

Original Paper

Identifying Patients With Heart Failure Who Are Susceptible to De Novo Acute Kidney Injury: Machine Learning Approach

Caogen Hong^{1,2}, MSc; Zhoujian Sun³, PhD; Yuzhe Hao², MSc; Zhanghuiya Dong², MSc; Zhaodan Gu², MSc; Zhengxing Huang⁴, PhD

¹College of Biomedical Engineering and Instrument Science, Zhejiang University, Hangzhou, China

²Jiangsu Automation Research Institute, Lianyungang, China

³Research Center for Applied Mathematics and Machine Intelligence, Zhejiang Lab, Hangzhou, China

⁴College of Computer Science and Technology, Zhejiang University, Hangzhou, China

Corresponding Author:

Zhoujian Sun, PhD

Research Center for Applied Mathematics and Machine Intelligence

Zhejiang Lab

Kechuang Ave, Zhongtai Subdistrict, Yuhang District

Hangzhou, 311121

China

Phone: 86 571 56390515

Fax: 86 518 85983716

Email: sunzj@zhejianglab.edu.cn

Abstract

Background: Studies have shown that more than half of patients with heart failure (HF) with acute kidney injury (AKI) have new-onset AKI, and renal function evaluation markers such as estimated glomerular filtration rate are usually not repeatedly tested during the hospitalization. As an independent risk factor, delayed AKI recognition has been shown to be associated with the adverse events of patients with HF, such as chronic kidney disease and death.

Objective: The aim of this study is to develop and assess of an unsupervised machine learning model that identifies patients with HF and normal renal function but who are susceptible to de novo AKI.

Methods: We analyzed an electronic health record data set that included 5075 patients admitted for HF with normal renal function, from which 2 phenogroups were categorized using an unsupervised machine learning algorithm called K-means clustering. We then determined whether the inferred phenogroup index had the potential to be an essential risk indicator by conducting survival analysis, AKI prediction, and the hazard ratio test.

Results: The AKI incidence rate in the generated phenogroup 2 was significantly higher than that in phenogroup 1 (group 1: 106/2823, 3.75%; group 2: 259/2252, 11.50%; $P < .001$). The survival rate of phenogroup 2 was consistently lower than that of phenogroup 1 ($P < .005$). According to logistic regression, the univariate model using the phenogroup index achieved promising performance in AKI prediction (sensitivity 0.710). The generated phenogroup index was also significant in serving as a risk indicator for AKI (hazard ratio 3.20, 95% CI 2.55-4.01). Consistent results were yielded by applying the proposed model on an external validation data set extracted from Medical Information Mart for Intensive Care (MIMIC) III pertaining to 1006 patients with HF and normal renal function.

Conclusions: According to a machine learning analysis on electronic health record data, patients with HF who had normal renal function were clustered into separate phenogroups associated with different risk levels of de novo AKI. Our investigation suggests that using machine learning can facilitate patient phenogrouping and stratification in clinical settings where the identification of high-risk patients has been challenging.

(*JMIR Med Inform* 2022;10(10):e37484) doi: [10.2196/37484](https://doi.org/10.2196/37484)

KEYWORDS

heart failure; acute kidney injury; unsupervised machine learning; risk stratification; phenogrouping

Introduction

Acute kidney injury (AKI) is a common disorder in patients with heart failure (HF), with the reported incidence rate varying from 7% to 38% in cardiology departments [1-3]. A recently conducted nationwide survey in China showed that about 85% of AKI incidents that occurred during cardiac hospitalization were ignored or were late to be identified [4,5]. As an independent risk factor, the delayed recognition of AKI has been proven to be associated with worse outcomes of patients with HF (eg, chronic kidney disease and mortality) [4,6]. To this end, the prompt identification of patients with HF at high-risk of AKI has great potential to improve clinical outcomes.

Although a few specific clinical markers (eg, estimated glomerular filtration rate [eGFR]) have been adopted to evaluate the renal function of patients with HF such that those at high risk of AKI can be identified, these markers lack the ability to screen de novo AKI patients who had normal renal function at admission [7,8]. Of note, several recently conducted population studies have indicated that more than half of the AKI that occurred in patients with HF were de novo [1-3]. To address this challenge, we attempted to clarify the characteristics of patients with HF who are susceptible to de novo AKI and developed a machine learning model for identification of HF patients with normal renal function but at high risk of de novo AKI.

As recently conducted cardiovascular studies have demonstrated that an unsupervised machine learning approach is able to model correlations among variables that contain prognostic information and cluster cohesive patients into 1 homogeneous phenogroup

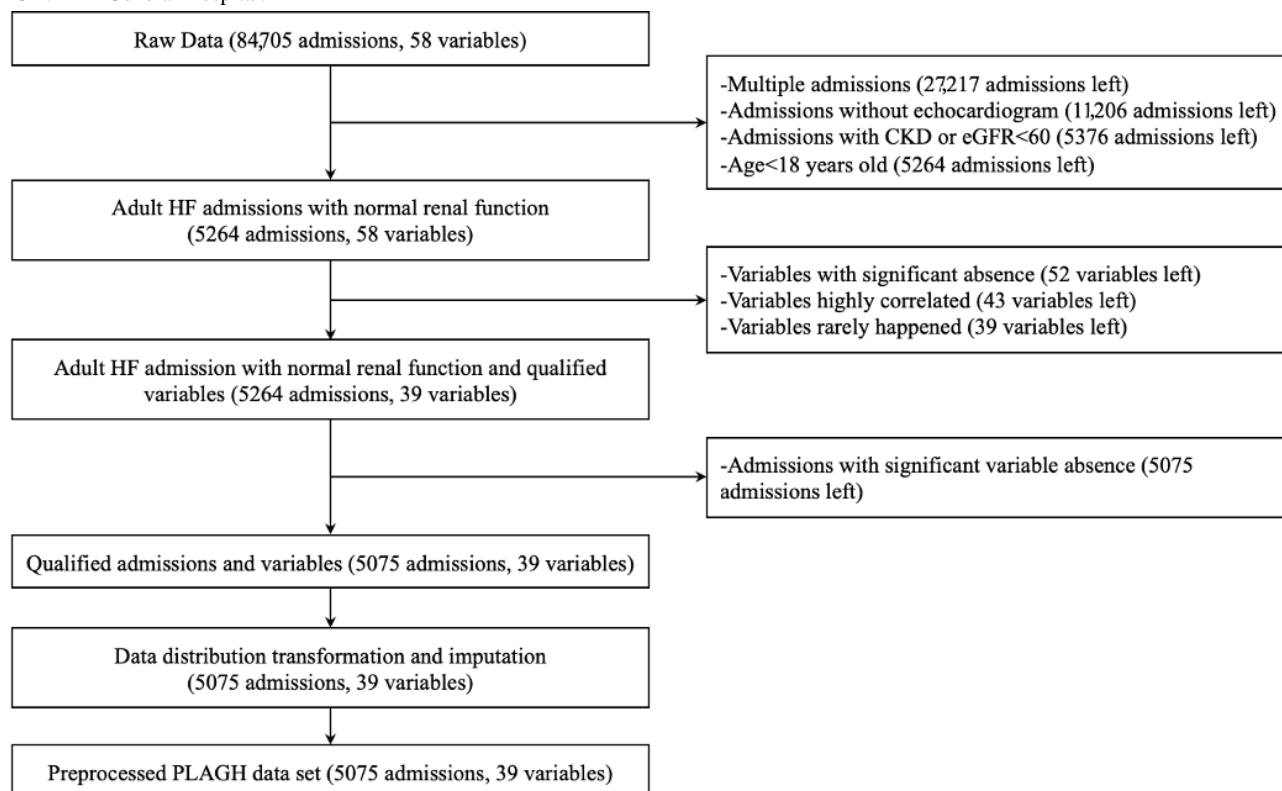
[9-11], we hypothesized that it can also be applied to identify patients with HF at high risk of de novo AKI. Recently, with the rapid development of hospital information systems, a large collection of electronic health records (EHRs) has become available that documents various types of patient information (eg, vital signs, laboratory test results) and treatments (eg, medication, surgery) and thus offers the considerable potential to implement a large-scale real-world analysis at a low expenditure. Therefore, in this study, we aimed to develop an EHR-based unsupervised machine learning analysis to group patients with HF and identify those who are susceptible to de novo AKI.

Methods

Study Population

The proposed retrospective study used a real-world data set obtained from the EHR system of the Chinese PLA General Hospital (PLAGH). The data set documented regular medical information in 84,705 hospitalizations of 29,699 patients who were diagnosed with HF in the PLAGH from 1998 to 2018. Adult patients with HF and normal renal function (eGFR >60 mL/min/1.73m² as calculated by the serum creatinine [SCr] version of the Chronic Kidney Disease Epidemiology Collaboration [CKD-EPI] equation [12] and without chronic kidney disease diagnosis) were considered for inclusion. Additionally, patients who did not have echocardiogram records were excluded. For patients with multiple hospitalizations, only the last hospitalization was reserved. The detailed preprocessing procedure is illustrated in Figure 1.

Figure 1. Preprocessing procedure of the PLAGH data set. CKD: chronic kidney disease; eGFR: estimate glomerular filtration rate; HF: heart failure; PLAGH: PLA General Hospital.



Ethics Approval

The study protocol was approved, with a waiver of consent granted on the basis of minimal harm and general impracticability by the health institutional review board of Zhejiang University (No. ZJU-2021-27).

Variable Selection and Machine Learning Model

In this study, 58 variables potentially associated with AKI, including demographics, vital sign measurements, medications, laboratories, operations, and echocardiogram exams, and routinely documented in EHRs at the admission stage of hospitalization were considered as candidates for analysis. To ensure that the most informative variables were selected and the correlation between variables could be diluted, we excluded variables with a missing rate larger than 30% or with a Pearson correlation coefficient >0.6 or that were documented fewer than 100 times in the raw EHR data set. As a result, 39 variables were included in the cohort. All continuous variables were transformed to standard normal distribution for the convenience of the unsupervised machine learning model (Table S1, [Multimedia Appendix 1](#)). Thereafter, we adopted multivariate imputation by chained equations [13] to impute the missing data.

We employed a simple yet effective unsupervised machine learning model called K-means clustering to categorize patients into different phenogroups [14]. The silhouette coefficient was applied to determine the optimal number of phenogroups [15]. We also adopted the nonlinear dimensionality reduction technique of t-distributed stochastic neighbor embedding [16] to visualize and evaluate the clustering results in a qualitative manner. The model was repeatedly run 1000 times to guarantee the achieved results stable.

Outcomes of Interest

The primary outcome was the incidence of AKI, which was defined according to the Kidney Disease: Improving Global Outcomes (KDIGO) standard [17], with the occurrence of AKI defined as the increase of SCr to ≥ 1.5 times the baseline in 7 days or the increase of SCr by ≥ 26.5 $\mu\text{mol/L}$ within 48 hours. The secondary outcome was in-hospital mortality.

Characterization of Phenogroups

Once patients with HF were categorized into separate phenogroups, we measured the differences of variables in different groups. Continuous variables are reported as median and IQR (interquartile range). Categorical variables are reported as the frequencies and counts. Differences between groups were tested using the 1-way analysis of variance, Kruskal-Wallis test, or the chi-square test where appropriate. A P value of $<.01$ was considered statistically significant.

Discrimination of Phenogroups

We validated whether the phenogroup index generated by K-means clustering correlated with outcomes of interests by carrying out the following 3 experiments. First, Kaplan-Meier estimators with log-rank tests were conducted to analyze the time-to-event characteristics in different phenogroups. Second, we compared the prediction performance on AKI and in-hospital mortality to check whether the inferred phenogroup index was

an effective risk predictor for outcomes of interest. Specifically, we selected the top-ranked 10 variables using a forward stepwise strategy with the Akaike information criterion and then developed 5 logistic regression (LR) models to predict the outcomes of interest. Model 1 used the phenogroup index as the univariate predictor. Model 2 used the top-ranked 10 variables as predictors. Model 3 used the top-ranked 10 variables and the phenogroup index. Model 4 used all 39 variables. Model 5 used all 39 variables and the phenogroup index. All models were trained by 70% of the data from the PLAGH data set and tested with the remaining 30% of data. Third, to evaluate whether the phenogroup index could achieve the competitive discriminative performance compared to the original variables with respect to the primary and secondary outcomes, we applied unadjusted Cox proportional hazard regression to examine hazard ratios (HRs), 95% CIs, and P values for all included original variables as well as the phenogroup index on both the whole PLAGH data set and the following subgroups: age (age <65 vs ≥ 65 years), sex, type of HF (acute vs chronic), diabetes mellitus, stroke, atrial fibrillation, coronary heart disease, anemia, and left ventricular ejection fraction ($<40\%$, $40\%-49\%$, and $\geq 50\%$). To assess continuous variables appropriately, we categorized all continuous variables in validation, and the cutoff points for these continuous variables are presented in online supplementary Table S2, [Multimedia Appendix 1](#).

External Validation

We externally validated our model on a well-known open-source database, Medical Information Mart for Intensive Care (MIMIC)-III [18]. After a requisite preprocessing procedure (online supplementary, Figure S1), we prepared a MIMIC-III data set that contained 1006 patients with HF who had normal renal function. The model trained by the PLAGH data set was directly transferred onto the MIMIC-III data set. In detail, we compared the distance between the data of each patient in the MIMIC-III data set and the centroids of the derived phenogroups from the PLAGH data set and then assigned the patient into a phenogroup with the minimum Euclidean distance. After that, we assessed the survival rate and prediction performance of AKI and in-hospital mortality of the generated phenogroups from the MIMIC-III data set. As patients contained in the PLAGH data set were mainly from general wards in the PLAGH and patients included in the MIMIC-III data set were from intensive care units in the United States, there inevitably were statistical differences between the baseline characteristics of patients in the 2 data sets (Table S3, [Multimedia Appendix 1](#)). In this sense, the external validation was able to evaluate the stability of the proposed model in diverse clinical settings.

In this study, statistical and machine learning analysis was based on sklearn, lifelines, scipy package [19-21], and Python. We also report the centroids of the generated phenogroups from the PLAGH data set (Table S4, [Multimedia Appendix 1](#)), which may be nontrivial knowledge to assist clinicians in identifying their patients with HF at high risk of de novo AKI.

Results

Phenogroup Results

After preprocessing, 5075 hospitalizations and 39 variables (Table 1) were reserved for the PLAGH data set (median age 61 years, IQR 51-70 years; female 1723/5075, 32.39%; acute

HF 1723/5075, 33.95%). Using K-means clustering, we naturally separated patients into 2 basically nonoverlapping phenogroups, where the number of clusters was suggested by the silhouette coefficient test (Figure S1, Multimedia Appendix 1). Similar results were found using t-distributed stochastic neighbor embedding visualization (Figure S2, Multimedia Appendix 1).

Table 1. Included variables for clustering.

Domain	Features
Demographic	Age, sex
Disease	Acute/chronic HF, atrial fibrillation, cardiomyopathy, coronary heart disease, diabetes, stroke, valvular heart disease
Medication	Angiotensin-converting enzyme inhibitor/angiotensin receptor blocker, anticoagulant, antiplatelet, beta blocker, calcium channel blocker, diuretic, positive inotropic drug, vasodilator
Echocardiography	Left ventricular ejection fraction
Laboratory result	Alanine aminotransferase, aspartate transaminase, estimated glomerular filtration rate, gamma-glutamyl transferase, hemoglobin, high-density lipoprotein cholesterol, low-density lipoprotein cholesterol, N-terminal probrain natriuretic peptide, serum calcium, serum potassium, serum sodium, serum urea, total bilirubin, total serum protein, triglyceride, troponin T
Operation	Angiography percutaneous coronary intervention
Vital sign	BMI, diastolic blood pressure, systolic blood pressure

^aOnly drugs used in the first 48 hours after admission were included to ensure the drug usage could reflect the patient admission status.

Characteristics of Phenogroups

Table 2 illustrates the baseline characteristics of the PLAGH data set and the 2 derived phenogroups. Compared to phenogroup 1, phenogroup 2 had a higher rates of AKI (group 1: 106/2823, 3.75%; group 2: 259/2252, 11.50%; $P<.001$) and in-hospital mortality (phenogroup 1: 21/2823, 0.74%; phenogroup 2: 118/2252, 5.24%; $P<.001$). In addition, patients in phenogroup 2 were generally older than those in phenogroup 1 (58 vs 65 years; $P<.001$).

As can be seen in Table 2, there are more patients diagnosed with acute HF in phenogroup 2 than those in phenogroup 1 (phenogroup 1: 738/2823, 26.14%; phenogroup 2: 985/2252, 43.74%; $P<.001$). Moreover, cardiac function of patients in phenogroup 2 was worse than that in phenogroup 1. Specifically, there were statistical differences between patients in phenogroup 1 and phenogroup 2 in terms of left ventricular ejection fraction (50% vs 41%; $P<.001$), diastolic blood pressure (77 mmHg vs 70 mmHg; $P<.001$), systolic blood pressure (130 mmHg vs 118 mmHg; $P<.001$), N-terminal pro-brain natriuretic peptide (572 pg/mL vs 2680 pg/mL; $P<.001$), hemoglobin (143 g/L vs 129 g/L; $P<.001$), atrial fibrillation (phenogroup 1: 526/2823, 18.63%; phenogroup 2: 595/2252, 26.42%; $P<.001$), diuretic usage (phenogroup 1: 1608/2823, 56.96%; phenogroup 2: 1799/2252, 79.88%; $P<.001$), and positive inotropic drug usage (phenogroup 1: 778/2823, 27.56%; phenogroup 2: 1089/2252, 48.36%; $P<.001$). Furthermore, phenogroup 2 had higher troponin T levels (0.01 ng/mL vs 0.02 ng/mL; $P<.001$), indicating that there were more patients in phenogroup 2 who

underwent myocardial damage. Patients in phenogroup 2 had higher values of gamma-glutamyl transferase (31.70 IU/L vs 40.30 IU/L; $P<.001$), total bilirubin (12.79 μ mol/L vs 15.85 μ mol/L; $P<.001$), and aspartate aminotransferase (19.60 IU/L vs 24.29 IU/L; $P<.001$), indicating that patients in phenogroup 2 might have worse liver function compared with phenogroup 1. Moreover, although we had excluded patients with renal dysfunction in advance, patients in phenogroup 2 had worse eGFR values (92.06 mL/min/1.73m² vs 81.85 mL/min/1.73 m²; $P<.001$) and urea (5.46 mmol/L vs mmol/L; $P<.001$). These findings demonstrated that patients in phenogroup 2 had relatively worse kidney function. Furthermore, patients in phenogroup 2 used less angiotensin-converting enzyme inhibitor/angiotensin receptor blocker (phenogroup 1: 1531/2823, 54.23%; phenogroup 2: 1016/2252, 45.11%; $P<.001$), calcium channel blocker (phenogroup 1: 789/2823, 27.95%; phenogroup 2: 321/2252, 14.25%; $P<.001$), and antiplatelets (phenogroup 1: 1914/2823, 67.80%; phenogroup 2: 1384/2823, 61.45%; $P<.001$). It was worth nothing that patients in phenogroup 2 had higher lipid levels (low-density lipoprotein cholesterol and triglyceride) and BMI (25.88 kg/m² vs 23.05 kg/m²; $P<.001$). Compared to phenogroup 1, phenogroup 2 also received less angiography (phenogroup 1: 1311/2823, 46.44%; phenogroup 2: 1311/2252, 30.95%; $P<.001$) and percutaneous coronary intervention (phenogroup 1: 640/2823, 21.96%; phenogroup 2: 349/2252, 15.50%; $P<.001$). Comprehensive baseline characteristics including all 58 candidate variables are listed in Table S5, Multimedia Appendix 1.

Table 2. Baseline characteristics of the PLA General Hospital data set and the generated phenogroups.

Feature	Population (N=5075)	Phenogroup 1 (n=2823)	Phenogroup 2 (n=2252)	P value
Feature of interest, n (%)				
AKI ^a	365 (7.19)	106 (3.75)	259 (11.50)	<.001
In-hospital mortality,	139 (2.74)	21 (0.74)	118 (5.24)	<.001
Demographic				
Age (years), median (IQR)	61 (51-70)	58 (48-67)	65 (55-75)	<.001
BMI (kg/m ²), median (IQR)	24.60 (22.46-27.08)	25.88 (23.87-28.08)	23.05 (20.95-25.01)	<.001
DBP ^b (mmHg), median (IQR)	74 (67-81)	77 (70-85)	70 (64-78)	<.001
SBP ^c (mmHg), median (IQR)	125 (113-138)	130 (119-143)	118 (106-130)	<.001
Male, n (%)	3431 (67.61)	1943 (68.83)	1488 (66.07)	<.001
Disease, n (%)				
HF^d				
Acute HF	1723 (33.95)	738 (26.14)	985 (43.73%)	<.001
Chronic HF	3352 (66.05)	2075 (73.86)	1267 (56.26%)	<.001
AF ^e	1121 (22.09)	526 (18.63)	595 (26.42)	<.001
Cardiomyopathy	941 (18.54)	497 (17.61)	444 (19.71)	<.001
CHD ^f	2928 (57.69)	1660 (58.80)	1268 (56.30)	.07
Diabetes	2002 (39.44)	1041 (36.88)	961 (42.67)	<.001
Stroke	485 (9.56)	282 (9.99)	233 (10.35)	.09
VHD ^g	616 (12.13)	336 (11.90)	280 (12.43)	.57
Medication, n (%)				
ACEI/ARB ^h	2547 (50.18)	1531 (54.23)	1016 (45.11)	<.001
Anticoagulant	1927 (37.97)	989 (35.03)	938 (41.65)	<.001
Antiplatelet	3298 (64.99)	1914 (67.80)	1384 (61.45)	<.001
Beta blocker	3428 (67.54)	1981 (70.17)	1447 (64.25)	<.001
CCB ⁱ	1110 (21.87)	789 (27.95)	321 (14.25)	<.001
Diuretic	3407 (67.13)	1608 (56.96)	1799 (79.88)	<.001
Positive inotropic drugs	1867 (36.79)	778 (27.56)	1089 (48.36)	<.001
Vasodilator	3103 (61.14)	1698 (60.15)	1405 (62.39)	.10
Echocardiogram				
LVEF ^j , median (IQR)	46 (35-56)	50 (39-58)	41 (31-54)	<.001
<40%, n (%)	1716 (33.81)	719 (25.47)	997 (44.27)	<.001
40%-50%, n (%)	1174 (23.13)	690 (24.44)	484 (21.49)	.05
≥50%, n (%)	2185 (42.86)	1414 (50.09)	771 (34.24)	<.001
Laboratory result, median (IQR)				
ALT ^k , (IU/L)	21.00 (14.39-33.79)	20.80 (14.70-31.99)	21.54 (13.80-36.49)	<.001
AST ^l , (IU/L)	21.29 (16.29-30.50)	19.60 (15.50-26.00)	24.29 (18.09-38.80)	<.001
Calcium (mmol/L)	2.24 (2.16-2.33)	2.28 (2.21-2.36)	2.19 (2.10-2.27)	<.001
eGFR ^m (mL/min/1.73 m ²)	87.62 (75.65-98.80)	92.06 (80.84-101.91)	81.85 (70.90-92.91)	<.001
GGT ⁿ (IU/L)	34.80 (21.90-63.79)	31.70 (21.30-54.89)	40.30 (23.09-75.00)	<.001

Feature	Population (N=5075)	Phenogroup 1 (n=2823)	Phenogroup 2 (n=2252)	P value
HDL-C ^o (mmol/L)	1.02 (0.85-1.22)	1.04 (0.88-1.22)	1.01 (0.82-1.22)	<.001
Hemoglobin, g/L	137 (124-150)	143 (132-154)	129 (116-142)	<.001
LDL-C ^p (mmol/L)	2.25 (1.79-2.81)	2.46 (1.96-3.05)	2.04 (1.62-2.48)	<.001
NT-pro-BNP ^q (pg/mL)	1216 (422-2950)	572 (225-1319)	2680 (1355-5188)	<.001
Potassium (mmol/L)	3.89 (3.62-4.17)	3.87 (3.62-4.13)	3.91 (3.61-4.20)	.005
Sodium (mmol/L)	140.70 (138.10-142.70)	141.30 (139.40-143.20)	139.40 (136.30-142.00)	<.001
Total bilirubin (μmol/L)	13.69 (9.80-19.90)	12.79 (9.40-17.40)	15.85 (10.39-24.60)	<.001
Total protein (g/L)	67.5 (63.3-71.8)	69.2 (65.8-73.3)	65.1 (60.4-69.0)	<.001
Triglyceride (mmol/L)	1.11 (0.82-1.59)	1.34 (0.98-1.87)	0.92 (0.72-1.21)	<.001
Troponin T (ng/mL)	0.01 (0.01-0.04)	0.01 (0.00-0.02)	0.02 (0.01-0.10)	<.001
Urea (mmol/L)	5.84 (4.73-7.25)	5.46 (4.51-6.60)	6.45 (5.11-8.12)	<.001
Operation, n (%)				
Angiography	2008 (29.57)	1311 (46.44)	697 (30.95)	<.001
PCI ^r	969 (19.09)	620 (21.96)	349 (15.50)	<.001

^aAKI: acuted kidney injury.

^bDBP: diastolic blood pressure.

^cSBP: systolic blood pressure.

^dHF: heart failure.

^eAF: atrial fibrillation.

^fCAD: coronary artery disease.

^gVHD: valvular heart disease.

^hACEI/ARB: angiotensin-converting enzyme inhibitor/angiotensin receptor blocker.

ⁱCCB: calcium channel blocker.

^jLVEF: left ventricular ejection fraction.

^kALT: alanine aminotransferase.

^lAST: aspartate transaminase.

^meGFR: estimated glomerular filtration rate.

ⁿGGT: gamma-glutamyl transferase.

^oHDL-C: high-density lipoprotein cholesterol.

^pLDL-C: low-density lipoprotein cholesterol.

^qNT-pro-BNP: N-terminal probrain natriuretic peptide.

^rPCI: percutaneous coronary intervention.

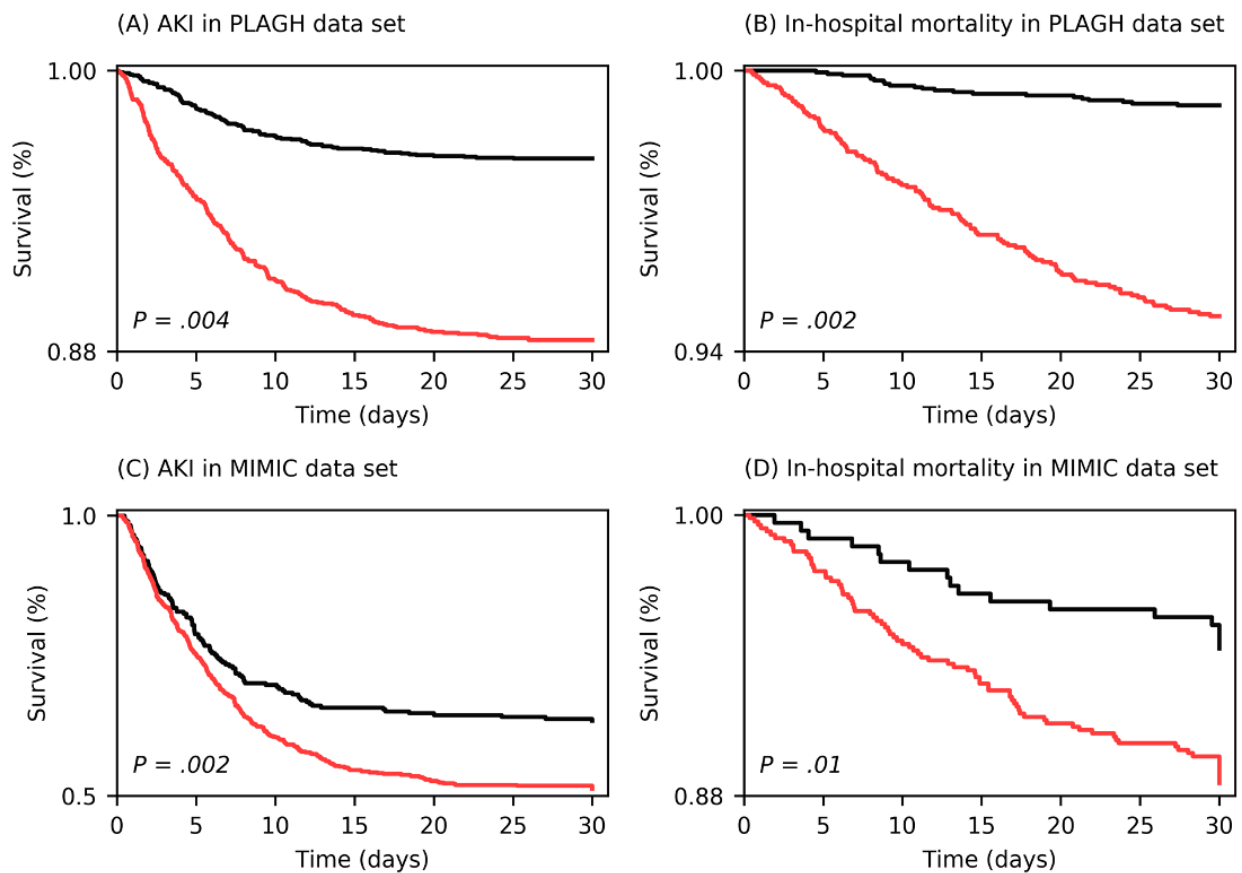
Survival Analysis

As the prevalence of AKI and in-hospital mortality had a significant difference between the generated phenogroups, phenogroup 1 was intuitively labeled as “low-risk” and phenogroup 2 as “high-risk.” We further investigated whether the generated phenogroup index could serve as an essential risk indicator for clinical outcomes of interest.

Figure 2 shows the survival difference with respect to AKI and in-hospital mortality between the generated “high-risk” and “low-risk” phenogroups from both the PLAGH data set and the external validation MIMIC-III data set. For AKI, the curves of

phenogroup 2 were lower than the curves of phenogroup 1 in both development and external validation data sets (PLAGH: $P=.004$; MIMIC-III: $P=.002$). In addition, we found that most AKI events often happened in the first few days of hospitalization in both the PLAGH and MIMIC-III data sets. This finding was in line with the literature [7,8]. For in-hospital mortality, the curves of phenogroup 2 were consistently lower than the curves of phenogroup 1 (PLAGH: $P=.002$; MIMIC-III: $P=.01$). In consideration of the baseline difference between the PLAGH data set and MIMIC-III data set, the results demonstrated that our model was robust in discriminating between high-risk and low-risk patients and easily transferable to different clinical settings.

Figure 2. Kaplan-Meier curves for AKI and in-hospital mortality in the development (PLAGH) and external validation (MIMIC-III) data sets. AKI: acute kidney injury; MIMIC: Medical Information Mart for Intensive Care; PLAGH: PLA General Hospital.



Outcome Prediction

Table 3 compares the prediction performances of the 5 LR models. Sensitivity, specificity, and concordance statistics are reported for the prediction performance evaluation. As the false-negative prediction (ie, neglecting AKI) may lead to extremely negative consequences, we mainly compared the sensitivity performance among the 5 models. The threshold of sensitivity and specificity was 0.5 in all experiments, and the selected top-10 variables are listed in Table S6, [Multimedia](#)

[Appendix 1](#). The results showed that the phenogroup index was an essential risk predictor of outcomes. For one, Model 1 used 1 variable (the phenogroup index) as the predictor and achieved promising sensitivity in terms of AKI (0.710) and in-hospital mortality (0.820) among the 5 prediction models with the PLAGH data set. For another, the prediction performance of Model 1 remained quite stable in the external validation (AKI sensitivity 0.760; in-hospital mortality sensitivity 0.826), while there existed significant degradation of performance in the other prediction models.

Table 3. Prediction performance comparison.

Model by task	PLAGH ^a data set (development)			MIMIC-III ^b data set (validation)		
	Sensitivity	Specificity	C-statistic ^c	Sensitivity	Specificity	C-statistics
AKI^d						
Model 1	0.710	0.577	0.643	0.760	0.342	0.551
Model 2	0.647	0.638	0.696	0.374	0.652	0.532
Model 3	0.679	0.723	0.756	0.478	0.562	0.546
Model 4	0.737	0.753	0.815	0.544	0.560	0.570
Model 5	0.718	0.746	0.816	0.573	0.540	0.575
In-hospital mortality						
Model 1	0.849	0.568	0.708	0.826	0.309	0.568
Model 2	0.791	0.736	0.824	0.530	0.672	0.622
Model 3	0.820	0.763	0.856	0.622	0.599	0.647
Model 4	0.835	0.809	0.899	0.490	0.746	0.646
Model 5	0.856	0.812	0.900	0.620	0.720	0.644

^aPLAGH: PLA General Hospital.

^bMIMIC-III: Medical Information Mart for Intensive Care III.

^cC-statistic: concordance statistic.

^dAKI: acute kidney injury.

HR Comparison

We used unadjusted Cox proportional hazard regression to determine whether the phenogroup index can act as an essential risk stratification indicator in comparison with the original 39 included variables. The top-ranked 10 variables with the highest HR are listed in Figure 3 (full list is available from Figure S3, Multimedia Appendix 1). The results showed that the HR of the phenogroup index was ranked second in AKI analysis and first in in-hospital mortality analysis, indicating that the phenogroup index can be an effective risk stratification indicator

compared with the original variables. Of further note, although troponin T was ranked first for AKI analysis, it was not appropriate for univariate risk indicators since only 16.73% (849/5075) of patients in the PLAGH data set had abnormal records in troponin T. Using troponin T as the indicator only achieved a sensitivity of 0.431, which was significantly lower than the performance of the phenogroup index (0.710). The association between the generated phenogroup index and risk of AKI (in-hospital mortality) was consistent in all examined subgroups (Figure 4).

Figure 3. Hazard ratios of of top-ranked 10 discriminative features for (a) acute kidney injury and (b) in-hospital mortality from the PLA General Hospital data set. AST: aspartate aminotransferase; eGFR: estimated glomerular filtration rate; NT-pro-BNP: N-terminal probrain natriuretic peptide. *Anemia was defined as hemoglobin <135 g/L for men and hemoglobin <120 g/L for women. All units of variables in this figure are same as the units in Table 2.

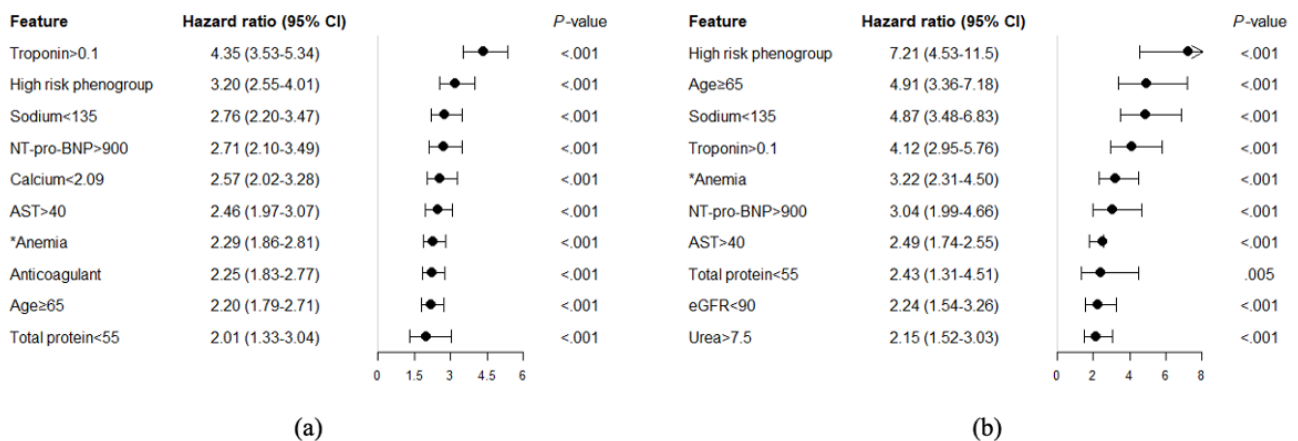
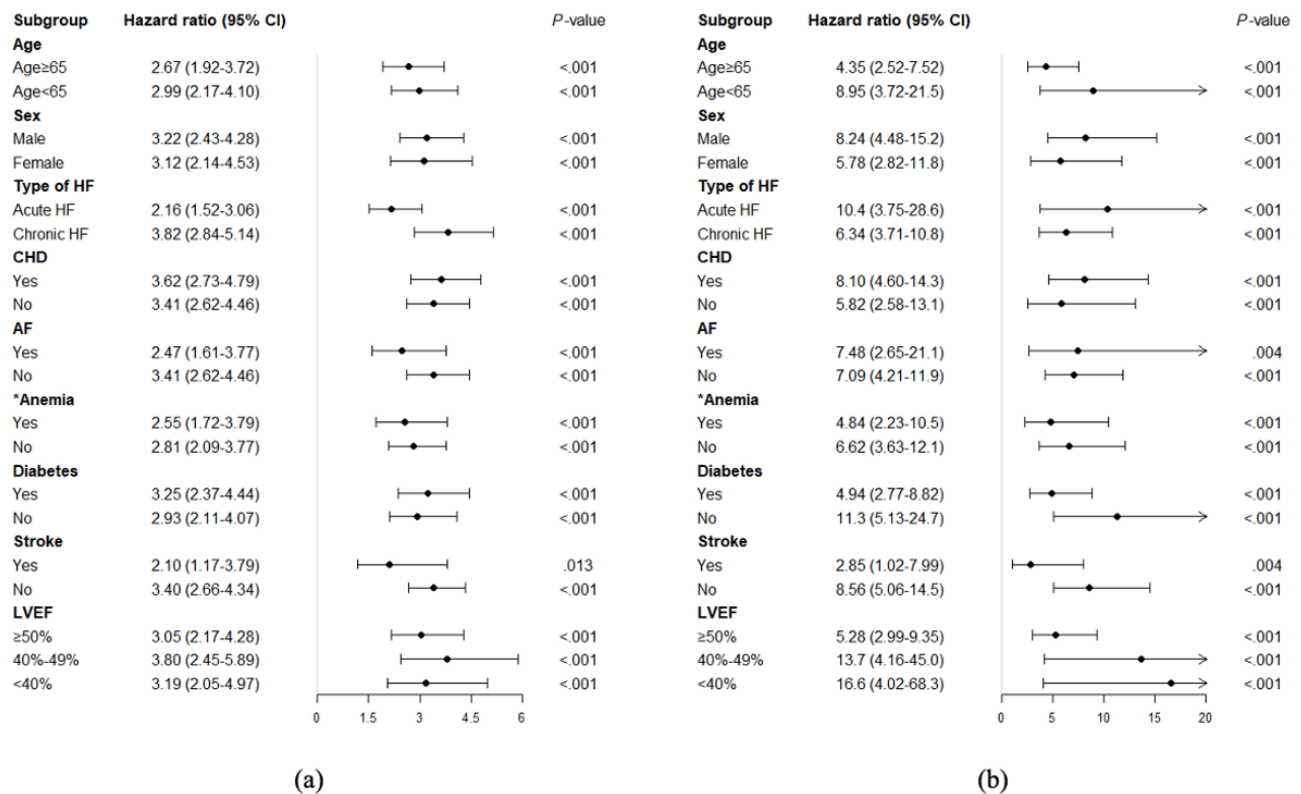


Figure 4. Subgroup analysis of the generated phenogroup index for (a) acute kidney injury and (b) in-hospital mortality. AF: atrial fibrillation; CHD: chronic heart disease; HF: heart failure; LVEF: left ventricular ejection fraction. *Anemia was defined as hemoglobin <135 g/L for men and hemoglobin <120 g/L for women. All units of variables in this figure are same as the units in Table 2.



Discussion

Principal Findings

We explored the potential of using a large volume of EHR data to cluster patients with HF and identify those with normal renal function but susceptible to de novo AKI via an unsupervised machine learning model. The experimental results showed that there was significant difference in AKI and in-hospital mortality occurrence between the 2 phenogroups generated from EHR data. As EHR is a real-world, readily available data source containing rich medical information of thousands of patients, our study demonstrated that it was possible for researchers to answer important clinical and scientific questions effectively by exploiting the huge potential of EHR data via machine learning techniques at a fraction of the resource cost that would have been required using traditional approaches [22,23].

We demonstrated that HF patients with normal renal function can be naturally separated into a “high-risk phenogroup,” of patients susceptible to de novo AKI and a “low-risk phenogroup” who were not. Patients in high-risk phenogroup were typically older, more susceptible to multi-organ dysfunction and anemia, and had significantly higher in-hospital mortality than did those in the low-risk phenogroup. These findings were in line with recent studies [17,24] and warrant further assessment. We found that patients in the high-risk phenogroup had lower levels of lipid and BMI than did those in the low-risk group. These findings are consistent with previous studies reporting that worse cardiac function may cause malnutrition [25] and a decrease of lipid level [26]. Of note,

worse cardiac function was also associated with hemodynamic instability, which influences the choice of oral medication strategies [27]. We observed that patients in the high-risk phenogroup received less medication (angiotensin-converting enzyme inhibitor, angiotensin receptor blocker, calcium channel blocker, and beta blockers) than did those in the low-risk phenogroup. On the contrary, we found that patients in low-risk phenogroup were likely to receive percutaneous coronary intervention (PCI) during their stay at the emergency care unit or in hospitalization to revascularize the stable hemodynamic level such that the perfusion of the kidney could be improved and the risk of AKI significantly alleviated. This finding is consistent with previous findings, emphasizing the benefit of timely revascularization [28].

Identification of patients with HF with normal renal function but at high-risk of de novo AKI is a major challenge in HF treatment management. Clinicians have highlighted the need for more effective methods to perform this important clinical task [29]. In this study, we illustrated that machine learning analysis can tackle this challenge by providing deep integration of the comprehensive clinical variables routinely documented in EHR data. As observed in the present study, the phenogroup index generated by an unsupervised machine learning approach, as a latent representation of 39 original variables and their interactions, exhibited a sensitivity of 0.710 and 0.760 on the development data set (PLAGH) and the external validation data set (MIMIC-III). In this sense, the generated phenogroups from raw EHR data are meaningful and can be translated into actionable information for clinical decision-making. On the

contrary, all other LR models met a serious overfitting problem due to the fact that the included variables had different distributions between the development (PLAGH) and external validation (MIMIC-III) data sets (as can be seen in Table S3, [Multimedia Appendix 1](#)). Inevitably, this issue caused a significant performance degeneration in the external validation. In consideration of the baseline difference between the PLAGH data set and the MIMIC-III data set, the results suggested that the generated phenogroup index was able to act as an essential de novo AKI risk indicator for patients with HF and normal renal function and be smoothly applied in different clinical settings and in different patient populations. In fact, machine learning algorithms can handle a large volume of variables and a vast number of variable-variable interactions in each patient. This merit effectively individualizes risk assessment and remedies many of the limitations of standard statistical models [22].

Our study has potentially important clinical ramifications. For one, as AKI risk is often underestimated or neglected in patient with HF, especially those with normal renal function [5], our study provided a new perspective for identifying patients with HF and normal renal function but who are at high risk of AKI. For another, in comparison with recent studies that focused on finding new biomarkers for AKI prediction or detection [30], we adopted an improved alternative strategy that used machine learning techniques to explore readily available clinical data to identify patients with HF at high risk of de novo AKI. Such meaningful use of EHR data may provide the best available evidence to assist clinical decision-making. It should be noted that these improvements may be enhanced by mining a large volume of readily available EHR data, which in turn may provide a new avenue for improving any given machine learning algorithm.

Limitations

Several limitations of this study should be acknowledged. First, this is a single-institution study. Although we have evaluated our model on an external validation data set extracted from MIMIC-III, the methods may perform less well in other situations due to the lack of sufficient external validation samples collected from different medical facilities and in different clinical settings. Second, our study was limited by its retrospective design, and all analyses were purely observational. Although we found that there were distinct variables associated with increased risks of de novo AKI and in-hospital mortality, these nonrandomized comparisons should be interpreted cautiously in this context, and the prognostic ability of our model needs to be supported by validation in prospective studies. Third, considering the sensitivity and the specificity for AKI forecasting, our model was relatively sensitive but not very specific. Despite the influence of false-positive classification being limited in this study, further study will be required to enable machine learning-based analysis to capture the salient features distinguishing high- from low-risk cases, such that the prediction performance of our model can be improved.

Conclusions

This study demonstrated that unsupervised machine learning-based EHR analysis is able to separate patients with HF and normal renal function into mutually exclusive phenogroups that correspond to saliently distinct AKI risk levels. Our investigation paves the way for developing an easy-to-use, broadly available model that allows the identification of patients with HF at high-risk of de novo AKI and may help improve outcomes in HF, offering a crucial advantage over traditional techniques for patient phenogrouping and clinical risk stratification.

Acknowledgments

The contribution of investigators and clinical coordinators are duly acknowledged.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Experimental data set introduction and detailed experiment results.

[\[DOC File , 2618 KB-Multimedia Appendix 1\]](#)

References

1. Roy AK, Mc Gorrian C, Treacy C, Kavanaugh E, Brennan A, Mahon NG, et al. A comparison of traditional and novel definitions (RIFLE, AKIN, and KDIGO) of acute kidney injury for the prediction of outcomes in acute decompensated heart failure. *Cardiorenal Med* 2013 Apr;3(1):26-37 [FREE Full text] [doi: [10.1159/000347037](https://doi.org/10.1159/000347037)] [Medline: [23801998](https://pubmed.ncbi.nlm.nih.gov/23801998/)]
2. Holgado JL, Lopez C, Fernandez A, Sauri I, Uso R, Trillo JL, et al. Acute kidney injury in heart failure: a population study. *ESC Heart Fail* 2020 Apr;7(2):415-422 [FREE Full text] [doi: [10.1002/ehf2.12595](https://doi.org/10.1002/ehf2.12595)] [Medline: [32059081](https://pubmed.ncbi.nlm.nih.gov/32059081/)]
3. Thakar CV, Parikh PJ, Liu Y. Acute kidney injury (AKI) and risk of readmissions in patients with heart failure. *Am J Cardiol* 2012 May 15;109(10):1482-1486. [doi: [10.1016/j.amjcard.2012.01.362](https://doi.org/10.1016/j.amjcard.2012.01.362)] [Medline: [22381163](https://pubmed.ncbi.nlm.nih.gov/22381163/)]
4. Yang L, Xing G, Wang L, Wu Y, Li S, Xu G, et al. Acute kidney injury in China: a cross-sectional survey. *The Lancet* 2015 Oct;386(10002):1465-1471. [doi: [10.1016/s0140-6736\(15\)00344-x](https://doi.org/10.1016/s0140-6736(15)00344-x)]

5. Tang X, Chen D, Yu S, Yang L, Mei C, ISN AKF 0 by 25 China Consortium. Acute kidney injury burden in different clinical units: Data from nationwide survey in China. *PLoS One* 2017 Feb 2;12(2):e0171202 [FREE Full text] [doi: [10.1371/journal.pone.0171202](https://doi.org/10.1371/journal.pone.0171202)] [Medline: [28152018](https://pubmed.ncbi.nlm.nih.gov/28152018/)]
6. Rangaswami J, Bhalla V, Blair JE, Chang TI, Costa S, Lentine KL, American Heart Association Council on the Kidney in Cardiovascular Disease Council on Clinical Cardiology. Cardiorenal syndrome: classification, pathophysiology, diagnosis, and treatment strategies: a scientific statement from the American Heart Association. *Circulation* 2019 Apr 16;139(16):e840-e878. [doi: [10.1161/CIR.0000000000000664](https://doi.org/10.1161/CIR.0000000000000664)] [Medline: [30852913](https://pubmed.ncbi.nlm.nih.gov/30852913/)]
7. Inohara T, Kohsaka S, Miyata H, Ueda I, Maekawa Y, Fukuda K, et al. Performance and validation of the U.S. NCDR acute kidney injury prediction model in Japan. *J Am Coll Cardiol* 2016 Apr 12;67(14):1715-1722 [FREE Full text] [doi: [10.1016/j.jacc.2016.01.049](https://doi.org/10.1016/j.jacc.2016.01.049)] [Medline: [27056778](https://pubmed.ncbi.nlm.nih.gov/27056778/)]
8. Abusaada K, Yuan C, Sabzwari R, Butt K, Maqsood A. Development of a novel score to predict the risk of acute kidney injury in patient with acute myocardial infarction. *J Nephrol* 2017 Jun;30(3):419-425. [doi: [10.1007/s40620-016-0326-1](https://doi.org/10.1007/s40620-016-0326-1)] [Medline: [27300206](https://pubmed.ncbi.nlm.nih.gov/27300206/)]
9. Cikes M, Sanchez-Martinez S, Claggett B, Duchateau N, Piella G, Butakoff C, et al. Machine learning-based phenogrouping in heart failure to identify responders to cardiac resynchronization therapy. *Eur J Heart Fail* 2019 Jan;21(1):74-85 [FREE Full text] [doi: [10.1002/ejhf.1333](https://doi.org/10.1002/ejhf.1333)] [Medline: [30328654](https://pubmed.ncbi.nlm.nih.gov/30328654/)]
10. Shah SJ, Katz DH, Selvaraj S, Burke MA, Yancy CW, Gheorghiade M, et al. Phenomapping for novel classification of heart failure with preserved ejection fraction. *Circulation* 2015 Jan 20;131(3):269-279. [doi: [10.1161/circulationaha.114.010637](https://doi.org/10.1161/circulationaha.114.010637)]
11. Segar MW, Patel KV, Ayers C, Basit M, Tang WW, Willett D, et al. Phenomapping of patients with heart failure with preserved ejection fraction using machine learning-based unsupervised cluster analysis. *Eur J Heart Fail* 2020 Jan;22(1):148-158 [FREE Full text] [doi: [10.1002/ejhf.1621](https://doi.org/10.1002/ejhf.1621)] [Medline: [31637815](https://pubmed.ncbi.nlm.nih.gov/31637815/)]
12. Levey AS, Stevens LA, Schmid CH, Zhang Y, Castro AF, Feldman HI, CKD-EPI (Chronic Kidney Disease Epidemiology Collaboration). A new equation to estimate glomerular filtration rate. *Ann Intern Med* 2009 May 05;150(9):604-612 [FREE Full text] [doi: [10.7326/0003-4819-150-9-200905050-00006](https://doi.org/10.7326/0003-4819-150-9-200905050-00006)] [Medline: [19414839](https://pubmed.ncbi.nlm.nih.gov/19414839/)]
13. Buuren SV, Groothuis-Oudshoorn K. mice: Multivariate Imputation by Chained Equations in R. *Journal of Statistical Software* 2011 Dec 12;45(3):1-67. [doi: [10.18637/jss.v045.i03](https://doi.org/10.18637/jss.v045.i03)]
14. James G, Witten D, Hastie T, Tibshirani R. *An Introduction to Statistical Learning*. New York, NY: Springer; 2013.
15. Rousseeuw PJ. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics* 1987 Nov;20:53-65. [doi: [10.1016/0377-0427\(87\)90125-7](https://doi.org/10.1016/0377-0427(87)90125-7)]
16. van der Maaten L, Hinton G. Visualizing data using t-SNE. *Journal of Machine Learning Research* 2008;9:2579-2605.
17. Khwaja A. KDIGO clinical practice guidelines for acute kidney injury. *Nephron Clin Pract* 2012;120(4):c179-c184 [FREE Full text] [doi: [10.1159/000339789](https://doi.org/10.1159/000339789)] [Medline: [22890468](https://pubmed.ncbi.nlm.nih.gov/22890468/)]
18. Johnson AE, Pollard TJ, Shen L, Lehman LH, Feng M, Ghassemi M, et al. MIMIC-III, a freely accessible critical care database. *Sci Data* 2016 May 24;3:160035 [FREE Full text] [doi: [10.1038/sdata.2016.35](https://doi.org/10.1038/sdata.2016.35)] [Medline: [27219127](https://pubmed.ncbi.nlm.nih.gov/27219127/)]
19. Abraham A, Pedregosa F, Eickenberg M, Gervais P, Mueller A, Kossaifi J, et al. Machine learning for neuroimaging with scikit-learn. *Front Neuroinform* 2014;8:14 [FREE Full text] [doi: [10.3389/fninf.2014.00014](https://doi.org/10.3389/fninf.2014.00014)] [Medline: [24600388](https://pubmed.ncbi.nlm.nih.gov/24600388/)]
20. Davidson-Pilon C. lifelines: survival analysis in Python. *JOSS* 2019 Aug;4(40):1317. [doi: [10.21105/joss.01317](https://doi.org/10.21105/joss.01317)]
21. Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, Polat, SciPy 1.0 Contributors. Author Correction: SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat Methods* 2020 Mar;17(3):352 [FREE Full text] [doi: [10.1038/s41592-020-0772-5](https://doi.org/10.1038/s41592-020-0772-5)] [Medline: [32094914](https://pubmed.ncbi.nlm.nih.gov/32094914/)]
22. Diller GP, Kempny A, Babu-Narayan SV, Henrichs M, Brida M, Uebing A, et al. Machine learning algorithms estimating prognosis and guiding therapy in adult congenital heart disease: data from a single tertiary centre including 10 019 patients. *Eur Heart J* 2019 Apr 01;40(13):1069-1077 [FREE Full text] [doi: [10.1093/eurheartj/ehy915](https://doi.org/10.1093/eurheartj/ehy915)] [Medline: [30689812](https://pubmed.ncbi.nlm.nih.gov/30689812/)]
23. Tokodi M, Schwertner W, Kovács A, Tóser Z, Staub L, Sárkány A, et al. Machine learning-based mortality prediction of patients undergoing cardiac resynchronization therapy: the SEMMELWEIS-CRT score. *Eur Heart J* 2020 May 07;41(18):1747-1756 [FREE Full text] [doi: [10.1093/eurheartj/ehz902](https://doi.org/10.1093/eurheartj/ehz902)] [Medline: [31923316](https://pubmed.ncbi.nlm.nih.gov/31923316/)]
24. Zymliński R, Sokolski M, Biegus J, Siwołowski P, Nawrocka-Millward S, Sokolska JM, et al. Multi-organ dysfunction/injury on admission identifies acute heart failure patients at high risk of poor outcome. *Eur J Heart Fail* 2019 Jun;21(6):744-750 [FREE Full text] [doi: [10.1002/ejhf.1378](https://doi.org/10.1002/ejhf.1378)] [Medline: [30561066](https://pubmed.ncbi.nlm.nih.gov/30561066/)]
25. Sze S, Pellicori P, Zhang J, Clark AL. Malnutrition, congestion and mortality in ambulatory patients with heart failure. *Heart* 2019 Feb;105(4):297-306. [doi: [10.1136/heartjnl-2018-313312](https://doi.org/10.1136/heartjnl-2018-313312)] [Medline: [30121635](https://pubmed.ncbi.nlm.nih.gov/30121635/)]
26. Pitt B, Loscalzo J, Ycas J, Raichlen JS. Lipid levels after acute coronary syndromes. *J Am Coll Cardiol* 2008 Apr 15;51(15):1440-1445 [FREE Full text] [doi: [10.1016/j.jacc.2007.11.075](https://doi.org/10.1016/j.jacc.2007.11.075)] [Medline: [18402897](https://pubmed.ncbi.nlm.nih.gov/18402897/)]
27. Ponikowski P, Voors AA, Anker SD, Bueno H, Cleland JG, Coats AJ, et al. 2016 ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure. *Kardiol Pol* 2016 Oct 13;74(10):1037-1147. [doi: [10.5603/kp.2016.0141](https://doi.org/10.5603/kp.2016.0141)]
28. Shacham Y, Steinvil A, Arbel Y. Acute kidney injury among ST elevation myocardial infarction patients treated by primary percutaneous coronary intervention: a multifactorial entity. *J Nephrol* 2016 Apr;29(2):169-174. [doi: [10.1007/s40620-015-0255-4](https://doi.org/10.1007/s40620-015-0255-4)] [Medline: [26861658](https://pubmed.ncbi.nlm.nih.gov/26861658/)]

29. Abusaada K, Yuan C, Sabzwari R, Butt K, Maqsood A. Development of a novel score to predict the risk of acute kidney injury in patient with acute myocardial infarction. *J Nephrol* 2017 Jun;30(3):419-425. [doi: [10.1007/s40620-016-0326-1](https://doi.org/10.1007/s40620-016-0326-1)] [Medline: [27300206](https://pubmed.ncbi.nlm.nih.gov/27300206/)]
30. Lassus JPE, Nieminen MS, Peuhkurinen K, Pulkki K, Siirilä-Waris K, Sund R, FINN-AKVA study group. Markers of renal function and acute kidney injury in acute heart failure: definitions and impact on outcomes of the cardiorenal syndrome. *Eur Heart J* 2010 Nov 27;31(22):2791-2798. [doi: [10.1093/eurheartj/ehq293](https://doi.org/10.1093/eurheartj/ehq293)] [Medline: [20801926](https://pubmed.ncbi.nlm.nih.gov/20801926/)]

Abbreviations

AF: atrial fibrillation
AKI: acute kidney injury
CHD: coronary heart disease
CKD: chronic kidney disease
CKD-EPI: Chronic Kidney Disease Epidemiology Collaboration
eGFR: estimated glomerular filtration rate
EHR: electronic health record
HF: heart failure
HR: hazard ratio
KDIGO: Kidney Disease: Improving Global Outcomes
LR: logistic regression
MIMIC: Medical Information Mart for Intensive Care
PLAGH: Chinese PLA General Hospital
SCr: serum creatinine

Edited by T Hao, B Tang, Z Li; submitted 23.02.22; peer-reviewed by H Monday, G Nneji, A Naser, K Uludag; comments to author 15.05.22; revised version received 31.05.22; accepted 05.06.22; published 14.10.22

Please cite as:

Hong C, Sun Z, Hao Y, Dong Z, Gu Z, Huang Z

Identifying Patients With Heart Failure Who Are Susceptible to De Novo Acute Kidney Injury: Machine Learning Approach
JMIR Med Inform 2022;10(10):e37484

URL: <https://medinform.jmir.org/2022/10/e37484>

doi: [10.2196/37484](https://doi.org/10.2196/37484)

PMID:

©Caogen Hong, Zhoujian Sun, Yuzhe Hao, Zhanghuiya Dong, Zhaodan Gu, Zhengxing Huang. Originally published in JMIR Medical Informatics (<https://medinform.jmir.org>), 14.10.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Medical Informatics, is properly cited. The complete bibliographic information, a link to the original publication on <https://medinform.jmir.org/>, as well as this copyright and license information must be included.