y \log(p) + (1 - y) \log(1 - p)]$ consists of 15 regular fully-connected layers (using ReLU activation), two dropout layers, one after the second and third fully-connected layers, and a binary output layer (Softmax)^{13,14}. (Fig. 2). The output of the k th neuron in a given layer can be written as



```

File Edit Selection Find View Goto Tools Project Preferences Help
Machine Learning and Deep Learning Workflow Proces
1 Machine Learning and Deep Learning Workflow Process Diagram
2
3 import tensorflow, numpy, pandas, keras, sklearn, ROCR, caret, gplots, mice, ROSE
4
5 set random number
6
7 load dataset
8 check data
9 preprocessing data
10 Data cleaning (e.g., check missing values, outliers, class imbalance)
11 Data transformation (normalization, Feature Selection using Holm-Bonferroni Method) (Step-down, familywise error rates)
12 Data imputation
13
14 Change to matrix
15 Split dataset into training and testing
16 Convert output values to one-hot encoding
17 Cross validation
18
19 Create models
20 Compile models
21 Fit models
22 Fine tuning parameters
23
24 Evaluate models
25 Calculate AUC, sensitivity, specificity
26 Generate ROC, confusion matrix
27
28 Compare models by the interaction test
29

```

Figure 1. Machine learning and deep learning workflow process diagram.

$y_k = f\left(\sum_{j=1}^n w_{kj}x_j + b_j\right)$ for a layer with n inputs and hence weights w_1 through w_n . Because the number of patients with SCAD was small and the mortality rate was low, a plethora of hidden layers may impair performance due to overfitting. The model was tested using an Adam optimizer with a learning rate of 0.01¹⁵. To optimize the model performance, the model was fine-tuned using grid search hyperparameter selection and optimally trained at 1,000 epochs¹⁶. Sensitivity analyses were performed using grid search for each hyperparameter selection, different data partitions, and different value of the class label. To minimize biases and optimize hyperparameters, we employed the nested cross-validation to fine-tune the model.

Statistical analysis. Model performance was assessed from the area under the receiver-operator characteristic curve (AUC) for accuracy and adjusted for event imbalance using balancing statistics. The bootstrap technique was used to estimate confidence intervals. Performance assessments were performed using Caret, Scikit-learn and Keras software (R and Python, respectively). We compared models' performance based on AUC values using the interaction test¹⁷.

Results

Of 30,425 patients with acute coronary syndromes identified in the EHR survey, 375 (1.2%) had a diagnosis of SCAD. Overall, the mean age was 52.2 ± 12.8 years and 64.3% were women. Table 1 summarized selected baseline clinical characteristics for SCAD patients. Among these, 43 patients died during the index hospitalization (mortality 11.5%). Based on feature selections and regression analysis, predictors of in-hospital mortality in SCAD patients include elevated c-reactive protein, atrial fibrillation, hypertension, and steroid use. The best-performing DL models predicted in-hospital mortality with AUC 0.98 (95% CI 0.97–0.99) with mean accuracy 97%, balanced accuracy 98%, sensitivity 98%, and specificity 96%, compared to other ML models or logistic regressions ($P < 0.0001$). Table 2 summarizes all model performances. The AdaBoost method yielded an AUC of 0.95 (95% CI 0.93–0.96), mean accuracy 94%, balanced accuracy 61%, sensitivity 25%, and specificity 97%, compared to logistic regression model ($P < 0.0001$). The AUC with the support vector machine method was 0.92 (95% CI 0.89–0.94), mean accuracy 93%, balanced accuracy 60%, sensitivity 25%, and specificity 96%, compared to logistic regression model ($P < 0.0001$). The K-nearest neighbors method generated an AUC of 0.91 (95% CI 0.88–0.93), and had a mean accuracy of 89%, balanced accuracy 50%, sensitivity 74%, and specificity 97%, compared to logistic regression model ($P < 0.0001$). Extreme gradient boosting resulted in an AUC of 0.90 (95% CI 0.86–0.93), mean accuracy 95%, balanced accuracy 54%, sensitivity 83%, and specificity 99%, compared to logistic regression model ($P < 0.0001$). The decision tree model had an AUC of 0.78 (95% CI 0.72–0.83), mean accuracy of 79%, balanced accuracy 53%, sensitivity 87%, and specificity 35%, compared to logistic regression model ($P < 0.0001$). The conventional logistic regression model was associated with an AUC of 0.59 (95% CI 0.51–0.67), mean accuracy 91%, balanced accuracy 59%, sensitivity 25%, and specificity 94%, compared to

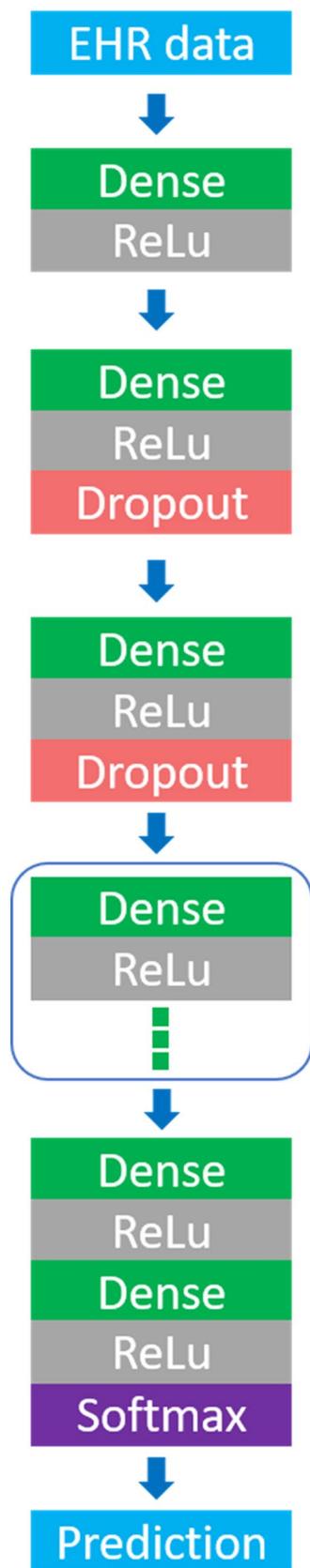


Figure 2. The architecture of the network.

Clinical characteristics	SCAD (N= 375)	Missing data (%)
Age (mean \pm SD)	52.2 \pm 12.8	0
Female	64.3%	0
White	44%	0
African-American	6.9%	0
Hispanic	4.8%	0
Asian	3.2%	0
Others/Unknown	36.3%	0
Hypertension	54.7%	0
History of smoking	5.6%	0
Overweight	2.1%	0
Secondary hypertension	0.53%	0
Malignant hypertension	1.8%	0
Chronic obstructive pulmonary disease	8.8%	0
Ischemic stroke	5.6%	0
Intracranial hemorrhage	2.1%	0
Peripheral artery disease	26.9%	0
Carotid artery disease	8.5%	0
Pulmonary hypertension	13.3%	0
Atrial fibrillation	19.7%	0
Chronic kidney disease	7.5%	0
FMD	1.9%	0
Type 1 diabetes	2.4%	0
Hyperthyroid	1.1%	0
Hypothyroid	6.4%	0
Anxiety	1.1%	0
Depression	8.5%	0
Emotional stress	5.1%	0
Migraine	1.6%	0
Rheumatoid arthritis	0.8%	0
Hypertrophic cardiomyopathy	0.5%	0
Ventricular arrhythmia	31%	0
Cardiac arrest	34.5%	0
Vital Signs		
Body mass index (kg/m ²),	23.4 \pm 7.9	1.50
Systolic blood pressure (mmhg)	130.6 \pm 16.2	1.65
Diastolic blood pressure (mmhg)	71.2 \pm 9.0	7.13
Heart rate	72.3 \pm 10.2	3.60
Body temperature	78.2 \pm 21.9	1.97
Respiratory rate	18.6 \pm 2.8	2.13
Oxygen saturation	95.8 \pm 3.8	7.68
Lab values		
Sodium	138.7 \pm 2.6	6.56
Potassium	4.1 \pm 0.4	6.61
Magnesium	2.1 \pm 0.3	7.40
Chloride	103.9 \pm 4.0	6.94
Blood urea nitrogen	20.5 \pm 10.3	6.54
Creatinine	1.29 \pm 0.6	8.62
Cholesterol	82.3 \pm 31.7	6.82
HDL	43.1 \pm 13.2	6.68
White blood cell count	8.5 \pm 4.2	6.58
Hemoglobin	12.3 \pm 1.8	6.21
C-reactive protein	15.5 \pm 7.9	8.30
Platelet count	219.5 \pm 62.7	6.54
Albumin	3.6 \pm 0.5	7.10
Vaccination		
Influenza vaccination	12.8%	7.60
Continued		

Clinical characteristics	SCAD (N=375)	Missing data (%)
Medications		
NSAID use	1.1%	0
Steroid use	2.4%	0
Cannabis use disorder	1.1%	0

Table 1. Selected baseline clinical characteristics of patients with SCAD. NSAIDs nonsteroidal anti-inflammatory drugs.

Model	Accuracy	Balanced accuracy	AUC (95% CI)
Deep learning	0.97	0.98	0.98 (0.97–0.99)
AdaBoost	0.94	0.61	0.95 (0.93–0.96)
Support vector machine	0.93	0.60	0.92 (0.89–0.94)
K-nearest neighbors	0.89	0.50	0.91 (0.88–0.93)
Extreme Gradient Boosting	0.95	0.54	0.90 (0.86–0.93)
Decision tree	0.79	0.53	0.78 (0.72–0.83)
Logistic regression	0.91	0.59	0.59 (0.51–0.67)
Random forest	0.93	0.52	0.50 (0.41–0.58)

Table 2. Comparison among machine learning and deep learning algorithms.

logistic regression model ($P < 0.0001$). The random forest ML model had an AUC of 0.50 (95% CI 0.41–0.58), mean accuracy 93%, balanced accuracy 52%, sensitivity 25%, and specificity 96%. The random forest ML model had no statistical difference from logistic regression model. Table 3 summarizes all statistical comparison among DL and ML models.

Discussion

There are three main conclusions from the present study. First, elevated c-reactive protein, atrial fibrillation, hypertension and steroid use are important predictors of SCAD mortality. SCAD is a unique and heterogeneous condition. Numerous studies suggested that SCAD is usually not associated with atherosclerosis or traditional cardiovascular risk factors (e.g., dyslipidemia, type 2 diabetes), but may be associated with connective tissue disease, autoimmune disease, stimulants, intense emotional stress, or intense physical exertion.

All the ML approaches, except the random forest model, outperformed conventional logistic regression models in predicting mortality during the index hospitalization for patients with ACS due to SCAD. Furthermore, the custom-built DL models outperformed both logistic regression and ML methods for predicting in-hospital mortality in patients with SCAD. Compared to either ML or DL, conventional logistic regression performed poorly at identifying predictors of early mortality in the population we examined. Even the relatively rudimentary decision tree ML approach, which had limited predictive power, had better predictive capacity. These observations are consistent with previous studies establishing that ML algorithms typically outperform regression models^{18–20}. Deep neural network models also generally perform better than regression, reflecting the mathematical complexity and non-linearity of medical diagnosis and prognosis that defy simple parametric methodologies. In well-established diseases with causative agents, regression models can be used to estimate the effect of an independent variable on a dependent outcome (e.g., ACS directly predicts ischemic cardiomyopathy) and may be better than ML or DL methods when relationships are linear. In contrast, heterogeneous conditions with obscure predictors such as those linking SCAD to mortality require nonlinear analytic methodology that pools a large number of multidimensional variables to identify predictors of an infrequent outcome. That may explain how DL and ML models outperform regression models in select circumstances.

In this investigation, DL outperformed both regression and ML models. Although several ML models were more robust than regression models, a low-bias ML model such as boosting, which is based on methodology designed to minimize bias, may be subject to overfitting when applied to small amounts of heterogeneous data (high variance). This may explain why DL performed better than boosting. The DL model perhaps used multidimensional variables (matrix multiplication) including weight and bias, capturing a greater proportion of interactions between variables than kernel and regularization penalties in SVM, improving its performance compared to SVM. The SVMs must tune relatively fewer parameters, while DL requires multiple parameter selections, entailing complexity when applied to more than 400 clinical variables. Pathophysiological between disease and mortality are non-linear, particularly in heterogeneous conditions with low frequency events^{21, 22}. This may be why a deep neural network using matrix multiplication for complex variable interactions outperformed all the ML models in the present study. To date, attempts to explain the iteration of stochastic gradient descent and cosine loss have yielded no reliable mathematical explanation as to how DL unravels such complex variable interactions^{23, 24}.

Model	AUC (95% CI)	Comparison	AUC (95% CI)	P-value
Deep learning	0.98 (0.97–0.99)	Support vector machine	0.92 (0.89–0.94)	< 0.0001
		K-nearest neighbors	0.91 (0.88–0.93)	< 0.0001
		Decision tree	0.78 (0.72–0.83)	< 0.0001
		Random forest	0.50 (0.41–0.58)	< 0.0001
		AdaBoost	0.95 (0.93–0.96)	0.001
		Extreme Gradient Boosting	0.90 (0.86–0.93)	< 0.0001
		Logistic regression	0.59 (0.51–0.67)	< 0.0001
AdaBoost	0.95 (0.93–0.96)	Extreme Gradient Boosting	0.90 (0.86–0.93)	0.01
		Logistic regression	0.59 (0.51–0.67)	< 0.0001
		Deep learning	0.98 (0.97–0.99)	0.0011
Support vector machine	0.92 (0.89–0.94)	K-nearest neighbors	0.91 (0.88–0.93)	0.58
		Decision tree	0.78 (0.72–0.83)	< 0.0001
		Random forest	0.50 (0.41–0.58)	< 0.0001
		AdaBoost	0.95 (0.93–0.96)	0.04
		Extreme Gradient Boosting	0.90 (0.86–0.93)	0.36
		Logistic regression	0.59 (0.51–0.67)	< 0.0001
		Deep learning	0.98 (0.97–0.99)	< 0.0001
K-nearest neighbors	0.91 (0.88–0.93)	Decision tree	0.78 (0.72–0.83)	< 0.0001
		Random forest	0.50 (0.41–0.58)	< 0.0001
		AdaBoost	0.95 (0.93–0.96)	0.01
		Extreme Gradient Boosting	0.90 (0.86–0.93)	0.65
		Logistic regression	0.59 (0.51–0.67)	< 0.0001
Extreme gradient boosting	0.90 (0.86–0.93)	Logistic regression	0.59 (0.51–0.67)	< 0.0001
		Deep learning	0.98 (0.97–0.99)	< 0.0001
		Decision tree	0.78 (0.72–0.83)	< 0.0001
Decision tree	0.78 (0.72–0.83)	Random forest	0.50 (0.41–0.58)	< 0.0001
		AdaBoost	0.95 (0.93–0.96)	< 0.0001
		Extreme Gradient Boosting	0.90 (0.86–0.93)	< 0.0001
		Logistic regression	0.59 (0.51–0.67)	< 0.0001
Logistic regression	0.59 (0.51–0.67)	Deep learning	0.98 (0.97–0.99)	< 0.0001
				< 0.0001
Random forest	0.50 (0.41–0.58)	AdaBoost	0.95 (0.93–0.96)	< 0.0001
		Extreme Gradient Boosting	0.90 (0.86–0.93)	< 0.0001
		Logistic regression	0.59 (0.51–0.67)	0.13
		Deep learning	0.98 (0.97–0.99)	< 0.0001

Table 3. Comparison among machine learning and deep learning algorithms.

In general, boosting models adjust for more parameters and are more suitable for objective function than random forests. In addition, random forests may cause biases related to different number of levels or correlated features of similar relevance. We found that boosting models (e.g., XGBoost) could potentially result in better performance than random forests perhaps due to optimal hyperparameter selection (aka hyperparameter tuning) and minimal noise.

We undertook this study as a proof-of-concept. The final DL model exhibited higher discrimination, better calibration, and greater classification accuracy than either logistic regression or ML models for predicting early mortality in patients with SCAD. We experimented with several DL models, including transfer learning, and found that the best results with the ReLU activation function, which outperformed the tanh and sigmoid activation functions. This is likely due to non-saturation of gradient, the inherent non-linearity, a reduction likelihood of vanishing gradient, and sparsity effects. Dropout layers on the upstream section of the deep learning may also help²⁵. Why ReLU exhibited better convergence performance than the tanh and sigmoid activation functions is less clear²⁵. These observations are consistent with earlier studies in which DL models outperformed ML algorithms when applied to other disease states^{26–29}. This suggests that DL may be better suited to non-linear, low frequency outcomes such as SCAD mortality or recurrent SCAD, because it benefits from multiparametric adjustment and successive model-fitting. Since SCAD is an uncommon, heterogeneous, and poorly understood clinical entity, DL modeling uses repeated model-fitting to discern patterns that were not exposed by the other staged analysis methods.

This study has several limitations. First, ML and DL methodologies depend upon mathematical relationships between variables (e.g., variable selection algorithms), rather than medical knowledge and biological plausibility.

Although we used clinical judgment to select variables and filter the algorithms, the DL decision process—the so-called “black box”—cannot be directly observed. While ML algorithms do not always yield information about effect size like the hazard ratio derived from Cox regression analysis, the contribution of individual variables can be determined and indicate signals in the data that are not causal associations and that should be interpreted cautiously pending validation in prospective epidemiological studies. Discrepancies between clinically and mathematically selected variables are complex, but DL methods such as quantum neural networks using systematic randomization of weights in each neuron rather than regularization/dropout/early stopping methods may be able to open the “black box” and shed light on these relationships.

The completeness, quality and consistency of data contribute to the success of ML algorithms. Missing value imputation may have biased the prediction algorithms, but there is at present no consensus on statistical standards to reflect this. Data quality is to some extent subjective and standards for assessment or regulation of quality control are lacking. Omission of potential confounders including localization of SCAD, severity of the coronary artery disease, proportion of intracoronary imaging, therapeutic management or proportion of PCI as well as life-style variables such as coffee consumption, sleep hygiene, emotional state or exercise could contribute to overfitting when applied to a SCAD cohort. Moreover, the temporal information of the predictors is a limitation. For instance, although we averaged lab values in laboratory values (e.g., troponin, BNP, NT-proBNP) for each subject, this may introduce biases for SCAD patients who did not present with an acute coronary syndrome or heart failure. Support for inclusion of several of the variables selected for the analysis we conducted, such as connective tissue disease and hormonal therapy, is inconsistent in the medical literature.

Although we employed pretrained models, hyper-parameter selection, and normalization, most of the models were prone to overfitting. Despite grid-searching, we sampled combinations of variables based on incremental significance, including hyper-parameters to control for dropping variables from the model that did not contribute to functionality. A key limitation of this proof of concept analysis is the small number of patients who developed the main outcome event of in-hospital mortality during the study period. Moreover, given a small number of SCAD patients, data splitting reduces the information available for development and possibly renders validation impotent. The model was trained for a particular period when survival after ACS was more frequent than demise. More training data, ultimately requiring a larger clinical sample, is needed to reduce the impact of these factors and others that detract from predictive power.

Given the rarity of the condition, diagnosis of SCAD patients is very challenging. Our findings require validation using additional clinical datasets from multiple health systems to assure generalizability to other populations with SCAD. Furthermore, while DL has been applied to other small clinical datasets^{24, 30, 31}, its performance in cohorts with a small number of outcome events, such as the SCAD population we studied for in-hospital mortality, may not be generalizable to other disease states. Adoption of ML methodology in clinical practice requires multiple formal replication and validation steps, given the host of factors affecting variability of data (e.g., laboratory collection, data cleaning) and models (e.g., hyperparameter selection). In the future, blockchain-encrypted ML models could be shared among institutions or EHR systems to validate the models using the same environmental controls³².

Hyperparameter selection is prone to confounding, and the performance of each algorithm varies depending on specific variables of the databases and parameters employed³³. There is currently no consensus around standards for reporting or interpreting ML or DL studies, limiting comparative analysis. In addition, confounders may arise in neural networks, and further methodological advances such as theoretical quantum neural network and systematic randomization within each neuron may prove more applicable than the traditional regularization methods currently employed³⁴.

Overall, this is a proposed proof-of-concept work of ML and DL models for SCAD patients. Further studies are needed to validate the models and results in different populations.

Conclusions

Elevated c-reactive protein, atrial fibrillation, hypertension and steroid use may be used as predictors of SCAD mortality. Although in this analysis a DL model was more predictive and discriminative than ML methods and logistic regression models to identify patients with heterogeneous clinical features such as SCAD, several limitations involving mathematical modeling, data structure and clinical integration must be addressed before these tools can be applied in clinical practice. Deep neural network models seem most promising for development to this purpose, but further methodological enhancements are needed to leverage data and develop valid predictors of early mortality in patients with SCAD. Data from prospective studies and randomized trials would greatly facilitate this effort to forecast clinical outcomes.

Received: 30 September 2020; Accepted: 23 March 2021

Published online: 26 April 2021

References

1. Gulshan, V. *et al.* Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA* **316**, 2402–2410 (2016).
2. Esteva, A. *et al.* Dermatologist-level classification of skin cancer with deep neural networks. *Nature* **542**, 115 (2017).
3. McKnight, L. K., Wilcox, A., Hripcsak, G. The effect of sample size and disease prevalence on supervised machine learning of narrative data. In *Proceedings AMIA Symposium* 519–22 (2002).
4. Krittanawong, C. *et al.* Conditions and factors associated with spontaneous coronary artery dissection (from a National Population-Based Cohort Study). *Am. J. Cardiol.* **123**, 249–253 (2019).

5. Krittanawong, C. *et al.* Trends in Incidence, characteristics, and in-hospital outcomes of patients presenting with spontaneous coronary artery dissection (From a National Population-Based Cohort Study Between 2004 and 2015). *Am. J. Cardiol.* **122**, 1617–1623 (2018).
6. Krittanawong, C., Zhang, H., Wang, Z., Aydar, M. & Kitai, T. Artificial intelligence in precision cardiovascular medicine. *J. Am. Coll. Cardiol.* **69**, 2657–2664 (2017).
7. Krittanawong, C. *et al.* Deep learning for cardiovascular medicine: a practical primer. *Eur. Heart J.* **40**, 2058–2073 (2019).
8. Holm, S. A simple sequentially rejective multiple test procedure. *Scand. J. Stat.* **5**, 65–70 (1979).
9. Kovatch, P., Costa, A., Giles, Z., Fluder, E., Cho, H. M., & Mazurkova, S. Big omics data experience. In *Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis: ACM*, 39 (2015).
10. Sv, B. & Groothuis-Oudshoorn, K. mice: Multivariate imputation by chained equations in R. *J. Stat. Softw.* **50**, 1–68 (2010).
11. Japkowicz, N. & Stephen, S. The class imbalance problem: a systematic study. *Intell. Data Anal.* **6**, 429–449 (2002).
12. Luo, G. A review of automatic selection methods for machine learning algorithms and hyper-parameter values. *Netw. Model. Anal. Health Inf. Bioinf.* **5**, 18 (2016).
13. Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I. & Salakhutdinov, R. Dropout: a simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.* **15**, 1929–1958 (2014).
14. Nair, V., & Hinton, G. E. Rectified linear units improve restricted boltzmann machines. In *Proceedings of the 27th international conference on machine learning (ICML-10)*, 807–814 (2010).
15. Kingma, D. P., & Ba, J. Adam: a method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014).
16. Ndiaye, E., Le, T., Fercoq, O., Salmon, J., & Takeuchi, I. Safe Grid Search with Optimal Complexity. arXiv preprint arXiv:181005471 (2018).
17. Altman, D. G. & Bland, J. M. Interaction revisited: the difference between two estimates. *BMJ (Clinical research ed)* **326**, 219 (2003).
18. Huang, H. H., Xu, T. & Yang, J. Comparing logistic regression, support vector machines, and permanent classification methods in predicting hypertension. *BMC Proc.* **8**, S96 (2014).
19. Manoochehri, Z. *et al.* Comparison of support vector machine based on genetic algorithm with logistic regression to diagnose obstructive sleep apnea. *J. Res. Med. Sci.* **23**, 65 (2018).
20. Liu, L. *et al.* An interpretable boosting model to predict side effects of analgesics for osteoarthritis. *BMC Syst. Biol.* **12**, 105 (2018).
21. Fuller, J. H., Shipley, M. J., Rose, G., Jarrett, R. J. & Keen, H. Mortality from coronary heart disease and stroke in relation to degree of glycaemia: the Whitehall study. *Br. Med. J.* **287**, 867–870 (1983).
22. Fa-Binefa, M. *et al.* Early smoking-onset age and risk of cardiovascular disease and mortality. *Prev. Med.* **124**, 17–22 (2019).
23. Tishby, N., & Zaslavsky, N. Deep learning and the information bottleneck principle. In *2015 IEEE Information Theory Workshop (ITW)*: IEEE, 2015:1–5.
24. Barz, B., & Denzler, J. Deep Learning on Small Datasets without Pre-Training using Cosine Loss. arXiv preprint arXiv:190109054 (2019).
25. Krizhevsky, A., Sutskever, I. & Hinton, G. E. Imagenet classification with deep convolutional neural networks. *Adv. Neural Inf. Process. Syst.* **67**, 1097–1105 (2012).
26. Cai, Y. *et al.* Concussion classification via deep learning using whole-brain white matter fiber strains. *PLoS ONE* **13**, e0197992 (2018).
27. Sonobe, T. *et al.* Comparison between support vector machine and deep learning, machine-learning technologies for detecting epiretinal membrane using 3D-OCT. *Int. Ophthalmol.* **39**, 1871–1877 (2019).
28. Younghak, S. & Balasingham, I. Comparison of hand-craft feature based SVM and CNN based deep learning framework for automatic polyp classification. *Conf. Proc.* **2017**, 3277–3280 (2017).
29. Rds, A. S. Deep learning based skin lesion segmentation and classification of melanoma using support vector machine (SVM). In *Asian Pacific journal of cancer prevention: APJCP* 2019;20:1555–1561.
30. Feng, S., Zhou, H. & Dong, H. Using deep neural network with small dataset to predict material defects. *Mater. Des.* **162**, 300–310 (2019).
31. He, G., & Lung, C. T. Imaging sign classification through deep learning on small data. arXiv preprint arXiv:190300183 (2019).
32. Krittanawong, C. R. A. *et al.* Integrating Blockchain technology with artificial intelligence for cardiovascular medicine. *Nat. Rev. Cardiol.* **5**, 69 (2010).
33. Statnikov, A., & Aliferis, C. F. Are random forests better than support vector machines for microarray-based cancer classification? In *AMIA Annual Symposium proceedings AMIA Symposium* 686–90 (2007).
34. Farhi, E., & Neven, H. Classification with quantum neural networks on near term processors. arXiv preprint arXiv:180206002 (2018).

Author contributions

C.K. wrote the main manuscript text. A.K. and J.L.H. edited the manuscript. H.H. prepared tables and figures. C.K. wrote new R and python codes, built machine learning and deep learning models. M.A. and M.P.S. tested and assisted with model architecture and mathematical computation. Z.W. performed statistical analyses and comparisons.

Funding

There was no funding for this work.

Competing interests

Dr. Krittanawong discloses the following relationships – Journal of the American Heart Association (Editorial board); European Heart Journal Digital Health (Editorial board); The Journal of Scientific Innovation in Medicine (Associate Editor); The Lancet Digital Health (Advisory board); JACC Asia (Section Editor); The American College of Cardiology Solution Set Oversight Committee (Committee); The ACCF/AHA Task Force on Performance Measures (Committee). None of the other authors have any disclosures.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-021-88172-0>.

Correspondence and requests for materials should be addressed to C.K.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021