



UNIVERSIDADE ESTADUAL PAULISTA
“JÚLIO DE MESQUITA FILHO”
Instituto de Ciência e Tecnologia
Câmpus de Sorocaba

RICARDO BOZOLLAN OLIVEIRA

**PREDIÇÃO DE MORTALIDADE CARDÍACA EM PACIENTES COM
INSUFICIÊNCIA CARDÍACA: UMA ABORDAGEM COMPARATIVA ENTRE
MODELOS TRADICIONAIS, *MACHINE LEARNING* E *DEEP LEARNING***

Sorocaba

2025

RICARDO BOZOLLAN OLIVEIRA

**PREDIÇÃO DE MORTALIDADE CARDÍACA EM PACIENTES COM
INSUFICIÊNCIA CARDÍACA: UMA ABORDAGEM COMPARATIVA ENTRE
MODELOS TRADICIONAIS, *MACHINE LEARNING* E *DEEP LEARNING***

Trabalho de Conclusão de Curso apresentado à
Universidade Estadual Paulista (UNESP),
Instituto de Ciência e Tecnologia, Sorocaba,
como parte dos requisitos para obtenção do
grau de Bacharel em Engenharia de Controle e
Automação.

Orientador: Prof. Dr. Leopoldo André Dutra
Lusquino Filho

Sorocaba

2025

O48p

Oliveira, Ricardo Bozollan

Predição de mortalidade cardíaca em pacientes com insuficiência cardíaca :
uma abordagem comparativa entre modelos tradicionais, machine learning e
deep learning / Ricardo Bozollan Oliveira. -- Sorocaba, 2025
65 p.

Trabalho de conclusão de curso (Bacharelado - Engenharia de Controle e
Automação) - Universidade Estadual Paulista (UNESP), Instituto de Ciência e
Tecnologia, Sorocaba

Orientador: Leopoldo André Dutra Lusquino Filho

1. Aprendizagem profunda (Aprendizado do computador). 2. Aprendizado
do computador. 3. Insuficiência cardíaca. 4. Inteligência artificial - Aplicações
médicas. I. Título.

RICARDO BOZOLLAN OLIVEIRA

**PREDIÇÃO DE MORTALIDADE CARDÍACA EM PACIENTES COM
INSUFICIÊNCIA CARDÍACA: UMA ABORDAGEM COMPARATIVA ENTRE
MODELOS TRADICIONAIS, *MACHINE LEARNING* E *DEEP LEARNING***

Trabalho de Conclusão de Curso apresentado ao Instituto de Ciência e Tecnologia de Sorocaba, Universidade Estadual Paulista (UNESP), como parte dos requisitos para obtenção do grau de Bacharel(a) em Engenharia de Controle e Automação.

Data da defesa: 12/06/2025

BANCA EXAMINADORA:

Prof. Dr. Leopoldo André Dutra Lusquino Filho
UNESP – Instituto de Ciência e Tecnologia – Campus de Sorocaba

Prof. Dr. Luis De Oro Arenas
UNESP – Instituto de Ciência e Tecnologia – Campus de Sorocaba

Andressa Maira Sardinha de Souza
UNESP – Instituto de Ciência e Tecnologia – Campus de Sorocaba

AGRADECIMENTOS

Aos meus pais, Eduardo Oliveira e Ana Maria Bozollan, por todo amor, dedicação e pelos valores que sempre me transmitiram ao longo da vida.

À minha irmã, Ana Paula Oliveira, pelo carinho, apoio constante e por estar sempre presente na minha trajetória.

À minha namorada, Nathalia Bersan, pelo amor incondicional, companheirismo e por sua ajuda fundamental durante a realização deste trabalho.

Aos amigos Caio Cortada, Christian Eduard, Giuliano Faria, Guilherme Gomes Leite, João Paulo Barbosa, Santiago Teixeira e Thiago Colombari, pela amizade sincera, pelo apoio e por compartilharem comigo os anos da graduação.

Ao meu orientador, Prof. Dr. Leopoldo André Dutra Lusquino Filho, por sua orientação dedicada, pelo compartilhamento de conhecimento e por todo o suporte técnico e humano ao longo deste projeto.

Às instituições Dinâmica Engenharia Júnior e Semana da Engenharia, e a todos os seus membros, pelas experiências que contribuíram para minha formação pessoal e profissional.

E, por fim, à Universidade Estadual Paulista (UNESP), pela oportunidade de cursar Engenharia de Controle e Automação e por proporcionar uma formação de excelência pautada no conhecimento, na ética e na transformação social.

RESUMO

Este trabalho tem como objetivo avaliar e comparar diferentes abordagens para a predição de mortalidade cardíaca em pacientes com insuficiência cardíaca, utilizando o banco de dados MUSIC. Inicialmente, foi aplicada uma versão adaptada do modelo clínico tradicional Seattle Heart Failure Model (SHFM), cuja implementação foi limitada pelas variáveis disponíveis no dataset. Essa versão apresentou desempenho modesto, com AUC-ROC de 0,637 e baixa sensibilidade, evidenciando suas limitações na predição individual de risco. Em seguida, foram implementados modelos supervisionados de aprendizado de máquina (Regressão Logística, Random Forest, SVM e XGBoost) que superaram significativamente o SHFM adaptado, alcançando AUC-ROC superiores a 0,71 e maior capacidade de detecção de pacientes de alto risco. Além disso, foi realizado o ajuste de hiperparâmetros nos modelos Random Forest e XGBoost, buscando melhorias adicionais, e implementados modelos de *deep learning*, incluindo MLP e MLP com camada Transformer. Embora os modelos de *deep learning* tenham superado o SHFM adaptado (AUC até 0,75), os melhores resultados gerais foram obtidos com Random Forest e XGBoost, indicando que modelos tradicionais bem calibrados podem, neste contexto, superar arquiteturas mais complexas. Esses resultados reforçam o potencial das abordagens baseadas em aprendizado de máquina para apoiar a prática clínica na estratificação de risco, ao mesmo tempo em que destacam a importância do balanceamento entre complexidade, tamanho amostral e capacidade de generalização.

Palavras-chave: Insuficiência cardíaca, mortalidade cardíaca, aprendizado de máquina, *deep learning*, SHFM, Random Forest, XGBoost, MLP, Transformer.

ABSTRACT

This study aims to evaluate and compare different approaches for predicting cardiac mortality in patients with heart failure using the MUSIC dataset. Initially, an adapted version of the traditional clinical Seattle Heart Failure Model (SHFM) was applied, whose implementation was limited by the variables available in the dataset. This adapted version showed modest performance, with an AUC-ROC of 0.637 and low sensitivity, highlighting its limitations in individual risk prediction. Subsequently, supervised machine learning models (Logistic Regression, Random Forest, SVM, and XGBoost) were implemented, significantly outperforming the adapted SHFM by achieving AUC-ROC values above 0.71 and greater ability to detect high-risk patients. Additionally, hyperparameter tuning was performed on the Random Forest and XGBoost models to seek further performance improvements, and deep learning models were implemented, including MLP and MLP with a Transformer layer. Although the deep learning models outperformed the adapted SHFM (AUC-ROC up to 0.75), the best overall results were obtained with Random Forest and XGBoost, indicating that well-calibrated traditional models can, in this context, surpass more complex architectures. These findings reinforce the potential of machine learning-based approaches to support clinical practice in risk stratification while highlighting the importance of balancing algorithmic complexity, sample size, and generalization capacity.

Keywords: heart failure, cardiac mortality, machine learning, deep learning, SHFM, Random Forest, XGBoost, MLP, Transformer.

LISTA DE FIGURAS

Figura 1- Representação esquemática Random Forest.....	24
Figura 2- Esquemático do Pipeline.....	31
Figura 3- Configuração ambiente de execução do Google Colab.....	32
Figura 4 - Distribuição dataset MUSIC.....	35
Figura 5 - Curva ROC e AUC SHFM Adaptado.....	41
Figura 6 - Curva ROC + AUC-ROC SHFM adaptado.....	42
Figura 7 - Matrizes de Confusão dos Modelos (Validação) (a) Regressão Logística, (b) SVM, (c) Random Forest e (d) XGBoost.....	44
Figura 8 - Matrizes de Confusão dos Modelos (Teste) (a) Regressão Logística, (b) SVM, (c) Random Forest e (d) XGBoost.....	45
Figura 9 - Curva ROC-AUC dos modelos Supervisionados (a) Conjunto de Validação (b) Conjunto de testes.....	47
Figura 10 - Curva ROC-AUC Validação Cruzada.....	48
Figura 11 - Curva ROC-AUC comparativa de todos os modelos.....	49
Figura 12 - Curvas de perdas do MLP (a) e MLP + Transformer (c), e curvas AUC do MLP (b) e MLP + Transformer (d).....	51
Figura 13 - Curvas ROC no conjunto de validação do MLP (a) e MLP + Transformer (c), e curvas ROC no conjunto de teste do MLP (b) e MLP + Transformer (d).....	52
Figura 14 - Curvas ROC com validação cruzada estratificada 5-fold dos modelos MLP (a) e MLP + Transformer (b).....	53
Figura 15 - Curvas ROC e AUC modelos Deep e SHFM adaptado.....	54

LISTA DE TABELAS

Tabela 1 - Métricas SHFM adaptado.....	42
Tabela 2 - Dados de Pacientes Aleatórios.....	42
Tabela 3 - Métricas no conjunto validação para os modelos supervisionados..	46
Tabela 4 - Métricas no conjunto de teste para os modelos supervisionados.....	46

LISTA DE ABREVIATURAS E SIGLAS

AUC-ROC	- Área sob a Curva Receiver Operating Characteristic
CNN	- Convolutional Neural Network
DL	- Deep Learning
ECG	- Eletrocardiograma
FEVE / LVEF	- Fração de Ejeção do Ventrículo Esquerdo
IECA	- Inibidor da Enzima Conversora da Angiotensina
ML	- Machine Learning
MLP	- Multi-Layer Perceptron
NYHA	- New York Heart Association
PFD	- Pump Failure Death
RF	- Random Forest
RL	- Regressão Logística
SCD	- Sudden Cardiac Death
SHFM	- Protocol Buffers
SMOTE	- Synthetic Minority Over-sampling Technique
SVM	- Support Vector Machine
TRC	- Terapia de Ressincronização Cardíaca
XGBoost	- Extreme Gradient Boosting

LISTA DE SÍMBOLOS

β	- Coeficiente estimado em modelos de regressão
γ	- Parâmetro de regularização no XGBoost
δ	- Derivada da função de perda
η	- Taxa de aprendizado no treinamento de redes neurais
i	- Índice de observação ou iterador
J	- Número de folhas de uma árvore de decisão
n	- Número de amostras (ou variáveis)
p	- Probabilidade estimada de ocorrência do evento
ξ_i	- Variável de folga na formulação do SVM
ω_j	- Valor do peso da folha j em uma árvore
$\hat{y}_i$	- Valor predito para a i -ésima amostra
x_i	- Valor da i -ésima variável de entrada
w	- Vetor de pesos nos modelos lineares
z	- Soma ponderada linear de entradas em um neurônio
a	- Saída de ativação de um neurônio

SUMÁRIO

1 INTRODUÇÃO.....	13
2 REVISÃO BIBLIOGRÁFICA.....	16
2.1 Machine Learning Aplicado ao Diagnóstico Médico.....	16
2.2 Seattle Heart Failure Model (SHFM).....	18
2.3 Modelos de Aprendizado de Máquina.....	21
2.3.1 Regressão Logística.....	22
2.3.2 Random Forest.....	23
2.3.3 Support Vector Machine.....	24
2.3.4 XGBoost.....	26
2.3.5 Redes Neurais e Transformer.....	27
2.4 Métricas.....	28
3 MATERIAIS E MÉTODOS.....	30
3.1 Abordagem Proposta.....	30
3.2 Configuração para os experimentos.....	31
3.3 Base de Dados e Pré-processamento.....	32
3.4 Implementação e Adaptação SHFM.....	35
3.5 Modelos de Machine Learning Supervisionados.....	37
3.6 Modelos Deep Learning e Transformer.....	37
3.7 Estratégia de Validação, Avaliação e Métricas.....	39
4 RESULTADOS EXPERIMENTAIS E DISCUSSÃO.....	40
4.1 Resultados SHFM Adaptado.....	40
4.2 Resultados Modelos Supervisionados.....	43
4.3 Resultados Redes Neurais Profundas.....	50
4.4 Custo Computacional.....	54
4.4 Discussão dos Resultados.....	55
5 CONCLUSÃO.....	58
REFERÊNCIAS.....	61

1 INTRODUÇÃO

As doenças cardiovasculares permanecem como a principal causa de mortalidade no mundo, sendo responsáveis por aproximadamente 20,5 milhões de óbitos em 2021, o que equivale a cerca de um terço de todas as mortes globais (World Heart Federation, 2023). No Brasil, apesar de ter ocorrido uma redução na taxa de mortalidade padronizada por idade (de 348,5 para 162,2 por 100 mil habitantes entre 1990 e 2021), as doenças cardiovasculares continuam a representar uma das maiores causas de óbito e hospitalização, com impacto significativo sobre os sistemas público e privado de saúde (Oliveira et al., 2023). Dentre essas doenças, a insuficiência cardíaca é particularmente preocupante, por conta das suas altas taxas de morbidade, reinternação e mortalidade, principalmente entre indivíduos idosos (Siqueira et al., 2020). Sua crescente prevalência está associada ao envelhecimento populacional, ao estilo de vida sedentário e à maior incidência de fatores de risco como hipertensão arterial, diabetes mellitus e infarto do miocárdio prévio (Ponikowski et al., 2016). Assim, nota-se que existe uma necessidade de desenvolvimento de estratégias clínicas e tecnológicas que aprimorem o rastreamento de risco e a tomada de decisões médicas, contribuindo para a redução das complicações e a melhor alocação de recursos em saúde.

A insuficiência cardíaca (IC) é uma síndrome clínica complexa, caracterizada pela incapacidade do coração em bombear sangue de maneira eficiente para atender às necessidades metabólicas do organismo, ou em fazê-lo apenas com pressões de enchimento elevadas (Ponikowski et al., 2016). Os principais sintomas incluem dispneia (falta de ar), fadiga e intolerância ao exercício (Heidenreich et al., 2022). Essa condição é considerada uma das principais causas de hospitalização em idosos e está associada a altas taxas de morbidade e mortalidade. O prognóstico de pacientes com insuficiência cardíaca pode ser bastante variável, sendo influenciado por fatores como idade, fração de ejeção ventricular, comorbidades e resposta ao tratamento (Ponikowski et al., 2016). Estimar com precisão o risco de morte desses pacientes representa um grande desafio clínico, uma vez que a evolução da doença é frequentemente não linear e multifatorial, exigindo a integração de múltiplos parâmetros clínicos, laboratoriais e demográficos.

Para lidar com esse problema, existem alguns modelos tradicionais de predição de risco cardiovascular, como o Framingham Risk Score (D'Agostino et al., 2008) e o Seattle Heart Failure Model (SHFM) (Levy et al., 2006), que são utilizados na prática clínica por sua simplicidade e capacidade de estratificação de pacientes com base em variáveis clínicas e laboratoriais. No entanto, a aplicação integral do SHFM pode ser inviável em alguns

contextos, especialmente quando há ausência de determinadas variáveis no banco de dados. Por esse motivo, este estudo utilizou uma versão adaptada do SHFM, construída a partir das variáveis disponíveis no dataset MUSIC. O SHFM, especificamente, estima a sobrevida de pacientes com insuficiência cardíaca crônica a partir de parâmetros como idade, sexo, fração de ejeção ventricular, níveis de sódio e creatinina séricos, uso de medicamentos e presença de dispositivos implantáveis. Mas tais modelos apresentam limitações relevantes, como por exemplo subestimarem o risco em pacientes hospitalizados com insuficiência cardíaca descompensada em alguns casos, tornando sua aplicabilidade a contextos clínicos mais complexos (Albuquerque et al., 2019). Além disso, esses escores possuem uma certa limitação por terem sido desenvolvidos com base em populações específicas (Pocock et al., 2013), e sua aplicação pode ser dificultada pela necessidade de inserção manual de diversas variáveis, principalmente em ambientes com recursos restritos. Essas dificuldades reforçam a importância de abordagens mais adaptáveis e precisas, como as técnicas de aprendizado de máquina, que permitem a incorporação de conjuntos de dados mais amplos e heterogêneos, capturando padrões complexos não identificáveis por métodos convencionais.

Diante deste cenário, este trabalho tem como objetivo a utilização de algoritmos de aprendizado de máquina para a predição da mortalidade em pacientes com insuficiência cardíaca crônica. O objetivo é avaliar a capacidade desses algoritmos de identificar relações não lineares entre variáveis clínicas, com o intuito de ser mais eficiente que o SHFM para o *dataset* tratado. Utilizou-se o banco de dados MUSIC (MUerte Súbita en Insuficiencia Cardíaca), disponível publicamente na plataforma PhysioNet, que contém informações clínicas e laboratoriais detalhadas de pacientes acompanhados ao longo de cinco anos (MARTÍN-YEBRA et al., 2024).

Inicialmente, conduziu-se uma análise exploratória dos dados, seguida pelo tratamento de valores ausentes com estratégias baseadas na distribuição estatística de cada variável. Colunas redundantes, com alto índice de ausência ou irrelevantes para a modelagem foram removidas para evitar ruídos em modelos de ML. Em paralelo, foi implementado o cálculo de uma versão adaptada do escore SHFM, baseada nos coeficientes originais descritos na literatura, utilizando apenas as variáveis clínicas e laboratoriais disponíveis no dataset, previamente padronizadas por meio do z-score. O risco de mortalidade em cinco anos foi estimado por meio da função de Gompertz, gerando um *baseline* clínico robusto para comparação com modelos de aprendizado de máquina.

Na etapa de modelagem preditiva, foi criada uma versão do conjunto de dados voltada à aplicação de algoritmos de aprendizado de máquina. Após um novo tratamento de dados e a

divisão em subconjuntos de treino, validação e teste, aplicou-se a técnica SMOTE para correção do desbalanceamento entre as classes (paciente vivo ou morto). Foram desenvolvidos e avaliados modelos baseados em Regressão Logística, Random Forest, XGBoost, Support Vector Machine (SVM) e MLP com e sem Transformer, utilizando métricas como AUC-ROC, precisão, *recall* e F1-score. Os modelos Random Forest e XGBoost foram testados com diferentes configurações por meio de ajustes de hiperparâmetros via GridSearchCV. As curvas ROC geradas para os conjuntos de validação e teste possibilitaram a comparação direta entre os modelos desenvolvidos e o SHFM adaptado, permitindo avaliar de forma visual e quantitativa o potencial superior das abordagens baseadas em aprendizado de máquina na predição de mortalidade cardíaca em cinco anos.

As próximas seções deste trabalho estão organizadas da seguinte forma: a Seção 2 - Revisão de Literatura aborda os fundamentos do aprendizado de máquina aplicado ao diagnóstico médico (Seção 2.1), detalha o SHFM e suas limitações (Seção 2.2), e apresenta os principais modelos de aprendizado de máquina utilizados no estudo, como Regressão Logística, Random Forest, SVM, XGBoost e Redes Neurais com Transformer (Seção 2.3). A Seção 3 - Materiais e Métodos descreve a abordagem proposta (Seção 3.1), o setup dos experimentos (Seção 3.2), a base de dados e o pré-processamento realizado (Seção 3.3), a implementação da versão adaptada do SHFM (Seção 3.4), a modelagem dos algoritmos supervisionados (Seção 3.5), os modelos de *deep Learning* e Transformer utilizados (Seção 3.6) e a estratégia de validação e métricas adotadas (Seção 3.7). A Seção 4 - Resultados e Discussões apresenta os resultados obtidos com o SHFM (Seção 4.1), com os modelos supervisionados (Seção 4.2), com as redes neurais profundas (Seção 4.3) e discute os resultados do trabalho (Seção 4.4). Por fim, a Seção 5 (conclusão) sintetiza as principais contribuições do estudo e propõe direções para trabalhos futuros.

2 REVISÃO BIBLIOGRÁFICA

2.1 *Machine Learning* Aplicado ao Diagnóstico Médico

O *Machine Learning* (ML) é um ramo da inteligência artificial que capacita sistemas computacionais a aprenderem padrões a partir de dados, o que possibilita realizar previsões ou classificações sem serem explicitamente programados para tarefas específicas. Existem três tipos principais técnicas categorizadas: aprendizado supervisionado, em que os modelos são treinados com dados rotulados; aprendizado não supervisionado, em que identifica-se padrões em dados sem rótulos; e aprendizado por reforço, o qual agentes aprendem a tomar decisões por meio de interações com o ambiente (Jiang et al., 2017). No campo da medicina, o aprendizado supervisionado é utilizado em tarefas no diagnóstico de doenças principalmente, em que algoritmos (como SVMs, redes neurais artificiais, etc.) são treinados para classificar imagens médicas ou sinais fisiológicos, como eletrocardiogramas (ECG), com alta precisão, especialmente no uso de redes neurais convolucionais (CNN) aplicadas à classificação de imagens dermatológicas (Esteve et al., 2017). Já o aprendizado não supervisionado é empregado para identificar padrões ocultos em grandes volumes de dados clínicos, como na descoberta de subgrupos de pacientes com características semelhantes, frequentemente utilizando técnicas como clustering ou análise de componentes principais (Miotto et al., 2016; Che et al., 2017). O aprendizado por reforço tem potencial para otimizar protocolos de tratamento, adaptando-se dinamicamente às respostas dos pacientes ao longo do tempo (Ali, 2022).

Historicamente, a aplicação da inteligência artificial (IA) na medicina teve início com os sistemas especialistas, como o MYCIN, desenvolvido na década de 1970 para diagnosticar infecções bacterianas e recomendar tratamentos com base em regras lógicas (Shortliffe, 1976). Porém, apenas nas últimas duas décadas que o ML começou a ser amplamente utilizado na prática clínica, muito por conta da digitalização de prontuários médicos, poder computacional maior e do uso grandes bases de dados públicos, como o MIMIC (Johnson et al., 2016) e o PhysioNet (Goldberger et al., 2000). Esses recursos trouxeram a possibilidade de treinar modelos em dados reais de pacientes, incentivando o avanço da medicina preditiva, personalizada e baseada em evidências. Além disso, o número de publicações científicas envolvendo ML em contextos clínicos vem crescendo, o que reflete o interesse global em integrar essas ferramentas aos sistemas de saúde (Jiang et al., 2017). Assim, o papel do ML na medicina evoluiu de uma abordagem experimental para uma solução prática cada vez mais

incorporada em plataformas diagnósticas, sistemas hospitalares e iniciativas de inovação em saúde digital.

Assim, nota-se como a aplicação do ML na medicina tem proporcionado uma melhora em termos de automação, escalabilidade e precisão diagnóstica. Pode-se dizer como exemplo que os modelos supervisionados são capazes de reduzir erros humanos ao identificar padrões complexos não perceptíveis por métodos estatísticos tradicionais, o que contribui para a diminuição de falsos positivos e negativos em exames diagnósticos. Existem trabalhos que demonstram como algoritmos de ML aplicados à interpretação de imagens médicas, como mamografias, podem superar a acurácia de especialistas humanos em determinadas tarefas, podendo reduzir a taxa de erros (Vedana et al., 2024). Além disso, o ML viabiliza a estratificação de risco e a personalização de tratamentos, otimizando recursos e promovendo condutas médicas mais direcionadas. Porém, ainda existem barreiras persistentes, como o viés algorítmico, que pode levar à perpetuação de desigualdades raciais e socioeconômicas nos cuidados de saúde, como demonstrado por estudos que identificaram menor precisão preditiva em populações sub-representadas (Obermeyer et al., 2019). Outra barreira que pode ser citada como exemplo é a interpretabilidade de modelos complexos, como redes neurais profundas, que operam como “caixas misteriosas” e dificultam a rastreabilidade das decisões tomadas, o que compromete sua aceitação por profissionais clínicos e reguladores (Jiang et al., 2017). Essas dificuldades reforçam a necessidade de técnicas de explicabilidade e validação rigorosas em ambientes clínicos reais, assegurando que as soluções baseadas em ML não apenas sejam eficazes, mas também transparentes, seguras e eticamente embasadas.

Nesse cenário, um dos campos que mais tem se beneficiado dos avanços em aprendizado de máquina é a cardiologia, visto que as doenças cardiovasculares possuem um grau elevado de complexidade e relevância clínica. Existem modelos supervisionados que têm sido aplicados com sucesso na análise de eletrocardiogramas (ECG), sendo capazes de detectar condições clínicas relevantes em pacientes assintomáticos com elevada acurácia; em um estudo, uma rede neural profunda treinada exclusivamente com sinais de ECG atingiu AUC de 0,87 na detecção de fibrilação atrial silenciosa, e até 0,90 com múltiplos exames por paciente (Attia et al., 2019). Não só isso, mas também técnicas de ML também vêm sendo utilizadas para interpretar imagens ecocardiográficas e distinguir etiologias de hipertrofia ventricular esquerda com desempenho superior ao de especialistas (Zhang et al., 2018). No contexto da insuficiência cardíaca, abordagens não supervisionadas permitem a identificação de fenótipos clínicos latentes, que auxiliam na personalização de condutas terapêuticas e prognóstico (Samson et al., 2016). Além disso, modelos baseados em árvores de decisão,

como o XGBoost, combinados com técnicas de explicabilidade como SHAP, vêm sendo empregados para prever mortalidade em pacientes com insuficiência cardíaca isquêmica, oferecendo interpretações clínicas valiosas sobre os determinantes individuais de risco (Wang, 2021). Esses avanços demonstram o potencial de como o ML na cardiologia moderna pode ser revolucionário ao fornecer ferramentas complementares que ampliam a capacidade diagnóstica e prognóstica da medicina tradicional.

2.2 Seattle Heart Failure Model (SHFM)

O Seattle Heart Failure Model é uma ferramenta prognóstica utilizada para estimar a sobrevida de pacientes com insuficiência cardíaca crônica que incorpora um conjunto de variáveis clínicas, laboratoriais e terapêuticas com o objetivo de estimar o risco de mortalidade. O modelo foi desenvolvido por Levy Wayne C. com base em uma coorte de 1.125 pacientes e posteriormente validado em cinco coortes externas, totalizando mais de 9.900 indivíduos e 17.000 anos-pessoa de acompanhamento, o que contribuiu para a robustez estatística de sua aplicação clínica (Levy et al., 2006). Entre as variáveis incluídas estão idade, sexo, classe funcional da New York Heart Association (NYHA), fração de ejeção do ventrículo esquerdo (FEVE), pressão arterial sistólica, uso de medicamentos como inibidores da enzima conversora de angiotensina (IECA), betabloqueadores, espironolactona, além de marcadores laboratoriais como hemoglobina, sódio sérico, colesterol total e linfócitos (Levy et al., 2006). O modelo permite a inserção de dispositivos como cardioversores desfibriladores implantáveis (CDI) e terapia de ressincronização cardíaca (TRC) como modificadores de risco também, sendo uma ferramenta útil na tomada de decisão terapêutica. Sua acurácia discriminativa foi demonstrada por meio de valores de AUC-ROC em torno de 0,73, mostrando uma boa performance preditiva (Levy et al., 2006). Existem mais estudos que validaram o SHFM em diferentes contextos clínicos, incluindo populações submetidas à TRC, nas quais o modelo mostrou capacidade significativa de estratificação de risco em múltiplos horizontes temporais (Regoli et al., 2013).

Apesar disso, pode-se dizer que existem limitações em populações específicas, como pacientes com idade avançada, em que o SHFM pode superestimar a sobrevida, levantando a necessidade de ajustes ou complementaridade com outras abordagens preditivas (Benbarkat et al., 2012). Além disso, é importante destacar que o SHFM é um modelo desenvolvido há quase duas décadas, baseado em coortes de pacientes tratados com estratégias terapêuticas vigentes na época. Desde então, avanços ocorreram no manejo da insuficiência cardíaca,

incluindo o surgimento de novas classes de medicamentos, melhorias no acesso a dispositivos e um aumento na expectativa de vida da população idosa. Esse cenário contribui para uma redução da acuracidade do SHFM, principalmente em pacientes com idade avançada, as quais o modelo tende a superestimar a sobrevida, como evidenciado em diversos estudos recentes (Benbarkat et al., 2012; Shiraishi et al., 2018). Tal limitação reforça a necessidade de desenvolvimento de abordagens preditivas mais atualizadas, capazes de incorporar as características demográficas e terapêuticas da população contemporânea. Ainda assim, o SHFM permanece uma das ferramentas mais consolidadas na predição de mortalidade em IC crônica, sendo amplamente incorporado em diretrizes e práticas clínicas internacionais.

A estrutura do SHFM fundamenta-se em um modelo de regressão multivariada baseado em riscos proporcionais de Cox, o qual cada variável inserida contribui com um coeficiente ponderado na estimativa final de sobrevida do paciente (Levy et al., 2006). A função de risco é expressa como:

$$h(t | X) = h_0(t) \cdot \exp(\beta_{1x1} + \beta_{2x2} + \dots + \beta_{n xn}) \quad (1)$$

Em que:

- $h(t | X)$ é o risco instantâneo de morte no tempo t , dado o vetor de covariáveis x .
- $h_0(t)$ é a função de risco basal (risco quando todas as variáveis são zero).
- i são os coeficientes ajustados para cada variável x_i , determinados a partir da análise da coorte original.

Dessa forma, a probabilidade de sobrevida pode ser obtida por meio de:

$$S(t | X) = S_0(t)^{\exp(\beta_{1x1} + \beta_{2x2} + \dots + \beta_{n xn})} \quad (2)$$

Em que:

- $S(t | X)$ Função de sobrevida basal no tempo t , estimada a partir da coorte original.

Originalmente, esses coeficientes foram derivados de análises estatísticas sobre uma coorte de pacientes ambulatoriais com insuficiência cardíaca crônica, o que possibilitou a quantificação do impacto relativo de cada fator clínico ou terapêutico sobre o risco de mortalidade. O escore é calculado por meio da soma ponderada dessas variáveis, em que a transformação matemática gera uma função de risco cumulativo, que é posteriormente

convertida em probabilidade de sobrevida em 1, 2, 3, 4 ou 5 anos. Um diferencial do modelo é sua capacidade de incorporar modificações terapêuticas prospectivas na estimativa de sobrevida. Ou seja, o escore pode ser recalculado para simular o impacto de intervenções clínicas hipotéticas, como a introdução de novos medicamentos ou dispositivos (Levy et al., 2006). Assim, o SHFM pode ser utilizado não apenas para estratificação prognóstica, mas também como ferramenta de apoio à decisão terapêutica. Além disso, o modelo pode ser aplicado com ou sem os dados de marcadores laboratoriais, mantendo um desempenho aceitável mesmo em cenários clínicos com dados limitados, o que amplia sua aplicabilidade em contextos com recursos restritos (Regoli et al., 2013). A funcionalidade computacional do modelo foi implementada em versões digitais e calculadoras clínicas, sendo integrada a sistemas hospitalares e protocolos de manejo em diversas instituições de saúde (University of Washington, 2023).

Existem diversos estudos que avaliam a performance do SHFM em diferentes contextos clínicos, demonstrando tanto suas fortalezas quanto limitações. Em uma coorte de 427 pacientes submetidos à terapia de ressincronização cardíaca (TRC), o SHFM demonstrou boa discriminação, com áreas sob a curva ROC de 0,74 no primeiro ano, 0,79 no segundo e 0,76 no quinto ano, o que indica uma acurácia preditiva consistente ao longo do tempo (Clemens et al., 2012). De forma semelhante, Regoli et al (2013) validou o modelo em 1.139 pacientes com TRC com o trabalho “Validation of Seattle Heart Failure Model for mortality risk prediction in patients treated with cardiac resynchronization therapy”, observando AUCs entre 0,64 e 0,68 nos anos subsequentes. Foi notado uma tendência do modelo em superestimar a sobrevida, principalmente em pacientes de alto risco. Já em populações com idade mais elevada (especialmente idades maior que 80 anos) o SHFM demonstrou limitações significativas. Benbarkat et al (2012) no trabalho “Application of the Seattle Heart Failure Model in Patients >80 Years of Age Enrolled in a Tertiary Care Heart Failure Clinic” relatou que, em uma coorte de 261 pacientes com média de idade de 85 anos, o modelo superestimou a sobrevida em aproximadamente 4,8 anos, sugerindo que o SHFM pode não ser adequado para estimativas precisas de prognóstico nessa faixa etária avançada. Além disso, em uma população japonesa de 2.470 pacientes com insuficiência cardíaca aguda, o SHFM apresentou desempenho satisfatório apenas após recalibração, mais voltado para pacientes com fração de ejeção reduzida, mostrando que existe uma necessidade de ajuste do modelo para diferentes populações (Shiraishi et al, 2018). Esses estudos mostram que, embora o SHFM seja uma ferramenta valiosa na estratificação de risco em insuficiência cardíaca, sua aplicabilidade

pode ser limitada em subgrupos específicos, exigindo adaptações ou complementações com outros modelos prognósticos.

A crescente aplicação de técnicas de aprendizado de máquina na predição de desfechos em insuficiência cardíaca tem levado à comparação entre modelos tradicionais, como o SHFM, e abordagens baseadas em inteligência artificial. Diversos estudos demonstraram que modelos de ML podem superar o SHFM em termos de acurácia preditiva. Por exemplo, um estudo desenvolvido no Japão utilizou dados administrativos de 10.175 pacientes com IC para construir modelos preditivos baseados em ML. Os modelos resultantes, incluindo o SMART-HF, apresentaram estatísticas superiores às do SHFM (SMART-HF: 0,765; SHFM: 0,713), indicando melhor desempenho na predição de mortalidade em um ano (Tohyama et al., 2021). Além disso, uma pesquisa conduzida na Mayo Clinic avaliou o desempenho do SHFM utilizando registros eletrônicos de saúde e desenvolveu um modelo baseado em ML que incorporou comorbidades adicionais, resultando em uma melhoria de 11% na AUC em comparação com o SHFM (Panahiazar et al., 2015). Outro estudo demonstrou que um modelo de ML aplicado a dados de monitoramento remoto foi capaz de prever hospitalizações por IC com uma AUC de 0,855, superando significativamente o algoritmo convencional, que apresentou uma AUC de 0,727 (Koehler et al., 2023). Esses resultados sugerem que modelos baseados em ML podem oferecer vantagens sobre o SHFM na predição de desfechos em IC, especialmente quando incorporam variáveis adicionais e dados em tempo real (Koehler et al., 2023). Ainda assim, o SHFM mantém sua relevância na prática clínica devido à sua simplicidade operacional e alta interpretabilidade, características que favorecem sua aceitação entre médicos não especialistas e facilitam sua aplicação em contextos com infraestrutura tecnológica limitada. Em muitos cenários, o modelo é utilizado como baseline interpretativo ou como elemento complementar em frameworks baseados em ML, o que reforça sua importância como ferramenta consolidada de apoio à decisão médica.

2.3 Modelos de Aprendizado de Máquina

Como dito anteriormente, a aplicação de técnicas de aprendizado de máquina na área da saúde tem se mostrado promissora na predição de desfechos clínicos, incluindo a mortalidade por insuficiência cardíaca. Assim, diversos algoritmos têm sido empregados com o objetivo de melhorar a acurácia preditiva e superar as limitações dos modelos tradicionais, como o SHFM. Esta subseção apresenta uma revisão dos principais modelos utilizados neste trabalho (Regressão Logística, Random Forest, SVM, XGBoost e Redes Neurais com

arquitetura MLP integrada a Transformer) destacando suas bases teóricas e aplicações reportadas na literatura. Busca-se, assim, contextualizar o uso dessas técnicas em cenários clínicos, discutindo suas potencialidades, limitações e resultados observados em estudos relacionados à estratificação de risco e predição de mortalidade em pacientes com IC.

2.3.1 Regressão Logística

Um dos modelos que “disputa” mais diretamente com o SHFM é a Regressão Logística. Trata-se de uma técnica estatística amplamente utilizada para modelar a probabilidade de ocorrência de um evento binário, como a presença ou ausência de uma doença, com base em variáveis independentes. A Regressão Logística estima os coeficientes que relacionam as variáveis preditoras ao logaritmo das chances do evento, permitindo a interpretação dos efeitos de cada variável sobre a probabilidade do desfecho (Hosmer et al., 2013). Ou seja, esse algoritmo modela a probabilidade de um evento binário $y \in \{0,1\}$ com base em variáveis preditoras, com a função logística mostrada na expressão 3:

$$P(y = 1 | X) = \frac{1}{1 + e^{-(\beta_{0x0} + \beta_{1x1} + \dots + \beta_{n xn})}} \quad (3)$$

Em que o termo $\beta_{0x0} + \sum_{i=1}^n \beta_{ixi}$ representa o logito na expressão 4:

$$\text{logit}(P) = \ln\left(\frac{P}{1-P}\right) \quad (4)$$

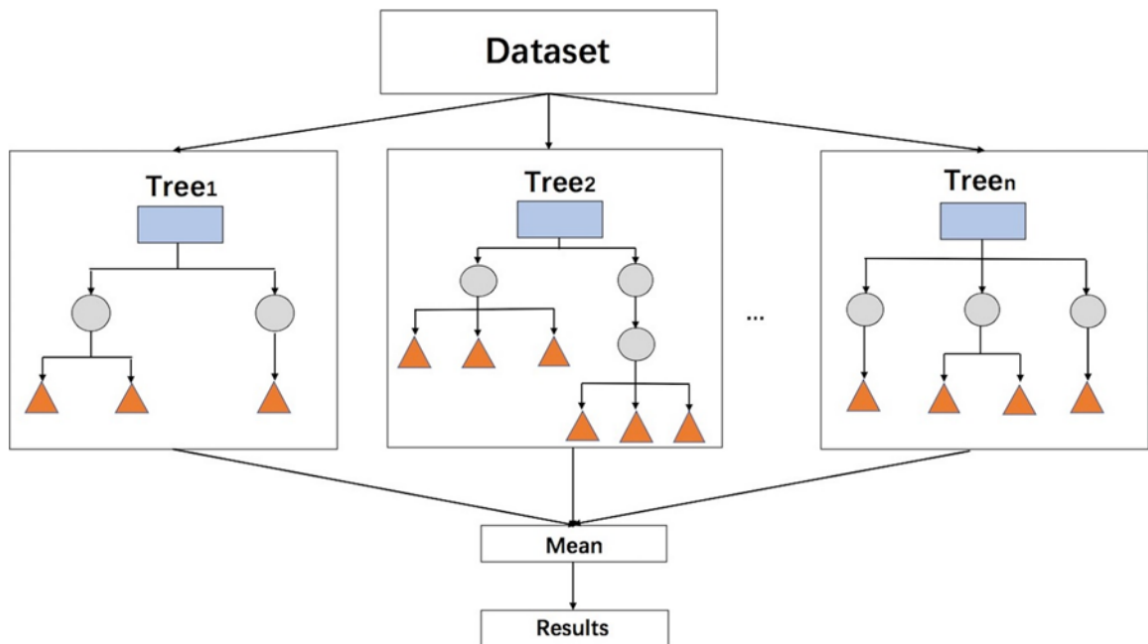
Sua simplicidade, interpretabilidade e eficiência computacional tornam esta técnica uma escolha frequente em contextos clínicos, especialmente quando se busca compreender a influência de fatores de risco específicos. O estudo “Logistic regression technique for prediction of cardiovascular disease” conduzido por Ambrish et al. (2022), a Regressão Logística foi aplicada ao conhecido conjunto de dados da UCI sobre doenças cardíacas, com técnicas de pré-processamento e seleção de características, resultando em uma acurácia de aproximadamente 87%, evidenciando seu desempenho robusto. De modo semelhante,

Anshori et al. (2022) em “Predicting Heart Disease using Logistic Regression” utilizaram a Regressão Logística para prever o risco de doenças cardiovasculares com base em registros clínicos, obtendo uma acurácia de 81,35% e AUC de 0,8936, demonstrando sua eficácia em contextos médicos reais. Esses trabalhos reforçam que, apesar de existir algoritmos mais complexos, a Regressão Logística permanece como uma ferramenta confiável, especialmente em aplicações que exigem transparência na tomada de decisões clínicas.

2.3.2 Random Forest

Outra técnica de ML de grande relevância nessa área é o Random Forest. Trata-se de uma técnica de aprendizado de máquina baseada em um conjunto de árvores de decisão, as quais múltiplos classificadores são treinados sobre subconjuntos aleatórios do conjunto de dados original e cujas previsões são combinadas por votação para produzir a saída final na qual múltiplos classificadores são treinados sobre subconjuntos aleatórios do conjunto de dados original e cujas previsões são combinadas por votação para produzir a saída final (Breiman, 2001). Essa abordagem é um exemplo da técnica conhecida como *ensemble*, que consiste na combinação de múltiplos modelos individuais (também chamados de "modelos fracos") para formar um modelo mais robusto e preciso. No caso do Random Forest, utiliza-se o método de *bagging* (bootstrap aggregating), em que várias árvores são treinadas de forma independente em diferentes subconjuntos de dados, e suas previsões são agregadas para gerar uma predição final mais estável. A Figura 1 representa esquematicamente este modelo.

Figura 1- Representação esquemática Random Forest.



Fonte: ZOU 2021

Além disso, estudos recentes demonstram a eficácia do Random Forest na predição de desfechos clínicos em pacientes com insuficiência cardíaca. Por exemplo, a revisão “Evaluation of ML methods for prediction of heart failure mortality and readmission: meta-analysis” sistemática conduzida por Hajishah et al. (2025) avaliou diversos modelos de aprendizado de máquina na predição de mortalidade e readmissão hospitalar em pacientes com insuficiência cardíaca, destacando o Random Forest como um dos algoritmos com desempenho robusto, com AUC média de 0,763 para predição de mortalidade e 0,736 para readmissão, superando modelos tradicionais como a regressão logística. Outro estudo chamado “Study of cardiovascular disease prediction model based on random forest in eastern China” conduzido por Yang et al. (2020) desenvolveu um modelo preditivo baseado em Random Forest para avaliar o risco de doenças cardiovasculares em uma coorte de mais de 25.000 indivíduos na China, alcançando uma AUC de 0,787, superando modelos tradicionais como a regressão logística.

2.3.3 Support Vector Machine

Outra técnica empregada em diversos estudos na predição de desfechos clínicos em pacientes com insuficiência cardíaca é a SVM. Trata-se de um algoritmo de aprendizado

supervisionado que busca identificar o hiperplano ótimo que separa classes distintas no espaço de características, maximizando a margem entre os dados de diferentes categorias (Cortes e Vapnik, 1995). Assim, o objetivo do SVM é encontrar o hiperplano $w \cdot x + b = 0$, que maximiza essa margem. A expressão 5 representa a formulação primária:

$$\frac{\min}{w, b} \frac{1}{2} ||w||^2 \text{ sujeito a } y_i(w \cdot x_i + b) \geq 1, \forall i \quad (5)$$

Porém, para dados não linearmente separáveis, usa-se a versão suavizada com penalidade C e variáveis de folga ξ_i , que pode ser visto na expressão 6:

$$\frac{\min}{w, b, \xi} \frac{1}{2} ||w||^2 + C \sum_{i=1}^n \xi_i \text{ sujeito a } y_i(w \cdot \phi(x_i) + b) \geq 1 - \xi_i \quad (6)$$

Em que $\phi(\cdot)$ é a função de mapeamento para o espaço de alta dimensionalidade.

Sua capacidade de lidar com conjuntos de dados de alta dimensionalidade e sua eficácia em contextos não lineares tornam a SVM uma ferramenta valiosa na área médica. Existem estudos que destacam a eficácia da SVM na predição de desfechos em insuficiência cardíaca. Por exemplo, o estudo “Heart Failure Prediction Using Support Vector Machine” de Ghasemi et al. (2025) desenvolveu um modelo preditivo baseado em SVM utilizando dados clínicos de 299 pacientes, incluindo variáveis como idade, fração de ejeção, níveis de sódio sérico e comorbidades como diabetes e hipertensão. O modelo alcançou uma acurácia de 99,1% nos testes, demonstrando sua robustez e aplicabilidade clínica. Além disso, Herrera Salazar et al. (2025) realizou uma comparação entre SVM e redes neurais artificiais na predição de insuficiência cardíaca, utilizando o conjunto de dados Heart Failure Clinical Records disponível na plataforma Kaggle no estudo “Predicting Heart Failure: A Comparative Approach between Artificial Neural Networks and Support Vector Machines”. Os resultados indicaram que a SVM obteve uma acurácia de 79%, superando as redes neurais, que alcançaram 77%. Isso reforça o potencial da SVM como uma ferramenta eficaz na estratificação de risco e na tomada de decisões clínicas em cardiologia.

2.3.4 XGBoost

Existe também o XGBoost (Extreme Gradient Boosting), que se trata de um algoritmo de aprendizado supervisionado baseado em árvores de decisão, que utiliza o método de *boosting* para combinar múltiplos modelos fracos e formar um modelo forte, otimizando a função de perda por meio de gradiente descendente (Chen e Guestrin, 2016). Assim, o XGBoost constrói modelos sequenciais $f(t)$, adicionando árvores f_t que corrigem os erros dos modelos anteriores. A expressão 7 mostra a função da previsão:

$$\hat{y}_i = \sum_{t=1}^T f_t(x_i), f_t \in F \quad (7)$$

Assim, a função objetivo combinando erro e regularização está representada na expressão 8:

$$\mathcal{L} = \sum_{i=1}^n l(y_i, \hat{y}_i) + \sum_{t=1}^T \Omega(f_t) \quad (8)$$

Em que, pela expressão 9, penaliza a complexidade da árvore:

$$\Omega(f_t) = \gamma T + \frac{1}{2} \lambda \sum_{j=1}^J \omega_j^2 \quad (10)$$

As árvores são treinadas com base na descida do gradiente da função de perda. A cada iteração, uma nova árvore é ajustada aos resíduos negativos do gradiente, que pode ser visto na expressão 11:

$$g_t = \frac{\delta l(y_i, \hat{y}_i^{(t-1)})}{\delta \hat{y}_i^{(t-1)}} \quad (11)$$

Alguns estudos demonstram a eficácia do XGBoost, como por exemplo “Optimizing Mortality Prediction for ICU Heart Failure Patients: Leveraging XGBoost and Advanced ML with the MIMIC-III Database” de Ashrafi et al. (2024), o qual aplicou o XGBoost para prever a mortalidade em pacientes com insuficiência cardíaca internados em UTI, utilizando o banco de dados MIMIC-III (Johnson et al., 2016). Após rigorosa seleção de características e otimização de hiperparâmetros, o modelo atingiu uma AUC de 0,9228, destacando-se como uma ferramenta promissora para suporte à decisão clínica. Existe também “Heart failure prediction: Exploratory analysis and modeling with XGBoost and deep neural networks” Gudiño-Ochoa et al. (2023), o qual desenvolveu um modelo preditivo baseado em XGBoost

para avaliar o risco de insuficiência cardíaca, utilizando dados clínicos de 918 pacientes de diferentes países. O modelo alcançou uma acurácia de 88,04%, demonstrando sua robustez e aplicabilidade clínica.

2.3.5 Redes Neurais e Transformer

Por fim, outro modelo utilizado neste trabalho e que se mostra robusto no campo da medicina são as redes neurais artificiais (RNAs). As RNAs são modelos computacionais inspirados na estrutura e funcionamento do cérebro humano, compostos por camadas de unidades interconectadas (neurônios) que processam informações por meio de pesos ajustáveis (Zell, 2003). Essas redes são capazes de modelar relações complexas entre variáveis, tornando-se ferramentas valiosas na predição de desfechos clínicos, como a insuficiência cardíaca. A arquitetura mais comum de RNA utilizada em tarefas de classificação é a perceptron multicamada (MLP – Multi-Layer Perceptron). Uma MLP é composta por uma camada de entrada, uma ou mais camadas ocultas e uma camada de saída. Cada neurônio aplica uma transformação linear dos dados de entrada, seguida por uma função de ativação não linear. A operação de um neurônio pode ser observada nas expressões 12 e 13 a seguir:

$$z = w^T x + b \quad (12)$$

$$a = f(z) \quad (13)$$

Em que:

- x representa o vetor de entrada
- w é o vetor de pesos
- b é o viés
- $f(z)$ é uma função de ativação

O treinamento da MLP é realizado pelo algoritmo de retropropagação (backpropagation), que utiliza gradiente descendente para atualizar os pesos com base no erro visto na expressão 14 a seguir:

$$w := w - \eta \cdot \frac{\delta \mathcal{L}}{\delta w} \quad (14)$$

Uma arquitetura emergente no campo das RNAs é o Transformer, introduzido por Vaswani et al. (2017). Diferentemente das redes recorrentes, os Transformers utilizam mecanismos de atenção para modelar dependências entre elementos de uma sequência, o que possibilita o processamento paralelo e eficiente de dados sequenciais. No contexto da predição de insuficiência cardíaca, existem estudos que têm explorado a integração de Transformers com outras arquiteturas de redes neurais para aprimorar a acurácia dos modelos. Por exemplo, o estudo “An explainable Transformer-based *deep learning* model for the prediction of incident heart failure” conduzido por Rao et al. (2021) desenvolveu um modelo baseado em Transformer para prever a incidência de insuficiência cardíaca, utilizando registros eletrônicos de saúde de 100.071 pacientes no Reino Unido. O modelo alcançou uma AUC de 0,93 na validação externa, superando modelos tradicionais e demonstrando sua eficácia na predição de desfechos clínicos. Além disso, o estudo “CardioTabNet: A Novel Hybrid Transformer Model for Heart Disease Prediction using Tabular Medical Data” de Islam et al. (2025) propôs o CardioTabNet, um modelo híbrido que combina Transformers com dados tabulares para prever doenças cardíacas. Utilizando um conjunto de dados com 1.190 instâncias, o modelo alcançou uma acurácia de 94,1% e uma AUC de 95,0%, destacando-se como uma ferramenta promissora para suporte à decisão clínica.

2.4 Métricas

Para avaliar o desempenho dos modelos de classificação utilizados neste estudo, foram adotadas as métricas de precisão, *recall*, F1-score e AUC-ROC. Cada uma dessas métricas fornece uma perspectiva complementar sobre a performance dos modelos em relação à tarefa de predição binária (morte ou sobrevivência).

A precisão indica a proporção de previsões positivas que estavam corretas. Em outras palavras, mede quantos dos pacientes que o modelo classificou como de alto risco realmente vieram a óbito. Essa métrica é particularmente útil quando o custo de falsos positivos é elevado, como em diagnósticos médicos, onde um alarme falso pode levar a tratamentos desnecessários (Saito & Rehmsmeier, 2015).

O *recall* (ou sensibilidade) representa a capacidade do modelo de identificar corretamente os pacientes que de fato morreram; ou seja, mostra quantos dos óbitos reais foram corretamente detectados. Alta sensibilidade é fundamental em contextos clínicos, como a predição de mortalidade, onde a falha em detectar um caso crítico pode ter consequências graves (Davis & Goadrich, 2006).

O F1-score é a média harmônica entre precisão e *recall*, servindo como uma métrica equilibrada quando se deseja considerar tanto a taxa de acertos quanto a taxa de omissões do modelo. Essa métrica é especialmente útil em cenários de classes desbalanceadas, como neste trabalho, pois penaliza fortemente os extremos em que uma das métricas é muito baixa (Powers, 2011).

Por fim, a AUC-ROC (Área sob a Curva ROC) representa a capacidade discriminativa global do modelo. A curva ROC relaciona a taxa de verdadeiros positivos com a taxa de falsos positivos para diferentes limiares de decisão. A AUC, por sua vez, é a área sob essa curva e varia entre 0,5 (modelo aleatório) e 1,0 (classificação perfeita). Quanto maior o valor da AUC, melhor o desempenho do modelo em distinguir entre os pacientes que sobreviveram e os que faleceram (Fawcett, 2006).

A combinação dessas métricas permite uma avaliação abrangente da eficácia dos modelos testados, considerando tanto a qualidade das classificações quanto a capacidade de generalização dos algoritmos em diferentes cenários clínicos.

3 MATERIAIS E MÉTODOS

Esta seção abordará com mais detalhes a respeito de como este trabalho foi implementado. Será abordado os objetivos do trabalho, bem como as etapas de execução e exploração dos dados, além da explicação da base de dados utilizada.

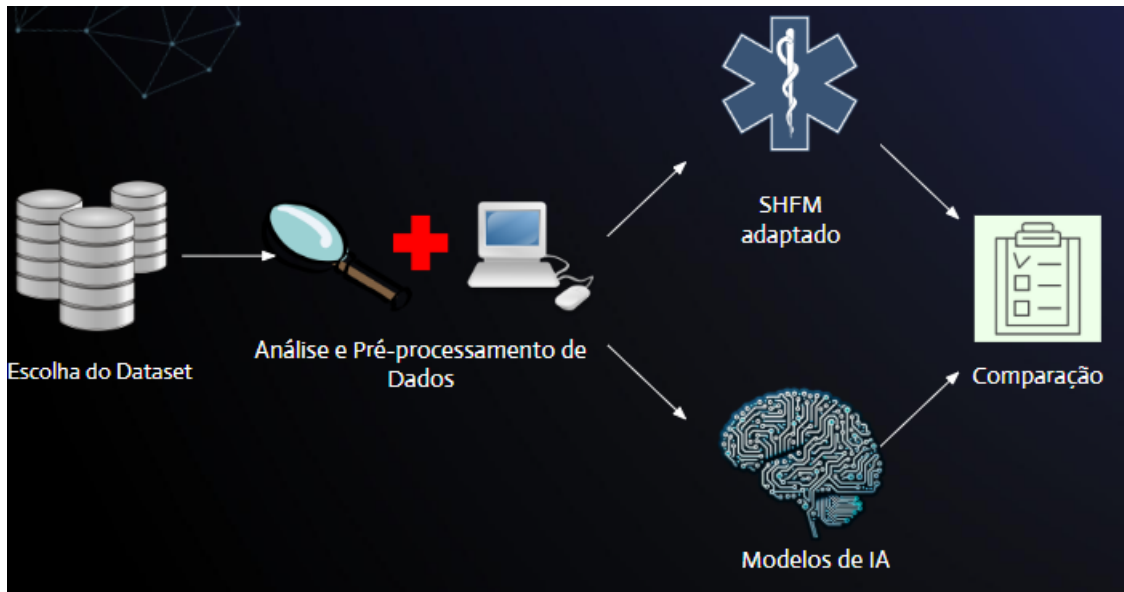
3.1 Abordagem Proposta

Este estudo propõe o uso de técnicas de ML para prever a mortalidade cardíaca de pacientes com insuficiência cardíaca, comparando seu desempenho com o modelo tradicional SHFM, amplamente utilizado na prática clínica. Para tal, foram utilizados dados do estudo MUSIC (MUerte Súbita en Insuficiencia Cardiaca), cuja descrição detalhada encontra-se na seção 3.3.

Foi realizada uma análise exploratória inicial, em que a eficácia do SHFM como *benchmark* também foi avaliada. Considerando que o SHFM utiliza variáveis clínicas para calcular a probabilidade de sobrevivência de pacientes com insuficiência cardíaca, sendo avaliado essencialmente por meio da curva ROC e da AUC, e dado que este modelo estatístico é considerado limitado em alguns contextos, optou-se por ajustar modelos de aprendizagem supervisionada. Estes modelos foram comparados de forma direta e equitativa ao SHFM, utilizando métricas rigorosas e apropriadas para avaliação de desempenho, uma vez que evidências em diversos contextos clínicos indicam que algoritmos de ML podem melhorar a estratificação de risco.

O *pipeline* adotado no trabalho compreende as seguintes etapas: análise exploratória inicial dos dados; pré-processamento, incluindo o tratamento de valores; implementação do SHFM para cálculo do escore de risco tradicional; treinamento de diferentes modelos de ML sobre os dados estruturados; e, finalmente, avaliação comparativa entre os métodos. A Figura 2 apresenta uma visão esquemática simplificada do *pipeline* metodológico proposto.

Figura 2- Esquemático do *Pipeline*.



Fonte: Autoria própria.

O *pipeline* adotado no trabalho compreende as seguintes etapas: análise exploratória inicial dos

3.2 Configuração para os experimentos

Os experimentos realizados neste trabalho foram desenvolvidos e executados utilizando a plataforma Google Colab, que oferece um ambiente de computação em nuvem gratuito, voltado principalmente para aplicações de programação utilizando a linguagem Python, R ou Julia. A linguagem de programação utilizada foi o Python 3, conforme o ambiente de execução configurado na plataforma. A execução dos códigos foi realizada majoritariamente com CPU virtualizada fornecida pelo Colab, sendo disponibilizados aproximadamente 2 vCPUs e cerca de 12 GB de memória RAM. A utilização de GPU não foi necessária para a realização dos experimentos, uma vez que o processamento dos dados e o treinamento dos modelos foram suficientemente rápidos em ambiente de CPU.

O desenvolvimento e a execução dos experimentos foram realizados com suporte de diversas bibliotecas da linguagem Python:

- **pandas e numpy:** para manipulação de dados e operações matemáticas.
- **matplotlib e seaborn:** para geração de gráficos e visualizações dos dados.

- **scikit-learn:** para tarefas de pré-processamento, treinamento e avaliação de modelos de aprendizado de máquina, além de estratégias de validação e ajuste de hiperparâmetros.
- **imbalanced-learn (imblearn):** para balanceamento de classes utilizando o método SMOTE (*Synthetic Minority Over-sampling Technique*).
- **tensorflow:** para a construção e treinamento de redes neurais profundas e modelos baseados na arquitetura Transformer.
- **xgboost:** para a implementação do algoritmo Extreme Gradient Boosting (XGBoost).

As versões específicas das bibliotecas utilizadas não foram explicitamente controladas, seguindo as versões pré-instaladas no ambiente do Google Colab no período de execução do projeto. A Figura 3 apresenta a configuração selecionada para o ambiente de execução no Google Colab, destacando a utilização de CPU como acelerador de hardware.

Figura 3- Configuração ambiente de execução do Google Colab.



Fonte: Autoria própria.

3.3 Base de Dados e Pré-processamento

O conjunto de dados utilizado neste trabalho foi obtido a partir do estudo MUSIC (MUerte Súbita en Insuficiencia Cardíaca), um estudo prospectivo, multicêntrico e

longitudinal que visou investigar preditores de mortalidade cardíaca, em especial morte súbita cardíaca (SCD) e morte por falência de bomba (PFD), em pacientes com insuficiência cardíaca crônica. O estudo foi conduzido em oito hospitais universitários da Espanha, com o recrutamento de 992 pacientes ambulatoriais entre abril de 2003 e dezembro de 2004, acompanhados por um período mediano de 44 meses até novembro de 2008. O MUSIC fornece uma base rica de dados clínicos, laboratoriais e de exames complementares (como ecocardiografia e eletrocardiografia), podendo ser utilizado para o desenvolvimento e validação de modelos prognósticos em cardiologia.

O banco de dados principal utilizado reúne variáveis clínicas (como idade, sexo, classe funcional da NYHA), laboratoriais (como níveis de creatinina, colesterol LDL, troponina), ecocardiográficas (como fração de ejeção do ventrículo esquerdo – LVEF, e diâmetros ventriculares), eletrocardiográficas (como ritmos de Holter e ECG de repouso) e informações sobre dispositivos implantáveis. As variáveis categóricas são codificadas por padrão, enquanto a definição completa das variáveis e suas unidades de medida foram fornecidas em um arquivo csv. A diversidade de variáveis disponíveis, tanto contínuas quanto categóricas, faz com que este *dataset* seja apropriado para a aplicação de técnicas de ML, que se beneficiam da riqueza e heterogeneidade de dados para a modelagem de padrões complexos.

A primeira etapa do pré-processamento consistiu em uma análise exploratória inicial, destinada a compreender a estrutura geral do *dataset*. Foram avaliadas a quantidade de registros e colunas, a presença de valores ausentes, a distribuição das variáveis numéricas e a frequência das variáveis categóricas. Com base nessa análise preliminar, procedeu-se à remoção de colunas consideradas irrelevantes para o objetivo de predição. Entre essas colunas estavam identificadores únicos (ID do Paciente), informações administrativas (Saída do Estudo), e variáveis que indicavam apenas a disponibilidade de exames (Holter disponível, ECG de alta resolução disponível), que não adicionariam valor preditivo ao modelo.

Em seguida, foi realizado o tratamento sistemático de valores ausentes. Três abordagens distintas foram aplicadas de acordo com a natureza e a distribuição esperada de cada variável, com estratificação por faixas etárias:

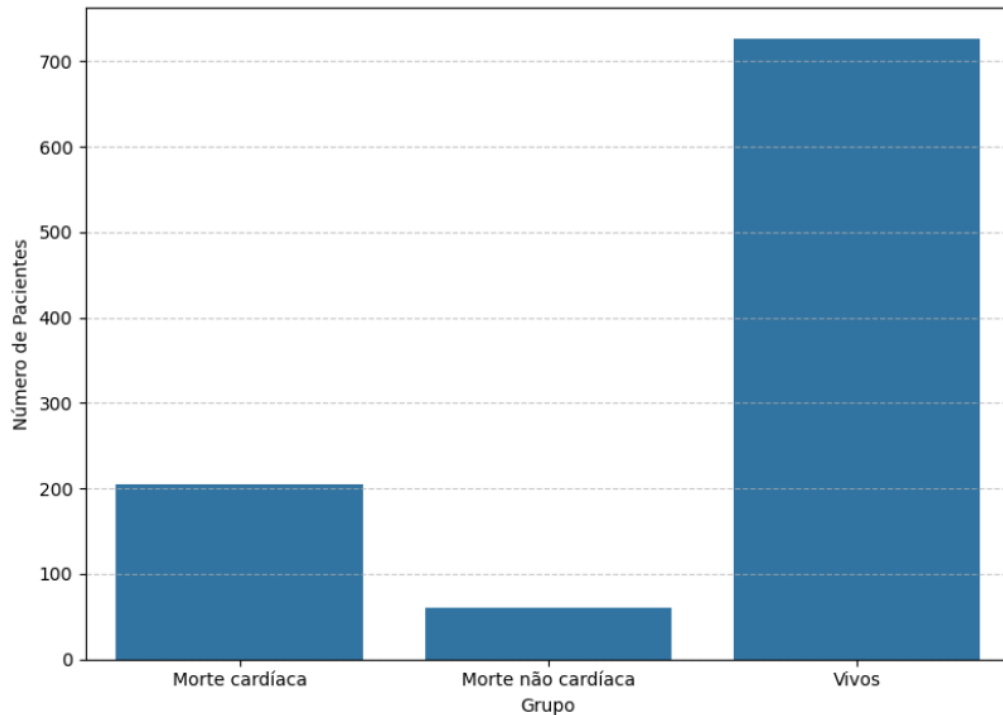
- **Preenchimento pela mediana:** adotado para variáveis susceptíveis à presença de outliers ou distribuição assimétrica, como Troponin (ng/mL), LVEF (%), Albumin (g/L) e medidas estruturais cardíacas.
- **Preenchimento pela média:** utilizado para variáveis cuja distribuição tende a ser mais simétrica na população geral, como Sódio (mEq/L), Creatinina (?mol/L) e Glucose (mmol/L).

- **Preenchimento por zero:** aplicado para variáveis cuja ausência de informação poderia ser interpretada como ausência do evento ou característica clínica, como Extrassístole ventricular, Taquiarritmia supraventricular paroxística e indicadores de anormalidades na condução elétrica cardíaca.

Além disso, foi realizada uma análise de redundância e multicolinearidade entre as variáveis, culminando na exclusão de atributos altamente correlacionados ou com informação sobreposta, como Troponina normalizada, RR mínimo (ms), RR máximo (ms), T3, T4 e TSH. Variáveis relacionadas a medicações específicas (ex.: Betabloqueadores, Inibidores da ECA) também foram removidas para evitar vazamento de informação ("*data leakage*"), uma vez que sua prescrição poderia refletir avaliações clínicas subjetivas realizadas após a inclusão no estudo.

Para garantir a robustez da análise, foi realizada a remoção de pacientes que faleceram por causas não cardíacas. Esta decisão teve como objetivo assegurar que o problema de classificação fosse restrito à mortalidade de origem cardíaca, tanto para a aplicação do SHFM quanto para o treinamento dos modelos de ML, evitando enviesamento na avaliação de desempenho. Após a exclusão desses pacientes, a variável original Causa do óbito foi transformada em uma nova variável binária, denominada Morte Cardíaca, em que o valor "1" indica morte de origem cardíaca (SCD ou PFD) e "0" indica sobrevivência ao longo do período de seguimento.

Durante a análise exploratória, observou-se um desbalanceamento na distribuição da variável alvo, com uma proporção menor de pacientes que evoluíram para óbito cardíaco em relação aos sobreviventes. A figura 4 ilustra como está o balanceamento do dataset. Para lidar com esse desbalanceamento de classes e melhorar a capacidade dos modelos de ML em identificar padrões associados à mortalidade, aplicou-se a técnica de *Synthetic Minority Over-sampling Technique* (SMOTE) exclusivamente no conjunto de treinamento, preservando a distribuição natural dos dados nos conjuntos de validação e teste.

Figura 4 - Distribuição *dataset* MUSIC.

Fonte: Autoria própria.

Nas etapas de preparação para os modelos de Redes Neurais Artificiais (MLP) e Transformer, foram aplicadas transformações adicionais: variáveis numéricas foram padronizadas utilizando o método de *StandardScaler*, enquanto variáveis categóricas foram codificadas por meio de *OneHotEncoder*, consolidando o *pipeline* de pré-processamento com o uso do *ColumnTransformer*. É importante ressaltar que o *dataset* foi dividido em três subconjuntos: treinamento (60%), validação (20%) e teste (20%), utilizando divisão estratificada para preservar a proporção da variável alvo entre os conjuntos.

Todo o processo de pré-processamento foi conduzido de maneira sistemática e replicável, com o objetivo de garantir a integridade dos dados, minimizar riscos de viés, preservar a generalização dos modelos e possibilitar a reprodutibilidade dos experimentos. Cabe ressaltar que todos os dados utilizados já se encontravam previamente anonimizados de acordo com as diretrizes da General Data Protection Regulation (GDPR), conforme estabelecido pelo repositório PhysioNet.

3.4 Implementação e Adaptação SHFM

Para servir como *baseline* comparativo, implementou-se o cálculo de uma versão adaptada do SHFM, um dos modelos tradicionais mais utilizados na predição de sobrevida em

pacientes com insuficiência cardíaca. O SHFM adaptado foi calculado para estimar o risco de mortalidade em um horizonte temporal de 5 anos, compatível com o período mediano de acompanhamento do estudo MUSIC, de 44 meses.

O escore do SHFM adaptado foi calculado utilizando os coeficientes originais da literatura, ajustados para variáveis padronizadas quando aplicável (z-score para pressão arterial sistólica, fração de ejeção do ventrículo esquerdo, sódio, hemoglobina e colesterol total). Ao todo, das 24 variáveis que compõem o SHFM original, 15 estavam disponíveis no dataset MUSIC e foram utilizadas no cálculo da versão adaptada. Essas incluíram variáveis demográficas e clínicas (idade, sexo, classe funcional da NYHA, etiologia isquêmica, fração de ejeção do ventrículo esquerdo, pressão sistólica e peso), laboratoriais (sódio, hemoglobina e colesterol total) e terapêuticas (uso de inibidor da ECA, betabloqueadores, bloqueadores de receptor de angiotensina, espironolactona e digoxina). As 9 variáveis que não puderam ser incluídas, devido à ausência de dados ou falta de especificidade na base, foram: percentual de linfócitos, ácido úrico, dose de diurético, uso de alopurinol e as variáveis relacionadas a dispositivos implantáveis, como marca-passo, cardioversor desfibrilador implantável (CDI), terapia de ressincronização cardíaca (TRC) e dispositivo de assistência ventricular. Essa limitação na disponibilidade de variáveis é um dos principais fatores que caracterizaram a adaptação metodológica do SHFM neste estudo.

O valor do escore individual de cada paciente foi convertido em uma estimativa de risco de morte em 5 anos utilizando uma função de Gompertz adaptada, conforme a expressão 15:

$$Risco(5\text{ anos}) = 1 - \exp(-\exp(SHFM\text{ Score})t) \quad (15)$$

Essa abordagem representou uma aproximação prática do modelo original do SHFM, considerando que a função de sobrevivência basal derivada da coorte original não estava disponível. Apesar disso, buscou-se manter a fidelidade conceitual ao modelo original, assegurando uma comparação robusta com os modelos de aprendizado de máquina testados.

Após a aplicação do escore, pacientes com morte não cardíaca foram excluídos da análise para manter a consistência na comparação entre métodos. A performance preditiva do SHFM foi avaliada por meio da construção da curva ROC e cálculo da AUC-ROC, o que possibilitou sua comparação direta com os modelos de ML supervisionados.

3.5 Modelos de *Machine Learning* Supervisionados

Além da implementação do SHFM, Foram implementados quatro modelos supervisionados distintos para a predição de mortalidade cardíaca, com o objetivo de comparar sua eficácia frente ao modelo tradicional SHFM e analisar o desempenho relativo entre eles. Os modelos foram:

- **Regressão Logística:** Configurado com tolerância de 1000 iterações.
- **Random Forest:** 100 árvores (n_estimators=100).
- **XGBoost:** Otimizado para AUC-ROC.
- **Support Vector Machine:** Kernel RBF e calibração de probabilidades (CalibratedClassifierCV).

Os modelos Random Forest e XGBoost passaram por um processo adicional de ajustes de hiperparâmetros com o objetivo de maximizar o desempenho preditivo. Para o Random Forest, foram ajustados parâmetros como o número de árvores (n_estimators), profundidade máxima (max_depth) e critério de divisão (min_samples_split). No caso do XGBoost, foram testadas combinações envolvendo o número de estimadores, profundidade das árvores, taxa de aprendizado (learning_rate), amostragem (subsample) e proporção de colunas utilizadas por árvore (colsample_bytree). A seleção das melhores configurações foi realizada por meio de GridSearchCV com validação cruzada estratificada, tendo a AUC-ROC como métrica principal.

3.6 Modelos *Deep Learning* e Transformer

Para explorar abordagens baseadas em redes neurais profundas, foram desenvolvidos dois modelos distintos de *deep learning*, com o objetivo de capturar padrões complexos e não lineares associados à mortalidade cardíaca.

O primeiro modelo foi um MLP Tradicional (Multi-Layer Perceptron), estruturado como uma rede totalmente conectada com as seguintes características:

- Duas camadas ocultas compostas por 128 e 64 neurônios, respectivamente.
- Função de ativação ReLU (Rectified Linear Unit) em todas as camadas ocultas, devido à sua eficiência no treinamento de redes profundas e à capacidade de mitigar o problema de gradientes desvanecentes.
- Camadas de *Dropout* com taxa de 30% aplicadas após cada camada oculta, para reduzir o risco de sobreajuste (*overfitting*).

- Camada final densa com ativação sigmoid, para saída binária (morte/sobrevivência).
- Otimização utilizando o algoritmo Adam com taxa de aprendizado de 0.0001.
- Função de perda: *binary_crossentropy*.
- Métrica de avaliação principal: AUC.

O segundo modelo integrou uma camada baseada em Transformer Encoder à arquitetura MLP, apresentando a seguinte estrutura:

- Camada inicial totalmente conectada com 128 neurônios
- LayerNormalization aplicada após a primeira camada (estabilizar a aprendizagem e acelerar a convergência).
- MultiHeadAttention com 4 cabeças e dimensão de chave de 32, aplicada sobre os *embeddings* expandidos para simular uma estrutura sequencial, permitindo capturar relações internas complexas entre as variáveis de entrada e detectar padrões de dependência cruzada.
- Conexões residuais entre a entrada e a saída do bloco de atenção, favorecendo o fluxo de gradientes e prevenindo degradação da performance.
- Nova camada de LayerNormalization.
- Camada densa com 128 neurônios, seguida por *Dropout* (30%).
- Camada densa adicional com 64 neurônios e *Dropout* (30%).
- Camada final densa com ativação sigmoid, responsável pela predição binária.
- Algoritmo de otimização, função de perda e métrica de avaliação idênticos ao modelo MLP tradicional (Adam, *binary_crossentropy* e AUC).

Ambos os modelos foram treinados com monitoramento de *overfitting* por meio da técnica EarlyStopping (paciência de 20 épocas) e avaliados quanto à capacidade discriminativa. A arquitetura baseada em Transformer foi adaptada para dados tabulares, explorando as vantagens do mecanismo de atenção para identificar dependências cruzadas entre atributos clínicos e laboratoriais, algo que arquiteturas tradicionais podem não capturar de forma tão eficiente.

3.7 Estratégia de Validação, Avaliação e Métricas

Para os modelos tradicionais, incluindo o SHFM e os algoritmos supervisionados de ML, foram construídas curvas ROC a partir das probabilidades preditas por cada modelo. A métrica AUC-ROC foi adotada como principal critério de desempenho, pois avalia de forma abrangente a capacidade discriminativa dos modelos ao longo de diferentes limiares de decisão. Relatórios adicionais contendo métricas de precisão, *recall* e F1-score também foram gerados, possibilitando uma análise detalhada do desempenho de cada modelo em aspectos mais específicos da tarefa de classificação. Além disso, utilizou-se também da matriz de confusão para uma maior clareza da quantidade de falsos positivos e falsos negativos.

Para os modelos supervisionados de ML, a avaliação inicial consistiu em uma validação simples sobre o conjunto de validação, permitindo ajustes e observação preliminar dos resultados. Para garantir maior robustez e reduzir o risco de *overfitting* a uma única divisão de dados, foi realizada também uma avaliação por validação cruzada estratificada com 5 folds, fornecendo uma estimativa mais estável e confiável da performance geral dos modelos. Conforme descrito em 3.3, o balanceamento via SMOTE foi restrito ao treino.

Nos modelos de *deep Learning*, a estratégia de avaliação seguiu princípios semelhantes, com monitoramento da performance tanto no conjunto de validação quanto no conjunto de testes. Foram calculadas curvas ROC e métricas de AUC-ROC separadamente em ambas as etapas, permitindo avaliar não apenas a performance durante o treinamento, mas também a capacidade de generalização dos modelos frente a novos dados. Essa separação foi fundamental para validar a eficácia dos modelos neurais, especialmente considerando a maior complexidade e capacidade de ajuste desses algoritmos.

Por fim, a comparação abrangente incluiu os modelos ajustados (Random Forest e XGBoost, com hiperparâmetros otimizados) e os modelos de *deep Learning* (MLP tradicional e MLP com Transformer), que foram avaliados lado a lado com o SHFM Score. A principal métrica de comparação permaneceu a AUC-ROC, assegurando um critério padronizado e robusto para mensurar a eficácia discriminativa de todas as abordagens testadas. Essa estratégia permitiu não apenas confirmar a viabilidade das técnicas de ML e *deep Learning* no contexto clínico, mas também posicionar esses modelos em relação ao método estatístico tradicional, oferecendo uma visão completa da evolução potencial no prognóstico de mortalidade cardíaca.

4 RESULTADOS EXPERIMENTAIS E DISCUSSÃO

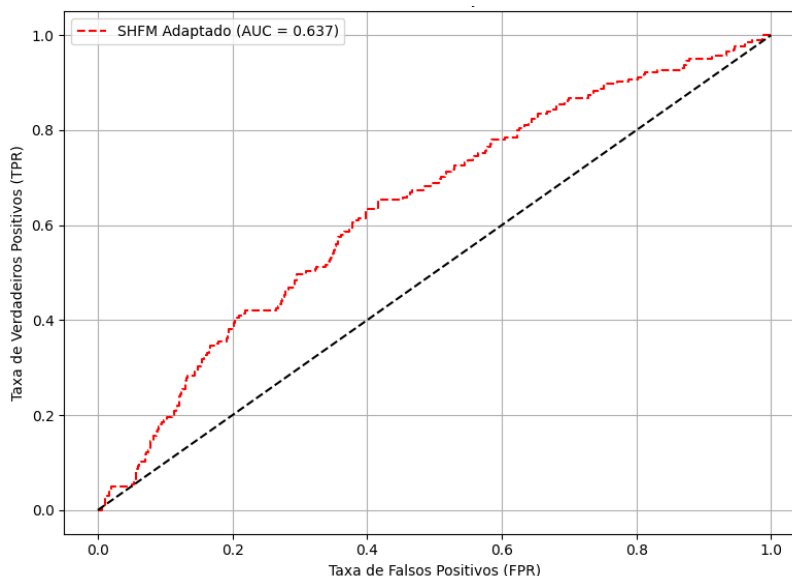
Esta seção apresenta os resultados obtidos a partir da aplicação e avaliação dos modelos propostos para a predição de mortalidade cardíaca em pacientes com insuficiência cardíaca, utilizando o banco de dados MUSIC. Inicialmente, são demonstrados os resultados do SHFM, considerado o *baseline* tradicional, com sua performance avaliada por meio de métricas clássicas como AUC-ROC, precisão, *recall* e F1-score. Em seguida, são apresentados os resultados dos modelos supervisionados de aprendizado de máquina, incluindo Regressão Logística, Random Forest, SVM e XGBoost, comparando seu desempenho entre si e em relação ao SHFM adaptado.

Posteriormente, a seção detalha os resultados dos modelos de *deep Learning*, abrangendo tanto a arquitetura tradicional Multi-Layer Perceptron (MLP) quanto o modelo híbrido que integra uma camada Transformer, avaliando sua eficácia frente aos demais modelos. Além da apresentação individual dos resultados, a seção realiza uma comparação global entre as abordagens tradicionais, os modelos supervisionados e os modelos de *deep Learning*, destacando os ganhos de performance, limitações observadas e implicações clínicas potenciais dos resultados. Por fim, são discutidas as limitações do estudo, como tamanho amostral, balanceamento das classes e limitações inerentes aos dados utilizados, bem como propostas para trabalhos futuros, sugerindo melhorias metodológicas e novas direções de pesquisa.

4.1 Resultados SHFM Adaptado

Para estabelecer uma linha de base de comparação, foi aplicado uma adaptação do SHFM, um escore prognóstico amplamente utilizado na estratificação de risco de pacientes com insuficiência cardíaca. Essa adaptação, como descrito anteriormente, foi necessária devido à ausência de algumas variáveis presentes no modelo original, como CDI, TRC e ácido úrico, entre outras. A figura 5 a seguir representa a curva ROC e AUC do SHFM:

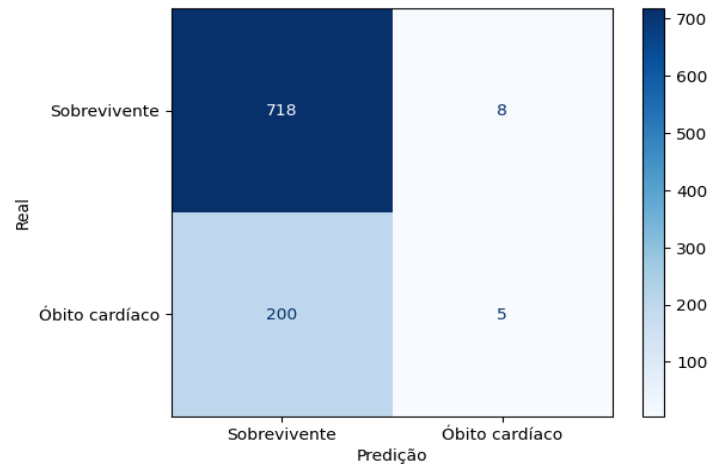
Figura 5 - Curva ROC e AUC SHFM Adaptado



Fonte: Autoria própria

Pela imagem, nota-se que a AUC-ROC foi de aproximadamente 0,637. Este valor indica uma capacidade discriminativa moderada, sendo coerente com os resultados da literatura, que situam a AUC do SHFM entre 0,63 e 0,68 para diferentes populações de insuficiência cardíaca. Apesar de superar a linha de aleatoriedade ($AUC = 0,5$), observa-se que a curva permanece relativamente próxima da diagonal, especialmente nas regiões intermediárias da taxa de falsos positivos, sugerindo limitações na sensibilidade e especificidade para níveis mais restritivos de decisão clínica. Este desempenho reforça o papel do SHFM como uma ferramenta clássica, porém com potencial limitado de predição individualizada, especialmente em cenários onde características populacionais ou fatores de risco adicionais possam influenciar significativamente a mortalidade. A Figura 6 confirma essa limitação por meio da matriz de confusão do método, a qual se observa uma alta quantidade de falsos negativos. Isto significa que o modelo tende a subestimar o risco de óbito, classificando erroneamente muitos pacientes que de fato vieram a óbito como de baixo risco. Esse comportamento ressalta a baixa sensibilidade do SHFM, evidenciando sua limitação prática para uma predição individual eficaz, especialmente quando a detecção precoce de alto risco é crucial.

Figura 6 - Curva ROC + AUC-ROC SHFM adaptado.



Fonte: Autoria própria

Nas Tabelas 1 e 2 são abordadas resumidamente algumas métricas para o método, bem como o risco estimado pelo SHFM adaptado para alguns pacientes do *dataset*.

Tabela 1 - Métricas SHFM adaptado

Métrica	Valor
AUC-ROC	0.637
Precisão	0.385
<i>Recall</i>	0.024
<i>F1-Score</i>	0.046
Acurácia	0.777

Fonte: Autoria própria

Tabela 2 - Dados de Pacientes Aleatórios

ID Paciente	Risco Estimado	Desfecho
458	2.41%	Sobrevivente
863	11.03%	Sobrevivente
832	7.25%	Sobrevivente
311	6.54%	Sobrevivente
78	4.33%	Óbito cardíaco
816	3.78%	Sobrevivente
283	25.20%	Óbito cardíaco
954	41.36%	Sobrevivente

Fonte: Autoria própria

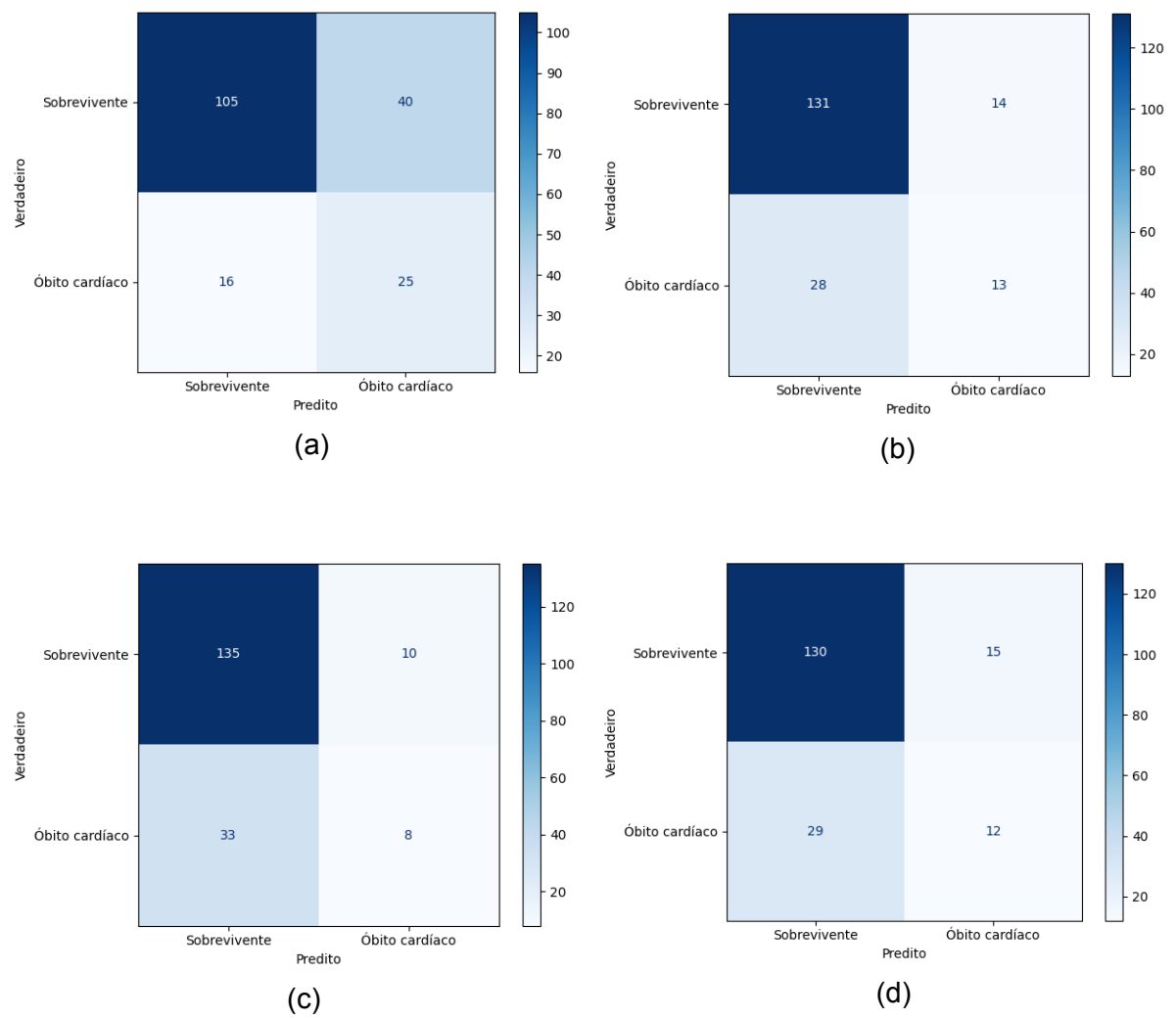
Assim, a Tabela 1 sintetiza as principais métricas do SHFM adaptado no conjunto avaliado, evidenciando o desempenho limitado do modelo, principalmente no que diz respeito à sensibilidade (*recall*) e ao F1-score, ambos notavelmente baixos. Já a Tabela 2 apresenta exemplos reais de pacientes aleatórios do *dataset*, ilustrando a variabilidade dos riscos estimados pelo escore e os respectivos desfechos clínicos, o que reforça a observação de que, apesar do modelo conseguir identificar pacientes com risco elevado em alguns casos, sua acurácia na predição de óbitos permanece insuficiente para aplicações críticas.

Diante destes resultados, é possível dizer que, embora o SHFM permaneça uma ferramenta relevante para avaliação de risco populacional, há espaço considerável para aprimoramento em predições individualizadas. Vale ressaltar novamente que, por limitações do banco de dados, algumas variáveis importantes do SHFM original, como CDI, TRC, creatinina e ácido úrico, não puderam ser incluídas na construção da versão adaptada aplicada neste estudo, o que pode ter impactado negativamente o desempenho observado. Nas próximas seções, serão apresentados os resultados dos modelos de aprendizado supervisionado, com o objetivo de avaliar se abordagens baseadas em ML podem superar as limitações identificadas e oferecer uma alternativa mais robusta para a estratificação de risco em insuficiência cardíaca.

4.2 Resultados Modelos Supervisionados

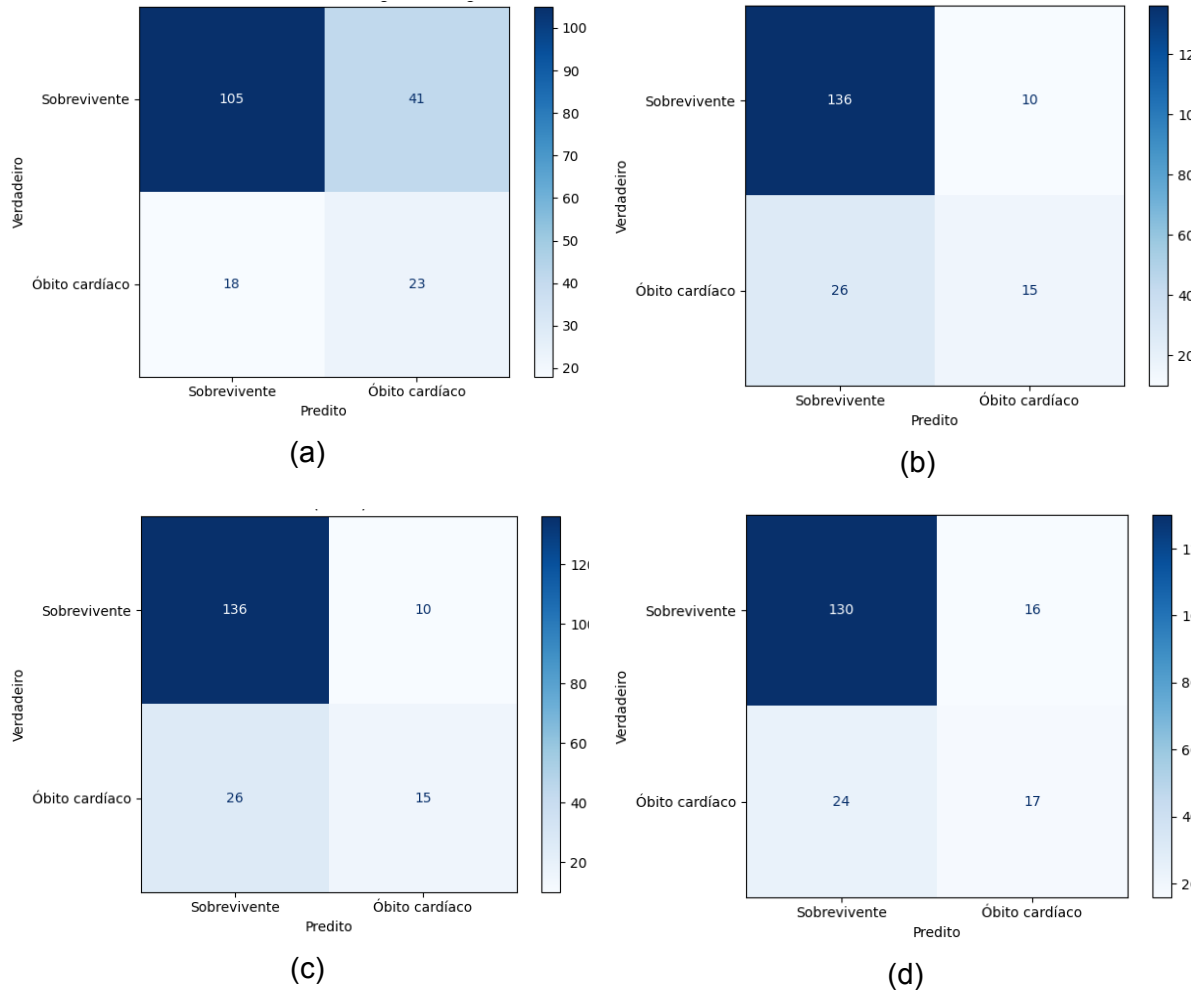
Para avaliar o potencial de modelos baseados em aprendizado supervisionado na predição de mortalidade cardíaca, foram implementados quatro modelos, como já dito nas seções anteriores: Regressão Logística, Suporte a Vetores de Máquinas, Random Forest e XGBoost. Todos os modelos foram treinados, validados e testados conforme descrito na seção de metodologia, utilizando técnicas adequadas de pré-processamento, balanceamento (SMOTE) e validação estratificada. As figuras 7 e 8 a seguir apresentam as matrizes de confusão para cada modelo, considerando respectivamente os conjuntos de validação e de teste, respectivamente:

Figura 7 - Matrizes de Confusão dos Modelos (Validação) (a) Regressão Logística, (b) SVM, (c) Random Forest e (d) XGBoost



Fonte: Autoria própria

Figura 8 - Matrizes de Confusão dos Modelos (Teste) (a) Regressão Logística, (b) SVM, (c) Random Forest e (d) XGBoost



Fonte: Autoria própria

De forma geral, nota-se que todos os modelos apresentam desempenho superior ao SHFM adaptado na detecção de óbitos cardíacos, com uma clara redução da quantidade de falsos negativos em comparação ao observado na Figura 6. Na validação (Figura 7), observa-se que o Random Forest, SVM e XGBoost exibem desempenho mais equilibrado entre sensibilidade (capacidade de identificar óbitos) e especificidade (capacidade de identificar sobreviventes), enquanto a Regressão Logística apresentara um maior número de falsos negativos, embora ainda superando o SHFM adaptado nesse aspecto. Esse padrão se mantém de forma consistente no conjunto de teste (Figura 8), reforçando a estabilidade dos modelos e a boa generalização.

É importante destacar que, ao contrário do SHFM adaptado, que exibiu elevada taxa de falsos negativos (subestimação do risco de óbito), os modelos supervisionados conseguiram mitigar esse problema, aumentando significativamente a detecção de eventos fatais, mesmo que com leve aumento de falsos positivos. Esse comportamento é particularmente relevante em contextos clínicos, nos quais a detecção precoce de risco elevado pode ter impacto direto na intervenção e na sobrevida dos pacientes.

De forma comparativa, tanto no conjunto de validação quanto no teste, a Random Forest e o XGBoost se mostraram mais robustos, apresentando menor número de falsos negativos e uma matriz de confusão mais balanceada. A Regressão Logística e o SVM, apesar de sua simplicidade, também mostraram avanços em relação ao SHFM adaptado, mas com resultados um pouco mais modestos, especialmente em termos de sensibilidade.

Esses resultados sugerem que a adoção de modelos de ML para predição de mortalidade cardíaca tem potencial de superar significativamente as limitações da adaptação do SHFM, principalmente no que diz respeito à detecção de pacientes de alto risco. Entretanto, destaca-se a necessidade de avaliar o impacto clínico do aumento de falsos positivos, que pode levar a intervenções desnecessárias, ressaltando a importância de ajustar os limiares de decisão conforme o contexto clínico específico. As tabelas 3 e 4 a seguir também confirmam essa maior robustez dos modelos em relação ao SHFM adaptado, em que os quatro modelos conseguem superar este método em todas as métricas:

Tabela 3 - Métricas no conjunto validação para os modelos supervisionados

Modelo	AUC-ROC	Precisão	Recall	F1-Score	Acurácia
Regressão Logística	0.712	0.385	0.610	0.472	0.699
SVM	0.710	0.481	0.317	0.382	0.774
Random Forest	0.676	0.429	0.22	0.29	0.769
XGBoost	0.680	0.444	0.293	0.353	0.763

Fonte: Autoria própria

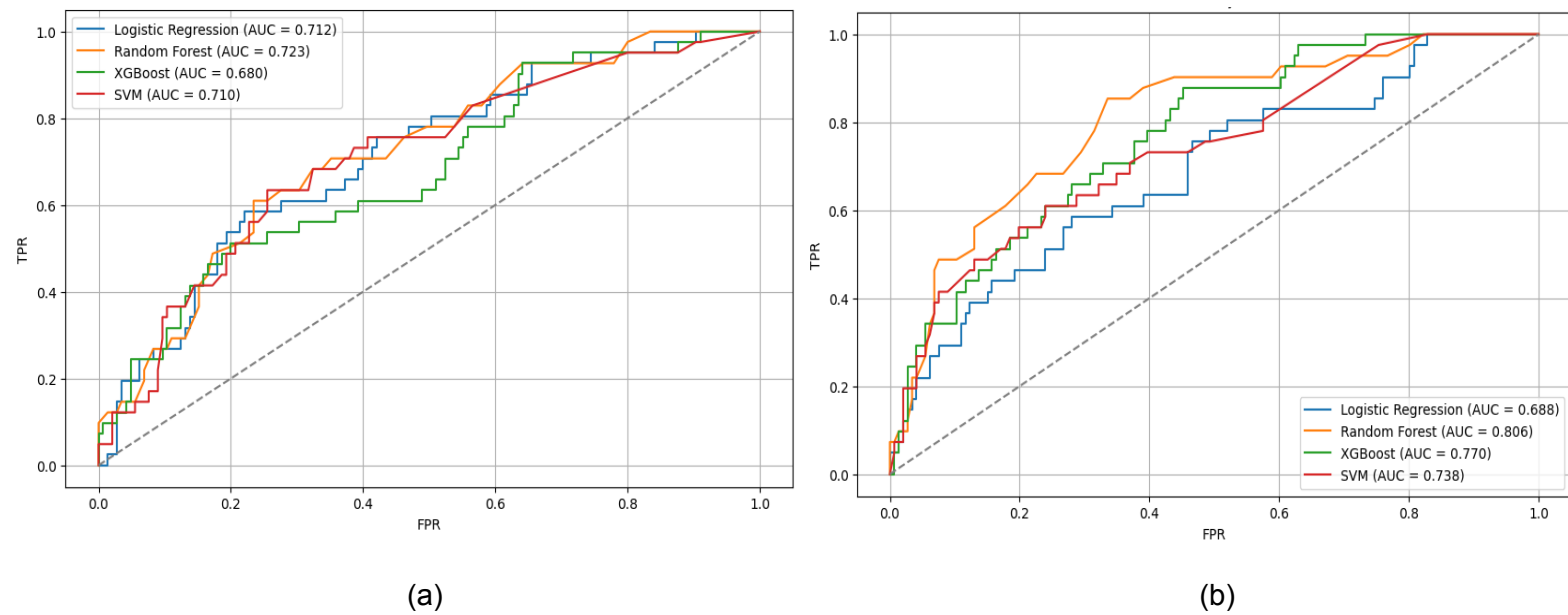
Tabela 4 - Métricas no conjunto de teste para os modelos supervisionados

Modelo	AUC-ROC	Precisão	Recall	F1-Score	Acurácia
Regressão Logística	0.688	0.359	0.561	0.438	0.684
SVM	0.738	0.407	0.268	0.455	0.807
Random Forest	0.806	0.750	0.293	0.455	0.807
XGBoost	0.770	0.515	0.415	0.459	0.786

Fonte: Autoria própria

Assim, a Figura 9 ilustra um comparativo das curvas ROC e AUC de cada um dos modelos entre si.

Figura 9 - Curva ROC-AUC dos modelos Supervisionados (a) Conjunto de Validação (b) Conjunto de testes



Fonte: Autoria própria

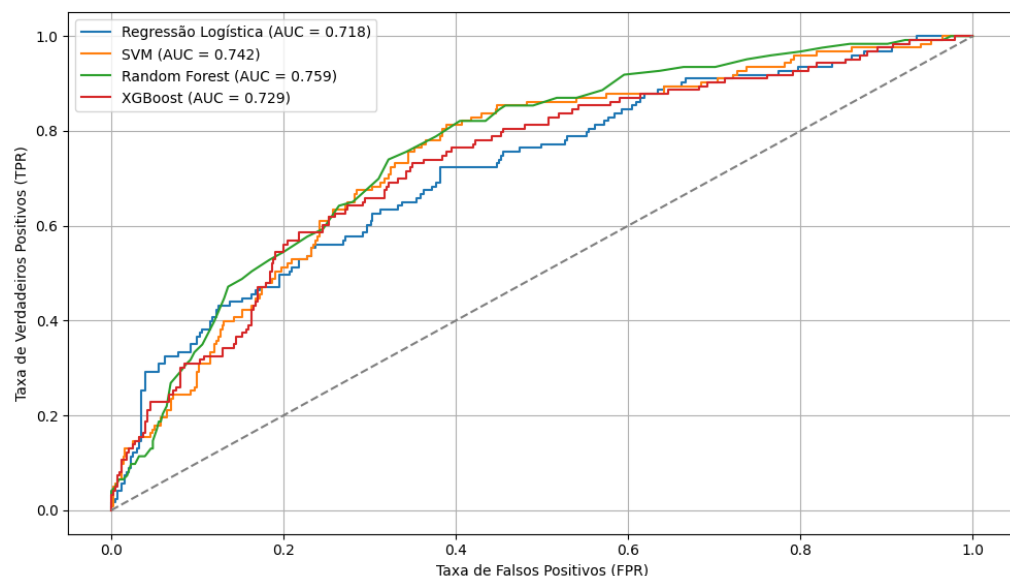
De forma geral, observa-se que todos os modelos mantiveram desempenho superior ao SHFM adaptado (AUC = 0.637), reforçando os ganhos trazidos pelas abordagens baseadas em ML. No conjunto de validação (Figura 9a), o Random Forest apresentou a maior AUC (0.723), seguida pela SVM (0.710), Regressão Logística (0.712) e XGBoost (0.680). Embora as diferenças entre os modelos sejam discretas, o Random Forest destacou-se levemente em termos de área sob a curva, o que corrobora seu bom desempenho ao lidar com dados tabulares. No conjunto de teste (Figura 9b), nota-se um comportamento ligeiramente diferente: a Random Forest apresentou a maior AUC (0.806), seguida pelo XGBoost (0.770), SVM (0.738) e Regressão Logística (0.788). Essa inversão de desempenho evidencia a boa capacidade de generalização dos modelos de *ensemble* (Random Forest e XGBoost), que superaram a Regressão Logística e SVM no cenário real de teste, possivelmente por capturar relações mais complexas entre as variáveis.

Outro ponto importante é que, apesar das diferenças numéricas, as curvas ROC dos quatro modelos se mostraram relativamente próximas, especialmente para taxas de falsos

positivos abaixo de 0.3, indicando que todos os modelos são capazes de distinguir bem entre sobreviventes e óbitos cardíacos nessa faixa crítica para decisão clínica.

Para garantir uma melhor validação dos resultados, foi realizada uma validação cruzada estratificada (*5-fold*) durante o treinamento dos modelos. Nessa etapa, foram coletadas as métricas médias de AUC-ROC, precisão, *recall* e F1-score, o que reforçou a estabilidade observada nos conjuntos de validação e teste. Os resultados da validação cruzada demonstraram consistência com os resultados já apresentados, destacando que Random Forest continuou a apresentar desempenho superior, confirmando sua robustez mesmo em diferentes partições dos dados. A figura 10 ilustra esses resultados.

Figura 10 - Curva ROC-AUC Validação Cruzada.

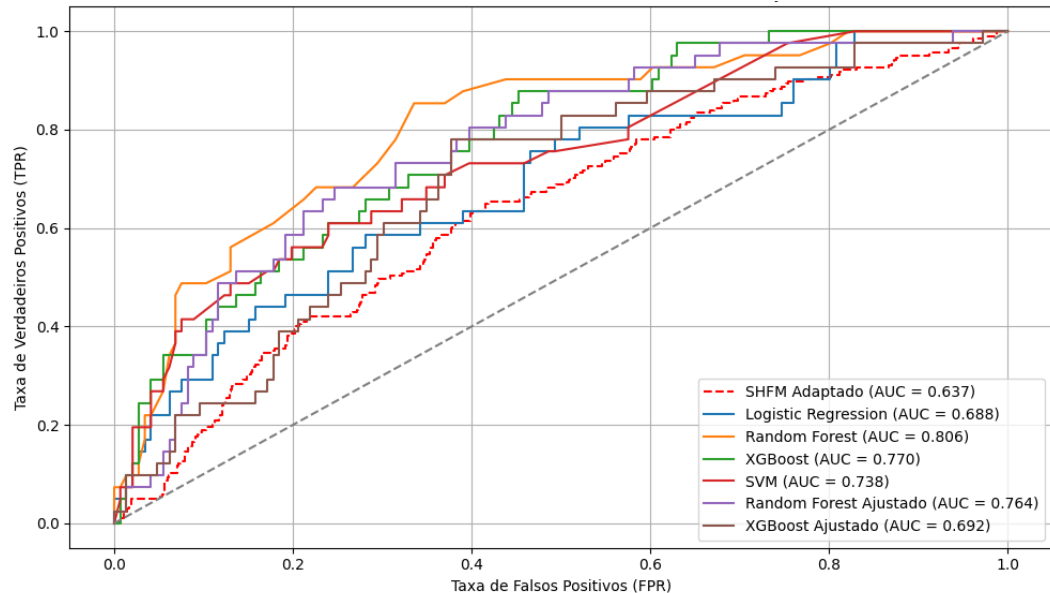


Fonte: Autoria própria

Para complementar as análises, foi realizado o ajuste dos hiperparâmetros dos modelos Random Forest e XGBoost, buscando melhorias adicionais de desempenho. Esse ajuste foi conduzido por meio de uma busca em grade (*Grid Search*) acoplada a um *pipeline* com SMOTE interno e validação cruzada estratificada (*3-fold*), garantindo um balanceamento adequado das classes durante o treinamento. No caso do Random Forest, foram avaliadas combinações dos seguintes hiperparâmetros: número de árvores ($n_estimators = 50, 100, 200$), profundidade máxima ($max_depth = \text{None}, 10, 20, 30$) e tamanho mínimo de divisão interna ($min_samples_split = 2, 4, 5$). Já para o XGBoost, foram ajustados o número de árvores ($n_estimators = 50, 100, 200$), profundidade das árvores ($max_depth = 3, 5, 7$), taxa de aprendizado ($learning_rate = 0.001, 0.01, 0.1$), fração de amostragem ($subsample = 0.8, 1$),

1.2) e fração de colunas (*colsample_bytree* = 0.6, 0.8, 1). Dessa forma, a Figura 11 apresenta o comparativo geral das curvas ROC envolvendo os quatro modelos supervisionados, suas versões ajustadas e o escore SHFM adaptado.

Figura 11 - Curva ROC-AUC comparativa de todos os modelos.



Fonte: Autoria própria

Observa-se claramente que todos os modelos de ML superaram o SHFM adaptado em termos de discriminação, apresentando AUC-ROC consideravelmente superiores (SHFM adaptado AUC = 0,637 versus modelos de ML acima de 0,71). Entre os modelos supervisionados, Random Forest (AUC = 0,806) e XGBoost (AUC = 0,770) mantiveram desempenho robusto, com equilíbrio entre sensibilidade e especificidade. A Regressão Logística (AUC = 0,688) e o SVM (AUC = 0,738), embora mais simples, também superaram a adaptação do SHFM de forma consistente. O ajuste de hiperparâmetros não representou uma virada significativa no desempenho geral, reforçando que os modelos já estavam bem calibrados previamente. Estes resultados consolidam a superioridade dos modelos baseados em ML em relação ao SHFM adaptado, principalmente na capacidade de identificar pacientes de alto risco, mostrando-se promissores como ferramentas complementares para a estratificação de risco em insuficiência cardíaca.

Adicionalmente, foi realizada uma análise das variáveis mais relevantes para os modelos Random Forest e XGBoost, com o objetivo de identificar os fatores clínicos de maior influência na predição. As 15 variáveis mais importantes foram:

- Extrassístoles ventriculares pareadas
- Níveis de Pro-BNP (ng/L)
- Diâmetro atrial esquerdo (mm)
- Creatinina (μ mol/L)
- Infarto do miocárdio prévio
- Fração de ejeção do ventrículo esquerdo (LVEF %)
- Doença vascular periférica
- Idade
- Ureia (mg/dL)
- Distúrbio de condução intraventricular
- Duração do QRS (ms)
- Relação cardiotorácica
- Pressão arterial diastólica (mmHg)
- Revascularização prévia
- Presença de dispositivos implantáveis prévios

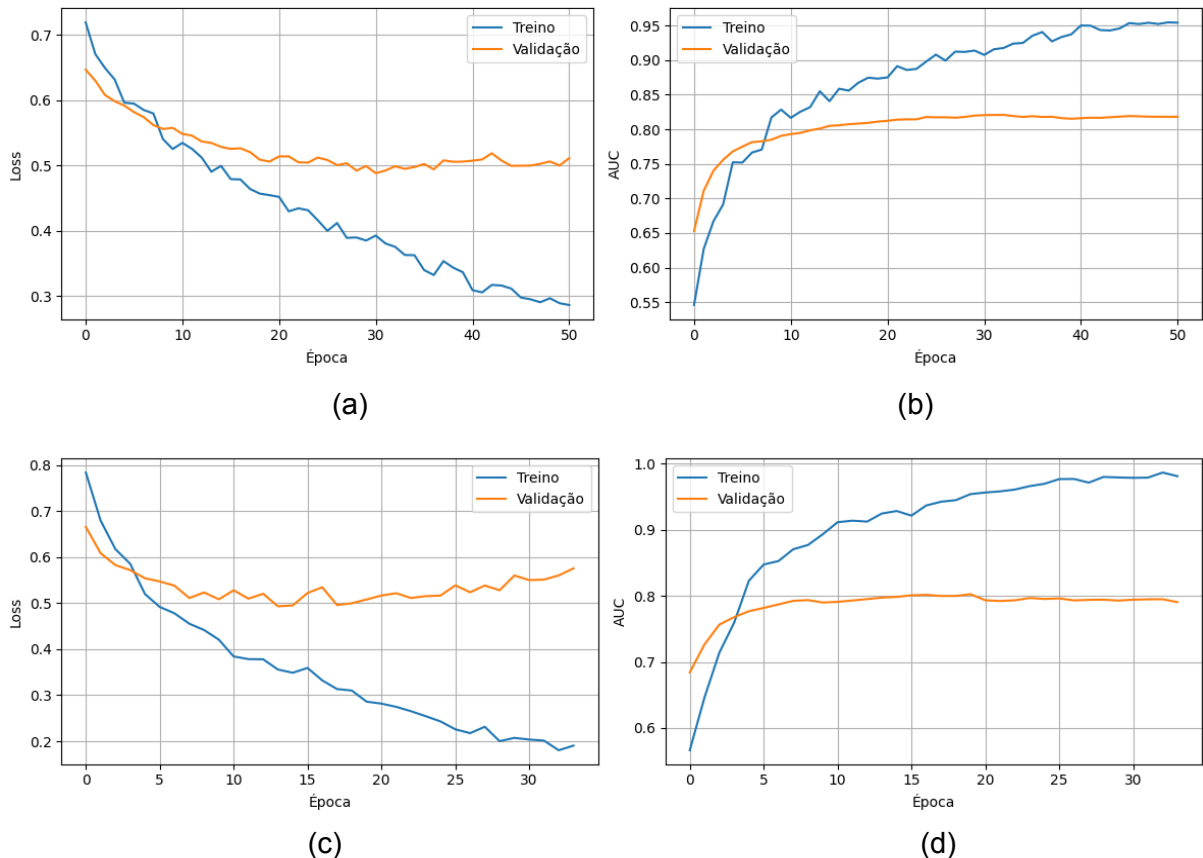
Essas variáveis, majoritariamente clínicas e laboratoriais, reforçam bem estabelecidos na literatura médica sobre insuficiência cardíaca, indicando que os modelos baseiam suas decisões em fatores fisiopatológicos relevantes e clinicamente interpretáveis. Outros fatores tradicionalmente associados à piora prognóstica em pacientes com insuficiência cardíaca, como distúrbios hidroeletrólíticos (ex.: sódio e potássio), variáveis de congestão pulmonar e biomarcadores inflamatórios, também poderiam ter surgido entre as variáveis mais importantes, caso estivessem melhor representados ou tivessem maior variabilidade dentro da base de dados utilizada.

4.3 Resultados Redes Neurais Profundas

Para os modelos de *deep Learning*, duas arquiteturas foram implementadas: um modelo MLP tradicional e uma versão MLP com camada Transformer Encoder. O modelo MLP consistiu em camadas densas com funções de ativação ReLU e *dropout* para evitar o *overfitting*, enquanto o modelo MLP + Transformer incorporou mecanismos de atenção *multi-head* para capturar relações entre atributos de entrada. Ambos os modelos foram treinados com o otimizador Adam, utilizando taxa de aprendizado de 0.0001 e monitoramento de *early stopping* com paciência de 20 épocas para prevenir *overfitting*, com mais detalhes na

seção 3.6. A figura 12 a seguir ilustra as curvas de treino dos dois modelos (curva de perda e AUC):

Figura 12 - Curvas de perdas do MLP (a) e MLP + Transformer (c), e curvas AUC do MLP (b) e MLP + Transformer (d).

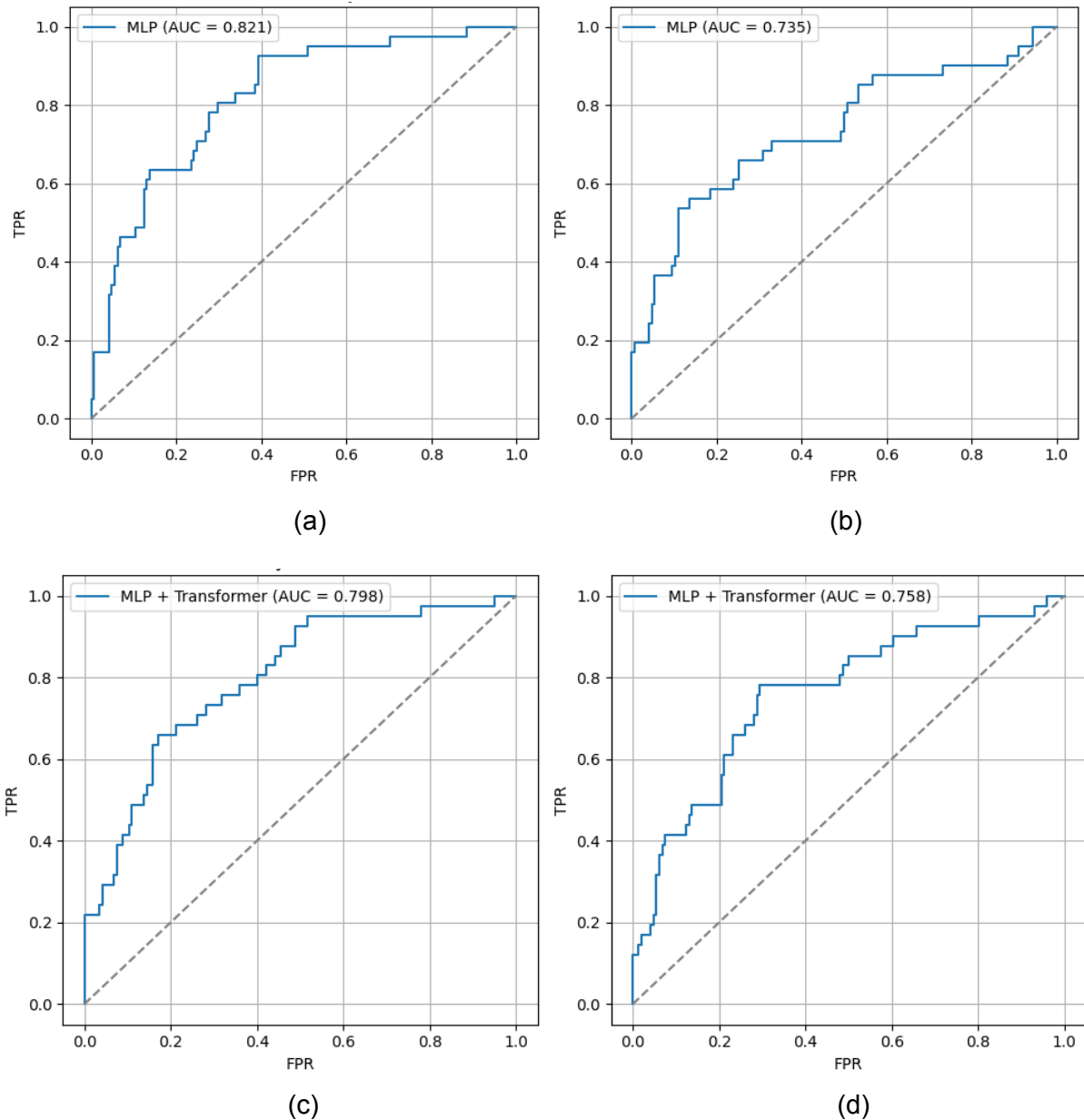


Fonte: Autoria própria.

As curvas de treino dos modelos deep revelam um bom aprendizado nos dados de treino, com reduções progressivas na perda e aumentos consistentes no AUC ao longo das épocas. O MLP simples mostrou queda contínua na perda de treino e aumento do AUC até aproximadamente 0,95, enquanto o conjunto de validação estabilizou mais cedo, sugerindo um leve *overfitting* após cerca de 20–30 épocas. O modelo MLP + Transformer apresentou comportamento similar, com melhora rápida no início e estabilização precoce no conjunto de validação, também indicando *overfitting* moderado. De modo geral, ambos os modelos atingiram boa performance no treino, mas a distância crescente entre as curvas de treino e validação sinaliza que ajustes adicionais (como mais regularização ou tuning) poderiam melhorar a generalização.

A figura 13 ilustra as curvas ROC dos modelos para o conjunto de teste e validação:

Figura 13 - Curvas ROC no conjunto de validação do MLP (a) e MLP + Transformer (c), e curvas ROC no conjunto de teste do MLP (b) e MLP + Transformer (d).



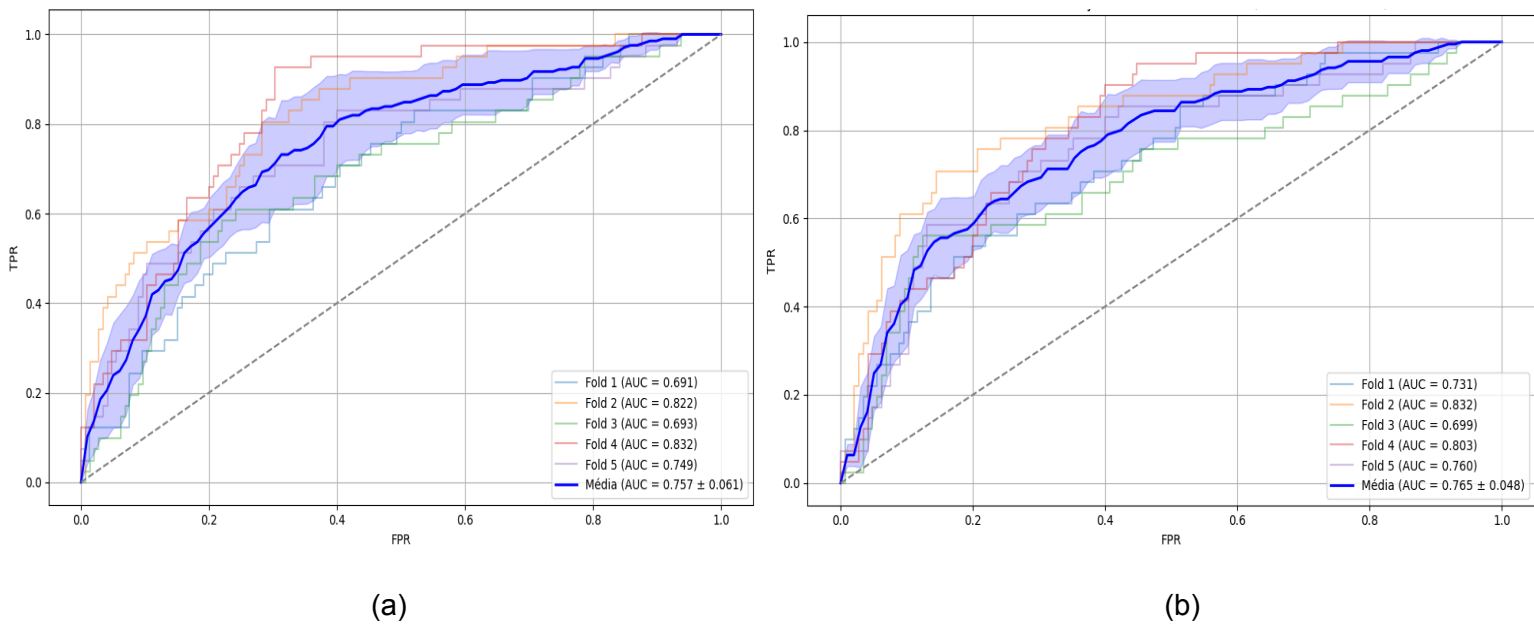
Fonte: Autoria própria.

As curvas ROC apresentadas indicam que os modelos *deep learning* tiveram desempenhos competitivos, especialmente em validação, mas com sinais claros de queda no conjunto de testes. O MLP simples alcançou AUC de 0,821 na validação e 0,735 no teste, enquanto o MLP + Transformer apresentou AUC de 0,798 na validação e 0,758 no teste. Isso pode significar que, embora ambos tenham aprendido padrões relevantes, houve leve *overfitting*, já que a performance caiu ao sair do conjunto de validação para o de teste. Assim, pode-se notar que o Transformer conseguiu manter um AUC um pouco superior no teste em

relação ao MLP puro, indicando que a arquitetura com atenção pode oferecer melhor generalização nesse contexto. Apesar disso, ambos ainda apresentaram desempenhos um pouco abaixo dos melhores modelos de ML (como Random Forest e XGBoost), reforçando que, neste caso, modelos tradicionais bem calibrados podem superar redes neurais mais complexas. Uma possível explicação se deve ao tamanho relativamente limitado do *dataset* e ao desbalanceamento das classes. Esse cenário reforça a necessidade de mais dados ou de estratégias de regularização para que os modelos deep expressem todo seu potencial.

Além da avaliação em validação e teste, também foi realizada uma validação cruzada estratificada (*5-fold*) para os modelos MLP e MLP + Transformer. Essa abordagem teve como objetivo avaliar a robustez média dos modelos frente à variação dos dados e evitar qualquer vazamento de informações entre treino e validação. A Figura 14 apresenta as curvas ROC médias obtidas com essa validação cruzada:

Figura 14 - Curvas ROC com validação cruzada estratificada *5-fold* dos modelos MLP (a) e MLP + Transformer (b)

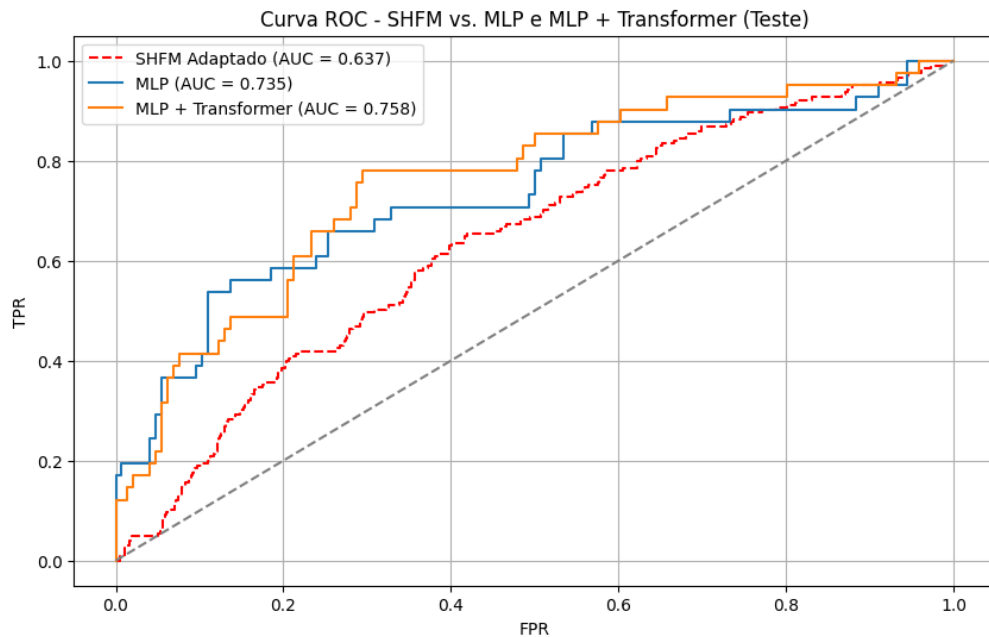


Fonte: Autoria própria.

Os resultados demonstraram AUC-ROC média de $0,757 \pm 0,061$ para o modelo MLP e $0,765 \pm 0,048$ para o MLP + Transformer. Nota-se que os resultados são similares aos obtidos na figura 13, podendo significar que a robustez do modelo para este cenário é válida.

Por fim, a figura 15 a seguir traz um comparativo entre os modelos de redes neurais com o SHFM adaptado:

Figura 15 - Curvas ROC e AUC modelos Deep e SHFM adaptado.



Fonte: Autoria própria.

A análise final das curvas ROC no conjunto de teste mostra que tanto o MLP quanto o MLP + Transformer superaram o modelo SHFM adaptado, que apresentou uma AUC de 0,637. O MLP alcançou AUC de 0,735, enquanto o MLP + Transformer atingiu 0,758, indicando melhorias relevantes na capacidade de discriminar casos de mortalidade cardíaca. Embora as redes neurais tenham apresentado uma performance um pouco inferior aos melhores modelos tradicionais de ML, como dito anteriormente, elas ainda representaram um avanço considerável em relação ao *baseline* clínico, representado aqui pela versão adaptada do SHFM. Esses resultados reforçam que métodos baseados em *deep learning* possuem potencial para oferecer suporte adicional à prática clínica, destacando-se principalmente em contextos onde a detecção precoce de alto risco é fundamental. Porém, eles também evidenciam que, sem ajustes adicionais ou *datasets* maiores, modelos tradicionais bem otimizados podem continuar a ser mais competitivos em aplicações práticas.

4.4 Custo Computacional

Com o objetivo de enriquecer a análise comparativa entre os modelos, foi realizada uma mensuração do custo computacional associado ao treinamento de cada um deles. Para isso, foi registrado o tempo total de treinamento, desde o início até a conclusão do ajuste dos

parâmetros, considerando as mesmas divisões de treino utilizadas nas etapas anteriores. Os tempos observados foram os seguintes:

- Regressão Logística: 0,03 segundos
- Random Forest: 0,40 segundos
- XGBoost: 0,31 segundos
- SVM: 0,39 segundos
- MLP: 25,76 segundos
- MLP + Transformer: 18,82 segundos

Observa-se que os modelos clássicos de aprendizado de máquina (Regressão Logística, Random Forest, XGBoost e SVM) apresentaram tempos de treinamento extremamente reduzidos, todos inferiores a 1 segundo, o que os torna bastante atrativos para aplicações clínicas onde há necessidade de rápida reavaliação de modelos com base em novos dados. Por outro lado, os modelos de *deep learning* (MLP e MLP + Transformer) demandaram tempos significativamente maiores, com o MLP levando cerca de 26 segundos e o MLP + Transformer aproximadamente 19 segundos. Esse aumento no custo computacional já era esperado, uma vez que redes neurais profundas, mesmo com arquiteturas relativamente simples como as utilizadas neste estudo, possuem maior número de parâmetros e exigem maior esforço de otimização durante o processo de aprendizado.

4.4 Discussão dos Resultados

Os resultados deste estudo mostram certa superioridade das abordagens de aprendizado de máquina e *deep learning* em relação ao modelo tradicional (representado pela adaptação do SHFM) para a predição de mortalidade cardíaca. O SHFM adaptado, apesar de consolidado na prática clínica, apresentou AUC-ROC de 0,637, indicando desempenho apenas moderado. Isso era esperado, já que ele foi desenvolvido a partir de coortes específicas e depende de variáveis limitadas, não capturando interações complexas ou padrões não lineares que podem existir entre os fatores clínicos e laboratoriais. Além disso, a adaptação SHFM apresentou sensibilidade muito baixa, resultando em muitos falsos negativos, o que limita sua utilidade para a detecção precoce de pacientes de alto risco. Apesar de se tratar de uma adaptação, o que foi apontado nos resultados também observa-se em aplicações mais fiéis ao modelo original na literatura, como foi explorado na seção 2.2. Ainda assim, é importante destacar que os resultados obtidos devem ser interpretados considerando essas

limitações metodológicas, uma vez que a ausência de algumas variáveis importantes pode ter impactado negativamente a performance do SHFM adaptado.

Os modelos supervisionados (Regressão Logística, SVM, Random Forest e XGBoost) mostraram ganhos consistentes sobre o SHFM adaptado, com destaque para Random Forest e XGBoost, que alcançaram AUCs próximas ou acima de 0,77 no conjunto de teste. Esses modelos se beneficiam de sua capacidade de capturar interações não lineares e complexas entre as variáveis, algo fora do alcance de modelos tradicionais. Random Forest, por exemplo, combina múltiplas árvores de decisão, reduzindo o viés e a variância por meio de *bagging*, enquanto o XGBoost se destaca por seu mecanismo de *boosting*, corrigindo erros sequenciais e explorando relações sutis no conjunto de dados.

A regressão logística apresentou desempenho robusto durante a validação, mas sofreu uma queda de aproximadamente 3,4% na AUC no conjunto de teste, indicando um leve problema de generalização. Ainda assim, esse comportamento reforça a ideia de que, em cenários com tamanho amostral limitado, modelos mais simples tendem a ser menos suscetíveis ao *overfitting* extremo quando comparados a modelos mais complexos. No entanto, no teste, os modelos ensemble (Random Forest e XGBoost) superaram a regressão logística, demonstrando melhor capacidade de generalização, possivelmente por conseguirem capturar interações não lineares e padrões complexos que um modelo linear não é capaz de representar adequadamente.

O ajuste de hiperparâmetros não trouxe ganhos significativos. Pelo contrário, fez com que o modelo apresentasse uma generalização um pouco inferior. Isso sugere que, mesmo com configurações padrão, os modelos já estavam razoavelmente calibrados, possivelmente porque o conjunto de dados, embora rico, não era grande o suficiente para explorar completamente ajustes mais finos sem risco de *overfitting*.

Os modelos de *deep learning* (MLP e MLP + Transformer) apresentaram AUCs competitivos, superando o SHFM adaptado (0,735–0,758), mas ficaram um pouco abaixo dos melhores modelos de ML. Isso pode ser explicado pelo fato de que redes neurais, especialmente arquiteturas complexas como Transformers, geralmente demandam grandes volumes de dados para atingir todo seu potencial. O *dataset* MUSIC, embora robusto para modelos clássicos, pode ter sido limitado para os modelos deep, que são mais propensos ao sobreajuste em amostras menores. As curvas de treino confirmam essa hipótese, mostrando indícios de *overfitting* após 20–30 épocas, com perda crescente de generalização.

Apesar disso, o Transformer demonstrou uma leve vantagem sobre o MLP puro no teste, podendo afirmar que os mecanismos de atenção conseguem capturar dependências entre

variáveis que as redes densas convencionais não conseguem, ainda que as melhorias tenham sido discretas.

Do ponto de vista clínico, a redução dos falsos negativos nos modelos de ML e *deep learning* é altamente relevante, especialmente quando comparada à elevada taxa de falsos negativos observada no SHFM adaptado, pois a falha em identificar pacientes de alto risco pode resultar em desfechos graves. No entanto, esse ganho muitas vezes veio com aumento de falsos positivos, o que precisa ser ponderado, já que intervenções desnecessárias também trazem custos e riscos. Outro aspecto importante a ser considerado é a relação entre a AUC-ROC e o recall (sensibilidade), métricas que, apesar de complementares, refletem diferentes perspectivas clínicas na avaliação de modelos preditivos. Enquanto a AUC-ROC fornece uma visão global da capacidade discriminativa do modelo em todos os limiares de decisão, o recall é particularmente crítico em cenários onde a identificação de pacientes de alto risco é prioritária, como no contexto da mortalidade por insuficiência cardíaca. No presente estudo, alguns modelos apresentaram AUC-ROC satisfatória, mas com recall relativamente baixo, como observado no desempenho do SHFM adaptado, que obteve sensibilidade de apenas 2,4%. Em contrapartida, os modelos Random Forest e XGBoost, além de alcançarem AUCs superiores a 0,77, também mostraram recall significativamente mais elevado, superando 29% e 40%, respectivamente, no conjunto de teste. Essa diferença é clínica e operacionalmente relevante, uma vez que a capacidade de reduzir falsos negativos pode impactar diretamente a intervenção precoce e o prognóstico dos pacientes. Porém, é importante destacar que, apesar da melhoria, os valores de recall obtidos ainda estão aquém do ideal para aplicações clínicas sensíveis, indicando que há espaço para avanços futuros na otimização dos modelos ou na incorporação de novas variáveis preditoras. Assim, reforça-se que a escolha do melhor modelo não deve se basear exclusivamente na AUC-ROC, mas considerar também métricas como o recall e o F1-score, além de possíveis ajustes nos limiares de decisão, de modo a alinhar o desempenho dos modelos às necessidades práticas de cada contexto clínico.

5 CONCLUSÃO

Este trabalho teve o objetivo de investigar o uso de técnicas de aprendizado de máquina (ML) na predição de mortalidade em pacientes com insuficiência cardíaca crônica, utilizando como *baseline* uma adaptação do modelo tradicional SHFM. A partir do banco de dados MUSIC, foi conduzido um pré-processamento rigoroso, incluindo análise exploratória, tratamento de valores ausentes, eliminação de colunas redundantes e balanceamento de classes por SMOTE, garantindo a robustez e a qualidade dos dados para os experimentos.

Os resultados obtidos sustentam a hipótese central deste estudo: os modelos baseados em ML podem superar o SHFM adaptado em termos de desempenho preditivo, principalmente no contexto em questão (pacientes com problemas de insuficiência cardíaca). Enquanto a adaptação do SHFM apresentou uma AUC-ROC de aproximadamente 0,637, os modelos supervisionados (especialmente Random Forest e XGBoost) alcançaram AUCs superiores a 0,77 no conjunto de teste, demonstrando uma melhora significativa na capacidade discriminativa. Além disso, os modelos de *deep Learning* (MLP tradicional e MLP com camada Transformer) também mostraram avanços relevantes, atingindo AUCs de até 0,750, superando o *baseline* clínico (representado pelo SHFM adaptado) e evidenciando seu potencial para aplicações futuras.

Observou-se ainda que o ajuste de hiperparâmetros e a integração de mecanismos de atenção, como o Transformer, trouxeram poucos ganhos, apesar dos modelos tradicionais de *ensemble* terem mantido desempenho superiores aos modelos ajustados. Um ponto de destaque foi a redução expressiva de falsos negativos nas predições, fator crítico em cenários clínicos onde a detecção precoce de pacientes de alto risco pode influenciar diretamente as decisões terapêuticas e os desfechos clínicos.

Quanto à aplicação prática, a implementação de modelos baseados em aprendizado de máquina para predição de mortalidade em insuficiência cardíaca poderia ser realizada em diferentes níveis do sistema de saúde, como em hospitais, clínicas especializadas ou até mesmo em plataformas de triagem remota. Porém, antes de uma adoção clínica, seria necessário ampliar a quantidade e a diversidade dos dados utilizados para o treinamento dos modelos, de forma a reduzir riscos de viés e garantir maior representatividade populacional. Não só isso, mas também seriam imprescindíveis validações externas e testes prospectivos, acompanhados de uma avaliação rigorosa por profissionais da saúde. Outro ponto relevante diz respeito à escolha da arquitetura e dos hiperparâmetros mais adequados para diferentes contextos clínicos, o que demanda estudos adicionais. Um fator importante a ser considerado

é o custo computacional envolvido. Em sistemas de saúde pública, esse pode representar um desafio, visto que a implementação exigiria maior capacidade de processamento e arquiteturas mais robustas para lidar com volumes de dados ampliados. Ainda assim, com o uso de estratégias de otimização e adequação de recursos computacionais, a aplicação prática dos modelos desenvolvidos pode ser viável. Dessa forma, qualquer futura implantação deve ser cuidadosamente acompanhada por equipes multidisciplinares, garantindo a segurança, a acurácia e a utilidade real do sistema proposto.

Apesar dos resultados satisfatórios, algumas limitações foram identificadas: o tamanho amostral relativamente restrito e o desbalanceamento das classes podem ter impactado a generalização dos modelos, principalmente nas arquiteturas mais complexas, como as redes neurais profundas. Além disso, o conjunto de dados, embora rico em informações, representa uma população específica, o que pode limitar a aplicabilidade dos modelos desenvolvidos em outros contextos clínicos sem validação externa. Outra limitação importante observada ao longo do desenvolvimento deste trabalho foi a escassez de bases de dados públicas nacionais com informações clínicas detalhadas sobre pacientes com insuficiência cardíaca, contendo registros de longo prazo e desfechos de mortalidade. A maior parte dos datasets amplamente utilizados para predição de mortalidade cardíaca, como o próprio MUSIC (MARTÍN-YEBRA et al., 2024) e o MIMIC-III (Johnson et al., 2016), são provenientes de instituições internacionais, com características demográficas e epidemiológicas distintas da população brasileira. Essa lacuna dificulta a validação de modelos preditivos no contexto nacional, podendo comprometer a aplicabilidade direta dos resultados obtidos em populações locais. Assim, reforça-se a necessidade de investimentos em iniciativas brasileiras voltadas à construção e disponibilização de grandes bases clínicas com desfechos acompanhados, o que permitiria o desenvolvimento de soluções de inteligência artificial mais alinhadas à realidade epidemiológica do país.

Como sugestão para aplicações futuras, ampliar a base de dados utilizada, incorporar variáveis adicionais como biomarcadores e dados de monitoramento remoto, realizar testes e validações externas, e explorar abordagens de aprendizado explicável podem ser boas práticas para garantir maior interpretação nas predições pelos profissionais de saúde. Além disso, estudos multicêntricos e análises prospectivas são fundamentais para validar a robustez e a utilidade clínica dos modelos em cenários reais.

Em resumo, este estudo demonstra o potencial transformador do aprendizado de máquina na medicina cardiovascular, reforçando a importância de integrar essas ferramentas

às práticas clínicas para aprimorar a estratificação de risco, personalizar intervenções e, potencialmente, melhorar os desfechos dos pacientes.

REFERÊNCIAS

- ALBUQUERQUE, D. C. et al. **Manual de Insuficiência Cardíaca**. Rio de Janeiro: SOCERJ, 2019. Disponível em: <https://cssjd.org.br/imagens/editor/files/2019/Manual%20de%20Insuficiência%20Cardíaca%20SOCERJ2019.pdf>. Acesso em: 10 abr. 2025.
- ALI, H. Reinforcement learning in healthcare: optimizing treatment strategies, dynamic resource allocation, and adaptive clinical decision-making. **International Journal of Computer Applications Technology and Research**, v. 11, n. 3, p. 88-104, 2022. DOI: <https://doi.org/10.7753/IJCATR1103.1007>.
- AMBRISH, G. et al. Logistic regression technique for prediction of cardiovascular disease. **Global Transitions Proceedings**, [S.l.], v. 3, p. 127–132, 2022. DOI: <https://doi.org/10.1016/j.gltp.2022.04.008>.
- ANSHORI, M.; HARIS, M. S. Predicting Heart Disease using Logistic Regression. **Knowledge Engineering and Data Science**, v. 5, n. 2, p. 188–196, 2022. Disponível em: https://www.researchgate.net/publication/371142044_Predicting_Heart_Disease_using_Logistic_Regression. Acesso em: 25 abr. 2025.
- ASHRAFI, Mahdi et al. Optimizing mortality prediction for ICU heart failure patients: leveraging XGBoost and advanced ML with the MIMIC-III database. **Informatics in Medicine Unlocked**, [S.l.], v. 40, art. 101300, 2024. Preprint. Disponível em: <https://arxiv.org/abs/2409.01685>. Acesso em: 3 mai. 2025.
- ATTIA, Z. I. et al. An artificial intelligence-enabled ECG algorithm for the identification of patients with atrial fibrillation during sinus rhythm. **The Lancet**, [S.l.], v. 394, n. 10201, p. 861–867, 2019. DOI: [https://doi.org/10.1016/S0140-6736\(19\)31721-0](https://doi.org/10.1016/S0140-6736(19)31721-0).
- BENBARKAT, H. et al. Application of the Seattle Heart Failure Model in patients >80 years of age enrolled in a tertiary care heart failure clinic. **American Journal of Cardiology**, v. 110, n. 11, p. 1663–1666, 2012. DOI: <https://doi.org/10.1016/j.amjcard.2012.07.034>.
- BREIMAN, L. Random Forests. **Machine Learning**, v. 45, n. 1, p. 5–32, 2001. DOI: <https://doi.org/10.1023/A:1010933404324>.
- CHE, Zhengping et al. Interpretable deep models for ICU outcome prediction. In: AMIA ANNUAL SYMPOSIUM, 2016, Chicago. **Proceedings**. Bethesda: American Medical Informatics Association, 2016. p. 371–380. Disponível em: <https://pmc.ncbi.nlm.nih.gov/articles/PMC5333206/>. Acesso em: 10 mai. 2025.
- CHEN, Tianqi; GUESTRIN, Carlos. XGBoost: A scalable tree boosting system. In: KDD '16: PROCEEDINGS OF THE 22ND ACM SIGKDD INTERNATIONAL CONFERENCE ON KNOWLEDGE DISCOVERY AND DATA MINING, San Francisco. **Proceedings [...]**. New York: ACM, 2016. p. 785–794. DOI: <https://doi.org/10.1145/2939672.2939785>.

CLEMENS, M. et al. The Seattle Heart Failure Model predicts survival in patients with cardiac resynchronization therapy: a validation study. **Journal of Cardiac Failure**, v. 18, n. 9, p. 682–687, 2012. DOI: <https://doi.org/10.1016/j.cardfail.2012.07.001>.

CORTES, C.; VAPNIK, V. Support-vector networks. **Machine Learning**, v. 20, n. 3, p. 273–297, 1995. DOI: <https://doi.org/10.1007/BF00994018>.

D'AGOSTINO, R. B. et al. General cardiovascular risk profile for use in primary care: the Framingham Heart Study. **Circulation**, v. 117, n. 6, p. 743–753, 2008. DOI: <https://doi.org/10.1161/CIRCULATIONAHA.107.699579>.

DAVIS, J.; GOADRIC, M. The relationship between Precision-Recall and ROC curves. In: INTERNATIONAL CONFERENCE ON MACHINE LEARNING, 23., 2006, Pittsburgh. **Proceedings [...]**. New York: ACM, 2006. p. 233–240. DOI: <https://doi.org/10.1145/1143844.1143874>.

ESTEVA, A. et al. Dermatologist-level classification of skin cancer with deep neural networks. **Nature**, v. 542, n. 7639, p. 115–118, 2017. DOI: <https://doi.org/10.1038/nature21056>.

FAWCETT, T. An introduction to ROC analysis. **Pattern Recognition Letters**, v. 27, n. 8, p. 861–874, 2006. DOI: <https://doi.org/10.1016/j.patrec.2005.10.010>.

GHASEMI, Fariba; SHARIFI, Seyedhassan. Heart failure prediction using support vector machine. **International Journal of Novel Research in Life Sciences**, [S.l.], v. 12, n. 1, p. 12–17, jan./fev. 2025. DOI: 10.5281/zenodo.14738316.

GOLDBERGER, A. L. et al. PhysioBank, PhysioToolkit, and PhysioNet. **Circulation**, v. 101, n. 23, p. E215–E220, 2000. DOI: 10.1161/01.cir.101.23.e215.

GUDIÑO-OCHOA, Alberto et al. Heart failure prediction: Exploratory analysis and modeling with XGBoost and deep neural networks. **Journal of Physiotherapy and Medical Technology**, v. 7, n. 18, p. 1-14, 2023. Disponível em: https://www.researchgate.net/publication/378367783_Heart_failure_prediction_Exploratory_analysis_and_modeling_with_XGBoost_and_deep_neural_networks. Acesso em: 3 mai. 2025.

HAJISHAH, Arian; TAVASOL, Arian; EPHELDE, Francisco. Evaluation of machine learning methods for prediction of heart failure mortality and readmission: meta-analysis. **BMC Cardiovascular Disorders**, [S.l.], v. 25, n. 1, p. 1–15, 2025. Disponível em: <https://bmccardiovascdisord.biomedcentral.com/articles/10.1186/s12872-025-04700-0>. Acesso em: 30 abr. 2025.

HEIDENREICH, P. A. 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. **Circulation**, v. 145, n. 18, p. e895–e1032, 2022. DOI: 10.1161/CIR.0000000000001063.

HERRERA SALAZAR, J. L. et al. Predicting heart failure: a comparative approach between artificial neural networks and support vector machines. **International Journal of Online and**

Biomedical Engineering (iJOE), [S.l.], v. 21, n. 1, p. 41–55, 2025. DOI: 10.3991/ijoe.v21i01.52869.

HOSMER, David W.; LEMESHOW, Stanley; STURDIVANT, Rodney X. **Applied Logistic Regression**. 3. ed. Hoboken: Wiley, 2013.

ISLAM, Md. Rabiul et al. CardioTabNet: **A Novel Hybrid Transformer Model for Heart Disease Prediction using Tabular Medical Data**. 2025. Preprint. Disponível em: <https://arxiv.org/abs/2503.17664>. Acesso em: 3 mai. 2025.

JIANG, F. et al. Artificial intelligence in healthcare: past, present and future. **Stroke and Vascular Neurology**, v. 2, n. 4, p. 230-243, 2017. DOI: 10.1136/svn-2017-000101.

JOHNSON, A. E. W. et al. MIMIC-III, a freely accessible critical care database. **Scientific Data**, v. 3, art. 160035, 2016. DOI: <https://doi.org/10.1038/sdata.2016.35>.

KOEHLER, F. et al. Artificial intelligence based real-time prediction of imminent heart failure hospitalisation in patients undergoing non-invasive telemedicine. **Frontiers in Cardiovascular Medicine**, v. 11, art. 1457995, 2024. DOI: <https://doi.org/10.3389/fcvm.2024.1457995>.

LEVY, W. C. et al. The Seattle Heart Failure Model: Prediction of survival in heart failure. **Circulation**, v. 113, n. 11, p. 1464–1470, 2006. DOI: <https://doi.org/10.1161/CIRCULATIONAHA.105.584102>.

MARTÍN-YEBRA, A. et al. The MUSIC Database: Sudden Cardiac Death in Heart Failure Patients. In: **COMPUTING IN CARDIOLOGY**, 51., 2024, Karlsruhe. **Anais [...]**. [S.l.]: [s.n.], 2024. v. 51. DOI: 10.22489/CinC.2024.355.

MIOTTO, R. et al. Deep Patient: An Unsupervised Representation to Predict the Future of Patients from the Electronic Health Records. **Scientific Reports**, v. 6, art. 26094, 2016. DOI: <https://doi.org/10.1038/srep26094>.

OBERMEYER, Z. et al. Dissecting racial bias in an algorithm used to manage the health of populations. **Science**, v. 366, n. 6464, p. 447–453, 2019. DOI: <https://www.science.org/doi/10.1126/science.aax2342>.

OLIVEIRA, G. M. M. et al. Cardiovascular Statistics – Brazil 2023. **Arquivos Brasileiros de Cardiologia**, v. 121, n. 1, p. e20230653, 2024. Disponível em: <https://www.scielo.br/j/abc/a/jzFMcdN5y3w6CtjVgdJdSdR/?lang=en>. Acesso em: 2 abr. 2025.

PANAHAZAR, M. et al. Using EHRs and machine learning for heart failure survival analysis. In: GEORGIOU, A.; SARKAR, I. N.; MAZZONCINI DE AZEVEDO MARQUES, P. (Ed.). **MEDINFO 2015: eHealth-Enabled Health – Proceedings of the 15th World Congress on Health and Biomedical Informatics**. Amsterdam: IOS Press, 2015. p. 40–44. DOI: <https://doi.org/10.3233/978-1-61499-564-7-40>.

POCOCK, S. J. et al. Predicting survival in heart failure: a risk score based on 39 372 patients. **European Heart Journal**, v. 34, n. 19, p. 1404–1413, 2013. DOI: <https://doi.org/10.1093/eurheartj/ehs337>.

PONIKOWSKI, P. et al. 2016 ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure: The Task Force for the diagnosis and treatment of acute and chronic heart failure of the European Society of Cardiology (ESC). **European Heart Journal**, v. 37, n. 27, p. 2129–2200, 2016. DOI: <https://doi.org/10.1093/eurheartj/ehw128>.

POWERS, D. M. W. Evaluation: From Precision, Recall and F-Measure to ROC, Informedness, Markedness & Correlation. **Journal of Machine Learning Technologies**, v. 2, n. 1, p. 37–63, 2011.

RAO, S. et al. An explainable Transformer-based deep learning model for the prediction of incident heart failure. **IEEE Journal of Biomedical and Health Informatics**, v. 26, n. 7, p. 3362–3372, 2022. DOI: 10.1109/JBHI.2022.3148820.

REGOLI, F. et al. Validation of Seattle Heart Failure Model for mortality risk prediction in patients treated with cardiac resynchronization therapy. **European Journal of Heart Failure**, v. 15, n. 3, p. 302–307, 2013. DOI: <https://doi.org/10.1093/eurjhf/hfs162>.

SAITO, T.; REHMSMEIER, M. The Precision-Recall Plot Is More Informative than the ROC Plot When Evaluating Binary Classifiers on Imbalanced Datasets. **PLOS ONE**, v. 10, n. 3, art. e0118432, 2015. DOI: <https://doi.org/10.1371/journal.pone.0118432>.

SAMSON, R. et al. Clinical phenotypes in heart failure with preserved ejection fraction. **Journal of the American Heart Association**, v. 5, n. 1, p. e002477, 2016. DOI: 10.1161/JAHA.115.002477.

SHIRAISHI, Y. et al. Validation and recalibration of Seattle Heart Failure Model in Japanese acute heart failure patients. **Journal of Cardiac Failure**, v. 25, n. 7, p. 561–567, 2019. DOI: <https://doi.org/10.1016/j.cardfail.2018.07.463>.

SHORTLIFFE, E. H. **Computer-Based Medical Consultations: MYCIN**. New York: Elsevier, 1976.

SIQUEIRA, A. C. et al. Reduction of mortality and predictions for acute myocardial infarction, stroke, and heart failure in Brazil until 2030. **Scientific Reports**, [S.l.], v. 10, n. 1, p. 1–10, 2020. Disponível em: <https://www.nature.com/articles/s41598-020-73070-8>. Acesso em: 2 abr. 2025.

TOHYAMA, T. et al. Machine learning-based model for predicting 1-year mortality of hospitalized patients with heart failure. **ESC Heart Failure**, v. 8, n. 5, p. 4077–4085, 2021. DOI: <https://doi.org/10.1002/ehf2.13556>.

UNIVERSITY OF WASHINGTON. **Seattle Heart Failure Model**. [2022]. Disponível em: <https://depts.washington.edu/shfm/?width=1920&height=1080>. Acesso em: 30 abr. 2025.

VASWANI, A. et al. Attention is all you need. In: CONFERENCE ON NEURAL INFORMATION PROCESSING SYSTEMS (NeurIPS), 30., 2017, Long Beach, CA.

Proceedings [...]. Red Hook: Curran Associates, 2017. p. 5998–6008. Disponível em: <https://arxiv.org/abs/1706.03762>. Acesso em: 30 abr. 2025.

VEDANA, A. B. et al. Inteligência artificial na medicina diagnóstica. **Brazilian Journal of Implantology and Health Sciences**, [S.l.], v. 6, n. 11, p. 765–794, 2024. Disponível em: <https://bjih.s.emnuvens.com.br/bjih/article/download/4235/4316/9407>. Acesso em: 25 abr. 2025.

WANG, Jing. Heart failure prediction with machine learning: a comparative study. In: SIGNAL PROCESSING AND COMPUTER SCIENCE CONFERENCE (SPCS 2021), 2., 2021. **Journal of Physics: Conference Series**, [S.l.], v. 2031, p. 012068, 2021. DOI: <https://doi.org/10.1088/1742-6596/2031/1/012068>.

WORLD HEART FEDERATION. Deaths from cardiovascular disease surged 60% globally over the last 30 years. 2023. Disponível em: <https://world-heart-federation.org/news/deaths-from-cardiovascular-disease-surged-60-globally-over-the-last-30-years-report/>. Acesso em: 30 mar. 2025.

YANG, Li et al. Study of cardiovascular disease prediction model based on random forest in eastern China. **Scientific Reports**, v. 10, art. 5245, 2020. DOI: 10.1038/s41598-020-62133-5.

ZELL, Andreas. **Simulation Neuronaler Netze**. Munich: Oldenbourg Verlag, 2003.

ZHANG, J. et al. Fully automated echocardiogram interpretation in clinical practice: feasibility and diagnostic accuracy. **Circulation**, v. 138, n. 16, p. 1623–1635, 2018. DOI: <https://doi.org/10.1161/CIRCULATIONAHA.118.034338>.