UNIVERSIDADE FEDERAL DA FRONTEIRA SUL
CAMPUS CHAPECÓ
CURSO DE CIÊNCIA DA COMPUTAÇÃO

LUAN BORTOLI

PREDIÇÃO DE INFARTO AGUDO DO MIOCÁRDIO E INSUFICIÊNCIA CARDÍACA COM APRENDIZADO DE MÁQUINA - UMA ANÁLISE COMPARATIVA NO MIMIC-IV

CHAPECÓ
2026

LUAN BORTOLI

PREDIÇÃO DE INFARTO AGUDO DO MIOCÁRDIO E INSUFICIÊNCIA CARDÍACA COM APRENDIZADO DE MÁQUINA - UMA ANÁLISE COMPARATIVA NO MIMIC-IV

Trabalho de Conclusão de Curso apresentado ao Curso de Ciência da Computação da Universidade Federal da Fronteira Sul (UFFS), como requisito para obtenção do título de Bacharel em Ciência da Computação.

Orientador: Prof. Dr. Giancarlo Dondoni Salton

CHAPECÓ
2026

Bortoli, Luan

PREDIÇÃO DE INFARTO AGUDO DO MIOCÁRDIO E INSUFICIÊNCIA CARDÍACA COM APRENDIZADO DE MÁQUINA – UMA ANÁLISE COMPARATIVA NO MIMIC-IV / Luan Bortoli – 2026.

20 f.

Orientador: Dr. Giancarlo Dondoni Salton

Trabalho de Conclusão de Curso (Graduacão) – Universidade Federal da Fronteira Sul, Curso de Ciência da Computação, Chapecó, SC, 2026.

1. Aprendizado de Máquina 2. Doenças Cardiovasculares 3. MIMIC-IV 4. Interpretabilidade 5. Métricas I. Salton, Dr. Giancarlo Dondoni , orient. II. Universidade Federal da Fronteira Sul. III. Título.

LUAN BORTOLI

PREDIÇÃO DE INFARTO AGUDO DO MIOCÁRDIO E INSUFICIÊNCIA CARDÍACA COM APRENDIZADO DE MÁQUINA - UMA ANÁLISE COMPARATIVA NO MIMIC-IV

Trabalho de Conclusão de Curso apresentado ao Curso de Ciência da Computação da Universidade Federal da Fronteira Sul (UFFS), como requisito para obtenção do título de Bacharel em Ciência da Computação.

Orientador: Prof. Dr. Giancarlo Dondoni Salton

Este Trabalho de Conclusão de Curso foi avaliado e aprovado pela banca avaliadora em: 26/06/2026.

BANCA AVALIADORA

Documento assinado digitalmente
gov.br
GIANCARLO DONDONI SALTON
Data: 06/07/2026 17:02:35-0300
Verifique em https://validar.iti.gov.br

Dr. Giancarlo Dondoni Salton - UFFS

Documento assinado digitalmente
gov.br
DENIO DUARTE
Data: 07/07/2026 08:29:06-0300
Verifique em https://validar.iti.gov.br

Prof. Dr. Denio Duarte - UFFS

Documento assinado digitalmente
gov.br
JOAO VICTOR GARCIA DE SOUZA
Data: 07/07/2026 14:09:07-0300
Verifique em https://validar.iti.gov.br

Prof. Me. João Victor Garcia de Souza - UFFS

DEDICATÓRIA

Dedico este trabalho à minha mãe Marinês, a meu pai Valdair, às minhas irmãs Franciane e Carla, minha sobrinha Amalle e meu cunhado Marcos Augusto, pelo incentivo, apoio e carinho durante a minha vida acadêmica. Cada um, à sua maneira, contribuiu para que esta conquista fosse possível.

AGRADECIMENTOS

Agradeço, primeiramente, a Deus, por ter me concedido o dom da vida, por me fortalecer diante das dificuldades e por me dar saúde para alcançar este sonho.

Aos meus pais, Marinês e Valdair, agradeço por todos os ensinamentos e valores transmitidos a mim, os quais não poderiam ser encontrados em nenhum outro lugar. Sou grato pela dedicação e pelo apoio em todas as etapas da minha vida, pois, sem eles, este sonho não teria se concretizado.

Às minhas irmãs, Franciane e Carla, agradeço por todo o apoio em cada passo que dei e ainda darei. Nunca deixaram de confiar em mim e em meu potencial, e sempre estiveram ao meu lado em todos os momentos da minha vida.

À minha sobrinha Amalle, que chegou recentemente à minha vida e, mesmo ainda tão pequena, já trouxe alegria, amor e um novo significado para os meus dias. Que este trabalho sirva como exemplo de que, com dedicação, fé e perseverança, é possível alcançar nossos sonhos.

Ao meu cunhado, Marcos Augusto, agradeço pelo apoio, incentivo e por também fazer parte desta caminhada.

Aos meus amigos, de dentro e fora da universidade, e aos colegas de curso, agradeço pela amizade, pelo apoio e pelo companheirismo ao longo desta trajetória.

Ao meu professor e orientador, Giancarlo, agradeço por estar sempre disposto a ajudar nos momentos em que precisei e por acreditar em meu potencial para o desenvolvimento deste projeto.

Ao meu líder do SENAI/SC - Chapecó, Fábio Junior Bet, agradeço pelo apoio, pela confiança e pela compreensão ao longo desta etapa. Sua flexibilidade foi fundamental para que eu pudesse conciliar minhas responsabilidades profissionais com a conclusão da minha formação acadêmica.

A todos os professores e demais servidores da UFFS, que contribuíram direta e indiretamente para minha formação e para meu crescimento profissional durante os cinco anos em que estive nesta instituição.

A todas as pessoas que, de alguma forma, fizeram parte da minha vida acadêmica, deixo minha gratidão e meu muito obrigado, de todo o coração.

ISSN:

Predição de Infarto Agudo do Miocárdio e Insuficiência Cardíaca com Aprendizado de Máquina - uma Análise Comparativa no MIMIC-IV

Prediction of Acute Myocardial Infarction and Heart Failure with Machine Learning - A Comparative Analysis in MIMIC-IV

Luan Bortoli [Universidade Federal da Fronteira Sul | *luan.bortoli@estudante.uffs.edu.br*]
Giancarlo Dondoni Salton [Universidade Federal da Fronteira Sul | *gian@uffs.edu.br*]

Universidade Federal da Fronteira Sul, Rodovia SC 484, km 02, Bairro Fronteira Sul, Chapecó, SC, CEP 89815-899. Tel. (49) 2049-2600 CNPJ 11.234.780/0007-46, Brasil

Resumo. As doenças cardiovasculares figuram entre as principais causas de morbidade e mortalidade, com destaque para o Infarto Agudo do Miocárdio e a Insuficiência Cardíaca, cuja gravidade clínica demanda identificação precoce de pacientes em risco. Neste trabalho, foram desenvolvidos e comparados quatro classificadores supervisionados (Regressão Logística, Support Vector Machine, Árvore de Decisão e K-Nearest Neighbors) para a predição de DCV a partir de dados do MIMIC-IV. Utilizaram-se 546.028 admissões, com variável-alvo binária definida por códigos CID-9/CID-10, das quais 16,3% corresponderam a casos positivos. As variáveis incluíram dados demográficos, sinais vitais e exames laboratoriais; o pré-processamento envolveu limpeza de valores incompatíveis, imputação pela mediana, padronização e divisão estratificada em treino (70%) e teste (30%). Os modelos foram ajustados por busca em grade com validação cruzada estratificada e avaliados por acurácia, precisão, Recall, F1-Score e AUC-ROC, além de matrizes de confusão e análise de interpretabilidade. A Árvore de Decisão apresentou o melhor equilíbrio, com AUC-ROC de 0,868, Recall de 0,788 e menor número de falsos negativos (5.664), enquanto o SVM obteve desempenho próximo, porém a maior custo computacional. A Regressão Logística mostrou desempenho consistente e interpretabilidade pelos coeficientes, e o KNN, embora com maior acurácia, exibiu Recall baixo, evidenciando limitações da métrica em bases desbalanceadas. As análises indicaram idade, ureia, troponina T e NT-proBNP como variáveis mais relevantes, reforçando a plausibilidade clínica dos padrões aprendidos.

Abstract. Cardiovascular diseases are among the leading causes of morbidity and mortality, particularly acute myocardial infarction and heart failure, whose clinical severity calls for early identification of at-risk patients. This study developed and compared four supervised classifiers (Logistic Regression, Support Vector Machine, Decision Tree, and K-Nearest Neighbors) to predict CVD using MIMIC-IV data. A total of 546,028 hospital admissions were analyzed, with a binary outcome defined by ICD-9/ICD-10 codes; 16.3% were positive cases. Features comprised demographic variables, vital signs, and laboratory tests; preprocessing included removal of physiologically implausible values, median imputation, standardization, and a stratified split into training (70%) and testing (30%). Models were tuned via grid search with stratified cross-validation and assessed using accuracy, precision, recall, F1-score, and AUC-ROC, complemented by confusion matrices and interpretability analyses. The decision tree achieved the best overall balance (AUC-ROC 0.868; recall 0.788; lowest false negatives, 5,664), while the SVM was competitive at higher computational cost. Logistic regression provided stable performance with coefficient-based interpretability, and KNN, despite the highest accuracy, showed low recall, underscoring the limits of accuracy under class imbalance. Interpretability results highlighted age, urea, troponin T, and NT-proBNP as the most influential variables, supporting the clinical plausibility of the learned patterns.

Palavras-chave: Aprendizado de Máquina, Doenças Cardiovasculares, MIMIC-IV, Interpretabilidade, Métricas

Keywords: Machine Learning, Cardiovascular Diseases, MIMIC-IV, Interpretability, Metrics

1 Introdução

As doenças cardiovasculares (DCV) representam uma das principais causas de morbidade e mortalidade em escala mundial, constituindo um desafio relevante para os sistemas de saúde. Entre essas condições, o Infarto Agudo do Miocárdio (IAM) e a Insuficiência Cardíaca (IC) destacam-se pela gravidade clínica, pela elevada demanda por recursos hospitalares e pela necessidade de identificação precoce de pacientes em risco [Organização Mundial da Saúde, 2025]. A detecção antecipada desses quadros pode contribuir para decisões clínicas mais oportunas, melhor direcionamento de condutas e redução de complicações associadas ao agravamento do estado do paciente.

Com o avanço dos prontuários eletrônicos e a crescente disponibilidade de bases clínicas estruturadas, tornou-se possível aplicar técnicas computacionais capazes de explorar grandes volumes de dados hospitalares. Nesse contexto, algoritmos de aprendizado de máquina têm sido amplamente investigados como ferramentas de apoio à decisão clínica, especialmente por sua capacidade de identificar padrões complexos em variáveis demográficas, laboratoriais e fisiológicas [Paixão *et al*., 2022; Singh *et al*., 2024]. Diferentemente de abordagens tradicionais baseadas exclusivamente em regras fixas, esses modelos podem aprender relações presentes nos dados e auxiliar na classificação de pacientes com maior probabilidade de apresentar determinados desfechos.

Apesar do crescimento das pesquisas envolvendo aprendizado de máquina na área da saúde, ainda existem desafios importantes para sua aplicação em problemas cardiovascula-

res. Muitos estudos utilizam bases de pequeno porte, como o conjunto do UCI Machine Learning Repository com 303 indivíduos [Costalat and Tavares, 2022] ou a base heart.csv do Kaggle com 1.025 pacientes [Oliveira *et al*., 2023], o que pode comprometer a generalização dos resultados. Além disso, parte dos trabalhos concentra-se em métricas globais como acurácia, sem considerar adequadamente o impacto do desbalanceamento entre classes e a relevância clínica dos falsos negativos. Em problemas de triagem, deixar de identificar pacientes com risco cardiovascular pode ser mais crítico do que classificar incorretamente um paciente sem a condição.

Outro aspecto relevante diz respeito à interpretabilidade dos modelos. Embora algoritmos mais complexos possam alcançar alto desempenho preditivo, sua adoção em contextos clínicos exige que os resultados sejam compreensíveis e auditáveis, o que torna necessário avaliar não apenas a capacidade de classificação, mas também se as variáveis que influenciam as predições possuem coerência com o conhecimento médico estabelecido[Ferreira *et al*., 2025].

Diante desse contexto, este trabalho tem como objetivo desenvolver e avaliar modelos de aprendizado de máquina para a predição de DCV, com foco em IAM e IC, utilizando dados clínicos reais do MIMIC-IV. Foram comparados quatro algoritmos supervisionados de classificação, sendo eles: Regressão Logística, Support Vector Machine, Árvore de Decisão e K-Nearest Neighbors, avaliados por métricas de acurácia, precisão, Recall, F1-Score e AUC-ROC, complementadas pela análise de matrizes de confusão e técnicas de interpretabilidade.

A escolha do MIMIC-IV justifica-se por sua ampla utilização em pesquisas científicas na área da saúde e por reunir registros clínicos anonimizados de pacientes hospitalizados, contemplando diagnósticos, sinais vitais, exames laboratoriais e informações demográficas. Essa base permite avaliar os algoritmos em um cenário próximo da prática hospitalar real, caracterizado pela heterogeneidade dos dados, valores ausentes e desbalanceamento entre classes.

A principal contribuição deste estudo consiste na comparação experimental de algoritmos clássicos de aprendizado supervisionado em uma base clínica ampla, considerando não apenas o desempenho global, mas também a capacidade de identificar corretamente pacientes com DCV e a interpretabilidade dos fatores associados às predições.

O artigo está organizado da seguinte forma. Na Seção 2 apresenta os trabalhos relacionados. Na Seção 3 descreve a metodologia. Na Seção 4 apresenta os resultados e a discussão, e, por fim, na seção 5 traz as conclusões e perspectivas de continuidade.

2 Trabalhos Relacionados

Paixão *et al*. [2022] apresentam uma revisão sistemática sobre o uso de Machine Learning na medicina, com o objetivo de introduzir seus conceitos fundamentais aos profissionais de saúde e discutir sua aplicabilidade clínica, com ênfase particular nas inovações em cardiologia. Nesse sentido, o artigo não se limita à proposição de um modelo específico, mas consolida o conhecimento sobre técnicas e modelos de aprendizado, como supervisionados e não supervisionados, servindo como fundamentação teórica para aplicações prognósticas e diagnósticas já documentadas na literatura.

A metodologia empregada pelos autores caracteriza-se por uma revisão sistemática da literatura, utilizando as bases de dados *PubMed* e *Medline* e selecionando estudos prospectivos e retrospectivos mediante a avaliação de dois investigadores. O artigo descreve as etapas dos algoritmos de *machine learning*, como o pré-processamento, o treinamento e a avaliação, e revisa os principais algoritmos utilizados, como redes neurais artificiais, regressão logística, árvore de decisão, *random forests*, redes *bayesianas*, *deep learning* e *support vector machines*.

Os resultados da revisão demonstraram um crescimento exponencial na produção científica sobre *machine learning* na medicina, com um aumento significativo no número de artigos publicados ao longo dos anos, evidenciando a relevância e o interesse crescente da área. O estudo destaca a aplicabilidade do ML em diversas especialidades médicas, com ênfase em cardiologia, onde as técnicas de ML tem sido utilizadas com sucesso no prognóstico, diagnóstico e tratamento das doenças. Os autores ressaltam que os modelos de ML, ao analisarem grandes volumes de dados e variáveis não tradicionais, identificam pacientes em risco que poderiam ser negligenciados pelas ferramentas convencionais de estratificação de riscos.

O estudo possui natureza didática e abrangente, alinhada ao objetivo de introduzir o machine learning a profissionais de saúde. O artigo aborda conceitos de ML, como a diferença entre aprendizado supervisionado e não supervisionado e o funcionamento das redes neurais, em linguagem acessível e com rigor técnico. Essa característica permite que pesquisadores iniciando estudos ou aplicações de ML na área da saúde utilizem o texto como referência inicial, com panorama técnico e exemplos práticos de aplicação voltados à cardiologia.

Costalat and Tavares [2022] apresentam uma alternativa para a detecção de doenças cardiovasculares baseada em inteligência artificial, com o objetivo de avaliar a eficácia de diferentes algoritmos de aprendizado supervisionado na predição de riscos cardíacos a partir de indicadores de saúde, como idade, sexo, glicose e colesterol. A relevância do estudo consiste no fato de que as doenças cardiovasculares são uma das principais causas de morte no mundo, o que torna essencial identificar a melhor alternativa para a construção de um preditor eficaz para auxiliar na prevenção e no manejo de pacientes em alto risco.

A metodologia empregada no estudo baseou-se na utilização do *Heart Disease Data Set*, disponível no Repositório *UCI Machine Learning*. Este banco de dados contém 14 atributos utilizados no estudo para um total de 303 indivíduos. Foram implementados e comparados quatro algoritmos de aprendizado supervisionado: *K-Nearest Neighbours*, árvore de decisão, regressão logística e o *voting classifier*. A implementação foi realizada utilizando a linguagem de programação Python e a biblioteca *Scikit-learn*. A avaliação dos modelos foi realizada por meio de uma metodologia de teste que inclui a análise da acurácia, sensibilidade e especificidade, com base na matriz de confusão de cada classificador.

Os resultados da comparação demonstram que o *voting classifier* alcançou a maior acurácia entre os modelos testa-

dos, atingindo um valor de 82,89%. Isoladamente, os algoritmos de regressão logística e KNN apresentaram o mesmo desempenho, ambos com acurácia de 80,26%. Já a árvore de decisão obteve o desempenho mais baixo, com acurácia de 73,68%. O desempenho superior do *voting classifier* sugere que a combinação de diferentes modelos de ML pode mitigar as fraquezas individuais de cada algoritmo, resultando em um preditor mais robusto e confiável para a detecção de doenças cardiovasculares.

Vale destacar a validação da abordagem de combinação de modelos proposta pelos autores. Ao demonstrar que o *voting classifier* superou as técnicas de desempenho de um único algoritmo, o trabalho reforça a ideia de que a combinação de múltiplos algoritmos é uma estratégia viável e superior para mitigar limitações de modelos únicos. Essa conclusão contribui para a área de diagnóstico médico auxiliado por ML, ao indicar uma possibilidade de aumentar a confiabilidade e a capacidade preditiva na identificação de pacientes em risco.

Oliveira *et al*. [2023] apresenta uma análise comparativa de algoritmos de *machine learning* na tarefa de classificar o estado de saúde de pacientes em relação à predisposição para doenças cardiovasculares. Reconhecendo que as doenças cardiovasculares são a principal causa de morte e são frequentemente silenciosas, os autores propuseram o uso de ML para agilizar e simplificar o diagnóstico prévio, reduzindo a intervenção humana em análises complexas. Para isso, o estudo aplicou e avaliou três modelos distintos de classificação em um conjunto de dados públicos de pacientes, buscando identificar o modelo mais eficiente para a classificação binária de presença ou ausência de doença cardiovascular.

Metodologicamente, o estudo implementou os algoritmos K-Nearest Neighbor, árvore de decisão e rede neural multicamada perceptron (MLP), utilizando a linguagem *python*, o ambiente *Google Colab* e as bibliotecas *scikit-learn* e *pandas* para a implementação e manipulação dos dados. A base de dados pública utilizada foi a *heart.csv*, disponibilizada na plataforma Kaggle, sobre doenças cardiovasculares, contendo informações de 1025 pacientes e 13 características biológicas, como idade, sexo, tipo de dor no peito e colesterol. Antes da aplicação dos modelos, os dados passaram por um pré-processamento que incluiu a normalização e a utilização da técnica de validação cruzada *K-Fold* (com 3 *folds*) para mitigar o risco de *overfitting*. A avaliação de desempenho foi realizada com base em métricas estatísticas como *Precision, Recall, F1-score e Support*.

Os resultados da análise comparativa demonstraram que a árvore de decisão foi o algoritmo com o melhor desempenho, apresentando uma precisão superior a 0,96 e um *F1-score* médio de 0,97 para ambas as classes (saudável e com doença cardiovascular). O algoritmo KNN também obteve resultados consideráveis, com um *F1-score* médio de 0,88, indicando sua adequação para problemas de classificação mais simplificados. Por outro lado, a Rede MLP apresentou desempenho mais baixo entre os três, com um *F1-score* médio de 0,80, apesar de utilizar técnicas de otimização como o *backpropagation*.

Vale ressaltar que o desempenho inferior da Rede MLP neste cenário específico está associado às características da base de dados utilizada, uma vez que arquiteturas baseadas em *backpropagation* geralmente exigem um volume de dados massivo e um maior custo computacional para calibrar seus pesos adequadamente. Assim, embora a árvore de decisão tenha se provado o algoritmo mais robusto para o conjunto de dados atual, a exploração completa do potencial de modelos mais complexos, como as redes neurais, depende da ampliação dos registros clínicos em pesquisas futuras, garantindo uma diversidade amostral que favoreça o aprendizado profundo.

Ferreira *et al*. [2025] apresentam um mapeamento sistemático da literatura sobre o uso de algoritmos de machine learning na previsão de doenças cardiovasculares, com o objetivo de sintetizar o conhecimento sobre as técnicas eficazes para o diagnóstico precoce dessas condições. Os autores realizaram uma investigação abrangente sobre diversos artigos que aplicam técnicas de ML, com foco na comparação de desempenho entre modelos de *ensemble*, *Deep Learning* e algoritmos tradicionais, como *Random Forest, Support Vector Machines* e redes neurais. O objetivo foi determinar quais metodologias se destacam em termos de eficiência e acurácia para o diagnóstico precoce de doenças cardiovasculares, um fator crucial para a redução da alta taxa de mortalidade global associada a essas condições.

A metodologia empregada para a realização deste estudo foi o mapeamento sistemático da literatura. O processo envolveu a definição de questões de pesquisa específicas, a formulação de termos de busca detalhados e a aplicação de critérios de inclusão e exclusão bem definidos. A busca foi realizada nas bases de dados *IEEE Xplore* e *Scopus*, resultando em uma seleção inicial de 90 trabalhos. Após uma filtragem em três etapas (busca inicial, filtragem por título e resumo, e leitura completa), apenas 14 artigos foram selecionados para a análise final, garantindo que apenas estudos que comparavam a eficiência de modelos de ML na previsão de doenças cardiovasculares fossem considerados.

Os resultados do mapeamento sistemático revelaram que os modelos híbridos e as técnicas de *Deep Learning* apresentam os mais altos níveis de acurácia na previsão de doenças cardiovasculares. Modelos de *ensemble*, como o *stacking* e o uso de *Gradient Boosting* como *meta-learner*, demonstraram ser particularmente promissores. Notavelmente, um modelo híbrido de *Hybrid Deep Neural Networks* (HDNNs) que combinou *Convolutional Neural Network* (CNN) e *Long Short-Term Memory* (LSTM) alcançou acurácia de 98,86%. Entre os algoritmos tradicionais, o *Random Forest* se manteve como um dos mais citados e eficientes, reforçando sua robustez e aplicabilidade na área.

Apesar da alta precisão alcançada, o trabalho destaca barreiras significativas, como a necessidade de maior qualidade e disponibilidade de dados, a importância da seleção de características e, principalmente, a questão da interpretabilidade. Modelos mais complexos, como as *Deep Neural Networks*, são frequentemente vistos como "caixas-pretas", o que dificulta sua adoção na prática médica. O estudo conclui que é essencial buscar uma maior integração entre a alta acurácia, a interpretabilidade e a usabilidade para que o *Machine Learning* possa ser amplamente incorporado no contexto clínico de previsão de doenças cardiovasculares.

Teoh *et al*. [2026] realizaram um estudo comparativo voltado à predição de insuficiência cardíaca a partir de regis-

tros eletrônicos de saúde, avaliando o desempenho de diferentes algoritmos de aprendizado de máquina e de aprendizado profundo. Os autores partem da constatação de que a insuficiência cardíaca representa uma das principais causas de morbidade e mortalidade no mundo, conforme apontado pela literatura da área, e que a identificação precoce dos pacientes em risco permanece um desafio na prática clínica. O propósito do trabalho foi determinar quais abordagens computacionais oferecem maior eficácia como apoio à decisão médica, confrontando técnicas tradicionais de Machine Learning com modelos mais complexos de Deep Learning sobre um mesmo conjunto de dados.

Para isso, o estudo utilizou as bases públicas MIMIC-IV e MIMIC-IV-ED, das quais foram extraídos 17.892 pacientes adultos, sendo 9.373 com diagnóstico de insuficiência cardíaca e 8.519 no grupo controle, com a condição identificada por códigos CID-9 e CID-10. O pré-processamento envolveu o tratamento de valores discrepantes por meio das cercas de Tukey, técnica que identifica valores muito distantes da maioria dos registros. Os dados ausentes foram imputados com o algoritmo K-Nearest Neighbors, e as variáveis foram padronizadas pela transformação Z-score. A seleção de atributos seguiu as diretrizes clínicas das sociedades AHA/ACC/HFSA e ESC.

Um cuidado metodológico relevante foi a divisão dos dados em treino, validação e teste na proporção de 70:20:10 antes da imputação, de modo a evitar o vazamento de informações entre os conjuntos. Além das variáveis estruturadas, as notas de admissão foram codificadas pelo modelo BioClinicalBERT, gerando representações de 768 dimensões que foram reduzidas a 32 componentes principais via PCA antes de serem combinadas aos dados tabulares. Os algoritmos avaliados incluíram Regressão Logística, Árvore de Decisão, Naïve Bayes, Adaptive Boosting e Random Forest, além dos modelos Multilayer Perceptron, Convolutional Neural Network e Dense Neural Network, com desempenho medido por acurácia, precisão, recall, F1-score e AUROC, todos acompanhados de intervalos de confiança estimados por bootstrap.

O Random Forest obteve o melhor desempenho entre todos os modelos testados, com acurácia de 89,94% e AUROC de 0,9659, superando tanto os demais algoritmos tradicionais quanto as arquiteturas de aprendizado profundo. Entre os modelos de Deep Learning, o Multilayer Perceptron foi o que apresentou melhor resultado, ainda assim com desempenho inferior ao do Random Forest. Os autores conduziram dois estudos de ablação: o primeiro mostrou que a inclusão das representações textuais do BioClinicalBERT melhorou de forma expressiva o desempenho do Random Forest, ainda que com ganhos mais discretos para os modelos neurais; o segundo, uma análise de sensibilidade, ampliou o grupo controle com pacientes portadores de outras condições cardíacas e verificou que o modelo manteve alto poder de discriminação mesmo nesse cenário mais ambíguo. A análise dos atributos apontou diabetes, frequência respiratória, índice de massa corporal e idade como os preditores mais influentes, em consonância com fatores de risco já consolidados na literatura.

Entre os pontos fortes do trabalho está a proximidade entre o cenário experimental e a realidade clínica, sustentada pelo uso de uma base ampla e de origem hospitalar como a MIMIC-IV, pelo rigor no pré-processamento e pela comparação sistemática entre múltiplas técnicas. O relato dos intervalos de confiança por bootstrap e a realização dos estudos de ablação conferem solidez estatística que diferencia o estudo de trabalhos baseados em conjuntos pequenos, como o UCI Cleveland, criticados pelos próprios autores. Como limitação, os autores reconhecem a ausência de validação externa em bases de outras instituições, o que mantém em aberto a generalização dos resultados entre diferentes contextos assistenciais, ainda que a análise de sensibilidade já evidencie robustez frente a variações na composição da coorte. Cabe observar ainda que o Support Vector Machine e o K-Nearest Neighbors, embora reconhecidos pelos autores na revisão da literatura e citados em estudos correlatos com bom desempenho, não foram incluídos na comparação experimental de classificadores. Essa lacuna reforça a pertinência de investigações que avaliem tais algoritmos sobre a mesma base, contribuindo para o desenvolvimento de ferramentas preditivas mais abrangentes e aplicáveis à prática clínica.

A Tabela 1 sintetiza os principais aspectos dos trabalhos relacionados, considerando a base de dados utilizada, os modelos avaliados, as variáveis clínicas consideradas e a relação de cada estudo com a presente pesquisa. Essa comparação permite evidenciar tanto a proximidade metodológica com estudos anteriores quanto as diferenças relacionadas ao uso de uma base hospitalar ampla e à composição do conjunto de variáveis adotado.

Além da comparação entre algoritmos, os trabalhos relacionados também evidenciam a importância da seleção das variáveis clínicas utilizadas na predição de doenças cardiovasculares. Estudos baseados em bases menores, como os conjuntos derivados do repositório UCI, frequentemente utilizam atributos demográficos e clínicos gerais, como idade, sexo, colesterol, glicose e indicadores de condição cardíaca. Já estudos baseados em registros hospitalares mais amplos, como os que utilizam o MIMIC-IV, permitem incorporar variáveis adicionais provenientes de exames laboratoriais, sinais vitais e biomarcadores específicos. Nesse sentido, o presente trabalho aproxima-se das abordagens que exploram registros clínicos reais de larga escala, ao combinar variáveis demográficas, sinais vitais, marcadores de função renal, eletrólitos, perfil lipídico e biomarcadores cardíacos, como troponina T e NT-proBNP. Essa composição amplia a caracterização clínica das admissões hospitalares e permite avaliar não apenas o desempenho dos modelos, mas também a coerência das variáveis identificadas como relevantes pelos métodos de interpretabilidade.

Os trabalhos analisados convergem ao demonstrar a eficácia de algoritmos de aprendizado de máquina na predição de doenças cardiovasculares, mas apresentam lacunas que delimitam o espaço desta pesquisa. Parte expressiva dos estudos recorre a conjuntos de dados públicos de pequeno porte, com poucas centenas de registros, o que limita a robustez estatística e a capacidade de generalização dos modelos; já os trabalhos que adotam bases hospitalares de larga escala tendem a concentrar-se em um único algoritmo de melhor desempenho ou em arquiteturas mais complexas, sem oferecer uma comparação sistemática entre os classificadores supervisionados clássicos. Observa-se também que, embora o Sup-

Tabela 1. Comparação entre os trabalhos relacionados e este trabalho

Trabalho	Base de dados	Modelos avaliados	Variáveis consideradas	Relação com este trabalho
Paixão *et al*. [2022]	Revisão sistemática	Diversos modelos de ML, incluindo Regressão Logística, Árvore de Decisão, Random Forest, SVM e redes neurais	Variáveis clínicas e fatores de risco cardiovasculares discutidos na literatura	Fundamenta o uso de ML em cardiologia e reforça a necessidade de modelos interpretáveis
Costalat and Tavares [2022]	Heart Disease Dataset (UCI)	KNN, Árvore de Decisão, Regressão Logística e Voting Classifier	Idade, sexo, glicose, colesterol e indicadores clínicos gerais	Relaciona-se pela comparação de classificadores clássicos e pelo uso de variáveis clínicas gerais
Oliveira *et al*. [2023]	*heart.csv* (Kaggle)	KNN, Árvore de Decisão e MLP	Idade, sexo, dor torácica, colesterol e demais características clínicas	Reforça a aplicação de modelos supervisionados em bases cardiovasculares públicas
Ferreira *et al*. [2025]	Mapeamento sistemático	Modelos tradicionais, *ensembles* e Deep Learning	Variáveis clínicas diversas, com ênfase em seleção de características e interpretabilidade	Evidencia a relevância de comparar desempenho, interpretabilidade e uso clínico dos modelos
Teoh *et al*. [2026]	MIMIC-IV e MIMIC-IV-ED	Regressão Logística, Árvore de Decisão, Naïve Bayes, AdaBoost, Random Forest e modelos de Deep Learning	Variáveis clínicas estruturadas, sinais vitais, IMC, idade, diabetes e notas clínicas codificadas por BioClinicalBERT	Aproxima-se pelo uso do MIMIC-IV e pela predição de insuficiência cardíaca em registros hospitalares reais
Este trabalho	MIMIC-IV v3.1	Regressão Logística, SVM, Árvore de Decisão e KNN	Idade, sexo, sinais vitais, função renal, eletrólitos, perfil lipídico, troponina T, NT-proBNP, IMC e proporção HDL/LDL	Diferencia-se pela comparação de quatro modelos clássicos em larga escala, com foco em IAM e IC e análise de interpretabilidade

port Vector Machine e o K-Nearest Neighbors sejam frequentemente reconhecidos na literatura como técnicas competitivas, raramente são avaliados em conjunto com a Regressão Logística e a Árvore de Decisão sobre uma mesma base ampla e de origem clínica. Soma-se a isso o predomínio de estudos voltados à classificação genérica de doença cardiovascular, em detrimento de desfechos clinicamente específicos. Diante desse panorama, o presente trabalho diferencia-se ao propor a comparação experimental desses quatro algoritmos supervisionados sobre a base MIMIC-IV, com foco em desfechos cardiovasculares específicos e aliando o tratamento sistemático dos dados à aplicação de técnicas de interpretabilidade, de modo a contribuir tanto para a acurácia quanto para a transparência das decisões do modelo.

3 Metodologia

Este estudo caracteriza-se como uma pesquisa aplicada, de abordagem quantitativa e natureza experimental, cujo objetivo consiste em desenvolver e avaliar modelos de aprendizado de máquina para a predição de doenças cardiovasculares a partir de dados clínicos reais. A base de dados utilizada foi o Medical Information Mart for Intensive Care IV (MIMIC-IV), repositório público amplamente empregado na literatura científica de saúde e inteligência artificial, que reúne registros anonimizados de pacientes admitidos em ambiente hospitalar, incluindo informações provenientes tanto das internações gerais quanto das unidades de terapia intensiva do Beth Israel Deaconess Medical Center, em Boston, Massachusetts.

A abordagem experimental permitiu a implementação, o treinamento e a comparação de quatro algoritmos supervisionados de classificação, que são Regressão Logística, Support Vector Machine (SVM), Árvore de Decisão e K-Nearest Neighbors (KNN), selecionados em razão de sua consolidada aplicação em problemas de classificação binária no contexto da saúde. O processo foi conduzido em etapas sequenciais, compreendendo a seleção dos pacientes, a extração das variáveis clínicas, o pré-processamento dos dados, a construção dos conjuntos de treinamento e teste, o ajuste de hiperparâmetros e a avaliação comparativa dos modelos por meio de múltiplas métricas e validação cruzada estratificada.

3.1 Base de dados

Os dados utilizados neste estudo foram obtidos do Medical Information Mart for Intensive Care IV (MIMIC-IV), base pública desenvolvida pelo Beth Israel Deaconess Medical Center em parceria com o Massachusetts Institute of Technology (MIT). O repositório reúne informações clínicas anonimizadas de pacientes atendidos em unidades hospitalares e de terapia intensiva, constituindo uma das principais fontes de dados para pesquisas em saúde e inteligência artificial aplicada à medicina.

Foi utilizada a versão 3.1 da base, que contempla registros coletados entre 2008 e 2022 e é organizada em módulos relacionais contendo dados demográficos, diagnósticos, procedimentos, resultados laboratoriais, prescrições médicas, sinais vitais e informações de admissão hospitalar. A disponibilidade de dados reais e heterogêneos, provenientes de diferentes contextos clínicos, favorece o desenvolvimento de modelos preditivos com maior potencial de generalização em comparação com datasets sintéticos ou de escopo restrito.

Para este estudo, foram utilizados dois módulos do MIMIC-IV: o módulo `hosp`, que contém dados administrativos e clínicos da internação hospitalar geral, incluindo diagnósticos codificados, resultados de exames laboratoriais e informações demográficas; e o módulo `icu`, que registra variáveis coletadas especificamente durante a permanência na unidade de terapia intensiva, como sinais vitais contínuos e parâmetros hemodinâmicos. A integração entre esses dois módulos, realizada por meio do identificador único de admissão, permitiu compor um conjunto de dados com maior abrangência clínica do que o obtido a partir de qualquer módulo isolado.

No módulo `hosp`, foram utilizadas as tabelas `patients`, `admissions`, `diagnoses_icd`, `labevents`,

`d_labitems` e `omr`, cada uma associada a um aspecto específico dos dados extraídos. Dados demográficos como idade e sexo vieram da tabela `patients`; as admissões hospitalares foram identificadas em `admissions`; e os diagnósticos, registrados nos formatos CID-9 e CID-10, foram recuperados a partir de `diagnoses_icd`. Os exames laboratoriais de interesse foram selecionados cruzando `labevents` com `d_labitems`, responsável por identificar cada tipo de exame. As variáveis de peso e altura foram obtidas inicialmente de `chartevents` e complementadas com registros de `omr` apenas quando estavam ausentes, mantendo-se os valores provenientes da UTI quando disponíveis.

No módulo `icu`, os dados da tabela `chartevents` foram filtrados previamente pelos `itemids` correspondentes aos sinais vitais, identificados com o auxílio da tabela `d_items`. Foram selecionados registros de frequência cardíaca, pressão arterial sistólica e diastólica, frequência respiratória e saturação periférica de oxigênio, além dos `itemids` de peso e altura mencionados anteriormente. Após esse filtro inicial, os registros foram associados às admissões hospitalares por meio do campo `hadm_id` e agregados pela mediana quando havia múltiplas medições para uma mesma admissão. A tabela `icustays` foi utilizada apenas na etapa exploratória inicial, para relacionar admissões hospitalares às respectivas estadias em UTI, mas não integrou o pipeline final de extração dos dados.

Cada linha do conjunto final corresponde a uma admissão hospitalar, identificada pelo campo `hadm_id`. Por isso, um mesmo paciente, identificado por `subject_id`, pode aparecer mais de uma vez no conjunto de dados quando estiver associado a múltiplas internações, já que diagnósticos, exames laboratoriais e sinais vitais estão vinculados ao contexto de cada admissão específica, e não ao paciente de forma geral.

A Tabela 2 apresenta as principais tabelas utilizadas na extração dos dados e a finalidade de cada uma no estudo.

Para complementar a descrição da extração dos dados, as Figuras 1 e 2 apresentam os modelos lógicos simplificados das tabelas relacionadas aos módulos `hosp` e `icu`. No módulo `hosp`, a tabela `admissions` atua como ponto central para relacionar pacientes, diagnósticos e exames laboratoriais. No módulo `icu`, os sinais vitais foram extraídos de `chartevents`, com identificação dos itens clínicos por meio de `d_items`.

A tabela `icustays` é apresentada para explicitar a relação entre admissões hospitalares e estadias em UTI. No entanto, no pipeline final, a consolidação dos sinais vitais foi realizada diretamente a partir de `chartevents`, utilizando o campo `hadm_id`.

Antes da extração das variáveis de interesse, foi conduzida uma etapa exploratória para verificar a integridade dos arquivos, identificar a estrutura da versão 3.1 e confirmar a disponibilidade dos itens laboratoriais e de sinais vitais necessários para o estudo. Essa etapa permitiu validar os identificadores utilizados na construção do conjunto de dados analítico e gerar relatórios auxiliares que documentam o processo, contribuindo para a consistência das análises.

O acesso ao MIMIC-IV é público, condicionado à certificação ética dos pesquisadores, e sua ampla adoção na literatura científica internacional reforça a adequação da base para investigações envolvendo aprendizado de máquina aplicado à predição de desfechos clínicos.

3.2 Seleção dos pacientes e definição da variável-alvo

A seleção dos pacientes foi realizada a partir dos registros de admissões hospitalares presentes no módulo hosp do MIMIC-IV, considerando todos os indivíduos com idade igual ou superior a 18 anos. Pacientes menores de idade foram excluídos em razão das particularidades fisiológicas e clínicas da população pediátrica, que diferem substancialmente do perfil de risco cardiovascular investigado.

A identificação dos casos positivos baseou-se nos códigos da Classificação Estatística Internacional de Doenças e Problemas Relacionados à Saúde (CID) registrados nos diagnósticos hospitalares. Foram selecionados pacientes com diagnóstico de Infarto Agudo do Miocárdio (IAM) ou Insuficiência Cardíaca (IC), por representarem as principais causas de morbimortalidade cardiovascular diretamente relacionadas aos objetivos deste estudo. Para o IAM, foram utilizados os códigos CID-10 I21 e I22 e o código CID-9 410; para a IC, os códigos CID-10 I50 e CID-9 428. A inclusão de ambas as versões da classificação visou contemplar admissões registradas em períodos distintos do intervalo temporal coberto pela base.

A partir desses critérios, foi construída uma variável alvo binária para utilização nos modelos supervisionados. As admissões hospitalares contendo pelo menos um dos códigos selecionados foram classificadas como pertencentes à classe positiva (valor igual a 1), indicando a presença de doença cardiovascular. Os demais registros foram classificados como pertencentes à classe negativa (valor igual a 0), representando a ausência dos diagnósticos investigados. Essa estratégia possibilitou a formulação do problema como uma tarefa de classificação binária, adequada à aplicação dos algoritmos de aprendizado de máquina selecionados.

3.3 Extração e seleção das variáveis clínicas

Após a definição da população do estudo, foram extraídas variáveis clínicas dos módulos `hosp` e `icu` do MIMIC-IV, selecionadas com base na literatura sobre fatores de risco e biomarcadores cardiovasculares e na disponibilidade dos registros na versão 3.1 da base.

As variáveis demográficas compreenderam idade e sexo biológico. O sexo foi convertido para representação binária, assumindo valor 1 para o sexo masculino e 0 para o feminino. Entre os sinais vitais, foram incluídos frequência cardíaca, pressão arterial sistólica e diastólica, frequência respiratória, saturação periférica de oxigênio, peso corporal e altura, medidas rotineiramente coletadas na avaliação clínica de pacientes cardiopatas.

Os exames laboratoriais selecionados abrangem marcadores de função renal (creatinina e ureia), hematológicos (hemoglobina, contagem de leucócitos e MCH), eletrolíticos (sódio, potássio e magnésio), metabólicos (glicose), lipídicos (colesterol total, HDL e LDL) e cardíacos específicos (troponina T e NT-proBNP). Esses marcadores permitem avaliar alterações metabólicas, renais e cardíacas associadas à pre-

Tabela 2. Tabelas utilizadas na extração dos dados

Módulo	Tabela	Informação utilizada
`hosp`	`patients`	Idade e sexo dos pacientes
`hosp`	`admissions`	Identificação das admissões hospitalares
`hosp`	`diagnoses_icd`	Códigos CID-9/CID-10 para definição da variável-alvo
`hosp`	`labevents`	Resultados de exames laboratoriais
`hosp`	`d_labitems`	Identificação dos exames laboratoriais
`hosp`	`omr`	Peso e altura, como complemento quando ausentes em `chartevents`
`icu`	`chartevents`	Sinais vitais, peso e altura registrados durante a UTI
`icu`	`d_items`	Identificação dos sinais vitais registrados em `chartevents`

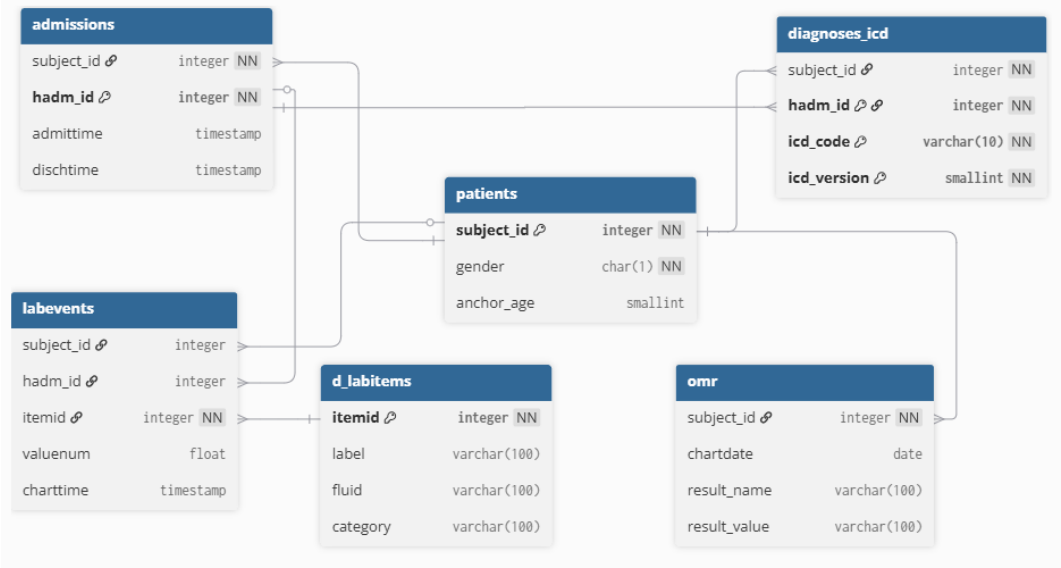

Figura 1. Modelo lógico simplificado das principais tabelas utilizadas no módulo `hosp`

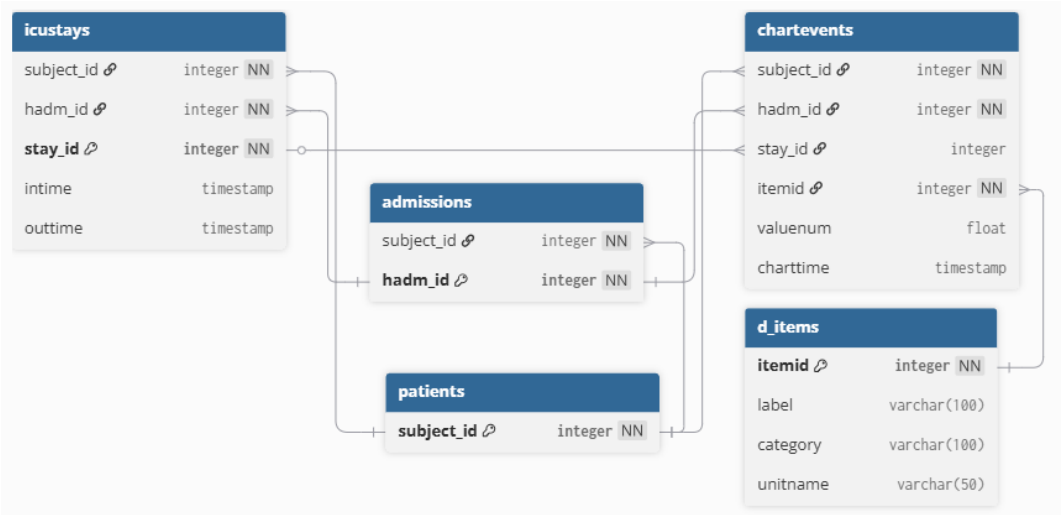

Figura 2. Modelo lógico simplificado das principais tabelas relacionadas ao módulo `icu`

sença de DCV.

Além das variáveis originalmente presentes na base, foram calculados atributos derivados. O Índice de Massa Corporal (IMC) foi obtido a partir das informações de peso e altura dos pacientes, enquanto a razão entre HDL e LDL foi incluída com o objetivo de representar de forma complementar o perfil lipídico dos indivíduos avaliados. A utilização conjunta dessas variáveis buscou fornecer aos modelos preditivos uma representação mais abrangente das condições clínicas dos pacientes.

A Tabela 3 apresenta o conjunto de atributos utilizado na construção dos modelos preditivos. Para cada variável, são indicados o nome descritivo adotado no texto, o nome correspondente no conjunto de dados, o módulo de origem, a tabela ou fonte utilizada e a forma de representação no conjunto final de dados. Essa organização permite relacionar as variáveis clínicas discutidas no trabalho com os atributos efetivamente utilizados no pipeline de aprendizado de máquina.

Observa-se que a maior parte dos exames laboratoriais foi extraída do módulo `hosp`, a partir da tabela `labevents`, e

Tabela 3. Atributos utilizados na construção dos modelos preditivos

Atributo	Variável no dataset	Módulo	Origem	Representação
Idade	idade	hosp	patients	original
Sexo masculino	sexo_m	hosp	patients	binário
Frequência cardíaca	fc	icu	chartevents	mediana por admissão
Pressão arterial sistólica	pas	icu	chartevents	mediana por admissão
Pressão arterial diastólica	pad	icu	chartevents	mediana por admissão
Frequência respiratória	fr	icu	chartevents	mediana por admissão
Saturação periférica de O_2	spo2	icu	chartevents	mediana por admissão
Troponina T	troponina_t	hosp	labevents	mediana por admissão
NT-proBNP	nt_probnp	hosp	labevents	mediana por admissão
Creatinina	creatinina	hosp	labevents	mediana por admissão
Ureia	ureia	hosp	labevents	mediana por admissão
Hemoglobina	hemoglobina	hosp	labevents	mediana por admissão
Leucócitos	leucocitos	hosp	labevents	mediana por admissão
Sódio	sodio	hosp	labevents	mediana por admissão
Potássio	potassio	hosp	labevents	mediana por admissão
Glicose	glicose	hosp	labevents	mediana por admissão
Magnésio	magnesio	hosp	labevents	mediana por admissão
MCH	mch	hosp	labevents	mediana por admissão
Colesterol total	colesterol_total	hosp	labevents	mediana por admissão
HDL	hdl	hosp	labevents	mediana por admissão
LDL	ldl	hosp	labevents	mediana por admissão
IMC	imc	Novo	peso e altura	calculado
Proporção HDL/LDL	proporcao_hdl_ldl	Novo	HDL e LDL	calculada

representada pela mediana dos valores registrados para cada admissão hospitalar. Já os sinais vitais foram extraídos do módulo `icu`, por meio de `chartevents`, seguindo a mesma estratégia de consolidação por mediana. As variáveis `imc` e `proporcao_hdl_ldl` foram calculadas a partir de atributos previamente extraídos, sendo tratadas como variáveis derivadas.

Além da descrição dos atributos, foi realizada uma caracterização estatística das variáveis antes da etapa de imputação dos valores ausentes. Para cada variável, foram calculados o número de registros presentes, o número de registros ausentes, o percentual de ausência, a média e a mediana. A Tabela 4 apresenta uma síntese das variáveis clínicas utilizadas no estudo.

A análise estatística evidencia diferenças importantes na disponibilidade das variáveis. As variáveis demográficas, idade e sexo masculino, não apresentaram valores ausentes. Exames laboratoriais gerais, como creatinina, ureia, hemoglobina, leucócitos, sódio, potássio e glicose, apresentaram maior disponibilidade no conjunto de dados. Em contrapartida, os sinais vitais provenientes do módulo `icu` apresentaram percentuais elevados de ausência, uma vez que dependem da existência de registros durante a permanência em unidade de terapia intensiva. Marcadores biológicos mais específicos, como troponina T e NT-proBNP, apresentaram percentuais elevados de valores ausentes, aspecto que requer uma interpretação cuidadosa no contexto da prática hospitalar.

No caso dos marcadores biológicos mais específicos, a ausência de registros não deve ser interpretada apenas como falha de preenchimento ou perda aleatória de informação. Esses exames tendem a ser solicitados conforme a suspeita clínica, a gravidade do quadro e a necessidade de investigação diagnóstica. Assim, em admissões nas quais não havia hipótese clínica inicial relacionada à lesão miocárdica ou insuficiência cardíaca, é possível que tais exames não tenham sido requisitados. Dessa forma, a ausência desses valores pode refletir o próprio fluxo de atendimento hospitalar, e não necessariamente inconsistência da base de dados.

A presença de percentuais elevados de valores ausentes nesses marcadores possui impacto direto na modelagem. Por um lado, a imputação pela mediana permitiu preservar o tamanho do conjunto de dados e evitar a exclusão de um grande número de admissões hospitalares. Por outro lado, essa estratégia pode reduzir a variabilidade de marcadores altamente específicos e atenuar sua influência individual em determinados algoritmos. Portanto, os resultados associados a esses atributos devem ser interpretados considerando que a disponibilidade do exame pode estar relacionada à suspeita clínica, à gravidade do quadro e ao perfil do paciente no momento da admissão.

3.4 Pré-processamento dos dados

O pré-processamento teve como finalidade garantir a qualidade e a consistência dos dados antes do treinamento dos modelos. Inicialmente, os exames laboratoriais e os sinais vitais foram submetidos a limites de plausibilidade clínica, definidos individualmente para cada variável. Valores fora desses limites foram tratados como ausentes, reduzindo a influência de possíveis erros de registro ou inconsistências presentes na base. O mesmo critério foi aplicado ao Índice de Massa Corporal calculado, com valores fora da faixa de 10 a 80 kg/m² também convertidos para ausentes.

Os dados foram então consolidados em um único conjunto analítico, reunindo as variáveis selecionadas e a variável-alvo. Como as bases clínicas frequentemente apresentam registros incompletos, os valores ausentes foram tratados por imputação com a mediana de cada variável numérica, estratégia escolhida por sua robustez à presença de valores extremos e ampla adoção em estudos com dados médicos.

Na sequência, o conjunto foi dividido em treinamento (70%) e teste (30%), de forma estratificada em relação à variável-alvo, preservando a distribuição original das classes em ambas as partições e evitando vieses decorrentes do desbalanceamento entre elas.

Por fim, os atributos utilizados pela Regressão Logística, o SVM e o KNN foram padronizados pela transformação Z-score, com média zero e desvio padrão unitário, já que esses algoritmos são sensíveis à escala das variáveis[James *et al.*, 2023]. Esse procedimento evita que variáveis de maior magnitude dominem o processo de aprendizagem. A Árvore de Decisão não foi submetida a essa etapa, pois suas divisões

Tabela 4. Caracterização estatística resumida das variáveis selecionadas

Variável	Registros presentes	Registros ausentes	% ausente	Média	Mediana
Idade	546028	0	0,00	56,89	58,00
Sexo masculino	546028	0	0,00	0,48	0,00
Frequência cardíaca	85237	460791	84,39	83,33	82,00
Pressão arterial sistólica	84281	461747	84,56	119,17	117,50
Pressão arterial diastólica	84280	461748	84,56	64,89	64,00
Frequência respiratória	85155	460873	84,40	18,92	18,00
Saturação periférica de O_2	85142	460886	84,41	96,84	97,00
Troponina T	52695	493333	90,35	0,41	0,07
NT-proBNP	24463	521565	95,52	5338,51	1797,00
Creatinina	415830	130198	23,84	1,24	0,90
Ureia	411868	134160	24,57	20,47	16,00
Hemoglobina	422276	123752	22,66	11,04	11,00
Leucócitos	421377	124651	22,83	8,54	7,80
Sódio	409794	136234	24,95	138,67	139,00
Potássio	411533	134495	24,63	4,10	4,10
Glicose	406252	139776	25,60	119,64	108,00
Magnésio	378840	167188	30,62	1,99	2,00
MCH	421286	124742	22,85	29,71	29,95
Colesterol total	31325	514703	94,26	160,82	156,00
HDL	30342	515686	94,44	48,52	45,00
LDL	29032	516996	94,68	91,80	87,00
IMC	42635	503393	92,19	28,69	27,47
Proporção HDL/LDL	29032	516996	94,68	0,64	0,52

internas são baseadas em limiares ordinais e independem da escala dos atributos.

3.5 Desenvolvimento e ajuste dos modelos

Com o conjunto de dados devidamente preparado, foram implementados quatro algoritmos supervisionados de classificação: Regressão Logística, SVM, Árvore de Decisão e KNN. A seleção desses métodos considerou tanto a ampla utilização de cada um na literatura de predição de doenças cardiovasculares quanto a diversidade de princípios de aprendizagem que representam, desde abordagens paramétricas e baseadas em margens de separação até métodos hierárquicos e por similaridade, o que permite uma comparação mais abrangente do que a análise de algoritmos de uma mesma família.

A Regressão Logística foi incluída por ser o método de referência em classificação binária, com interpretabilidade direta por meio dos coeficientes. O SVM foi escolhido por sua capacidade de identificar fronteiras de decisão não-lineares em espaços de alta dimensionalidade, característica relevante, dado o volume e a heterogeneidade das variáveis clínicas utilizadas. A Árvore de Decisão, além de competitiva em desempenho, oferece uma estrutura de regras interpretável, permitindo identificar quais atributos exercem maior influência na classificação. O KNN representa uma abordagem não paramétrica baseada em similaridade entre observações, cuja inclusão permite avaliar o comportamento desse tipo de método frente ao desbalanceamento de classes característico do conjunto de dados.

Embora o Random Forest seja amplamente utilizado em problemas de classificação e possa apresentar desempenho superior por combinar múltiplas árvores de decisão, optou-se por utilizar a Árvore de Decisão simples neste estudo em razão de sua maior interpretabilidade direta. Como um dos objetivos do trabalho é comparar modelos clássicos e compreender a influência das variáveis clínicas nas predições, a Árvore de Decisão permite analisar regras, limiares e importância dos atributos de forma mais transparente. Além disso, considerando o volume do conjunto de dados e a inclusão de modelos com maior custo computacional, como SVM e KNN, a escolha contribuiu para manter a viabilidade experimental e a clareza na comparação entre algoritmos de diferentes naturezas.

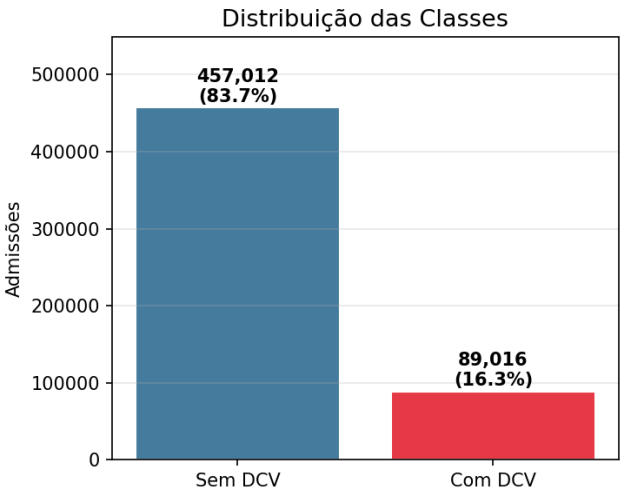

Figura 3. Distribuição das classes no conjunto de dados.

Dado o desbalanceamento entre as classes, 83,7% das admissões sem DCV contra 16,3% com DCV, foi adotada a estratégia de ponderação durante o treinamento. Para a Regressão Logística, o SVM e a Árvore de Decisão, o parâmetro *class_weight='balanced'* atribui pesos inversamente proporcionais à frequência de cada classe, reduzindo o viés do modelo em direção à classe majoritária. O KNN, por não dispor de um mecanismo nativo equivalente, recebeu ponderação por distância (*weights='distance'*), embora essa abordagem se mostre menos eficaz para compensar desbalanceamentos acentuados.

Os melhores hiperparâmetros identificados para cada modelo estão apresentados na Tabela 5.

O ajuste dos hiperparâmetros foi realizado exclusivamente sobre o conjunto de treinamento, por meio da técnica GridSearchCV com validação cruzada estratificada, evitando o vazamento de informações do conjunto de teste. Para o SVM e o KNN, cujo custo computacional inviabilizaria a busca exaustiva sobre o dataset completo, foi empregada amostragem estratificada durante essa etapa, preservando a

Tabela 5. Melhores hiperparâmetros identificados para cada modelo

Modelo	Hiperparâmetros selecionados
Regressão Logística	*C* = 10; *penalty* = l2; *class_weight* = balanced
Árvore de Decisão	*max_depth* = 10; *min_samples_leaf* = 50; *class_weight* = balanced
SVM	*C* = 1; *gamma* = 0,01; *kernel* = rbf; *class_weight* = balanced
KNN	*n_neighbors* = 51; *weights* = distance

representatividade das classes sem comprometer a viabilidade do processo. A implementação foi realizada em Python com as bibliotecas Pandas, NumPy e Scikit-learn, e o parâmetro random_state foi fixado em 42 em todas as etapas aplicáveis para garantir a reprodutibilidade dos experimentos.

3.6 Avaliação de desempenho dos modelos

A avaliação do desempenho dos modelos foi realizada por meio de validação cruzada estratificada. Para a Regressão Logística, Árvore de Decisão e K-Nearest Neighbors, foram empregadas dez partições (10-fold cross-validation), enquanto para o Support Vector Machine foram utilizadas cinco partições (5-fold cross-validation). Essa adaptação foi adotada em virtude do elevado custo computacional do SVM em conjuntos de dados extensos, permitindo reduzir o tempo de processamento sem comprometer a robustez da estimativa de desempenho. Em ambos os casos, o conjunto de treinamento foi dividido em subconjuntos mutuamente exclusivos, com preservação da proporção entre as classes, e o processo foi repetido até que todas as partições tivessem sido utilizadas para validação. Ao final, os modelos foram avaliados sobre o conjunto de teste previamente separado para estimar sua capacidade de generalização em dados não observados durante o treinamento.

As métricas utilizadas foram acurácia, precisão, Recall, F1-Score e área sob a curva ROC (AUC-ROC). A acurácia representa a proporção global de classificações corretas. A precisão indica a fração das predições positivas que correspondem efetivamente a casos positivos. O Recall, também denominado sensibilidade, avalia a capacidade do modelo de identificar corretamente os indivíduos da classe positiva, métrica de especial relevância no contexto médico, onde falsos negativos têm consequências clínicas mais graves do que falsos positivos [Teodoro *et al*., 2025]. O F1-Score corresponde à média harmônica entre precisão e Recall, sendo particularmente informativo em cenários com distribuição desbalanceada das classes [Sokolova and Lapalme, 2009]. Por fim, a AUC-ROC mede a capacidade discriminativa do modelo ao longo de diferentes limiares de decisão. Essa métrica é preferível à acurácia simples em problemas com desbalanceamento acentuado Fawcett [2006].

Neste estudo, o Recall recebeu atenção especial na comparação dos modelos, pois o objetivo principal não se limita a maximizar o número total de classificações corretas, mas sim identificar admissões hospitalares associadas a IAM ou IC. Nesse contexto, um falso negativo representa uma admissão com doença cardiovascular classificada incorretamente como ausência da condição, situação que poderia atrasar a investigação clínica ou o direcionamento do paciente para acompanhamento adequado. Dessa forma, a análise dos modelos considerou o Recall em conjunto com F1-Score, AUC-ROC e matriz de confusão, evitando que a acurácia isolada conduzisse a uma interpretação superestimada do desempenho em uma base desbalanceada.

3.7 Interpretabilidade dos modelos

Além da avaliação preditiva, foram empregadas técnicas de interpretabilidade para compreender a influência das variáveis clínicas no processo de decisão de cada modelo.

A importância da permutação foi aplicada à Regressão Logística, à Árvore de Decisão e ao KNN. O SVM não foi submetido a essa análise devido ao elevado custo computacional associado ao procedimento de permutação. A técnica consiste em avaliar a redução do desempenho do modelo após a aleatorização dos valores de cada atributo individualmente, permitindo identificar quais variáveis exercem maior impacto sobre as predições.

Para a Regressão Logística, foram analisados adicionalmente os coeficientes associados a cada variável explicativa, possibilitando interpretar a direção e a magnitude da influência de cada atributo sobre a probabilidade de ocorrência do desfecho investigado.

No caso da Árvore de Decisão, foram examinadas as importâncias baseadas no Índice Gini e as regras de decisão extraídas da estrutura da árvore treinada, fornecendo uma representação direta do processo de classificação e das combinações de atributos associadas à presença ou ausência de DCV.

4 Resultados e Discussão

Esta seção apresenta os resultados obtidos a partir da aplicação dos algoritmos de aprendizado de máquina para a predição de doenças cardiovasculares, bem como a discussão dos principais resultados do estudo. Inicialmente, são comparados os desempenhos dos modelos por meio das métricas de avaliação utilizadas. Em seguida, são analisadas suas capacidades discriminativas, os erros e acertos observados nas matrizes de confusão e, por fim, a relevância das variáveis clínicas identificadas pelos métodos de interpretabilidade. Os resultados obtidos são comparados aos apresentados nos trabalhos relacionados, com discussão das implicações para aplicações clínicas de apoio à decisão.

4.1 Desempenho dos modelos preditivos

Após o treinamento e o ajuste dos hiperparâmetros, os modelos foram avaliados sobre o conjunto de teste independente, correspondente a 30% dos dados. Os resultados obtidos para as métricas de acurácia, precisão, Recall, F1-Score e AUC-ROC estão apresentados na Tabela 6.

A Figura 4 ilustra comparativamente o desempenho dos quatro modelos nas métricas avaliadas.

O KNN apresentou a maior acurácia (85,71%) e precisão (73,35%) entre os modelos analisados. Entretanto, seu Recall foi significativamente inferior ao dos demais, atingindo apenas 19,37%, o que indica baixa capacidade de identificação dos pacientes pertencentes à classe positiva e torna o modelo inadequado para triagem clínica, contexto em que falsos negativos têm consequências mais graves do que falsos

Tabela 6. Desempenho dos modelos no conjunto de teste

Modelo	Acurácia	Precisão	Recall	F1-Score	AUC-ROC
Regressão Logística	0,764	0,384	0,735	0,504	0,839
Árvore de Decisão	0,779	0,408	0,788	0,538	0,868
KNN	0,857	0,734	0,194	0,306	0,832
SVM	0,776	0,402	0,769	0,528	0,855

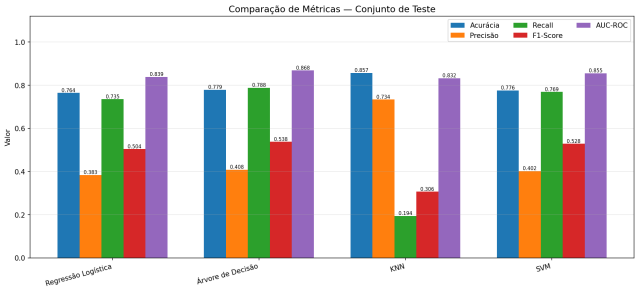

Figura 4. Comparação das métricas de desempenho dos modelos no conjunto de teste

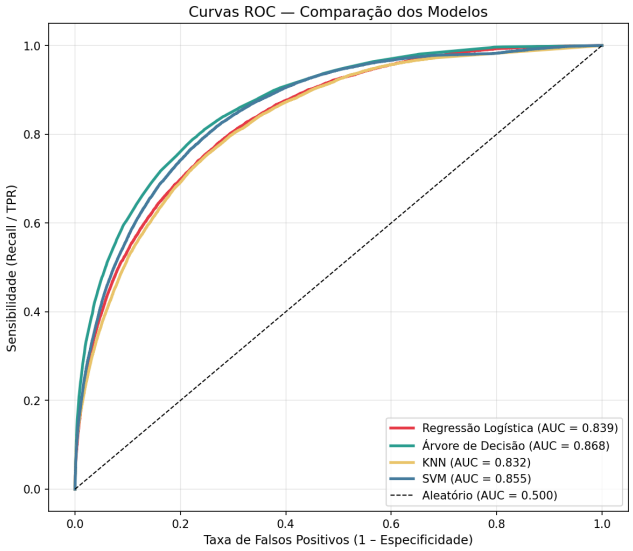

Figura 5. Curvas ROC dos modelos avaliados no conjunto de teste

positivos.

A Árvore de Decisão apresentou o melhor desempenho global, com Recall de 78,79%, F1-Score de 53,77% e AUC-ROC de 86,81%. Esses valores refletem boa capacidade de identificação dos indivíduos com DCV, mantendo simultaneamente uma adequada capacidade discriminativa entre as classes.

A Regressão Logística alcançou Recall de 73,51%, F1-Score de 50,40% e AUC-ROC de 83,88%. Apesar de sua natureza linear e maior simplicidade em relação aos demais algoritmos, o modelo demonstrou desempenho consistente e competitivo, indicando que abordagens estatísticas tradicionais ainda produzem resultados satisfatórios em problemas de classificação com dados clínicos estruturados.

O SVM obteve Recall de 76,89%, F1-Score de 52,83% e AUC-ROC de 85,50%, desempenho próximo ao da Árvore de Decisão. Vale destacar que esses resultados foram alcançados com validação cruzada de cinco partições e otimização de hiperparâmetros restrita a uma amostra do conjunto de treinamento, em razão do custo computacional do algoritmo, que sugere que recursos computacionais mais amplos poderiam permitir uma investigação mais aprofundada de suas configurações, embora os resultados já indiquem desempenho competitivo.

Considerando especificamente o Recall, observa-se que a Árvore de Decisão foi o modelo mais adequado ao objetivo clínico do estudo, pois identificou 78,79% das admissões pertencentes à classe positiva. O SVM apresentou desempenho próximo, com Recall de 76,89%, seguido pela Regressão Logística, com 73,51%. Esses três modelos demonstraram maior capacidade de reconhecer admissões com DCV, ainda que à custa de maior número de falsos positivos. Em contrapartida, o KNN apresentou Recall de apenas 19,37%, apesar de ter obtido a maior acurácia entre os modelos avaliados. Esse resultado evidencia que, em cenários com desbalanceamento entre classes, a acurácia isolada pode conduzir a interpretações equivocadas, tornando métricas como Recall, F1-Score e AUC-ROC mais adequadas para avaliar a identificação correta dos casos positivos e o equilíbrio entre sensibilidade e precisão.

4.2 Capacidade discriminativa dos modelos

A análise das curvas ROC foi realizada com o objetivo de verificar a capacidade discriminativa de cada algoritmo em diferentes limiares de decisão. A curva ROC relaciona a taxa de verdadeiros positivos (sensibilidade) à taxa de falsos positivos (1 − especificidade), e a área sob essa curva (AUC-ROC) fornece uma medida global da capacidade do modelo em distinguir admissões com e sem DCV, independentemente do limiar adotado.

A Figura 5 apresenta as curvas ROC dos quatro modelos no conjunto de teste. Todos os algoritmos apresentaram desempenho superior ao classificador aleatório, representado pela linha diagonal tracejada, confirmando que os modelos aprenderam padrões relevantes a partir das variáveis clínicas selecionadas.

A Árvore de Decisão obteve a maior AUC-ROC (0,868), seguida pelo SVM (0,855), pela Regressão Logística (0,839) e pelo KNN (0,832). A proximidade entre os três primeiros modelos é visível na figura, com suas curvas sobrepostas na região de baixa taxa de falsos positivos — justamente onde o desempenho clínico é mais relevante, pois corresponde a limiares mais restritivos de classificação positiva.

A análise das curvas ROC também evidencia a limitação da acurácia como métrica isolada em cenários de desbalanceamento. O KNN, que apresentou a maior acurácia na seção anterior (85,71%), obteve a menor AUC-ROC entre os modelos avaliados. Esse resultado confirma que seu desempenho global foi inflado pela predominância da classe negativa, enquanto sua capacidade discriminativa real, medida

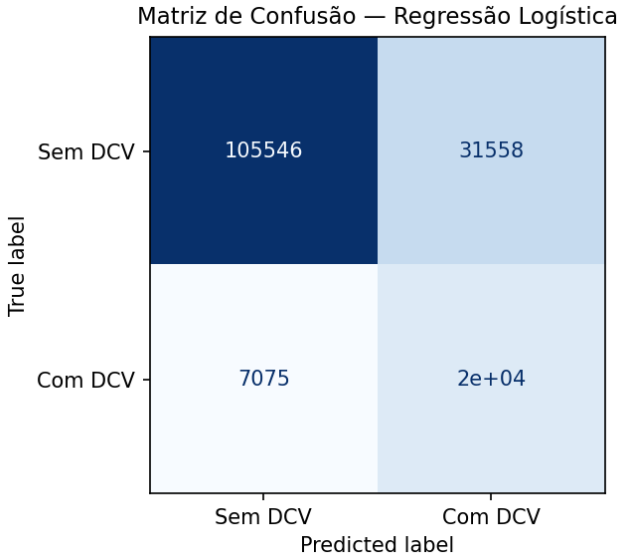

Figura 6. Matriz de confusão - Regressão Logística

independentemente da distribuição das classes, foi inferior à dos demais algoritmos.

4.3 Análise das matrizes de confusão

Com o objetivo de compreender de forma mais detalhada os tipos de acertos e erros cometidos pelos modelos, foram analisadas as matrizes de confusão obtidas no conjunto de teste. Essa análise é especialmente relevante em classificação médica, pois permite verificar não apenas a quantidade total de predições corretas, mas também a distribuição entre falsos positivos e falsos negativos. No contexto deste estudo, os falsos negativos merecem atenção particular, pois representam admissões com DCV classificadas incorretamente como ausência da condição, situação clinicamente mais grave do que um falso alarme.

A Regressão Logística classificou corretamente 105.546 admissões sem DCV e 19.625 com DCV, registrando 31.558 falsos positivos e 7.075 falsos negativos, conforme apresentado na Figura 6. O modelo demonstrou capacidade razoável de identificação da classe positiva, alcançando Recall de 73,51%.

A Árvore de Decisão classificou corretamente 106.593 admissões sem DCV e 21.041 com DCV, com 30.511 falsos positivos e 5.664 falsos negativos, conforme apresentado na Figura 7. Esse resultado é diretamente refletido no Recall de 78,79%, o maior do estudo.

O KNN apresentou comportamento distinto dos demais. Embora tenha acertado 135.225 admissões sem DCV e registrado apenas 1.879 falsos positivos, identificou corretamente somente 5.172 casos com DCV, deixando de reconhecer 21.533 pacientes da classe positiva, conforme apresentado na Figura 8.

O SVM classificou corretamente 106.609 admissões sem DCV e 20.534 com DCV, registrando 30.095 falsos positivos e 6.171 falsos negativos, conforme apresentado na Figura 9. O desempenho foi próximo ao da Árvore de Decisão, com Recall de 76,89% e F1-Score de 52,83%.

A comparação entre as matrizes de confusão reforça que a avaliação dos modelos não deve se basear apenas na acurácia. O KNN apresentou o maior número total de classi-

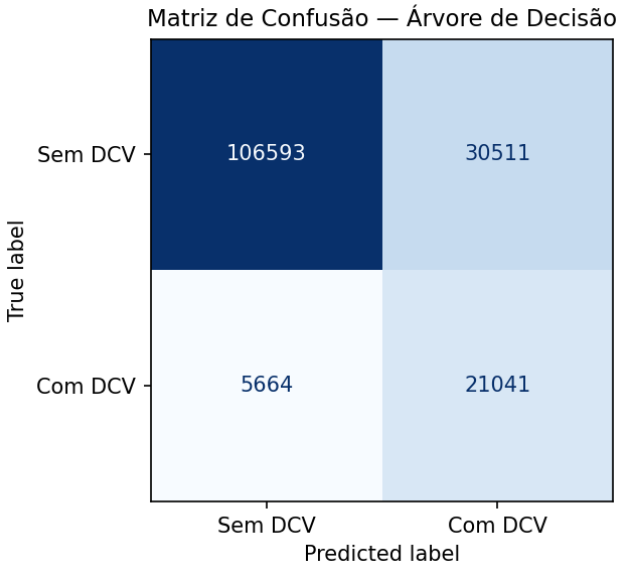

Figura 7. Matriz de confusão - Árvore de Decisão

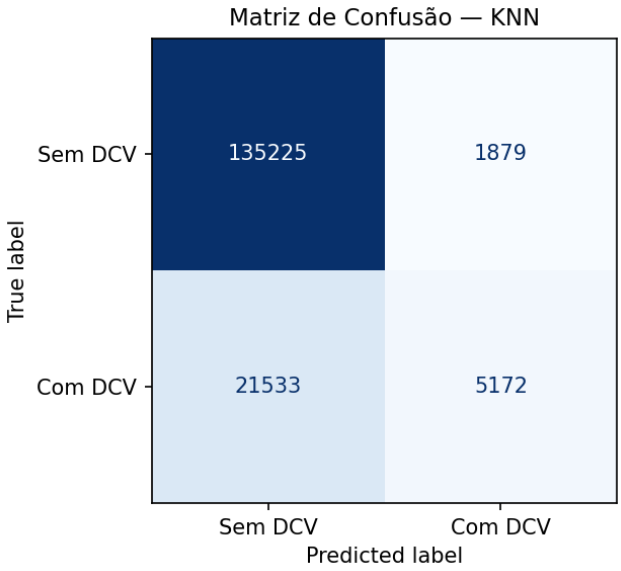

Figura 8. Matriz de confusão - KNN

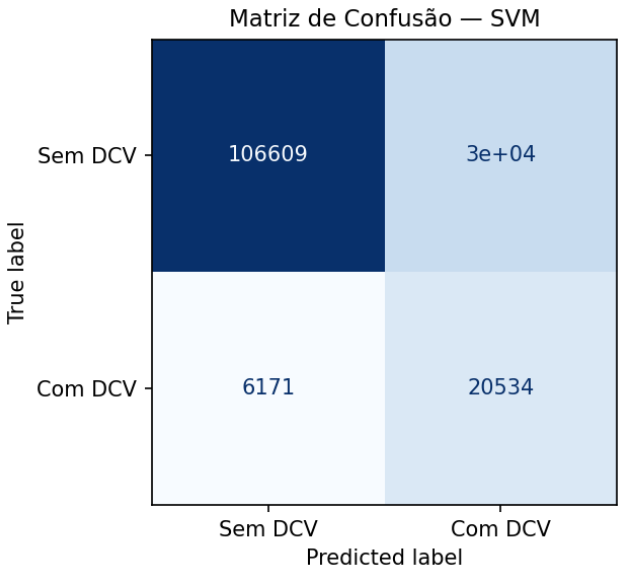

Figura 9. Matriz de confusão - SVM.

ficações corretas devido ao predomínio da classe negativa no conjunto de dados, mas falhou na identificação da maior parte dos casos com doenças cardiovasculares. Por outro lado, a Árvore de Decisão e o SVM apresentaram maior número de falsos positivos, mas reduziram consideravelmente os falsos negativos, característica desejável em aplicações clínicas nas quais a não identificação de pacientes em risco pode trazer consequências mais graves do que a indicação de avaliação adicional para pacientes sem a condição.

Dessa forma, a análise das matrizes de confusão confirma a superioridade da Árvore de Decisão no contexto deste estudo, pois o modelo apresentou o menor número de falsos negativos e o melhor equilíbrio entre sensibilidade e capacidade geral de classificação. O SVM também apresentou desempenho robusto, enquanto a Regressão Logística se manteve como uma alternativa simples e consistente. Já o KNN, apesar de sua elevada acurácia, demonstrou limitação relevante para a identificação da classe positiva.

4.4 Interpretabilidade dos modelos

Além da análise de desempenho, foram aplicadas técnicas de interpretabilidade para identificar quais variáveis clínicas exerceram maior influência nas predições de cada modelo. Em aplicações de apoio à decisão clínica, essa etapa é tão relevante quanto as métricas preditivas, pois permite verificar se os padrões aprendidos pelos algoritmos correspondem a fatores clinicamente plausíveis.

A importância por permutação foi calculada para a Regressão Logística, a Árvore de Decisão e o KNN, e os resultados estão apresentados na Figura 10. A técnica avalia a redução do desempenho do modelo quando os valores de determinada variável são aleatorizados individualmente, estimando sua contribuição relativa para a predição. Observa-se que idade, ureia, troponina T e NT-proBNP figuraram entre os atributos de maior relevância nos três modelos, indicando que essas variáveis contribuíram de forma expressiva para a distinção entre admissões com e sem DCV.

Para a Regressão Logística, foram analisados adicionalmente os coeficientes das variáveis explicativas, apresenta-

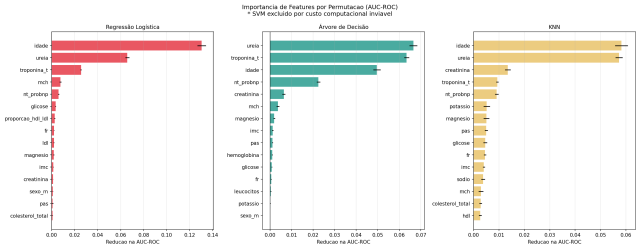

Figura 10. Importância por permutação dos modelos.

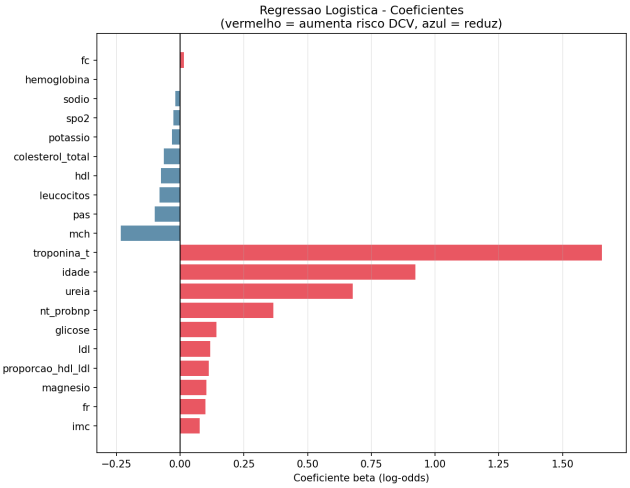

Figura 11. Coeficientes da Regressão Logística.

dos na Figura 11. Os maiores coeficientes positivos foram observados para troponina T, idade, ureia e NT-proBNP, indicando que valores mais elevados dessas variáveis aumentaram a probabilidade estimada de DCV. Esse resultado é clinicamente coerente. A troponina está associada à lesão miocárdica, e o NT-proBNP é um marcador estabelecido de insuficiência cardíaca. A idade, por sua vez, é um fator de risco cardiovascular já consolidado na literatura. Já a ureia pode refletir alterações renais associadas a um quadro clínico mais grave.

A análise da Árvore de Decisão pelo Índice Gini, apresentada na Figura 12, apontou o mesmo conjunto de variáveis (ureia, troponina T, idade e NT-proBNP) como atributos de maior relevância. A convergência entre métodos distintos de interpretabilidade reforça a consistência dos achados e sugere que os modelos não apenas capturaram padrões estatísticos nos dados, mas atribuíram relevância a fatores com respaldo clínico estabelecido na literatura sobre risco cardiovascular.

Além desses atributos, a creatinina merece atenção específica, pois sua relevância variou entre os métodos de interpretabilidade. Embora não tenha figurado entre os atributos de maior destaque em todos os métodos analisados, sua presença nos resultados deve ser interpretada em conjunto com outros marcadores de função renal, especialmente a ureia. Ambos os exames refletem aspectos relacionados à função renal e ao estado clínico geral do paciente, podendo apresentar informações parcialmente sobrepostas. Na Árvore de Decisão, a ureia assumiu maior importância relativa, o que pode ter reduzido o destaque individual da creatinina nesse modelo. Em contrapartida, em métodos sensíveis à distribuição dos atributos no espaço de características, como o KNN, a creatinina pode contribuir de forma diferente para a separa-

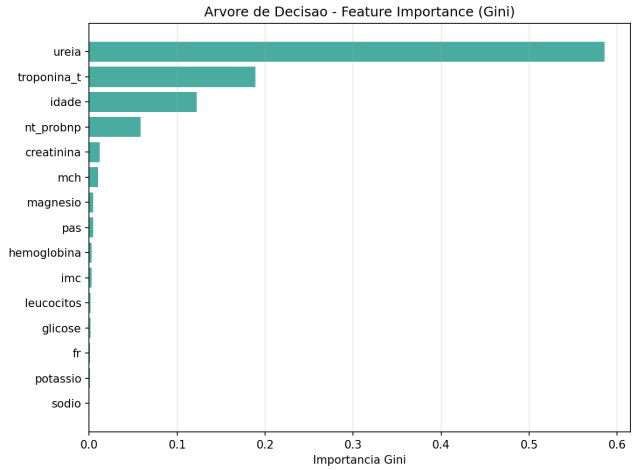

Figura 12. Importância dos atributos na Árvore de Decisão pelo Índice Gini.

ção entre admissões com e sem DCV. Dessa forma, a variação da importância da creatinina entre os modelos não indica ausência de relevância clínica, mas sim diferenças na forma como cada algoritmo utiliza as relações entre as variáveis.

O SVM não foi submetido à análise de importância por permutação em razão do elevado custo computacional do procedimento, que exige sucessivas reavaliações do modelo após a permutação individual dos valores de cada variável. Durante a execução, o processo foi interrompido após 24 horas de processamento sem que o cálculo fosse concluído, inviabilizando sua inclusão nos resultados de interpretabilidade. Ainda assim, a convergência entre os três modelos analisados foi suficiente para identificar idade, ureia, troponina T e NT-proBNP como os principais fatores associados à presença de DCV no conjunto de dados.

4.5 Discussão geral

Embora todos os modelos tenham apresentado capacidade discriminativa superior ao classificador aleatório, a análise conjunta das métricas evidencia que o desempenho de cada algoritmo varia de forma expressiva dependendo da métrica considerada, e que essa variação tem implicações diretas para a aplicabilidade clínica dos modelos.

O KNN ilustra bem esse ponto. Sua maior acurácia (85,71%) poderia sugerir, em uma análise superficial, desempenho superior aos demais. No entanto, o Recall de 19,37% revela que o modelo classificou corretamente menos de um quinto dos pacientes com DCV, gerando 21.533 falsos negativos no conjunto de teste. Em um contexto de triagem clínica, esse comportamento é clinicamente inaceitável, falsos negativos representam pacientes em risco que deixam de receber atenção ou investigação complementar.

A Árvore de Decisão apresentou o melhor equilíbrio entre as métricas analisadas, com o maior Recall (78,79%), maior F1-Score (53,77%) e maior AUC-ROC (86,81%), além do menor número de falsos negativos (5.664). A interpretabilidade intrínseca do modelo, expressa nas regras de decisão e nas importâncias pelo Índice Gini, representa um diferencial adicional em aplicações de apoio à decisão clínica, nas quais a transparência do processo decisório é tão relevante quanto o desempenho preditivo.

O SVM obteve resultados próximos aos da Árvore de Decisão em todas as métricas relevantes, constituindo uma alternativa competitiva. Sua principal limitação prática é o custo computacional, a otimização de hiperparâmetros foi restrita a uma subamostra do conjunto de treinamento e a análise de importância por permutação não pôde ser concluída após 24 horas de processamento, o que restringe sua aplicabilidade em bases clínicas extensas como o MIMIC-IV.

A Regressão Logística, por sua vez, apresentou desempenho consistente e interpretabilidade direta pelos coeficientes, sendo uma alternativa relevante quando há restrições computacionais ou exigências de auditabilidade do modelo.

Ao comparar os resultados com os trabalhos relacionados, observa-se convergência com Oliveira *et al*. [2023], que também identificaram a Árvore de Decisão como o algoritmo de melhor desempenho em classificação de DCV, e com Costalat and Tavares [2022], que evidenciaram a importância da comparação entre diferentes algoritmos supervisionados para a identificação do modelo mais adequado ao problema. Em relação a Teoh *et al*. [2026], que utilizaram o MIMIC-IV e o MIMIC-IV-ED para predição de insuficiência cardíaca, modelos baseados em árvores se destacaram em ambos os trabalhos. A diferença entre os resultados, Random Forest em Teoh *et al*. [2026] contra Árvore de Decisão simples neste estudo, é esperada, uma vez que o Random Forest combina múltiplas árvores para ampliar a capacidade preditiva às custas da interpretabilidade individual, enquanto a Árvore de Decisão simples favorece a transparência do processo decisório.

A análise de interpretabilidade reforçou a plausibilidade clínica dos modelos desenvolvidos. Variáveis como idade, ureia, troponina T e NT-proBNP foram identificadas de forma recorrente como os atributos mais relevantes por métodos distintos. Essa convergência indica que os modelos capturaram padrões coerentes com o conhecimento médico estabelecido.

Dessa forma, os resultados indicam que a Árvore de Decisão foi o modelo mais adequado ao objetivo deste estudo, por apresentar o melhor equilíbrio entre sensibilidade, capacidade discriminativa, F1-Score e interpretabilidade. Embora o KNN tenha obtido a maior acurácia, sua baixa capacidade de identificação da classe positiva limita sua utilização em cenários de triagem clínica. Dessa forma, a escolha do modelo mais apropriado não deve considerar apenas o desempenho global, mas também o impacto clínico dos erros produzidos, aspecto especialmente relevante em bases desbalanceadas, nas quais métricas agregadas podem obscurecer limitações críticas de sensibilidade.

5 Conclusão

Este trabalho desenvolveu e avaliou quatro algoritmos supervisionados de aprendizado de máquina para a predição de Infarto Agudo do Miocárdio e Insuficiência Cardíaca a partir de dados clínicos reais do MIMIC-IV, respondendo ao objetivo geral proposto. A Árvore de Decisão apresentou o melhor equilíbrio entre sensibilidade, capacidade discriminativa e interpretabilidade, com AUC-ROC de 86,81%, Recall de 78,79% e o menor número de falsos negativos entre os modelos avaliados, características determinantes para aplicações de triagem clínica. O SVM demonstrou capaci-

dade preditiva competitiva, mas com restrições práticas de escalabilidade. A Regressão Logística manteve desempenho consistente e auditável pelos coeficientes. O KNN, apesar da maior acurácia, mostrou-se inadequado para o problema ao classificar corretamente menos de um quinto dos pacientes com DCV.

A convergência entre métodos distintos de interpretabilidade na identificação de idade, ureia, troponina T e NT-proBNP como variáveis mais relevantes reforça a plausibilidade clínica dos modelos e contribui para sua transparência em contextos de apoio à decisão médica.

Este estudo apresenta limitações que devem orientar pesquisas futuras. A avaliação foi restrita ao MIMIC-IV, sem validação externa em outras bases clínicas, o que limita inferências sobre generalização para outros contextos hospitalares. O desbalanceamento entre classes foi tratado exclusivamente por ponderação, sem técnicas de reamostragem. A análise de interpretabilidade não pôde ser aplicada ao SVM devido ao custo computacional elevado, e a definição dos casos de DCV baseou-se exclusivamente em codificação CID, sem validação por critérios diagnósticos adicionais. Além disso, a comparação foi restrita a quatro algoritmos supervisionados, deixando em aberto a avaliação de outras abordagens de aprendizado de máquina que poderiam complementar ou superar os resultados obtidos.

Como continuidade deste estudo, pretende-se aprofundar a investigação sobre modelos preditivos aplicados a doenças cardiovasculares, explorando novas estratégias de balanceamento de classes, seleção de atributos, validação externa e comparação com modelos ensemble e técnicas de aprendizado profundo. Além disso, busca-se avançar no desenvolvimento de soluções preditivas mais precisas, confiáveis e interpretáveis, capazes de apoiar a identificação precoce de pacientes em risco e contribuir para aplicações futuras de suporte à decisão clínica em cenários reais de saúde.

Referências

Costalat, T. R. M. and Tavares, G. F. (2022). Machine learning techniques comparison for risk assessment of cardiovascular disease development by health indicators / comparação de técnicas de aprendizagem de máquinas para avaliação de risco de desenvolvimento de doença cardiovascular por indicadores de saúde. *Brazilian Journal of Development*, 8:6851–6862. Acesso em: 10 dez. 2025.

Fawcett, T. (2006). Introduction to roc analysis. *Pattern Recognition Letters*, 27:861–874. Acesso em: 12 nov. 2025. DOI: 10.1016/j.patrec.2005.10.010.

Ferreira, A., Almeida, M., and Castilho, F. (2025). Comparativo de precisão entre algoritmos de machine learning para previsão de doenças cardiovasculares: um mapeamento sistemático. In *Anais da I Escola Regional de Sistemas de Informação de Mato Grosso*, pages 65–72, Porto Alegre, RS, Brasil. SBC. Acesso em: 10 dez. 2025. DOI: 10.5753/ersimt.2025.8779.

James, G., Witten, D., Hastie, T., and Tibshirani, R. (2023). *An Introduction to Statistical Learning: with Applications in Python*. Springer Texts in Statistics. Springer, New York. Acesso em: 04 dez. 2025.

Oliveira, M. C., Ferreira, L. V. B., and de Oliveira Barreiros, M. (2023). Classificação de doenças cardiovasculares utilizando aprendizado de máquina. *Revista Multidebates*, 7. Acesso em: 10 dez. 2025.

Organização Mundial da Saúde (2025). Cardiovascular diseases (cvds). Acesso em: 10 nov. 2025.

Paixão, G. M. d. M., Santos, B. C., Araujo, R. M. d., Ribeiro, M. H., Moraes, J. L. d., and Ribeiro, A. L. (2022). Machine learning na medicina: Revisão e aplicabilidade. *Arquivos Brasileiros de Cardiologia*, 118(3):550–559. Acesso em: 12 nov. 2025.

Singh, M., Kumar, A., Khanna, N. N., Laird, J. R., Nicolaides, A., Faa, G., Johri, A. M., Mantella, L. E., Fernandes, J. F. E., Teji, J. S., Singh, N., Fouda, M. M., Singh, R., Sharma, A., Kitas, G., Rathore, V., Singh, I. M., Tadepalli, K., Al-Maini, M., Isenovic, E. R., Chaturvedi, S., Garg, D., Paraskevas, K. I., Mikhailidis, D. P., Viswanathan, V., Kalra, M. K., Ruzsa, Z., Saba, L., Laine, A. F., Bhatt, D. L., and Suri, J. S. (2024). Artificial intelligence for cardiovascular disease risk assessment in personalised framework: a scoping review. *eClinical Medicine*, 73:102660. Acesso em: 12 nov. 2025.

Sokolova, M. and Lapalme, G. (2009). A systematic analysis of performance measures for classification tasks. *Information Processing Management*, 45:427–437. Acesso em: 12 nov. 2025.

Teodoro, G. G. F., Lopes, M. d. C., Christ, Y. K., Lucas, T. J., and Bertazzoli, C. E. S. (2025). Análise do desempenho de classificadores supervisionados na detecção de fraude em transações por cartões de crédito. *RETEC – Revista de Tecnologias*, 18(1):24–39. Acesso em: 12 nov. 2025.

Teoh, J. R., Hasikin, K., Wong, J. H. D., Ng, W. L., Lee, K. W., Kiew, L. V., Wu, X., and Lai, K. W. (2026). Predicting heart failure using MIMIC-IV and MIMIC-IV-ED: a comparative study of machine learning and deep learning models. *PeerJ Computer Science*, 12:e3743. Acesso em: 01 jun. 2026. DOI: 10.7717/peerj-cs.3743.