MEDINFO 2017: Precision Healthcare through Informatics
A.V. Gundlapalli et al. (Eds.)

doi:10.3233/978-1-61499-830-3-476

Using Machine Learning Models to Predict In-Hospital Mortality for ST-Elevation Myocardial Infarction Patients

Xiang Li[a], Haifeng Liu[a], Jingang Yang[b], Guotong Xie[a], Meilin Xu[c], Yuejin Yang[b]

[a] *IBM Research – China, Beijing, China*
[b] *Department of Cardiology, Fuwai Hospital, National Center for Cardiovascular Diseases, Beijing, China*
[c] *Pfizer Investment Co. Ltd., Beijing, China*

Abstract

Acute myocardial infarction is a major cause of hospitalization and mortality in China, where ST-elevation myocardial infarction (STEMI) is more severe and has a higher mortality rate. Accurate and interpretable prediction of in-hospital mortality is critical for STEMI patient clinical decision making. In this study, we used interpretable machine learning approaches to build in-hospital mortality prediction models for STEMI patients from Chinese Acute Myocardial Infarction (CAMI) registry data. We first performed cohort construction and feature engineering on CAMI data to generate an available dataset and identify potential predictors. Then several supervised learning methods with good interpretability, including generalized linear models, decision tree models, and Bayes models, were applied to build prediction models. The experimental results show that our models achieve higher prediction performance (AUC = 0.80~0.85) than the previous in-hospital mortality prediction STEMI models and are also easily interpretable for clinical decision support.

Keywords:

Myocardial Infarction; Hospital Mortality; Machine Learning

Introduction

Acute myocardial infarction (AMI) is a major cause of hospitalization and mortality in China, where the in-hospital mortality rate for ST-elevation myocardial infarction (STEMI) is even higher than that for non-ST-elevation myocardial infarction (NSTEMI) [1,2]. Because there is considerable variability in mortality risk among patients with STEMI, it is critical to accurately predict the risks of in-hospital mortality for STEMI patients at the time of hospital presentation, in order to decide on the allocation of clinical resources and the choice of interventional and medical therapies [3,4].

Current in-hospital mortality risk models for STEMI, such as TIMI score [3] and GRACE score [4], use predictors that are ground in previous known evidence, including age, heart rate, systolic blood pressure (SBP), Killip levels, weight, history of hypertension, diabetes and angina, anterior STEMI, time to treatment, serum creatinine, and cardiac arrest. These risk scores were derived from logistic regression models, which are well understood and easy to apply in clinical practice. However, the performance of in-hospital mortality prediction for STEMI still has room for improvement. Besides the predictors used in the existing models, there are other potential predictors that are highly related to STEMI in-hospital mortality, which can also be used in risk prediction to improve the performance. Besides, some other machine learning models may achieve better prediction performance or interpretability than regression analysis models like logistic regression.

Therefore, this study investigated the modeling of in-hospital mortality prediction models in STEMI that have good prediction ability and interpretability using machine learning methods. The study was based on the Chinese Acute Myocardial Infarction (CAMI) [2] data, which collected the patients' demographics, symptoms, medical history, results of physical examination and laboratory test, details of in-hospital treatments, and clinical events including mortality. Because the dataset is heterogeneous and redundant, dimensionality reduction methods and model learning algorithms that can handle redundant feature sets should be used for predictive modeling. Moreover, the objective of this study was to build human understandable and applicable risk prediction models. Though many previous works used feature engineering and supervised learning methods to build high accuracy risk prediction models for cardiovascular diseases and diabetes [5-10], the learning methods in which resulting models are difficult to interpret (e.g., principle component analysis, support vector machine, and deep neural network) are not preferable in building clinically interpretable risk models.

In this study, we integrated interpretable machine learning approaches to build in-hospital mortality risk prediction models for STEMI patients from CAMI data. Feature engineering methods, including feature construction, missing data imputation and filter-based feature selection, were used to generate available dataset, identify potential predictors and reduce redundancy. Then we applied different categories of supervised learning methods that have good interpretability, including generalized linear models, decision trees, and Bayes models, to build risk prediction models that are suitable for different clinical scenarios. The experimental results show that our models can achieve higher prediction performance than the previous risk prediction models for STEMI, and are also easily interpretable for clinical decision support.

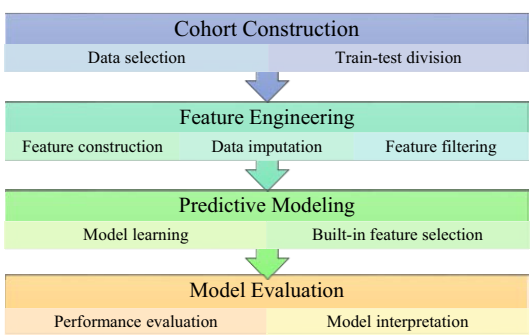

Figure 1 – Pipeline of building in-hospital mortality risk prediction models for STEMI patients

Methods

Figure 1 shows our pipeline of building in-hospital mortality risk prediction models for STEMI patients. We first constructed the cohort of interest, and applied feature engineering to identify potential predictors. Then we trained prediction models using supervised learning algorithms, and evaluated their prediction performance and model interpretability.

Cohort Construction

The dataset used in this study were collected in the CAMI registry project [2]. The project started in 2013 and 26,103 patients with AMI were registered until 2014. From the CAMI data, we identified 18,744 patients who were hospitalized due to STEMI in 2013 and 2014, where 1,263 patients were cases who died in hospital (the in-hospital mortality rate is 6.74%), and the others were control instances who survived in hospital. The features used in this study were those collected at the time of hospital presentation (i.e., before any in-hospital treatment). In the CAMI data, 132 originial features meet this criteria, including demographics, medical and treatment histories, life styles, onset symptoms, initial in-hospital vital signs, laboratory test results, etc.

In this study, we used the data of patients hospitalized in 2014 as the training set (9,619 patients, whose mortality rate is 6.78%) to develop the risk prediction models and the data of patients hospitalized in 2013 as the testing set (9,125 patients, whose mortality rate is 6.70%) to valid the prediction performance of the models.

Feature Engineering

In the raw CAMI data, some original features are not appropriate to be directly used in predictive modeling (e.g., birth year), and a large proportion of original features (more than 95%) have missing values. Moreover, not all original features are highly related to in-hospital mortality of STEMI. Therefore, we performed feature engineering to construct and select the potential predictors for risk prediction. We first transformed the original features to features that are easy to analyze by feature construction. For example, the "birth year" of each patient was transformed to "age". Then, we employed data imputation to fill-in missing values and applied filter-based feature selection algorithms to identify potential predictors from the imputed features.

The raw CAMI data has significant omissions due to the questionnaire structure; unknown values or errors in data collection. Therefore, we performed missing data imputation before predictive analysis. Firstly, the features with too many missing entries (more than 20%) were discarded, because their distributions are difficult to estimate. For the remaining features, every missing value of a numeric feature (e.g., SBP) was replaced with the mean of the feature's observed values, every missing value of an ordinal feature (e.g., Killip level) was replaced with the median of its observed values, and every missing value of an unordered nominal feature (e.g., history of diabetes) was replaced with the mode of its observed values. In this study, after feature construction and missing data imputation, 93 available candidate features were produced to the following feature selection step.

In machine learning, feature selection methods automatically test and select predictive features from a large number of candidate features. There are three main supervised feature selection strategies: filter, wrapper, and embedded models [11]. The filter models separate feature selection from model learning, and the wrapper and embedded models integrate feature selection in learning process. In the step of feature engineering, we performed filter models to remove the features that do not provide useful information and select the features that have high relevancy against the outcome. Concretely, the close-to-constant features, in which 99% of the instances have identical values, were first removed. Then we employed and compared two feature filtering models.

- **Univariate filter.** A univariate filter method calculates a score to represent the relevancy of a feature against the outcome, and filter the feature based on the score independently. In this study, we used the p-value from two standard statistical tests, the Chi-square test for categorical features and the ANOVA F test for numeric features, as the relevancy scores and selected the features whose p-value < 0.05.
- **Multivariate filter.** Different from the univariate filter method, a multivariate filter method evaluates the input features as a batch producing a subset of features that have the highest overall score. In this study, we used the correlation-based feature subset selection (CFS) method [12] to obtain the subset of features highly correlated with the outcome while having low intercorrelation between the features.

In this study, we also combined the features that were automatically selected by the filter-based feature selection algorithm with features that are well-known risk factors from prior knowledge [3,4], but were not automatically selected (e.g., anterior STEMI, time to treatment) as the predictors to build risk prediction models.

Predictive Modeling

In this study, we applied and compared different categories of machine learning models that have good interpretability, including generalized linear models (GLM), decision tree models, and Bayes models, to develop in-hospital mortality prediction models for STEMI patients. Besides, built-in feature selection strategies, including wrapper and embedded selection, were employed in some modeling processes.

- **Generalized linear model.** GLM generalizes ordinary linear regression by allowing the linear model to be related to the response variable via a link function, which is widely used in both medical statistics and machine learning due to its good prediction performance and interpretability. In this study, we applied logistic regression (LR), which is a GLM with a logit link function and a binomial distribution, and Cox proportional hazards model [13], which is a semi-parametric GLM that takes into account the time of censoring. We also employed forward stepwise feature selection, which is a wrapper selection model, to evaluate the performance and statistical significance of the LR and Cox models under different selection of features.
- **Decision tree model.** Decision trees are very easy to interpret and therefore have been successfully applied in healthcare. Decision tree learning algorithms usually embed feature selection into the learning process of a model when splitting the source data set into subsets. In this study, we employed the Chi-squared automatic interaction detector (CHAID) [14] method to build tree-based prediction model. CHAID is an efficient statistical technique that uses significance of Chi-squared test as a criterion for tree growing. We also employed random forest [15], which constructs multiple random decision trees and integrate the outputs of the trees for prediction. Compared to single decision tree models like CHAID, random forest reduces the problem of over-fitting, but has worse interpretability.
- **Bayes model.** Bayes models can learn probabilistic relationships among features and an outcome; computing

the probabilities of the outcome given the features. The Bayes models are also interpretable based on Bayes theorem. In this study, we employed naive Bayes [16], which is a Bayes model with strong independence assumptions between the input features and therefore can be trained very efficiently. We also applied Bayes network [17], which is a probabilistic graphical model that represents features and their conditional dependencies via a directed acyclic graph, where both the dependencies between outcome and input features and the interdependencies among input features can be modeled.

Results

We evaluated the performance of our approaches in building in-hospital mortality risk prediction models for STEMI patients from the CAMI dataset. The area under the receiver operating characteristic curve (AUC) were used to evaluate the prediction performance of models. We performed feature engineering and model learning on the training set, applied the learned models on the testing set, and evaluated the AUC on both datasets.

We first generated four different feature sets by feature engineering:

1. None of the feature filtering methods were applied, all original 93 features were kept.
2. Univariate filter selection was performed to select 51 features
3. CFS, which is a multivariate filter algorithm, selected 19 features
4. A combination of features from CFS and prior knowledge [3,4].

Then we built different learning models described above using different feature sets and evaluated their prediction performance. The results are shown in Table 1, where the models with less features and higher AUC on the testing set are highlighted. Random forest and the Bayes network achieved the best performance when applied on all candidate features, but their performance cannot be increased by filter-based feature selection. In comparison, after performing filter-based feature selection, the performance of GLM methods (LR and Cox) increased. Moreover, the combination of auto-selected features by CFS and prior knowledge based features improved the prediction performance for the majority of learning models.

We also compared the performance of our approaches to the state-of-the-arts risk models: TIMI score [3] and GRACE score [4]. We applied both previous models and our trained models on the same testing set and computed the AUC. As shown in Figure 4, the prediction performance of most of our models outweighed the TIMI and GRACE models.

Table 1 – AUC of different learning models on different feature sets

Feature selection	1) None			2) Univariate filter			3) CFS			4) Combination		
Learning model	**No. feature**	**AUC (train)**	**AUC (test)**	**No. feature**	**AUC (train)**	**AUC (test)**	**No. feature**	**AUC (train)**	**AUC (test)**	**No. feature**	**AUC (train)**	**AUC (test)**
LR	93	0.858	0.836	51	0.852	**0.842**	19	0.840	**0.841**	26	0.844	**0.843**
LR stepwise	**21**	0.846	0.839	19	0.845	0.839	15	0.840	0.840	17	0.842	**0.843**
Cox	93	0.853	0.829	51	0.849	0.838	19	0.839	0.840	26	0.842	0.842
Cox stepwise	**21**	0.842	0.835	**18**	0.843	0.835	**14**	0.838	0.838	**16**	0.840	0.839
CHAID	**11**	0.818	0.796	**11**	0.818	0.794	**7**	0.811	0.801	**7**	0.811	0.801
Random forest	93	0.917	**0.849**	51	0.915	**0.842**	19	0.898	**0.843**	26	0.901	**0.846**
Naive Bayes	93	0.820	0.818	51	0.820	0.823	19	0.817	0.820	26	0.821	0.825
Bayes network	93	0.872	**0.846**	51	0.867	0.840	19	0.865	0.835	26	0.868	0.835

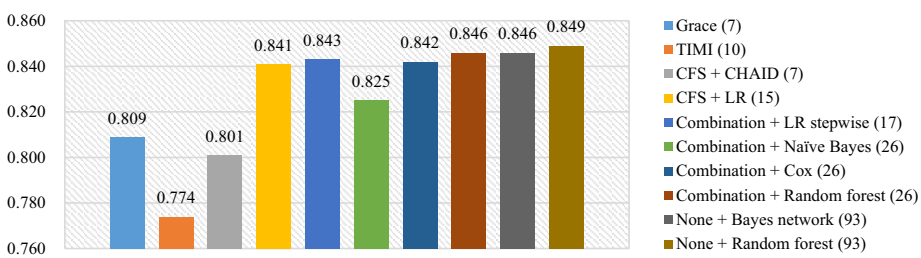

Figure 2 – AUC of different models, evaluated on the same testing set.

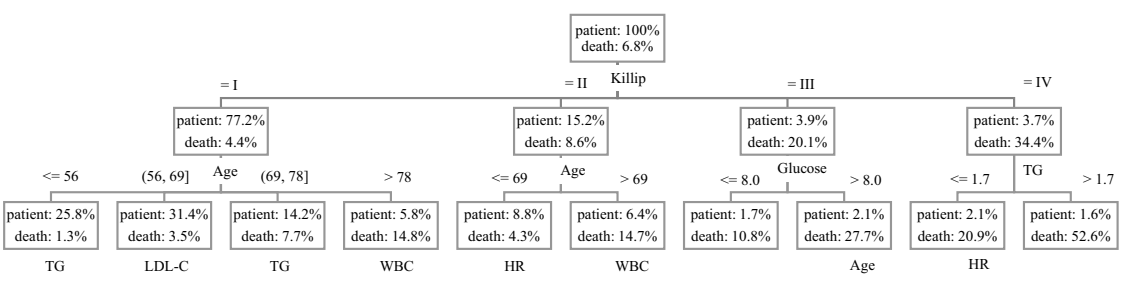

Figure 3 – CHAID tree model on CFS feature set

Since the objective of this work was to develop in-hospital mortality prediction models that can be used in real world clinical practices, the interpretability of the models was as important as the prediction performance. Table 2 shows the stepwise LR model built on the Combination feature set (AUC: 0.842, 95% confidence interval (CI):0.826-0.859). As a traditional regression analysis model, the contribution of each selected feature can be represented by the odds ratio (OR), and its statistical significance can be evaluated using 95% CI and p-value. For example, in the LR model of Table 2, for every 10-year increase in age the odds of in-hospital mortality multiplies by 1.654. The cox regression models are similar for interpretation, where the contribution of each feature can be represented by the hazard ratio.

Table 2 – Stepwise LR Model on Combination feature set

Feature	OR	95% CI		p
History of CABG	5.278	1.482	18.795	0.010
Cardiac shock	2.346	1.631	3.374	<0.001
Killip = I (referent)				<0.001
Killip = II	0.892	0.679	1.173	0.414
Killip = III	1.449	1.012	2.074	0.043
Killip = IV	1.910	1.317	2.770	0.001
Atypical presentation	1.776	1.305	2.418	<0.001
Age (per 10)	1.654	1.515	1.807	<0.001
Malignant arrhythmia	1.496	1.137	1.969	0.004
Anterior STE	1.346	1.111	1.631	0.002
Heart failure	1.326	1.019	1.727	0.036
Heart rate (per 10)	1.224	1.172	1.277	<0.001
Potassium (per 1)	1.210	1.037	1.413	0.016
WBC (per 10^9)	1.083	1.058	1.109	<0.001
Glucose (per 1)	1.051	1.026	1.075	<0.001
Creatinine (per 10)	1.039	1.025	1.053	<0.001
Weight (per 10)	0.888	0.804	0.982	0.020
SBP (per 10)	0.872	0.839	0.907	<0.001
Living with spouse	0.743	0.604	0.913	0.005
Sex (male)	0.666	0.539	0.821	<0.001

Though the AUC of the CHAID tree models are not as good as our other models, the CHAID model has the very clear interpretation. As shown in Figure 3, in the CHAID model built on the CFS feature set, the whole dataset can be split into four subgroup with very different mortality rates by Killip level. Also, the patient group with Killip = I can be further divided into four smaller subgroups by age. Therefore, each patient belong to a leaf node in the tree that can be uniquely defined using a set of rules, and the mortality rate of this node (subgroup) is explicitly used to predict the patient's risk. In contrast, though random forest achieved the best prediction performance on every feature set of this study, each random forest model has many different decision trees (100 trees in our setting) and is not easy to interpret.

A Bayes model can represent the conditional dependencies between input features and outcome, and also has good interpretability. Because the interdependencies between features are complex in the Bayes network models built in previous experiments, for demostration purpose we show the Bayes network model developed on the 12 well-known predictors using our training dataset in Figure 4. Both the dependencies between the outcome and the predictors (e.g., the death outcome has strong direct dependencies on creatinine, Killip level, heart rate, SBP and hypertension) as well as the interdependencies between the predictors (e.g., dependency between heart rate and Killip level, dependency between SBP and hypertension, etc.) are clearly represented in the Bayes network model.

Discussion

In this study, we compared the prediction of several feature

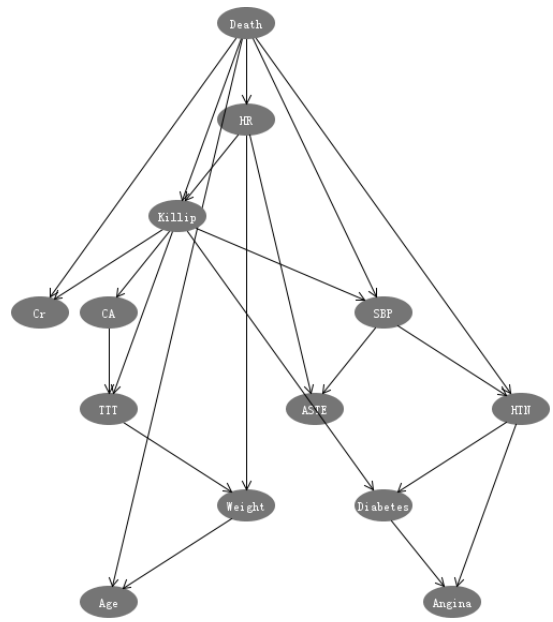

Figure 4 – Bayes network model built on knowledge-based features. HR = heart rate; Cr = creatinine; CA = Cardiac arrest, TTT = time to treatment; ASTE = anterior STEMI; HTN = hypertension

selection and supervised learning methods in building in-hospital mortality prediction models for STEMI patients. For GLM models (LR and Cox), appropriate feature selection can not only reduce the model complexity, but also improves the prediction performance. This is probably because GLM makes the assumption of no multicollinearity between the features, but the whole feature set is redundant, which negatively affects prediction performance. The feature selection methods, including CFS, which minimizes the intercorrelation and stepwise selection that optimizes the performance and statistical significance, can reduce the redundancy of features and therefore increase prediction performance. In contrast, the random forest and Bayes network methods can handle the redundant and intercorrelated features and therefore achieved the best prediction performance on the whole feature set. Moreover, the known predictors from prior knowledge [3,4] were grounded in previous evidence. Adding them to the auto-selected features from our dataset essentially includes the information outside the dataset, and therefore improved the prediction performance.

We also compared the interpretability of different machine learning models for risk prediction. As there is a trade-off between prediction performance and model interpretability, the choice of machine learning models may vary depending on the real-world clinical scenario.

5. Clinicians need to quickly estimate a patient's risk without any decision support tool requiring a human-memorable risk prediction model. Though the prediction performance of a single decision tree model (e.g., the CHAID model, Figure 3) is usually not as good as other machine learning models, it is human understandable and memorable, and is very suitable for this scenario.
6. Clinicians can predict a patient's risk with an independent risk prediction tool, such as a risk calculator, but still needs to manually input the predictor values. For these cases, a model that is

developed by combining feature selection and GLM (e.g., the stepwise LR model in Table 2) can provide decent prediction performance while keeping the input workload acceptable.

7. Clinicians can directly load a patient's data from the health information system (HIS) to perform risk prediction. For these cases, the complex risk prediction models with higher prediction performance (e.g., Bayes network and random forest models built on the whole feature set) can be used in clinical decision support.

For the purpose of interpretability, we did not apply more complex machine learning models such as deep neural network (DNN). However, there already have been attempts to make traditionally uninterpretable models interpretable. For example, Che et al. [18] developed a mimic learning approach, which can derive interpretable decision tree models from DNN models and maintain DNN's strong prediction performance. For future work, we would follow this direction in order to develop interpretable risk prediction models with higher prediction performance.

Another limitation of this work is that we only used a standard data imputation method based on column mean, median, and mode to remedy missing values. Some more advanced statistical imputation methods like multiple imputation, as well as the machine learning based imputation methods such as k-nearest neighbors imputation and neural network imputation, could be tried in the future, to make a more accurate estimation for missing values.

Conclusions

ST-elevation myocardial infarction (STEMI) is a major cause of hospitalization and has high in-hospital mortality rate. Accurate and interpretable prediction of in-hospital mortality is critical for clinical decision making to STEMI patients. In this study, we used integrated machine learning approaches, including the feature engineering and supervised learning methods that have good interpretability, to build in-hospital mortality prediction models for STEMI patients from CAMI data. The experimental results show that our models achieve higher prediction performance than previous models, and are also easily interpretable for clinical decision support.

References

[1] J. Li, Q. Wang, S. Hu, Y. Wang, F.A. Masoudi, J.A. Spertus, H.M. Krumholz, L. Jiang, China Peace Collaborative Group, ST-segment elevation myocardial infarction in China from 2001 to 2011 (the China PEACE-Retrospective Acute Myocardial Infarction Study): a retrospective analysis of hospital data, *Lancet* **385** (2015):441-51.

[2] H. Xu, J. Yang, S.D. Wiviott, M.S. Sabatine, E.D. Peterson, Y. Xian, M.T. Roe, W. Zhao, Y. Wang, X. Tang, The China Acute Myocardial Infarction (CAMI) Registry: A national long-term registry-research-education integrated platform for exploring acute myocardial infarction in China, *Am Heart J* **175** (2016):193-201.

[3] D.A. Morrow, E.M. Antman, A. Charlesworth, R. Cairns, S.A. Murphy, J.A. de Lemos, R.P. Giugliano, C.H. McCabe, E. Braunwald, TIMI risk score for ST-elevation myocardial infarction: A convenient, bedside, clinical score for risk assessment at presentation an intravenous nPA for treatment of infarcting myocardium early II trial substudy, *Circulation* **102** (2000):2031-2037.

[4] C.B. Granger, R.J Goldberg, O. Dabbous, K.S. Pieper, K.A. Eagle, C.P. Cannon, F. Van De Werf, A. Avezum, S.G. Goodman, M.D. Flather, K.A. Fox, Predictors of hospital mortality in the global registry of acute coronary events, *Arch Intern Med* **163** (2003):2345-2353.

[5] A. Khosla, Y. Cao, C.C. Lin, H.K. Chiu, J. Hu, H. Lee, An integrated machine learning approach to stroke prediction, *In: Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining* (2010):183-192.

[6] H. Neuvirth, M. Ozery-Flato, J. Hu, J. Laserson, M.S. Kohn, S. Ebadollahi, M. Rosen-Zvi, Toward personalized care management of patients at risk: the diabetes case study, *In: Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining* (2011):395-403.

[7] J. Sun, J. Hu, D. Luo, M. Markatou, F. Wang, S. Edabollahi, Z. Daar, W.F. Stewart, Combining knowledge and data driven insights for identifying risk factors using electronic health records, *In: proceedings of the American Medical Informatics Association Symposium* **2012** (2012):901-10.

[8] O. Ogunyemi, D. Kermah, Machine learning approaches for detecting diabetic retinopathy from clinical and public health records, *In: proceedings of the American Medical Informatics Association Symposium* **2105** (2015): 983-90.

[9] X. Li, H. Liu, X. Du, P. Zhang, G. Hu, G. Xie, Integrated machine learning approaches for predicting ischemic stroke and thromboembolism in atrial fibrillation, *In: proceedings of the American Medical Informatics Association Symposium* **2016** (2016):799-807.

[10] E. Choi, A. Schuetz, W.F. Stewart, J. Sun, Using recurrent neural network models for early detection of heart failure onset, *J Am Med Inform Assoc* **2016** (2016):799.

[11] J. Tang, S. Alelyani, H. Liu, Feature selection for classification: A review, *Dat Class Algorithms Appl* (2014):37.

[12] H.Z. Hamilton, *Correlation-Based Feature Subset Selection for Machine Learning*, New Zealand, 1998.

[13] D.R. Cox, Regression models and life-tables, *J R Stat Soc B* **34** (1972):187–220.

[14] G. Kass, An exploratory technique for investigating large quantities of categorical data, *Appl Stat* **29** (1980):119–27.

[15] L. Breiman, Random forests, *Mach Learn* **45** (2001):5-32.

[16] G.H. John, P. Langley, Estimating continuous distributions in Bayesian classifiers, *Uncertainty in Artificial Intelligence Proc* (1995):338-345.

[17] J. Pearl, Bayesian networks: a model of self-activated memory for evidential reasoning. Proceedings of the 7th Conference of the, *In: Proceedings of the 7th Conference of the Cognitive Science Society* (1985):329–34.

[18] Z. Che, S. Purushotham, R. Khemani, Y. Li, Interpretable deep models for ICU outcome prediction, *In: proceedings of the American Medical Informatics Association Symposium* **2016** (2016):371-80.

Address for correspondence

Xiang Li: IBM Research – China, Building 19, Zhongguancun Software Park, Haidian District, Beijing, China, 100193. Email: lixiang@cn.ibm.com

Yuejin Yang: Department of Cardiology, Fuwai Hospital, National Center for Cardiovascular Diseases, Beijing, China, 100037. Email: yangyjfw@126.com