ORIGINAL ARTICLE

Prognostic Comparison between Risk Scores and Neural Networks to Predict Short- and Mid-Term Mortality in Acute Heart Failure

Comparación pronóstica entre scores de riesgo y la aplicación de redes neuronales para la predicción de la mortalidad a corto y mediano plazo en pacientes con insuficiencia cardíaca

MARIA JIMENA GAMBARTE, RAÚL ALFREDO BORRACCI[†], CLAUDIO HIGA, FEDOR NOVO, GRACIANA MARIA CIAMBRONE, OMAR DARIO TUPAYACHI VILLAGOMEZ, AGUSTINA GINESI, MARIA SOL DONATO, IGNACIO NOGUES

ABSTRACT

Background: Heart failure (HF) risk scores to assess all-cause mortality during the first year have areas under the ROC curve (AUC) ranging between 0.59 and 0.80
Objective: to develop and validate a neural network (NN) algorithm-based model to improve traditional scores' performance for predicting short- and mid-term mortality of patients with acute HF.
Methods: A prospective clinical database was analyzed including 483 patients admitted with diagnosis of acute HF in a coronary care unit community hospital of Buenos Aires, between June 2005 and June 2019. Among 181 demographic, laboratory, treatment and follow-up variables, only 25 were selected to calculate five acute heart failure risk scores aimed to predict 30-day, 6-month and 1-year mortality: EFFECT, ADHERE, GWTG-HF, 3C-HF, and ACUTE-HF.
Results: Mean age was 78±11.1 years, 58% were men, 35% had ischemic necrotic HF and median left ventricular ejection fraction was 52% (35-60). At 30 days, the EFFECT score (AUC:0.68) and the 3C-HF score (AUC: 0.68) showed better performance than the ACUTE-HF score (AUC: 0.54). At 6-month and 1-year follow-up, the EFFECT score (ROC: 0.69 and 0.69) outperformed the ADHERE score (AUC: 0.53 and 0.56), and EFFECT (AUC: 0.69 and 0.69), GWRG-HF (AUC=0.68 and 0.66), and 3C-HF (AUC:0.67 and 0.67) scores outperformed the ACUTE-HF score (AUC:0.53 and 0.56). The best results with NN algorithms were obtained with a two-hidden layer multilayer perceptron. A 24-9-7-2-layer architecture NN was used with the following results: AUC: 0.82, negative predictive value (NPV) 93.2% and positive predictive value (PPV) 66.7% for 30-day mortality; AUC: 0.87, NPV: 89.1% and PPV: 78,6% for 6-month mortality; and AUC: 0.85, NPV: 85.6% and PPV: 78.9% for 1-year mortality. In terms of discrimination, NN algorithms outperformed all the traditional scores ($p < 0.001$). For this algorithm, the most influential factors in descending order that scored ≥50% normalized importance to predict 30-day mortality were serum creatinine, hemoglobin, respiratory rate, blood urea nitrogen, serum sodium, age and systolic blood pressure. Also, NYHA functional class III-IV and dementia added prognostic capacity to 6-month mortality, and heart rate and chronic kidney disease to 1-year mortality.
Conclusions: The models with NN algorithms were significantly superior to traditional risk scores in our population of patients with HF. These findings constitute a working hypothesis to be validated with a larger and multicenter sample of cases.

Key Words: heart failure, prognosis, mortality, risk score, deep learning, artificial intelligence

RESUMEN

Introducción: En el contexto de la insuficiencia cardíaca (IC) existen scores de riesgo para evaluar la mortalidad por cualquier causa durante el primer año, con áreas bajo la curva ROC que oscilan entre 0,59 y 0,80.
Objetivo: Desarrollar y validar un modelo basado en algoritmos de redes neuronales (RN) destinado a mejorar el rendimiento de los modelos tradicionales para predecir mortalidad a corto y mediano plazo de pacientes con IC aguda.
Material y método: Se analizó una base de datos con 181 variables de 483 pacientes con IC aguda en un hospital de comunidad de la Ciudad de Buenos Aires (junio de 2005 -junio de 2019). Se utilizaron 25 variables para calcular 5 modelos de riesgo validados para predecir la mortalidad a 30 días, 6 meses y un año: EFFECT, ADHERE, GWTG-HF, 3C-HF y ACUTE-HF.
Resultados: La edad media fue 78±11,1años, 58% eran varones, el 35% de las IC eran de etiología isquémico necrótica, y la fracción de eyección media fue 52% (35-60). En término de discriminación a 30 días, fueron mejores el score EFFECT (ROC: 0,68) y el 3C-HF (ROC: 0,67) que el ACUTE- HF (ROC: 0,54). A los 6 meses y al año, el score EFFECT (ROC: 0,69 y 0,69) superó al ADHERE (ROC: 0,53 y 0,56), y los scores EFFECT (ROC: 0,69 y 0,69), GWRG-HF (ROC: 0,68 y 0,66 y 3C-HF (ROC: 0,67 y 0,67) superaron al score ACUTE-HF (ROC: 0,53 y 0,56). De los algoritmos de RN los mejores resultados se obtuvieron con un perceptrón multicapa (PMC)

Rev Argent Cardiol 2021;89:416-426. http://dx.doi.org/10.7775/rac.v89.i5.20434
SEE RELATED ARTICLE: Rev Argent Cardiol 2021;89:369-370. http://dx.doi.org/10.7775/rac.v89.i5.20449

Received: 02/18/2021 – *Accepted:* 09/06/2021

Financing: None
This work received the 2020 Cardiology Congress Award

[1]Hospital Alemán, Buenos Aires, Argentina
Corresponding author: María Jimena Gambarte

con dos capas ocultas. Se usó una RN de arquitectura de capas 24-9-7-2 con los siguientes resultados: ROC: 0,82, valor predictivo negativo (VPN) 93,2% y valor predictivo positivo (VPP) 66,7% para mortalidad a 30 días; ROC: 0,87, VPN: 89,1% y VPP: 78,6% para mortalidad a 6 meses; y ROC: 0,85, VPN: 85,6% y VPP: 78,9% para mortalidad al año. En términos de discriminación, los algoritmos de RN superaron a los scores tradicionales ($p<0,001$). Los factores que obtuvieron ≥50% de importancia estandarizada para predecir la mortalidad a los 30 días fueron en orden descendente la creatinina sérica, la hemoglobina, la frecuencia respiratoria, la urea, el sodio, la edad y la presión arterial sistólica. Agregaron capacidad pronóstica la clase III-IV NYHA y la demencia para mortalidad a 6 meses, y la frecuencia cardíaca y la disfunción renal crónica para mortalidad al año.
Conclusiones: Los modelos con algoritmos de RN fueron significativamente superiores a los scores de riesgo tradicionales en nuestros pacientes con IC. Estos hallazgos constituyen una hipótesis de trabajo a validar con una mayor muestra de casos y en forma multicéntrica.

Palabras claves: insuficiencia cardíaca, pronóstico, mortalidad, score de riesgo, redes neuronales, inteligencia artificial, mortalidad.

Abbreviations

ACCF	American College of Cardiology Foundation	**HF**	Heart failure
ADHERE	Acute Decompensated Heart Failure National Registry	**OPTIMIZE-HF**	Organized Program to Initiate Lifesaving Treatment in Hospitalized Patients with Heart Failure
AHA	American Heart Association	**MLP**	Multilayer perceptron
EFFECT	Enhanced Feedback for Effective Cardiac Treatment	**NN**	Neural network
ESCAPE	Evaluation Study of Congestive Heart Failure and Pulmonary Artery Catheterization Effectiveness	**IQR**	Interquartile range
RBF	Radial basis function network	**ROC**	Receiver Operating Characteristic
GWTG-HF	Get With the Guidelines-Heart Failure	**AUC**	Area under the ROC curve
HANBAH	Hemoglobin, age, sodium, blood urea nitrogen, atrial fibrillation, and high-density lipoprotein	**NPV**	Negative predictive value
		PPV	Positive predictive value

INTRODUCTION

Heart failure (HF) is a highly prevalent worldwide disease associated with considerable morbidity, high costs and poor mid-term prognosis. Therefore, the evaluation of HF risk is relevant for clinical decision making. The American College of Cardiology Foundation / American Heart Association (ACCF/AHA) HF guidelines mention the usefulness of validated scores to estimate mortality risk in patients hospitalized for HF (class IIa, level of evidence B). (1) However, a recent critical evaluation on these scores utility concluded that their application is still limited in clinical practice. (2)

In acute HF, there are risk scores with multiple variables of clinical use, with areas under the Receiver Operating Characteristic (ROC) curve (AUC) ranging between 0.59 and 0.80 to assess all-cause mortality during the first year. (3-14) The evaluation of these models, based on external validation cohorts, showed AUC between 0.69 and 0.81 for the GWTG-HF (Get With the Guidelines-Heart Failure), (15-18) 0.69 and 0.70 for the EFFECT (Enhanced Feedback for Effective Cardiac Treatment), (19, 20) 0.64 and 0.68 for the ADHERE (Acute Decompensated Heart Failure National Registry), (18, 19) and 0.74 for the OPTIMIZE-HF (Organized Program to Initiate Lifesaving Treatment in Hospitalized Patients with Heart Failure) (21) scores, most with evidence of adequate calibration.

A systematic review on HF predictive models identified 117 scores including 249 different variables, blood urea nitrogen and sodium levels being the most important predictors. (22) Mortality was better predicted in prospective registries, using a larger number of clinical variables. In this review, the average global AUC for all models was 0.66, and 0.71, 0.68 and 0.63 for those which predicted mortality, hospitalization for HF or both, respectively.

Since prediction of mortality in HF patients is still only moderately successful, artificial intelligence algorithms have been postulated to assess risk in acute conditions. Recently, a model based on deep learning attained 0.88 and 0.79 AUC to predict in-hospital and 12-month mortality for acute HF. (16) Despite these results, another contemporary study demonstrated that although automatic learning algorithms outperformed logistic regression to predict 30-day readmissions for decompensated HF, the improvements were only marginal, with AUC between 0.61 and 0.78, deep learning and a Naïve Bayes algorithm yielding the best results. (23-26)

The aim of this study was to develop and validate a model based on neural network (NN) algorithms to improve the performance of traditional models to predict short- and mid-term (30 days, 6 months and 1-year) mortality in patients with acute HF.

METHODS

A prospective clinical database was analyzed including 483 patients admitted with diagnosis of acute HF in the coronary care unit of a community hospital of Buenos Aires, between June 2005 and June 2019. The entire database comprised 181 demographic, laboratory, imaging, treatment and follow-up variables, among which only 25 variables were selected to calculate five acute heart failure risk scores aimed to predict 30-day, 6-month and 1-year mortality.

The models used for the analysis were: EFFECT

(4), ADHERE (3), GWTG-HF (5), 3C-HF (7) and ACUTE-HF (8). Table 1 shows the variables included in each calculation. Two models, ESCAPE (Evaluation Study of Congestive Heart Failure and Pulmonary Artery Catheterization Effectiveness) (6) and OPTIMIZE-HF (10, 13, 21) were excluded from the study due to the absence of complete data, such as BNP values on discharge in the ESCAPE risk score, and because patients had to be treated with milrinone for hemodynamic instability during 48-72 hours in the case of the OPTIMIZE-HF score. The primary endpoint was 30-day (in-hospital), 6 month and 1-year all-cause mortality

Among the 25 predictors included in the five traditional model, "black ethnicity" was excluded, and the other 24 were used to test the models based on NN algorithms: one or two-hidden layer multilayer perceptron (MLP), and a radial basis function (RBF) network. To this end, the database was divided into two groups: 70% of cases to test the NN algorithms and 30% for validation. An NN algorithm is a special type of non-linear regression that presents multiple minimum local values, and hence, each time the training algorithm is executed it will converge into a different model. In order to choose the best model, training was repeated 50 times for each NN model. Simultaneously, each time the validation cohort models were tested, accuracy, AUC, and negative (NPV) and positive predictive value (PPV) were recorded. Only NN models with the best discrimination power were selected for comparison with predictions from the five traditional models.

All models based on MLP NN were implemented with the input layer covariate normalization method, hidden layer hyperbolic tangent activation functions, an output layer *softmax* activation function and a cross-entropy error function. The RBF model was also implemented with the input layer covariate normalization method, a hidden layer Gaussian or *softmax* activation function, an output "identity" activation function and a sum of squares error function.

Statistical analysis

Categorical variables were expressed as absolute frequencies and percentages, and continuous variables as mean, standard deviation or median and interquartile range (IQR). The Kolmogorov-Smirnov Goodness-of-fit test was used to analyze normal distribution. The Hanley-MacNeil test was applied to compare AUC with their corresponding 95% confidence intervals. Calibration was evaluated with the Hosmer-Lemeshow chi-square test. All the models were also compared with respect to their predictive variables through a hierarchical cluster analysis, in order to identify subgroups sharing the same predictors. IBM SPSS 23.0 Statistics (IBM Corporation, Armonk, NY) software package was used for

Table 1. Variables included in heart failure risk scores to predict 30-day, 6-month and 1-year mortality

Variables	EFFECT	GWTG-HF	ADHERE	3C-HF	ACUTE-HF
Age	X	X		X	X
Heart rate		X			
Respiratory frequency	X				
NYHA FC III-IV				X	
Systolic blood pressure	X	X	X		
Hypertension				X	
Blood urea nitrogen	X	X	X		
Creatinine			X		X
Sodium	X	X			
Cerebrovascular disease*	X				X
Dementia	X				
COPD	X	X			
Liver Cirrhosis	X				
Cancer	X				
Hemoglobin/anemia	X			X	
Black ethnicity		X			
Target organ damage				X	
Chronic kidney disease				X	
Atrial fibrillation				X	
No beta blocker				X	
No ACEI				X	
Low LVEF				X	X
Severe valvular heart disease**				X	X
Non-invasive ventilation					X
Prior hospitalization					X

*It includes stroke and transient ischemic attack.
**It includes moderate mitral regurgitation.
FC: Functional class. COPD: Chronic obstructive pulmonary disease. ACEI: Angiotensin-converting enzyme inhibitor. LVEF: Left ventricular ejection fraction.

statistical analysis and NN modeling. A p value ≤0.05 was considered statistically significant.

Ethical considerations

The Institutional Ethics Committee approved the study, waiving an informed consent due to the observational nature of the study.

RESULTS

Table 2 shows baseline population characteristics used to calculate traditional models and for NN algorithm training and validation.

Figure 1 shows ROC curves and Table 3 summarizes the performance of 30-day, 6-month and 1-year mortality predictive models for HF patients. In terms of 30-day discrimination, the EFFECT score was better than the ACUTE-HF score (Hanley-McNeil p=0.041) and the 3C-HF score better than the AUTE-HF score (p=0.047). At 6-month and 1-year follow-up, the EFFECT score outperformed the ADHERE score (p=0.011 and p=0.003), respectively, while the EFFECT (p<0.001 and p<0.001), GWTG-HF (p=0.001 and p=0.006) and 3C-HF (p=0.001 and p=0.002) scores outperformed the ACUTE-HF score, respectively, in the same time periods.

The best results with NN algorithms were obtained with the two-hidden layer MLP (29-9-7-2-layer architecture). This NN model characteristics are shown in Figure 2, and its performance is summarized in Table 4.

In terms of discrimination, NN algorithms were superior to traditional models (Hanley-McNeil p <0.001) for 30-day, 6-month and 1-year mortality, respectively. With respect to the rest of NN algorithms, the AUC varied between 0.81 and 0.82 and between 0.75 and 0.78 for one-hidden layer LMP and the RBF model, respectively, for the same time periods (Figure 3).

Figure 4 shows the independent normalized importance of NN algorithm variables. For this algorithm, the most influential factors in descending order that obtained ≥50% "normalized importance" to predict 30-day mortality were: serum creatinine and blood urea nitrogen, hemoglobin, respiratory rate, sodium

Table 2. Baseline population characteristics (n=483)

Variables	n (%)
Age, years (mean±SD)	78±11.1
Male gender	279 (57.8)
NYHA functional class III-IV dyspnea	90 (18.6)
Hypertension	438 (90.7)
Diabetes	102 (21.1)
Etiology:	
Ischemic	169 (35.0)
Hypertensive	123 (25.5)
Valvular heart disease	96 (19.9)
Other etiologies	95 (19.7)
Chronic kidney disease	78 (16.1)
Stroke	58 (12.0)
Chronic pulmonary disease	77 (15.9)
Anemia	77 (15.9)
Atrial fibrillation	239 (49.5)
Moderate/severe mitral regurgitation	75 (15.5)
Non-invasive ventilation	286 (59.2)
Dementia	36 (4.6)
Cancer	71 (14.7)
Beta blocker	239 (49.5)
ACEI	149 (30.8)
Blood urea nitrogen, mg% (median and IQR)	51 (38-71)
Hemoglobin, g% (mean±SD)	12.9±3.9
Serum creatinine, mg% (mean±SD)	1.3±0.96
Serum sodium, mEq/L (mean±SD)	136±10.1
Admission systolic blood pressure, mmHg (median and IQR)	142 (130-160)
Heart rate, bpm (mean±SD)	91±25
Respiratory frequency/min (mean±SD)	19±2
Left ventricular ejection fraction, % (median and IQR)	52 (35-60)

SD: Standard deviation. IQR: Interquartile range. ACEI: Angiotensin-converting enzyme inhibitor.

Table 3. Predictive models of 30-day, 6-month and 1-year mortality in heart failure patients

	EFFECT	GWTG-HF	ADHERE	3C-HF	ACUTE-HF
30-day mortality:					
Area under the ROC curve	0.68	0.66	0.59	0.67	0.54
95% CI	0.59-0.77	0.58-0.74	0.51-0.68	0.59-0.75	0.44-0.63
Hosmer-Lemeshow χ^2	9.68	10.0	3.10	12.3	16.7
degrees of freedom	8	8	2	8	6
p value	0.289	0.262	0.212	0.138	0.011
6-month mortality:					
Area under the ROC curve	0.69	0.68	0.58	0.67	0.53
95% CI	0.63-0.75	0.62-0.74	0.52-0.64	0.61-0.73	0.47-0.60
Hosmer-Lemeshow χ^2	6.05	6.55	2.93	5.59	9.84
degrees of freedom	8	8	2	8	6
p value	0.641	0.586	0.231	0.693	0.132
1-year mortality:					
Area under the ROC curve	0.69	0.66	0.57	0.67	0.56
95% CI	0.64-0.74	0.61-0.72	0.51-0.63	0.62-0.72	0.51-0.62
Hosmer-Lemeshow χ^2	4.65	11.8	3.03	2.98	6.68
degrees of freedom	8	8	2	8	6
p value	0.794	0.163	0.220	0.936	0.352

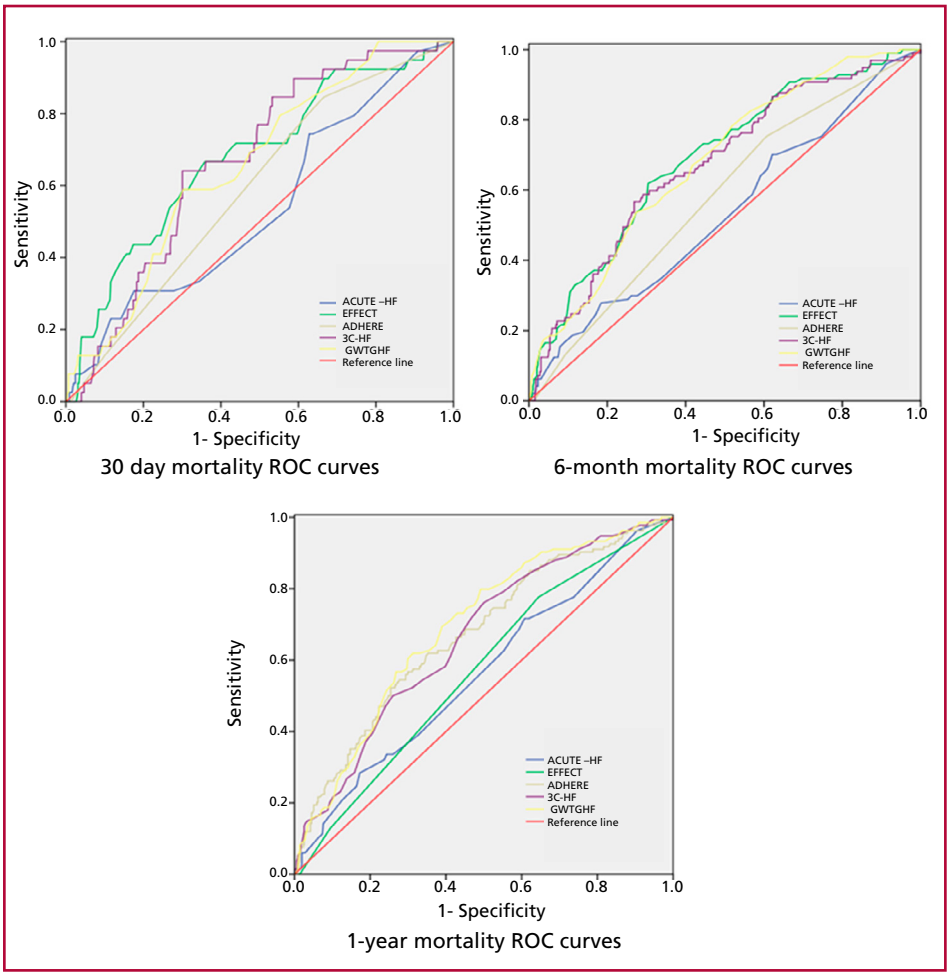

Fig. 1. ROC curves of the different predictive models.

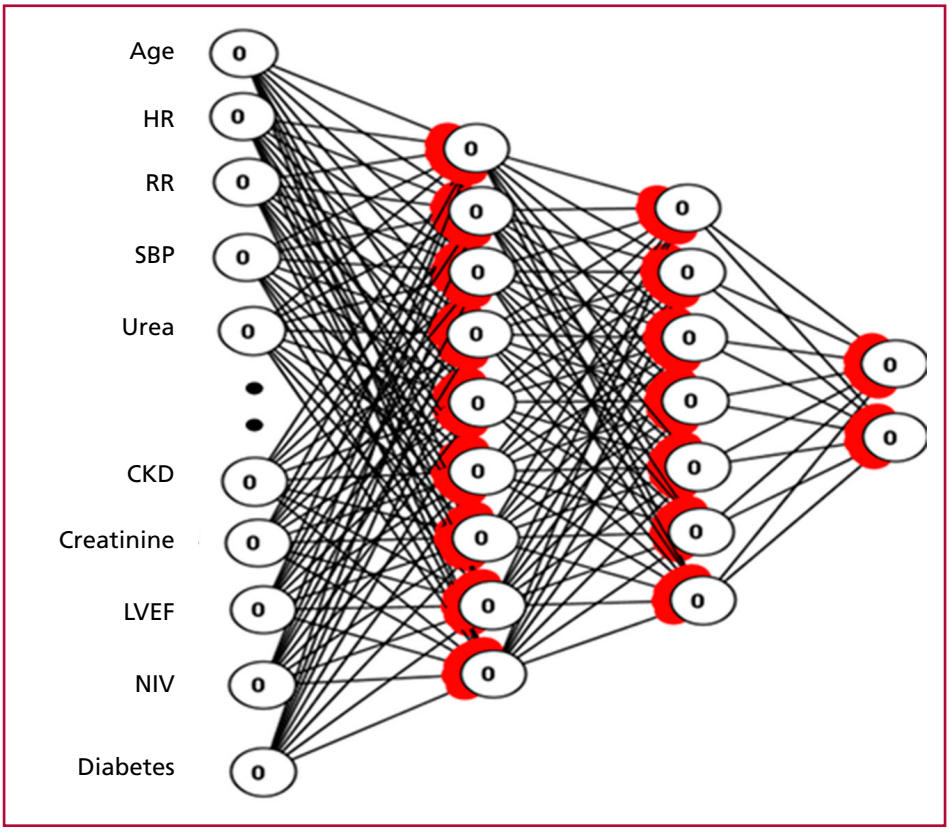

Fig. 2. Two-hidden layer multilayer perceptron neural network architecture

HR: heart rate. RR: respiratory rate. SBP: systolic blood pressure. CKD: chronic kidney disease. LVEF: left ventricular ejection fraction. NIV: non invasive ventilation

Table 4. Neural network model performance to predict 30-day, 6-month and 1-year heart failure mortality

	Accuracy (95% CI)	Area under the ROC curve (95% CI)	NPV (95% CI)	PPV (95% CI)
30-day mortality:				
two-hidden layer MLP	92.9%	0.82	93.2%	66.7%
(29-9-7-2-layer architecture)	(90.5-95.3%)	(0.79-0.85)	(90.9-95.6%)	(28.9-100%)
6-month mortality				
two-hidden layer MLP	87.7%	0.87	89.1%	78.6%
(29-9-7-2-layer architecture)	(84.7-90.8%)	(0.84-0.90)	(85.9-92.2%)	(67.8-89.3%)
1-year mortality				
two-hidden layer MLP	84.4%	0.85	85.6%	78.9%
(29-9-7-2-layer architecture)	(81.0-87.8%)	(0.81-0.88)	(81.9-89.2%)	(69.8-88.1%)

MLP multilayer perceptron

concentration, age and systolic blood pressure. In addition, NYHA functional class III-IV and dementia were associated with higher 6-month mortality, and heart rate and chronic kidney failure with 1-year mortality.

Finally, a hierarchical cluster analysis was performed on the prediction variables of all scores, to identify model subgroups with similar characteristics. The subgroups created are presented in the Figure 5 dendrogram showing the average link between scores. Based on this dendrogram, the NN algorithm revealed more similitudes with the ADHERE, EFFECT and GWTG-HF scores. When only the most influential factors obtaining ≥50% "standardized importance" were considered, the NN algorithm shared 100% of predictors with the ADHERE score, 64% with the EFFECT score and 57% with the GWTG-HF score.

DISCUSSION

In the present study, the model based on a NN algorithm outperformed traditional models in predicting short-and mid-term mortality in patients with acute HF. The two-hidden layer MLP perceptron model not

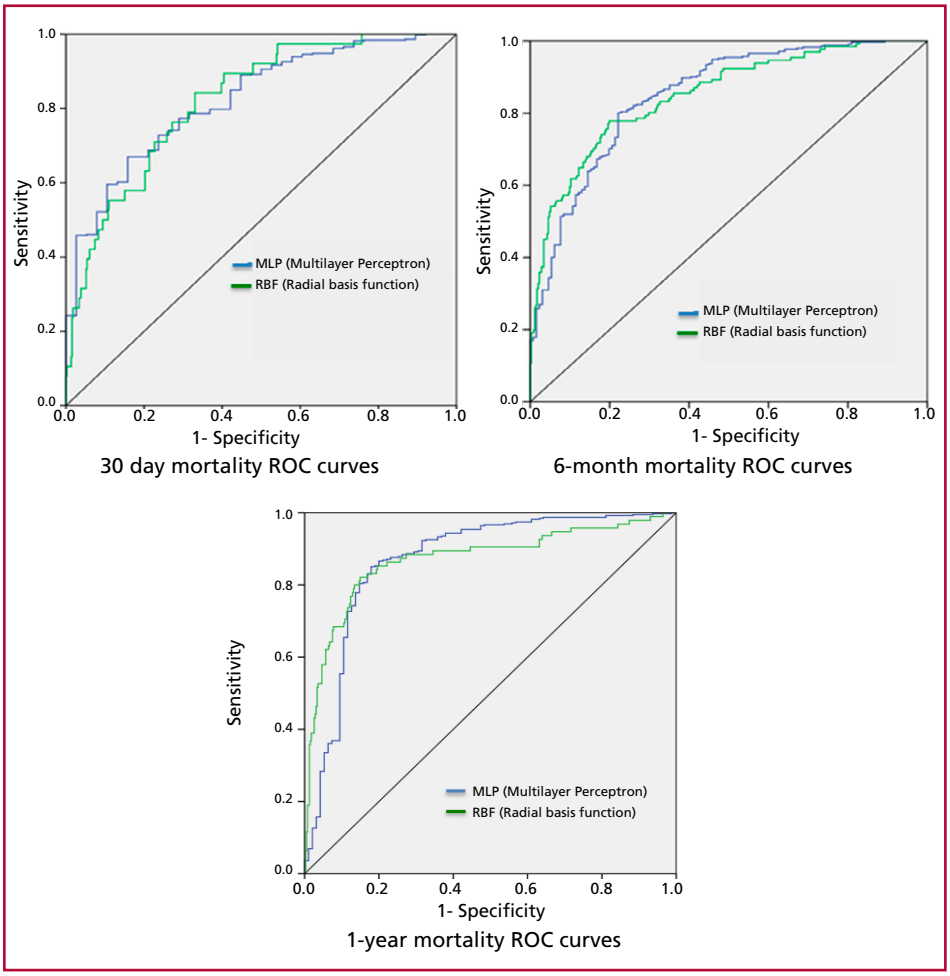

Fig. 3. ROC curves for multilayer perceptron (MLP) and radial basis function (RBF) models

only statistically improved overall discrimination, but also preserved a good NPV and PPV performance up to 1-year follow-up. This point is crucial, as most NN algorithms tend to improve their results mainly based on an increase in NPV instead of PPV. Although only the same 24 predictors included in any of the 5 traditional models were used for NN algorithm training and validation, this novel methodological approach was enough to significantly improve the prediction of results.

Among the 24 variables, serum creatinine and blood urea nitrogen, hemoglobin, respiratory rate, sodium concentration, age and admission systolic blood pressure were the most influential ones to predict short- and mid-term mortality in NN algorithms. In addition to these variables, NYHA functional class III-IV and dementia were associated with higher 6-month mortality, and heart rate and chronic kidney disease with 1-year mortality.

On many occasions, NN algorithms have been criticized for being considered as a "black box" with limited capacity to identify possible causal relationships. In the present study, the most influential factors were identified through values of standardized relevance. The analysis of independent predictive variable importance calculates their individual weight in the NN algorithm through a sensitivity analysis based on sample testing. In addition, as a result of hierarchical cluster analysis, similarities and relationships between the NN algorithm and the rest of predictive scores were determined.

An improvement in the predictive accuracy would be useful for HF patients, mainly for those with worse prognosis that could benefit with a more focused, aggressive treatment and closer follow-up. These improved scores could also help in the design of clinical trials, facilitating the choice of the population with higher potential rate of events. Up to the present, most studies have demonstrated that the prediction of mortality, particularly in patients hospitalized for HF, still has limited success, without significant differences in the discrimination value between patients with chronic or acute HF. A systematic review making reference to the discrimination power of HF risk scores, demonstrated that 69 of 117 models did not present external validation. These models probably overestimated the predictive capacity by using an internal validation based on the bootstrap method. (22)

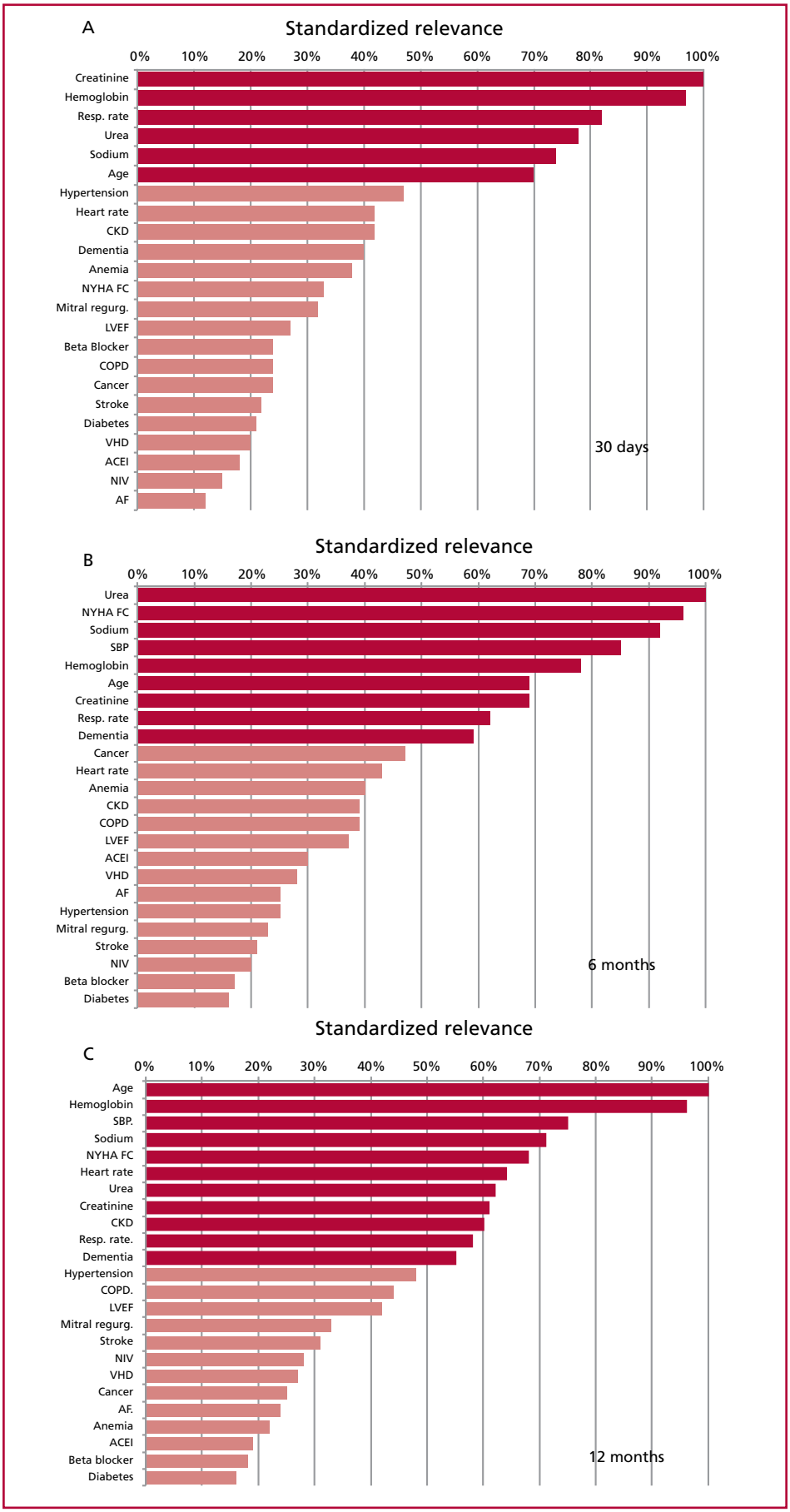

Fig. 4. Normalized importance of two-hidden layer multilayer perceptron variables to predict (a) 30-day, (b) 6-month and (c) 1-year mortality in patients with acute heart failure.
CKD: chronic kidney disease. FC: functional class. LVEF: left ventricular ejection fraction. COPD: Chronic obstructive pulmonary disease. VHD: valvular heart disease. ACEI: angiotensin converting enzyme inhibitors. NIV: non invasive ventilation. AF: atrial fibrillation. SBP: systolic blood pressure

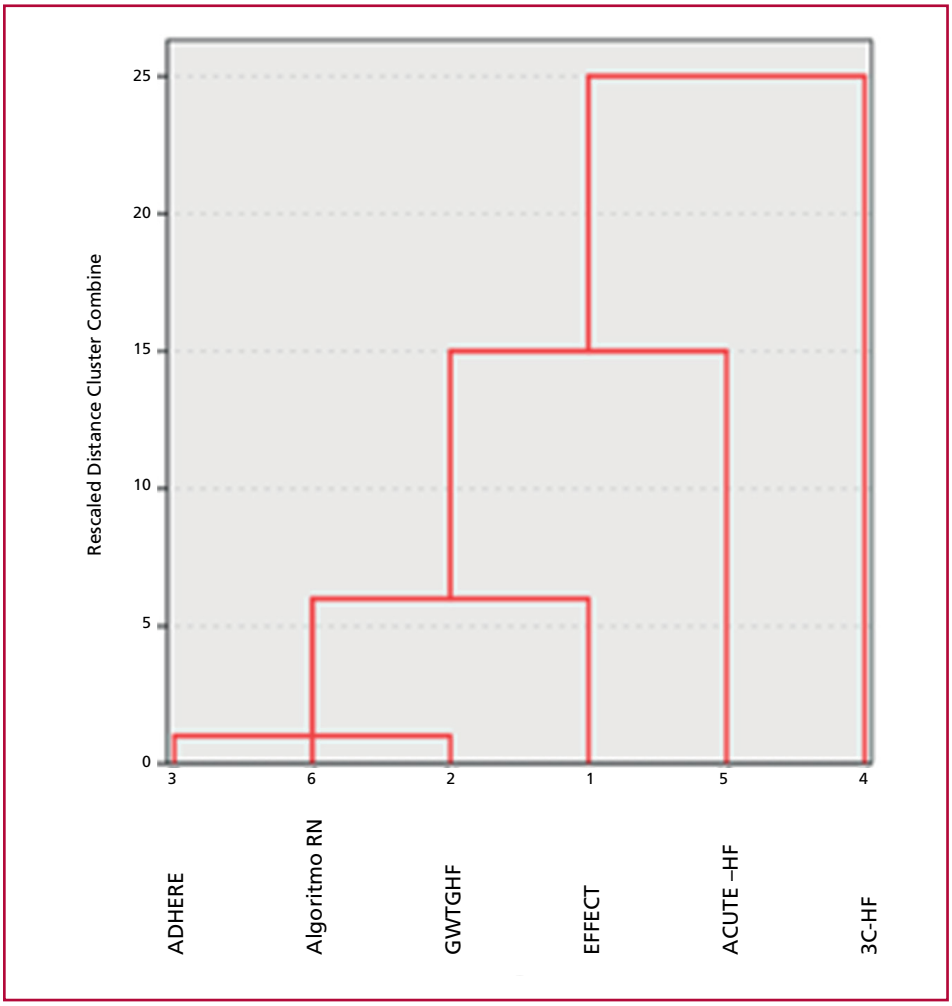

Fig. 3. ROC curves for multilayer perceptron (MLP) and radial basis function (RBF) models

It is expected that the latter models will report higher AUC than studies with models validated in a different patient population.

Similarly, cohort and prospective studies usually generate higher AUC than models based on retrospective analyses. The AUC of our predictive model were 0.82 and 0.85 for 30-day and 1-year mortality, respectively, results which were similar to those obtained by Kwon et al. (16), who used a deep learning algorithm (AUC: 0.88 and 0.79 for the same follow-up period, respectively). Moreover, Kwon's NN algorithm outperformed other machine learning methods, as logistic regression, random forest, support vector machine (SVM) and Bayesian networks. To the best of our knowledge, only four other studies have used NN algorithms or traditional machine learning methods in HF patients, but in these cases 30-day readmissions instead of all-cause mortality were evaluated. (23-26) Recently, a machine learning technique using SVM with Gaussian nucleus was employed to validate a new and simple model able to predict short- and long-term mortality in patients with acute HF. However, this 6-factor score called HANBAH (acronym for hemoglobin, age, sodium, blood urea nitrogen, atrial fibrillation and high-density lipoprotein) only attained an AUC of 0.75 at its best performance. (27) Considering that one of our NN algorithms, such as RBF, was equivalent to SVM with Gaussian nucleus, our AUC values between 0.75 and 0.78 were similar to those obtained with the HANBAH score.

As previously mentioned, in medical literature, artificial intelligence methods have generally demonstrated that although most of their models perform with better discrimination and accuracy level than traditional scores, they also evidence high NPV and low PPV. Improved precision for predicting in-hospital and mid-term mortality after admission for acute HF is important to identify persons needing intensified treatment and care. But the individual risk could be better calculated by means of sufficiently precise scores showing higher PPV instead of high NPV. In medical conditions of low incidence of adverse results or events, the overall precision of predictive risk models can be exaggerated with high NPV, though with low PPV. Consequently, the best models should be based specially on PPV and sensitivity. In the present study, the NN algorithm showed high precision levels with an acceptable PPV, approximately between 67% and 79%.

Different from the conventional statistical approach, the NN algorithm does not require the preselection of significant variables, since the least important factors are automatically ignored in the model adjust process. Moreover, this model does not limit the number of input predictors, and can use all the available information from a database without losing power. However, in the present study, we guided variable selection based only on predictors included in previous scores. Even though the same input variables were used, the improved performance can be explained because NN algorithms can detect non-linear relationships between independent and dependent variables beyond the reach of logistic regression.

Our study has certain limitations. Firstly, although the most relevant variables upon which the NN predictive algorithm was based could be identified, as well as the relationships between this new model and traditional scores using a dendrogram, in a certain way the NN algorithm is still a "black box", since we cannot interpret the approach used to classify individual patient risk. Secondly, this predictive model was developed with a limited number of variables obtained from a single center database. Thirdly, candidate variables for predictive models of HF are originally selected from clinical variables acquired from previously published studies, where multiple logistic regression analyses were performed to eliminate those factors not associated with short- or mid-term mortality. Therefore, the input variables used in the current NN algorithm should be considered to have been guided at least by a previous logistic regression analysis. This paradox could generate some bias when comparing the performance between the NN algorithm and traditional scores based on logistic regressions.

CONCLUSION

The present study analyzed the utility of NN to predict in-hospital and mid-term mortality in patients with acute HF. Using the individual predictors included in the 5 traditional scores (EFFECT, ADHERE, GWTG-HF, 3C-HF and ACUTE-HF) as NN input variables, it was shown that the algorithm based on artificial intelligence was more accurate and with better discrimination power than the above-mentioned scores. The NN-based analysis constitutes an alternative model that improves the results of traditional scores.

Conflicts of interest

None declared.

(See authors' conflict of interests forms on the web/Additional material.)

REFERENCES

1. Yancy CW, Jessup M, Bozkurt B, Butler J, Casey DE Jr, Drazner MH, et al. 2013 ACCF/AHA guideline for the management of heart failure: executive summary: a report of the American College of Cardiology Foundation/American Heart Association Task Force on Practice Guidelines. J Am Coll Cardiol 2013;62:1495-539. https://doi.org/10.1016/j.jacc.2013.05.020

2. Ferrero P, Iacovoni A, D'Elia E, Vaduganathan M, Gavazzi A, Senni M. Prognostic scores in heart failure - Critical appraisal and practical use. Int J Cardiol 2015;188:1-9.https://doi.org/10.1016/j.ijcard.2015.03.154

3. Fonarow GC, Adams KF Jr, Abraham WT, Yancy CW, Boscardin WJ; ADHERE Scientific Advisory Committee, Study Group, and Investigators. et al. Risk stratification for in-hospital mortality in acutely decompensated heart failure: classification and regression tree analysis. JAMA 2005;293:572-80. https://doi.org/10.1001/jama.293.5.572

4. Lee DS, Austin PC, Rouleau JL, Liu PP, Naimark D, Tu JV, et al. Predicting mortality among patients hospitalized for heart failure: derivation and validation of a clinical model. JAMA 2003;290:2581-7. https://doi.org/10.1001/jama.290.19.2581

5. Peterson PN, Rumsfeld JS, Liang L, Hernandez AF, Peterson ED, Fonarow GC, 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. https://doi.org/10.1161/CIRCOUTCOMES.109.854877

6. O'Connor CM, Hasselblad V, Mehta RH, Tasissa G, Califf RM, Fiuzat M, et al. Triage after hospitalization with advanced heart failure: the ESCAPE (Evaluation Study of Congestive Heart Failure and Pulmonary Artery Catheterization Effectiveness) risk model and discharge score. J Am Coll Cardiol 2010;55:872-8.https://doi.org/10.1016/j.jacc.2009.08.083

7. Senni M, Parella P, De Maria R, Cottini C, Böhm M, Ponikowski P, et al. Predicting heart failure outcome from cardiac and comorbid conditions: The 3C-HF score. Int J Cardiol 2013;163:206-11.https://doi.org/10.1016/j.ijcard.2011.10.071

8. Cameli M, Pastore MC, De Carli G, Henein MY, Mandoli GE, Lisi E, et al. ACUTE HF score, a multiparametric prognostic tool for acute heart failure: A real-life study. Int J Cardiol 2019;296:103-8. https://doi.org/10.1016/j.ijcard.2019.07.015

9. Lee DS, Stitt A, Austin PC, Stukel TA, Stukel TA, Schull MJ, Chong A, et al. Prediction of heart failure mortality in emergent care: a cohort study. Ann Intern Med 2012;156:767-75. https://doi.org/10.7326/0003-4819-156-11-201206050-00003

10. Felker GM, Leimberger JD, Califf RM, Cuffe MS, Massie BM, Adams KF Jr, et al. Risk stratification after hospitalization for decompensated heart failure. J Card Fail 2004;10:460-6. https://doi.org/10.1016/j.cardfail.2004.02.011

11. Hsieh M, Auble TE, Yealy DM. Validation of the Acute Heart Failure Index. Ann
Emerg Med 2008;51:37-44. https://doi.org/10.1007/s12035-008-8015-2

12. O'Connor CM, Mentz RJ, Cotter G, Metra M, Cleland JG, Davison BA, et al. The PROTECT in-hospital risk model: 7-day outcome in patients hospitalized with acute heart failure and renal dysfunction. Eur J Heart Fail 2012;14:605-12. https://doi.org/10.1093/eurjhf/hfs029

13. Abraham WT, Fonarow GC, Albert NM, Stough WG, Gheorghiade M, Greenberg BH, et al. OPTIMIZE-HF Investigators and Coordinators. 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; 29:347-56. https://doi.org/10.1016/j.jacc.2008.04.028

14. 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;64:441-9. https://doi.org/10.1016/j.jjcc.2014.03.002

15. Shiraishi Y, Kohsaka S, Abe T, Mizuno A, Goda A, Izumi Y, et al; West Tokyo Heart Failure Registry Investigators. Validation of the Get With The Guideline-Heart Failure risk score in Japanese patients and the potential improvement of its discrimination ability by the inclusion of B-type natriuretic peptide level. Am Heart J 2016;171:33-9. https://doi.org/10.1016/j.ahj.2015.10.008

16. Kwon JM, Kim KH, Jeon KH, Lee SE, Lee HY, Cho HJ, et al. Artificial intelligence algorithm for predicting mortality of patients with acute heart failure. PLoS One 2019;14: e0219302. https://doi.org/10.1371/journal.pone.0219302

17. Yagyu T, Kumada M, Nakagawa T. Novel risk stratification with time course assessment of in-hospital mortality in patients with acute heart failure. PLoS One 2017;12:e0187410. https://doi.

org/10.1371/journal.pone.0187410

18. Win S, Hussain I, Hebl VB, Dunlay SM, Redfield MM. Inpatient Mortality Risk Scores and Postdischarge Events in Hospitalized Heart Failure Patients: A Community-Based Study. Circ Heart Fail 2017;10:e003926. https://doi.org/10.1161/CIRCHEARTFAILURE.117.003926

19. Lagu T, Pekow PS, Shieh MS, Stefan M, Pack QR, Kashef MA, et al.Validation and Comparison of Seven Mortality Prediction Models for Hospitalized Patients With Acute Decompensated Heart Failure. Circ Heart Fail 2016;9:e002912. https://doi.org/10.1161/CIRCHEARTFAILURE.115.002912

20. Martín-Sánchez FJ, Gil V, Llorens P, Herrero P, Jacob J, Fernández C, et al. Acute Heart Failure Working Group of the Spanish Society of Emergency Medicine Investigation Group. Barthel Index-Enhanced Feedback for Effective Cardiac Treatment (BI-EFFECT) Study: contribution of the Barthel Index to the Heart Failure Risk Scoring System model in elderly adults with acute heart failure in the emergency department. J Am Geriatr Soc 2012;60:493-8. https://doi.org/10.1111/j.1532-5415.2011.03845.x

21. Yap J, Lim FY, Chia SY, Allen JC Jr, Jaufeerally FR, Macdonald MR, et al. Prediction of Survival in Asian Patients Hospitalized With Heart Failure: Validation of the OPTIMIZE-HF Risk Score. J Card Fail 2019;25:571-5. https://doi.org/10.1016/j.cardfail.2019.02.016

22. Ouwerkerk W, Voors AA, Zwinderman AH. Factors influencing the predictive power of models for predicting mortality and/or heart failure hospitalization in patients with heart failure. JACC Heart Fail 2014;2:429-36. https://doi.org/10.1016/j.jchf.2014.04.006

23. Mortazavi BJ, Downing NS, Bucholz EM, Dharmarajan K, Manhapra A, Li SX, et al. Analysis of Machine Learning Techniques for Heart Failure Readmissions. Circ Cardiovasc Qual Outcomes 2016;9:629-40. https://doi.org/10.1161/CIRCOUTCOMES.116.003039

24. Frizzell JD, Liang L, Schulte PJ, Yancy CW, Heidenreich PA, Hernandez AF, et al. Prediction of 30-Day All-Cause Readmissions in Patients Hospitalized for Heart Failure. JAMA Cardiol 2017;2:204. https://doi.org/10.1001/jamacardio.2016.3956

25. Golas SB, Shibahara T, Agboola S, Otaki H, Sato J, Nakae T, 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. https://doi.org/10.1186/s12911-018-0620-z

26. Shameer K, Johnson KW, Yahi A, Miotto R, Li LI, Ricks D, et al. Predictive modeling of hospital readmission rates using electronic medical record-wide machine learning: a case-study using Mount Sinai Heart Failure Cohort. Pac Symp Biocomput 2017;22:276-87. https://doi.org/10.1142/9789813207813_0027

27. Guo CY, Chan CH, Chou YC, Sung SH, Cheng HM. A Statistical Predictive Model Consistent Within a 5-Year Follow-up Period for Patients with Acute Heart Failure. J Chin Med Assoc 2020;83:1008-13. https://doi.org/10.1097/JCMA.0000000000000403