ESC HEART FAILURE
ESC Heart Failure 2025; **12**: 2200–2209
Published online 13 February 2025 in Wiley Online Library (wileyonlinelibrary.com) **DOI:** 10.1002/ehf2.15244

ORIGINAL ARTICLE

Prediction of 90 day mortality in elderly patients with acute HF from e-health records using artificial intelligence

Joconde Weller[1]*, Johann Gutton[1], Guillaume Hocquet[1], Leïla Pellet[1], Marie-José Aroulanda[2], Amélie Bruandet[3], Didier Theis[3], Fabio Boudis[3], Romain Cador[2], Pierre Zweigenbaum[4], Anne Buronfosse[1], Pascal de Groote[5,6] and Michel Komajda[2,7]

[1]*Direction of Medical Information, Prospects and Data Sciences, Hôpitaux Paris Saint-Joseph and Marie-Lannelongue, Paris, France;* [2]*Department of Cardiology, Hôpital Paris Saint-Joseph, Paris, France;* [3]*Department of Medical Information, CHU Lille, Lille, France;* [4]*LISN Research Center, Université Paris-Saclay, Paris, France;* [5]*Department of Cardiology, CHU Lille, Lille, France;* [6]*Inserm U1167, Institut Pasteur de Lille, Lille, France; and* [7]*Paris Sorbonne University, Paris, France*

Abstract

Aims Mortality risk after hospitalization for heart failure (HF) is high, especially in the first 90 days. This study aimed to construct a model automatically predicting 90 day post-discharge mortality using electronic health record (EHR) data 48 h after admission and artificial intelligence.

Methods All HF-related admissions from 2015 to 2020 in a single hospital were included in the model training. Comprehensive EHR data were collected 48 h after admission. Natural language processing was applied to textual information. Deaths were identified from the French national database. After variable selection with least absolute shrinkage and selection operator, a logistic regression model was trained. Model performance [area under the receiver operating characteristic curve (AUC)] was tested in two independent cohorts of patients admitted to two hospitals between March and December 2021.

Results The derivation cohort included 2257 admissions (248 deaths after hospitalization). The evaluation cohorts included 348 and 388 admissions (34 and 38 deaths, respectively). Forty-two independent variables were selected. The model performed well in the derivation cohort [AUC: 0.817; 95% confidence interval (CI) (0.789–0.845)] and in both evaluation cohorts [AUC: 0.750; 95% CI (0.672–0.829) and AUC: 0.723; 95% CI (0.644–0.803]), with better performance than previous models in the literature. Calibration was good: 'low-risk' (predicted mortality $\leq$8%), 'intermediate-risk' (8–12.5%) and 'high-risk' ($>$12.5%) patients had an observed 90 day mortality rate of 3.8%, 8.4% and 19.4%, respectively.

Conclusions The study proposed a robust model for the automatic prediction of 90 day mortality risk 48 h after hospitalization for decompensated HF. This could be used to identify high-risk patients for intensification of therapeutic management.

Keywords heart failure; machine learning; mortality; natural language processing; risk prediction

Received: 10 October 2024; Revised: 19 December 2024; Accepted: 26 January 2025
**Correspondence to: Joconde Weller, Direction of Medical Information, Prospects and Data Sciences, Hôpitaux Paris Saint-Joseph and Marie-Lannelongue, 185 rue Raymond Losserand 75014 Paris, France. Email: jweller@ghpsj.fr*
All the authors take responsibility for all aspects of the reliability and freedom from bias of the data presented and their discussed interpretation.

Introduction

Heart failure (HF) is a common medical condition associated with high mortality and rehospitalization rates. In 2019, there were 56.2 million prevalent cases of HF worldwide, with an age-standardized rate of 7.12 per 1000 individuals.[1] In Europe, the prevalence rate of HF in 2019 was approximately 17 per 1000.[2] In the United States, according to data from the National Health and Nutrition Examination Survey (NHANES), the prevalence rate was 23 per 1000 between 2017 and 2020 and the average incidence of hospitalized HF was 11.6 per 1000 patient years with a readmission rate of 6.6 per 1000 patient years for patients aged 55 years and older.[3]

Rates of death and rehospitalization in the first 60 to 90 days after discharge from a hospitalization for decompensated HF are high, and this period has been referred to as the 'vulnerable phase'.[4] Over this period, rates of readmission and death range from 15% to 30% and 7% to 11%, respectively.[5–8] Although advances in medical therapy for HF have had a beneficial impact on mortality rates, the number of HF-related deaths and readmissions remains high.[9]

Various approaches have been studied to reduce early readmission and/or mortality rates in patients with HF, including a personalized medical journey with coordination of care, early visits by specialist nurse, general practitioners or cardiologists and/or remote monitoring, but results are inconsistent and inconclusive.[10–13] An alternative approach would be to identify individuals at high risk of early readmission and/or death before they are discharged from hospital to adapt their care management.

Several risk scores using a combination of clinical and laboratory data and/or other parameters have been developed[14] in different types of HF (chronic, acute and advanced) such as the PROTECT (Placebo-Controlled Randomized Study of the Selective A1 Adenosine Receptor Antagonist Rolofylline for Patients Hospitalized with Acute Decompensated Heart Failure and Volume Overload to Assess Treatment Effect on Congestion and Renal Function) score,[15] the Seattle Heart Failure Model (SHFM),[16] the Heart Failure Survival Score (HFSS),[17] the MAGGIC (Meta-Analysis Global Group in Chronic Heart Failure) score[18] and the Acute Decompensated Heart Failure National Registry (ADHERE) model[19] (*Table* S1). However, despite their potential to help identify high-risk individuals, existing scores usually require manual calculation, which limits their usefulness in clinical practice. In addition, currently available options are not very accurate in determining the risk of readmission or 1 year mortality.[20] Furthermore, there is a lack of data on risk prediction during the vulnerable phase after HF-related hospitalization.

Therefore, there is an unmet need for a new approach to automatically identify high-risk individuals at the beginning of their HF-related hospitalization. One approach would involve exploiting the wealth of data available in the electronic health record (EHR) without a priori to automatically predict 90 day mortality. Based on routine care data, this model could be implemented in the hospital information system and calculated automatically, without the need for re-entry or additional data collection. Such an approach is now possible by artificial intelligence (AI) methods. AI can exploit these large amounts of data to capture complex relationships and select relevant variables.

The objective of this study was to develop and validate a model for the automatic prediction of mortality during the vulnerable phase—90 day post-discharge—in patients hospitalized for decompensated HF using data available in the EHR up to 48 h after admission.

Methods

Study design

This study was conducted over the period from 1 January 2015 to 31 December 2021. The research was conducted in two hospitals on retrospective data collected during hospital stays: Paris Saint-Joseph Hospital and Lille University Hospital. The study followed the French National Commission for Information Technology and Civil Liberties (CNIL) reference methodology MR004 and was approved by the institutional ethics committee (institutional review board number IRB00012157). Patients were informed of the study by mail and given the opportunity to refuse to participate.

Study populations

The derivation cohort included all individuals hospitalized for HF at Paris Saint-Joseph Hospital between 1 January 2015 and 31 December 2020. The evaluation cohorts included all individuals hospitalized for HF at Paris Saint-Joseph Hospital (evaluation Cohort A) and Lille University Hospital (evaluation Cohort B) between 1 March 2021 and 31 December 2021. HF hospitalizations were identified by the International Classification of Diseases, 10th Revision (ICD-10) HF codes (I500–I501) and by HF keywords in the discharge summary. Keywords were searched with a natural language processing (NLP) tool considering misspellings and negations. The NLP tool was validated on 100 randomly selected discharge summaries of patients hospitalized at Paris Saint-Joseph with an ICD-10 code for HF and verified by two HF specialists. The NLP tool confirmed HF hospitalization with 97% specificity. Patients with at least one of the following criteria were excluded: aged over 95 years, receiving palliative care, with documented cardiac amyloidosis, who died during the hospitalization or who expressed their opposition to the study (*Figure* S1). Every admission meeting the inclusion criteria was included, meaning that a given patient could contribute more than one admission to the analysis.

Data extraction and management

Data available up to 48 h after admission in the EHRs of both hospitals were extracted. Two cardiologists selected an initial set of 151 clinical and biological variables considered relevant and obtainable routinely by the two hospitals for the management of HF. A least absolute shrinkage and selection operator[21] (LASSO) regression was then used to select the final 42 variables in the model. For example, calcaemia, phosphoraemia and eosinophil count were not extracted. One hundred and fifty-one variables from 11 categories were

extracted: demographics; hospital admission mode; month of hospitalization; medications (based on the anatomical therapeutic chemical classification); biology; comorbidities; electrocardiogram (EKG); left ventricular ejection fraction (LVEF) determined by echocardiography; vital signs; clinical signs of HF; and HF aetiology (Figure S2).

Structured data were preferred but, when necessary, NLP tools were used to search for keywords. For discrete variables, the first values were retained and for continuous variables, the mean values were used. A programme was developed to automate the extraction and cleaning of the data and to ensure the homogeneity of the data between the derivation and the evaluation cohorts, and between the two hospitals calculated (Supporting information Methods). Only variables with less than 25% missing data were retained. Missing values were imputed using the variable-specific statistics: for continuous variables, missing values were substituted with the mean value, and for categorical variables, missing values were replaced with the most frequent value.

The 90 day post-discharge vital status was obtained from the hospital information system, supplemented by the database of the French National Institute of Statistics and Economic Studies (INSEE). Matching between the study population and the INSEE database was performed using the INSEEHOP[22] open-source tool.

Model development and testing

Logistic regression was used to predict the probability of 90 day post-discharge mortality by creating a model that correlates EHR variables and mortality with an output between 0 and 1. The model was developed in the derivation cohort and then independently tested on the two evaluation cohorts. Learning was performed using five-fold cross-evaluation.

Variables significantly associated with 90 day post-discharge mortality in a univariate model were included in a multivariate logistic regression. Given the large number of variables and the resulting risk of multicollinearity and overfitting, a LASSO[21] regression was applied, with a threshold of 0.08, set to maximize cross-validation performance. LASSO is a regularization technique that promotes sparse patterns and limits overfitting, by automatically selecting variables on the basis of a penalty term derived from the absolute values of the coefficients. The variables selected by LASSO were then validated by clinical experts (M. K., P. d. G.). The final model was built with 42 variables (Table S3). This selection of variables not only improved the robustness of the model, but also facilitated its integration into the hospital information system.

The performance of the logistic regressions was evaluated with the area under the curve (AUC) of the receiver operating characteristics (ROC) curve for each of the three cohorts. AUC was also determined in subgroups: according to LVEF ($<$50% vs. $\geq$50%), and according to the number of medications recommended for HF (0 or 1 medication vs. $\geq$2 medications). Pharmacological therapies included at least diuretics, renin–angiotensin system (RAS) inhibitors, beta-blockers, mineralocorticoid receptor antagonists (MRAs) (*Table* 1).

The coherence of predictions obtained with logistic regression was verified with calibration and Kaplan–Meier curves. To plot the Kaplan–Meier curves, the evaluation population of both cohorts was pooled and then divided into three subgroups according to the terciles of 90 day mortality prediction [$\leq$8.0%; (8.0%–12.5%); $>$12.5%]. Patients with a 90 day mortality probability of less than or equal to 8.0% were classified in the 'low-risk' subgroup, those with a probability between 8.0% and 12.5% were classified to the 'intermediate-risk' subgroup, and those with a probability greater than 12.5% were classified to the 'high-risk' subgroup. Kaplan–Meier curves were plotted up to 6 months of follow-up to assess the consistency of the risk groups over time.

Statistical analysis

After testing for normality of distribution, data were presented as mean ± standard deviation (normal distribution) or median [interquartile range] (non-normal distribution); categorical variables were presented as number (percentage). Characteristics of the derivation and of the two evaluation cohorts were compared using unpaired Student's *t*-test or Mann–Whitney test for continuous variables and the χ^2 test for categorical variables. Performance of the prediction model was determined by calculating the AUC; 95% confidence intervals (CIs) for AUC values were calculated by performing bootstrapping resampling. Differences in survival curves were tested using log-rank. A *P* value $\leq$0.05 was considered statistically significant. All experiments were performed using the Python package Scikit-learn version 1.3.0.

The AUC of our model was compared with that of the PROTECT score to predict the 90 day mortality post-discharge in the study population. The PROTECT score was calculated without the uric acid variable, as it was not available in the study population.

Results

Characteristics of the derivation cohort

The derivation cohort included 1733 patients (2257 hospitalizations) (*Table* 1). Mean age was 81 years. Half of the cohort

Table 1 Demographic and clinical characteristics of participants in the derivation cohort, based on 90 day survival or death and in the validation cohorts.

		Outcome after hospitalization for decompensated HF				
	Derivation cohort (n = 1733)	Survived to 90 days (n = 1568)	Death within 90 days (n = 223)	P value	Evaluation Cohort A (n = 312)	Evaluation Cohort B (n = 353)	P value
Number of hospitalization for HF	2257	2009	248		348	388	
Age[a], years	81 [73–87]	81 [72–87]	85 [79–89]	<0.001	80 [71–86]	72 [61–79]	<0.001
Male sex, n (%)	1251 (55.4)	1203 (55.4)	138 (55.6)	1.000	189 (54.3)	231 (59.5)	0.301
HF aetiology, n (%)							
Ischaemic	1014 (44.9)	885 (44.1)	129 (52.0)	0.056	143 (41.1)	122 (31.4)	0.021
Non-ischaemic	747 (33.1)	674 (33.5)	73 (29.4)		107 (30.7)	132 (34.0)	
Other	496 (22.0)	450 (22.4)	46 (18.5)		98 (28.2)	134 (34.5)	
LVEF, %	46.9 ± 15.4	47.2 ± 15.4	44.8 ± 15.3	0.020	44.3 ± 19.5	43.7 ± 21.0	0.682
LVEF category, n (%)							
≥50%	1102 (48.8)	993 (49.4)	109 (44.0)		138 (39.7)	173 (44.6)	0.051
40% to <50%	485 (21.5)	429 (21.4)	56 (22.6)	0.240	64 (18.4)	47 (12.1)	
<40%	670 (29.7)	587 (29.2)	83 (33.5)		146 (42.0)	168 (43.3)	
BNP[a], pg/mL	836.0 [441.0–1532.0]	808.0 [431.0–1438.5]	1262.0 [696.5–2192.0]	<0.001	879.5 [409.5–1739.8]	952.9 [481.5–1879.8]	0.001
Serum creatinine, μmol/L	125.0 ± 67.5	122.1 ± 65.9	148.5 ± 76.0	<0.001	122.1 ± 63.9	124.2 ± 58.6	0.643
Comorbidities, n (%)							
Hypertension	1723 (76.3)	1534 (76.4)	189 (76.2)	0.937	233 (67.0)	267 (68.8)	0.635
Atrial fibrillation	1301 (57.6)	1155 (57.5)	146 (58.9)	0.734	201 (57.8)	217 (55.9)	0.541
Diabetes	820 (36.3)	732 (36.4)	88 (35.5)	0.834	133 (38.2)	153 (39.4)	0.762
Cancer	473 (21.0)	417 (20.8)	56 (22.6)	0.509	75 (21.6)	31 (8.0)	<0.001
COPD	518 (23.0)	459 (22.8)	59 (23.8)	0.749	83 (23.9)	81 (20.9)	0.375
Pacemaker, n (%)	621 (27.5)	531 (26.4)	90 (36.3)	0.001	116 (33.3)	59 (15.2)	<0.001
Automatic implantable defibrillator, n (%)	281 (12.5)	248 (12.3)	33 (13.3)	0.683	51 (14.7)	88 (22.7)	0.006
Medications, n (%)							
Diuretics	1797 (79.6)	1596 (79.4)	201 (81.0)	0.616	287 (82.5)	358 (92.3)	<0.001
Beta-blocker	1266 (56.1)	1146 (57.0)	120 (48.4)	0.118	216 (62.1)	207 (53.4)	0.021
RAS inhibitor	977 (43.3)	903 (44.9)	74 (29.8)	<0.001	191 (54.9)	193 (49.7)	0.237
MR antagonist	344 (15.2)	310 (15.4)	34 (13.7)	0.513	61 (17.5)	99 (25.5)	0.009
CCB	329 (14.6)	301 (15.0)	28 (11.3)	0.128	86 (24.7)	101 (26.0)	0.735
Cardiac glycosides	204 (4.6)	97 (4.8)	7 (2.8)	0.198	13 (3.7)	26 (6.7)	0.098
SGLT2i	1 (0.0)	1 (0.0)	0 (0.0)		28 (8.0)	47 (12.1)	0.087
Anticoagulants	1123 (49.8)	1005 (50.0)	118 (47.6)	0.501	225 (64.7)	190 (49.0)	<0.001
Antiplatelets	887 (39.3)	788 (39.2)	99 (39.9)	0.836	112 (32.2)	118 (30.4)	0.633
Amiodarone	547 (24.2)	488 (24.3)	59 (23.8)	0.937	42 (12.1)	60 (15.5)	0.200
Number of recommended HF medications, n (%)							
0 or 1	1483 (65.7)	1299 (64.7)	184 (74.2)	0.003	198 (56.9)	227 (58.5)	0.659
≥2	774 (34.3)	710 (35.3)	64 (25.8)		150 (43.1)	161 (41.5)	

Note: Values are mean ± standard deviation, median [interquartile range], or number of hospitalizations (%).
Abbreviations: BNP, B-type natriuretic peptide; CCB, calcium channel blocker; COPD, chronic obstructive pulmonary disease; HF, heart failure; LVEF, left ventricular ejection fraction; MR, mineralocorticoid receptor; RAS, renin–angiotensin system; SGLT2i, sodium-glucose cotransporter 2 inhibitor.
[a]BNP and age values do not follow a normalized distribution.

had an ischaemic HF (44.9%) and nearly half of the cohort had a preserved LVEF (48.8%). Almost half of the cohort (44.0%) had been admitted to intensive care. The most common HF medications were diuretics (79.6% of hospitalizations), followed by beta-blockers (56.1%) and RAS inhibitors (43.3%). MRAs were used in 15% of patients. Due to the recruitment period of this cohort, only one patient was treated with a Sodium Glucose Cotransporter 2 Inhibitor (SGLT2i). The rate of comorbidities was high, especially hypertension (76.3%), atrial fibrillation (57.6%) and diabetes mellitus (36.3%). Patients in the '0 or 1 medication recommended for HF' subgroup were more likely to have preserved LVEF than those in the '≥2 medications recommended for HF' subgroup (53% vs. 40% respectively, $P < 0.05$).

Two hundred and twenty-three patients (12.9%) died within 90 days of hospital discharge (248/2257 hospitalizations, 11.0%). They were characterized by older age, higher levels of B-type natriuretic peptide (BNP), serum creatinine and pacemaker implantation rate, as well as lower LVEF and prescription rate of RAS inhibitors. Furthermore, these patients were more likely to have 0 or 1 medication recommended for HF (*Table* 1).

Only three variables had more than 14% missing data. Of these, aspartate aminotransferase (ASAT) and BNP were retained in the final model, despite missing values rate of respectively 24.15% and 14.18%. The other variables in the dataset had missing values of less than 10% (Table S2).

Characteristics of the evaluation cohorts

Evaluation Cohort A included 312 patients (348 hospitalizations) and evaluation Cohort B included 353 patients (388 hospitalizations). There were several statistically significant differences between Cohort A and Cohort B. Patients of Cohort B were characterized by younger age, higher levels of BNP, higher prescription rate of diuretics and MRAs, as well as lower prescription rate of beta-blocker and anticoagulant, and pacemaker implantation rate. In addition, these patients were less likely to have cancer, whereas ischaemic HF and the presence of an implantable cardiac defibrillator (ICD) were more common in cohort B (*Table* 1).

Performance of the prediction model

The 90 day mortality rates were similar in the three cohorts (11.0% of the patients discharged in the derivation cohort, 10.9% in the evaluation Cohort A, and 10.3% in the evaluation Cohort B) (*Table* 2). Across all cohorts, the mortality rate tended to be higher in the 'LVEF <50%' subgroup and in the '0 or 1 medication recommended for HF' subgroup.

Table 2 Ninety-day mortality and model performance for prediction of 90 day post-discharge mortality

	N[a]	Mortality rate, %	Model performance, AUC (95% CI)
Derivation cohort			
Overall	2257	11.0	0.817 (0.789–0.845)
LVEF subgroups			
≥50%	1102	9.9	0.779 (0.689–0.817)
<50%	1115	11.6	0.827 (0.764–0.859)
Number of HF therapies subgroups			
0 or 1	1483	12.4	0.807 (0.750–0.835)
≥2	774	8.3	0.793 (0.686–0.841)
Sex subgroups			
Female	1006	10.	0.810 (0.766–0.854)
Male	1251	11.0	0.823 (0.787–0.860)
Age subgroups			
>80 years	1222	14.2	0.792 (0.754–0.829)
<80 years	1035	7.1	0.843 (0.797–0.889)
Evaluation Cohort A			
Overall	348	10.9	0.750 (0.672–0.829)
LVEF subgroups			
≥50%	138	10.1	0.673 (0.413–0.794)
<50%	210	11.4	0.793 (0.618–0.860)
Number of HF therapies subgroups			
0 or 1	140	10.2	0.721 (0.509–0.802)
≥2	150	8.3	0.791 (0.589–0.895)
Sex subgroups			
Female	159	8.8	0.742 (0.643–0.841)
Male	189	12.7	0.753 (0.641–0.864)
Age subgroups			
>80 years	163	1	0.701 (0.586–0.817)
<80 years	185	x	0.795 (0.689–0.902)
Evaluation Cohort B			
Overall	388	10.3	0.723 (0.644–0.803)
LVEF subgroups			
≥50%	173	9.8	0.673 (0.453–0.776)
<50%	215	10.7	0.763 (0.589–0.844)
Number of HF therapies subgroups			
0 or 1	128	11.7	0.691 (0.513–0.778)
≥2	161	8.1	0.765 (0.540–0.853)
Sex subgroups			
Female	157	11.5	0.703 (0.577–0.829)
Male	231	9.5	0.748 (0.643–0.853)
Age subgroups			
>80 years	89	18.0	0.741 (0.617–0.866)
<80 years	299	8.0	0.697 (0.588–0.805)

Abbreviations: AUC, area under the receiver operating characteristic curve; CI, confidence interval; HF, heart failure; LVEF, left ventricular ejection fraction.
[a]*N* indicates the number of hospitalizations included in each group (patients could contribute more than one hospitalization event to the dataset).

The prediction model had an AUC: 0.817; 95% CI (0.789–0.845), in the derivation cohort. Performance in predicting 90 day post-discharge mortality was consistent in evaluation Cohort A [AUC: 0.750; 95% CI (0.672–0.829)] and evaluation Cohort B [AUC: 0.723; 95% CI (0.644–0.803)] (*Table* 2 and *Figure* 1). Across the derivation and the two evaluation cohorts, model performance was better in the 'LVEF < 50%' and in the '0 or 1 medication recommended for HF' subgroups (*Table* 2).

The calibration curves showed that there was a good consistency between predicted probability and actual probability

Figure 1 ROC curves of the risk models for the derivation and evaluation cohorts. The performance of logistic regression predicting 90 day mortality after hospitalization for heart failure was evaluated with the AUC of the ROC curves. The model had a good discriminative ability (AUC = 0.82) and performance was consistent in the evaluation cohorts. AUC, area under the curve; ROC, receiver operating characteristics.

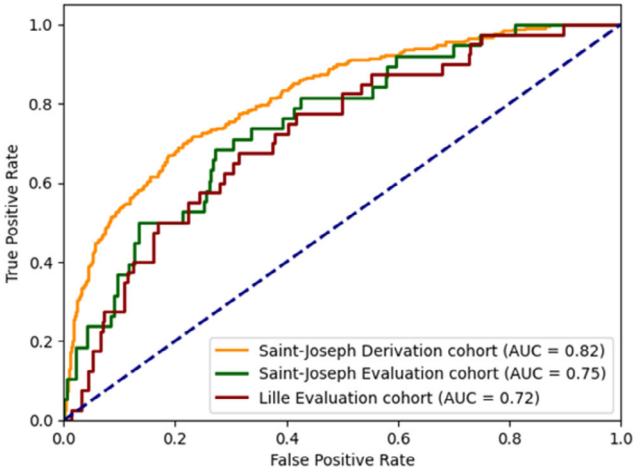
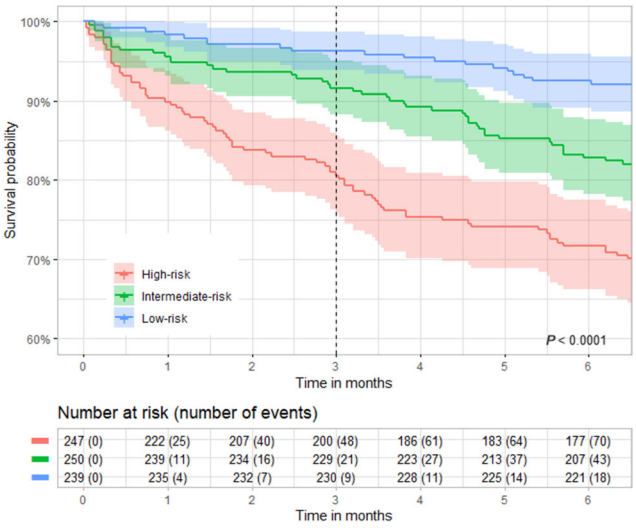


Figure 2 Observed survival in the evaluation cohorts by risk prediction subgroups. The coherence of 90 day mortality predictions was verified using Kaplan–Meier curves. To plot the Kaplan–Meier curves, the evaluation population was divided into three subgroups: low risk: predicted 90 day mortality $\leq$ 8.0%; intermediate risk: 8.0% $<$ predicted 90 day mortality $<$ 12.5%; high risk: predicted 90 day $\geq$ 12.5%. The risk subgroups had different survival curves and 90 day mortality rate.



(Figure S3). Moreover, the three risk subgroups ('low-risk', 'intermediate-risk' and 'high-risk') had different survival curves and 90 day mortality rate (3.8%, 8.4% and 19.4% respectively) (*Figure* 2).

The PROTECT model had an AUC: 0.710; 95% CI (0.676–0.744), in the derivation cohort, and an AUC: 0.687; 95% CI (0.591–0.783), and an AUC: 0.655; 95% CI (0.571–0.739) in the evaluation Cohorts A and B, respectively.

Discussion

To the best of our knowledge, this is the first model developed to predict 90 day post-discharge mortality for HF decompensation episodes using data from the EHR available up to 48 h after admission. The model is based on a limited number of routine care data (42 variables) extracted from the EHR data. The model performed well in the derivation

cohort [AUC: 0.817; 95% CI (0.789–0.845)], indicating good discriminative ability. Evaluation in two independent cohorts from different hospitals [Cohort A AUC: 0.750; 95% CI (0.672–0.829), and Cohort B AUC: 0.723; 95% CI (0.644–0.803)] indicated that there was no overfitting of the model. Furthermore, despite some differences in clinical profiles or management modalities between the two evaluation cohorts, the overall model performance remained consistent. This indicates a good level of robustness and provides further support for the reliability and generalizability of the model.

However, subgroup analyses suggest that model performance may differ according to patient profile in the evaluation cohorts: model performance in patients with LVEF $< 50\%$ was higher than in patients with preserved ejection fraction (LVEF $\geq 50\%$). This could be explained by the difference in pathophysiology and treatment modalities according to ejection fraction. A higher model performance was also observed in the '$\geq$2 medications recommended for HF' subgroup. This may be related to lower proportion of patients with preserved LVEF in the '$\geq$2 medications recommended for HF' subgroup, as there is less evidence of benefit from currently available pharmacological therapies in HF with preserved ejection fraction.

It is difficult to compare our model with other predictive models because the study populations are different, as are the primary outcomes (mortality at 30 days, 180 days or 1 year). Previous scores have predicted mortality at 90 days, the Organized Program To Initiate lifesaving treatMent In hospitaliZEd patients (OPTIMIZE-HF),[23] the Acute Physiology And Chronic Health Evaluation in Heart Failure (APACHE-HF)[24] and the PROTECT scores. However, the methodology was different, with a pre-specified selection of variables, no evaluation cohort and standard statistical analyses. We compared our score with the PROTECT score and obtained a better performance [Cohort A AUC: 0.750; 95% CI (0.672–0.829), and Cohort B AUC: 0.723; 95% CI (0.644–0.803), vs. Cohort A AUC: 0.687; 95% CI (0.591–0.783), and Cohort B AUC: 0.655; 95% CI (0.571–0.739), respectively]. However, we calculated the PROTECT score without the uric acid variable, which was not available in our data. However, the effect of the uric acid may be small, as its hazard ratio (95% CI) was low [1.01 (1.00–1.02)] in the PROTECT model. The performance of the current model is also better than that of the ADHERE[19] score used to predict in-hospital mortality in acutely decompensated HF, which had an AUC of 0.63. In addition, although the SHFM model applied on discharge after a decompensated HF hospitalization had good risk discrimination for predicting 1 year mortality in patients aged $<$65 years (AUC 0.704), it showed poor performance in older individuals,[25] which is an important limitation given the age of HF patients referred to hospitals. Other scores have been developed to predict in-hospital mortality,[26–29] but this is out of the scope of our work as we were interested in studying the post-discharge mortality during the vulnerable phase.

Other researchers have applied machine learning approaches to determine risk in patients with HF. The AUC in our model was slightly lower than that of a machine learning model developed to predict 1 year mortality risk in HF patients [Machine learning Assessment of RisK and EaRly mortality in Heart Failure (MARKER-HF)], which had an AUC 0.81–0.88.[30] However, the training population of the MARKER-HF study included both hospitalized and stable ambulatory patients and focussed on 1 year mortality. Moreover, the mean age of the derivation cohort was 59 years, which is not representative of the older HF population usually seen in clinical practice. Another machine learning model with a LASSO algorithm used data on 27 continuous and 44 categorical variables from patients hospitalized for acute HF syndrome.[31] The model's AUC for 30 day mortality in this study was 0.78 in the derivation set and 0.76 in the evaluation set. However, this study included only data from East-Asian patients, and therefore, its usefulness in individuals from other ethnic groups is not known.

In contrast to previous machine-learning models that looked at mortality and/or readmission risk over longer time periods, the model developed in this study was designed to specifically predict mortality risk in the vulnerable 90 day period after discharge from a HF-related hospitalization. This is particularly important as it has recently been suggested that early uptitration of guideline-directed medical therapies in decompensated HF is associated with significant clinical benefit.[32]

It has been suggested that the ideal risk score for HF would have the highest level of 'goodness of fit' and the lowest level of complexity.[33] This means that the variables included in the model must be easy to obtain, routinely available in clinical practice, and as objective as possible.[33] Therefore, our risk prediction model was designed to be easily integrated into the hospital information system using data already available in the first hours after admission. It can then be used to provide an automated calculation of the risk before a patient is discharged. This automation is a major advantage, as it does not add to the workload of physicians or other healthcare professionals involved in patient management.

Study strengths

This study has several strengths. The model was developed using data from all hospitalized HF patients, regardless of LVEF, and validated in two independent cohorts from different hospitals, indicating the model's robustness for different HF populations. As 90 day mortality was obtained by matching with the French national mortality database, very few patients were lost to follow-up. The calibration of the model was assessed with concordance between predicted and actual

90 day mortality to ensure accuracy. NLP was used to complement ICD-10 codes and improve the accuracy of diagnostic and comorbidity information, as previously demonstrated in HF.[34]

Study limitations

Several limitations need to be acknowledged when interpreting our findings. The model was developed using data from a single centre in France, and although it was subsequently validated on two independent and clinically different populations, the possibility of location and selection bias cannot be completely excluded. Our model requires therefore further external validation and will be tested in multicentre studies to confirm its robustness and applicability across different healthcare facilities and hospitals.

Although the model gives promising results, the lower limit of the CI, falling below 0.7, could indicate a limitation in the model's discriminative power. A larger sample size of the evaluation cohorts could contribute to strengthening the model's robustness as well as including other populations from different cities and countries.

SGLT2 inhibitors have been shown to significantly improve mortality in HF in several large outcome clinical trials. The timeframe of development of our study did not allow to enrol many patients treated by SGLT2 inhibitors, which were not yet widely used. They were approved for the management of HF in our country in August 2021. The relevance of our model will therefore need to be reassessed in populations treated by this new class of HF. Nevertheless, practices for HF management in the current datasets are more contemporary than those used in older HF risk models.

Our population was made of elderly patients. However, subgroup analyses suggested that the algorithm performed well for the age group $<$80 years [Cohort A AUC 0.795 (0.689–0.902) and Cohort B AUC 0.697 (0.588–0.805)]. Whether our model is relevant for a younger HF population needs therefore further studies.

The current model is specific to the setting in which the data informing the model were obtained, that is, the prediction of 90 day mortality after hospitalization for decompensated HF, and it does not address prediction of rehospitalization or longer-term mortality. However, the 6 month survival curves show consistency in risk prediction.

Conclusions

The model developed in this study provides a prediction of early post-discharge mortality in patients with decompensated HF. This could help to identify patients at very high risk of mortality after a HF decompensation episode and allow appropriate intensification of their medical management before and after discharge.

To minimize performance gaps between different populations and improve replicability, further model training could be undertaken on cohorts from different hospitals. This could be achieved using innovative machine learning approaches such as federated learning. In addition, this algorithm must be prospectively tested in clinical practice to measure its real impact on the 90 day survival rates.

Acknowledgements

Medical writing assistance was provided by Nicola Ryan, independent medical writer, funded by Hôpital Paris Saint-Joseph, Paris, France. We thank Xavier Maynadier for his help with data verification and Olivier Billuart for his help with data management.

Conflict of interest statement

None declared.

Funding

This work was supported by the French National Research Agency (ANR) under the PREDHIC project (ANR-21-CE23-0039-03).

Supporting information

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

Table S1. Main risk scores in acute heart failure.
Table S2. Variables with missing values.
Figure S1. Inclusion flow chart for the derivation cohort.
Figure S2. Data selection through the process of model development.
Figure S3. Calibration curves for predicting 90 day mortality risk after hospitalization for decompensated heart failure.
Table S3. Variables in the model prediction 90 day mortality risk after hospitalization for decompensated heart failure.

References

1. Yan T, Zhu S, Yin X, Xie C, Xue J, Zhu M, *et al*. Burden, trends, and inequalities of heart failure globally, 1990 to 2019: a secondary analysis based on the Global Burden of Disease 2019 study. *JAHA* 2023;**12**:e027852. doi:10.1161/JAHA.122.027852
2. Seferović PM, Vardas P, Jankowska EA, Maggioni AP, Timmis A, Milinković I, *et al*. The Heart Failure Association atlas: heart failure epidemiology and management statistics 2019. *Eur J Heart Fail* 2021;**23**:906-914. doi:10.1002/ejhf.2143
3. Tsao CW, Aday AW, Almarzooq ZI, Anderson CAM, Arora P, Avery CL, *et al*. Heart disease and stroke statistics-2023 update: a report from the American Heart Association. *Circulation* 2023;**147**:e93-e621. doi:10.1161/CIR.0000000000001123
4. Greene SJ, Fonarow GC, Vaduganathan M, Khan SS, Butler J, Gheorghiade M. The vulnerable phase after hospitalization for heart failure. *Nat Rev Cardiol* 2015;**12**:220-229. doi:10.1038/nrcardio.2015.14
5. O'Connor CM, Miller AB, Blair JEA, Konstam MA, Wedge P, Bahit MC, *et al*. Causes of death and rehospitalization in patients hospitalized with worsening heart failure and reduced left ventricular ejection fraction: results from Efficacy of Vasopressin Antagonism in Heart Failure Outcome Study with Tolvaptan (EVEREST) program. *Am Heart J* 2010;**159**:841-849.e1. doi:10.1016/j.ahj.2010.02.023
6. Krumholz HM. Post-hospital syndrome—an acquired, transient condition of generalized risk. *N Engl J Med* 2013;**368**:100-102. doi:10.1056/NEJMp1212324
7. Llorens P. Risk assessment in emergency department patients with acute heart failure: we need to reach beyond our clinical judgment. *Emergencias* 2018;**30**:75-76.
8. Rizzi MA, Sarasola AG, Arbé AA, Mateo SH, Gil V, Llorens P, *et al*. Factors associated with in-hospital mortality and adverse outcomes during the vulnerable post-discharge phase after the first episode of acute heart failure: results of the NOVICA-2 study. *Clin Res Cardiol* 2021;**110**:993-1005. doi:10.1007/s00392-020-01710-0
9. Kimmoun A, Takagi K, Gall E, Ishihara S, Hammoum P, El Bèze N, *et al*. Temporal trends in mortality and readmission after acute heart failure: a systematic review and meta-regression in the past four decades. *Eur J Heart Fail* 2021;**23**:420-431. doi:10.1002/ejhf.2103
10. Galinier M, Roubille F, Berdague P, Brierre G, Cantie P, Dary P, *et al*. Telemonitoring versus standard care in heart failure: a randomised multicentre trial. *Eur J Heart Fail* 2020;**22**:985-994. doi:10.1002/ejhf.1906
11. Koehler F, Koehler K, Deckwart O, Prescher S, Wegscheider K, Winkler S, *et al*. Telemedical Interventional Management in Heart Failure II (TIM-HF2), a randomised, controlled trial investigating the impact of telemedicine on unplanned cardiovascular hospitalisations and mortality in heart failure patients: study design and description of the intervention. *Eur J Heart Fail* 2018;**20**:1485-1493. doi:10.1002/ejhf.1300
12. Logeart D, Berthelot E, Bihry N, Eschalier R, Salvat M, Garcon P, *et al*. Early and short-term intensive management after discharge for patients hospitalized with acute heart failure: a randomized study (ECAD-HF). *Eur J Heart Fail* 2022;**24**:219-226. doi:10.1002/ejhf.2357
13. Murtaugh CM, Deb P, Zhu C, Peng TR, Barrón Y, Shah S, *et al*. Reducing readmissions among heart failure patients discharged to home health care: effectiveness of early and intensive nursing services and early physician follow-up. *Health Serv Res* 2017;**52**:1445-1472. doi:10.1111/1475-6773.12537
14. Kansagara D, Englander H, Salanitro A, Kagen D, Theobald C, Freeman M, *et al*. Risk prediction models for hospital readmission: a systematic review. *JAMA* 2011;**306**:1688-1698. doi:10.1001/jama.2011.1515
15. Davison BA, Metra M, Senger S, Edwards C, Milo O, Bloomfield DM, *et al*. Patient journey after admission for acute heart failure: length of stay, 30-day readmission and 90-day mortality. *Eur J Heart Fail* 2016;**18**:1041-1050. doi:10.1002/ejhf.540
16. Levy WC, Mozaffarian D, Linker DT, Sutradhar SC, Anker SD, Cropp AB, *et al*. The Seattle heart failure model: prediction of survival in heart failure. *Circulation* 2006;**113**:1424-1433. doi:10.1161/CIRCULATIONAHA.105.584102
17. Aaronson KD, Schwartz JS, Chen TM, Wong KL, Goin JE, Mancini DM. Development and prospective validation of a clinical index to predict survival in ambulatory patients referred for cardiac transplant evaluation. *Circulation* 1997;**95**:2660-2667. doi:10.1161/01.CIR.95.12.2660
18. Pocock SJ, Ariti CA, McMurray JJV, Maggioni A, Køber L, Squire IB, *et al*. Predicting survival in heart failure: a risk score based on 39 372 patients from 30 studies. *Eur Heart J* 2013;**34**:1404-1413. doi:10.1093/eurheartj/ehs337
19. Fonarow GC, Adams KF, Abraham WT, Yancy CW, Boscardin WJ, ADHERE Scientific Advisory Committee, Study Group, and Investigators. Risk stratification for in-hospital mortality in acutely decompensated heart failure: classification and regression tree analysis. *JAMA* 2005;**293**:572-580. doi:10.1001/jama.293.5.572
20. Bo X, Zhang Y, Liu Y, Kharbuja N, Chen L. Performance of the heart failure risk scores in predicting 1 year mortality and short-term readmission of patients. *ESC Heart Fail* 2023;**10**:502-517. doi:10.1002/ehf2.14208
21. Tibshirani R. Regression shrinkage and selection via the lasso. *J R Stat Soc B Methodol* 1996;**58**:267-288.
22. Guardiolle V, Bazoge A, Morin E, Daille B, Toublant D, Bouzillé G, *et al*. Linking biomedical data warehouse records with the National Mortality Database in France: large-scale matching algorithm. *JMIR Med Inform* 2022;**10**:e36711. doi:10.2196/36711
23. O'Connor CM, Abraham WT, Albert NM, Clare R, Gattis Stough W, Gheorghiade M, *et al*. Predictors of mortality after discharge in patients hospitalized with heart failure: an analysis from the Organized Program to Initiate Lifesaving Treatment in Hospitalized Patients with Heart Failure (OPTIMIZE-HF). *Am Heart J* 2008;**156**:662-673. doi:10.1016/j.ahj.2008.04.030
24. Okazaki H, Shirakabe A, Hata N, Yamamoto M, Kobayashi N, Shinada T, *et al*. New scoring system (APACHE-HF) for predicting adverse outcomes in patients with acute heart failure: evaluation of the APACHE II and modified APACHE II scoring systems. *J Cardiol* 2014;**64**:441-449. doi:10.1016/j.jjcc.2014.03.002
25. Li S, Marcus P, Núñez J, Núñez E, Sanchis J, Levy WC. Validity of the Seattle Heart Failure Model after heart failure hospitalization. *ESC Heart Fail* 2019;**6**:509-515. doi:10.1002/ehf2.12427
26. Auble TE, Hsieh M, Gardner W, Cooper GF, Stone RA, McCausland JB, *et al*. A prediction rule to identify low-risk patients with heart failure. *Acad Emerg Med* 2005;**12**:514-521. doi:10.1197/j.aem.2004.11.026
27. Abraham WT, Fonarow GC, Albert NM, Stough WG, Gheorghiade M, Greenberg BH, *et al*. Predictors of in-hospital mortality in patients hospitalized for heart failure: insights from the Organized Program to Initiate Lifesaving Treatment in Hospitalized Patients with Heart Failure (OPTIMIZE-HF). *J Am Coll Cardiol* 2008;**52**:347-356. doi:10.1016/j.jacc.2008.04.028
28. Peterson PN, Rumsfeld JS, Liang L, Albert NM, Hernandez AF, Peterson ED, *et al*. A validated risk score for in-hospital mortality in patients with heart failure from the American Heart Association Get With The Guidelines program. *Circ Cardiovasc Qual Outcomes* 2010;**3**:25-32. doi:10.1161/CIRCOUTCOMES.109.854877
29. Zhao H-L, Gao X-L, Liu Y-H, Li S-L, Zhang Q, Shan W-C, *et al*. Validation

and derivation of short-term prognostic risk score in acute decompensated heart failure in China. *BMC Cardiovasc Disord* 2022;**22**:307. doi:10.1186/s12872-022-02743-1

30. Adler ED, Voors AA, Klein L, Macheret F, Braun OO, Urey MA, *et al*. Improving risk prediction in heart failure using machine learning. *Eur J Heart Fail* 2020; **22**:139-147. doi:10.1002/ejhf.1628
31. Kim W, Park JJ, Lee H-Y, Kim KH, Yoo B-S, Kang S-M, *et al*. Predicting survival in heart failure: a risk score based on machine-learning and change point algorithm. *Clin Res Cardiol* 2021;**110**: 1321-1333. doi:10.1007/s00392-021-01870-7
32. Mebazaa A, Davison B, Chioncel O, Cohen-Solal A, Diaz R, Filippatos G, *et al*. Safety, tolerability and efficacy of up-titration of guideline-directed medical therapies for acute heart failure (STRONG-HF): a multinational, open-label, randomised, trial. *Lancet* 2022; **400**:1938-1952. doi:10.1016/S0140-6736(22)02076-1
33. Wussler D, Mueller C. Biomarker-driven prognostic model for risk prediction in heart failure: ready for prime time? *Eur Heart J* 2021;**42**:4465-4467. doi:10.1093/eurheartj/ehab645
34. Ambrosy AP, Parikh RV, Sung SH, Narayanan A, Masson R, Lam P-Q, *et al*. A natural language processing-based approach for identifying hospitalizations for worsening heart failure within an integrated health care delivery system. *JAMA Netw Open* 2021;**4**: e2135152. doi:10.1001/jamanetworkopen.2021.35152