

Um Novo Escore de Risco Baseado em Aprendizado de Máquina (Machine Learning) em Pacientes com Insuficiência Cardíaca Aguda: O Escore ML-HF

A New Risk Score Based on Machine Learning in Patients with Acute Heart Failure: The ML-HF score

Matheus Bissa Duarte Ferreira,¹ Jorge Tadashi Daikubara Neto,¹ Gustavo S. Pereira da Cunha,¹ Rafael Moretti,¹ Jessica Tamires Reichert,¹ Lucas Müller Prado,¹ Raphael Henrique Déa Cirino,¹ Sidney C. Smith Jr,² Fábio Papa Taniguchi,³ Andrei C. Sposito,⁴ Odilson M. Silvestre,⁵ Wilson Nadruz,⁵ Miguel Morita Fernandes-Silva¹

Universidade Federal do Paraná,¹ Curitiba, PR – Brasil

University of North Carolina At Chapel Hill, Chapel Hill,² North Carolina – EUA

Hospital do Coração (HCor),³ São Paulo, SP – Brasil

Universidade Estadual de Campinas,⁴ Campinas, SP – Brasil

Universidade Federal do Acre,⁵ Rio Branco, AC – Brasil

Resumo

Fundamento: Os escores de avaliação prognóstica convencionais muitas vezes não têm desempenho suficiente para prever a mortalidade em pacientes com insuficiência cardíaca aguda (ICA).

Objetivo: Desenvolver e validar um escore prognóstico baseado em aprendizado de máquina (*Machine Learning* - escore ML-HF) para prever morte hospitalar em pacientes com ICA e comparar seu desempenho com os principais escores tradicionais.

Métodos: Pacientes admitidos por ICA em hospitais brasileiros do “Programa de Boas Práticas em Cardiologia” de 2016 a 2022 foram incluídos. Dados clínicos, resultados laboratoriais e o questionário de Qualidade de Vida da Organização Mundial da Saúde (*World Health Organization Quality of Life*) WHOQOL-Bref foram coletados na admissão hospitalar. O desfecho foi óbito hospitalar. O modelo foi treinado usando 70% das admissões (conjunto de treinamento) e validado com os 30% restantes (conjunto de teste). A área sob a curva ROC (AUC) do escore ML-HF foi comparada com a AUC dos escores tradicionais *Acute Decompensated Heart Failure National Registry* (ADHERE) e *Get With the Guidelines–Heart Failure* (GWTG-HF). O nível de significância foi de $p < 0,05$.

Resultados: Foram incluídos mil cento e cinquenta e sete pacientes hospitalizados por ICA. As cinco variáveis mais importantes do escore ML-HF foram: Qualidade do Domínio de Saúde Física (WHOQOL-BREF), sódio sérico, ureia sérica, creatinina sérica e pressão arterial sistólica na admissão hospitalar. No conjunto de teste, o escore ML-HF apresentou calibração de modelo adequada (valor de p do teste de Hosmer-Lemeshow = 0,056) e discriminação (AUC = 0,722 [IC95%, 0,661-0,783]), que foi superior aos escores GWTG-HF (AUC = 0,616 [IC95%, 0,529-0,702; $p = 0,014$]) e ADHERE (AUC = 0,601 [IC95%, 0,511-0,691; $p = 0,006$]).

Conclusão: Desenvolvemos e validamos um escore usando aprendizado de máquina para prever morte hospitalar em pacientes com ICA, que superou os escores tradicionais.

Palavras-chave: Insuficiência Cardíaca; Aprendizado de Máquina; Morte.

Abstract

Background: Conventional prognostic assessment scores often fall short in performance to predict mortality in patients with acute heart failure (AHF).

Objective: To develop and validate a machine learning-based prognostic score to predict in-hospital death in patients with AHF and compare its performance with the main traditional scores.

Methods: Patients admitted for AHF in Brazilian hospitals from the “Best Practices in Cardiology Program” from 2016 - 2022 were included. Clinical data, laboratory results, and the World Health Organization Quality of Life (WHOQOL-Bref) at hospital admission were collected. The outcome was in-hospital death. The model was trained using 70% of admissions (training set) and validated with the remaining 30% (test set). The ML-HF score area under the ROC curve (AUC) was compared with the AUC of the traditional scores *Acute Decompensated Heart Failure National Registry* (ADHERE) and *Get With the Guidelines–Heart Failure* (GWTG-HF). The level of significance was $p < 0.05$.

Correspondência: Miguel Morita Fernandes-Silva •

Universidade Federal do Paraná – Rua General Carneiro, 181, 10º andar. CEP 80060-900, Curitiba, PR – Brasil

E-mail: miguelmorita@ufpr.br

Artigo recebido em 24/02/2025, revisado em 18/06/2025, aceito em 28/07/2025

Editor responsável pela revisão: Gláucia Maria Moraes de Oliveira

DOI: <https://doi.org/10.36660/abc.20250136>

Results: One thousand and one hundred fifty-seven patients hospitalized for AHF were included. The five most important variables of the ML-HF score were: Physical Health Domain Quality (WHOQOL-BREF), serum sodium, serum urea, serum creatinine, and systolic blood pressure at hospital admission. In the test set, the ML-HF score showed an adequate model calibration (Hosmer-Lemeshow test p value=0.056) and discrimination (AUC=0.722 [CI95%, 0.661-0.783]), which was superior to GWTG-HF (AUC=0.616 [CI95%, 0.529-0.702; $p=0.014$]) and the ADHERE (AUC=0.601 [CI95%, 0.511-0.691; $p=0.006$]) scores.

Conclusion: We developed and validated a score using machine learning to predict in-hospital death in patients with AHF, which outperformed the traditional scores.

Keywords: Heart Failure; Machine Learning; Death.

Full texts in English - <https://abccardiol.org/en/>

Figura Central: Um Novo Escore de Risco Baseado em Aprendizado de Máquina (Machine Learning) em Pacientes com Insuficiência Cardíaca Aguda: O Escore ML-HF



UM NOVO ESCORE DE RISCO BASEADO EM APRENDIZADO DE MÁQUINA EM PACIENTES COM INSUFICIÊNCIA CARDÍACA AGUDA: O ESCORE ML-HF

POPULAÇÃO

- Pacientes com ICA
- Dezenove hospitais no Brasil
- Março de 2016 - Dezembro de 2022

Desenvolvimento de modelo de aprendizado de máquina para prever óbito hospitalar, incluindo:

- Dados sociodemográficos, clínicos, ecocardiográficos e laboratoriais
- Questionário WHOQOL-Brief

Algoritmo de aprendizado de máquina utilizado:

- Random Forest

RESULTADOS

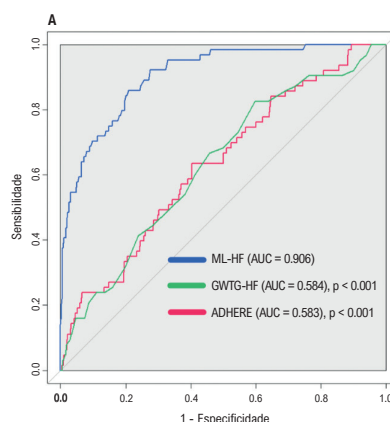
- 1.157 pacientes incluídos
- 58% homens
- $61,4 \pm 14,5$ anos
- LVEF $42\% \pm 17$
- 117 (10,1%) óbitos hospitalares

17 variáveis selecionadas para construir o modelo ML-HF. As mais importantes:

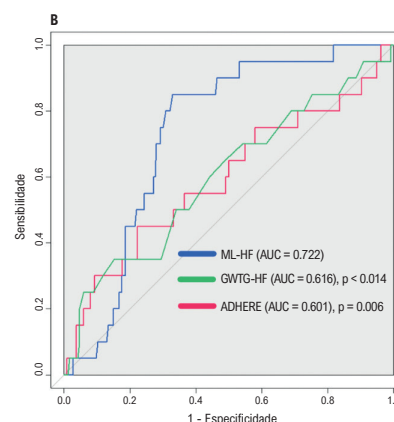


Desempenho do modelo

Conjunto de treinamento



Conjunto de teste



Arq Bras Cardiol. 2025; 122(11):e20250136

Introdução

A insuficiência cardíaca (IC) afeta 25 milhões de pessoas em todo o mundo e apresenta altas taxas de mortalidade e hospitalização. No Brasil, ocorrem aproximadamente 250.000 internações hospitalares por IC por ano, resultando em uma taxa de mortalidade hospitalar de 10,9%.¹

As ferramentas de classificação de risco ajudam a direcionar recursos e medidas intensivas para pacientes de alto risco. No entanto, estimar o prognóstico dos pacientes na admissão ou durante a hospitalização tem sido desafiador no cenário da IC. Estudos anteriores desenvolveram escores, como o *Acute Decompensated Heart Failure National Registry* (ADHERE) e o *Get With The Guidelines-Heart Failure* (GWTG-HF), que podem ser derivados de umas poucas variáveis laboratoriais e clínicas, para prever a mortalidade hospitalar nesses pacientes.^{2,3} Infelizmente, esses escores têm precisão abaixo do ideal e têm sido subutilizados na prática clínica.^{4,5}

Uma alternativa a essas abordagens convencionais são os modelos baseados em aprendizado de máquina (*Machine Learning* - ML), que têm sido estudados nos últimos anos para o manejo de IC.^{6,7} A capacidade dos algoritmos de ML de analisar de forma mais profunda, multidimensional e não linear uma variedade de variáveis aumenta seu potencial para melhorar os modelos de estratificação de pacientes.⁷ Modelos baseados em ML foram construídos em pacientes com IC crônica e aguda usando registros médicos eletrônicos (RME), registros e ensaios clínicos.⁸ Esses modelos parecem ter precisão superior à dos escores tradicionais ADHERE e GWTG-HF.⁹ No entanto, a maioria desses estudos foi desenvolvida e validada em países de alta renda. Diferenças na população de pacientes, determinantes sociais e prestação de cuidados de saúde podem impedir a implementação desses modelos em países de baixa e média renda (PBMR). De fato, já foi demonstrado que a

capacidade dos modelos de ML de prever desfechos de IC varia significativamente entre diferentes subgrupos étnicos e grupos socioeconômicos.¹⁰

Portanto, desenvolvemos e validamos um algoritmo baseado em ML para prever mortes hospitalares em pacientes com IC aguda, utilizando um amplo estudo de coorte em hospitais públicos no Brasil. Comparamos ainda o algoritmo final baseado em ML com os modelos ADHERE e GWGT.

Métodos

População

Trata-se de uma análise do programa “Boas Práticas em Cardiologia (BPC)”, adaptado da iniciativa “Get With The Guidelines” (GWGT) para melhorar a qualidade do atendimento cardiovascular e os desfechos dos pacientes com síndrome coronariana aguda, IC e fibrilação atrial em hospitais públicos brasileiros.¹¹ Resumidamente, foram selecionados hospitais públicos terciários em todas as cinco macrorregiões brasileiras desde janeiro de 2016.¹² Esses hospitais foram incluídos em um estudo quase experimental que implementou uma intervenção em nível institucional para melhorar a adesão às diretrizes e que avaliou seu impacto nas medidas de desempenho institucional. Dados sociodemográficos, clínicos, laboratoriais e ecocardiográficos foram coletados durante a admissão dos pacientes. Os pacientes também foram solicitados a preencher o questionário de Qualidade de Vida da Organização Mundial da Saúde (WHOQOL-Bref), que é um instrumento de 26 itens que consiste em quatro domínios: saúde física (7 itens), saúde psicológica (6 itens), relações sociais (3 itens) e saúde ambiental (8 itens). Os itens do WHOQOL-Bref consistem em escalas Likert de 5 pontos que, combinadas, fornecem cada domínio de saúde. O procedimento para calcular o WHOQOL-Bref foi descrito em detalhe em outro lugar.¹³ Como parte do programa BPC, todos os participantes foram acompanhados com visitas na alta, um mês e seis meses. Este estudo foi aprovado pelo comitê de ética do Centro Coordenador (número 48561715.5.1001.0060) e de cada hospital participante. Todos os participantes assinaram um termo de consentimento livre e esclarecido para participação no estudo.

Critérios de inclusão e exclusão

Para esta análise, foram incluídos pacientes com 18 anos ou mais internados com diagnóstico primário de IC aguda (códigos CID-10 I50; I50.0; I50.1 ou I50.9) entre 14/03/2016 e 07/12/2022 em qualquer hospital participante do programa BPC, que concordaram em participar do estudo e assinaram o termo de consentimento livre e esclarecido conforme ilustrado na Figura Central. Os participantes precisavam ter ficado hospitalizados por pelo menos 24 horas para serem elegíveis.

Excluímos pacientes transferidos para outro hospital ou com estado vital ausente na alta. Pacientes que estavam em lista de espera para transplante cardíaco ou em uso de dispositivo de assistência ventricular (DAV) também foram excluídos. Também excluímos pacientes com dados ausentes para as variáveis selecionadas do modelo (veja a seleção de variáveis abaixo).

Para construir e validar o modelo, dividimos o conjunto de dados em duas coortes: Coorte 1, pacientes internados de 16/01/2016 a 15/10/2021, e coorte 2, pacientes internados de 16/10/2021 a 07/12/2022. A coorte 1 foi utilizada nos processos de seleção de variáveis, construção do modelo e validação. A coorte 2 foi utilizada apenas para validação externa (veja Análise estatística para obter detalhes).

Desfecho

O desfecho foi morte por qualquer causa durante a hospitalização índice.

Seleção de variáveis

Utilizamos todas as variáveis na admissão no modelo preditivo, que incluiu dados sociodemográficos, clínicos, laboratoriais e ecocardiográficos, juntamente com os escores derivados das respostas do questionário WHOQOL-Bref. Primeiramente, excluímos variáveis com mais de 40% de dados ausentes.¹⁴ Veja a tabela suplementar S1 para todas as variáveis e respectivas taxas de dados ausentes. Em seguida, aplicamos o método de ML “Boruta” no conjunto de treinamento (veja abaixo) para detectar automaticamente as variáveis mais relevantes que foram selecionadas para o algoritmo de treinamento. O método Boruta gera uma distribuição de probabilidade binomial do número de vezes que uma variável excede a importância de uma variável fictícia criada, que tem um valor aleatório (não preditivo).¹⁵ As variáveis foram selecionadas se tivessem um valor de $p < 0,05$ na cauda superior da distribuição de probabilidade. Durante esse processo de seleção, aplicamos um método de imputação K-vizinhos mais próximos, selecionando os K vizinhos mais próximos ($K = 5$) com dados completos com base em uma distância de Gower e imputando o valor médio da variável completa aos vizinhos com valores ausentes.¹⁶

Após a seleção das variáveis mais relevantes, elas foram utilizadas para construir o algoritmo de ML no banco de dados não imputado. A Figura S1 ilustra o processo de seleção de variáveis (Figura Suplementar S1). Realizamos uma Análise de Casos Completos, incluindo apenas pacientes com dados completos para as variáveis selecionadas. Assim, pacientes com dados ausentes para qualquer uma das variáveis selecionadas foram excluídos (Figura 1).

Análise estatística

Estatísticas descritivas

As variáveis contínuas foram avaliadas quanto à normalidade por meio da inspeção visual dos histogramas, assimetria e curtose, e pela aplicação do teste de Kolmogorov-Smirnov. Os dados são apresentados como média $\pm$ desvio-padrão ou mediana (percentil 25 - percentil 75), conforme apropriado. As comparações entre os grupos foram realizadas utilizando o teste-t de Student não pareado para variáveis com distribuição normal e o teste U de Mann-Whitney para variáveis com distribuição anormal. As variáveis categóricas foram expressas como contagens e porcentagens e comparadas utilizando o teste exato de Fisher.

Desenvolvimento e validação de modelos

A coorte 1 foi dividida aleatoriamente em conjuntos de treinamento (70%) e teste (30%). Este procedimento foi realizado por meio de amostragem aleatória gerada por computador, sem reposição, garantindo igual probabilidade de seleção de cada paciente. O conjunto de treinamento foi usado para treinamento do modelo e validação interna. O conjunto de teste combinou 30% da coorte 1 com a coorte 2 completa e foi usado para validação externa. Uma análise de sensibilidade foi realizada usando apenas a coorte 2 como conjunto de teste. Este processo é ilustrado na Figura 2.

Seleção de modelos

Utilizamos um método de validação cruzada de 5 etapas para selecionar o algoritmo de ML mais adequado. Este método consiste em subdividir o conjunto de treinamento em 5 grupos aleatórios. Quatro desses subconjuntos foram usados para treinar o modelo e o subconjunto restante para testá-lo. Trata-se de um processo iterativo que ocorre 5 vezes, alterando o subconjunto de teste, conforme ilustrado na Figura 2. O algoritmo com a melhor AUC média foi escolhido como modelo final.

Desempenho de modelo

Para avaliar a discriminação do modelo, construímos as curvas ROC (*Receiver Operating Characteristic*) e comparamos a área sob a curva ROC (AUC) usando a morte hospitalar como desfecho. A calibração do modelo foi avaliada plotando a probabilidade prevista versus a observada de morte hospitalar em decis e calculando o teste de qualidade de ajuste de Hosmer-Lemeshow. Um valor de $p < 0,05$ indicou ajuste inadequado dos dados, enquanto valores de p mais altos indicaram calibração adequada.

Comparação com escores tradicionais

Os escores ADHERE e GWTG-HF foram calculados para cada paciente. Como o escore GWTG-HF não fornece diretamente as probabilidades de morte, derivamos uma regressão logística de seus resultados numéricos.^{2,3} Comparamos a AUC-ROC entre os escores usando o teste de DeLong nos conjuntos de treinamento e teste.¹⁷

Seguimos a diretriz de Relatório Transparente de um Modelo de Predição Multivariável para Prognóstico ou Diagnóstico Individual (TRIPOD) para conduzir nossa pesquisa e escrever este artigo.¹⁸

Todas as análises e treinamento do modelo foram conduzidos usando o software R. Consideramos estatisticamente significativos valores de $p < 0,05$.

Resultados

De janeiro de 2016 a dezembro de 2022, um total de 3.532 pacientes com IC aguda foram incluídos no programa BPC. Após a exclusão daqueles com DAV, na lista de espera para transplante ou sem estado vital na alta, 3.043 foram

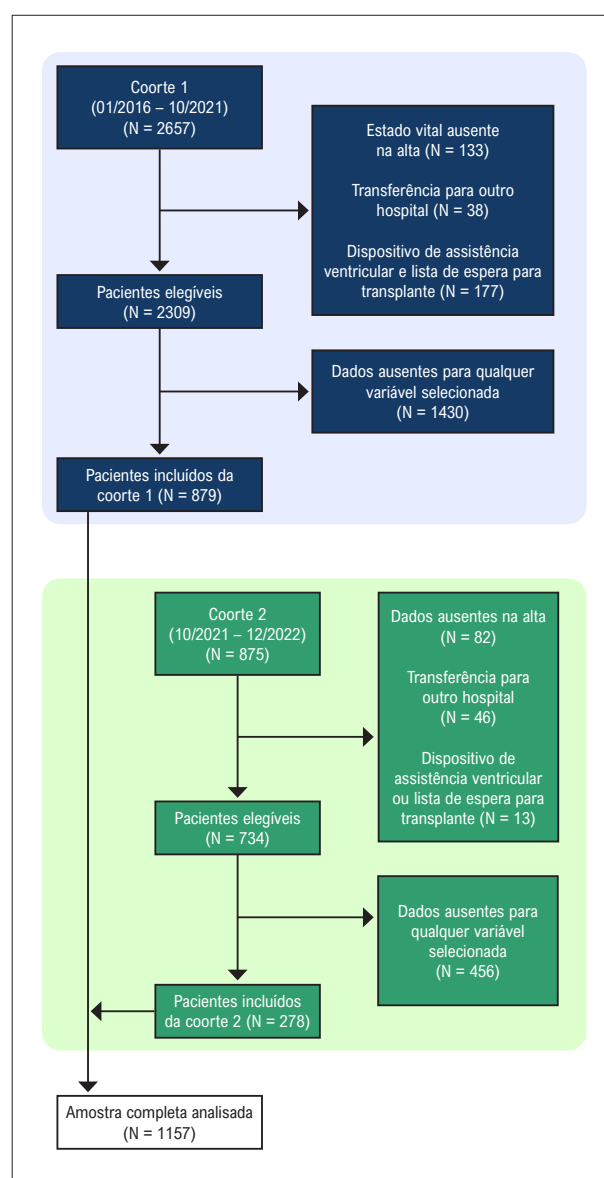


Figura 1 – Fluxograma de pacientes incluídos na análise.

considerados elegíveis para o estudo (2.309 na coorte 1 e 734 na coorte 2, como visto na Figura 1).

Das 113 variáveis disponíveis na admissão, nove foram excluídas devido à ausência de mais de 40% de dados (Tabela Suplementar S1). O algoritmo de Boruta foi aplicado a 70% dos pacientes ($n = 1.616$) selecionados aleatoriamente da coorte 1, resultando em 17 variáveis selecionadas para construir o modelo de ML.

Dos 3.043 pacientes elegíveis, 1.157 foram incluídos na análise final devido à disponibilidade completa de dados para todas as 17 variáveis. O conjunto de treinamento consistiu em 621 pacientes, o que representou 70% dos 879 indivíduos da coorte 1, enquanto o conjunto de teste consistiu em 536 pacientes, compreendendo os 258 pacientes restantes da coorte 1 e 278 pacientes mais da coorte 2.

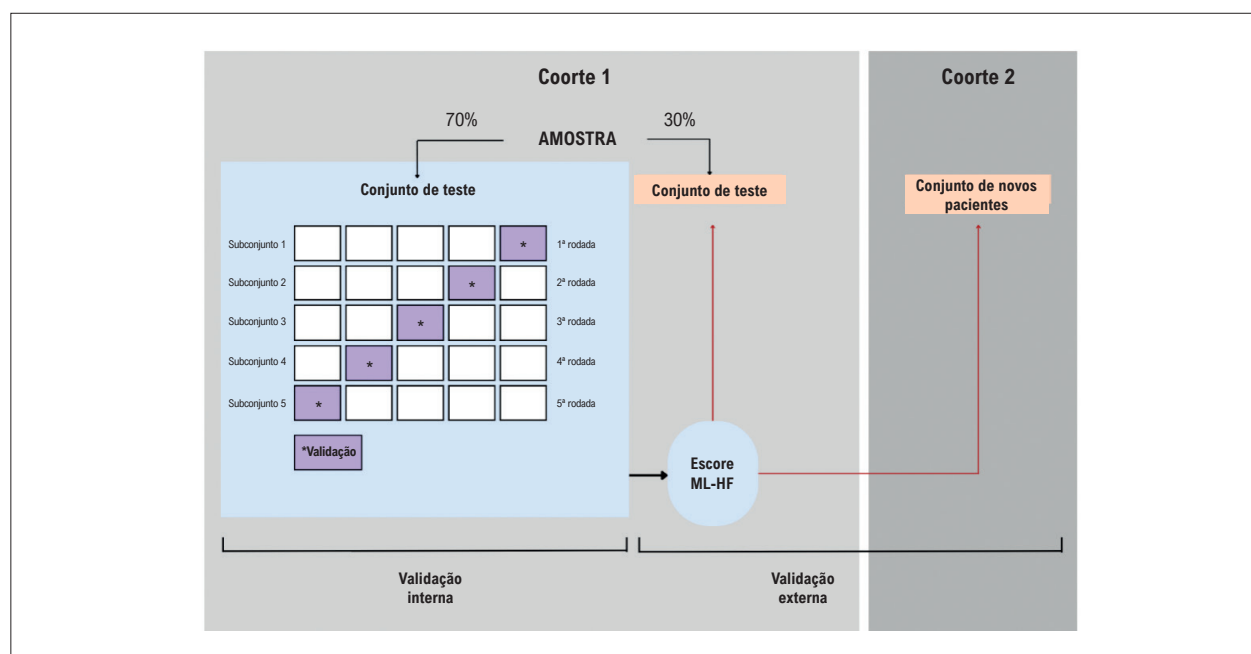


Figura 2 – Fluxograma ilustrando a validação interna e externa do modelo de escore ML-HF.

A taxa de mortalidade hospitalar foi de 10,1%, com um total de 117 óbitos entre os 1.157 pacientes examinados. A idade média desses pacientes foi de 61 ± 15 anos e a fração de ejeção do ventrículo esquerdo (FEVE) foi de $42 \pm 17\%$ e 58% eram do sexo masculino (Tabela 1). A Tabela 2 exibe as variáveis selecionadas no escore ML-HF e aquelas usadas nos escores ADHERE e GWTG para comparação.

Seleção de algoritmo

Dentre os algoritmos de ML avaliados, o modelo Random Forest apresentou o melhor desempenho na predição de mortalidade hospitalar em comparação com Gradient Boosting Machines e Extreme Gradient Boosting (Tabela 3). Dessa forma, o algoritmo Random Forest foi escolhido para construir o modelo final.

Validação interna

No conjunto de treinamento ($n=621$), o escore ML-HF demonstrou precisão de boa a excelente para prever mortalidade hospitalar, com uma AUC de 0,906, superando os escores ADHERE e GWTG-HF. [Figura 3A] [Figura Central].

Validação externa

No conjunto de teste ($n=536$), o escore ML-HF manteve uma discriminação aceitável, com uma AUC de 0,722. Embora inferior ao conjunto de treinamento, seu desempenho permaneceu superior aos escores ADHERE e GWTG. [Figura 3B, Figura Central]. Resultados semelhantes foram observados ao utilizar apenas a coorte 2 completa como conjunto de teste ($n=280$), conforme mostrado na Figura S2 Suplementar.

Calibração do modelo

A Figura 4 mostra a mortalidade hospitalar observada versus prevista para cada escore no conjunto de teste. Embora o escore ML-HF tenha apresentado boa calibração do modelo, a mortalidade prevista divergiu significativamente da mortalidade observada nos outros dois escores convencionais. Especificamente, o escore ADHERE subestimou o risco, enquanto o escore GWTG o superestimou.

Importância da variável no escore ML-HF

A Figura 5 mostra a importância relativa das variáveis do escore ML-HF em ordem decrescente. A variável com maior importância foi o domínio Saúde Física do WHOQOL, seguido pelos níveis sanguíneos de sódio, ureia e creatinina, conforme mostrado na Figura Central. A pressão arterial na admissão também estava entre as dez variáveis mais importantes, com a pressão arterial sistólica e a pressão arterial diastólica ocupando a quinta e a oitava posições, respectivamente. Notavelmente, a FEVE não teve importância suficiente para prever a mortalidade hospitalar no método de Boruta. O único parâmetro ecocardiográfico no modelo foi o tamanho do átrio esquerdo, ocupando o sexto lugar em ordem de importância.

Discussão

Este estudo apresenta três principais achados. Desenvolvemos um novo escore baseado em ML – o escore ML-HF – para prever a mortalidade hospitalar em pacientes com ICA em uma PBMR, que foi validado externamente com acurácia aceitável e boa calibração. O escore ML-HF superou os escores convencionais ADHERE e GWTG-HF na mesma população.²⁻⁵ O estudo revelou que as informações coletadas

Tabela 1 – Características basais dos pacientes incluídos

Parâmetros	Alta com vida (n=1040)	Morte hospitalar (n=117)	Valor-p
Idade, anos	61,3 ± 14,5	61,6 ± 14,4	0,829
Sexo feminino, n(%)	424 (40,8)	53 (45,3)	0,398
Raça branca, n(%)	674 (64,8)	78 (66,7)	0,766
Educação, n(%)			
Analfabeto	84 (8,1)	9 (7,7)	0,242
Ensino fundamental incompleto	399 (38,4)	47 (40,2)	
Ensino fundamental completo	148 (14,2)	26 (22,2)	
Ensino médio incompleto	72 (6,9)	8 (6,8)	
Ensino médio completo	225 (21,6)	20 (17,1)	
Ensino superior incompleto	30 (2,9)	2 (1,7)	
Ensino superior completo	82 (7,9)	5 (4,3)	
Renda média, n(%)			
≤ 1 salário-mínimo	721 (69,3)	97 (82,9)	0,009
Fumante, n(%)			
Não	520 (50)	56 (47,9)	0,889
Atual	438 (42,1)	52 (44,4)	
Passado	82 (7,9)	9 (7,7)	
Alcoólatra, n(%)			
Não	654 (62,9)	80 (68,4)	0,200
Atual	246 (23,7)	28 (23,9)	
Passado	140 (13,5)	9 (7,7)	
Etiologia da IC, n(%)			
Isquêmica	200 (19,2)	22 (18,8)	< 0,001
Doença de Chagas	95 (9,1)	13 (11,1)	
Doença valvar	123 (11,8)	30 (25,6)	
Outro	622 (59,8)	52 (44,4)	
Prótese valvar, n(%)	338 (32,5)	52 (44,4)	0,013
Insuficiência renal crônica, n(%)	152 (14,6)	27 (23,1)	0,024
Hipotireoidismo, n(%)	77 (7,4)	19 (16,2)	0,002
FEVE, %	41,7 ± 17,1	41,90 ± 18,5	0,911
Diâmetro do átrio esquerdo, mm	45,8 ± 8,5	48,1 ± 9,4	0,005
Número de comorbidades [†]	4 (3 - 5)	4 (2 - 5)	0,665
Número de procedimentos cirúrgicos anteriores, n(%)[‡]			0,002
0	702 (67,5)	65 (55,6)	
1	288 (27,7)	36 (30,8)	
2	46 (4,4)	14 (12,0)	
3 ou mais	4 (0,3)	2 (1,7)	

Características clínicas na admissão

Perfil hemodinâmico, n(%)			0,020
---------------------------	--	--	-------

A	78 (7,5)	8 (6,8)	
B	797 (76,6)	77 (65,8)	
C	130 (12,5)	25 (21,4)	
L	35 (3,4)	7 (6,0)	
Pressão arterial sistólica, mmHg	123 ± 29	115 ± 27	0,004
Pressão arterial diastólica, mmHg	76 ± 17	71 ± 16	0,001
Baixa perfusão periférica, n(%)	127 (12,2)	20 (17,1)	0,175
Ortopneia, n(%)	757 (74,6)	78 (70,9)	0,471
Ascite, n(%)	242 (24,6)	39 (35,5)	0,019
B3, n(%)	103 (10,3)	13 (11,5)	0,806
Edema periférico, n(%)	779 (75,6)	99 (85,3)	0,025

Resultados laboratoriais na admissão (níveis sanguíneos)

Creatinina, mg/dL	1,71 ± 1,51	1,93 ± 1,30	0,136
BUN, mg/dL	33,07 ± 22,8	46,0 ± 30,4	< 0,001
Sódio, mmol/L	136,6 ± 5,0	134,2 ± 6,9	< 0,001
Potássio, mmol/L	4,46 ± 0,77	4,66 ± 0,92	0,010
Hemoglobina, g/dL	12,6 ± 2,28	12,32 ± 2,4	0,207

Questionário WHOQOL-BREF

Domínio da saúde física	50,4 ± 19,4	39,5 ± 15,8	< 0,001
Domínio de relações sociais	64,2 ± 16,5	61,1 ± 17,3	0,059
Saúde ambiental	58,9 ± 12,6	54,5 ± 12,9	< 0,001
Saúde psicológica	77,3 ± 17,9	68,8 ± 18,0	< 0,001

Variáveis categóricas são apresentadas como n (%) e variáveis contínuas como média ± desvio padrão ou mediana (percentil 25 - percentil 75). IC: insuficiência cardíaca; FEVE: fração de ejeção do ventrículo esquerdo; WHOQOL-BREF: Versão resumida do questionário de qualidade de vida da Organização Mundial da Saúde; BUN: Nitrogênio ureico no sangue. † As comorbidades incluíram hipertensão, diabetes, dislipidemia, acidente vascular cerebral, doença carotídea, doença arterial coronariana, doença vascular periférica, infarto agudo do miocárdio, fibrilação atrial, doença de Chagas, doença cardíaca reumática, doença valvar, doença pulmonar obstrutiva crônica, apneia do sono, doença hepática, insuficiência renal crônica, anemia, hipertireoidismo, hipotireoidismo, depressão, câncer de órgão sólido e câncer hematológico. ‡ Procedimentos cirúrgicos anteriores incluíram angioplastia coronária, cirurgia de revascularização do miocárdio, marcapasso ou desfibrilador cardíaco implantável, terapia de ressincronização cardíaca e prótese valvar.

a partir de um questionário de qualidade de vida emergiram como o preditor mais importante de mortalidade. No geral, este estudo fornece uma nova ferramenta para prever o risco em pacientes com IC aguda e destaca o potencial do ML para aprimorar os modelos de estratificação de risco nessa população.

Até onde sabemos, este é o primeiro estudo a incorporar uma medida de qualidade de vida pessoal em um modelo baseado em ML para prever mortalidade na IC. Surpreendentemente, o domínio saúde física do questionário

WHOQOL-BREF emergiu como o preditor mais importante de mortalidade, superando biomarcadores séricos e parâmetros ecocardiográficos. Esse achado novo e inesperado provavelmente só foi possível graças ao uso de técnicas de ML, pois elas permitem a análise de variáveis de maneira não linear e predefinida. Este é um exemplo de como o ML pode analisar padrões complexos em dados para encontrar preditores contraintuitivos da perspectiva humana. Este resultado ressalta a importância de incorporar os resultados relatados pelo paciente (RRP) e considerar a autopercepção

Tabela 2 – Comparação das variáveis utilizadas nos escores ADHERE, GWTG e ML-HF

VARIÁVEL	ADHERE	GWTG	ML-HF
Demográfico			
Idade, anos	✓	✓	
Raça negra		✓	
Características clínicas			
BUN, mg/dL	✓	✓	✓
Creatinina, mg/dL	✓		✓
Diâmetro do átrio esquerdo			✓
DPOC		✓	
Doença Renal Crônica			✓
Etiologia valvar da IC			✓
Frequência cardíaca		✓	
Frequência respiratória		✓	
Hemoglobina, g/dL			✓
Hipotireoidismo			✓
Número de procedimentos cardíacos anteriores			✓
Pressão arterial sistólica, mmHg	✓	✓	✓
Pressão Arterial Diastólica, mmHg			✓
Perfil hemodinâmico B			✓
Perfil hemodinâmico C			✓
Perfil hemodinâmico L			✓
Má perfusão periférica			✓
Prótese valvar			✓
Sódio, mmol/L		✓	✓
Qualidade de vida			
Domínio da saúde física – WHOQOL-BREF			✓

BUN: Nitrogênio ureico sanguíneo; DPOC: doença pulmonar obstrutiva crônica; WHOQOL-BREF: Versão resumida do questionário de qualidade de vida da Organização Mundial da Saúde.

Tabela 3 – Desempenho de discriminação no conjunto de treinamento de acordo com o algoritmo de Aprendizado de Máquina

Algoritmo de Aprendizado de Máquina	AUC (Média ± DP) N = 810
Random Forest	0,772 ± 0,07
Gradient Boosting Machines	0,703 ± 0,08
Extreme Gradient Boosting	0,724 ± 0,09

DP: desvio padrão; AUC: área sob a curva de operação de recepção.

do paciente sobre sua saúde. Em termos mais simples, destaca a importância de realmente ouvir o paciente.

Estudos anteriores demonstraram o valor prognóstico do RRP em pacientes com IC crônica. No registro CHAMP-HF (*Change the Management of Patients With Heart Failure*), uma melhora de 5 pontos no Questionário de Cardiomiopatia de Kansas City foi associada a uma redução de 41% no risco de morte por todas as causas.¹⁹ Notavelmente, o RRP provou ser um melhor preditor de eventos futuros em comparação com a classe funcional da NYHA, ressaltando a importância de avaliar a perspectiva do paciente sobre seu estado de saúde, além da avaliação clínica.^{19,20} De fato, as diretrizes atuais de IC recomendam a avaliação padronizada do RRP como uma

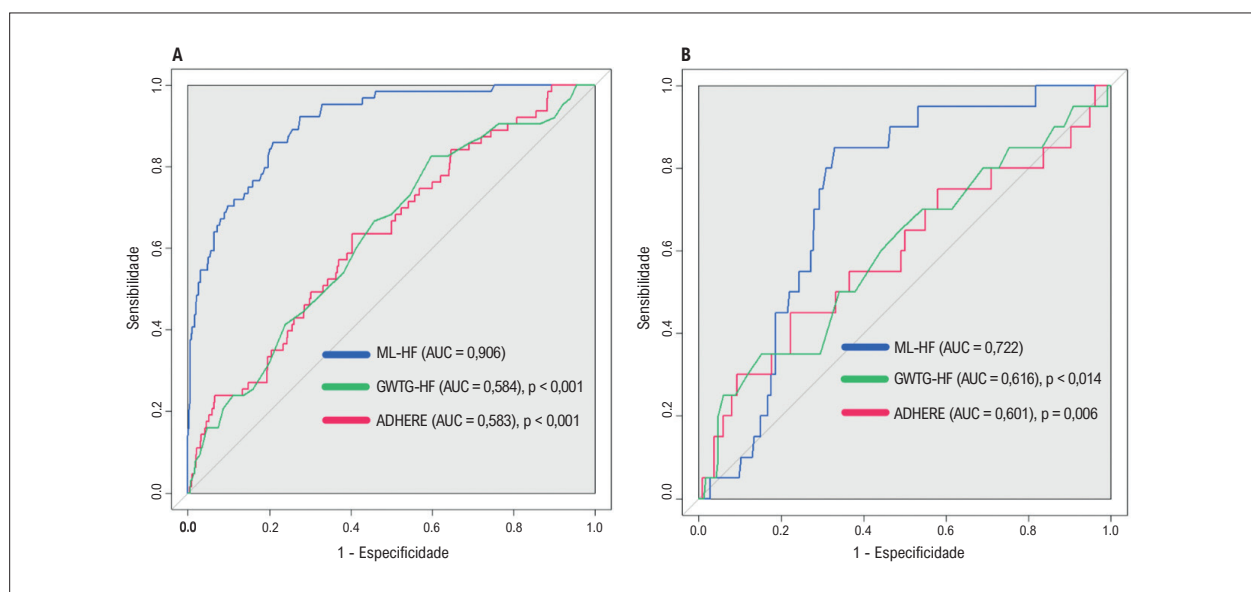


Figura 3 – Comparação das curvas ROC (Receiver Operating Characteristic) entre os escores ML-HF, GWTG-HF e ADHERE nos conjuntos de treinamento (A) e teste (B).

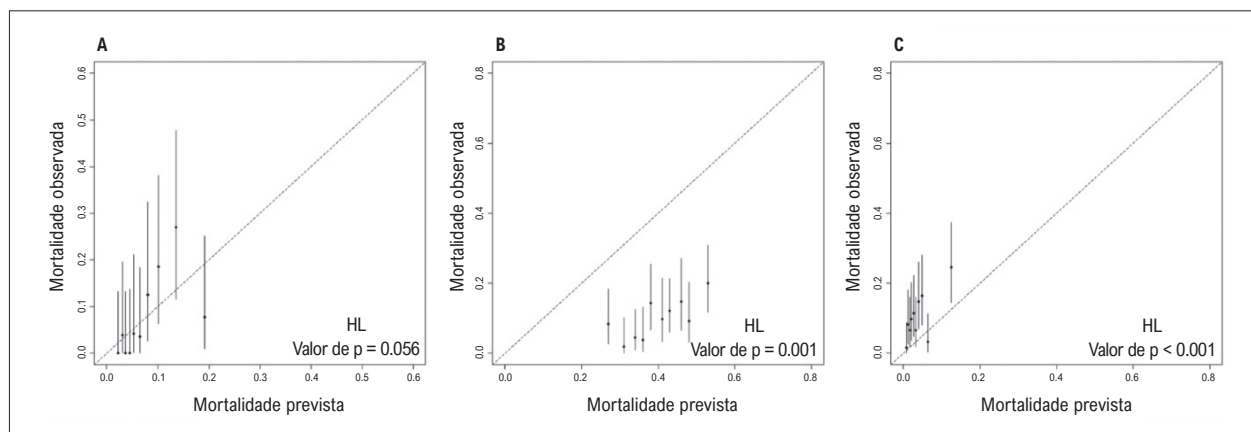


Figura 4 – Comparação da mortalidade hospitalar observada e prevista para avaliar a calibração do modelo dos escores ML-HF (A), GWTG-HF (B) e ADHERE (C) no conjunto de teste. HL: Teste de Hosmer-Lemeshow. Um valor de p menor que 0,05 indica um ajuste inadequado dos dados.

recomendação de classe 2a.¹⁰ Nosso estudo contribui para a literatura ao demonstrar o papel prognóstico independente da medição do RRP na admissão de pacientes com IC descompensada.

É possível destacar o fato de que novas variáveis, além do domínio de saúde física do WHOQOL-Brief, foram relevantes para a predição de óbito hospitalar em nosso estudo, fatores que não são avaliados em outros escores, como hemoglobina, diâmetro do átrio esquerdo, presença de prótese valvar, número de procedimentos cirúrgicos e presença de hipotireoidismo. A relevância da presença de prótese valvar pode ser devida à alta prevalência de febre reumática em nossa população. Isso, mais uma vez, destaca a importância do desenvolvimento de escores personalizados para PBM.²¹

Modelos de ML foram testados anteriormente para prever a mortalidade hospitalar de pacientes com IC aguda em outros países. O modelo MARKER-HF (Machine learning Assessment of Risk and Early mortality in Heart Failure) demonstrou excelente poder discriminatório, com uma AUC de 0,88. No entanto, esse modelo foi desenvolvido usando pacientes com IC crônica e aguda, com durações de acompanhamento variadas.⁹ Segar et al. utilizaram dados do registro GWTG nos Estados Unidos para construir um modelo baseado em ML para prever a mortalidade hospitalar, que foi validado na coorte ARIC (Atherosclerosis Risk in Community).⁵ Assim como em nossa investigação, eles observaram que o modelo baseado em ML superou o escore GWTG-HF, apresentando um índice C de aproximadamente 0,80. Eles puderam

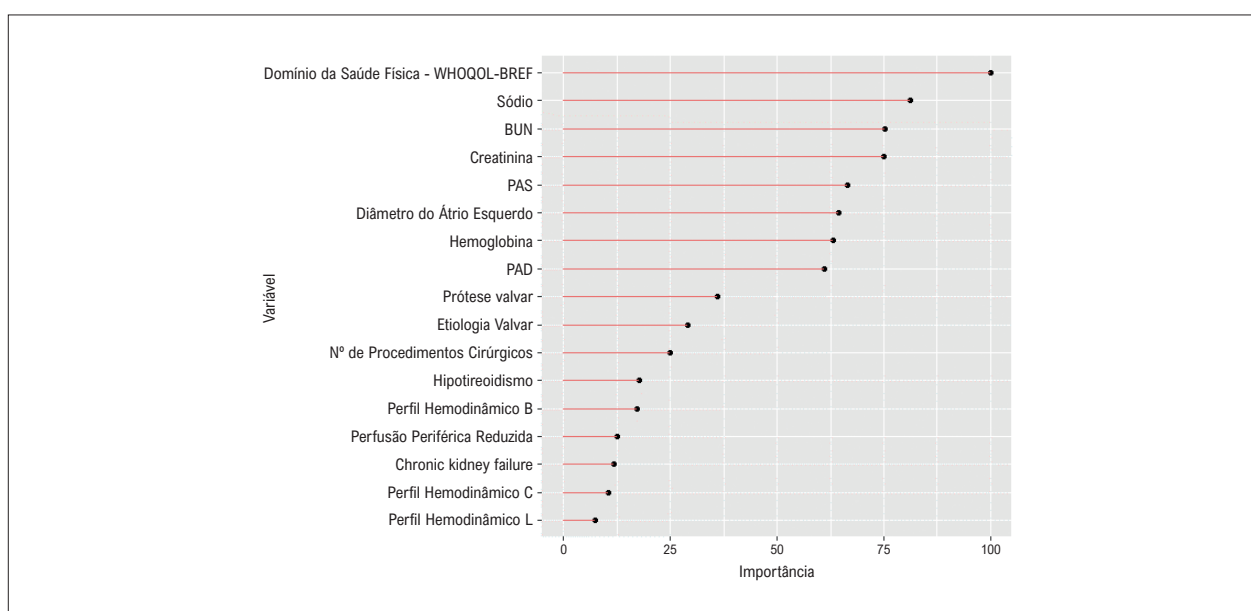


Figura 5 – Importância das variáveis incluídas no escore ML-HF. WHOQOL-BREF: Versão resumida do questionário de qualidade de vida da Organização Mundial da Saúde; PAS: pressão arterial sistólica; PAD: pressão arterial diastólica; BUN: Nitrogênio ureico no sangue.

incorporar biomarcadores séricos, como níveis sanguíneos de troponina e peptídeos natriuréticos, e determinantes sociais de saúde em nível contextual, o que ajuda a reduzir o viés do algoritmo e melhorar a equidade entre populações carentes.^{4,5}

Diferenças nas populações de pacientes, no manejo clínico e na disponibilidade de biomarcadores, como peptídeos natriuréticos, podem explicar, em parte, a maior precisão do modelo de Segar et al.⁵ em comparação ao nosso; no entanto, nosso modelo pode ser particularmente valioso em cenários com acesso limitado a esses biomarcadores, especialmente em PBMR. Em nosso estudo, determinantes sociais individuais foram utilizados, mas não selecionados a partir do algoritmo baseado em ML. Por outro lado, demonstramos pela primeira vez que uma medida de qualidade de vida ajuda a melhorar a previsão de morte hospitalar de pacientes com IC em um modelo baseado em ML.

Nosso estudo tem pontos fortes que devem ser destacados. Primeiramente, analisamos os determinantes sociais da saúde individualmente, embora não tenham sido selecionados como variáveis importantes. Esse achado está em linha com a literatura atual, que afirma que os determinantes sociais podem ser relevantes apenas em subgrupos mais homogêneos, o que pode não ser o caso da população brasileira.⁴ Em segundo lugar, também realizamos validação externa usando uma coorte diferente daquela usada para o treinamento do modelo, o que deve ser incentivado em estudos que utilizam ML.⁷ Em terceiro lugar, nossa amostra contém pacientes de todas as macrorregiões do país, abrangendo uma ampla gama de características demográficas e status sociais. Essa diversidade aumenta a generalização e a validação de nossos achados em diferentes populações. Em quarto lugar, nossos dados foram coletados sistematicamente como parte de um estudo de coorte prospectivo, em vez de depender de RME, que são mais suscetíveis ao viés de classificação incorreta.²²⁻²⁶

Este estudo também apresenta limitações que devem ser consideradas. Dados ausentes levaram a uma redução no tamanho da amostra inicial, diminuindo-a para aproximadamente 25% do tamanho original, pois removemos pacientes com dados ausentes em qualquer uma das variáveis mais significativas. Incluímos apenas casos completos, assumindo que os dados estavam completamente ausentes aleatoriamente. Optamos por não imputar múltiplas variáveis para evitar a introdução de viés potencial de estimativas baseadas em modelos em um ambiente preditivo. Além disso, os peptídeos natriuréticos não foram utilizados para construir o modelo devido à falta de disponibilidade na maioria dos centros, por se tratar de um banco de dados em que a grande maioria dos locais é financiada publicamente. Reconhecemos que a ausência desses marcadores pode prejudicar a análise, visto que atualmente são testes essenciais para a confirmação do diagnóstico de IC. Em países com maior disponibilidade, acreditamos que essa variável pode ser ainda mais relevante e aumentar a precisão do algoritmo. Por outro lado, isso poderia melhorar a aplicabilidade do escore ML em outros PBMR, onde o acesso ao BNP também pode ser limitado.

Apesar dos resultados promissores, a implementação de escores de risco baseados em ML na prática clínica continua desafiadora. Ao contrário dos escores de risco tradicionais, que podem usar tabelas ou calculadoras simples, os modelos de ML exigem infraestrutura digital para entrada de dados, processamento e integração de resultados em fluxos de trabalho clínicos. Em nosso estudo, o desenvolvimento e a validação do modelo foram realizados em um computador centralizado contendo dados de todos os centros participantes, e a análise foi computacionalmente intensiva. Para uso no mundo real, a implantação exigiria sistemas compatíveis e entrada de dados padronizada. Estratégias como o desenvolvimento de aplicativos móveis offline ou a integração do modelo em

prontuários eletrônicos de saúde de baixa complexidade poderiam ser exploradas para facilitar a implementação em ambientes com recursos limitados.

Embora o modelo atual possa ser aplicado sem retreinamento, os modelos baseados em ML podem ser atualizados ou retreinados periodicamente para melhorar o desempenho e se adaptar à evolução das populações clínicas. Isso levanta importantes considerações éticas e regulatórias, incluindo privacidade de dados, consentimento informado para uso secundário de dados de pacientes e transparência em relação aos procedimentos de atualização do modelo. A implementação eficaz dependerá não apenas de uma infraestrutura técnica adequada para atender às demandas computacionais, mas também de fortes estruturas de supervisão e governança para garantir o uso seguro, equitativo e responsável.

Conclusão

Nós desenvolvemos e validamos um modelo baseado em aprendizado de máquina - o escore ML-HF - para prever o risco de morte hospitalar em pacientes com IC aguda. Essa ferramenta preditiva apresentou poder discriminativo superior aos escores tradicionais ADHERE e GWTC.

Agradecimentos

Os autores gostariam de expressar sua gratidão aos colegas participantes que apoiaram o recrutamento para o programa BPC.

Contribuição dos autores

Concepção e desenho da pesquisa: Ferreira MBD, Daikubara Neto JT, Taniguchi FP, Fernandes-Silva MM; Obtenção de dados: Ferreira MBD, Daikubara Neto JT, Cunha GSP, Moretti R, Reichert JT, Prado LM; Análise e interpretação dos dados: Ferreira MBD, Daikubara Neto JT, Fernandes-Silva MM; Análise estatística: Ferreira MBD, Daikubara Neto JT;

Obtenção de financiamento: Taniguchi FP; Redação do manuscrito: Ferreira MBD, Daikubara Neto JT, Fernandes-Silva MM; Revisão crítica do manuscrito quanto ao conteúdo: Cunha GSP, Cirino RHD, Smith Jr SC, Taniguchi FP, Sposito A, Silvestre OM, Nadruz W, Fernandes-Silva MM.

Potencial conflito de interesse

Não há conflito com o presente artigo.

Fontes de financiamento

O presente estudo foi financiado por American Heart Association, Hospital do Coração (HCor) e Ministério da Saúde.

Vinculação acadêmica

Este artigo é parte de dissertação de mestrado de Gustavo Sarot Pereira pela Universidade Federal do Paraná.

Aprovação ética e consentimento informado

Este estudo foi aprovado pelo Comitê de Ética do Hospital do Coração - HCor / Associação Beneficente Síria sob o número de protocolo 48561715.5.1001.0060. Todos os procedimentos envolvidos nesse estudo estão de acordo com a Declaração de Helsinki de 1975, atualizada em 2013. O consentimento informado foi obtido de todos os participantes incluídos no estudo.

Uso de Inteligência Artificial

Os autores não utilizaram ferramentas de inteligência artificial no desenvolvimento deste trabalho.

Disponibilidade de Dados

Os dados estarão disponíveis sob demanda aos autores.

Referências

1. Albuquerque DC, Barros e Silva PGM, Lopes RD, Hoffmann-Filho CR, Nogueira PR, Reis H, et al. In-Hospital Management and Long-Term Clinical Outcomes and Adherence in Patients with Acute Decompensated Heart Failure: Primary Results of the First Brazilian Registry of Heart Failure (BREATHE). *J Card Fail.* 2024;30(5):639-50. doi: 10.1016/j.cardfail.2023.08.014.
2. Royston P, Altman DG. Risk Stratification for in-Hospital Mortality in Acutely Decompensated Heart Failure. *JAMA.* 2005;293(20):2467-8. doi: 10.1001/jama.293.20.2467-c.
3. 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(1):25-32. doi: 10.1161/CIRCOUTCOMES.109.854877.
4. Li Y, Wang H, Luo Y. Improving Fairness in the Prediction of Heart Failure Length of Stay and Mortality by Integrating Social Determinants of Health. *Circ Heart Fail.* 2022;15(11):e009473. doi: 10.1161/CIRCHEARTFAILURE.122.009473.
5. Segar MW, Hall JL, Jhund PS, Powell-Wiley TM, Morris AA, Kao D, et al. Machine Learning-Based Models Incorporating Social Determinants of Health vs Traditional Models for Predicting in-Hospital Mortality in Patients with Heart Failure. *JAMA Cardiol.* 2022;7(8):844-54. doi: 10.1001/jamacardio.2022.1900.
6. Deo RC. Machine Learning in Medicine. *Circulation.* 2015;132(20):1920-30. doi: 10.1161/CIRCULATIONAHA.115.001593.
7. Shin S, Austin PC, Ross HJ, Abdel-Qadir H, Freitas C, Tomlinson G, et al. Machine Learning vs. Conventional Statistical Models for Predicting Heart Failure Readmission and Mortality. *ESC Heart Fail.* 2021;8(1):106-15. doi: 10.1002/ehf2.13073.
8. Angraal S, Mortazavi BJ, Gupta A, Khera R, Ahmad T, Desai NR, et al. Machine Learning Prediction of Mortality and Hospitalization in Heart Failure with Preserved Ejection Fraction. *JACC Heart Fail.* 2020;8(1):12-21. doi: 10.1016/j.jchf.2019.06.013.
9. 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(1):139-47. doi: 10.1002/ehf.1628. Epub 2019 Nov 12.
10. Heidenreich PA, Bozkurt B, Aguilar D, Allen LA, Byun JJ, Colvin MM, et al. 2022 AHA/ACC/HFSA Guideline for the Management of Heart Failure: A Report of the American College of Cardiology/American Heart Association

- Joint Committee on Clinical Practice Guidelines. *J Am Coll Cardiol*. 2022;79(17):e263-e421. doi: 10.1016/j.jacc.2021.12.012.
11. Taniguchi FP, Bernardez-Pereira S, Silva SA, Ribeiro ALP, Morgan L, Curtis AB, et al. Implementation of a Best Practice in Cardiology (BPC) Program Adapted from Get with the Guidelines® in Brazilian Public Hospitals: Study Design and Rationale. *Arq Bras Cardiol*. 2020;115(1):92-9. doi: 10.36660/abc.20190393.
 12. Taniguchi FP, Bernardez-Pereira S, Ribeiro ALP, Morgan L, Curtis AB, Taubert K, et al. A Nationwide Initiative to Improve Cardiology Quality: The Best Practice in Cardiology Program in Brazil. *Arq Bras Cardiol*. 2023;120(10):e20230375. doi: 10.36660/abc.20230375.
 13. Haugen AS, Sevdalis N, Søfteland E. Impact of the World Health Organization Surgical Safety Checklist on Patient Safety. *Anesthesiology*. 2019;131(2):420-5. doi: 10.1097/ALN.0000000000002674.
 14. Jakobsen JC, Gluud C, Wetterslev J, Winkel P. When and How Should Multiple Imputation be Used for Handling Missing Data in Randomised Clinical Trials - A Practical Guide with Flowcharts. *BMC Med Res Methodol*. 2017;17(1):162. doi: 10.1186/s12874-017-0442-1.
 15. Kursa MB, Rudnicki WR. Feature Selection with the Boruta Package. *J Stat Softw*. 2010;36(11):1-13. doi: 10.18637/jss.v036.i11.
 16. Kowarik A, Templ M. Imputation with the R Package VIM. *J Stat Softw*. 2016;74(7):1-16. doi: 10.18637/jss.v074.i07.
 17. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the Areas Under Two or More Correlated Receiver Operating Characteristic Curves: A Nonparametric Approach. *Biometrics*. 1988;44(3):837-45.
 18. Collins GS, Reitsma JB, Altman DG, Moons KGM; members of the TRIPOD group. Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD): The TRIPOD Statement. *Eur Urol*. 2015;67(6):1142-51. doi: 10.1016/j.eururo.2014.11.025.
 19. Greene SJ, Butler J, Spertus JA, Hellkamp AS, Vaduganathan M, DeVore AD, et al. Comparison of New York Heart Association Class and Patient-Reported Outcomes for Heart Failure with Reduced Ejection Fraction. *JAMA Cardiol*. 2021;6(5):522-31. doi: 10.1001/jamacardio.2021.0372.
 20. Hawwa N, Vest AR, Kumar R, Lahoud R, Young JB, Wu Y, et al. Comparison between the Kansas City Cardiomyopathy Questionnaire and New York Heart Association in Assessing Functional Capacity and Clinical Outcomes. *J Card Fail*. 2017;23(4):280-5. doi: 10.1016/j.cardfail.2016.12.002.
 21. Coffey S, Cairns BJ, Lung B. The Modern Epidemiology of Heart Valve Disease. *Heart*. 2016;102(1):75-85. doi: 10.1136/heartjnl-2014-307020.
 22. 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.
 23. Desai RJ, Wang SV, Vaduganathan M, Evers T, Schneeweiss S. Comparison of Machine Learning Methods with Traditional Models for Use of Administrative Claims with Electronic Medical Records to Predict Heart Failure Outcomes. *JAMA Netw Open*. 2020;3(1):e1918962. doi: 10.1001/jamanetworkopen.2019.18962.
 24. Negassa A, Ahmed S, Zolty R, Patel SR. Prediction Model Using Machine Learning for Mortality in Patients with Heart Failure. *Am J Cardiol*. 2021;153:86-93. doi: 10.1016/j.amjcard.2021.05.044.
 25. Lv H, Yang X, Wang B, Wang S, Du X, Tan Q, et al. Machine Learning-Driven Models to Predict Prognostic Outcomes in Patients Hospitalized with Heart Failure Using Electronic Health Records: Retrospective Study. *J Med Internet Res*. 2021;23(4):e24996. doi: 10.2196/24996.
 26. Barbalho IMP, Fernandes F, Barros DMS, Paiva JC, Henriques J, Morais AHF, et al. Electronic Health Records in Brazil: Prospects and Technological Challenges. *Front Public Health*. 2022;10:963841. doi: 10.3389/fpubh.2022.963841.

*Material suplementar

Para informação adicional, por favor, clique aqui.



Este é um artigo de acesso aberto distribuído sob os termos da licença de atribuição pelo Creative Commons