


Machine learning-based model for predicting 1 year mortality of hospitalized patients with heart failure

Takeshi Tohyama^{1,2*} , Tomomi Ide^{2*}, Masataka Ikeda², Hidetaka Kaku², Nobuyuki Enzan², Shouji Matsushima², Kouta Funakoshi¹, Junji Kishimoto¹, Koji Todaka¹ and Hiroyuki Tsutsui²

¹Center for Clinical and Translational Research of Kyushu University Hospital, Higashi-ku, Fukuoka-shi, Fukuoka, Japan; and ²Department of Cardiovascular Medicine, Faculty of Medical Sciences, Kyushu University, Higashi-ku, Fukuoka-shi, Fukuoka, Japan

Abstract

Aims Individual risk stratification is a fundamental strategy in managing patients with heart failure (HF). Artificial intelligence, particularly machine learning (ML), can develop superior models for predicting the prognosis of HF patients, and administrative claim data (ACD) are suitable for ML analysis because ACD is a structured database. The objective of this study was to analyse ACD using an ML algorithm, predict the 1 year mortality of patients with HF, and finally develop an easy-to-use prediction model with high accuracy using the top predictors identified by the ML algorithm.

Methods and results Machine learning-based prognostic prediction models were developed from the ACD on 10 175 HF patients from the Japanese Registry of Acute Decompensated Heart Failure with 17% mortality during 1 year follow-up. The top predictors for prognosis in HF were identified by the permutation feature importance technique, and an easy-to-use prediction model was developed based on these predictors. The c-statistics and Brier scores of the developed ML-based models were compared with those of conventional risk models: Seattle Heart Failure Model (SHFM) and Meta-Analysis Global Group in Chronic Heart Failure (MAGGIC). A voting classifier algorithm (ACD-VC) achieved the highest c-statistics among the six ML algorithms. The permutation feature importance technique enabled identification of the top predictors such as Barthel index, age, body mass index, duration of hospitalization, last hospitalization, renal disease, and non-loop diuretics use (feature importance values were 0.054, 0.025, 0.010, 0.005, 0.005, 0.004, and 0.004, respectively). Upon combination of some of the predictors that can be assessed from a brief interview, the Simple Model by ARTificial intelligence for HF risk stratification (SMART-HF) was established as an easy-to-use prediction model. Compared with the conventional models, SMART-HF achieved a higher c-statistic {ACD-VC: 0.777 [95% confidence interval (CI) 0.751–0.803], SMART-HF: 0.765 [95% CI 0.739–0.791], SHFM: 0.713 [95% CI 0.684–0.742], MAGGIC: 0.726 [95% CI 0.698–0.753]} and better Brier scores (ACD-VC: 0.121, SMART-HF: 0.124, SHFM: 0.139, MAGGIC: 0.130).

Conclusions The ML model based on ACD predicted the 1 year mortality of HF patients with high accuracy, and SMART-HF along with the ML model achieved superior performance to that of the conventional risk models. The SMART-HF model has the clear merit of easy operability even by non-healthcare providers with a user-friendly online interface (<https://hfriskcalculator.herokuapp.com/>). Risk models developed using SMART-HF may provide a novel modality for risk stratification of patients with HF.

Keywords Artificial intelligence; Heart failure; Machine learning; Risk prediction

Received: 24 February 2021; Revised: 22 July 2021; Accepted: 28 July 2021

*Correspondence to: Takeshi Tohyama, Center for Clinical and Translational Research of Kyushu University Hospital, 3-1-1 Maidashi, Higashi-ku, Fukuoka-shi, Fukuoka 812-8582, Japan. Tel: +81-92-642-6054; Fax: +81-92-642-5374. Email: tohyama@cardiol.med.kyushu-u.ac.jp

Tomomi Ide, Department of Cardiovascular Medicine, Faculty of Medical Sciences, Kyushu University, 3-1-1 Maidashi, Higashi-ku, Fukuoka-shi, Fukuoka 812-8582, Japan. Tel: +81-92-642-5360; Fax: +81-92-642-5374. Email: tomomi_i@cardiol.med.kyushu-u.ac.jp

Introduction

Heart failure (HF) is a growing public health issue, with a prevalence that increases with age.¹ Approximately 0.4–2.2% of the population in industrialized countries suffer from HF.² In a meta-study of 39 372 patients from 30 studies with a 2.5 year median follow-up, comprising the largest available database of HF patients, Pocock *et al.* found that the mortality rate was 40.2%.³ The high morbidity and mortality are burdening the global economy with an estimated \$108 billion annually.⁴ Risk stratification and mortality assessment for HF patients provide not only fundamental strategies for clinical decision-making^{5,6} but also practical information for health policy and insurance services.

Several researchers have developed risk score models to stratify HF patients, such as the Seattle Heart Failure Model (SHFM) and Meta-Analysis Global Group in Chronic Heart Failure (MAGGIC).^{3,7,8} Nevertheless, Allen *et al.* reported that these models do not necessarily predict the mortality of individual HF patients.⁹ Hence, a superior prediction model is imperative for more accurate mortality prediction for HF patients.

To predict HF patient mortality more accurately, artificial intelligence, particularly machine learning (ML), may be a useful tool because ML algorithms can improve accuracy by analysing large volumes of medical data. Recently, several studies have demonstrated that ML approaches are superior to conventional risk models.^{10–12} For ML analysis, administrative claim data (ACD) are suitable because they are structured and can be directly fed to the ML algorithm. Desai *et al.* recently reported that an ML analysis to predict the mortality of HF patients based on ACD in the USA presented a satisfactory c-statistic of 0.727.¹³ However, a major concern regarding the versatility of ML analyses is that their utility depends on the quality and quantity of the variables in the input data. In particular, it is difficult to universally apply ACD-based ML models for risk stratification of HF on other databases owing to variation in ACD variables, indicating that ML model performance can be inferior to that of conventional methods under conditions of insufficient input data. Therefore, an ML-based prediction model was required that could be universally applied in practice.

In Japan, the Diagnosis Procedure Combination Casemix patient classification system is used for national ACD. As ACD are used for insurance reimbursement, they contain basic patient profiles, diagnoses with International Classification of Diseases 10th revision (ICD-10) codes, last hospitalization at individual facilities, activities of daily living (Barthel index), medications, and medical procedures that are usually entered into the database by hospital clerical staff; they do not contain detailed clinical variables related to HF, such as vital, laboratory, and physiological data (the details of the variables included in the Japanese ACD are presented in the Methods section and Supporting Information, *Table S1*).

Hence, the prediction model developed from ML analysis based on the Japanese ACD might have the advantage of operability by non-healthcare providers without any professional medical assessment.

In this study, we aimed to predict the 1 year mortality of HF patients using ML algorithms and the Japanese ACD and to compare the model performance with that of conventional risk models. We further sought to identify the top predictors by using the permutation feature importance technique and establish an easy-to-use ML-based prediction model with high prediction accuracy equivalent or superior to that of conventional risk models.

Methods

Trial and participants

The data utilized in this study were obtained from the Japanese Registry of Acute Decompensated Heart Failure (JROADHF).¹⁴ Across Japan, 128 facilities enrolled in the JROADHF study were selected by random cluster sampling. The inclusion criterion was acute HF in patients aged 16 years or older who were admitted to the enrolled facilities in 2013, excluding those with acute myocardial infarction. In this study, using the JROADHF registry and the accompanying ACD, we developed a predictive model for the patients discharged from hospitalization due to acute decompensated HF. This prognostic study complied with the Declaration of Helsinki and was approved by the institutional review board of Kyushu University Hospital (Approval Number 2020-335).

Outcome

The primary outcome of this study was the all-cause mortality within 1 year after discharge. It was based on the following premise: a known high probability of death within 1 year will prompt clinicians to plan frequent follow-ups or implement more aggressive life-saving therapies (e.g. implantable cardioverter–defibrillators or circulatory assist devices).^{5,6} We obtained the outcome data from the JROADHF follow-up survey.

Explanatory variables from administrative claim data

The JROADHF includes the Japanese ACD in addition to the general HF registry data, such as patient characteristics and prognosis survey results. As of 2014, the ACD covered approximately 70% of hospital discharges in Japan, that is, 10 million discharges per year.¹⁵ ACD have been collected during numerous projects to construct large-scale data for research

purposes^{16,17} and include fundamental data such as age, sex, hospitalization histories at the participating facilities, duration of hospitalization, discharge status, diagnoses, New York Heart Association (NYHA) functional class, medical procedures, and drug prescriptions. These data were recorded using a uniform submission format. Because of the generalized nature of the ACD, detailed clinical data related to HF such as vital, laboratory, and physiological data are not included.

The diagnoses obtained from the ACD were coded into Charlson comorbidities according to the ICD-10.^{18,19} Variables were extracted considering the Japanese HF guidelines and previously reported information regarding treatment during hospitalization and prescription at discharge. We extracted a total of 89 explanatory variables available from the ACD (Supporting Information, *Table S1*).

Machine learning model development

To extract the dataset for ML development (training and validation) and testing, the JROADHF study facilities were randomly divided in an 8:2 ratio with balancing of the number of subjects per facility. We substituted the missing values with the median or mode values, a common approach to dealing with missing values in ML algorithms.²⁰ Categorical data were converted into binary data using a label encoder or one-hot encoder. Prior to training, the data were standardized to an average of zero and a standard deviation of one.

Using the training set, we constructed six common ML algorithms: logistic lasso regression, support vector machine, random forest, gradient boosting tree, voting classifier, and neural network. The hyperparameters of each model were determined using a grid search and five-fold cross-validation with the training set. Using the obtained hyperparameters, we trained the models with the entire training set and evaluated the performance using the test set. Further details on the ML models are presented in the Supporting Information, *Text S1*. We implemented the ML models using Python Version 3.7.4 with scikit-learn Version 0.23.1, LightGBM Version 2.3.0, and TensorFlow Version 2.0.0. We evaluated the performances of these six ML models using the test set, as described in the Statistical analysis section.

Significant predictors and simple model development

The significant variables of the developed ML model that achieved the best performance were evaluated by permutation feature importance, which is defined as the decrease in the model score when a single feature value is randomly

shuffled.²¹ In this study, we evaluated the decrease in the c-statistic or area under the receiver operating characteristic curve. The permutation was repeated 100 times for each variable. Furthermore, via selection of the significant predictors of the original ML model obtained from this result, we proposed a simple model for universal utilization.

Comparison with conventional risk models

To compare the ACD-based ML models (the original ML model and the simple model) with the conventional HF risk models, SHFM and MAGGIC, we extracted the variables of those models not only from ACD but also from the JROADHF registry data (Supporting Information, *Tables S2 and S3*) because these conventional models include detailed clinical variables related to HF, which the Japanese ACD do not contain. We evaluated the performance of the conventional risk models SHFM and MAGGIC derived from the previously validated formulae^{3,7} and compared the discrimination and calibration with those of the ACD-based ML models using the test set. In addition, to evaluate the net effects of the ACD variables, we compared the ACD-based ML models with the ML models based on the variables in the conventional risk models. We pre-processed each variable in the conventional HF risk models described earlier and utilized the ML algorithm that achieved the highest performance for ML model development. Because the JROADHF registry included many elderly patients, we evaluated the prediction performance for the subgroup aged ≤ 80 years.

Statistical analysis

For model comparison, we evaluated discrimination and calibration indices. The discrimination was evaluated by plotting receiver operating characteristic curves and calculating the c-statistic,²² whereas the model calibration was assessed using a calibration plot and the Brier score.²³ The Brier score is the mean squared difference between the observed and predicted outcomes and ranges from '0' to '1', with '0' representing the best calibration. Using these evaluations, we identified the most suitable model for predicting the 1 year mortality in HF patients. To address the classification with imbalanced data (i.e. low proportions of events), we chose the prediction threshold based on the Youden index²⁴ used to classify disease status in the framework of medical diagnosis. The 95% confidence interval for the c-statistic was estimated using the DeLong method. We performed all statistical analyses using SAS 9.4 (SAS Institute, Cary, North Carolina).

Results

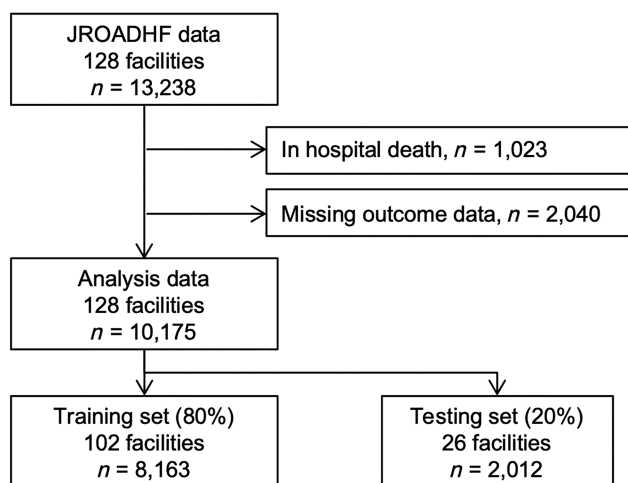
Patient characteristics and events

A total of 13 238 patients with acute decompensated HF from 128 facilities were enrolled in the JROADHF study (Figure 1). We excluded 1023 patients who died in hospital and 2040 patients without follow-up data during 1 year after discharge. Hence, this study comprised a total of 10 175 patients. The 1 year mortality was 17.0% ($n = 1727$). The baseline characteristics of the patients obtained from the JROADHF study are summarized in Table 1. The median age was 80 years (interquartile range = 70–86 years), and 45.3% of the study population were female. Patients with NYHA functional class symptoms of III or IV at discharge accounted for 9.0%, whereas 38.0% of the study population presented reduced left ventricular ejection fraction (LVEF). The JROADHF study facilities were divided into 102 facilities with 8163 patients for the training set and 26 facilities with 2012 patients for the testing set.

Machine learning model development

The discrimination and calibration of all ML models are shown in Figure 2 and Table 2. Although there are no apparent differences between the six ML models, the voting classifier achieved the highest *c*-statistic of 0.777 [95% confidence interval 0.751–0.803] for 1 year mortality. Each ML model was well calibrated, as shown in Figure 2B, and the minimum value of the Brier score was achieved by the voting classifier. We selected the voting classifier as the most promising approach among the six ML algorithms. We refer to the

Figure 1 Data extraction from the Japanese Registry of Acute Decompensated Heart Failure (JROADHF) study.



developed voting classifier model consisting of 89 ACD variables as the ACD-based voting classifier (ACD-VC) model in the following sections.

Significant predictors and development of the simplified model

We identified the most significant predictors via the permutation feature importance technique to translate the ACD-VC model into a universally applicable prediction model (Figure 3). We also assembled a simplified ML model called the Simple Model by ARTificial intelligence for HF risk stratification (SMART-HF), incorporating high-ranking and readily available variables, namely, the Barthel index, age, body mass index, duration of hospitalization, last hospitalization, furosemide dose, non-loop diuretics, and gender. The SMART-HF achieved a *c*-statistic of 0.765 [95% confidence interval 0.739–0.791] and a Brier score of 0.124.

Table 1 Baseline patient characteristics

Characteristics	<i>N</i> = 10 175
Age, years	80 [70–86]
Female, <i>n</i> (%)	4606 (45.3)
Body mass index, kg/m ²	22.4 [19.9–25.3]
Aetiologies, <i>n</i> (%)	
Ischaemic	3657 (35.9)
Hypertensive disease	2578 (25.3)
Valvular disease	3587 (35.3)
Cardiomyopathy	1571 (15.4)
Comorbidities, <i>n</i> (%)	
Hypertension	7411 (72.8)
Diabetes mellitus	3637 (35.7)
Dyslipidaemia	3329 (32.7)
COPD	663 (6.5)
Chronic kidney disease	3980 (39.1)
NYHA class on admission, <i>n</i> (%)	
I or II	1566 (15.6)
III or IV	8461 (84.4)
NYHA class at discharge, <i>n</i> (%)	
I or II	8800 (91.0)
III or IV	868 (9.0)
Blood pressure, mmHg	
Systolic	137 [118–160]
Diastolic	78 [65–92]
Pulse rate, b.p.m.	88 [73–107]
LVEF, %	46 [32–61]
LVEF < 40%, <i>n</i> (%)	3444 (38.0)
Serum sodium, mmol/L	139 [137–141]
Serum potassium, mmol/L	4.3 [4.0–4.7]
Estimated GFR, mL/min/1.73 m ²	44.1 [29.4–59.5]
Haemoglobin, g/dL	11.5 [10.1–13.2]
Barthel index	100 [65–100]
Medication, <i>n</i> (%)	
Diuretics	8408 (82.8)
RAS inhibitor	6140 (60.5)
Beta-blockers	6099 (60.1)
MRA	4512 (44.4)

COPD, chronic obstructive pulmonary disease; GFR, glomerular filtration rate; LVEF, left ventricular ejection fraction; MRA, mineralocorticoid receptor antagonist; NYHA, New York Heart Association; RAS, renin–angiotensin–aldosterone system.

Values are medians [interquartile ranges] or *n* (%).

Figure 2 (A) Comparison of the receiver operating characteristic curves of the six machine learning models when applied to the test set. All six models have similar predictive performance. (B) Calibration plot of six models. The data points are subdivided into deciles of predicted probability. Each mean observed mortality is plotted against the mean model probability. Perfect model calibration corresponds with the $y = x$ line.

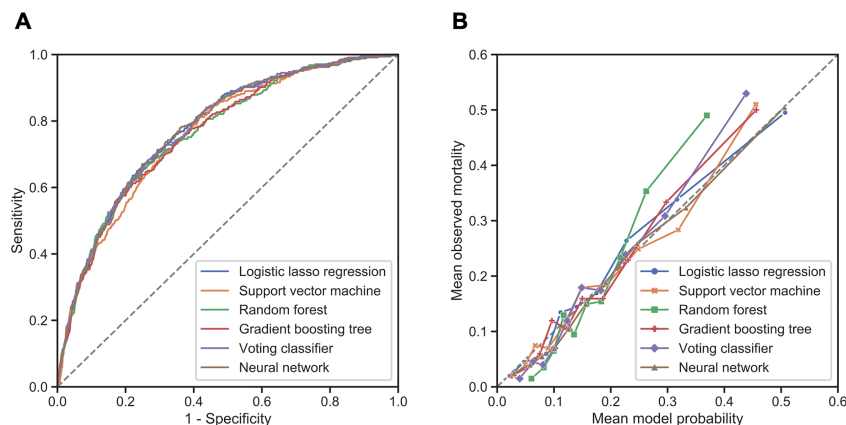


Table 2 Predictive performances of the machine learning models and conventional risk models

Model	Accuracy	Sensitivity	Specificity	c-statistic	Brier score
Logistic lasso regression	73.3	66.8	74.7	0.776 [0.750–0.802]	0.121
Support vector machine	69.6	69.7	69.6	0.764 [0.738–0.791]	0.123
Random forest	72.4	66.5	73.6	0.767 [0.740–0.794]	0.125
Gradient boosting tree	75.0	61.0	77.9	0.765 [0.738–0.792]	0.123
Voting classifier (ACD-VC)	70.7	71.1	70.6	0.777 [0.751–0.803]	0.121
Neural network	73.8	66.8	75.2	0.776 [0.750–0.802]	0.121
SMART-HF	71.2	67.9	71.8	0.765 [0.739–0.791]	0.124
SHFM	80.5	26.6	91.7	0.713 [0.684–0.742]	0.139
MAGGIC	70.4	58.7	72.9	0.726 [0.698–0.753]	0.130
SHFM-VC	73.5	64.5	75.3	0.760 [0.734–0.785]	0.125
MAGGIC-VC	64.8	75.1	62.6	0.754 [0.728–0.781]	0.126

ACD-VC, voting classifier based on administrative claim data; MAGGIC, meta-analysis global group in chronic heart failure risk score; MAGGIC-VC, voting classifier based on the variables in MAGGIC; NYHA, New York Heart Association; SHFM, Seattle Heart Failure Model; SHFM-VC, voting classifier based on the variables in SHFM; SMART-HF, simple model by artificial intelligence for heart failure risk stratification.

Values are means [95% confidence intervals].

Comparison with conventional risk models

To validate the developed ACD-VC and SMART-HF models, we compared them with conventional risk models. As shown in *Figure 4* and *Table 2*, the c-statistics of the conventional models were 0.713 [0.684–0.742] and 0.726 [0.698–0.753] for SHFM and MAGGIC, respectively, indicating that the performance of the ML models was significantly superior to that of the conventional HF risk models. The Brier scores of the ML models were better than those of SHFM and MAGGIC (0.139 and 0.130, respectively). The calibration plot in *Figure 4B* shows that the SHFM underestimated the mortality risk. In addition, we compared the developed ML models to the modified conventional risk models with the present voting classifier algorithm. Both conventional risk models improved their discrimination (c-statistic values: SHFM-VC, 0.760;

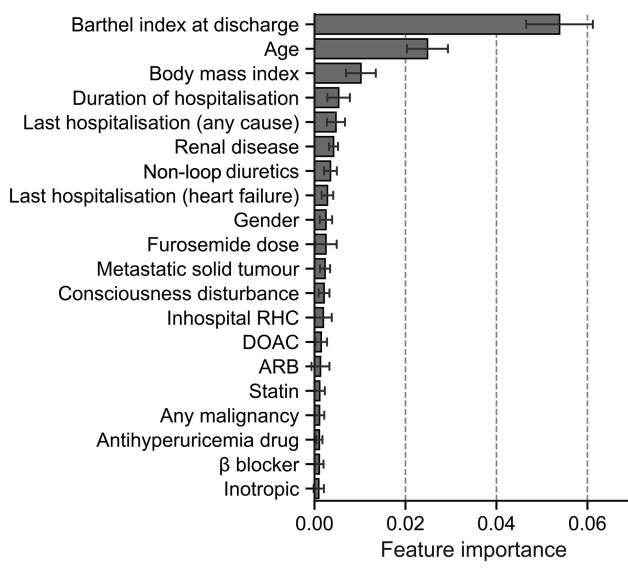
MAGGIC-VC, 0.754) and calibration (Brier scores: SHFM-VC, 0.125; MAGGIC-VC, 0.126) when recalculated with an ML algorithm. However, the performance of the developed ACD-VC model and SMART-HF remained equal or superior to that of both of the modified conventional models with the VC algorithm.

Furthermore, we analysed the predictive performance of the developed ML models for the subgroup aged ≤ 80 years. The results demonstrated a similar trend (Supporting Information, *Table S4*). Compared with the conventional models, both ACD-based ML models achieved higher c-statistics (ACD-VC, 0.777; SMART-HF, 0.768; SHFM, 0.728; MAGGIC, 0.725; SHFM-VC, 0.764; MAGGIC-VC, 0.750) and lower Brier scores (ACD-VC, 0.084; SMART-HF, 0.086; SHFM, 0.088; MAGGIC, 0.092; SHFM-VC, 0.087; MAGGIC-VC, 0.086).

Discussion

In this paper, we showed that the performance of the ML approach using the Japanese ACD is superior to that of conventional risk models. Furthermore, combining some of the top variables uncovered by the ML analysis, we developed a new easy-to-use prediction model, SMART-HF. The performance of SMART-HF was equivalent or superior to those of the conventional risk models. This new model requires a

Figure 3 Top 20 significant predictors of the developed machine learning model for 1 year mortality in heart failure patients calculated by permutation feature importance. ARB, angiotensin II receptor blocker; DOAC, direct oral anticoagulant; RHC, right heart catheterization.

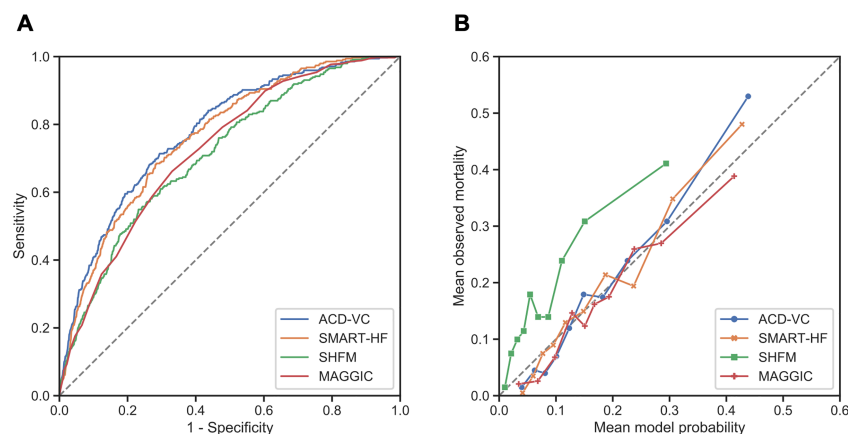


small number of easily available variables, which can be assessed from a brief interview even by non-healthcare providers.

All six ML algorithms considered in this study are fully established and commonly used for artificial intelligence prediction tasks. Among them, the voting classifier algorithm, which is designed to overcome the limitations of individual classifiers,²⁵ achieved the best performance in our study by a narrow margin. However, the difference between the ML models was not significant for predictive modelling of mortality in HF patients.

We then sought to develop a refined prediction model using the top predictors identified by the ACD-VC model. The permutation feature importance technique is an effective way to uncover the predictive power of each variable in an ML model. This technique showed that the top predictive variables were the Barthel index, age, body mass index, duration of hospitalization, last hospitalization, renal disease, diuretics use, and gender. In particular, the Barthel index played the most important role in the ACD-VC model performance. The Barthel index represents activities of daily living and frailty, and it is reportedly critical for predicting HF patient mortality.^{26,27} This result agrees with the ML-based findings of Desai *et al.*, who also identified frailty as the most significant predictor of 1 year mortality in HF patients.¹³ In the conventional risk models, the potential relationships between symptoms and daily activity are evaluated using NYHA classification. However, the Barthel index, which aids in accurate evaluation of daily activity, may be more useful than the NYHA classification for mortality prediction. Given that these variables incorporated in SMART-HF—excluding the Barthel index—are common in pre-existing models such as SHFM and MAGGIC,^{3,7} the high accuracy of prognostic prediction

Figure 4 (A) Comparison of ACD-based models (ACD-VC and SMART-HF) with the conventional risk models of SHFM and MAGGIC when applied to the test set. (B) Calibration plot of the four models. Data points are subdivided into deciles of predicted probability. Each mean observed mortality is plotted against the mean model probability. ACD, administrative claim data; ACD-VC, ACD-based voting classifier; MAGGIC, Meta-Analysis Global Group in Chronic Heart Failure; SHFM, Seattle Heart Failure Model; SMART-HF, Simple Model by ARTificial intelligence for HF risk stratification.



in SMART-HF is dependent on the Barthel index, and this underlines the significance of frailty in HF, as uncovered by the permutation feature importance technique.

We demonstrated the superiority of the ACD-VC and SMART-HF models over the conventional SHFM and MAGGIC risk models, as shown in *Figure 3*. The c-statistics of the SHFM and MAGGIC models were previously reported as 0.729 and 0.734, respectively,^{7,28} which are consistent with the present results, and a similar trend was evident in the results for the subgroup aged ≤ 80 years. However, given that their superiorities in prediction accuracy might be marginal, the higher c-statistics in the ACD-VC and SMART-HF models do not have much impact on clinical utility for predicting the prognosis of HF patients. Indeed, several researchers have reported that the ML approaches predicted mortality with over 0.8 c-statistics.^{29,30} Nevertheless, the SMART-HF model has two distinct advantages over conventional risk models. First, the variables in SMART-HF are readily available because the Japanese ACD has no clinical variables. Most previous models were developed based on the registry and electronic health record containing detailed clinical data, such as echocardiographic and blood biochemistry data. Some clinical variables incorporated in pre-existing models are often lacking in clinical practice, and the lack of these parameters has a critical negative effect on prediction accuracy. In contrast, SMART-HF was developed only on the ACD and could be applied to the information obtained from a brief interview. As most of the variables in the Japanese ACD were generally entered by hospital clerical staff, SMART-HF can be easily operated even by non-healthcare providers. Second, a prediction model based on ACD would prove useful for a national medical policy or an insurance market. As there are many such claim databases available with the government or insurance companies, appropriate decisions can be made for medical policy and actuarial services by utilizing ACD-based prediction models. This advantage may enable those in the government and actuarial services to potentially optimize the medical services in accordance with the prediction model based on ACD and reduce the increasing economic burden. These advantages will further promote research on ML-based modelling using ACD in many other diseases in the future.

SMART-HF, developed from the ACD-VC model, can yield an accurate prediction of the 1 year prognosis in patients with HF through the inputting of easily available variables, although the original ACD-VC could vary between databases, and the performance of the ML approach largely depends on the quality of the ACD. To facilitate the usability of SMART-HF by healthcare and non-healthcare providers, we have launched an online calculator implementing SMART-HF with a user-friendly interface (<https://hfriskcalculator.herokuapp.com/>). This online system provides a novel modality for risk stratification of HF patients.

Limitations

This study has some limitations. First, Japan has the largest proportion of older adults (>65 years old) worldwide, because of which it is referred to as a 'super-ageing' society (<https://population.un.org/wpp/Download/Standard/Population/>). Therefore, the overall profile of patients with HF in the JROADHF also showed a greater degree of ageing than that observed in the previous cohort database,¹⁴ and the median age of the participants in this study was 80 years. This difference in the ageing population between the JROADHF and previous studies might lead to an increased proportion of HF patients with preserved ejection fraction and average of LVEF in the JROADHF. Although the 'ageing society' phenomenon is progressing worldwide, especially in many developed countries, SMART-HF should be carefully utilized with consideration for the age distribution of the population in each region. Second, we excluded 2040 subjects with missing prognosis data for 1 year, which can potentially cause selection bias. Third, the Barthel index was not documented in many previous studies, and thus, further investigation will be required to validate the efficacy of SMART-HF by analysing ulterior databases using SMART-HF in the future. Finally, the ACD-VC and SMART-HF models showed high prediction accuracy without the requirement for any detailed clinical parameters, which have been proven to play pivotal roles in predicting the prognosis of patients with HF, such as in terms of vital signs, echocardiography, laboratory data, and optimal medical treatment. This might be because the accessibility to medical care is firmly ensured, and medical treatment for HF is well optimized in most HF patients in accordance with frequently documented clinical parameters and guidelines for HF. This medical scenario in Japan might lead to lower predictive values of LVEF and optimal medical treatment; thus, the lack of fully examined clinical parameters and guideline-based optimal medical treatment can negatively affect the prediction accuracy of SMART-HF in some populations where clinical parameters are not well assessed and medical therapy is not optimized.

Conclusions

The voting classifier model developed using the Japanese ACD can predict 1 year mortality more reliably than conventional risk models. Further, the SMART-HF model derived from the ACD-based model delivers high prediction performance and only requires a small number of readily available variables with easy operability. Our new prediction model is a useful tool for accurately predicting the 1 year mortality and optimizing the management of HF patients.

Acknowledgements

We thank the JROADHF Investigators (<https://www.jroadhf.jp/about/researcher.html>) for data collection.

Conflict of interest

T.T., T.I., M.I., K.F., H.K., N.E., S.M., J.K., and K.T. have nothing to declare. H.T. reports personal fees from MSD, Astellas, Pfizer, Bristol-Myers Squibb, Otsuka Pharmaceutical, Daiichi-Sankyo, Mitsubishi Tanabe Pharma, Nippon Boehringer Ingelheim, Takeda Pharmaceutical, Bayer Yakuhin, Novartis Pharma, Kowa Pharmaceutical, Teijin Pharma, Medical Review Co., and Japanese Journal of Clinical Medicine; non-financial support from Actelion Pharmaceuticals, Mitsubishi Tanabe Pharma, Nippon Boehringer Ingelheim, Daiichi-Sankyo, IQVIA Services Japan, and Omron Healthcare Co.; and grants from Astellas, Novartis Pharma, Daiichi-Sankyo, Takeda Pharmaceutical, Mitsubishi Tanabe Pharma, and Teijin Pharma, MSD, outside the submitted work.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: this work was supported by Health Sciences Research Grants from the Japanese Ministry of Health, Labour and Welfare (Comprehensive Research on Cardiovascular Diseases); Japan Agency for Medical Research and

Development (AMED) (grant numbers 19ek0109367h0002, 20ek0109367h0003); and the Japan Society for Promotion of Science (JSPS), KAKENHI (grant number 19K17529).

Author contributions

T.T., T.I., M.I., and H.T. contributed to the conception or design of the work. N.E., H.K., and T.I. contributed to the acquisition of data. T.T., M.I., and T.I. contributed to the analysis or interpretation of data for the work. T.T., T.I., and M.I. drafted the manuscript and prepared the figures. J.K. and K.F. made statistical proposals. K.T. and H.T. critically revised the manuscript. All authors gave final approval and agreed to be accountable for all aspects of the work ensuring integrity and accuracy.

Supporting information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Table S1. Variables of administrative claims data (ACD) used for predicting one-year mortality in heart failure patients.

Text S1. Development of six machine learning models.

Table S2. SHFM variables.

Table S3. MAGGIC variables.

Table S4. Comparison of predictive performance with conventional risk models for patients aged 80 years or less.

References

- Savarese G, Lund LH. Global public health burden of heart failure. *Card Fail Rev.* 2017; **03**: 7.
- Go AS, Mozaffarian D, Roger VL, Benjamin EJ, Berry JD, Blaha MJ, Dai S, Ford ES, Fox CS, Franco S, Fullerton HJ, Gillespie C, Hailpern SM, Heit JA, Howard VJ, Huffman MD, Judd SE, Kissela BM, Kittner SJ, Lackland DT, Lichtman JH, Lisabeth LD, Mackey RH, Magid DJ, Marcus GM, Marelli A, Matchar DB, McGuire DK, Mohler ER 3rd, Moy CS, Mussolino ME, Neumar RW, Nichol G, Pandey DK, Paynter NP, Reeves MJ, Sorlie PD, Stein J, Towfighi A, Turan TN, Virani SS, Wong ND, Woo D, Turner MB, American Heart Association Statistics Committee and Stroke Statistics Subcommittee. Heart disease and stroke statistics—2014 update: a report from the American Heart Association. *Circulation* 2014; **129**: e28–e292.
- Pocock SJ, Ariti CA, McMurray JJV, Maggioni A, Køber L, Squire IB, Swedberg K, Dobson J, Poppe KK, Whalley GA, Doughty RN, Meta-Analysis Global Group in Chronic Heart Failure. Predicting survival in heart failure: A risk score based on 39 372 patients from 30 studies. *Eur Heart J.* 2013; **34**: 1404–1413.
- Cook C, Cole G, Asaria P, Jabbour R, Francis DP. The annual global economic burden of heart failure. *Int J Cardiol.* 2014; **171**: 368–376.
- Yancy CW, Jessup M, Bozkurt B, Butler J, Casey DE, Drazner MH, Fonarow GC, Geraci SA, Horwich T, Januzzi JL, Johnson MR, Kasper EK, Levy WC, Masoudi FA, McBride PE, McMurray JJ, Mitchell JE, Peterson PN, Riegel B, Sam F, Stevenson LW, Tang WH, Tsai EJ, Wilkoff BL, American College of Cardiology Foundation, American Heart
- Association Task Force on Practice Guidelines. 2013 ACCF/AHA guideline for the management of heart failure: A report of the American College of Cardiology Foundation/American Heart Association Task Force on Practice Guidelines. *J Am Coll Cardiol.* 2013; **62**: e147–e239.
- Federal T, Find R. Medicare and Medicaid programs: Hospice conditions of participation. Final rule. *Fed Regist.* 2008; **73**: 32087–32220.
- Levy WC, Mozaffarian D, Linker DT, Sutradhar SC, Anker SD, Cropp AB, Anand I, Maggioni A, Burton P, Sullivan MD, Pitt B, Poole-Wilson PA, Mann DL, Packer M. The Seattle Heart Failure Model: Prediction of survival in heart failure. *Circulation.* 2006; **113**: 1424–1433.
- Collier TJ, Pocock SJ, McMurray JJV, Zannad F, Krum H, van Veldhuisen DJ,

- Swedberg K, Shi H, Vincent J, Pitt B. The impact of eplerenone at different levels of risk in patients with systolic heart failure and mild symptoms: Insight from a novel risk score for prognosis derived from the EMPHASIS-HF trial. *Eur Heart J*. 2013; **34**: 2823–2829.
9. Allen LA, Matlock DD, Shetterly SM, Xu S, Levy WC, Portalupi LB, McIlvennan CK, Gurwitz JH, Johnson ES, Smith DH, Magid DJ. Use of risk models to predict death in the next year among individual ambulatory patients with heart failure. *JAMA Cardiol*. 2017; **2**: 435–441.
 10. Mortazavi BJ, Downing NS, Bucholz EM, Dharmarajan K, Manhapra A, Li SX, Negahban SN, Krumholz HM. Analysis of machine learning techniques for heart failure readmissions. *Circ Cardiovasc Qual Outcomes*. 2016; **9**: 629–640.
 11. Angraal S, Mortazavi BJ, Gupta A, Khera R, Ahmad T, Desai NR, Jacoby DL, Masoudi FA, Spertus JA, Krumholz HM. Machine learning prediction of mortality and hospitalization in heart failure with preserved ejection fraction. *JACC Heart Fail*. 2020; **8**: 12–21.
 12. Bazoukis G, Stavrakis S, Zhou J, Bollepalli SC, Tse G, Zhang Q, Singh JP, Armoundas AA. Machine learning versus conventional clinical methods in guiding management of heart failure patients—a systematic review. *Heart Fail Rev*. 2020; **1**–2.
 13. Desai RJ, Wang SV, Vaduganathan M, Evers T, Schneeweiss S. Comparison of machine learning methods with traditional models for use of administrative claims with electronic medical records to predict heart failure outcomes. *JAMA Netw Open*. 2020; **3**: e1918962.
 14. Ide T, Kaku H, Matsushima S, Tohyama T, Enzan N, Funakoshi K, Sumita Y, Nakai M, Nishimura K, Miyamoto Y, Tsuchihashi-Makaya M, Hatano M, Komuro I, Tsutsui H, JROADHF Investigators. Clinical characteristics and outcomes of hospitalized patients with heart failure from the large-scale Japanese Registry Of Acute Decompensated Heart Failure (JROADHF). *Circ J* 2021. <https://doi.org/10.1253/circj.CJ-20-0947>
 15. Ishikawa KB. Medical big data for research use: Current status and related issues. *Japan Med Assoc J*. 2016; **59**: 110–124.
 16. Aso S, Matsui H, Fushimi K, Yasunaga H. The effect of intraaortic balloon pumping under venoarterial extracorporeal membrane oxygenation on mortality of cardiogenic patients: An analysis using a nationwide inpatient database. *Crit Care Med*. 2016; **44**: 1974–1979.
 17. Masuda K, Chikuda H, Yasunaga H, Hara N, Horiguchi H, Matsuda S, Takeshita K, Kawaguchi H, Nakamura K. Factors affecting the occurrence of pulmonary embolism after spinal surgery: Data from the national administrative database in Japan. *Spine J*. 2012; **12**: 1029–1034.
 18. Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: Development and validation. *J Chronic Dis*. 1987; **40**: 373–383.
 19. Quan H, Sundararajan V, Halfon P, Fong A, Burnand B, Luthi J-C, Saunders LD, Beck CA, Feasby TE, Ghali WA. Coding algorithms for defining comorbidities in ICD-9-CM and ICD-10 administrative data. *Med Care*. 2005; **43**: 1130–1139.
 20. Batista GEAPA, Monard MC. An analysis of four missing data treatment methods for supervised learning. *Appl Artif Intell*. 2003; **17**: 519–533.
 21. Breiman L. Random forests. *Mach Learn*. 2001; **45**: 5–32.
 22. Fawcett T. An introduction to ROC analysis. *Pattern Recognit Lett*. 2006; **27**: 861–874.
 23. Brier GW. Verification of forecasts expressed in terms of probability. *Mon Weather Rev*. 1950; **78**: 1–3.
 24. Youden WJ. Index for rating diagnostic tests. *Cancer*. 1950; **3**: 32–35.
 25. Tumer K, Ghosh J. Error correlation and error reduction in ensemble classifiers. *Conn Sci*. 1996; **8**: 385–404.
 26. Formiga F, Chivite D, Conde A, Ruiz-Laiglesia F, Franco AG, Bocanegra CP, Manzano L, Perez-Barquero MM, RICA Investigators. Basal functional status predicts three-month mortality after a heart failure hospitalization in elderly patients - The prospective RICA study. *Int J Cardiol*. 2014; **172**: 127–131.
 27. Aimo A, Barison A, Mammini C, Emdin M. The Barthel index in elderly acute heart failure patients. Frailty matters. *Int J Cardiol*. 2018; **254**: 240–241.
 28. Khanam SS, Choi E, Son JW, Lee JW, Youn YJ, Yoon J, Lee SH, Kim JY, Ahn SG, Ahn MS, Kang SM, Baek SH, Jeon ES, Kim JJ, Cho MC, Chae SC, Oh BH, Choi DJ, Yoo BS. Validation of the MAGGIC (Meta-Analysis Global Group In Chronic Heart Failure) heart failure risk score and the effect of adding natriuretic peptide for predicting mortality after discharge in hospitalized patients with heart failure. *PLoS One*. 2018; **13**: 1–13.
 29. Ahmad T, Lund LH, Rao P, Ghosh R, Warier P, Vaccaro B, Dahlström U, O'Connor CM, Michael Felker G, Desai NR. Machine learning methods improve prognostication, identify clinically distinct phenotypes, and detect heterogeneity in response to therapy in a large cohort of heart failure patients. *J Am Heart Assoc*. 2018; **7**: 1–14.
 30. Taslimitehrani V, Dong G, Pereira NL, Panahiazar M, Pathak J. Developing EHR-driven heart failure risk prediction models using CPXR (Log) with the probabilistic loss function. *J Biomed Inform*. 2016; **60**: 260–269.