Revista Colombiana de Cardiología
Órgano oficial de la Asociación Sociedad Colombiana de Cardiología & Cirugía Cardiovascular
Permanyer

Check for updates

ARTÍCULO ORIGINAL

Ensamble de redes neuronales artificiales ponderado mediante características operativas para el pronóstico de la insuficiencia cardiaca aguda

Ensemble of artificial neural networks weighted by operational characteristics for prognosis of acute heart failure

John J. Sprockel[1,2,3], Andrés Fandiño[1,2], Walter G. Chaves[1,2], Christian O. Benavides[4] y Juan J. Diaztagle[1,2,5]*

[1]Departamento de Medicina Interna, Hospital de San José de Bogotá; [2]Facultad de Medicina, Fundación Universitaria de Ciencias de la Salud; [3]Instituto de Investigaciones, Fundación Universitaria de Ciencias de la Salud; [4]Departamento de Ingeniería de Sistemas, Universidad de San Buenaventura; [5]Departamento de Ciencias Fisiológicas, Universidad Nacional de Colombia. Bogotá, Colombia

Resumen

Introducción: *La insuficiencia cardíaca es una condición común, progresiva y potencialmente mortal cuyo riesgo suele sobreestimarse. Se requieren herramientas eficaces para discriminar el riesgo, por lo que se entrenó un sistema basado en el ensamble de redes neuronales para este fin.* **Objetivo:** *Exponer los resultados del entrenamiento y la validación interna de un sistema basado en el ensamble de un conjunto de redes neuronales artificiales para el pronóstico de la mortalidad a un mes de los pacientes hospitalizados por insuficiencia cardíaca aguda, y comparar los resultados de cada una de las redes individuales desarrolladas y cuatro sistemas de ensamble, la votación simple y AdaBoost.* **Materiales y método:** *A partir de una cohorte de 462 pacientes con diagnóstico de insuficiencia cardiaca descompensada se entrenaron once redes que luego se ensamblaron en cuatro sistemas: votación simple, dos sistemas ponderados por características operativas (valores predictivos y likelihood ratios) y Boosting. Se calcularon las características operativas para el pronóstico de muerte a 30 días y se compararon con el de dos reglas clínicas y una regresión logística aplicada a la misma población.* **Resultados:** *Los diversos métodos de ensamble obtuvieron un mejor rendimiento pronóstico que el de cada una de las redes que lo componían. La votación ponderada por valores predictivos muestra el mejor desempeño, con una exactitud del 89.0% (IC 95%: 82.6-93.2%), aunque los intervalos de confianza se superponían entre los resultados.* **Conclusiones:** *El ensamble de redes neuronales mediante votación ponderada por valores predictivos demostró un adecuado rendimiento para el pronóstico de muerte a treinta días en insuficiencia cardiaca aguda.*

Palabras clave: *Inteligencia artificial. Redes neuronales. Insuficiencia cardiaca. Pronóstico. Mortalidad.*

Abstract

Introduction: *Heart failure is a common, progressive, and life-threatening condition whose risk is often overestimated. Effective tools are required to discriminate the risk and therefore a system based on the assembly of neural networks was trained for this purpose.* **Objective:** *To present the results of the training and internal validation of a system based on a set of artificial*

***Correspondencia:**
John J. Sprockel
E-mail: jjsprockel@fucsalud.edu.co

Fecha de recepción: 27-08-2022
Fecha de aceptación: 31-03-2023
DOI: 10.24875/RCCAR.22000065

Disponible en internet: 06-11-2023
Rev Colomb Cardiol. 2023;30(5):235-242
www.rccardiologia.com

0120-5633 / © 2023 Sociedad Colombiana de Cardiología y Cirugía Cardiovascular. Publicado por Permanyer. Este es un artículo *open access* bajo la licencia CC BY-NC-ND (http://creativecommons.org/licenses/by-nc-nd/4.0/).

neural networks for the prognosis of one-month mortality in patients hospitalized for acute heart failure, and to compare the results of each of the individual networks developed and four set systems simple voting and AdaBoost. **Materials and method:** *From a cohort of 462 patients diagnosed with decompensated heart failure, 11 networks were trained and then assembled using four systems: simple voting, two systems weighted by operating characteristics (predictive values and likelihood ratios) and Boosting. Operating characteristics for the 30-day prognosis of death were calculated and compared with two clinical rules and logistic regression applied to the same population.* **Results:** *The various ensemble methods had a better prognostic performance than each of the networks that composed them. Voting weighted by predictive values performed best, with an accuracy of 89.0% (95% CI: 82.6-93.2%) although the results' confidence intervals overlapped.* **Conclusions:** *The ensemble of neural networks through voting weighted by predictive values showed an adequate performance for predicting 30-day mortality in acute heart failure.*

Keywords: *Artificial intelligence. Neural networks. Heart failure. Prognosis. Mortality.*

Introducción

La insuficiencia cardiaca (IC) es una condición clínica que alcanza niveles de epidemia y que tiene un alto impacto en los sistemas de salud pública. A través de los últimos años se ha documentado un aumento en su incidencia y prevalencia, sin ningún cambio en sus desenlaces clínicos[1]. Por tratarse de un síndrome clínico complejo, los pacientes con insuficiencia cardíaca tienen un amplio espectro de riesgo de mortalidad. Ante esta situación, la evaluación del pronóstico se convierte en un aspecto fundamental para la atención de esta entidad, en especial en el contexto de las hospitalizaciones[2].

La inteligencia artificial se ha posicionado como una alternativa a las herramientas estadísticas, principalmente a la regresión logística, dentro del campo de las reglas de predicción clínicas aplicables al diagnóstico o al pronóstico, con la promesa de alcanzar una mayor capacidad de generalización de sus resultados al ser aplicado a otras poblaciones[3]. Al interior de este campo de acción, ocupan un puesto preponderante el uso de redes neuronales artificiales y, de forma más reciente, los sistemas de ensamble.

Una de las herramientas más reconocidas en el mundo para el pronóstico de la IC es el árbol de decisiones del ADHERE, compuesta por tres variables (hemoglobina, creatinina y BUN) desarrollada en el 2005 mediante la técnica del análisis de árbol de clasificación y regresión (CARD, una estrategia de sistemas inteligentes), la cual alcanzó un AUC de 0.68 en el desarrollo[4] y entre 0.58 y 0.76 en diversos estudios de validación cruzada[5-7].

El objetivo de este trabajo es exponer los resultados del entrenamiento y validación interna de un sistema basado en el ensamble de un conjunto de redes neuronales artificiales para el pronóstico de la mortalidad a un mes, de los pacientes hospitalizados por insuficiencia cardíaca aguda, y comparar los resultados de cada una de las redes individuales desarrolladas y cuatro sistemas de ensamble, la votación simple y AdaBoost.

Materiales y método

Se utilizaron los datos de una cohorte prospectiva realizada en el Hospital San José de Bogotá, Colombia, entre febrero de 2010 y marzo de 2013[8]. Se incluyeron pacientes con diagnóstico de IC descompensada, hospitalizados por el servicio de medicina interna, mayores de 18 años, que cumplían criterios diagnósticos de Framingham. Como criterios de exclusión se tuvieron: descompensación aguda de diabetes *mellitus*, urgencia dialítica, cirrosis avanzada o insuficiencia hepática aguda, síndrome nefrótico, choque hipovolémico, choque séptico de cualquier origen y enfermedad neoplásica terminal.

Se recolectaron los datos sociodemográficos, clínicos, comorbilidades y examen físico, así como los siguientes laboratorios: nitrógeno ureico, creatinina, NT-proBNP, hemoglobina, sodio, troponina I de alta sensibilidad y electrocardiograma. Del reporte ecocardiográfico se obtuvo el valor de la fracción de eyección del ventrículo izquierdo. Los pacientes fueron seguidos hasta el egreso hospitalario y se documentó la mortalidad intrahospitalaria y a 30 días mediante llamada telefónica o verificación del sistema de registro nacional de defunciones (RUAF).

La selección de los datos de entrada se llevó a cabo a partir de los resultados de un estudio de cohorte previo en el cual se identificaron los factores asociados con la mortalidad mediante un análisis de regresión logística multivariada[8], complementado con las variables conocidas por su relación con mortalidad que han sido incluidas en la elaboración de las

reglas de predicción GWTG (get with the guidelines)[9], OPTIMIZE[10] y ADHERE[4].

Se incluyeron las siguientes 15 variables como datos de entrada: sexo, edad mayor de 70 años, hospitalizaciones previas, antecedentes de diabetes *mellitus*, enfermedad renal crónica, EPOC, enfermedad coronaria, tensión arterial sistólica mayor de 150 mmHg o menor de 100 mmHg, creatinina mayor de 2.7 mg/dl, sodio inferior a 130 meq/l, nitrógeno ureico mayor de 43 mg/dl, troponina mayor de 0.1 μg/dl, NT-proBNP mayor de 4.630 pg/ml y fracción de eyección del ventrículo izquierdo menor del 30%.

La población se dividió en dos partes: el 70% (323 pacientes) para entrenamiento y pruebas de las diferentes redes neuronales y el 30% restante (139 pacientes) para la validación del mecanismo de ensamble. Las redes neuronales fueron entrenadas mediante un algoritmo genético, compartían, dentro de su arquitectura, el estar conformadas por 15 variables de entrada, una capa oculta, una neurona de salida y función de activación sigmoidea. El número de neuronas de la capa oculta variaba entre 4, 6, 8, 10 y 12. El algoritmo genético seleccionó los pesos de las redes y fue programado en lenguaje Java. A su vez, se dividió a los 323 pacientes en un 70% para entrenamiento y un 30% para pruebas (con los 139 pacientes elegidos para la validación del mecanismo de ensamble, también se validaron las redes neuronales individuales).

La estrategia evolutiva estuvo organizada así:

- Representación de los individuos: dos vectores con los pesos que relacionan las neuronas de entrada y las de la capa oculta, y entre esta y la neurona de salida.
- Tamaño de la población: 250 individuos, 50 por cada grupo de neuronas en la capa oculta (4, 6, 8, 10 y 12).
- Inicialización de la población: se asignó de manera aleatoria el valor de los pesos.
- Mecanismo de evolución: se realizó mediante mutación (5%) y recombinación (95%).
- Selección: por torneo.
- Función de *fitness*: consistió en evaluar el resultado de cada red en cada uno de los ejemplos de entrenamiento; posteriormente, se calculaba el desempeño evaluando la precisión de cada una de ellas (número de casos correctamente discriminados sobre el total de casos evaluados), para luego seleccionar las más altas dentro de cada uno de los grupos.
- Criterio de terminación: 100 iteraciones.

Tras un gran número de experimentos con múltiples ejecuciones del algoritmo genético se seleccionaron 11 redes neuronales con el mejor desempeño basado en la exactitud, seguido por el resultado de la sensibilidad para realizar con ellas los ensambes.

Si se tiene la siguiente ecuación:

$$v_t(x_i)\begin{cases}1 & si \quad \rho \geq \eta \\ 0 & de\ otra\ forma\end{cases} \qquad (1)$$

donde $v_t(x_i)$ es el resultado del ensamble (0 negativo y 1 positivo).

Se pueden expresar los sistemas de ensamble de la siguiente forma:

- Votación simple: se compara la sumatoria de los positivos contra los negativos asignando como resultado el que tenga el mayor número de votos.
  En la ecuación 1, ρ es el contador de los resultados positivos y η el contador de los casos negativos.
- Votación ponderada por valores predictivos: se suma a un contador de positivos o negativos el respectivo valor predictivo positivo (VPP) multiplicado por dos o negativo (VPN) obtenido en la fase de pruebas.
  En este caso se reemplaza en la ecuación 1, $\rho = \sum_{si\ h_t=1} w_p$ y $\eta = \sum_{si\ h_t=0} w_n$, siendo h_t el resultado de la red neuronal. En este caso $w_p = 2 * VPP$ para los resultados positivos y $w_n = VPN$ para los negativos.
- Votación ponderada por razón de verosimilitud (likelihood ratio, LR): se efectuó obteniendo un valor que ponderaba por igual los resultados positivos o negativos de cada red neuronal; el factor de ponderación se obtenía de dividir el LR positivo sobre el LR negativo.
  Para la ecuación 1, $\rho = \sum_{si\ h_t=1} w_p$ y $\eta = \sum_{si\ h_t=0} w_n$, en este caso $w_p = w_n = LR^+ / LR^-$.
- AdaBoost: consiste en una votación ponderada cuyos pesos se obtenían a través del algoritmo de boosting más popular; entrenado a partir de la maximización de ε_t, la suma ponderada del error para los puntos mal clasificados:

$\epsilon_t = \sum_{h_t(x_i)\neq y_i}^{n} i = 1\ w_{i,t}$; donde $h_t(x_i)$ es el resultado dado por cada una de las redes neuronales para cada caso (x_i) del entrenamiento, y_i son las salidas esperadas para cada caso y $w_{i,t}$ son los pesos para cada uno de los clasificadores en el momento de entrenamiento *t*.

Luego se calcula α_t a partir de la fórmula $\alpha_t = \frac{1}{2} ln\left(\frac{1-\epsilon_t}{\epsilon_t}\right)$, con lo que se procede a actualizar los pesos $w_{i,t+1} = w_{i,t} e^{-y_i \alpha_t h_t(x_i)}$.

El análisis estadístico se realizó en el paquete estadístico STATA 12. La descripción de las variables continuas se hizo mediante medidas de tendencia central y de dispersión, mientras que las variables categóricas con frecuencias absolutas y relativas. Se calcularon las características operativas (sensibilidad, especificidad, valores predictivos y exactitud) para el pronóstico de muerte a 30 días de cada una de las redes y de los sistemas de ensamble. Se realizó el cálculo de los respectivos intervalos de confianza del 95% para cada uno de los resultados obtenidos.

Resultados

La base de datos de los pacientes estuvo constituida por 462 pacientes con promedio de edad de 72.4 (± 12.7) años, de los cuales 240 (51.9%) fueron mujeres; las comorbilidades que se presentaron con mayor frecuencia fueron hipertensión arterial crónica (80.5%), enfermedad pulmonar obstructiva crónica (43.7%) y diabetes *mellitus* tipo 2 (23.3%) (Tabla 1). La mortalidad a 30 días fue del 13.8% (64 pacientes). En promedio, la presión arterial sistólica al ingreso fue de 131.6 mmHg, 375 (81.1%) tuvieron una frecuencia cardiaca superior a 70 lpm. Entre los hallazgos de laboratorio relevantes, 47 (10.2%) presentaron niveles de creatinina mayor de 2 mg/dl y 98 (21.2%) hiponatremia.

En la población de pruebas, las redes neuronales individuales mostraron resultados de precisión entre 66 y 81%, sensibilidad entre 25 y 50% y especificidad entre 68 y 90% (Tabla 2). La mejor red neuronal alcanzó una exactitud del 81%, con una sensibilidad del 90% y una especificidad del 25% en la población de validación.

Los diversos métodos de ensamble obtuvieron un mejor rendimiento pronóstico que el de cada una de las redes que lo componían. La votación ponderada por valores predictivos mostró una tendencia hacia un mejor desempeño, con una exactitud del 89%, una sensibilidad del 26% y una especificidad del 99% en la población de validación; aunque existe una amplia superposición entre los intervalos de confianza de las diferentes herramientas de sistemas inteligentes (Tabla 3 y Fig. 1), se debe anotar que tuvo la mayor razón de disparidad diagnóstica, con 41.4.

Tabla 1. Descripción de las características de la población

Característica		
Edad en años, promedio (DE)	72.4	(12.7)
Sexo femenino, n (%)	240	(51.9)
Comorbilidades, n (%)		
Hipertensión arterial	372	(80.5)
Enfermedad coronaria	87	(18.8)
Diabetes *mellitus* tipo 2	108	(23.3)
Enfermedad renal crónica	66	(14.2)
Enfermedad pulmonar obstructiva crónica	202	(43.7)
Arritmia cardiaca	84	(18.1)
Hallazgos paraclínicos		
Creatinina, mediana (RIQ) mg/dl	1.0	(0.8-1.4)
Creatinina > 2 mg/dl, n (%)	47	(10.2)
Sodio, promedio (DE) meq/l	138.1	(5.5)
Hiponatremia, n (%)	98	(21.2)
BUN, mediana (RIQ) mg/dl	24	(18-35)
Troponina I, mediana (RIQ) µg/dl	0.04	(0.015-0.09)
Troponinas positivas, n (%)	41	(9.7)
NTproBNP, mediana (RIQ) pg/ml	4630	(1780-12068)
NTproBNP > 125 pg/ml, n (%)	426/438	(97.2)
Hemoglobina, promedio (DE) g/dl	13,6	(2.9)
Hemoglobina menor 12 g/dl, n (%)	144	(31.1)
FEVI menor del 40%, n (%)	143	(32.9)
Clase funcional ingreso (%)		
II	26	(5.6)
III	167	(36.1)
IV	269	(58.2)
Hospitalización previa por IC (%)	223	(47.3)
Medicación previa al ingreso (%)		
Betabloqueador	190	(41.1)
ARA II	163	(35.2)
IECA	155	(33.5)
Antagonista de la aldosterona	107	(23.1)
Marcapaso/Cardiodesfibrilador (%)	12/461	(2.6)

DE: desviación estándar; RIQ: rango intercuartílico; BUN: nitrógeno ureico; NTproBNP: fracción N terminal del propéptido natriurético auricular tipo B; IECA: inhibidores de la enzima convertidora de angiotensina; ARA II: antagonistas de los receptores de angiotensina II; IC: insuficiencia cardiaca; rpm: respiraciones por minuto; FEVI: fracción de eyección del ventrículo izquierdo.

Discusión

Los sistemas inteligentes se definen como aquellos intentos de emular la inteligencia humana, mientras favorecen el razonamiento y aprendizaje llevado a cabo en ambientes inciertos o con alto grado de imprecisión[11]; es un término que es sinónimo, hasta cierto punto, con los de inteligencia computacional y reconocimiento de patrones. Los sistemas basados en ensamble han sido postulados como una solución teórica para varios problemas de las reglas de predicción, principalmente la capacidad de generalización, pero también en aquellas situaciones en las que se cuenta con datos muy escasos[12].

En la tabla 4 se presentan los resultados de los diez estudios hallados en la literatura que evalúan diversos modelos de predicción de mortalidad intrahospitalaria y

Tabla 2. Características operativas de las once mejores redes neuronales entrenadas por el algoritmo genético (los resultados expresan el comportamiento en las diferentes poblaciones así: entrenamiento/pruebas/validación)

	Red neuronal artificial										
	1	2	3	4	5	6	7	8	9	10	11
Entradas	15	15	15	15	15	15	15	15	15	15	15
Neuronas en capa oculta	12	8	12	10	10	8	8	10	6	8	12
Sensibilidad	41/40/45	27/53/40	58/46/35	55/66/50	37/46/60	51/20/15	31/46/35	34/40/30	31/33/35	44/46/45	17/33/25
Especificidad	73/80/77	78/82/87	75/80/78	71/71/68	71/78/70	85/90/89	79/80/77	75/80/79	82/81/83	71/73/79	90/93/90
VPP	18/27/25	16/36/34	25/30/21	21/30/21	16/28/25	34/27/18	18/30/20	16/27/20	20/25/25	18/24/27	20/50/25
VPN	88/89/88	88/90/89	92/89/87	91/92/89	88/88/91	92/86/86	88/89/87	88/88/87	89/87/88	89/88/89	88/88/87
Exactitud	69/74/72	72/78/80	73/75/71	69/71/66	67/73/69	80/79/78	73/75/71	69/74/72	75/74/76	68/69/74	80/84/81

VPP: valor predictivo positivo; VPN: valor predictivo negativo.

Tabla 3. Características operativas de las herramientas evaluadas (en porcentajes con sus respectivos intervalos de confianza)

	Mejor red neuronal	Sistemas de ensamble			
		Votación simple	Ponderada por valores predictivos	Ponderado por LR	AdaBoost
Número de entradas	15	11	11	11	11
Características operativas, % (IC 95%)					
Sensibilidad	25.0 (11.2-46.9)	31.6 (15.4-54.0)	26.3 (11.8-48.8)	42.1 (23.1-63.7)	25.0
Especificidad	90.8 (84.2-94.8)	94.0 (88.2-97.1)	99.1 (95.3-99.8)	94.0 (88.2-97.1)	94.1
VPP	31.3 (14.2-55.6)	46.2 (23.2-70.9)	83.3 (43.6-97.0)	53.3 (30.1-75.2)	41.7
VPN	87.8 (80.9-92.5)	89.4 (82.8-93.7)	89.2 (82.7-93.5)	90.0 (84.5-94.8)	88.2
Exactitud	81.3 (74.0-86.9)	85.3 (78.4-90.3)	89.0 (82.6-93.2)	86.8 (80.0-91.5)	84.2 (78.1-90.2)

LR: likelihood ratio; VPP: valor predictivo positivo; VPN: valor predictivo negativo; IC: intervalo de confianza.

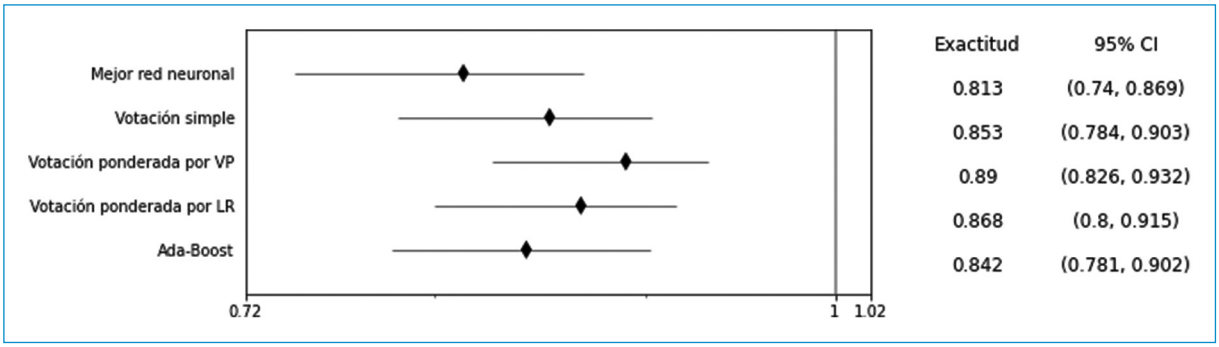

Figura 1. Forest plot con los resultados de la exactitud de la mejor red neuronal individual y de los modelos de ensamble con sus respectivos intervalos de confianza del 95%.

a 30 días mediante el uso de sistemas inteligentes[4,6,13-20]. Se encuentra, con frecuencia, el uso de árboles de decisiones (en 6 casos)[4,6,15-17,20] incluyendo 4 casos con *ramdom forest*[14,15], un modelo innovador (Patient specific Markov blanket global structure algoritm[19]), dos casos de Boosting[15,18] y dos utilizaron máquinas de vectores

Tabla 4. Estudios que evalúan diversos modelos de predicción de mortalidad en insuficiencia cardiaca mediante el uso de sistemas inteligentes

Autor y fecha	Herramienta de sistemas inteligentes	Desenlace	Tiempo	BD u hospital	n	C (o AUC)
Gambarte, 2021[13]	Red neuronal (perceptrón multicapa con dos capas ocultas)	Mortalidad a 30 días	2005-2019	Hospital Alemán de Buenos Aires, Argentina	483	0.82
Kwon, 2019[14]	Deep neural network (DAHF, con tres capas ocultas)	Mortalidad intrahospitalaria	2016-2017	Korean Acute Heart Failure (KorAHF) registry	Entrenamiento: 6724 Pruebas: 4.759.	0.880
	Ramdom forest					0.756
	Logistic regression					0.720
	Support vector machine					0.723
	Bayesian network					0.730
Panahiazar, 2015[15]	Ramdom forest	Mortalidad a 1 año	1993-2013	Mayo Clinic	Entrenamiento: 1.560 Pruebas: 3.484	0.62 (0.80)
	Support vector regression					0.56 (0.46)
	Decision tree					0.6 (0.66)
	Ada boost					0.59 (0.74)
	Logistic regression					0.68 (0.81)
Zhang, 2013[16]	Chi-square automatic interaction detector (CHAID) decision trees	Muerte y/o hospitalización por empeoramiento de la ICC a 1 año (muerte a 1 año)	ND	Trans-European Network-Home-Care Management System (TEN-HMS) Study	Entrenamiento: 284 Pruebas: 160	0.797 (0.892)
	LR					0.738 (0.858)
Tomcikova, 2013[17]	classification and regression tree analysis (CART)	Muerte intrahospitalaria	2006-2009	Acute Heart Failure Database–Main registry	Entrenamiento: 2.543 Pruebas: 1.387	0.823
						0.832
Austin, 2011[18]	Boosted classification trees (AdaBoost)	Muerte a 30 días	1999-2001	EFFECT study	8.240	Sens: 13% Espec: 98%
Austin, 2010[6]	Cinco modelos de regresion logística	Mortalidad intrahospitalaria	1999-2001	EFFECT study	Entrenamiento: 8.236 Pruebas: 7.608	0.747-0.775
	Tres modelos de árboles de regresión					0.620-0.651
Visweswaran, 2010[19]	Patient specific Markov blanket global structure algorithm (PSMBI-MS)	Muerte a 90 días	1999	ND	Entrenamiento: 7.453 Pruebas: 3.735	
	PSMBg-MA					
	LR					
Abraham, 2008[20]	Classification and regression tree analysis (CART)	Muerte intrahospitalaria	ND	OPTIMIZE HF	37.548	0.683
Fonarow, 2005[4]	Classification and regression tree analysis (CART)	Mortalidad intrahospitalaria	2001 2003	ADHERE	Entrenamiento: 33.046 Pruebas: 32.229	0.687
						0.668

ND: no hay datos disponibles.
El artículo de Visweswaran et al. no describe los resultados con las características operativas sino por el error.

de soporte[14,15]; las redes neuronales fueron aplicadas en dos casos[13,14]. Tres estudios compararon varios modelos de inteligencia artificial[6,14,15], mientras que en 5 casos fueron comparadas con regresiones logísticas[4,6,14,15,19]. Dos trabajos contaron con una población inferior a 500 pacientes[13,16]. El mejor desempeño lo alcanzó DAHF, un modelo de red neuronal profunda (3 capas ocultas con 33 neuronas) desarrollado a partir de un gran registro multicéntrico coreano por el grupo de Kwong[14] en 2019, que alcanzó un AUC de 0.88.

Aunque no pueda ser calificado del todo como un modelo de inteligencia artificial, un trabajo de 2010 presenta un modelo que realiza el ensamble mediante bootstrapping de un conjunto de regresiones logísticas con 16 variables procedente del análisis de 1.372 pacientes; así mismo, demostró buena discriminación de la mortalidad a 30 días (estadístico C de 0.86) y la readmisión (estadístico C de 0.72), que fue mejor que la regla del ADHERE[21].

Los resultados sugieren que el sistema desarrollado tiene un buen comportamiento en la determinación del pronóstico de mortalidad de la IC a 30 días, con un rendimiento comparable al de los sistemas con mejor desempeño encontrados en la revisión bibliográfica. A diferencia de ellos, se incluyó como entrada la medición del péptido natriurético de tipo B (NT-proBNP), que podría ser una de las posibles explicaciones de dicho resultado. En un trabajo realizado en esta misma población se evaluó el desempeño de tres de las reglas de predicción clínica más conocidas, con resultados muy pobres, con AUC de 0.63 (IC 95% 0.53-0.73) para el OPTIMIZE, de 0.57 (IC 95% 0.49-0.65) para el GWTG y de 0.58 (IC 95% 0.47-0.68) para el árbol de decisiones del ADHERE[22]. Los resultados obtenidos mostraron un patrón inverso en cuanto a las sensibilidades y especificidades, siendo en el presente ensayo más alta la especificidad con baja sensibilidad, lo cual podría ser importante en la toma de decisiones al definir la disposición de los pacientes.

En un trabajo anterior, se describió una nueva estrategia de ensamble que se basaba en el desempeño estadístico a través de las características operativas de una prueba, aplicado al diagnóstico del infarto agudo de miocardio[23]. En esa oportunidad se obtuvieron resultados similares a los de este trabajo, con rendimientos superiores de los ensambles al de los componentes individuales, pero no se comparó con otras herramientas populares de ensamble. Esta vez se utilizó boosting como comparador, que es reconocido ampliamente como una de las estrategias más difundidas de ensamble, y se obtuvo un resultado muy similar.

Es preciso aceptar que existen mejores algoritmos de entrenamiento para las redes neuronales que el algoritmo genético; aquí, su uso se explica por la posibilidad de entrenar un número mayor de redes en cada corrida de éste. Además, en nuestro trabajo la investigación se centraba en la comparación de los mecanismos de ensamble y, para poder hacer visible el efecto de la combinación, se requería que los elementos constituyentes tuvieran un desempeño regular.

Como limitaciones del presente trabajo se reconoce que la procedencia de los pacientes haya sido de un único centro, y que, además, puede ser posible que el número de desenlaces (13.8%) fuera pequeño para la elaboración de un modelo predictivo, con lo cual se pudieron dejar subrepresentadas poblaciones de interés, pero precisamente esta es una de las ventajas teóricas de las estrategias de ensamble.

Se convierten en fortalezas el uso de un amplio número de experimentos para la construcción y selección de las redes neuronales (a diferencia de la conducta *a priori* que suele usarse), así como la comparación de diferentes mecanismos de ensamble, que incluyó al AdaBoost.

Como trabajo futuro se deberá buscar la inclusión del modelo en una herramienta para la toma de decisiones aplicable en la práctica diaria de la disposición de los pacientes que consultan por descompensación de la IC, comparándolo con otras escalas o la decisión del médico tratante.

Conclusiones

El ensamble de redes neuronales mediante un sistema de votación ponderada centrado en las características operativas de la prueba mediante los valores predictivos demostró un adecuado rendimiento para el pronóstico de muerte a 30 días en IC descompensada, mejorando el desempeño de cada unas de las redes neuronales que la componían; el desempeño se comportó de forma similar al del sistema de ensamble por AdaBoost.

Financiamiento

Los autores declaran que el presente trabajo no recibió financiación para su conducción, análisis ni escritura.

Conflicto de intereses

Los autores declaran que no tienen ningún conflicto de intereses referido al presente trabajo de investigación.

Responsabilidades éticas

Protección de personas y animales. Los autores declaran que los procedimientos seguidos se conformaron a las normas éticas del comité de experimentación humana responsable y de acuerdo con la Asociación Médica Mundial y la Declaración de Helsinki.

Confidencialidad de los datos. Los autores declaran que han seguido los protocolos de su centro de trabajo sobre la publicación de datos de pacientes.

Derecho a la privacidad y consentimiento informado. Los autores han obtenido la aprobación del Comité de Ética para el análisis y publicación de datos clínicos obtenidos de forma rutinaria. El consentimiento informado de los pacientes no fue requerido por tratarse de un estudio observacional retrospectivo.

Uso de inteligencia artificial para generar textos. Los autores declaran que no han utilizado ningún tipo de inteligencia artificial generativa en la redacción de este manuscrito ni para la creación de figuras, gráficos, tablas o sus correspondientes pies o leyendas.

Bibliografía

1. Liu L, Eisen HJ. Epidemiology of heart failure and scope of the problem. Cardiol Clin. 2014;32(1):1-8. DOI: 10.1016/j.ccl.2013.09.009.
2. Rahimi K, Bennett D, Conrad N, Williams TM, Basu J, Dwight J, et al. Risk prediction in patients with heart failure: a systematic review and analysis. JACC Heart Fail. 2014;2(5):440-6. DOI: 10.1016/j.jchf.2014.04.008.
3. James CA, Wachter RM, Woolliscroft JO. Preparing clinicians for a clinical world influenced by artificial intelligence. JAMA. 2022;327(14):1333-4. DOI:10.1001/jama.2022.3580.
4. Fonarow GC, Adams KF Jr, 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(5):572-80. DOI: 10.1001/jama.293.5.572.
5. Auble TE, Hsieh M, McCausland JB, Yealy DM. Comparison of four clinical prediction rules for estimating risk in heart failure. Ann Emerg Med. 2007;50(2): 127-35. e1–2. DOI: 10.1016/j.annemergmed.2007.02.017.
6. Austin PC, Tu JV, Lee DS. Logistic regression had superior performance compared with regression trees for predicting in-hospital mortality in patients hospitalized with heart failure. J Clin Epidemiol. 2010;63(10):1145-55. DOI: 10.1016/j.jclinepi.2009.12.004.
7. Scrutinio D, Ammirati E, Passantino A, Guida P, D'Angelo L, Oliva F, et al. Predicting short-term mortality in advanced decompensated heart failure - role of the updated acute decompensated heart failure/N-terminal pro-B-type natriuretic Peptide risk score. Circ J. 2015;79(5):1076-83. DOI: 10.1253/circj.CJ-14-1219.
8. Chaves WG, Diaztagle JJ, Sprockel JJ, Hernandez JI, Benavidez J, Henao D, et al. Factores asociados a mortalidad en pacientes con falla cardiaca descompensada. Acta Med Colomb. 2014; 39(4):314–20.
9. 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.
10. 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;52:347-56. DOI: 10.1016/j.jacc.2008.04.028.
11. Negnevitsky M. Artificial Intelligence: A Guide to Intelligent Systems. 2nd. ed. Addison Wesley; 2004.
12. Polikar R. Ensemble based systems in decision making. Circuits and Systems Magazine, IEEE. 2006;6(3):21-45. DOI: 10.1109/MCAS.2006.1688199.
13. Gambarte MJ, Higa C, Novo F, Ciambrone GM, Tupayachi Villagomez OD, Ginesi A, et al. 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. Rev Argent Cardiol. 2021;89(5):435-46. DOI: 10.7775/rac.es.v89.i5.20434.
14. Panahiazar M, Taslimitehrani V, Pereira N, Pathak J. Using EHRs and machine learning for heart failure survival analysis. Stud Health Technol Inform. 2015;216:40-4.
15. 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(7):e0219302. DOI: 10.1371/journal.pone.0219302.
16. Zhang J, Goode KM, Rigby A, Balk AHMM, Cleland JG. Identifying patients at risk of death or hospitalisation due to worsening heart failure using decision tree analysis: evidence from the Trans-European Network-Home-Care Management System (TEN-HMS) study. Int J Cardiol. 2013;163(2):149-56. DOI: 10.1016/j.ijcard.2011.06.009.
17. Tomcikova D, Felsoci M, Spinar J, Miklik R, Mikusova T, Vitovec J, et al. Risk of in-hospital mortality identified according to the typology of patients with acute heart failure: Classification tree analysis on data from the Acute Heart Failure Database–Main registry. J Crit Care. 2013;28:250-8. DOI: 10.1016/j.jcrc.2012.09.014.
18. Austin P, Lee D. Boosted classification trees result in minor to modest improvement in the accuracy in classifying cardiovascular outcomes compared to conventional classification trees. Am J Cardiovasc Dis. 2011;1(1):1-15.
19. Visweswaran S, Angus DC, Hsieh M, Weissfeld L, Yealy D, Cooper GF. Learning patient-specific predictive models from clinical data. J Biomed Inform. 2010;43(5): 669-85. DOI: 10.1016/j.jbi.2010.04.009.
20. 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(5):347-56. DOI: 10.1016/j.jacc.2008.04.028.
21. Amarasingham R, Moore BJ, Tabak YP, Drazner MH, Clark CA, Zhang S, et al. An automated model to identify heart failure patients at risk for 30-day readmission or death using electronic medical record data. Med Care. 2010;48(11):981-8. DOI: 10.1097/MLR.0b013e3181ef60d9.
22. Sprockel J, Alfaro L, Cifuentes J, Jimenez M, Baron R, Chaves Santiago WG. Prognostic scores for risk stratification in patients with acute heart failure. Rev Argent Cardiol. 2016;84(6):574-80. DOI: 10.7775/rac.es.v84.i6.9823.
23. Sprockel JJ, Gonzalez E. Assembly of neural networks within a federation of rational agents for the diagnosis of acute coronary syndromes. In: Bramer M, Petridis M. (eds.). Research and Development in Intelligent Systems XXXII [Internet]. Springer International Publishing; 2015. p. 289-300. http://dx.doi.org/10.1007/978-3-319-25032-8_22.