

UNIVERSIDADE FEDERAL DE JUIZ DE FORA
INSTITUTO DE CIÊNCIAS EXATAS
BACHARELADO EM SISTEMAS DE INFORMAÇÃO

Predição de Insuficiência Cardíaca em Tempo Real: Uma Abordagem para Ambientes de Vida Assistida

Lucas Duarte Chaves

JUIZ DE FORA
JANEIRO, 2026

Predição de Insuficiência Cardíaca em Tempo Real: Uma Abordagem para Ambientes de Vida Assistida

LUCAS DUARTE CHAVES

Universidade Federal de Juiz de Fora

Instituto de Ciências Exatas

Departamento de Ciência da Computação

Bacharelado em Sistemas de Informação

Orientador: Luciana Conceição Dias Campos

Coorientador: Victor Stroele de Andrade Menezes

JUIZ DE FORA

JANEIRO, 2026

Aos meus amigos e irmãos.

Aos pais, pelo apoio e sustento.

Resumo

Este trabalho de conclusão de curso investiga o uso de modelos de aprendizado de máquina para a predição de insuficiência cardíaca em tempo real, com foco na realidade do sistema de saúde brasileiro. O objetivo principal é treinar e avaliar modelos preditivos utilizando uma base de dados que reflita os parâmetros clínicos adotados nacionalmente. A solução considera o uso de computação de borda para realizar o processamento local dos sinais vitais, reduzindo a dependência de conexão contínua com a internet e a latência na geração de alertas. Os resultados demonstraram que o modelo *XGBoost* obteve o melhor desempenho geral, com acurácia de 82% e *F1-Score* de 83%, enquanto o modelo SVM destacou-se pela maior sensibilidade (86% de *recall*). A implementação em dispositivos móveis validou a eficiência da computação de borda, reduzindo a latência de inferência em 99,3% em comparação ao processamento remoto e garantindo respostas imediatas para o monitoramento de pacientes.

Palavras-chave: Predição de insuficiência cardíaca, Aprendizado de máquina, Computação de borda, Internet das Coisas (IoT).

Abstract

This final project investigates the use of machine learning models for real-time heart failure risk prediction, with a focus on the context of the Brazilian healthcare system. The main objective is to train and evaluate predictive models using a dataset that reflects the clinical parameters commonly adopted in national medical practice. The proposed solution incorporates edge computing to enable local processing of vital signs, reducing dependence on continuous internet connectivity and minimizing latency in alert generation. The results demonstrated that the XGBoost model achieved the best overall performance, with 82% accuracy and an 83% F1-Score, while the SVM model stood out with the highest sensitivity (86% recall). Implementation on mobile devices validated the efficiency of edge computing, reducing inference latency by 99.3% compared to remote processing and ensuring immediate responses for patient monitoring.

Keywords: Heart failure prediction, Machine learning, Edge computing, Internet of Things (IoT).

Conteúdo

Lista de Figuras	6
Lista de Tabelas	7
Lista de Abreviações	8
1 Introdução	9
1.1 Apresentação do tema	9
1.2 Motivação	10
1.3 Objetivos	12
1.3.1 Objetivos gerais	12
1.3.2 Objetivos específicos	12
1.4 Organização do trabalho	13
2 Fundamentação teórica	14
2.1 Doenças cardiovasculares	14
2.2 Aprendizado de máquina para predição em saúde	15
2.2.1 Balanceamento de classes	17
2.2.2 Validação Cruzada	18
2.2.3 Métricas de avaliação	19
2.2.4 <i>ONNX</i>	20
2.3 Computação de borda	21
3 Revisão bibliográfica	23
3.1 PICO(t)	23
3.2 Questões de Pesquisa	24
3.3 Palavras-chave	24
3.4 String de busca	25
3.5 Critérios de inclusão e exclusão	26
3.6 Seleção e avaliação	26
3.7 Resultados	27
3.8 Análise dos resultados	32
4 Materiais e métodos	34
4.1 Descrição dos dados	35
4.2 Modelagem Preditiva para Insuficiência Cardíaca	39
4.2.1 Pré-processamento dos dados	39
4.2.2 Treinamento dos modelos	44
4.2.3 Implementação em Dispositivos Móveis	46
5 Resultados	49
5.1 Resultados do treinamento dos modelos	49
5.2 Resultados da aplicação em Dispositivos Móveis	52
6 Considerações Finais e Trabalhos Futuros	56

Lista de Figuras

3.1	Fluxograma PRISMA 2020	27
4.1	Fluxograma das etapas do trabalho	35
4.2	Matriz de correlação entre variáveis numéricas	38
4.3	Distribuição original das classes	42
4.4	Distribuição das classes após a junção	43
5.1	Matrizes de Confusão dos Modelos Classificatórios	50
5.2	Curvas ROC	52
5.3	Funcionamento do aplicativo em <i>smartphone</i>	53
5.4	Funcionamento do aplicativo em <i>smartwatch</i>	54
5.5	Comparação da latência de execução (em ms) em diferentes ambientes . . .	55

Lista de Tabelas

3.1	Strings de busca por biblioteca.	25
3.2	Lista de estudos selecionados.	28
4.1	Análise das variáveis utilizadas na modelagem	41
4.2	Hiperparâmetros avaliados durante o treinamento dos modelos	45
5.1	Desempenho dos modelos de classificação	51
5.2	Melhores hiperparâmetros selecionados para cada modelo	51

Lista de Abreviações

DCC Departamento de Ciência da Computação

UFJF Universidade Federal de Juiz de Fora

1 Introdução

1.1 Apresentação do tema

As doenças cardiovasculares continuam sendo a principal causa de morte no Brasil (USP, 2024), representando um grave problema de saúde pública. Nesse contexto, soluções tecnológicas capazes de promover monitoramento contínuo e preditivo da saúde cardíaca são cada vez mais importantes. A utilização de modelos de aprendizado de máquina embarcados em dispositivos móveis tem se mostrado uma abordagem promissora para assistência médica em tempo real (ABDELLATIF et al., 2019). Essa combinação permite reduzir significativamente a latência de resposta dos modelos e a dependência de infraestruturas centralizadas, viabilizando o monitoramento preditivo em regiões com infraestrutura de rede limitada ou instável.

Em alguns sistemas de monitoramento de saúde, o processamento das informações ocorre de forma centralizada, a partir da transmissão dos dados coletados para servidores em nuvem (AMIN; HOSSAIN, 2021). Embora essa abordagem apresente um maior poder computacional, ela traz limitações para aplicações em tempo real, como a predição de insuficiência cardíaca. Entre os principais obstáculos estão a latência da rede, a dependência de conectividade estável, a sobrecarga nos servidores e os riscos associados à privacidade dos dados sensíveis dos pacientes (ROMAN; LOPEZ; MAMBO, 2018).

Em contextos como os ambientes de vida assistida, nos quais o tempo de resposta a episódios clínicos pode ser decisiva, esses desafios comprometem a eficácia dos sistemas baseados em nuvem. Atrasos na transmissão ou interrupções na conexão podem impedir a emissão de alertas em tempo hábil, colocando em risco a segurança dos usuários. Além disso, regiões com infraestrutura de rede precária — como áreas remotas ou economicamente desfavorecidas — enfrentam dificuldades adicionais para a adoção de soluções que dependem da internet.

Estudos recentes que aliam o aprendizado de máquina à computação de borda, como o que motiva este trabalho (ALMEIDA et al., 2025), apresentam resultados pro-

missores na detecção precoce de insuficiência cardíaca em tempo real. Contudo, a predominância do uso de bases de dados internacionais nos estudos existentes compõe um obstáculo na aplicabilidade prática dessas soluções no contexto clínico brasileiro, visto que os perfis clínicos das populações em questão nas bases de dados internacionais podem divergir significativamente do perfil clínico nacional. A escassez de pesquisas baseadas em dados nacionais evidencia, portanto, uma lacuna a ser explorada para que essas ferramentas possam ser adotadas de forma segura e eficaz no sistema de saúde brasileiro.

1.2 Motivação

A motivação central deste trabalho é contribuir em uma solução tecnológica para a predição de insuficiência cardíaca em tempo real, voltada para a população de áreas com acesso limitado à internet no Brasil. Apesar do aumento da cobertura de internet no Brasil, ainda existem lacunas no acesso e na qualidade da conexão, especialmente em áreas rurais (IBGE, 2025). Nesse contexto, a proposição de soluções que realizem o processamento local dos dados mostra-se essencial para mitigar os impactos da conectividade limitada em determinadas regiões do país. Ao reduzir a dependência de conectividade plena com servidores remotos, torna-se viável o monitoramento contínuo de sinais vitais mesmo em cenários de infraestrutura precária. Dessa forma, a adoção de tecnologias embarcadas em dispositivos móveis pode ampliar o alcance do monitoramento preventivo da saúde cardíaca, promovendo maior inclusão e equidade no campo da saúde.

O artigo *Real-Time Heart Failure Prediction: An Approach for Ambient Assisted Living* (ALMEIDA et al., 2025), que motiva este trabalho, propõe uma arquitetura focada na predição de insuficiência cardíaca em tempo real, integrando aprendizado de máquina e computação de borda. O estudo parte da limitação de abordagens tradicionais baseadas em servidores centralizados, que sofrem com latência, dependência de conectividade e riscos à privacidade dos dados. Como alternativa, os autores desenvolveram uma ferramenta para dispositivos móveis que permite que as predições sejam realizadas localmente. Para isso, foram previamente treinados e avaliados algoritmos como *Random Forest*, *K-Nearest Neighbors* e *Logistic Regression*, utilizando um conjunto de dados clínicos.

O conjunto de dados adotado no estudo resulta da integração de quatro bases de

dados distintas, as quais compartilham variáveis em comum. Essas bases de dados são: um conjunto de dados fornecido pelo *Hungarian Institute of Cardiology* de Budapeste; um conjunto de dados fornecido pelo *University Hospital Zurich* da Suíça; um outro conjunto de dados de origem suíça, fornecido pelo *University Hospital Basel*; e um conjunto de dados fornecido pelo *V.A. Medical Center*, em conjunto com a *Cleveland Clinic Foundation*, dos Estados Unidos.

Após os testes, o modelo com melhor desempenho foi exportado e implantado em um aplicativo para dispositivos móveis, onde é possível realizar predições de maneira local, sem a necessidade de conexão com um servidor.

Embora os resultados do estudo tenham demonstrado a viabilidade técnica da solução proposta, o conjunto de dados utilizado para o treinamento e a validação dos modelos não contempla variáveis plenamente condizentes com os parâmetros empregados no diagnóstico de insuficiência cardíaca na prática clínica brasileira. Essa constatação fundamenta-se no desenvolvimento do presente trabalho, durante o qual foi realizada uma Entrevista com Especialista (*Subject Matter Expert – SME Interview*) com a médica cardiologista Dra. Lilian Marzullo de Carvalho Bramante. A partir dessa entrevista, foi possível identificar que poucos dos atributos presentes no conjunto de dados são realmente utilizados no diagnóstico de insuficiência cardíaca, enquanto a outra parcela das variáveis não são empregadas na prática médica no Brasil. Ademais, foi possível constatar que diversos parâmetros clínicos relevantes, comumente avaliados pelos profissionais de saúde no Brasil, não estão contemplados no referido conjunto de dados.

Diante desse cenário, o presente trabalho propõe a validação da predição de insuficiência cardíaca a partir de uma base de dados que contenha atributos clínicos e faixas de valores alinhados às recomendações fornecidas pela especialista entrevistada. A adoção de variáveis compatíveis com as práticas do sistema de saúde nacional visa ampliar a aplicabilidade prática da ferramenta proposta, potencializando sua utilização em cenários reais de monitoramento e apoio à decisão médica.

1.3 Objetivos

1.3.1 Objetivos gerais

O objetivo deste trabalho é desenvolver uma ferramenta de predição de insuficiência cardíaca em tempo real, baseada em modelos de *machine learning* e computação de borda, que apresente uma latência de resposta mínima, em decorrência da realização do processamento local em vez da execução em nuvem. Esse objetivo leva em conta a utilização de um conjunto de dados com atributos compatíveis com parâmetros clínicos empregados na prática médica brasileira, de modo a investigar sua aplicabilidade e limitações quando considerada a realidade do sistema de saúde nacional.

1.3.2 Objetivos específicos

Para alcançar os objetivos gerais propostos, este trabalho se desdobra nos seguintes objetivos específicos:

1. **Seleção da base de dados:** Buscar bases de dados públicas que contenham atributos e valores compatíveis com os parâmetros clínicos utilizados no sistema de saúde brasileiro, priorizando registros públicos ou de instituições parceiras que respeitem os princípios éticos e a Lei Geral de Proteção de Dados (LGPD)(Presidência da República, 2018).
2. **Desenvolvimento dos modelos:** Selecionar, treinar e ajustar modelos de aprendizado de máquina adequados ao problema de predição de insuficiência cardíaca, considerando diferentes abordagens e configurações. Depois de treinados, os modelos serão avaliados quanto ao desempenho preditivo e à eficiência computacional, com foco em sua posterior implementação em ambientes de computação de borda, garantindo viabilidade de execução em dispositivos com recursos limitados.
3. **Análise dos resultados:** Avaliar o desempenho dos modelos de aprendizado de máquina por meio de métricas apropriadas, como acurácia, precisão, *recall*, *F1-score* e curvas ROC, de modo a identificar o modelo mais adequado para a predição

de insuficiência cardíaca. Por fim, serão discutidas as limitações da abordagem proposta e apresentadas possíveis direções para aprimoramento da ferramenta.

1.4 Organização do trabalho

Diante da introdução apresentada, este trabalho de conclusão de curso está organizado da seguinte forma: no Capítulo 2 é apresentada a fundamentação teórica, com os conceitos essenciais para a compreensão do estudo. No Capítulo 3 é realizada a revisão bibliográfica, abordando as referências e estudos relacionados a este trabalho de conclusão de curso. No Capítulo 4 é detalhado o processo de desenvolvimento do trabalho, e o Capítulo 5 mostra os resultados obtidos e as análises realizadas. Por fim, o Capítulo 6 apresenta as conclusões do estudo, bem como sugestões para possíveis trabalhos futuros.

2 Fundamentação teórica

Neste capítulo são apresentados os conceitos para a compreensão dos temas abordados neste trabalho. Inicialmente, discutiremos sobre doenças cardiovasculares e insuficiência cardíaca. Também será abordado o uso de Aprendizado de Máquina para predição em saúde e sua relação com Computação de Borda.

2.1 Doenças cardiovasculares

As doenças cardiovasculares representam uma das principais causas de morbidade e mortalidade em nível mundial (World Health Organization, 2024), constituindo um dos maiores desafios para os sistemas de saúde atuais. No território brasileiro, essas enfermidades figuram entre as principais causas de óbitos. Entre essas condições, a insuficiência cardíaca destaca-se por seu caráter crônico, progressivo e pela elevada taxa de reinternações hospitalares (KHAN et al., 2021).

A insuficiência cardíaca é caracterizada pela incapacidade do coração de bombear sangue de forma adequada para suprir as necessidades metabólicas do organismo. Essa condição pode ser agravada por diversos fatores de risco, como hipertensão arterial, diabetes mellitus, obesidade, sedentarismo e histórico familiar de doenças cardiovasculares. Em muitos casos, a evolução da doença ocorre de forma silenciosa, tornando o diagnóstico precoce um fator essencial para a redução de complicações e mortalidade. (Sociedade Brasileira de Cardiologia, 2018)

Nesse cenário, estratégias de monitoramento contínuo e preditivo tornam-se fundamentais para identificar o agravamento da condição antes que eventos críticos ocorram. A utilização de dados clínicos e fisiológicos, coletados de maneira contínua, possibilita o monitoramento preventivo no acompanhamento de pacientes em risco cardiovascular, contribuindo para intervenções mais rápidas (STEHLIK et al., 2020).

2.2 Aprendizado de máquina para predição em saúde

O aprendizado de máquina (*Machine Learning*) é uma área da inteligência artificial dedicada ao desenvolvimento de métodos computacionais capazes de identificar padrões em dados e melhorar seu desempenho em tarefas específicas a partir da experiência. Diferentemente de abordagens tradicionais baseadas em regras explícitas, os algoritmos de *Machine Learning* aprendem automaticamente a partir dos dados, ajustando seus parâmetros de forma iterativa para produzir previsões ou decisões cada vez mais precisas (MITCHELL, 1997).

Em aplicações na área da saúde, o aprendizado de máquina pode ser utilizado para apoiar o diagnóstico e o monitoramento de pacientes, uma vez que permite explorar grandes volumes de dados clínicos e fisiológicos de maneira eficiente. A capacidade de extrair relações complexas entre variáveis torna essas técnicas especialmente adequadas para problemas médicos, nos quais múltiplos fatores podem influenciar o estado de saúde de um indivíduo (RAJKOMAR; DEAN; KOHANE, 2019).

De forma geral, os algoritmos de aprendizado de máquina podem ser organizados em três categorias principais: aprendizado supervisionado, aprendizado não supervisionado e aprendizado por reforço. No aprendizado supervisionado, o modelo é treinado a partir de um conjunto de dados rotulado, no qual cada instância de entrada está associada a uma saída conhecida. O objetivo do algoritmo é aprender uma função que mapeie corretamente as entradas para as saídas, possibilitando a realização de previsões sobre novos dados. Essa abordagem é amplamente utilizada em problemas de classificação e regressão. Entre os algoritmos mais conhecidos desse grupo destacam-se a Regressão Logística, Máquinas de Vetores de Suporte (*Support Vector Machines* – SVM), Árvores de Decisão, *Random Forest*, *Gradient Boosting* e Redes Neurais Artificiais (RUSSELL; NORVIG, 2020).

Os modelos de aprendizado de máquina adotados utilizam abordagens distintas para a identificação de padrões nos dados.

A **Regressão Logística** se baseia na combinação linear das variáveis de entrada, cujos resultados são transformados por meio de uma função logística, que mapeia os valores para o intervalo entre 0 e 1. Esse valor pode ser interpretado como a probabilidade de

uma instância pertencer a uma determinada classe. Durante o treinamento, os parâmetros do modelo são ajustados de forma a minimizar uma função de custo, buscando maximizar a separação entre as classes (JR; LEMESHOW; STURDIVANT, 2013).

O ***Random Forest*** é um algoritmo baseado em um conjunto de árvores de decisão. Seu princípio fundamental consiste em treinar múltiplas árvores de forma independente, utilizando subconjuntos aleatórios dos dados e das variáveis de entrada. Cada árvore produz uma predição individual e, no caso de classificação, o resultado final é obtido por meio de votação majoritária entre as árvores. Essa estratégia reduz o risco de sobreajuste (*overfitting*) e torna o modelo mais robusto a variações nos dados, além de melhorar sua capacidade de generalização (BREIMAN, 2001).

O ***Gradient Boosting*** é um método de aprendizado em conjunto que constrói o modelo de forma sequencial, adicionando árvores de decisão uma a uma. Diferentemente do *Random Forest*, cada nova árvore é treinada para corrigir os erros cometidos pelo conjunto de árvores anteriores. Esse processo é orientado pela minimização de uma função de perda, utilizando o gradiente dessa função para guiar o aprendizado, o que dá origem ao nome do algoritmo. Como resultado, o *Gradient Boosting* é capaz de produzir modelos altamente expressivos, embora seja sensível à escolha de hiperparâmetros e ao risco de sobreajuste quando não adequadamente configurado (FRIEDMAN, 2001).

A ***Support Vector Machine*** (SVM) é um algoritmo cujo objetivo é encontrar um hiperplano que separe as classes de forma ótima no espaço das características. Esse hiperplano é definido de maneira a maximizar a margem, ou seja, a distância entre os exemplos mais próximos de cada classe, conhecidos como vetores de suporte. Por meio do uso de funções de *kernel*, a SVM pode projetar os dados em espaços de maior dimensionalidade, permitindo a separação de padrões não linearmente separáveis no espaço original (CORTES; VAPNIK, 1995).

O ***XGBoost*** (*Extreme Gradient Boosting*) é uma implementação otimizada do método de *Gradient Boosting*, projetada para oferecer maior eficiência computacional e melhor desempenho. O algoritmo incorpora técnicas avançadas, como regularização para controle da complexidade do modelo, paralelização do treinamento e tratamento eficiente de dados ausentes. Assim como no *Gradient Boosting tradicional*, o *XGBoost* constrói o

modelo de forma incremental, adicionando árvores que corrigem os erros das anteriores, porém com mecanismos adicionais que tornam o processo mais rápido, escalável e robusto (CHEN; GUESTRIN, 2016).

Dentro do contexto deste trabalho, o foco está no aprendizado supervisionado aplicado a tarefas de classificação. A classificação consiste em atribuir uma instância de dados a uma ou mais classes previamente definidas, com base em suas características de entrada. Em problemas de classificação binária, o objetivo é distinguir entre duas classes, enquanto na classificação multiclasse existem três ou mais categorias possíveis. No caso da predição de risco de insuficiência cardíaca, a classificação permite identificar se um paciente pertence a um grupo de alto risco ou não.

2.2.1 Balanceamento de classes

O desbalanceamento entre classes é um problema recorrente em tarefas de classificação, especialmente em aplicações da área da saúde, nas quais eventos de interesse — como doenças ou condições críticas — tendem a ocorrer com menor frequência. Esse cenário pode levar os modelos de aprendizado de máquina a apresentarem viés em favor da classe majoritária, comprometendo a capacidade de identificar corretamente os casos mais relevantes (HE; GARCIA, 2009).

Entre as técnicas existentes para mitigar esse problema, destaca-se o *Synthetic Minority Over-sampling Technique* (SMOTE). O SMOTE consiste na geração de novas amostras sintéticas da classe minoritária a partir da interpolação entre exemplos reais e seus vizinhos mais próximos no espaço de atributos. Diferentemente de métodos de *oversampling* baseados na simples replicação de instâncias, o SMOTE cria exemplos artificiais mais variados, contribuindo para uma melhor representação da classe minoritária (CHAWLA et al., 2002).

Ao equilibrar a distribuição das classes, essa técnica favorece um processo de aprendizado mais justo, permitindo que os classificadores aprendam padrões relevantes de ambas as classes. Como resultado, observa-se uma melhora na sensibilidade do modelo para a classe minoritária, aspecto fundamental em sistemas de apoio à decisão médica, nos quais a detecção de casos críticos é prioritária.

2.2.2 Validação Cruzada

A validação cruzada é uma técnica utilizada em aprendizado de máquina para estimar o desempenho de um modelo de maneira mais confiável. Diferentemente da simples divisão dos dados em conjuntos de treino e teste, a validação cruzada particiona o conjunto de dados em k subconjuntos (ou *folds*), de modo que o modelo é treinado $k-1$ vezes em subconjuntos distintos e validado no subconjunto restante. Esse processo é repetido até que todos os *folds* tenham sido utilizados como conjunto de validação, e o desempenho final é obtido a partir da média das métricas avaliadas. Essa abordagem reduz o viés associado a uma única divisão dos dados e fornece uma estimativa mais estável da capacidade de generalização do modelo (KOHAVI, 1995).

Uma variação importante dessa técnica é a validação cruzada estratificada, conhecida como *Stratified K-Fold*. Nesse método, a divisão dos dados em *folds* preserva a proporção original das classes em cada partição. Essa característica é especialmente relevante em problemas de classificação, principalmente quando há desbalanceamento entre as classes, pois garante que cada *fold* seja representativo da distribuição global dos dados, evitando avaliações enviesadas do desempenho do modelo (Scikit-learn, 2024b).

O processo de ajuste de hiperparâmetros é fundamental para maximizar o desempenho dos algoritmos de aprendizado de máquina. Nesse contexto, o *Grid Search* é uma estratégia que consiste em avaliar exaustivamente todas as combinações possíveis de um conjunto pré-definido de hiperparâmetros. Para cada combinação, o modelo é treinado e avaliado, permitindo identificar aquela que produz os melhores resultados de acordo com uma métrica específica (BERGSTRA; BENGIO, 2012).

O *GridSearchCV* (*Grid Search Cross-Validation*) estende essa abordagem ao integrar o *Grid Search* com a validação cruzada. Essa técnica automatiza simultaneamente o processo de busca pelos melhores hiperparâmetros e a avaliação do modelo por meio de validação cruzada, garantindo uma seleção mais robusta e menos suscetível a variações aleatórias nos dados. Na prática, o *GridSearchCV* executa o *Grid Search* utilizando validação cruzada — frequentemente combinada com *Stratified K-Fold* em problemas de classificação — assegurando que a configuração escolhida apresente bom desempenho médio em diferentes subconjuntos dos dados (Scikit-learn, 2024a).

2.2.3 Métricas de avaliação

A avaliação do desempenho de modelos de classificação é realizada por meio de métricas derivadas da matriz de confusão, uma estrutura que sintetiza a relação entre as classes reais e as classes preditas pelo modelo. A matriz de confusão é composta por quatro elementos fundamentais: verdadeiros positivos (*True Positives* – TP), verdadeiros negativos (*True Negatives* – TN), falsos positivos (*False Positives* – FP) e falsos negativos (*False Negatives* – FN). A partir desses valores, é possível calcular diferentes métricas que fornecem uma visão abrangente do desempenho do classificador. A partir desses elementos, algumas métricas podem ser calculadas, como as descritas a seguir:

Acurácia: Representa a proporção total de previsões corretas em relação ao número total de amostras avaliadas, sendo calculada pela Equação 2.1.

$$\text{Acurácia} = \frac{TP + TN}{TP + TN + FP + FN} \quad (2.1)$$

Precisão: Mede a proporção de previsões positivas corretas em relação ao total de previsões positivas realizadas pelo modelo, conforme a Equação 2.2. Essa métrica indica o quão confiáveis são as predições positivas.

$$\text{Precisão} = \frac{TP}{TP + FP} \quad (2.2)$$

Recall: Também conhecido como sensibilidade, representa a proporção de casos positivos corretamente identificados pelo modelo em relação ao total de casos positivos reais, conforme a Equação 2.3. Essa métrica é particularmente relevante em aplicações médicas, nas quais a não detecção de casos críticos pode ter consequências severas.

$$\text{Recall} = \frac{TP}{TP + FN} \quad (2.3)$$

F1-Score: Corresponde à média harmônica entre a precisão e o recall, sendo uma métrica adequada para cenários com classes desbalanceadas. Seu cálculo é apre-

sentado na Equação 2.4.

$$\text{F1-Score} = 2 \times \frac{\text{Precisão} \times \text{Recall}}{\text{Precisão} + \text{Recall}} \quad (2.4)$$

Curva ROC-AUC: Avalia a capacidade do modelo em distinguir entre as classes ao considerar diferentes limiares de decisão. A Curva ROC (*Receiver Operating Characteristic*) é obtida a partir da relação entre a taxa de verdadeiros positivos (*True Positive Rate* – TPR) e a taxa de falsos positivos (*False Positive Rate* – FPR), definidas pelas Equações 2.5 e 2.6.

$$\text{TPR} = \frac{TP}{TP + FN} \quad (2.5)$$

$$\text{FPR} = \frac{FP}{FP + TN} \quad (2.6)$$

A área sob a Curva ROC, denominada AUC (*Area Under the Curve*), resume o desempenho do modelo em um único valor escalar, representando a probabilidade de o classificador atribuir uma pontuação maior a uma instância positiva do que a uma instância negativa. Valores de AUC próximos de 1 indicam excelente capacidade de discriminação entre as classes, enquanto valores próximos de 0,5 sugerem desempenho equivalente a uma classificação aleatória.

2.2.4 ONNX

O *Open Neural Network Exchange* (ONNX) é um formato aberto e padronizado para a representação de modelos de aprendizado de máquina e aprendizado profundo, desenvolvido com o objetivo de promover a interoperabilidade entre diferentes *frameworks* e plataformas. Por meio do ONNX, modelos treinados em bibliotecas amplamente utilizadas, como *scikit-learn*, *PyTorch* e *TensorFlow*, podem ser exportados para um formato comum, independente da ferramenta de origem, facilitando sua reutilização e implantação em diferentes ambientes computacionais (BAI et al., 2019).

O formato ONNX descreve a arquitetura do modelo, seus parâmetros treinados e o grafo computacional das operações envolvidas, permitindo que o modelo seja executado

de forma consistente em diversos dispositivos, incluindo servidores, computadores pessoais e sistemas embarcados. Essa característica é particularmente relevante em aplicações que demandam portabilidade, eficiência e baixo acoplamento entre as etapas de treinamento e inferência.

O *ONNX Runtime* é um mecanismo de inferência de alto desempenho projetado para executar modelos no formato ONNX de maneira otimizada. Ele oferece suporte a múltiplos provedores de execução, como CPU, GPU e aceleradores específicos, além de realizar otimizações automáticas no grafo computacional, incluindo fusão de operações e gerenciamento eficiente de memória. Essas otimizações contribuem para a redução da latência e do consumo de recursos, tornando o ONNX Runtime adequado para aplicações em tempo real (Microsoft, 2021).

No contexto de dispositivos móveis e computação de borda, o uso combinado do ONNX e do ONNX Runtime possibilita a execução local de modelos de aprendizado de máquina, eliminando a necessidade de comunicação constante com servidores remotos. Essa abordagem reduz a latência, melhora a confiabilidade do sistema em ambientes com conectividade limitada e contribui para a preservação da privacidade dos dados, uma vez que as informações sensíveis permanecem no próprio dispositivo do usuário.

2.3 Computação de borda

Como evidenciado por (SHI; DUSTDAR, 2016), computação de borda é um paradigma computacional no qual o processamento e a análise dos dados são realizados próximos à sua fonte de geração, como dispositivos finais, sensores inteligentes ou servidores locais, em vez de depender exclusivamente de centros de dados remotos em nuvem. A computação de borda busca estender os recursos computacionais e de armazenamento para a extremidade da rede, reduzindo a latência, o consumo de banda e a dependência de conectividade contínua com a infraestrutura centralizada.

Em sistemas convencionais centralizados, os dados coletados são enviados para a nuvem para processamento, o que pode introduzir atrasos significativos na geração de respostas, além de tornar o sistema dependente de uma conexão estável com a internet. Em cenários críticos, como o monitoramento de condições de saúde, essa latência pode

comprometer a emissão de alertas e, consequentemente, a tomada de decisões. Além disso, o envio constante de grandes volumes de dados pode sobrecarregar a infraestrutura de rede e aumentar os custos operacionais. (WANG et al., 2020)

A computação de borda mitiga essas limitações ao permitir que parte ou a totalidade do processamento seja realizada localmente. Com isso, as decisões podem ser tomadas de forma mais rápida e independente da conectividade com a nuvem, garantindo maior continuidade do serviço mesmo em ambientes com acesso limitado à internet. Essa característica é particularmente relevante para aplicações que exigem respostas imediatas e operação contínua.

Outro aspecto relevante da computação de borda diz respeito à segurança e à privacidade dos dados. Ao reduzir a necessidade de transmissão de informações sensíveis para servidores remotos, diminui-se a exposição a riscos de interceptação e vazamento de dados (ROMAN; LOPEZ; MAMBO, 2018).

3 Revisão bibliográfica

Este capítulo apresenta o processo de mapeamento sistemático conduzido com o objetivo de identificar e organizar evidências científicas relacionadas à predição de insuficiência cardíaca em tempo real, por meio de técnicas de aprendizado de máquina embarcadas em dispositivos móveis e computação de borda.

A seguir, são descritos os elementos do modelo PICO(T), as questões de pesquisa que guiaram a busca, as palavras-chave utilizadas, as strings aplicadas em bases científicas, os critérios de inclusão e exclusão, o processo de seleção e avaliação dos estudos e o preenchimento do protocolo PRISMA, que registra as etapas de identificação, triagem e inclusão dos estudos.

3.1 PICO(t)

Para guiar a formulação das questões de pesquisa e estruturar o processo de busca, foi adotado o modelo PICO(T), que auxilia na definição dos principais elementos do problema investigado:

P (População / Problema): Pacientes em risco de insuficiência cardíaca.

I (Intervenção): Predição de insuficiência cardíaca usando modelos de aprendizado de máquina embarcados em dispositivos móveis (computação de borda).

C (Comparação): Predição de insuficiência cardíaca usando modelos baseados em nuvem (computação centralizada).

O (Outcome / Desfecho): Maior precisão na detecção precoce e menor latência para alertas.

Assim, a pergunta de pesquisa pode ser formulada da seguinte forma: *“Em pacientes com risco de insuficiência cardíaca (P), o uso de modelos embarcados de aprendizado de máquina (I), em comparação com modelos baseados em nuvem (C), resulta em maior precisão e menor latência na detecção precoce (O)?”*

3.2 Questões de Pesquisa

Com base nos objetivos deste estudo e nas lacunas identificadas na literatura, foram elaboradas as seguintes questões de pesquisa que orientam o mapeamento sistemático:

1. Quais são os principais algoritmos e técnicas de aprendizado de máquina utilizados na predição de insuficiência cardíaca em tempo real?
2. Como a computação de borda (*edge computing*) tem sido utilizada para reduzir latência e melhorar a confiabilidade em sistemas de monitoramento de saúde?
3. Quais sensores vestíveis e dispositivos IoT são mais utilizados para coleta de dados em sistemas de predição de insuficiência cardíaca?
4. Quais métricas e métodos de avaliação são usados para validar a acurácia e a responsividade desses modelos em ambientes de computação de borda?
5. Quais *frameworks*, plataformas e ferramentas são utilizados para implementar sistemas de predição em dispositivos móveis?
6. Quais são os principais desafios e limitações para implementar modelos de aprendizado de máquina embarcados em dispositivos móveis para detecção de insuficiência cardíaca?
7. Como a integração de sensores, dispositivos móveis e computação de borda influencia a escalabilidade e a privacidade dos dados de saúde?

3.3 Palavras-chave

Os principais termos que refletem os temas centrais do trabalho podem ser listados como: *Heart failure prediction*, *Real-time*, *Machine learning*, *Edge computing*, *Internet of Things (IoT)*.

3.4 String de busca

A partir da elaboração da pergunta de pesquisa e da escolha das palavras-chave, foram feitas buscas em três bibliotecas digitais – *Scopus*, *IEEE* e *Web of Science* – consideradas por sua abrangência, rigor na seleção de periódicos e relevância em pesquisas interdisciplinares que envolvem tecnologias emergentes e inovação em saúde.

Cabe destacar que nem todas as palavras-chave apresentadas na seção 3.3 foram incluídas na versão final da *string* de busca. A utilização integral desses termos resultou em uma restrição excessiva do escopo, limitando significativamente a quantidade de estudos recuperados. Por esse motivo, foi necessário realizar um processo iterativo de refinamento da *string*, visando equilibrar a abrangência e a relevância dos resultados obtidos. Considerando que cada base de dados adota uma sintaxe própria para a formulação de consultas, foram elaboradas diferentes strings de busca específicas para cada biblioteca, mantendo-se as palavras-chave e as relações semânticas entre elas. As variações adotadas estão detalhadas na Tabela 3.1.

Tabela 3.1: Strings de busca por biblioteca.

SCOPUS	ABS ("heart failure"AND "prediction"AND ("edge computing"OR "wearable"OR "iot") AND ("machine learning"OR "artificial intelligence")) AND (LIMIT-TO (DOCTYPE , "ar"))
IEEE	((("Abstract":"heart failure") AND ("Abstract":"prediction") AND ("Abstract":"edge computing"OR "Abstract":"wearable"OR "Abstract":"iot") AND ("Abstract":"machine learning"OR "Abstract":"artificial intelligence")))
Web of Science	AB = ("heart failure"AND "prediction"AND ("edge computing"OR "wearable"OR "iot") AND ("machine learning"OR "artificial intelligence")))

3.5 Critérios de inclusão e exclusão

Os seguintes critérios de inclusão (CI) e exclusão (CE) foram definidos para este mapeamento sistemático:

CI1: O estudo apresenta uma abordagem de predição de insuficiência cardíaca associada a aprendizado de máquina ou inteligência artificial.

CI2: O estudo apresenta integração com dispositivos IoT e/ou sensores vestíveis para coleta e monitoramento de dados fisiológicos.

CI3: O estudo descreve a implementação de predição de insuficiência cardíaca em dispositivos móveis ou em arquiteturas de computação de borda (*edge computing*).

CE1: O estudo não está disponível para leitura.

CE2: O estudo não está escrito em inglês ou português.

CE3: O estudo não é primário.

CE4: O estudo não está relacionado à predição de insuficiência cardíaca ou não contém elementos de monitoramento de saúde por sensores, dispositivos IoT ou aprendizado de máquina.

3.6 Seleção e avaliação

Para o processo de seleção e avaliação, foi adotada uma abordagem em quatro estágios. No estágio 1, os 42 estudos encontrados a partir da aplicação da string de busca foram importados para a ferramenta *Rayyan* (OUZZANI et al., 2016), que detectou 22 duplicatas. Após a remoção desses artigos duplicados, obteve-se um conjunto de 31 estudos preparados para a fase de triagem. No estágio 2, foi verificada a disponibilidade dos textos completos e o idioma de publicação de cada estudo, aplicando-se os critérios de exclusão CE1 e CE2. Após essa filtragem, 7 estudos foram excluídos, resultando em 24 estudos. Em seguida, no estágio 3, realizou-se a leitura do título, resumo e palavras-chave de cada artigo, com base nos critérios CE3 e CE4. Como resultado, 13 estudos foram excluídos,

totalizando 11 artigos elegíveis para a próxima etapa. Finalmente, no estágio 4, procedeu-se à leitura completa dos estudos remanescentes, com a verificação do atendimento aos critérios de inclusão. Essa etapa garantiu a seleção final dos artigos mais relevantes e alinhados aos objetivos deste trabalho. Nessa etapa, 1 artigo foi removido por não estar relacionado ao tema do trabalho, resultando em 10 artigos para revisão. Além disso, o artigo motivador deste trabalho *Real-Time Heart Failure Prediction: An Approach for Ambient Assisted Living* (ALMEIDA et al., 2025) foi incluído como estudo obtido através de outras fontes, resultando em 11 estudos finais.

A Figura 3.1 ilustra o fluxograma PRISMA preenchido, detalhando as etapas percorridas.

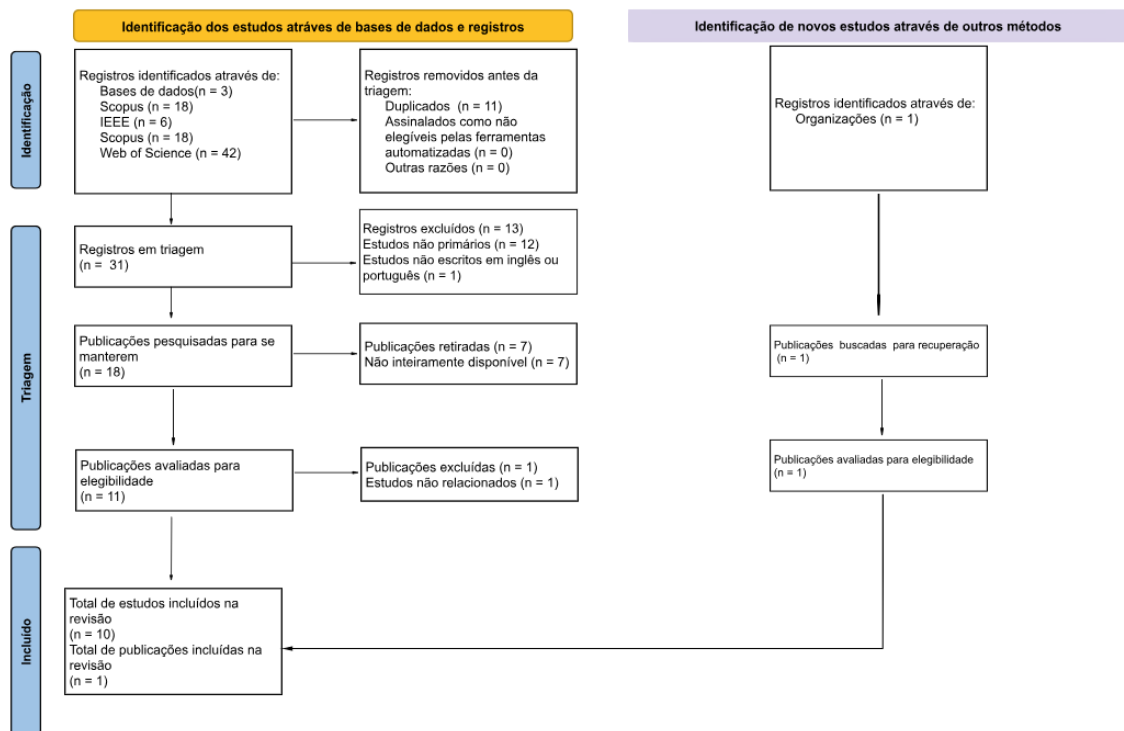


Figura 3.1: Fluxograma PRISMA 2020

3.7 Resultados

Os 11 artigos resultantes foram selecionados para responder as questões de pesquisa a seguir, e são apresentados na Tabela 3.2.

Tabela 3.2: Lista de estudos selecionados.

P1	A Comprehensive Review of Heart Disease Prediction using Cloud-Driven Machine Learning	(SURESH; DAYANA, 2024)
P2	An Intelligent Cardiovascular Diseases Prediction System Focused on Privacy	(KOLHAR; MISFER, 2023)
P3	Cloud-Based Machine Learning Platform to Predict Clinical Outcomes at Home for Patients With Cardiovascular Conditions Discharged From Hospital: Clinical Trial	(YANG et al., 2024)
P4	Continuous Wearable Monitoring Analytics Predict Heart Failure Hospitalization: The LINK-HF Multi-center Study	(STEHLIK et al., 2020)
P5	Designing a Smart Healthcare framework based on disease prediction using Optimizer based AI	(VYAS; KALSHETTY; PAREEK, 2023)
P6	Enhanced cardiovascular disease prediction model using random forest algorithm	(SUMWIZA et al., 2023)
P7	Healthcare diagnostics with an adaptive deep learning model integrated with the Internet of medical Things (IoMT) for predicting heart disease	(BASEER et al., 2024)
P8	IoT Based Smart Monitoring of Patients' with Acute Heart Failure	(UMER et al., 2022)
P9	Machine learning applied to wearable fitness tracker data and the risk of hospitalizations and cardiovascular events	(KUNDRICK et al., 2025)
P10	Optimized Heart Failure Prediction using Support Vector Machine Algorithms	(RANI, 2024)
P11	Real-Time Heart Failure Prediction: An Approach for Ambient Assisted Living	(ALMEIDA et al., 2025)

QP1: Quais são os principais algoritmos e técnicas de aprendizado de máquina utilizados na predição de insuficiência cardíaca em tempo real?

A predição de insuficiência cardíaca em tempo real tem sido realizada por meio de uma ampla gama de algoritmos de aprendizado de máquina. O algoritmo *Support Vector Machine* (SVM) foi amplamente destacado por sua eficácia em classificação binária de doenças cardíacas, com acurácia de até 87,5% (P10). O *Random Forest* (RF) também demonstrou desempenho notável, atingindo até 99% de acurácia ao integrar múltiplas técnicas de seleção de atributos (P6). Técnicas mais recentes incluem o uso combinado de TabNet (ARIK; PFISTER, 2019) e CatBoost (DOROGUSH; ERSHOV; GULIN, 2018), que foram utilizados com sucesso na predição precisa de doenças cardíacas em ambientes integrados com IoMT (P7). O uso de CNNs e RNNs, especialmente em contextos de computação em nuvem com sensores vestíveis, foi explorado em P3 e P4, assim como em P8, onde o CNN superou outros modelos com acurácia de 92,89%. Técnicas de *gradient boosting*, MLP e *federated learning* também foram utilizadas (P5, P7 e P9), indicando uma tendência crescente em modelos otimizados para dados heterogêneos e em tempo real.

QP2: Como a computação de borda (*edge computing*) tem sido utilizada para reduzir latência e melhorar a confiabilidade em sistemas de monitoramento de saúde?

A computação de borda tem sido utilizada para reduzir latência e aumentar a confiabilidade por meio da descentralização do processamento de dados. Em P2 e P4, o *edge computing* é incorporado para processar dados de sensores em dispositivos locais, evitando atrasos de comunicação com a nuvem. Em P8, sensores IoT enviam dados diretamente para a nuvem, mas com pré-processamento realizado localmente para aumentar a eficiência. Já em P7, o modelo prevê eventos cardíacos a partir de dados transmitidos continuamente por dispositivos vestíveis, com ênfase na resposta em tempo real. Embora alguns artigos usem computação em nuvem (P3 e P6), a tendência observada é integrar camadas de *fog* ou *edge computing* para análise preliminar e redução da sobrecarga dos servidores centrais.

QP3: Quais sensores vestíveis e dispositivos IoT são mais utilizados

para coleta de dados em sistemas de predição de insuficiência cardíaca?

Os sensores vestíveis e dispositivos IoT mais utilizados incluem *smartwatches*, *patches* multissensores, e dispositivos como Arduino. P3 e P4 destacam o uso de dispositivos vestíveis que coletam ECG, frequência respiratória, postura e outros sinais vitais. Em P6 e P10, sensores são conectados a plataformas móveis para adquirir sinais fisiológicos como pressão arterial, colesterol e frequência cardíaca. Já P9 analisa dados de *Fitbit* combinados com registros eletrônicos de saúde (EHR), provando a eficácia de *wearables* comerciais em detecção precoce de eventos cardiovasculares. A combinação de sensores ECG, pressão arterial e sensores de movimento também é comum em modelos propostos em P5 e P7, consolidando os dispositivos IoT como infraestrutura essencial.

QP4: Quais métricas e métodos de avaliação são usados para validar a acurácia e a responsividade desses modelos em ambientes de computação de borda?

As métricas mais comuns para validação incluem acurácia, sensibilidade, especificidade, AUROC (área sob a curva ROC) e F1-score. P4 utilizou medidas como tempo médio de alerta (mediana de 6,5 dias antes da hospitalização) e índice de mudança multivariada. Já P10 utilizou AUROC e acurácia (87,53%), enquanto P6 apresentou resultados com 99% de acurácia, testando diferentes algoritmos com validação cruzada. P9 também aplicou AUROC e F1-score, com *Random Forest* atingindo AUROC de 0.95 na previsão de hospitalizações. Além disso, métodos como matriz de confusão, precisão e curva de aprendizagem foram usados para comparar modelos (P3, P5 e P8). As avaliações demonstram a busca por equilíbrio entre desempenho preditivo e aplicabilidade clínica em tempo real.

QP5: Quais *frameworks*, plataformas e ferramentas são utilizados para implementar sistemas de predição em dispositivos móveis?

Diversas ferramentas são empregadas, incluindo plataformas em nuvem e *frameworks* como WEKA(HALL et al., 2009), Scikit-Learn(Scikit-learn, 2025b), e ambientes baseados em *Android-Arduino*. P2 e P5 apresentam sistemas que utilizam microcontroladores e sensores integrados a apps móveis para monitoramento remoto. P9 utilizou a plataforma *All of Us* com integração a APIs do *Fitbit* e modelos em *Python* para análise

de dados. P8 implementou um sistema completo de monitoramento inteligente com dispositivos IoT e processamento via *cloud computing*. Em P7, a combinação do *CatBoost* e *TabNet* reforça o uso de *frameworks* modernos de *deep learning* para dados tabulares em dispositivos de saúde conectados. Esses *frameworks* são selecionados conforme o tipo de dado (sinal fisiológico, imagem, histórico clínico) e os recursos computacionais disponíveis nos dispositivos móveis ou embarcados.

QP6: Quais são os principais desafios e limitações para implementar modelos de aprendizado de máquina embarcados em dispositivos móveis para detecção de insuficiência cardíaca?

Os principais desafios incluem limitações de hardware, heterogeneidade dos dados, segurança e privacidade, além da necessidade de modelos interpretáveis. Em P5, destaca-se a limitação de processamento e armazenamento dos dispositivos móveis. P2 enfatiza questões de privacidade, enquanto P9 reforça a dificuldade de generalização dos modelos para diferentes perfis clínicos. Já P10 aponta para a complexidade na seleção de atributos e necessidade de melhorar a interpretabilidade dos modelos SVM. Além disso, a latência e o volume de dados de sensores contínuos exigem algoritmos otimizados (P3, P6 e P8). A escalabilidade dos sistemas também é limitada pela necessidade de integração entre múltiplos dispositivos, algo que os artigos reconhecem como área de pesquisa contínua.

QP7: Como a integração de sensores, dispositivos móveis e computação de borda influencia a escalabilidade e a privacidade dos dados de saúde?

A integração de sensores, dispositivos móveis e computação de borda possibilita escalabilidade com resposta em tempo real e maior controle sobre a privacidade dos dados. P2 e P5 indicam que, ao manter o processamento próximo ao paciente, reduz-se o tráfego de dados sensíveis. P7 apresenta uma abordagem baseada em IoMT que permite análises locais e imediatas, promovendo intervenções personalizadas com menor risco de exposição de dados. Em P9, os dados do *Fitbit* são anonimizados e utilizados com EHRs, demonstrando que é possível manter privacidade sem comprometer o valor preditivo dos modelos. Apesar disso, poucos artigos tratam diretamente de *federated learning* ou criptografia ponta-a-ponta, sugerindo que essas são áreas em desenvolvimento (P2 e P10). A tendência é clara: sensores integrados a dispositivos móveis e conectados a bordas com-

putacionais permitem sistemas escaláveis, confiáveis e com menor dependência de centros de dados centralizados.

3.8 Análise dos resultados

Observando os 11 estudos deste Mapeamento Sistemático da Literatura (MSL), percebe-se que foram publicados entre os anos de 2022 e 2025, o que demonstra que a predição de insuficiência cardíaca por meio de *machine learning* (ML) embarcado é um tema altamente atual e em crescimento na comunidade científica.

Em relação aos algoritmos de aprendizado de máquina utilizados, observa-se uma predominância de técnicas clássicas como *Support Vector Machine* (SVM) (P1, P6, P10), *Random Forest* (RF) (P2, P4, P6), *Logistic Regression* e KNN (P6, P9) e o uso crescente de modelos de deep learning, especialmente Redes Neurais Convolucionais (CNN), LSTM e *TabNet* (P3, P5, P7, P8). Essa diversidade de abordagens revela uma busca ativa por modelos que conciliem precisão, velocidade de inferência e interpretabilidade clínica, algo fundamental para aplicação em sistemas embarcados e cenários de monitoramento em tempo real.

No que diz respeito à infraestrutura tecnológica e sensores utilizados, destacam-se os dispositivos vestíveis (*wearables*) e IoT, especialmente *smartwatches*, *patches* multi-sensores e plataformas como *Fitbit*, que aparecem em diversos estudos (P3, P4, P7, P9). Os sensores mais frequentes envolvem ECG, frequência cardíaca, pressão arterial, postura, temperatura e sinais metabólicos. Em termos de conectividade e arquitetura de sistema, observa-se a predominância de abordagens que mesclam IoT com *cloud computing* (P2, P3, P8), e mais recentemente, a incorporação de computação de borda (*edge/fog computing*) como estratégia para reduzir latência, aliviar a carga dos servidores e permitir resposta clínica mais rápida (P2, P4, P7).

As métricas de avaliação dos modelos são heterogêneas, mas predominam acurácia, AUROC, F1-score e sensibilidade (P4, P6, P9, P10). Embora todos os artigos apresentem avaliação quantitativa dos modelos, há ausência de padronização metodológica, o que dificulta comparações diretas entre os trabalhos e impede conclusões amplas sobre qual técnica é mais eficaz em ambientes embarcados. Apenas dois estudos (P4 e P9) avaliam

responsividade clínica direta, como tempo de alerta antes da hospitalização ou previsão de eventos cardiovasculares reais, o que sugere a necessidade de mais estudos focados em impacto clínico real.

Do ponto de vista da implementação prática, os estudos utilizam desde ferramentas acadêmicas como WEKA, até *frameworks* robustos como *Python* (*scikit-learn*, *TensorFlow*, *CatBoost*) e integração com dispositivos móveis por meio de plataformas como *Android-Arduino* (P2, P5, P9). Apesar disso, nenhum dos trabalhos relata implementação real embarcada em hardware comercial, o que aponta para uma lacuna entre o desenvolvimento em ambiente experimental e a aplicação no mundo real.

Por fim, questões como escalabilidade, privacidade e segurança de dados são mencionadas, mas pouco exploradas em profundidade, embora alguns estudos cite o uso de dados anonimizados (P9) e estratégias como *edge computing* ou criptografia (P2, P10).

Conclui-se, portanto, que o campo de *machine learning* embarcado para predição de insuficiência cardíaca está em rápido crescimento, com forte ênfase em técnicas precisas e sensores vestíveis. Contudo, há desafios críticos relacionados à validação clínica, privacidade, escalabilidade, acessibilidade e uso em dispositivos reais, que devem ser abordados em futuras pesquisas para que tais soluções possam ser efetivamente aplicadas em larga escala no monitoramento remoto da saúde cardíaca.

4 Materiais e métodos

Este capítulo apresenta o processo seguido para desenvolver e implementar o modelo de aprendizado de máquina proposto para predição de insuficiência cardíaca. O intuito é descrever as etapas envolvidas na construção da ferramenta, desde a escolha da base de dados até o desenvolvimento do modelo e o teste em sistemas embarcados.

A Figura 4.1 contém as etapas seguidas na elaboração do trabalho. Inicialmente, foi feita a escolha da base de dados e o seu download de uma fonte pública.

A escolha da base de dados para este estudo ocorreu de maneira conscienciosa. Para isso, foi conduzida uma entrevista com especialista (médica cardiologista), a partir da qual foram identificadas as variáveis e faixas de valores mais frequentemente empregadas na prática clínica nacional para a avaliação de pacientes com insuficiência cardíaca. Entre os atributos considerados relevantes, destacam-se a pressão arterial, segmentada em pressão sistólica e pressão diastólica; os níveis de colesterol, contemplando colesterol total, HDL e LDL; a frequência cardíaca; características relacionadas à dor torácica, incluindo intensidade e possível irradiação para regiões como mandíbula, braços e outras partes do corpo; o nível de estresse; a qualidade do sono; a presença de dores em repouso e durante esforço físico; a variabilidade da frequência cardíaca; medidas corporais, como peso e altura, utilizadas para o cálculo do índice de massa corporal (IMC); o uso de substâncias ilícitas, como anfetaminas; e a existência de histórico familiar de acidente vascular cerebral (AVC).

Além disso, foi possível constatar que variáveis associadas a resultados de exames de eletrocardiograma, presentes em algumas bases de dados, não são obtíveis de forma remota. Assim, deu-se preferência para bases de dados que apresentassem os outros atributos mencionados, pois possuem relevância para serem adotados no processo de alerta em tempo real.

Após a seleção do *dataset*, os dados passaram por uma etapa de pré-processamento, na qual foi feito o tratamento de valores ausentes, a engenharia das *features*, a codificação

e normalização dos dados, etc. Em seguida, os dados tratados foram utilizados para o treinamento e teste dos modelos. Após os testes, o modelo com melhor desempenho foi escolhido e exportado para teste em dispositivos móveis, onde foi possível avaliar o desempenho do modelo em cenários simulados de uso. Nas seções a seguir, cada uma dessas etapas é detalhada, explicando como foram desenvolvidas e implementadas.

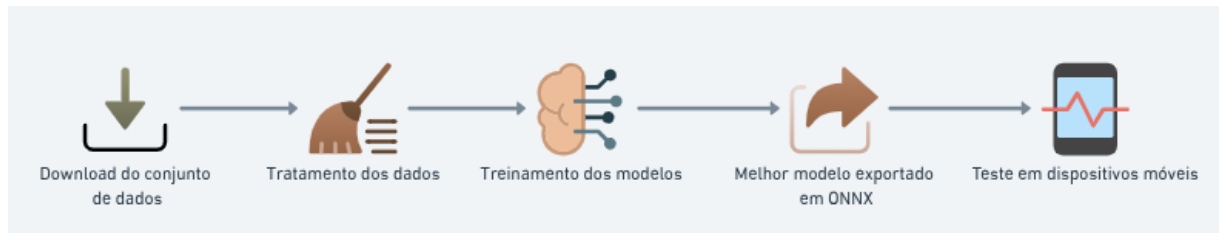


Figura 4.1: Fluxograma das etapas do trabalho

Na etapa de pré-processamento, os dados passaram por uma análise detalhada com o objetivo de identificar e corrigir inconsistências como valores ausentes e a presença de *outliers*. Após o tratamento adequado do conjunto de dados, diferentes modelos de aprendizado de máquina foram treinados e avaliados, visando a seleção daquele que apresentou o melhor desempenho. Por fim, o modelo escolhido foi submetido a testes em um ambiente de dispositivos móveis, com o objetivo de avaliar o desempenho da ferramenta em ambientes de *hardware* limitado. Nesse ambiente, a inferência é realizada localmente, a partir de dados vitais simulados, permitindo a emissão de alertas sempre que a predição indicar um elevado risco de insuficiência cardíaca.

4.1 Descrição dos dados

O objetivo desta etapa é detalhar o conjunto de dados utilizado para o treinamento dos modelos preditivos.

Inicialmente, foi visada a utilização de uma base de dados de origem brasileira que contemplasse as variáveis relevantes para este estudo. No entanto, diante da indisponibilidade de conjuntos de dados nacionais com essas características, a busca foi estendida para bases internacionais compatíveis. Nesse contexto, foi escolhida a base de

dados *CAIR-CVD-2025: An Extensive Cardiovascular Disease Risk Assessment Dataset from Bangladesh* (NIROB et al., 2025) , disponibilizada na plataforma *Mendeley Data*.

A base de dados é composta por 1.529 registros de pacientes, que foram coletados no Hospital da Faculdade de Medicina de Jamalpur, localizado na cidade de Jamalpur, em Bangladesh, entre 20 de janeiro de 2024 e 1 de janeiro de 2025. A escolha dessa base justifica-se pela presença de variáveis alinhadas às recomendações obtidas junto à profissional de saúde consultada, incluindo medidas de pressão arterial segmentadas em valores sistólicos e diastólicos, níveis de colesterol (total, HDL e LDL), peso, altura e informações relacionadas ao estilo de vida do paciente, como hábito de fumar, nível de atividade física e histórico familiar de doenças cardiovasculares. Além disso, o dataset apresenta uma variável categórica que representa o nível de risco de doença cardiovascular, elemento fundamental para o treinamento e avaliação dos modelos classificatórios.

A base de dados foi disponibilizada gratuitamente pelo *Mendeley Data* em formato *.csv*. Após o download, o arquivo foi submetido à etapa de pré-processamento, na qual foram realizadas as análises e transformações necessárias para garantir a qualidade dos dados antes da aplicação do treinamento nos modelos.

Abaixo, apresentamos as principais variáveis utilizadas e sua respectiva descrição:

- **Idade (*Age*)**: Representa a idade do indivíduo em anos.
- **Sexo (*Sex*)**: Variável categórica que representa o sexo biológico do indivíduo (M – Masculino, F – Feminino).
- **Peso (*Weight*)**: Representa o peso corporal do indivíduo em quilogramas.
- **Altura (*Height*)**: Representa a altura do indivíduo em metros.
- **IMC (*BMI*)**: Representa o índice de massa corporal, calculado a partir do peso e da altura, medido em kg/m^2 .
- **Circunferência abdominal (*Abdominal Circumference*)**: Representa a medida da circunferência da cintura em centímetros.
- **Proporção de cintura para altura (*Waist-to-Height Ratio*)**: Representa a razão entre a circunferência da cintura e a altura.

- **Pressão arterial (*Blood Pressure*)**: Representa a pressão arterial do indivíduo em mmHg.
- **Pressão sistólica (*Systolic BP*)**: Representa a pressão máxima exercida nas artérias durante a sístole, medida em mmHg.
- **Pressão diastólica (*Diastolic BP*)**: Representa a pressão mínima exercida nas artérias durante a diástole em mmHg.
- **Categoria de pressão arterial (*Blood Pressure Category*)**: Variável categórica que representa a classificação clínica da pressão arterial (*Normal, Elevated, Hypertension Stage 1, Hypertension Stage 2*).
- **Colesterol total (*Total Cholesterol*)**: Representa a concentração total de colesterol no sangue, medida em mg/dL.
- **HDL (*HDL*)**: Representa o nível de colesterol de alta densidade (HDL) no sangue, medido em mg/dL.
- **LDL estimado (*Estimated LDL*)**: Representa o nível estimado de colesterol de baixa densidade (LDL), medido em mg/dL.
- **Açúcar no sangue em jejum (*Fasting Blood Sugar*)**: Representa o nível de glicose no sangue em jejum, medido em mg/dL.
- **Hábito de fumar (*Smoking Status*)**: Variável categórica que representa se o indivíduo é fumante (Y – Sim, N – Não).
- **Diabético (*Diabetes Status*)**: Variável categórica que representa se o indivíduo possui diagnóstico de diabetes (Y – Sim, N – Não).
- **Nível de atividade física (*Physical Activity Level*)**: Variável categórica que representa o nível de prática de atividades físicas do indivíduo (*Low, Moderate, High*).
- **Histórico de doença cardiovascular na família (*Family History of CVD*)**: Variável categórica que representa a existência de histórico familiar de doenças cardiovasculares (Y – Sim, N – Não).

- **Nível de risco de doença cardiovascular (*CVD Risk Level*):** Variável categórica que representa a classe de risco cardiovascular atribuída ao indivíduo (*Low*, *Intermediary*, *High*).

A Figura 4.2 mostra a matriz de correlação das variáveis numéricas que compõem a base de dados. Nela, é possível observar como os atributos se relacionam entre si de forma bivariada. As cores indicam tanto o sentido quanto a intensidade dessas relações: tons em vermelho representam correlações positivas, enquanto tons em azul indicam correlações negativas. Quanto mais intensa a cor, mais forte é a correlação. Os valores apresentados em cada célula correspondem ao coeficiente de correlação de Pearson, que varia de -1 a $+1$.

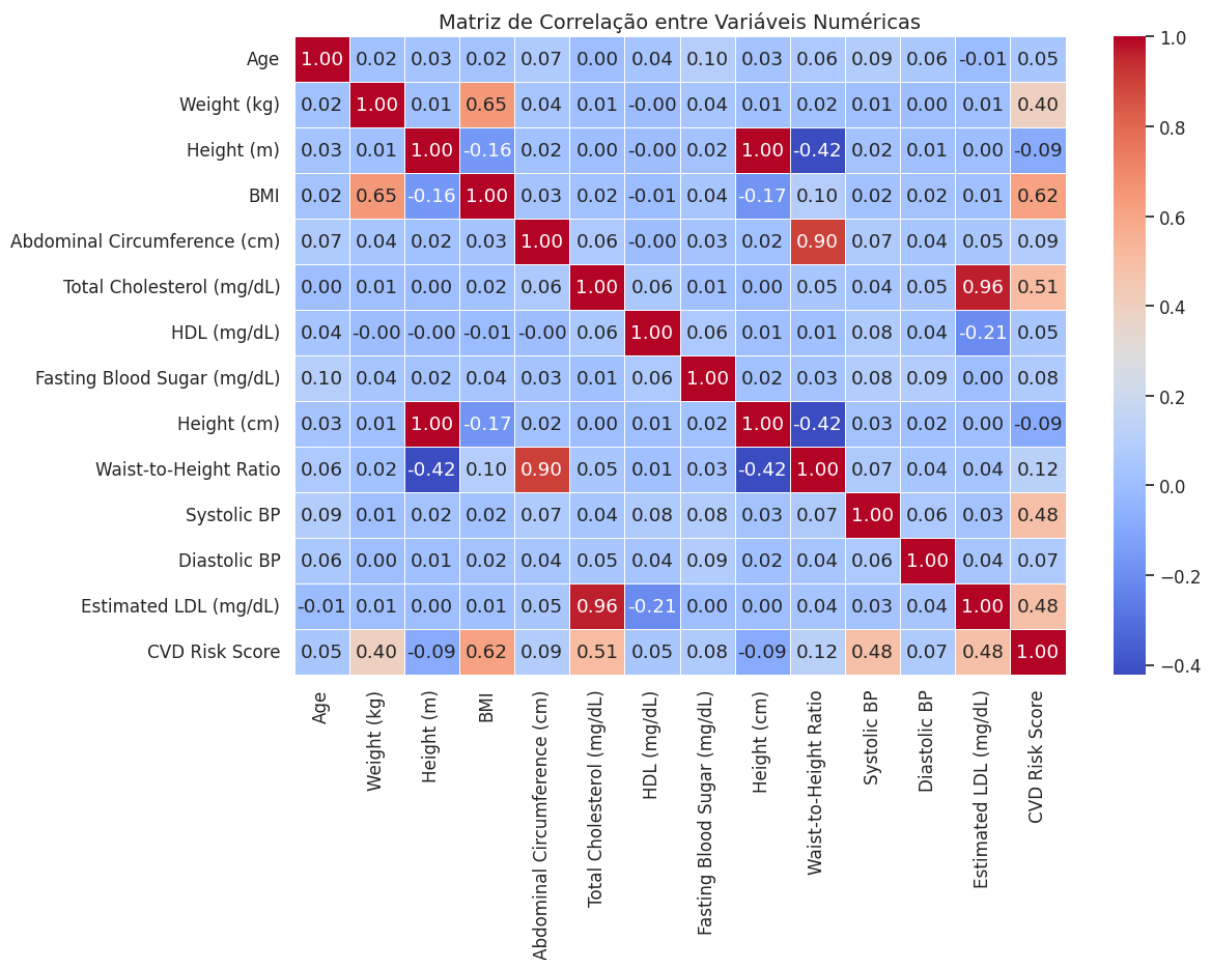


Figura 4.2: Matriz de correlação entre variáveis numéricas

É possível observar que proporção cintura-altura têm correlação positiva muito

alta com circunferência abdominal (quanto maior a circunferência abdominal, maior a razão) e correlação negativa com razão altura (quanto maior a altura, menor a razão).

O colesterol total e o LDL estimado têm correlação fortíssima, pois o LDL deriva do total. Ambos atributos correlacionam positivamente com *CVD Risk Score*, indicando possível causa de insuficiência cardíaca. Peso e IMC tem correlação considerável, e ambos têm correlação positiva *CVD Risk Score*. Excesso de peso contribui para hipertensão, resistência à insulina e sobrecarga do ventrículo.

4.2 Modelagem Preditiva para Insuficiência Cardíaca

O objetivo desta seção é apresentar a abordagem metodológica utilizada para a construção de modelos preditivos capazes de classificar o risco de insuficiência cardíaca do paciente com base em seus dados.

A modelagem preditiva tem como objetivo identificar padrões relevantes nos dados clínicos e fisiológicos dos pacientes, possibilitando a compreensão dos fatores associados ao risco de insuficiência cardíaca. A partir dessa análise, torna-se possível estimar a probabilidade de ocorrência de eventos adversos, oferecendo suporte para a tomada de decisão tanto por parte dos profissionais de saúde quanto dos próprios pacientes. Dessa forma, as previsões geradas pelo modelo contribuem para um acompanhamento mais eficiente e para a realização de intervenções preventivas de maneira mais oportuna.

4.2.1 Pré-processamento dos dados

Tratamento de valores ausentes

Durante a etapa de análise do conjunto de dados, identificou-se a presença de valores ausentes em diferentes atributos. Entretanto, em alguns casos, esses atributos apresentavam relações diretas com outras variáveis preenchidas, o que possibilitou a realização de inferências para o preenchimento dos valores faltantes. As inferências realizadas são descritas a seguir.

No que se refere às variáveis *Systolic BP* e *Diastolic BP*, foram encontrados registros em que uma ou ambas se encontravam ausentes, enquanto a variável *Blood*

Pressure (mmHg) estava preenchida. Como essa última representa a pressão arterial no formato “Sistólica/Diastólica” (por exemplo, 125/79), foi possível extrair diretamente os valores correspondentes e completar as colunas ausentes.

Em relação à altura do paciente, a base de dados disponibiliza essa informação por meio das variáveis **Height (m)** e **Height (cm)**. Dessa forma, sempre que uma dessas colunas apresentava valor ausente e a outra estava preenchida, a inferência foi realizada por meio de conversão direta entre as unidades, utilizando o fator de multiplicação ou divisão por 100.

Para os atributos **BMI**, **Weight (kg)** e **Height**, explorou-se a relação matemática conhecida entre essas variáveis, expressa pela equação $IMC = \text{peso} / \text{altura}^2$. Assim, nos registros em que apenas uma dessas variáveis estava ausente e as outras duas estavam disponíveis, foi possível calcular o valor faltante por meio de substituição algébrica.

Procedimento semelhante foi adotado para as variáveis **Abdominal Circumference (cm)**, **Height** e **Waist-to-Height Ratio**. Considerando a relação definida pela razão entre a circunferência abdominal e a altura, foi possível inferir a variável ausente sempre que dois dos três atributos estivessem presentes no registro.

Após a etapa de inferência, atributos redundantes ou altamente correlacionados foram removidos com o objetivo de simplificar o treinamento dos modelos sem perda significativa de informação. Nesse sentido, uma das features de altura (**Height (cm)**) foi excluída, visto que, após o preenchimento dos valores ausentes, todos os registros passaram a conter as informações de altura tanto em metros quanto em centímetros, tornando desnecessária a presença de ambas.

A variável **Waist-to-Height Ratio** foi removida, pois apresentou correlação forte com a circunferência abdominal, atributo que foi mantido no conjunto de dados por ser amplamente utilizado em avaliações clínicas. De forma semelhante, o índice de massa corporal (IMC) foi excluído devido à sua forte correlação com as variáveis peso e altura, que permaneceram no conjunto.

Além disso, a variável **Blood Pressure (mmHg)** foi retirada, visto que os valores de pressão arterial já estavam representados nas colunas de pressão sistólica (**Systolic BP**) e diastólica (**Diastolic BP**). Por fim, após a conclusão das etapas de inferência e remoção de

atributos redundantes, todos os registros que ainda apresentavam valores ausentes foram descartados. Como resultado desse processo, do total inicial de 1.529 registros, foram preservados 1.193 registros para as etapas subsequentes de modelagem.

A Tabela 4.1 apresenta todos os atributos do conjunto que foram mantidos.

Tabela 4.1: Análise das variáveis utilizadas na modelagem

Variável	Situação
Idade (<i>Age</i>)	Mantida
Sexo (<i>Sex</i>)	Mantida
Peso (<i>Weight</i>)	Mantida
Altura (m) (<i>Height (m)</i>)	Mantida
Altura (cm) (<i>Height (cm)</i>)	Removida
IMC (<i>BMI</i>)	Removida
Circunferência abdominal (<i>Abdominal Circumference</i>)	Mantida
Proporção-cintura-altura (<i>Waist-to-Height Ratio</i>)	Removida
Pressão arterial (<i>Blood Pressure</i>)	Removida
Pressão sistólica (<i>Systolic BP</i>)	Mantida
Pressão diastólica (<i>Diastolic BP</i>)	Mantida
Categoria de pressão arterial (<i>Blood Pressure Category</i>)	Mantida
Colesterol total (<i>Total Cholesterol</i>)	Mantida
HDL (<i>HDL</i>)	Mantida
LDL estimado (<i>Estimated LDL</i>)	Mantida
Açúcar no sangue em jejum (<i>Fasting Blood Sugar</i>)	Mantida
Hábito de fumar (<i>Smoking Status</i>)	Mantida
Diabético (<i>Diabetes Status</i>)	Mantida
Nível de atividade física (<i>Physical Activity Level</i>)	Mantida
Histórico de doença cardiovascular na família (<i>Family History of CVD</i>)	Mantida
Nível de risco de doença cardiovascular (<i>CVD Risk Level</i>)	Mantida

Transformação do problema multiclasse em classificação binária

Inicialmente, o problema foi abordado como uma tarefa de classificação multiclasse, considerando os três níveis de risco cardiovascular disponíveis na base de dados: *Low*, *Intermediary* e *High*. A Figura 4.3 exibe a distribuição das classes dentre os 1193 registros, sendo 178 pertencentes à classe *Low*, 441 à classe *Intermediary* e 574 à classe *High*.

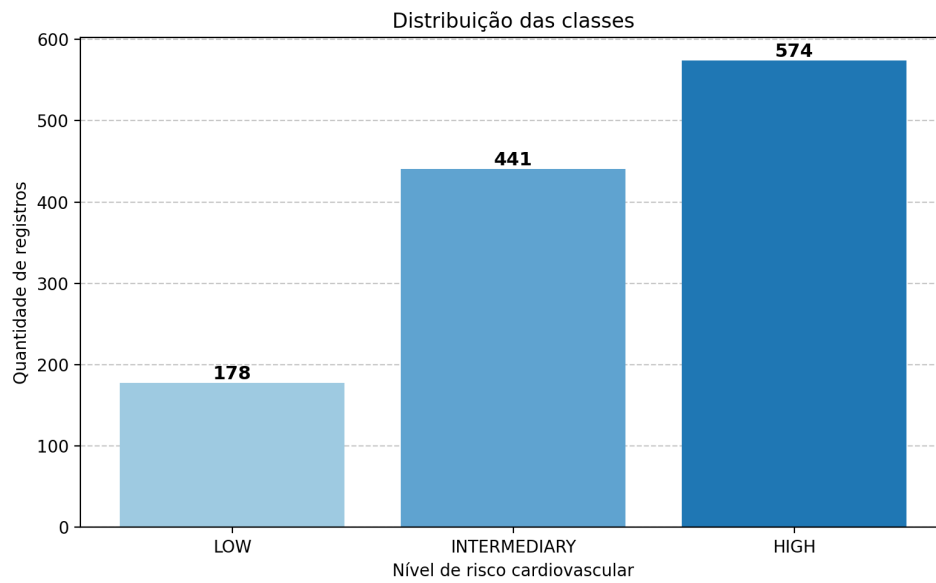


Figura 4.3: Distribuição original das classes

Durante os experimentos preliminares, foram realizados testes de predição considerando as três classes originais. Entretanto, ao ponderar sobre o objetivo central da aplicação — identificar indivíduos com alto risco de insuficiência cardíaca — optou-se por reformular o problema como uma tarefa de classificação binária. Para isso, as classes minoritárias (*Low* e *Intermediary*) foram agrupadas em uma única categoria, denominada *Non-High*, enquanto a classe *High* foi mantida de forma isolada. Essa estratégia permitiu direcionar o processo de aprendizado para a distinção entre pacientes de alto risco, representados pela classe *High*, e aqueles que não se enquadram nessa condição, correspondentes à classe *Non-High*, alinhando o modelo à aplicação prática do sistema, cujo foco é a identificação precoce de casos críticos.

A Figura 4.4 apresenta a distribuição das classes após a junção das classes minoritárias, evidenciando que 619 registros passaram a compor a classe *Non-High*, enquanto 574 registros foram mantidos na classe *High*.

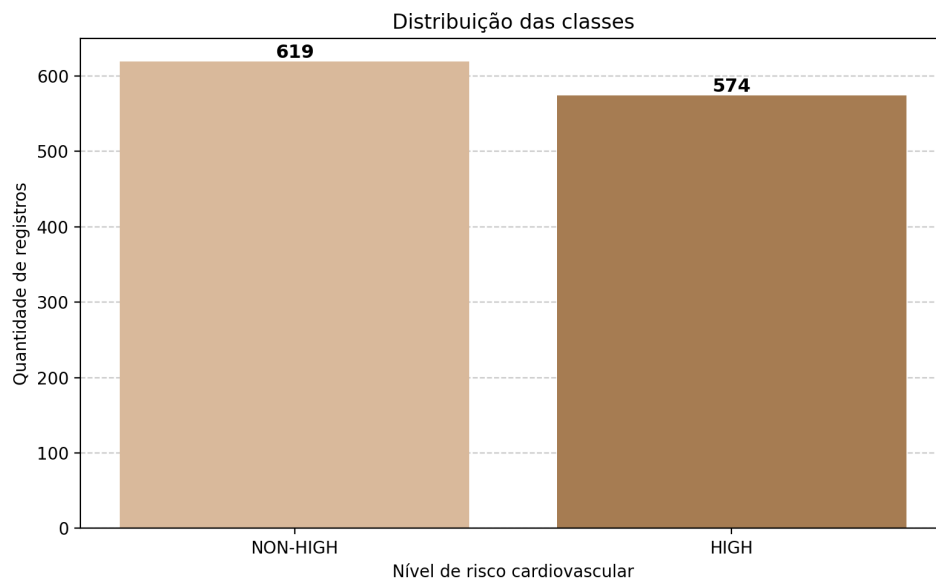


Figura 4.4: Distribuição das classes após a junção

Codificação das variáveis categóricas

As variáveis categóricas presentes no conjunto de dados — como *Sex*, *Smoking Status*, *Diabetes Status*, *Physical Activity Level*, *Family History of CVD* e *Blood Pressure Category* — foram tratadas por meio de codificação *One-hot Encoding* (Scikit-learn, 2025a). Essa técnica transforma cada categoria distinta em uma nova variável binária, assumindo valor 1 quando o registro pertence àquela categoria e 0 caso contrário.

Por exemplo, a variável *Physical Activity Level*, que pode assumir os valores *Low*, *Moderate* e *High*, foi convertida em três novas variáveis binárias: *PhysicalActivity_Low*, *PhysicalActivity_Moderate* e *PhysicalActivity_High*. Dessa forma, um indivíduo classificado como *Moderate* apresentará valor 1 apenas na variável correspondente (*PhysicalActivity_Moderate*), enquanto as demais assumirão valor 0. Esse procedimento elimina qualquer interpretação ordinal implícita entre as categorias, que poderia induzir o modelo a inferir relações numéricas inexistentes.

Normalização das variáveis numéricas

As variáveis numéricas do conjunto de dados, como idade, peso, circunferência abdominal, níveis de colesterol, glicemia em jejum, altura e valores de pressão arterial, foram normalizadas utilizando a técnica de padronização *Standard Scaling* (Scikit-learn, 2025c). Esse

método transforma os dados de forma que cada variável apresente média zero e desvio padrão unitário.

A normalização é especialmente importante para algoritmos sensíveis à escala dos dados, como *Logistic Regression* e *Support Vector Machine*, além de contribuir para uma convergência mais estável durante o treinamento dos modelos.

Divisão dos dados em conjuntos de treino e teste

A base de dados foi dividida em conjuntos de treino e teste, utilizando a proporção de 80% dos dados para treinamento e 20% para teste. Os dados foram divididos de maneira randômica e estratificada, isto é, preservando a proporção entre as classes *HIGH* e *NON-HIGH* em ambos os conjuntos. Essa estratégia garante que o conjunto de teste seja representativo da distribuição original das classes, para garantir uma avaliação confiável do desempenho dos modelos.

Balanceamento das classes

Como a classe *HIGH* representava uma parcela menor dos registros, a base de dados apresentava leve desbalanceamento entre as classes. Para mitigar esse problema, foi aplicada a técnica *SMOTE* exclusivamente sobre o conjunto de treinamento.

4.2.2 Treinamento dos modelos

O treinamento dos modelos foi realizado por meio de *pipelines* que integram as etapas de pré-processamento, balanceamento das classes e aprendizado propriamente dito, garantindo que todas as transformações fossem aplicadas de forma consistente durante o processo de validação. Foram avaliados cinco algoritmos de aprendizado de máquina: *Logistic Regression*, *Random Forest*, *Gradient Boosting*, *Support Vector Machine (SVM)* e *XGBoost*.

A otimização dos modelos foi realizada utilizando o método de busca em grade (*Grid Search*), por meio do *GridSearchCV*. Para a avaliação dos diferentes conjuntos de hiperparâmetros, foi empregada validação cruzada do tipo *k-fold* estratificada, com $k = 5$ partições, de forma a preservar a proporção entre as classes em cada subconjunto

de treinamento e validação. Essa estratégia permitiu uma estimativa mais confiável do desempenho dos modelos. A seleção dos hiperparâmetros considerou valores tipicamente adotados em implementações de referência, permitindo avaliar diferentes níveis de regularização e complexidade dos modelos. Os hiperparâmetros testados para cada modelo são descritos na Tabela 4.2.

Tabela 4.2: Hiperparâmetros avaliados durante o treinamento dos modelos

Modelo	Hiperparâmetros avaliados
Logistic Regression	$C \in \{0.01, 0.1, 1, 10\}$
Random Forest	Número de árvores $\in \{100, 200\}$; Profundidade máxima $\in \{None, 10, 20\}$
Gradient Boosting	Número de estimadores $\in \{100, 200\}$; Taxa de aprendizado $\in \{0.05, 0.1\}$; Profundidade $\in \{3, 5\}$
SVM	$C \in \{0.1, 1, 10\}$; Gamma $\in \{scale, auto\}$
XGBoost	Número de estimadores $\in \{100, 200\}$; Taxa de aprendizado $\in \{0.05, 0.1\}$; Profundidade $\in \{3, 5\}$; $Subsample \in \{0.8, 1.0\}$

No modelo de **Logistic Regression**, o parâmetro C representa o inverso da força de regularização. Valores menores de C impõem uma regularização mais forte, resultando em modelos mais simples e menos propensos a *overfitting*, enquanto valores maiores reduzem o efeito da regularização, permitindo modelos mais complexos e potencialmente mais sensíveis aos dados de treinamento.

Para o **Random Forest**, foram considerados o número de árvores ($n_estimators$) e a profundidade máxima das árvores (max_depth). O parâmetro $n_estimators$ define quantas árvores de decisão compõem a floresta; um número maior tende a aumentar a estabilidade das previsões, embora implique maior custo computacional. Já o max_depth limita o crescimento das árvores individuais: valores menores produzem árvores mais rasas, reduzindo o risco de *overfitting*, enquanto a ausência de limite permite árvores mais profundas.

No **Gradient Boosting**, o parâmetro $n_estimators$ indica o número de estágios

de *boosting*, ou seja, quantas árvores são adicionadas sequencialmente ao modelo. O *learning_rate* controla a contribuição de cada nova árvore para o modelo final; valores menores tornam o aprendizado mais gradual e conservador, geralmente exigindo um maior número de árvores para alcançar bom desempenho. O parâmetro *max_depth* define a profundidade das árvores individuais, sendo comum a utilização de árvores rasas para evitar o *overfitting* em métodos de *boosting*.

Para a **Support Vector Machine (SVM)**, foram avaliados os parâmetros *C* e *gamma*. O parâmetro *C* controla o compromisso entre maximizar a margem de separação entre as classes e minimizar os erros de classificação no conjunto de treinamento. Valores maiores de *C* tornam o modelo menos tolerante a erros, enquanto valores menores permitem uma margem mais suave. O parâmetro *gamma*, define o alcance de influência de cada amostra de treinamento: valores maiores fazem com que a influência seja mais local, enquanto valores menores resultam em uma influência mais global. As opções *scale* e *auto* ajustam esse valor automaticamente com base no número de atributos e na variância dos dados.

Por fim, no **XGBoost**, além dos parâmetros *n_estimators*, *learning_rate* e *max_depth*, também foi avaliado o parâmetro *subsample*. Esse parâmetro define a fração de amostras utilizada para o treinamento de cada árvore. Valores menores que 1.0 introduzem aleatoriedade no processo de treinamento, contribuindo para a redução do *overfitting* e aumentando a capacidade de generalização do modelo.

Ao final do processo de busca e validação, o modelo que apresentou o melhor resultado das métricas acurácia, precisão, *recall*, *F1-Score* e curva ROC-AUC foi selecionado como modelo final, sendo posteriormente utilizado na etapa de implantação em ambiente simulado de dispositivo móvel.

4.2.3 Implementação em Dispositivos Móveis

O modelo com melhor desempenho foi exportado para o formato *ONNX* e integrado a um aplicativo móvel por meio do *ONNX Runtime*. Essa abordagem possibilitou a execução do modelo diretamente nos dispositivos, viabilizando a predição local sem a necessidade de comunicação com servidores externos.

O aplicativo móvel desenvolvido permite que o usuário insira dados compatíveis com o conjunto de atributos do *dataset*. A partir dessas informações, o modelo embarcado realiza a predição em tempo real. Quando identificado um risco elevado de insuficiência cardíaca, o sistema exibe um alerta automaticamente.

Para a avaliação da aplicação em dispositivos móveis, foram realizados testes em dois cenários distintos. No caso do *smartphone*, utilizou-se um dispositivo virtual com 2 GB de RAM, 4 núcleos de CPU e sistema operacional Android 16. Já para os testes em *smartwatch*, foi empregado um dispositivo virtual com sistema operacional Wear OS, 512 MB de RAM, tela AMOLED de 1,4 polegadas e Android 16.

Em ambos os dispositivos, o aplicativo foi configurado para receber os dados do paciente, processá-los localmente por meio do modelo embarcado e gerar predições em tempo real. Para fins de experimentação e validação da aplicação, os valores utilizados nos testes foram previamente definidos e inseridos manualmente no aplicativo, de modo a simular as características presentes no *dataset*. Esses dados contemplaram dois cenários distintos: um paciente classificado com baixo risco de insuficiência cardíaca e outro com alto risco. Em um cenário de uso real, tais informações seriam obtidas automaticamente a partir de sensores integrados aos dispositivos móveis, como sensores de frequência cardíaca.

No que se refere a avaliação da latência, foram conduzidos testes comparativos entre a execução local do modelo e a execução remota em servidores. Para essa análise, foram utilizadas 15 amostras extraídas do conjunto de testes, representando diferentes perfis de pacientes. Esses testes tiveram como finalidade evidenciar os desafios associados à latência e à dependência de conectividade em arquiteturas centralizadas, em contraste com a inferência realizada diretamente nos dispositivos móveis.

No cenário de execução remota, o modelo no formato *ONNX* foi previamente carregado em máquinas virtuais na nuvem, em instâncias provisionadas na plataforma *Amazon Web Services* (AWS). Nessas instâncias, foi disponibilizado um serviço responsável por receber requisições de inferência. O tempo de resposta medido correspondeu ao intervalo entre o envio da requisição e o recebimento da predição retornada pelo servidor, englobando tanto o tempo de processamento do modelo quanto os atrasos internet.

A avaliação da latência foi conduzida considerando as seguintes configurações de execução:

- **Dispositivo Móvel (Execução Local):** *Smartphone* com 2 GB de RAM, 4 núcleos de CPU e sistema operacional *Android*, utilizado para a realização das inferências localmente.
- **Ambiente Remoto (SA-EAST-1):** Servidor com 1 GB de RAM, 2 vCPUs, sistema operacional Linux, localizado em São Paulo – Brasil.
- **Ambiente Remoto (US-EAST-2):** Servidor com 1 GB de RAM, 2 vCPUs, sistema operacional Linux, localizado em Ohio – Estados Unidos.
- **Ambiente Remoto (US-WEST-2):** Servidor com 1 GB de RAM, 2 vCPUs, sistema operacional Linux, localizado em Oregon – Estados Unidos.

5 Resultados

5.1 Resultados do treinamento dos modelos

Os cinco modelos foram treinados com 80% dos dados e sucessivamente testados com os 20% dos dados restantes. A Figura 5.1 apresenta as matrizes de confusão de cada modelo para o conjunto de teste. No contexto desse trabalho, as matrizes expressam a eficácia dos modelos em identificar corretamente casos de alto risco de insuficiência cardíaca (verdadeiros positivos) e em classificar corretamente pacientes com baixo risco (verdadeiros negativos). Os modelos *Random Forest* e *XGBoost* tiveram os melhores desempenhos, manifestando valores baixos de falsos negativos, o que é essencial em um sistema de monitoramento de saúde para evitar que casos graves de insuficiência cardíaca passem despercebidos.

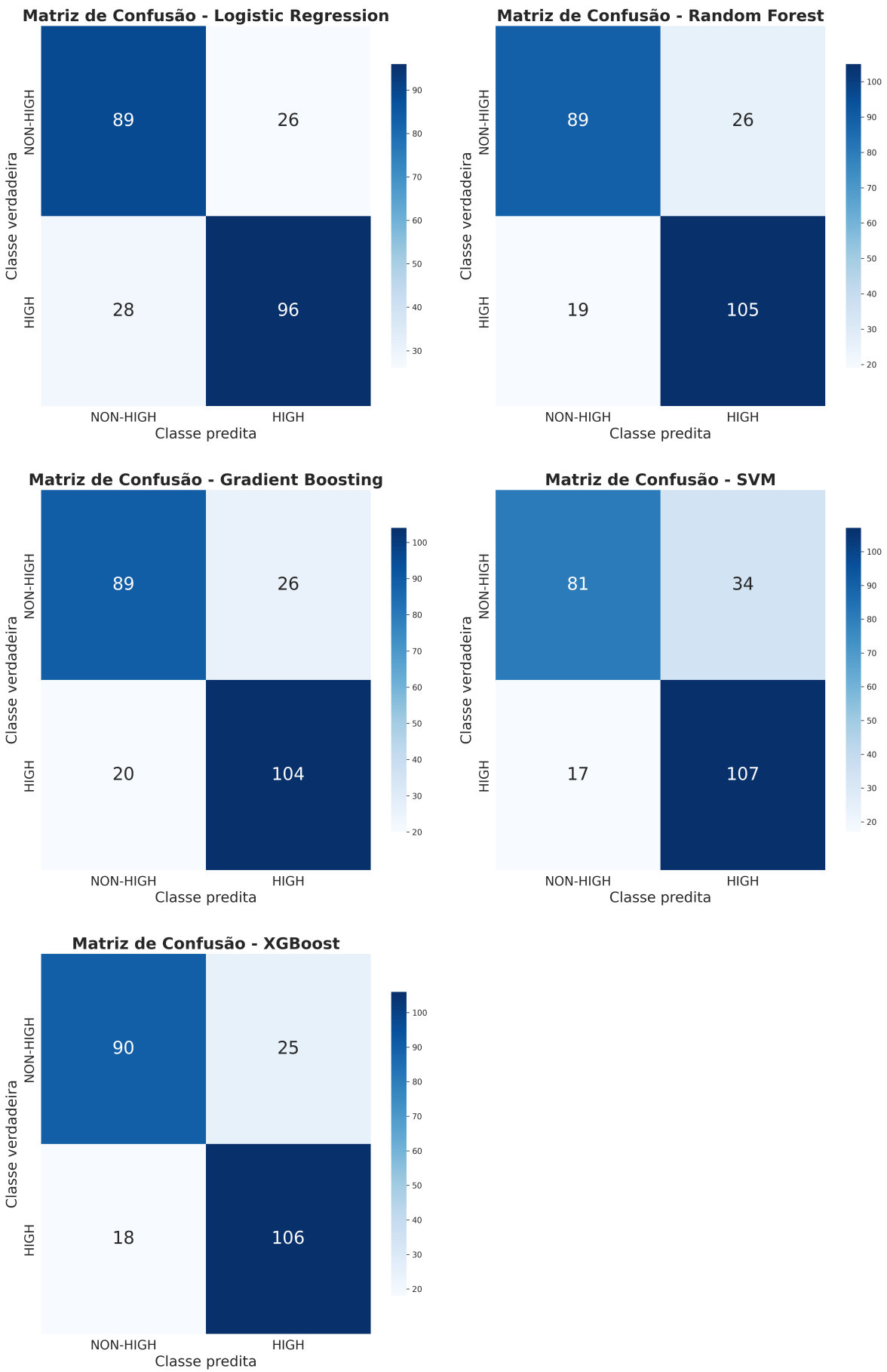


Figura 5.1: Matrizes de Confusão dos Modelos Classificatórios

A Tabela 5.1 apresenta as métricas de avaliação, onde é possível observar que o *XGBoost* foi o modelo que teve resultados ligeiramente melhores no geral.

Tabela 5.1: Desempenho dos modelos de classificação

Modelo	Acurácia (%)	Precisão (%)	Recall (%)	F1-Score (%)
Logistic Regression	0.77	0.79	0.77	0.78
Random Forest	0.81	0.80	0.85	0.82
Gradient Boosting	0.81	0.80	0.84	0.82
SVM	0.79	0.76	0.86	0.81
XGBoost	0.82	0.81	0.85	0.83

A Tabela 5.2 apresenta os hiperparâmetros de melhor desempenho no treinamento para cada modelo.

Tabela 5.2: Melhores hiperparâmetros selecionados para cada modelo

Modelo	C	max_depth	n_estimators	learning_rate	gamma	subsample
Logistic Regression	10	–	–	–	–	–
Random Forest	–	10	200	–	–	–
Gradient Boosting	–	3	200	0.1	–	–
SVM	1	–	–	–	scale	–
XGBoost	–	5	100	0.05	–	0.8

As curvas ROC dos modelos considerados aqui são apresentadas na Figura 5.2. Em geral, todos os modelos apresentaram curvas ROC com inclinação significativa em direção ao canto superior esquerdo, indicando alta sensibilidade e especificidade. Isso explicita a capacidade do modelo na detecção de insuficiência cardíaca, oferecendo segurança para alertas em situações de risco.

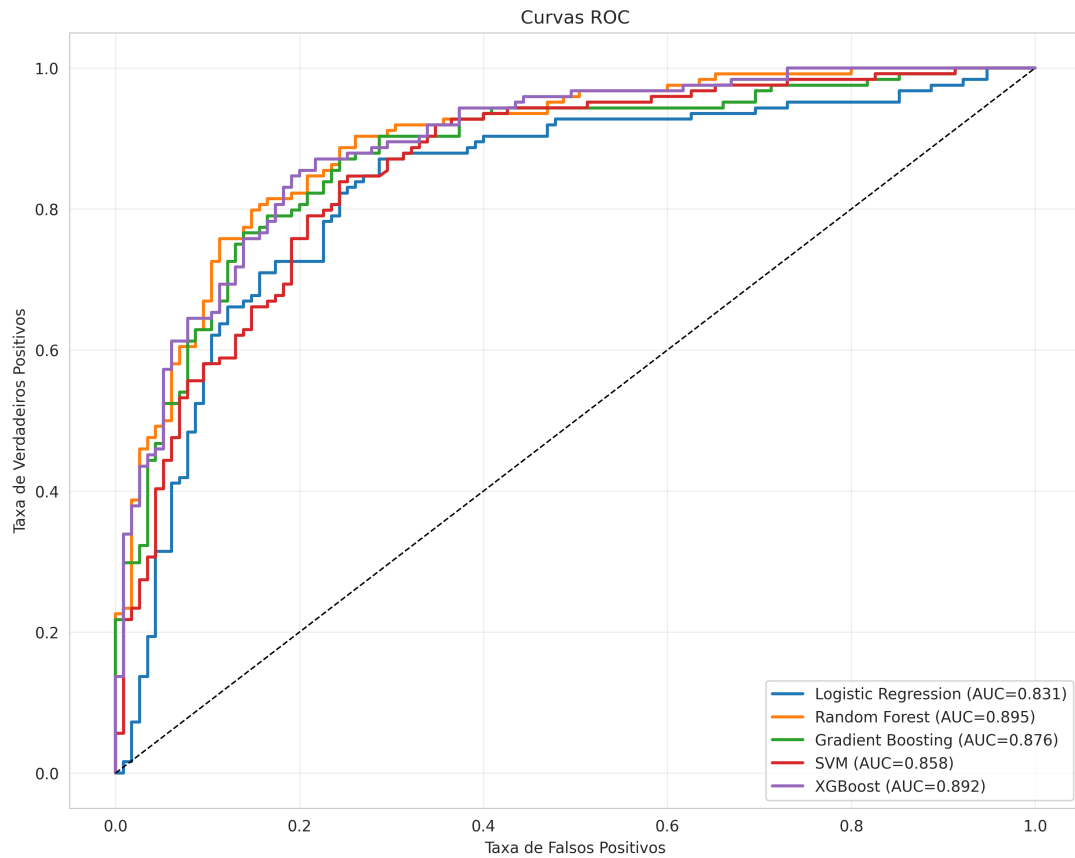


Figura 5.2: Curvas ROC

5.2 Resultados da aplicação em Dispositivos Móveis

Os testes de predição no aplicativo móvel foram executados para os cenários de pacientes com baixo e alto risco de insuficiência cardíaca.

A Figura 5.3 ilustra os resultados observados para as simulações no *smartphone*.



Figura 5.3: Funcionamento do aplicativo em *smartphone*

A Figura 5.4 ilustra os resultados observados para as simulações no *smartwatch*.



Figura 5.4: Funcionamento do aplicativo em *smartwatch*

Nos testes, o aplicativo demonstrou alta responsividade e precisão, exibindo o resultado imediato após o fornecimento da entrada de dados.

A Figura 5.5 ilustra a comparação dos tempos de latência (em milissegundos) para diferentes quantidades de predições executadas nos cenários local e remoto. Nota-se que a inferência realizada localmente apresentou tempos de resposta ínfimos e estáveis ao longo das amostras avaliadas, enquanto a execução em ambiente remoto evidenciou um crescimento praticamente linear da latência conforme aumentou o número de predições.

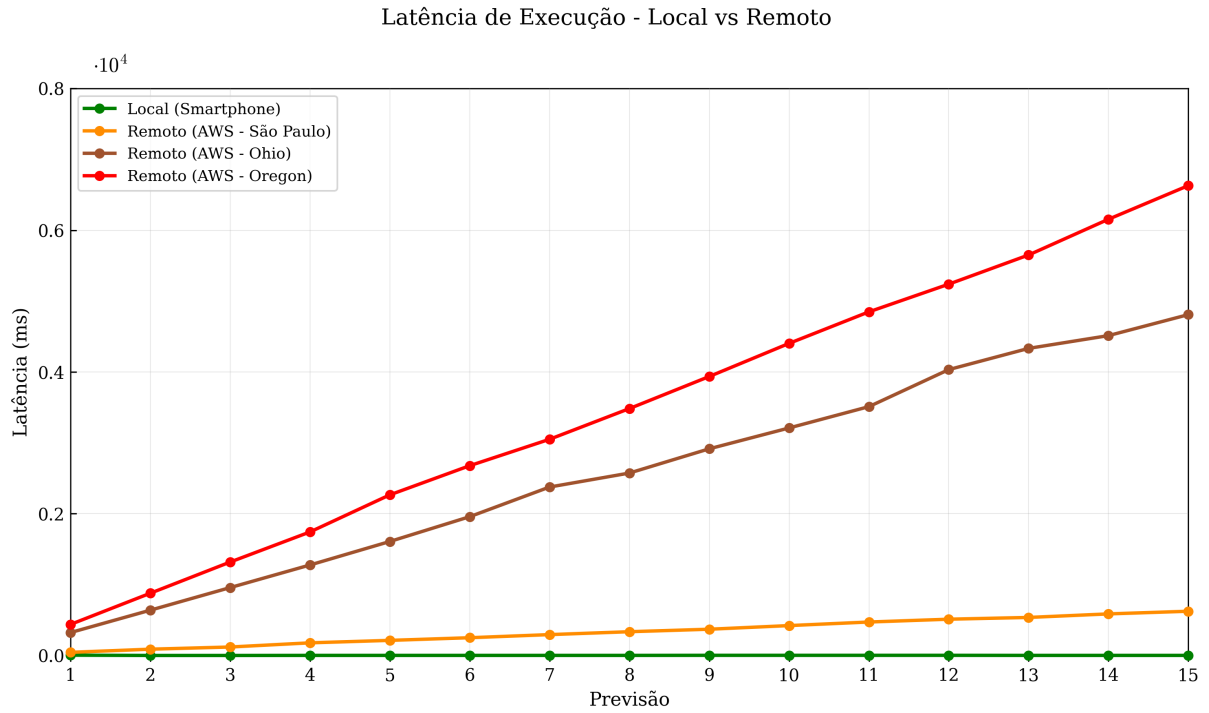


Figura 5.5: Comparação da latência de execução (em ms) em diferentes ambientes

Os resultados indicam que a execução local apresentou latência na média de 1 ms, caracterizando um tempo de resposta praticamente imediato do modelo embarcado no dispositivo móvel. Em contrapartida, no cenário de execução remota, observou-se que a latência aumentou de forma proporcional ao número de predições realizadas. De maneira quantitativa, o tempo médio de resposta das inferências locais foi aproximadamente 99,3% menor quando comparado ao ambiente remoto localizado em Oregon, evidenciando a eficiência da computação de borda na redução do tempo de inferência. Além disso, constatou-se a influência direta da localização geográfica do servidor de processamento em relação à origem dos dados, com uma diminuição significativa da latência à medida que o processamento ocorre mais próximo da fonte. Esse comportamento confirma a premissa de que a proximidade entre os dados e o ambiente de inferência contribui para respostas mais rápidas. Em conjunto, esses resultados reforçam a viabilidade da abordagem de computação de borda para aplicações de saúde, destacando seu potencial como uma solução prática para o monitoramento contínuo em ambientes com infraestrutura de conectividade limitada.

6 Considerações Finais e Trabalhos Futuros

Os resultados obtidos neste trabalho demonstram a viabilidade do uso de modelos de aprendizado de máquina aliados à computação de borda para a predição do risco de insuficiência cardíaca em tempo real. A abordagem proposta mostrou-se eficiente ao realizar a inferência localmente em dispositivos móveis, reduzindo a latência e a dependência de conectividade contínua com a internet, características fundamentais para aplicações em ambientes de vida assistida e monitoramento contínuo da saúde. Ainda que os experimentos tenham sido conduzidos em ambiente simulado, os resultados indicam o potencial da solução como ferramenta de acompanhamento preventivo de indivíduos em risco cardiovascular.

Como trabalhos futuros, destaca-se a necessidade de validação do sistema em cenários reais, envolvendo a participação de profissionais da área da saúde e pacientes. Testes empíricos em dispositivos móveis com diferentes capacidades computacionais também são fundamentais para avaliar a confiabilidade da ferramenta em condições de uso reais.

Outra possibilidade relevante de expansão do trabalho consiste na ampliação do conjunto de variáveis monitoradas. Dados como Variabilidade da Frequência Cardíaca (HRV), qualidade do sono, nível de estresse já são medidos automaticamente por dispositivos de marcas como *Garmin* e *Polar*. A incorporação dessas informações pode enriquecer o modelo preditivo, permitindo uma análise mais sensível às variações diárias do estado de saúde do paciente.

Para cenários em que o usuário não dispõe de dispositivos vestíveis com sensores específicos, uma alternativa viável é o desenvolvimento de um aplicativo móvel complementar. Nesse modelo, além do cadastro inicial com informações como sexo, idade, altura, o usuário poderia informar diariamente dados mais voláteis, como peso corporal, nível de estresse percebido, qualidade do sono, presença de dores, fadiga e outros sintomas relevantes.

No que se refere aos modelos de aprendizado de máquina, embora os algoritmos avaliados tenham apresentado desempenho satisfatório, futuras investigações podem ex-

plorar técnicas mais avançadas, como redes neurais profundas, incluindo redes neurais convolucionais (CNN) e recorrentes (RNN), capazes de capturar padrões temporais e não lineares mais complexos. A comparação entre essas abordagens e os modelos tradicionais pode auxiliar na identificação da estratégia mais adequada para aplicações em tempo real e dispositivos com recursos limitados.

Por fim, o aprendizado federado surge como um caminho inovador para trabalhos futuros, ao permitir o treinamento colaborativo de modelos diretamente nos dispositivos dos usuários, preservando a privacidade dos dados sensíveis. Essa abordagem possibilitaria a atualização contínua do modelo com dados reais de uso, promovendo uma adaptação dinâmica às características individuais dos pacientes e às variações populacionais, sem a necessidade de centralização das informações.

Em síntese, as perspectivas apresentadas reforçam a relevância de uma abordagem multidisciplinar que integre avanços em aprendizado de máquina, computação de borda e dispositivos vestíveis às necessidades clínicas reais. A evolução dessas direções pode contribuir significativamente para o desenvolvimento de soluções de monitoramento de saúde mais precisas, acessíveis e adaptáveis, fortalecendo o cuidado preventivo em ambientes de vida assistida e no contexto do sistema de saúde como um todo.

Bibliografia

- ABDELLATIF, A. A. et al. Edge computing for smart health: Context-aware approaches, opportunities, and challenges. *IEEE Network*, v. 33, n. 3, p. 196–203, 2019.
- ALMEIDA, C. P. de et al. Real-time heart failure prediction: An approach for ambient assisted living. In: *Simpósio Brasileiro de Sistemas de Informação (SBSI 2025)*. Recife, PE, Brasil: SBC, 2025. Disponível em: [⟨https://www.sbc.org.br⟩](https://www.sbc.org.br).
- AMIN, S. U.; HOSSAIN, M. S. Edge intelligence and internet of things in healthcare: A survey. *IEEE Access*, v. 9, p. 45–59, 2021.
- ARIK, S. O.; PFISTER, T. Tabnet: Attentive interpretable tabular learning. *arXiv pre-print arXiv:1908.07442*, 2019.
- BAI, J. et al. *ONNX: Open Neural Network Exchange*. [S.l.]: GitHub, 2019. [⟨https://github.com/onnx/onnx⟩](https://github.com/onnx/onnx).
- BASEER, K. et al. Healthcare diagnostics with an adaptive deep learning model integrated with the internet of medical things (iomt) for predicting heart disease. *Biomedical Signal Processing and Control*, v. 92, 2024. Cited by: 31. Disponível em: [⟨https://www.scopus.com/inward/record.uri?eid=2-s2.0-85185391601&doi=10.1016%2fj.bspc.2024.105988&partnerID=40&md5=8ecdad0a950afd24b953555ddf273fbf⟩](https://www.scopus.com/inward/record.uri?eid=2-s2.0-85185391601&doi=10.1016%2fj.bspc.2024.105988&partnerID=40&md5=8ecdad0a950afd24b953555ddf273fbf).
- BERGSTRA, J.; BENGIO, Y. Random search for hyper-parameter optimization. *Journal of Machine Learning Research*, v. 13, n. 2, p. 281–305, 2012.
- BREIMAN, L. Random forests. *Machine Learning*, v. 45, p. 5–32, 2001.
- CHAWLA, N. V. et al. Smote: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, v. 16, p. 321–357, 2002.
- CHEN, T.; GUESTRIN, C. Xgboost: A scalable tree boosting system. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. [S.l.: s.n.], 2016. p. 785–794.
- CORTES, C.; VAPNIK, V. Support-vector networks. *Machine Learning*, v. 20, p. 273–297, 1995.
- DOROGUSH, A. V.; ERSHOV, V.; GULIN, A. Catboost: Unbiased boosting with categorical features. In: *Advances in Neural Information Processing Systems*. [S.l.: s.n.], 2018. v. 31.
- FRIEDMAN, J. H. Greedy function approximation: a gradient boosting machine. *Annals of Statistics*, p. 1189–1232, 2001.
- HALL, M. et al. The weka data mining software: An update. *SIGKDD Explorations*, ACM, v. 11, n. 1, p. 10–18, 2009.
- HE, H.; GARCIA, E. A. Learning from imbalanced data. *IEEE Transactions on Knowledge and Data Engineering*, v. 21, n. 9, p. 1263–1284, 2009.

IBGE. *Internet chega a 74,9 milhões de domicílios do país em 2024*. 2025. Agência IBGE Notícias. Acesso em: 8 jan. 2026. Disponível em: <https://agenciadenoticias.ibge.gov.br/agencia-noticias/2012-agencia-de-noticias/noticias/44050-internet-chega-a-74-9-milhoes-de-domicilios-do-pais-em-2024>.

JR, D. W. H.; LEMESHOW, S.; STURDIVANT, R. X. *Applied Logistic Regression*. [S.l.]: John Wiley & Sons, 2013.

KHAN, M. S. et al. Trends in 30- and 90-day readmission rates for heart failure. *Circulation: Heart Failure*, v. 14, n. 4, p. e008335, 2021. Disponível em: <https://www.ahajournals.org/doi/abs/10.1161/CIRCHEARTFAILURE.121.008335>.

KOHAVI, R. A study of cross-validation and bootstrap for accuracy estimation and model selection. *Proceedings of the 14th International Joint Conference on Artificial Intelligence*, 1995.

KOLHAR, M.; MISFER, M. An intelligent cardiovascular diseases prediction system focused on privacy. *Intelligent Automation and Soft Computing*, v. 36, n. 1, p. 529 – 542, 2023. Cited by: 2; All Open Access, Hybrid Gold Open Access. Disponível em: <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85139202597&doi=10.32604/2fasc.2023.030098&partnerID=40&md5=a26e0001e490a49110a80371f76a1326>.

KUNDRICK, J. et al. Machine learning applied to wearable fitness tracker data and the risk of hospitalizations and cardiovascular events. *American Journal of Preventive Cardiology*, v. 22, 2025. Cited by: 0. Disponível em: <https://www.scopus.com/inward/record.uri?eid=2-s2.0-105005457341&doi=10.1016%2fj.ajpc.2025.101006&partnerID=40&md5=8f441b50bfa69fdfe7043be98c37453>.

Microsoft. *ONNX Runtime: cross-platform, high performance ML inferencing and training accelerator*. 2021. Acessado em: 09/01/2026. Disponível em: <https://onnxruntime.ai/>.

MITCHELL, T. M. *Machine Learning*. New York: McGraw-Hill Education, 1997.

NIROB, M. A. S. et al. *CAIR-CVD-2025: An Extensive Cardiovascular Disease Risk Assessment Dataset from Bangladesh*. Mendeley Data, 2025. Disponível em: <https://doi.org/10.17632/d9scg7j8fp.1>.

OUZZANI, M. et al. Rayyan—a web and mobile app for systematic reviews. *Systematic Reviews*, BioMed Central, v. 5, n. 1, p. 210, 2016.

Presidência da República. *Lei n.º 13.709, de 14 de agosto de 2018 – Lei Geral de Proteção de Dados Pessoais (LGPD)*. 2018. https://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/l13709.htm. Acesso em 16 de julho de 2025.

RAJKOMAR, A.; DEAN, J.; KOHANE, I. Machine learning in medicine. *New England Journal of Medicine*, v. 380, n. 14, p. 1347–1358, 2019.

RANI, R. Optimized heart failure prediction using support vector machine algorithms. In: . [s.n.], 2024. p. 1265 – 1268. Cited by: 1. Disponível em: <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85209895593&doi=10.1109%2fICOCSEC61587.2024.10722131&partnerID=40&md5=942b6df8c2bb4102dd78e7e6f8e2a6b7>.

ROMAN, R.; LOPEZ, J.; MAMBO, M. Mobile edge computing, fog et al.: A survey and analysis of security threats and challenges. *Future Generation Computer Systems*, v. 78, p. 680–698, 2018. ISSN 0167-739X. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0167739X16305635>.

RUSSELL, S.; NORVIG, P. *Artificial Intelligence: A Modern Approach*. 4. ed. [S.l.]: Pearson, 2020.

Scikit-learn. *GridSearchCV — scikit-learn documentation*. 2024. Disponível em: https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.GridSearchCV.html.

Scikit-learn. *StratifiedKFold — scikit-learn documentation*. 2024. Disponível em: https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.StratifiedKFold.html.

Scikit-learn. *One-Hot Encoding*. 2025. <https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.OneHotEncoder.html>. Acesso em 09 de janeiro de 2026.

Scikit-learn. *Scikit-learn*. 2025. <https://scikit-learn.org/stable/index.html>. Acesso em 09 de janeiro de 2026.

Scikit-learn. *StandardScaler*. 2025. <https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.StandardScaler.html>. Acesso em 09 de janeiro de 2026.

SHI, W.; DUSTDAR, S. The promise of edge computing. *Computer*, v. 49, n. 5, p. 78–81, 2016.

Sociedade Brasileira de Cardiologia. Diretriz brasileira de insuficiência cardíaca crônica e aguda. *Arquivos Brasileiros de Cardiologia*, São Paulo, v. 111, n. 3, p. 436–539, 2018.

STEHLIK, J. et al. Continuous wearable monitoring analytics predict heart failure hospitalization: The link-hf multicenter study. *Circulation: Heart Failure*, v. 13, n. 3, p. E006513, 2020. Cited by: 212. Disponível em: <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85083855271&doi=10.1161%2fCIRCHEARTFAILURE.119.006513&partnerID=40&md5=e373bac0a66f00c492e5741386aa32cd>.

SUMWIZA, K. et al. Enhanced cardiovascular disease prediction model using random forest algorithm. *Informatics in Medicine Unlocked*, v. 41, 2023. Cited by: 23; All Open Access, Gold Open Access. Disponível em: <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85166942213&doi=10.1016%2fj.imu.2023.101316&partnerID=40&md5=caae3298ee8950456e0ff00b142f20b>.

SURESH, R.; DAYANA, T. N. A comprehensive review of heart disease prediction using cloud-driven machine learning. In: . [s.n.], 2024. p. 160 – 167. Cited by: 3. Disponível em: <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85195111571&doi=10.1109%2fICC-ROBINS60238.2024.10533949&partnerID=40&md5=353690a8af07a824d526941d04ac4c73>.

UMER, M. et al. Iot based smart monitoring of patients' with acute heart failure. *Sensors*, v. 22, n. 7, 2022. Cited by: 55; All Open Access, Gold Open Access, Green Open Access. Disponível em: <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85126797804&doi=10.3390%2fs22072431&partnerID=40&md5=84f1e1fcd89c26d6bcb568f900cb7c4b>.

USP, J. *Doenças cardiovasculares são a principal causa de mortes no Brasil, segundo o Ministério da Saúde*. 2024. Publicado em 06/08/2024. Acesso em: 14 jul. 2025. Disponível em: <https://jornal.usp.br/?p=789024>.

VYAS, G.; KALSHETTY, J. N.; PAREEK, P. K. Designing a smart healthcare framework based on disease prediction using optimizer based ai. In: . [s.n.], 2023. Cited by: 0. Disponível em: <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85184821929&doi=10.1109%2fMysuruCon59703.2023.10397017&partnerID=40&md5=71f02cbe5fe2fbc691bf62c906f460fd>.

WANG, X. et al. Convergence of edge computing and deep learning: A comprehensive survey. *IEEE Communications Surveys and Tutorials*, PP, p. 1–1, 01 2020.

World Health Organization. *The top 10 causes of death*. 2024. <https://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death>. Acesso em 09 de janeiro de 2026.

YANG, P. C. et al. Cloud-based machine learning platform to predict clinical outcomes at home for patients with cardiovascular conditions discharged from hospital: Clinical trial. *JMIR Cardio*, v. 8, 2024. Cited by: 1; All Open Access, Gold Open Access, Green Open Access. Disponível em: <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85192776829&doi=10.2196%2f45130&partnerID=40&md5=cfbb26f43233a2bb392521a96e951651>.