

Machine learning-based prognostic modeling of patients with acute heart failure receiving furosemide in intensive care units

DIGITAL HEALTH
Volume 9: 1–12
© The Author(s) 2023
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/20552076231194933
journals.sagepub.com/home/dhj



Tadashi Kamio¹ , Masaru Ikegami², Yoshihito Machida²,
Tomoko Uemura², Naotaka Chino² and Masao Iwagami^{3,4,5}

Abstract

Purpose: This study developed machine learning models to predict in-hospital mortality, initiation of acute renal replacement therapy, and mechanical ventilation in patients with acute heart failure receiving furosemide in intensive care units.

Method: An extensive database comprising static and dynamic features obtained from a Japanese hospital chain was used to construct and train the machine learning models.

Results: The results revealed that the proposed machine learning models predict in-hospital mortality, initiation of acute renal replacement therapy, and mechanical ventilation with good accuracy. However, the optimal models vary depending on the predicted outcomes. The linear support vector machine classification models exhibited the highest in-hospital mortality and mechanical ventilation prediction accuracy, with the area under the receiver operating characteristic curve of 0.73 and 0.73, respectively, whereas the multi-layer neural network exhibited the highest accuracy for acute renal replacement therapy initiation prediction with an area under the receiver operating characteristic curve of 0.70.

Conclusions: In conclusion, this study demonstrated that machine learning models could help predict the clinical outcomes of patients with acute heart failure receiving furosemide. However, the optimal models may differ depending on the outcome of interest.

Keywords

Machine learning models, acute heart failure, furosemide, intensive care units

Submission date: 27 April 2023; Acceptance date: 28 July 2023

Introduction

Acute heart failure (AHF) remains a leading global health concern, with an estimated prevalence exceeding 37.7 million individuals worldwide.¹ AHF is a heterogeneous clinical syndrome with a complex and incomplete pathophysiology. Recommended treatments for the early management of AHF, including intravenous diuretics, oxygen, and vasodilators, have been associated with better outcomes.^{2–4}

Diuretics are essential for patients with AHF because they provide rapid symptomatic relief and control fluid retention. However, determining whether patients with AHF actually benefit from these drugs is complex in

¹Division of Critical Care, Shonan Kamakura General Hospital, Kanagawa, Japan

²Terumo Corporation R and D Center, Shonan Center, Ashigarakami-gun, Kanagawa, Japan

³Department of Health Services Research, University of Tsukuba, Ibaraki, Japan

⁴Health Services Research and Development Center, University of Tsukuba, Ibaraki, Japan

⁵Faculty of Epidemiology and Population Health, London School of Hygiene and Tropical Medicine, UK

Corresponding author:

Tadashi Kamio, Division of Critical Care, Shonan Kamakura General Hospital, 1370-1 Okamoto, Kamakura, Kanagawa 247-8533, Japan.

Email: tadashi-kamio@mail.goo.ne.jp



clinical practice. Relevant factors include age, ejection fraction, renal function, and disease severity. Therefore, accurate prediction of outcomes is a challenging task for physicians. If patients with AHF do not respond adequately to furosemide therapy, mechanical ventilation, and dialysis may become necessary.

Predicting outcomes and determining the risks of events help identify patients that are most likely to benefit from targeted interventions, thereby preventing future events. Recently, machine learning (ML) approaches have been applied to various aspects of medicine to predict the risks of clinical events.^{5,6} Electronic health records (EHRs) contain important information about individual patients and have, thus, been widely used to predict patient outcomes. They commonly contain diagnostic information, drug administration, vital signals, and details of laboratory tests. Based on learning from existing data, ML approaches have also been applied to improve the accuracy and efficiency of predictions of diagnosis, mortality, and rehospitalization in patients with HF.^{7–9} Establishing an accurate prediction model for patients with AHF based on EHRs may help physicians with early alerts and judgment; thus, ML techniques may play an essential role in the contemporary management of patients with AHF.¹⁰

Several studies have been conducted on this topic based on data from a single center or a few centers. In addition, most existing studies use only time-independent (static) data, such as demographics and background disease, to analyze and predict clinical outcomes. Few studies have predicted cases of invasive treatment, such as acute hemodialysis or mechanical ventilation.

In this context, this study used ML methods to develop and validate models for predicting the need for in-hospital mortality, invasive mechanical ventilation, and acute kidney injury requiring dialysis among patients with AHF undergoing furosemide therapy in intensive care units (ICUs). This study aimed to determine the optimal prediction model for each outcome based on data, including static and dynamic features obtained from an extensive database of a Japanese hospital chain.

Method

Data source

We conducted a retrospective cohort study using data obtained from the Tokushukai Medical Database between 2013 and 2019. The Tokushukai Medical Group operates 70 hospitals and 340 facilities throughout Japan and is Japan's most prominent private medical service operator (<https://www.tokushukai.or.jp/en/>, Accessed May 30, 2022). For the current study, we used the data from 20 hospitals with ICUs (Supplemental Table S1), participating in the Diagnosis Procedure Combination system, retrospectively.¹¹ The database includes the following information: patient

age, sex, vital signs, all measured in-patient laboratory values, admission and discharge dates, discharge outcome (alive or dead), primary admission diagnosis, comorbidities recorded by the attending physician at the time of discharge using International Classification of Disease of 10th revision (ICD-10) codes, comorbidities recorded by the doctor on admissions, post-admission illnesses, drugs, and hospital procedures, such as acute dialysis and ventilators.^{12,13} In most cases, laboratory data measurements were performed every 24 h, and additional measurements were conducted within a 24-h period if deemed necessary by the clinician.

This study was approved by the Tokushukai Group Joint Ethics Review Committee (approval number: TGE01426-024) and the Terumo Research Ethics Review Committee (processing number: CR20-R006) and was conducted in agreement with the principles of the Declaration of Helsinki. Information on this study, including its objectives, has been made publicly available.

Study participants

The study participants included adult patients (≥ 20 years of age) with AHF who underwent emergency or unscheduled admissions between 2013 and 2019 in 20 hospitals and who received (either bolus or continuous) furosemide in ICUs. We considered emergency care units (ECUs) as ICUs because, in the current Japanese system, ECUs tend to serve a similar role to ICUs, and there is no clear distinction between them. Patients with AHF were defined by DPC codes, with “heart failure” as the leading cause of hospitalization. Patients undergoing hemodialysis, who had an implanted pacemaker, underwent surgical or interventional treatment, or underwent intra-aortic balloon pumping or extracorporeal membrane oxygenation (ECMO) prior to ICU admission were excluded. Patients who exhibited acute renal replacement therapy (RRT) or died in the ICU on the same day of furosemide administration were also excluded.

In the current study, each patient was administered furosemide several times in ICUs. We regarded each furosemide administration as an independent event; therefore, multiple datasets could be generated for each patient (Supplemental Figure S1A). For all bolus and continuous administrations of furosemide, the recorded administration time was regarded as the start of administration.

Outcome variables

We evaluated (i) in-hospital mortality, (ii) initiation of acute RRT, and (iii) initiation of mechanical ventilation within 72 h of receiving furosemide. Initiation of RRT was defined as the initiation of RRT within 72 h of furosemide administration in the ICU. The initiation of mechanical ventilation was defined as the introduction of ventilators within 72 h of furosemide administration in the ICU.

Predictor variables

Thirty-four variables expected to predict the study outcomes were obtained from the database: age, sex; height; SPO₂; systolic, diastolic, and mean blood pressure; respiratory rate; body temperature; pulse rate; platelets; hemoglobin; leukocytes; blood urea nitrogen; potassium; HCO₃; PCO₂; PO₂; lactate; calcium; sodium; chlorine; troponin I; high-sensitivity troponin I; brain natriuretic peptide; creatinine; urine output amount; furosemide dosage; pulmonary sound; and cyanosis. We used two variables for systolic and diastolic blood pressure and urine output because they were recorded in the database under two different names or in different tables. For furosemide dosage, the active ingredient and volume were used as variables. The dosage at the time of administration was used as a variable for bolus administration. For continuous administration, we used the amount between the times recorded at the start and end of administration as a variable. Processed data were used for variables dependent on measurement time (e.g. blood pressure). We used the data obtained before furosemide administration in the ICU.

The variables were handled as three types of data: (i) time-series data, (ii) static data, and (iii) a combination of time-series and static data. When used as time-series data, the data were processed via extrapolation, comprising interfall spline interpolation and the median of five neighboring points. Thus, 24 data points were obtained for each hour following furosemide administration. Urinary volume was converted to an hourly figure. The missing values between measurements were interpolated using subsequent measured values. Extrapolation was performed as described previously.

The most recent data measured before furosemide administration were treated as static data. When using a combination of time-series and static data, variables related to vital signs, balanced outputs, and patient states were treated as time-series data. In contrast, the patient, laboratory, and medication information were treated as static data. The total number of 34 variables treated by each imputation method was 678 for the time-series data and 310 for the combination of time-series and static data.

Datasets

As mentioned above, separate datasets were created corresponding to each furosemide administration in the ICU (Supplemental Figure S1A). When explanatory variables were unavailable for 24 h, datasets whose objective variable period exceeded the ICU discharge were excluded. The datasets were divided into training datasets, used for model construction, and test datasets, used for accuracy assessment, in a ratio of 6:4 (Supplemental Figure S2B).

Oversampling was performed on the training data to adjust for an imbalance in the number of data points in

the objective variable. To predict the initiation of acute RRT, data were randomly selected and removed to achieve a 1:1 ratio of induction case data to non-induction case data.

Construction of the prognostic model

A prognostic model was constructed using the multi-neural network, linear support vector machine classification (LSVC), and eXtreme gradient boosting (XGB).

Data processing and manipulation were performed using Python (version 3.8.10).

Performance evaluation

Standard performance metrics were used to evaluate the models: F1, precision, recall, receiver operating characteristic (ROC), area under the curve (AUC), and precision–recall (PR) AUC. Five different split-test datasets were used for evaluation, and the accuracy was averaged over the five cases. The Youden index was used as the threshold for each model, and the learning curve ensured that the multi-neural network was adequately trained.

For the models with the highest ROC AUC, we calculated the contributions of the top five highest predictor variables to the model using feature importance for the LSVC and XGB models and Shapley Additive explanations for multi-neural network models.

Results

Study participants

In aggregate, 1416 patients were admitted to the ICU for AHF and treated with furosemide. Among them, 446, 503, and 1056 patients were excluded because they did not satisfy the eligibility criteria for the predictive tasks of in-hospital mortality, dialysis initiation, and ventilator initiation, respectively (Figure 1). Consequently, 970, 913, and 360 patients were used for model construction and evaluation of the prediction tasks for in-hospital mortality, acute RRT initiation, and ventilator initiation, respectively, comprising 6632, 5787, and 1689 data points, respectively, starting from furosemide administration (Figure 1). Table 1 lists the patient backgrounds in the datasets used to construct and evaluate the prediction models.

Prediction of in-hospital mortality

The performance evaluation of the in-hospital mortality prediction models is presented in Table 2 and Figure 2.

The LSVC model with data imputed based on combining time series and static data exhibited the best ROC AUC (0.74) (Table 2). The model with data imputed from static data also exhibited an almost identical ROC AUC

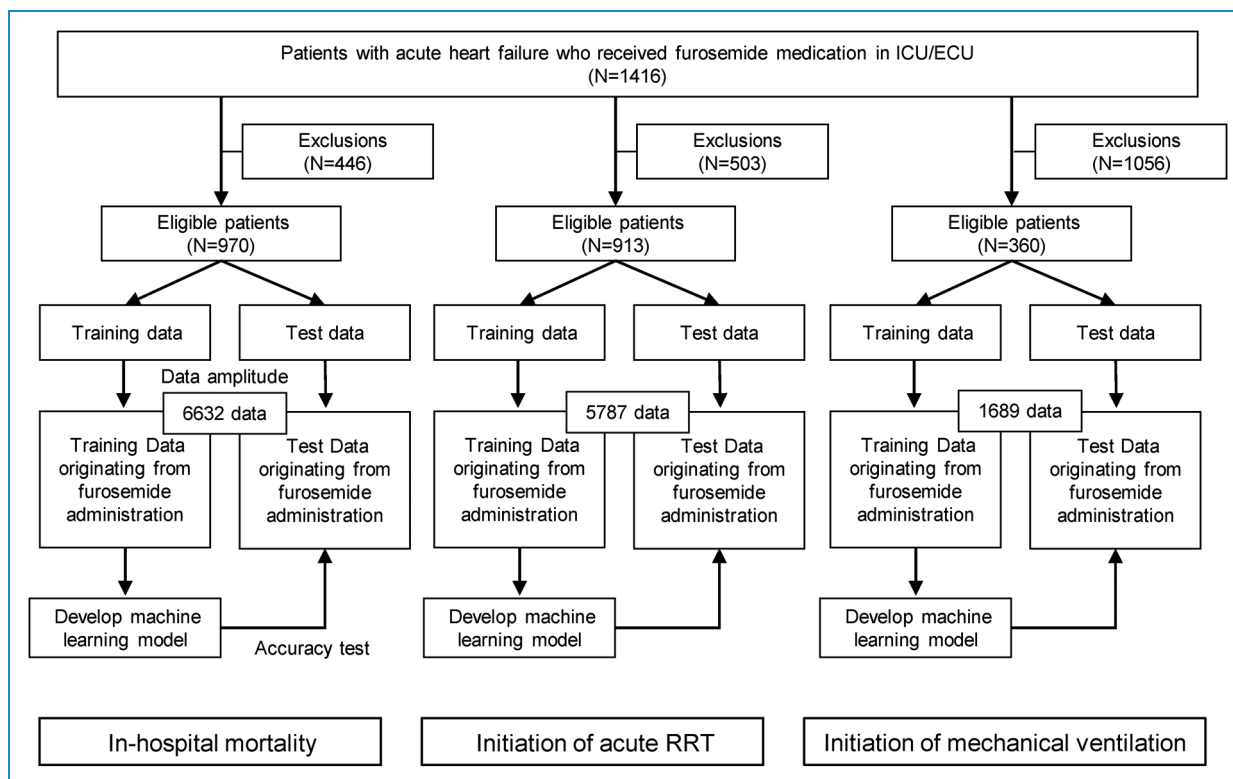


Figure 1. Data processing and model construction procedure.

(0.74). The contributions of these variables to the model are discussed in Table 5. Of the 310 variables treated by combining 34 variables with time series and static data, the contributions of lactate, blood urea nitrogen, platelets, creatinine, and age to the model were the highest.

Prediction of acute RRT initiation

The performance score results for the initiation of the acute RRT prediction model are presented in Table 3 and Figure 3. The multi-neural network model with data imputed as static data exhibited the best performance with an ROC AUC of 0.70 (Table 4). This model exhibited a similar or better F1 score (0.04) and PR AUC (0.05) compared to other data imputation methods and constructed models, with recall being the only exception. The contributions of these variables to the model are described in Table 5. Of the 34 variables treated as static data, sex, furosemide dosage, amount of active ingredient, sodium, and creatinine were the highest contributors to model performance.

Because the incidence outcome of initiating acute RRT was more imbalanced than that of other prediction tasks, with an incidence outcome of 1.1% (Table 1), oversampling was performed to compare the accuracy of the data. F1 and precision scores were similar, recall and ROC AUC improved, and PR AUC decreased (Supplemental Table S2).

Prediction of mechanical ventilation initiation

The performance evaluation of the initiation of the mechanical ventilation prediction model is presented in Table 4 and Figure 4. The LSVC model with data imputed as time-series data exhibited the best performance, with an ROC AUC of 0.732 (Table 4). This model exhibited the highest F1 score (0.15), precision (0.09), and PR AUC (0.07). However, the model did not exhibit the highest recall (Table 4).

The contributions of these variables to the model are presented in Table 5. Of the 678 variables treated as time-series data, age, urine output, diastolic blood pressure, SPO₂, furosemide amount of the active ingredient, and SPO₂ contributed the most to the model.

Discussion

In this study, ML models were developed to predict in-hospital mortality, initiation of acute RRT, and mechanical ventilation using data obtained from patients with AHF undergoing furosemide therapy after ICU admission. We demonstrated that machine learning-based models could predict the outcomes accurately. However, optimal models vary depending on the outcome to be predicted. The LSVC model exhibited the highest accuracy (ROC AUC: 0.73) for in-hospital mortality prediction, the multi-layer neural network exhibited the highest accuracy (ROC

Table 1. Patient characteristics.

	Study population for in-hospital mortality	Study population for dialysis initiation	Study population for ventilator initiation
	<i>n</i> = 970	<i>n</i> = 913	<i>n</i> = 360
	6632 data	5787 data	1689 data
Patient information			
Sex			
Female (%)	2934 (44.2%)	2674 (46.2%)	777 (46.0%)
Male (%)	3698 (55.8%)	3113 (53.8%)	912 (54.0%)
Age (years)	78.2 ± 11.2	78.3 ± 11.2	79.8 ± 12.0
Height (cm)	152.5 ± 29.5	152.6 ± 28.9	149.9 ± 32.8
Vital signs			
SPO ₂ (%)	96.6 ± 3.4	96.6 ± 3.2	96.2 ± 3.5
Systolic blood pressure (mmHg)	117.0 ± 24.5	116.8 ± 24.0	119.7 ± 22.9
Diastolic blood pressure (mmHg)	65.0 ± 16.3	65.1 ± 16.3	69.5 ± 18.0
Mean blood pressure (mmHg)	86.6 ± 18.9	88.0 ± 19.2	87.7 ± 18.4
Respiratory rate (/min)	19.9 ± 6.2	19.7 ± 5.9	20.5 ± 5.2
Body temperature (°C)	36.8 ± 1.1	36.8 ± 1.0	36.7 ± 1.2
Pulse rate (/min)	85.2 ± 19.5	84.9 ± 19.3	87.3 ± 20.7
Laboratory data			
Platelet (× 10 ⁴ /μL)	17.5 ± 8.8	17.8 ± 8.6	17.8 ± 8.6
Hemoglobin (g/dL)	11.0 ± 2.4	11.1 ± 2.4	11.1 ± 2.8
Leukocyte (× 10 ² /μL)	92.3 ± 33.7	91.7 ± 33.4	87.3 ± 33.7
Blood urea nitrogen (mg/dL)	36.3 ± 19.9	34.6 ± 19.0	35.6 ± 20.4
Potassium (mmol/L)	3.9 ± 0.7	3.9 ± 0.7	3.9 ± 0.7
HCO ₃ (mmol/L)	27.6 ± 7.1	28.3 ± 7.0	25.3 ± 6.3
PCO ₂ (mmHg)	43.2 ± 14.7	44.3 ± 15.1	35.1 ± 8.2
PO ₂ (mmHg)	93.8 ± 35.0	93.6 ± 36.1	89.9 ± 27.3
Lactate (mg/dL)	12.9 ± 16.0	11.2 ± 9.9	13.8 ± 14.3
Calcium (mg/dL)	8.1 ± 0.7	8.1 ± 0.6	8.1 ± 0.7

(continued)

Table 1. Continued.

	Study population for in-hospital mortality	Study population for dialysis initiation	Study population for ventilator initiation
Sodium (mmol/L)	140.0 ± 5.2	140.2 ± 5.1	139.2 ± 5.5
Chlorine (mmol/L)	102.1 ± 6.5	102.0 ± 6.6	101.8 ± 6.6
Troponin I (ng/mL)	5.6 ± 12.3	4.3 ± 9.9	4.5 ± 11.2
High-sensitive troponin I (ng/mL)	10.3 ± 7.9	10.3 ± 7.9	10.8 ± 5.6
Brain natriuretic peptide (pg/mL)	864.0 ± 758.3	823.1 ± 716.7	966.6 ± 727.5
Creatinine (mg/dL)	1.9 ± 1.6	1.7 ± 1.3	1.9 ± 1.9
Balance outputs			
Amount of urine output (mL/day)	150.6 ± 119.7	155.6 ± 120.9	167.4 ± 128.8
Medication (furosemide dosing)			
Dose of furosemide (mg)	25.4 ± 33.3	23.7 ± 31.2	25.0 ± 32.3
Patient state			
Pulmonary sound			
No	7844 (63.3%)	7214 (65.3%)	1960 (69.6%)
Mild symptom	2500 (20.2%)	2153 (19.5%)	455 (16.2%)
Yes	2042 (16.5%)	1685 (15.2%)	400 (14.2%)
Cyanosis			
No	10670 (98.0%)	9507 (98.3%)	2199 (98.0%)
Yes	217 (2.0%)	163 (1.7%)	44 (2.0%)
Incidence of outcome	1524 (23.0%)	65 (1.1%)	48 (2.8%)

SPO₂: oxygen saturation of peripheral artery; HCO₃: hydrogen carbonate; PO₂: partial pressure of oxygen; PCO₂: partial pressure of carbon dioxide. Data are presented as *n* (%) or mean.

AUC: 0.70) for acute RRT initiation prediction, and the LSVC model exhibited the highest accuracy (ROC AUC: 0.73) for mechanical ventilation initiation prediction.

Several studies have reported the use of ML, including deep neural networks, to predict prognoses such as mortality, cardiogenic shock, and hospitalization in patients with AHF.^{9,14,15} However, to the best of our knowledge, this is the first study to predict acute RRT initiation and mechanical ventilation initiation among AHF patients receiving furosemide therapy using ML models. There are many approaches to improving the accuracy of machine learning models, but even extremely accurate predictive models may not lead to better clinical care. Even if a study does not achieve the desired level of

accuracy, ML models predicting clinical outcomes may potentially assist in the risk stratification of patients with AHF and aid clinical decision-making for furosemide medication. We also identified factors that contributed to the accuracy of each predictive model in order to verify their consistency with clinical practice. Predictive algorithms using machine learning methods may also contribute to stratifying patient populations by risk in order to inform discrete decision-making.¹⁶

Thirty-four variables were considered in this study to predict the three aforementioned outcomes. These variables were treated as (i) time-series data, (ii) static data, and (iii) a combination of time-series and static data. Using the Gini index, we identified systolic blood pressure, sex, and age

Table 2. Predictive performance of in-hospital mortality.

Data	Method	F1	Precision	Recall	ROC AUC	PR AUC
Time-series	LSVC	0.45	0.36	0.61	0.68	0.41
	XGB	0.46	0.35	0.67	0.69	0.41
	Multi-neural network	0.36	0.36	0.38	0.58	0.31
Static	LSVC	0.50	0.37	0.76	0.74	0.47
	XGB	0.47	0.35	0.75	0.71	0.45
	Multi-neural network	0.46	0.34	0.69	0.69	0.41
Time-series & static	LSVC	0.49	0.38	0.69	0.74	0.47
	XGB	0.49	0.42	0.64	0.73	0.48
	Multi-neural network	0.37	0.41	0.52	0.65	0.37

LSVC: linear support vector machine classification; XGB: eXtreme gradient boosting; ROC: receiver operating characteristic; AUC: area under the curve; PR: precision-recall.

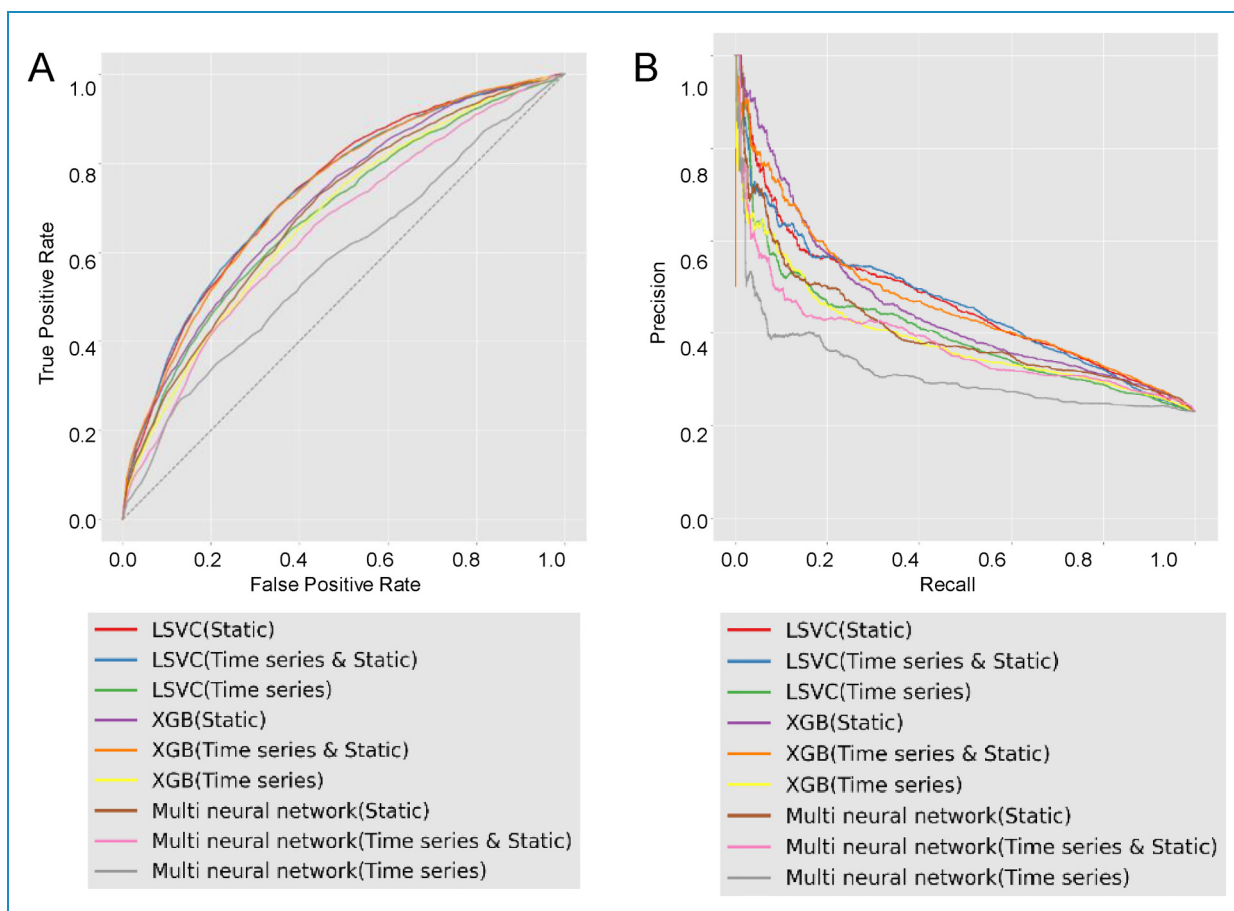


Figure 2. Final model performance. The best ROC curves of each ML model were obtained by testing various configurations in terms of feature extraction, data balancing, feature selection, and filtering for in-hospital mortality. (A) ROC curve for the crisis prediction task. (B) Precision-recall curve for the crisis prediction task with the same characteristics as (A). ROC: receiver operating characteristic; ML: machine learning.

Table 3. Predictive performance of initiation of acute RRT.

Data	Method	F1	Precision	Recall	ROC AUC	PR AUC
Time series	LSVC	0.04	0.02	0.65	0.57	0.02
	XGB	0.04	0.02	0.60	0.65	0.02
	Multi-neural network	0.04	0.02	0.69	0.57	0.01
Static	LSVC	0.03	0.02	0.73	0.55	0.01
	XGB	0.03	0.02	0.87	0.62	0.03
	Multi-neural network	0.04	0.02	0.76	0.70	0.05
Time series & static	LSVC	0.04	0.02	0.61	0.62	0.02
	XGB	0.04	0.02	0.67	0.66	0.03
	Multi-neural network	0.03	0.02	0.87	0.69	0.04

LSVC: linear support vector machine classification; XGB: eXtreme gradient boosting; ROC: receiver operating characteristic; AUC: area under the curve; PR: precision-recall; RRT: renal replacement therapy.

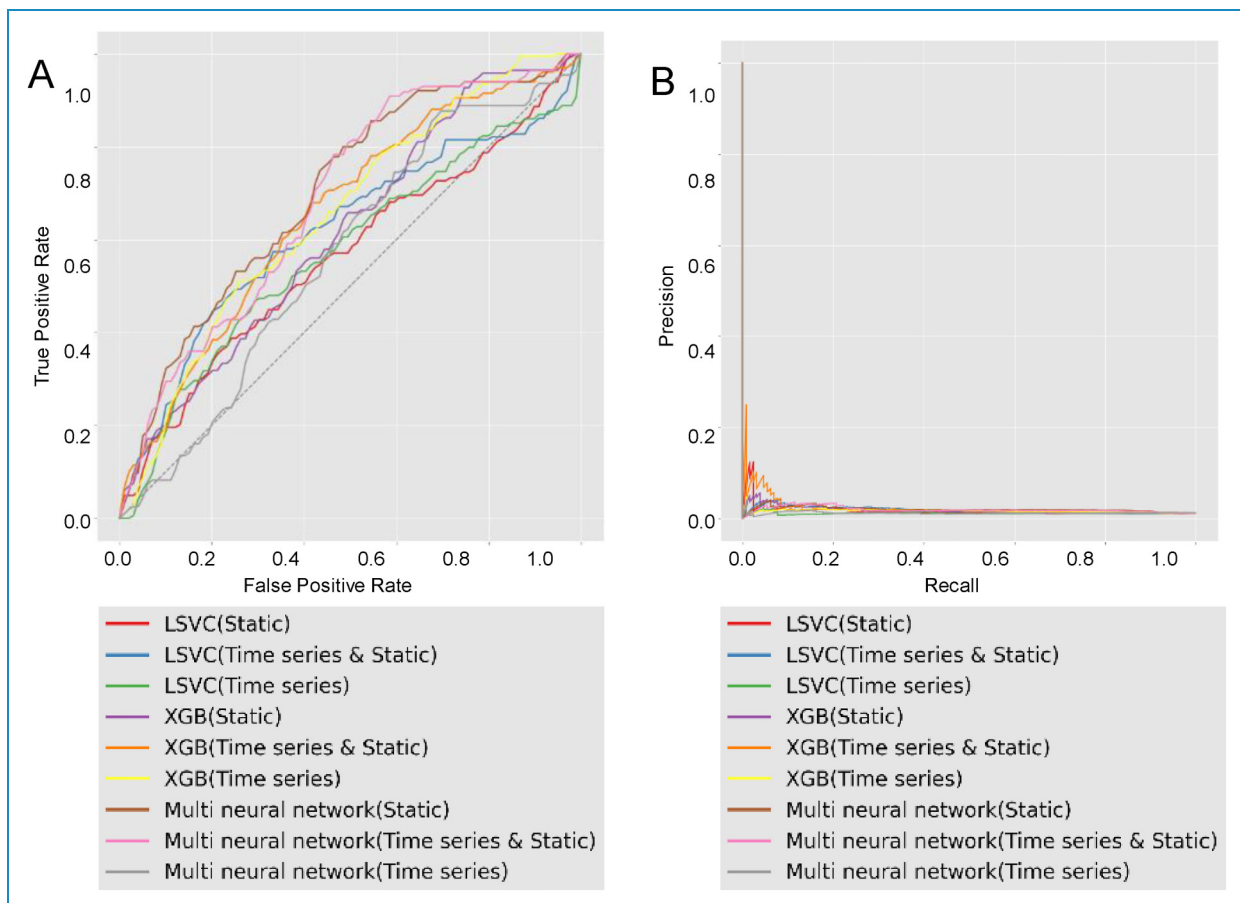


Figure 3. Final model performance. The best ROC curves for each ML model were obtained by testing various configurations in terms of feature extraction, data balancing, feature selection, and filtering for acute RRT initiation. (A) ROC curve for the crisis prediction task. (B) Precision-recall curve for the crisis prediction task with the same characteristics as (A).

ROC: receiver operating characteristic; ML: machine learning; RRT: renal replacement therapy.

Table 4. Predictive performance of initiation of mechanical ventilation.

Data	Method	F1	Precision	Recall	ROC AUC	PR AUC
Time series	LSVC	0.15	0.09	0.73	0.73	0.07
	XGB	0.10	0.05	0.79	0.71	0.06
	Multi-neural network	0.14	0.08	0.77	0.70	0.05
Static	LSVC	0.09	0.05	0.89	0.62	0.05
	XGB	0.08	0.04	0.88	0.63	0.05
	Multi-neural network	0.07	0.04	0.68	0.52	0.03
Time series & static	LSVC	0.11	0.06	0.76	0.64	0.05
	XGB	0.10	0.06	0.70	0.67	0.05
	Multi-neural network	0.12	0.07	0.83	0.69	0.04

LSVC: linear support vector machine classification; XGB: eXtreme gradient boosting; ROC: receiver operating characteristic; AUC: area under the curve; PR: precision-recall.

Table 5. List of contributing variables for each predictive model.

No.	In-hospital mortality	Initiation of acute RRT	Initiation of mechanical ventilation
1	Lactate	Sex	Age
2	Blood urea nitrogen	Dose of furosemide (mg)	Amount of urine output
3	Platelet	Potassium	Diastolic blood pressure
4	Creatinine	Furosemide dosing volume (mL)	Dose of furosemide (mg)
5	Age	Creatinine	SPO ₂

The contributions to each prediction model were calculated as follows: For in-hospital mortality prediction, the LSVC model with data imputed based on a combination of time series and static data was used. For the initiation of acute RRT prediction, a multi-neural network model with data imputed as static data was used. To predict the initiation of mechanical ventilation, the LSVC model with data imputed as time-series data was used.

LSVC: linear support vector machine classification; XGB: eXtreme gradient boosting; RRT: renal replacement therapy; SPO₂: oxygen saturation of peripheral artery.

as the most significant predictors of in-hospital mortality, dialysis, and ventilation, respectively. These findings are consistent with recent evidence suggesting that elderly

patients exhibit worse outcomes following AKI and the male sex is associated with lower survival rates.¹⁷ The association between systolic blood pressure and in-hospital mortality also agrees with the results of previous studies.¹⁸

A small variation was observed in the accuracy of the ROC AUCs for the three prediction tasks. This could be attributed to the variations in dataset sizes and the imbalanced distribution of outcomes within the dataset. Specifically, the number of data points used for the mechanical ventilation prediction task was the lowest, comprising just under 2000 instances. Previous studies have indicated that prediction models utilizing machine learning demonstrate improved accuracy when trained on larger datasets for ICU cases involving mortality and complications.¹⁹ Hence, the disparity in dataset sizes could have influenced the results. Additionally, the distribution of outcomes for acute RRT and mechanical ventilation predictions was imbalanced, constituting 1.1% and 2.8% of the data, respectively. Imbalanced data are known to pose challenges for accurate classification, particularly when the number of instances is small.²⁰ Consequently, this factor may have impacted the overall prediction accuracy.

One study applied deep learning to structured and unstructured patient data obtained from EMR and developed a risk prediction model to predict 30-day readmission in patients with HF.²¹ The developed model performed moderately well, with an AUC of 0.705, compared to previously published models. Although only structured data were used in our study, its accuracy may be improved by including both structured and unstructured data as features. Further, although other studies have been conducted to predict RRT, most of them were conducted on patients

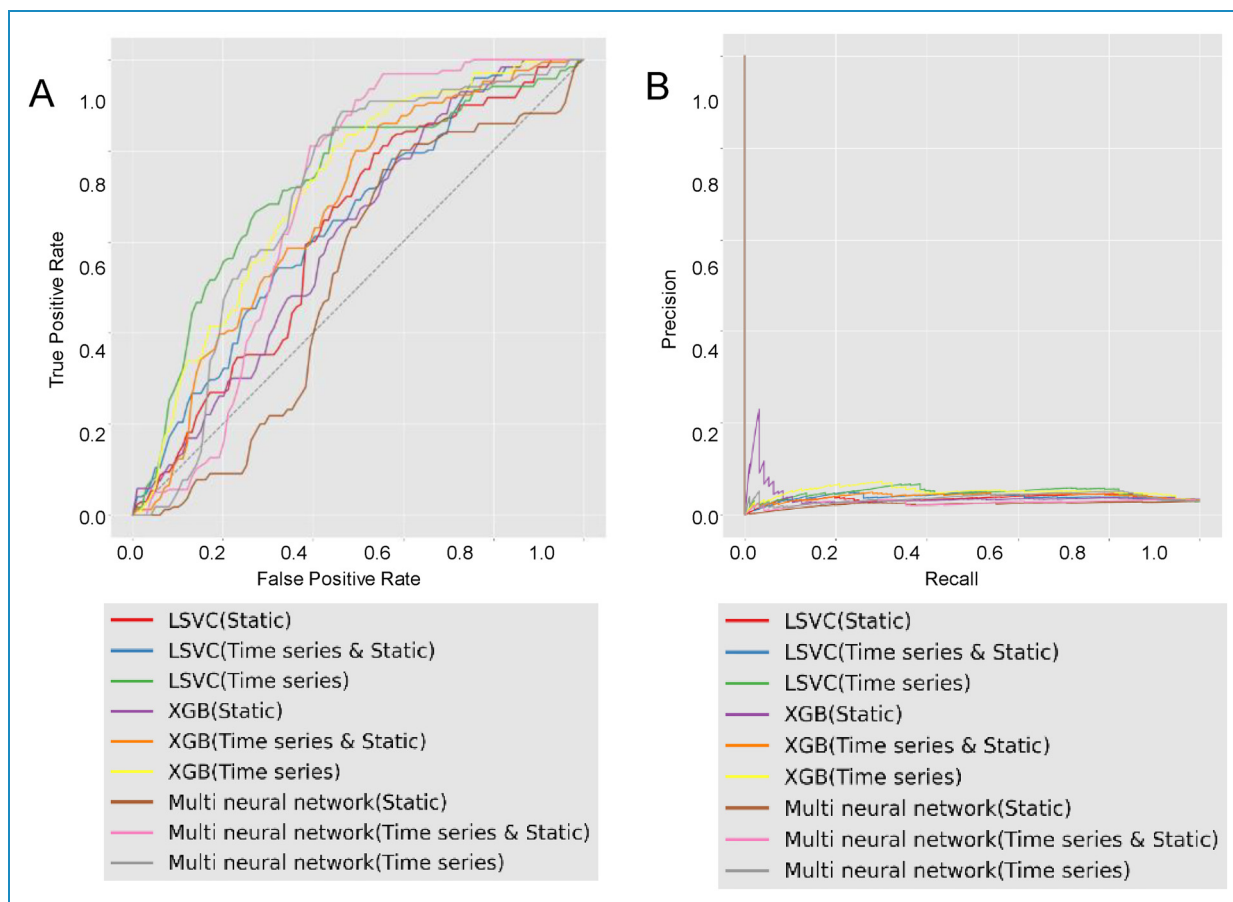


Figure 4. Final model performance. The best ROC curves for each ML model were obtained by testing various configurations in terms of feature extraction, data balancing, feature selection, and filtering for mechanical ventilation initiation. (A) ROC curve for the crisis prediction task. (B) Precision-recall curve for the crisis prediction task with the same characteristics as (A). ROC: receiver operating characteristic; ML: machine learning.

with chronic kidney disease.²² In particular, Zhang et al.²³ analyzed 2935 patients in the MIMIC-III database and 499 in a local database to develop and validate an AKI risk prediction model. The authors used five methods to develop risk prediction models for AKI: XGBoost, adaptive boosting, random forest, logistic regression, and multi-layer perception. They demonstrated that XGBoost exhibited a higher AUROC (0.88); however, they used only the clinical factors in the first 24 h after admission. AKI and AHF are dynamic diseases that require predictive models fed by dynamic parameters of the clinical course.

This study highlighted the value of ML models for prognostic prediction. The study had the following strengths and limitations. The strengths included (i) the use of data obtained from an extensive database of 20 hospitals belonging to a major Japanese hospital chain, (ii) the use of data including both static and dynamic features, and (iii) the exploration of the contributing factors of the models to reveal important predictors of several outcomes over different periods. On the other hand, the following limitations

were noted. Firstly, we used a relatively small sample size from the perspective of ML analysis. Because ML models scale very well with data, a significant improvement in diagnostic accuracy can be expected by increasing the sample size. Secondly, data replication was performed starting with furosemide administration. Because of the limited data available in the medical field, data replication is sometimes performed by treating the clinical examination as independent events.²⁴ However, this introduces the risk of biased data because data from the same cases may be replicated. Finally, the models used in this study may have been biased due to unclear reporting, assumptions made to manage missing values, and model overfitting.

Conclusions

We developed ML models to predict in-hospital mortality, initiation of acute RRT, and mechanical ventilation using data obtained from patients with AHF starting furosemide therapy after ICU admission. Our results suggested that

ML models can help predict the clinical outcomes of patients with AHF. However, the optimal models may differ depending on the outcome of interest. Further research is required to determine the generalizability of these conclusions to other populations and settings.

Acknowledgements: The authors thank the 20 hospitals (Supplemental Table S1) for contributing to the Tokushukai Medical Database and Mr Seiichiro Ishii for technical assistance.

Contributorship: TK was responsible for the manuscript content, including figures and tables. MI, YM, TU, and NC performed data pre-processing, feature engineering, model construction, hyperparameter tuning, and evaluation metric calculations. MI contributed to the preparation and revision of the manuscript. All the authors have read and approved the final version of the manuscript.

Consent to participate: Not applicable.

Consent for publication: Not applicable.

Code availability: Not applicable.

Declaration of conflicting interests: The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Ethics approval: This study was approved by the Institutional Review Board of the Shonan Kamakura General Hospital. The hospital review committee waived the need for a review because of the anonymity of the data collected from the database and the lack of a link to individual patient information.

Funding: The authors(s) received no financial support for the research, authorship, or publication of this article.

Guarantor statement: Dr Tadashi Kamio was responsible for the content of this manuscript, including the data, analysis, and draft and final versions.

ORCID iD: Tadashi Kamio  <https://orcid.org/0000-0002-8987-0103>

Supplemental material: Supplemental material for this article is available online.

References

- Ziaecian B and Fonarow GC. Epidemiology and aetiology of heart failure. *Nat Rev Cardiol* 2016; 13: 368–378.
- Tsutsui H, Ide T, Ito H, et al. JCS/JHFS 2021 guideline focused update on diagnosis and treatment of acute and chronic heart failure. *J Card Fail* 2021; 27: 1404–1444.
- Choi KH, Choi JO, Jeon ES, et al. Guideline-directed medical therapy for patients with heart failure with midrange ejection fraction: a patient-pooled analysis from the Kor HF and Kor AHF registries. *J Am Heart Assoc* 2018; 7: e009806.
- Grewal D, Partow-Navid R, Garcia D, et al. Role of guideline directed medical therapy doses and optimization in patients hospitalized with decompensated systolic heart failure. *Am J Card* 2021; 151: 64–69.
- Handelman GS, Kok HK, Chandra RV, et al. eDoctor: machine learning and the future of medicine. *J Intern Med* 2018; 284: 603–619.
- Olsen CR, Mentz RJ, Anstrom KJ, et al. Clinical applications of machine learning in the diagnosis, classification, and prediction of heart failure. *Am Heart J* 2020; 229: 1–17.
- Tabassian M, Sunderji I, Erdei T, et al. Diagnosis of heart failure with preserved ejection fraction: machine learning of spatiotemporal variations in left ventricular deformation. *J Am Soc Echocardiogr* 2018; 31: 1272–1284.e9.
- Jing L, Ulloa Cerna AE, Good CW, et al. A machine learning approach to management of heart failure populations. *JACC Heart Fail* 2020; 8: 578–587.
- Angraal S, Mortazavi BJ, Gupta A, et al. Machine learning prediction of mortality and hospitalization in heart failure with preserved ejection fraction. *JACC Heart Fail* 2020; 8: 12–21.
- Bazoukis G, Stavrakis S, Zhou J, et al. Machine learning versus conventional clinical methods in guiding management of heart failure patients—a systematic review. *Heart Fail Rev* 2021; 26: 23–34.
- Wang K, Li P, Chen L, et al. Impact of the Japanese diagnosis procedure combination-based payment system in Japan. *J Med Syst* 2010; 34: 95–100.
- Iwagami M, Moriya H, Doi K, et al. Seasonality of acute kidney injury incidence and mortality among hospitalized patients. *Nephrol Dial Transplant* 2018; 33: 1354–1362.
- Miyake K, Iwagami M, Ohtake T, et al. Association of pre-operative chronic kidney disease and acute kidney injury with in-hospital outcomes of emergency colorectal surgery: a cohort study. *World J Emerg Surg* 2020; 15: 22.
- Radhachandran A, Garikipati A, Zelin NS, et al. Prediction of short-term mortality in acute heart failure patients using minimal electronic health record data. *BioData Min* 2021; 14: 23.
- Mpanya D, Celik T, Klug E, et al. Predicting mortality and hospitalization in heart failure using machine learning: a systematic literature review. *Int J Cardiol Heart Vasc* 2021; 34: 100773.
- Chen JH and Asch SM. Machine learning and prediction in medicine – beyond the peak of inflated expectations. *N Engl J Med* 2017; 376: 2507–2509.
- Abdel-Rahman EM, Turgut F, Gautam JK, et al. Determinants of outcomes of acute kidney injury: clinical predictors and beyond. *J Clin Med* 2021 Mar 11; 10: 1175. doi:10.3390/jcm10061175
- Parissis JT, Ikonomidis I, Rafouli-Stergiou P, et al. Clinical characteristics and predictors of in-hospital mortality in acute heart failure with preserved left ventricular ejection fraction. *Am J Cardiol* 2011; 107: 79–84.
- Shillan D, Sterne JAC, Champneys A, et al. Use of machine learning to analyse routinely collected intensive care unit data: a systematic review. *Crit Care* 2019 Aug 22; 23: 284. doi:10.1186/s13054-019-2564-9

20. Kaur H, Pannu H and Malhi A. A systematic review on imbalanced data challenges in machine learning. *Appl Solut ACM Comput Surveys (CSUR)* 2019; 52: 1–36.
 21. Golas SB, Shibahara T, Agboola S, et al. A machine learning model to predict the risk of 30-day readmissions in patients with heart failure: a retrospective analysis of electronic medical records data. *BMC Med Inform Decis Mak* 2018; 18: 44.
 22. Dovgan E, Gradišek A, Luštrek M, et al. Using machine learning models to predict the initiation of renal replacement therapy among chronic kidney disease patients. *PLoS One* 2020; 15: e0233976.
 23. Zhang X, Chen S, Lai K, et al. Machine learning for the prediction of acute kidney injury in critical care patients with acute cerebrovascular disease. *Ren Fail* 2022; 44: 43–53.
 24. Kwon J, Kim KH, Jeon KH, et al. Artificial intelligence algorithm for predicting mortality of patients with acute heart failure. *PLoS One* 2019 Jul 8; 14: e0219302. doi:10.1371/journal.pone.0219302
-