

Uso da Inteligência Artificial Explicável aplicada à predição de doenças cardíacas

Use of Explainable Artificial Intelligence applied to the prediction of heart diseases

Valéria Barbosa de Araújo¹, Flávio Rosendo da Silva Oliveira¹

¹ Análise e Desenvolvimento de Sistemas – Instituto Federal de Pernambuco Campus Paulista (IFPE) Paulista – PE – Brasil

vba@discente.ifpe.edu.br, flavio.oliveira@paulista.ifpe.edu.br

Resumo. O coração é o maior órgão do sistema circulatório, ele é fundamental para bombear o sangue rico em oxigênio e nutrientes para todas as células do corpo humano. Problemas que ocorrem no funcionamento do coração, como palpitações, dor no peito e falta de ar, são sintomas que, quando graves, são indícios da presença de doenças cardíacas. Anualmente, as doenças cardiovasculares estão entre as doenças que mais matam no mundo e muitas dessas mortes poderiam ser evitadas com medidas efetivas de prevenção da doença. Atualmente, o uso da Mineração de Dados tem sido empregado na criação de sistemas que visam auxiliar no diagnóstico e prevenção de doenças cardíacas. Por isso, nesse artigo, será utilizada a Mineração de Dados aliada à Inteligência Artificial Explicável para analisar quais são as características mais importantes que um modelo considera ao predizer doenças cardíacas. O Random Forest e o XGBoost foram os classificadores que obtiveram os melhores resultados, apresentando acurácias e recalls acima de 85%. Os resultados sugerem que o uso do SHapley Additive exPlanations como modelo da Inteligência Artificial Explicável foi eficaz ao apontar que as características Inclinação do segmento ST do pico do exercício e Oldpeak, por exemplo, tiveram uma forte influência nos classificadores ao predizer a presença de doença cardíaca. Portanto, o uso da Inteligência Artificial Explicável pode ser um recurso importante a ser utilizado no processo de classificação de doenças.

Palavras-chave: Doenças cardíacas; Mineração de Dados; Inteligência Artificial Explicável.

Abstract. The heart is the largest organ in the circulatory system and is essential for pumping blood rich in oxygen and nutrients to all cells in the human body. Problems that occur in the functioning of the heart, such as palpitations, chest pain and shortness of breath, are symptoms that, when severe, are indications of the presence of heart disease. Every year, cardiovascular diseases are among the diseases that kill the most people in the world and many of these deaths could be avoided with effective measures to prevent the disease. Currently, the use of Data Mining has been used to create systems that aim to assist in the

diagnosis and prevention of heart disease. Therefore, in this article, Data Mining will be used combined with Explainable Artificial Intelligence to analyze which are the most important characteristics that a model considers when predicting heart disease. Random Forest and XGBoost were the classifiers that obtained the best results, presenting accuracies and recalls above 85%. The results suggest that the use of SHapley Additive exPlanations as an Explainable Artificial Intelligence model was effective in pointing out that the features ST segment slope of the exercise peak and Oldpeak, for example, had a strong influence on the classifiers when predicting the presence of heart disease. Therefore, the use of Explainable Artificial Intelligence can be an important resource to be used in the disease classification process.

Keywords: Heart Disease; Data Mining; Explainable Artificial Intelligence.

1. Introdução

As doenças cardíacas podem ser definidas como um conjunto de doenças que afetam o coração e os vasos sanguíneos. O aparecimento delas está geralmente associado ao acúmulo de gordura dentro das artérias e ao aumento de coágulos sanguíneos (Nhs, 2022). Entre as doenças cardiovasculares mais comuns estão os ataques cardíacos, os acidentes vasculares cerebrais (AVC), a doença isquêmica do coração e as doenças hipertensivas (Dattani *et al.*, 2023).

As doenças cardiovasculares estão situadas entre as maiores causas de mortalidade de adultos no mundo. Conforme dados da Organização Mundial da Saúde apresentados em 2020, as doenças ligadas ao coração ficaram entre os primeiros lugares num comparativo de 10 doenças que mais matam globalmente, com mais de 8 milhões de mortes entre os anos de 2000 a 2019 (Who, 2020).

De acordo com dados da *British Heart Foundation*, em 2021, havia aproximadamente 620 milhões de pessoas no mundo vivendo com doenças cardíacas e, a cada ano, cerca de 60 milhões de pessoas desenvolvem doenças cardiovasculares. Esse crescimento se deve a mudança no estilo de vida e ao envelhecimento e crescimento da população mundial (British heart foundation, 2024). Outros fatores de risco que mais contribuem para o aparecimento de doenças cardíacas são “o uso de tabaco, dietas inadequadas e obesidade, sedentarismo e o uso nocivo do álcool, hipertensão, diabetes e hiperlipidemia” (Opas, 2024).

No Brasil, a situação não é diferente, conforme dados do Instituto Nacional de Cardiologia (INC), houve um aumento significativo de 161,93% no número de internações por infarto agudo do miocárdio durante o período de 2008 a 2022 (Inc, 2024). Além disso, segundo dados da Sociedade Brasileira de Cardiologia, ocorrem mais de 1.100 mortes diárias no Brasil por doenças cardiológicas e muitas dessas mortes poderiam ser evitadas por meio de medidas de prevenção e tratamentos adequados (Sbc, 2023).

Atualmente, o uso da Mineração de Dados tem crescido e se expandido para outros campos que vão além da computação. O uso dela em conjunto com a medicina tem

ajudado a criar sistemas de apoio a decisão que tem auxiliado equipes médicas na detecção e prevenção de diversas doenças, principalmente as doenças cardíacas.

A avaliação e confiabilidade dos sistemas de saúde criados com Inteligência Artificial (IA) depende da ação conjunta entre os especialistas na área e os cientistas de dados. Esse processo de avaliação dos modelos de IA, de acordo com Costa (2023), nem sempre é um processo fácil, principalmente quando envolve entender quais são as variáveis mais importantes que levam um modelo a ter uma predição mais correta possível. Por causa disso, houve a necessidade da criação de métodos para explicar os resultados gerados pelos modelos.

Dentre os métodos mais conhecidos está a Inteligência Artificial Explicável (XAI). A XAI pode ser definida como um conjunto de “métodos e técnicas que produzem modelos explicáveis e precisos, evidenciando o porquê de um modelo chegar a uma decisão específica. Desse modo, soluções obtidas por sistemas artificialmente inteligentes podem ser entendidas por humanos” (Fernandes, 2022, p. 9–10). O uso da XAI auxilia os pesquisadores, cientistas de dados e *stakeholders* a compreender melhor os resultados gerados pelos modelos de IA.

Apesar de existirem diversos estudos que utilizam a Mineração de Dados aplicada a predição de doenças cardíacas, muitos desses estudos não avaliam a importância das *features* utilizadas nos modelos. Nesse trabalho, será empregado o método de explicabilidade XAI para avaliar quais são as características mais importantes que levam um modelo a prever se uma pessoa tem ou não uma doença cardíaca.

Por isso, o objetivo desse trabalho é verificar como o uso da XAI pode ser utilizada para adicionar camadas de explicação em classificadores tradicionais na predição de doenças cardíacas. Como contribuições, o seguinte trabalho propõe comparar o desempenho de cinco modelos classificadores com o uso do modelo da XAI, SHAP.

O SHAP é um modelo que utiliza “conceitos da teoria dos jogos cooperativos para calcular a contribuição de cada atributo em uma previsão gerada [por um] modelo” (Cesaro, 2021, p. 11). O objetivo do uso do SHAP é avaliar de forma global quais são as características mais importantes apontadas pelos modelos ao prever a presença de doença cardíaca.

Este artigo está organizado da seguinte forma: A Seção 2 apresenta o referencial teórico do trabalho, abordando os conceitos de Aprendizagem de Máquina, da XAI, do SHAP, assim como os conceitos dos modelos utilizados nesse trabalho. Essa seção aborda também os trabalhos relacionados a proposta deste estudo.

A Seção 3 apresenta a metodologia utilizada no trabalho. Nessa seção são abordadas como algumas etapas da metodologia *Cross Industry Standard Process for Data Mining* (CRISP-DM) são utilizadas nos processos de análise exploratória dos dados, preparação dos dados e modelagem dos classificadores.

A Seção 4 apresenta os resultados obtidos com os modelos que obtiveram o melhor desempenho entre os classificadores e o uso do SHAP nos modelos escolhidos. Por fim, a Seção 5 apresenta as considerações finais do trabalho e trabalhos futuros.

2. Referencial Teórico

2.1. Aprendizagem de Máquina

A Aprendizagem de Máquina é uma subárea da IA que ajuda a melhorar a performance de sistemas por meio da aprendizagem com dados e da identificação de padrões. Com isso, esses sistemas podem gerar previsões e, a partir dos resultados obtidos, tomar decisões sobre um problema. Com os algoritmos de aprendizagem, é possível construir modelos alimentados com dados que podem fazer previsões (Zhou; Liu, 2021). A Aprendizagem de Máquina visa ensinar as máquinas a aprender com os dados e usá-los da forma mais eficiente possível (Mahesh, 2020).

Os três tipos mais conhecidos de Aprendizagem de Máquina são a Aprendizagem Supervisionada, Aprendizagem Não Supervisionada e a Aprendizagem por Reforço. A Aprendizagem Supervisionada é usada quando se deseja prever algo. O modelo recebe um conjunto de dados rotulados, esses dados são divididos em dados de treinamento e teste. No treinamento, o modelo aprende padrões com os dados. Na etapa de teste, os dados são usados para avaliar o algoritmo e a avaliação é feita por meio de cálculos para verificar a taxa de erros e acertos que o modelo obteve (Silva, 2017).

A Aprendizagem Não Supervisionada, diferentemente da Aprendizagem Supervisionada, os algoritmos aprendem pouco com os dados, o objetivo principal é reconhecer padrões e analisar o relacionamento que os dados possuem entre si. Um exemplo de algoritmos de Aprendizagem Não Supervisionada são os *k-means*.

Na Aprendizagem por Reforço, como o próprio nome define, os modelos aprendem por reforço ou repetição, baseada num sistema de erros e acertos. Segundo Schiaffino (2020, p.32), “existe um retorno sobre o sucesso do processo de aprendizagem, permitindo que o sistema treine com base na crítica sobre seu desempenho”.

2.2. Aprendizagem de Máquina aplicada na saúde

O uso da IA aplicada a saúde vem crescendo nos últimos anos devido a sua capacidade de predição de diagnóstico de doenças. Uma das áreas da medicina em que algoritmos de IA são usados é a área da Cardiologia. Diversos estudos e pesquisas são feitos para desenvolver algoritmos que auxiliem na identificação precoce de doenças e ataques cardíacos. Um exemplo desses sistemas é o biomarcador criado pelos pesquisadores da Universidade de *Oxford*.

Esse biomarcador, chamado de *Fat Radiomic Profile* (FRP), ou Perfil Radiômico de Gordura, foi desenvolvido para detectar “sinais de alerta biológicos no espaço que reveste os vasos sanguíneos que fornecem sangue ao coração. Ele identifica inflamação, cicatrizes e alterações nesses vasos sanguíneos, que são todos indicadores de um futuro ataque cardíaco” (Jones, 2019).

O estudo foi feito com dados de 5.487 indivíduos. Dentre eles, os pesquisadores analisaram 101 pessoas que precisaram ir para o hospital por apresentarem sintomas como dor no peito. Os pacientes realizaram um exame específico para esse caso, a angiografia coronária por tomografia computadorizada. Os resultados desse exame não trouxeram indícios de que as pessoas analisadas tinham risco de ataque cardíaco.

Os pesquisadores de *Oxford* analisaram e compararam os resultados dos exames desses pacientes que, em um período de 5 anos após terem feito a angiografia, tiveram um ataque cardíaco ou tiveram morte cardiovascular. Eles analisaram as mudanças que houve nesse período no “espaço perivascular que indicam que alguém tem maior risco de um ataque cardíaco” (Jones, 2019). O biomarcador desenvolvido captura o nível de risco de ataque cardíaco que uma pessoa tem com base nos exames médicos adicionados no algoritmo.

Outro estudo realizado pela Associação Americana do Coração teve por objetivo criar um estetoscópio digital desenvolvido com IA para registrar eletrocardiogramas e sons cardíacos e a partir desses dados, usar um algoritmo para detectar a existência de um músculo cardíaco enfraquecido.

O estudo foi feito com 1.195 mulheres nigerianas que estavam grávidas ou que estavam no terceiro semestre de gravidez, com objetivo de detectar se elas apresentavam sintomas de cardiomiopatia periparto, que é um tipo de insuficiência cardíaca que afeta mulheres nos meses finais da gravidez e após a gravidez. Essa doença “enfraquece o coração, resultando em uma diminuição na quantidade de sangue que é bombeado do coração para outras partes do corpo” (American heart association, 2023).

No estudo desenvolvido, as mulheres foram divididas em grupos de controle, ou seja, as que fizeram testes de eletrocardiograma tradicionais sem o uso do estetoscópio digital com IA e o grupo que teve intervenção assistida por IA. Os resultados do estudo mostraram que no grupo de controle, a detecção da doença foi de 1,8%, enquanto o grupo que utilizou o estetoscópio digital a detecção da doença foi de 4%.

Segundo a principal autora do estudo, a pesquisa pode mudar a prática médica atual “para uma abordagem mais proativa de identificação de disfunção cardíaca relacionada a gravidez usando uma ferramenta de triagem simples, de baixo custo e eficaz” (American heart association, 2023).

2.3. Modelos de Aprendizado de Máquina

2.3.1. Logistic Regression

O *Logistic Regression* é um modelo estatístico que utiliza a função logística para mapear, por meio da função matemática sigmoide, a relação entre duas variáveis. O modelo tenta prever o valor de uma variável alvo, ou variável dependente, com base nas variáveis independentes. Alguns dos benefícios do uso do *Logistic Regression* são a simplicidade de aplicação do modelo e a velocidade de processamento com uma grande quantidade de dados.

Algumas aplicações que podem ser feitas com o *Logistic Regression* segundo Carneiro, Sorrentino e Rocha (2023) são: a área da saúde, com a predição de doenças e a área financeira, onde é possível prever a inadimplência de clientes através de vários fatores, como a renda. Além da detecção de fraudes em cartões, onde é possível classificar as transações feitas como fraudulentas e não fraudulentas (Silva, 2022).

2.3.2. KNN

O KNN (*k-nearest neighbors*) é um modelo de classificação de Aprendizagem Supervisionada. Ele analisa pontos de semelhança que existem entre agrupamentos e um ponto individual analisado. Ele tem como principal objetivo identificar, em um conjunto

de dados, os vizinhos mais próximos do elemento que será classificado. Quanto mais semelhantes os elementos forem, mais próximos eles serão (Torres, 2022).

Para calcular a distância entre o ponto que está sendo consultado e o agrupamento dos outros pontos são utilizadas algumas funções matemáticas, como a distância euclidiana. O valor de k no algoritmo é o número de vizinhos mais próximos que devem ser utilizados para classificar o ponto que está sendo analisado.

2.3.3. MLP

O *MultiLayer Perceptron* é um modelo de rede neural artificial que simula o comportamento de um cérebro humano. Os neurônios no cérebro humano possuem um potencial de ativação onde recebem sinais elétricos, e esses sinais são propagados para os outros neurônios para gerar uma saída ou ação. De maneira similar, no MLP, cada neurônio na camada de entrada da rede neural, chamados de receptores, recebem os dados e para cada entrada são atribuídos pesos calculados nas camadas intermediárias, ou camadas ocultas da rede.

Esses valores, ou sinais, são propagados nas outras camadas da rede para que, por fim, uma função de ativação calcule as saídas previstas pela rede neural (Desai; Shah, 2021). Esses procedimentos são chamados de *feedforward*. Para treinar o MLP é utilizada a técnica chamada de *Backpropagation*, onde os pesos da rede são recalculados para minimizar a função de perda. Essa função mede a diferença entre as previsões geradas pelo modelo e os valores-alvo esperados.

2.3.4. XGBoost

O *XGBoost (Extreme Gradient Boost)* é um modelo de Aprendizagem de Máquina que realiza uma combinação de múltiplas árvores de decisão utilizando uma estrutura de *Gradient boosting*. Algumas vantagens de implementação do modelo *XGBoost* são a rapidez do processamento que ele possui, sua capacidade de escalabilidade, além de possuir uma etapa de otimização para cada árvore que o modelo utiliza.

Por causa disso, o modelo consegue gerar menos alarmes falsos, rotular e classificar melhor os dados. O *XGBoost* possui também um sistema de regularização de dados para minimizar o *overfitting* nos dados (Dhaliwal; Nahid; Abbas, 2018).

2.3.5. Random Forest

O *Random Forest* é um modelo de Aprendizagem de Máquina que combina um conjunto de árvores de decisão para formar uma subamostra de conjunto de dados, além de usar “a média para melhorar a precisão preditiva e controlar o sobreajuste” (Scikitlearn, 2024).

O algoritmo seleciona aleatoriamente algumas características e a característica escolhida é posicionada no nó raiz da árvore. Após isso, nós filhos são gerados até que a quantidade de árvores desejada seja obtida. As previsões feitas são obtidas a partir do consenso entre as decisões tomadas pelas árvores no modelo (Pessanha, 2019).

2.4. XAI

A XAI pode ser definida como um conjunto de técnicas usadas para auxiliar na interpretação dos modelos de Aprendizagem de Máquina. Os modelos de IA, assim como os modelos de Aprendizagem Profunda, não é possível saber quais foram os passos ou fatores que mais influenciaram o modelo nas tomadas de decisão. De acordo com Jagati (2023), “a complexidade desses modelos e as transformações não lineares que realizam tornam extremamente desafiador rastrear a lógica por trás de suas saídas”.

Por causa disso, surgiu a necessidade da criação da XAI, visando compreender melhor os resultados obtidos com os modelos complexos, também conhecidos como modelos caixa-preta. A utilização da XAI em modelos de Inteligência Artificial traz vários benefícios, entre eles se destacam o aumento da confiança nos modelos, a transparência algorítmica e a mitigação de vieses nos sistemas inteligentes (Mohseni; Zarei; Ragan, 2021 apud Ferraz *et al.*, 2024). Apesar de existirem várias técnicas de XAI, nesse trabalho será abordada a técnica SHAP.

2.4.1. SHAP

O *SHapley Additive exPlanations* (SHAP) pode ser definido como uma “abordagem teórica de jogos para explicar a saída de qualquer modelo de aprendizado de máquina. Ele conecta a alocação ideal de crédito com explicações locais usando os valores clássicos de *Shapley* da teoria dos jogos e suas extensões relacionadas” (Lundberg, 2018). O conceito do valor *Shapley* tem sua origem na teoria dos jogos, onde o valor de um jogador individual é medido em uma equipe. Ao final do jogo, a teoria propõe que os jogadores recebam uma recompensa conforme os resultados que eles obtiveram durante o jogo.

De maneira similar, o SHAP mede o valor individual de cada variável e explica como elas impactam nos resultados de um modelo. Um dos benefícios que o SHAP possui é a interpretabilidade global, pois por meio dele é possível avaliar o nível de influência, positiva ou negativa, que cada preditor do modelo tem na variável alvo. Outro benefício que o SHAP possui é a interpretabilidade local, ou seja, com ele é possível explicar como cada preditor contribui para as previsões do modelo.

O SHAP pode ser utilizado para qualquer modelo, porém, dependendo do tipo de modelo que for utilizado, um explicador SHAP é usado. Por exemplo, modelos baseados em árvores utilizam o *TreeExplainer*, enquanto modelos *Deep Learning* utilizam o *DeepExplainer* e os outros tipos de modelos utilizam o *KernelExplainer* (Kuo, 2019).

2.5. Trabalhos Relacionados

Nessa seção, seção abordados trabalhos científicos feitos utilizando a Mineração de Dados aplicada a previsão de doenças cardíacas.

O estudo feito por Bhatt *et al.* (2023) teve como objetivo principal utilizar vários algoritmos de aprendizado de máquina para investigar a eficácia deles na previsão de doenças cardíacas. O diferencial que o trabalho propôs foi utilizar métodos de agrupamento, ou *clusters*.

O algoritmo escolhido foi o *k-modes* devido a sua melhor eficiência ao lidar com dados categóricos. O conjunto de dados foi dividido por gênero. Essa divisão teve a proposta de verificar padrões e fatores de risco que determinado gênero possui, uma vez que homens e mulheres possuem riscos distintos de desenvolver a doença. Para

determinar qual o conjunto ideal de *clusters* nos conjuntos de dados por gênero, foi utilizado o método da curva de cotovelo.

Os modelos de Aprendizagem de Máquina utilizados no estudo foram a Árvore de Decisão, a Floresta Aleatória, o *Perceptron* multicamadas e o *XGBoost*. Para ajustar os hiperparâmetros nos modelos, foi utilizado o método *GridSearchCV* e o comportamento dos modelos foi avaliado com e sem o uso dos hiperparâmetros.

O classificador melhor avaliado nos resultados do estudo com o uso dos hiperparâmetros foi o *Perceptron* multicamadas, obtendo uma acurácia de 87,28%, *precision* de 88,70%, *recall* de 84,85% e *f1-score* de 86,71%.

Os autores enfatizam a capacidade que o uso do agrupamento *k-modes* tem no processo de previsão de doenças cardiovasculares e como o uso do algoritmo pode ser utilizado para desenvolver sistemas para diagnóstico da doença.

O trabalho de Ahamad *et al.* (2023) propõe fazer uma avaliação experimental com seis modelos de Aprendizagem de Máquina visando verificar o desempenho dos algoritmos com e sem o uso da técnica de ajuste de hiperparâmetros *GridSearchCV* ao prever doenças cardíacas. Os autores utilizaram duas bases de dados para avaliar a acurácia em cada um dos algoritmos utilizados.

No conjunto de dados de teste, utilizando a base de dados *UCI Kaggle Cleveland* os algoritmos usados sem o ajuste de hiperparâmetros, obtiveram as seguintes acurácias: SVM (87,91%), *Logistic Regression* (86,91%), KNN (86,81%), *Random Forest* (82,42%), *XGBoost* (82,42%) e *Decision Tree* (78,02%).

Com os hiperparâmetros ajustados, os modelos obtiveram as seguintes acurácias: KNN (87,91%), *Logistic Regression* (85,71%), *Random Forest* (84,62%), SVM (84,62%), *Decision Tree* (81,32%) e *XGBoost* (79,12%).

Com o uso do conjunto de dados *Comprehensive UCI Kaggle Cleveland, Hungary, Switzerland, and Long Beach V* os autores tiveram os seguintes resultados com os algoritmos usados sem o ajuste de hiperparâmetros: *XGBoost* (99,03%), *Random Forest* (98,05%), *Decision Tree* (97,08%), SVM (90,26%), KNN (81,82%), *Logistic Regression* (81,82%).

Com o uso de hiperparâmetros os algoritmos tiveram os seguintes resultados: *XGBoost* (99,03%), *Random Forest* (98,05%), SVM (98,05%), *Decision Tree* (97,08%), KNN (83,44%) e *Logistic Regression* (82,14%). Os autores apontaram como o aumento nas taxas de acurácia dos modelos que tiveram os hiperparâmetros ajustados podem ser usados para aumentar a precisão dos classificadores na previsão de doenças cardíacas.

Sun e Pan (2023) avaliaram o desempenho de 5 algoritmos de Aprendizagem de Máquina utilizando indicadores de condição física auto mensuráveis, ou seja, indicadores obtidos pelos próprios pacientes, em comparação com indicadores obtidos no hospital, com objetivo de prever doenças cardíacas. A base de dados utilizada foi o conjunto de dados do coração de *Cleveland* do repositório UCI. Foram usados os seguintes algoritmos no estudo: Regressão Logística, *K-Nearest Neighbors*, SVM, Árvore de Decisão e Floresta Aleatória.

Entre os modelos que utilizaram os indicadores de condições físicas obtidos no hospital, a Regressão Logística e o SVM tiveram os melhores resultados, com acurácias

de 88,52% nos dois modelos. Em relação ao uso de indicadores auto mensuráveis, o modelo que obteve a melhor acurácia foi a Regressão Logística com 73,77%.

Outra métrica avaliada entre os modelos foi a taxa de falsos negativos. No conjunto de teste, o modelo Árvore de Decisão teve a menor taxa entre os algoritmos que utilizaram indicadores auto mensuráveis, com uma taxa de 17,65% de falsos negativos. Entre os modelos que utilizaram todos os indicadores de condição física, o modelo *Logistic Regression* obteve a menor taxa de falsos negativos, com 8,82%.

Portanto, os autores apontaram nos resultados do estudo que os algoritmos que utilizaram indicadores de condição física obtidos no hospital tiveram um desempenho melhor do que os algoritmos que utilizaram indicadores obtidos pelos próprios pacientes.

Paudel *et al.* (2023) avaliaram o desempenho de cinco modelos de Aprendizagem de Máquina para prever ataque cardíaco usando a XAI com o LIME. Os autores utilizaram o coeficiente de correlação de *Pearson* para avaliar o grau de correlação que os recursos têm entre si, porém, como os resultados obtidos foram baixos, os autores optaram por usar todas as variáveis nos modelos.

Os algoritmos utilizados no estudo foram o *AdaBoost*, *Gradient Boosting Classifier*, *LightGBM* e o *Random Forest*. Os resultados obtidos mostraram que os classificadores que tiveram o maior destaque foram o *Random Forest*, *Gradient Boosting Classifier* e o LGBM.

Os modelos apresentaram um nível de *precision* alto ao classificar a classe “Normal” (sem risco de ataque cardíaco), obtendo os seguintes valores com o conjunto de dados de teste: O *Random Forest* obteve 98,02%, o *Gradient Boosting Classifier* obteve 97,98% e o LGBM obteve 97,98%.

Foram avaliados também o *recall* na classe “Ataque” (risco de ataque cardíaco) obtendo os seguintes valores com os dados de teste: *Random Forest* 98,77%, *Gradient Boosting Classifier* 98,77% e o LGBM 98,77%. Ao avaliar a acurácia com o conjunto de teste, os três modelos citados acima obtiveram desempenho semelhante. O *Random Forest* obteve 98,48%, o *Gradient Boosting Classifier* teve uma acurácia de 97,72% e o LGBM atingiu 97,72%.

Ao avaliar o desempenho dos modelos com os dados de treinamento, os autores utilizaram o *Cross Validation K-fold* (10 vezes) e o modelo LGBM obteve o melhor desempenho entre os modelos, com 99,33% de acurácia. Ao usar o algoritmo de XAI, LIME, para avaliar a importância dos recursos localmente, as variáveis “*Troponin*” e “*kcm*” foram as variáveis que mais influenciaram os modelos ao apontar o risco de ataque cardíaco.

O trabalho de Sahu *et al.* (2023) propõe analisar duas bases de dados diferentes, a primeira base avalia a probabilidade de ter ataque cardíaco ou não, enquanto a outra base usada, verifica se há probabilidade de morte devido a problemas cardíacos. Foram utilizados seis algoritmos de *Machine Learning* (ML) tradicionais: SVM, *Decision Tree*, *Random Forest*, *Naive Bayes*, *Logistic Regression*, KNN e o algoritmo de Rede Neural Convolutiva Unidimensional (1DCNN).

Os autores compararam o desempenho dos modelos tradicionais com o modelo de rede neural. A Tabela 1 mostra as bases de dados utilizadas no estudo.

Tabela 1. Tabela comparativa dos algoritmos analisados nos trabalhos relacionados

Referências	Base de dados	XAI	Algoritmo escolhido
Paudel <i>et al.</i> (2023)	<i>Heart Disease Classification Dataset (Kaggle)</i>	LIME	LGBM 99,33% de acurácia
Ahamad <i>et al.</i> (2023)	<i>UCI Kaggle Cleveland (I) e UCI Kaggle Cleveland, Hungary Switzerland, and Long Beach V (II)</i>	Não	KNN 87.91% de acurácia (Base I) XGBoost 99.03% de acurácia (Base II)
Sun e Pan (2023)	Conjunto de dados do coração de <i>Cleveland</i> do repositório UCI.	Não	<i>Logistic Regression</i> e SVM com acurácias de 88.52%. O <i>Logistic Regression</i> teve o menor percentual de falsos negativos com uma taxa de 8.82%.
Sahu <i>et al.</i> (2023)	<i>UCI repository – Heart Failure Clinical Records Kaggle - Heart Attack Analysis Prediction Dataset</i>	Não	1DCNN - 99.9% de acurácia
Bhatt <i>et al.</i> (2023)	<i>Cardiovascular Disease dataset (Kaggle)</i>	Não	<i>Perceptron</i> multicamadas: Acurácia de 87.28% <i>precision</i> de 88.70%, <i>recall</i> de 84.85% e <i>f1-score</i> de 86.71%.

Fonte: Elaboração própria (2024)

Os resultados obtidos mostraram que o modelo tradicional que teve o melhor desempenho foi o *Random Forest*, com acurácias de 88,33% e 80,33% nos dois *datasets* utilizados, enquanto a rede neural obteve acurácias de 99,9% também nas mesmas bases de dados. Portanto, os autores apontaram como o uso de uma rede neural pode ser um recurso importante a ser utilizado no processo de previsão de ataque cardíaco.

Portanto, esse trabalho se assemelha aos estudos citados acima com relação ao uso dos modelos utilizados, por serem classificadores geralmente usados na literatura no processo de previsão de doenças cardíacas. A técnica de ajuste de hiperparâmetros *GridSearchCV* utilizada nesse trabalho se assemelha aos trabalhos de Bhatt *et al.* (2023)

e Ahamad *et al.* (2023), no qual foram os únicos trabalhos que usaram a técnica para melhorar a performance nos modelos. Outro aspecto de semelhança desse estudo com os trabalhos de Sun e Pan (2023) e Ahamad *et al.* (2023) foi a utilização da mesma base de dados ou semelhante.

Entretanto, a maioria dos trabalhos não utiliza modelos da XAI para verificar quais as características que mais tiveram importância nas saídas dos modelos. O único trabalho que utilizou a XAI foi o dos autores Paudel *et al.* (2023) que usaram o LIME para avaliar as instâncias somente de forma local.

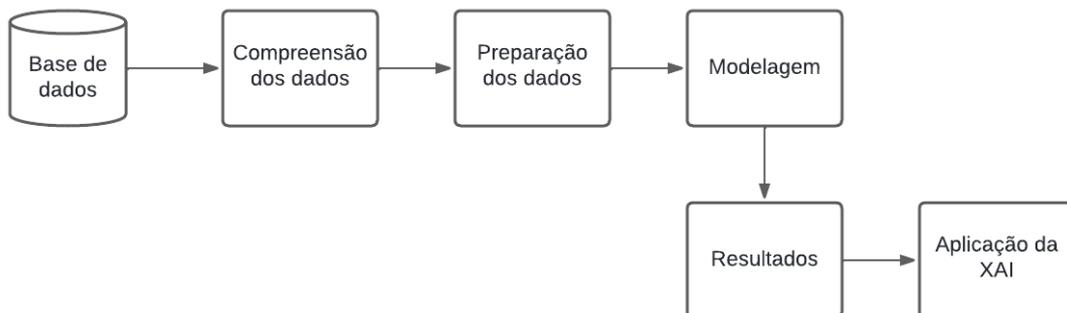
Por isso, esse trabalho utilizou o SHAP como um modelo de explicabilidade para verificar quais foram as variáveis que mais tiveram impacto nas previsões dos modelos globalmente. A Tabela 1 mostra uma comparação entre os resultados obtidos na literatura com a predição de doenças cardíacas.

3. Metodologia

A metodologia do trabalho consiste em utilizar o modelo de processo de Mineração de Dados CRISP-DM, também conhecido como Processo Padrão Inter-Indústrias para Mineração de Dados. Esse modelo é bastante empregado por pesquisadores e especialistas para construção de projetos de *data mining*. Essa metodologia é composta por seis etapas, não rígidas e cíclicas, ou seja, as etapas do modelo podem ser avançadas ou não, dependendo da necessidade do projeto (Ramos *et al.*, 2020).

As etapas no modelo CRISP-DM são as seguintes: Entendimento do negócio, compreensão dos dados, preparação dos dados, modelagem, avaliação e implantação. Nessa etapa do trabalho, as etapas do CRISP-DM que serão abordadas serão a compreensão dos dados, preparação dos dados e modelagem. A etapa de avaliação será abordada na seção de resultados do estudo. A Figura 1 mostra o fluxograma da metodologia abordada nesse trabalho.

Figura 1. Fluxograma da Metodologia



Fonte: Elaboração própria (2024)

3.1. Compreensão dos dados

3.1.1. Coleta dos dados

Os dados foram obtidos do repositório *Kaggle*, a partir da combinação de cinco bases de dados diferentes: *Cleveland, Hungarian, Switzerland, Long Beach VA e Stalog (Heart) Data Set*, disponibilizadas pelo repositório UCI (*UC Irvine Machine Learning Repository*). Fedesoriano (2021) utilizou essas bases para criar uma base com mais dados sobre doenças cardíacas. A base final obtida possui 11 características clínicas utilizadas para prever doenças cardíacas. A variável alvo a ser observada é a *Heart Disease*, que verifica se o paciente possui ou não doença cardíaca.

3.2. Descrição dos dados

Foram observados 918 padrões distribuídos em 12 colunas. Foram encontradas 5 variáveis do tipo numérica, 6 categóricas e 1 booleana. Não foram encontrados dados faltantes nem duplicados no *dataset*. A descrição das variáveis do *dataset* pode ser observada na Tabela 2.

Tabela 2. Descrição das variáveis

Variável	Descrição
Idade	Idade do paciente (anos)
Sexo	Sexo do paciente (M: Masculino, F: Feminino)
<i>ChestPainType</i>	Tipo de dor torácica (TA: Angina típica, ATA: Angina atípica, NAP: Dor não anginosa, ASY: Assintomática)
PA em repouso	Pressão arterial em repouso (mmHg)
Colesterol	Colesterol sérico (mm/dl)
BS em jejum	Glicemia em jejum (1: se BS em jejum > 120 mg/dl, 0: caso contrário)
ECG em repouso	Resultados do eletrocardiograma em repouso. Normal: Normal, ST: com anormalidade das ondas ST-T (inversões das ondas T e/ou elevação ou depressão de ST > 0,05 mV), HVE: mostrando hipertrofia ventricular esquerda provável ou definitiva pelos critérios de Estes.
<i>MaxHR</i>	Frequência cardíaca máxima alcançada (valor numérico entre 60 e 202)
<i>ExerciseAngina</i>	Angina induzida por exercício (S: Sim, N: Não)
<i>Oldpeak</i>	<i>Oldpeak</i> = ST (Valor numérico medido na depressão)
<i>ST Slope</i>	Inclinação do segmento ST do pico do exercício (<i>Up: up, Flat: flat, Down: downsloping</i>)

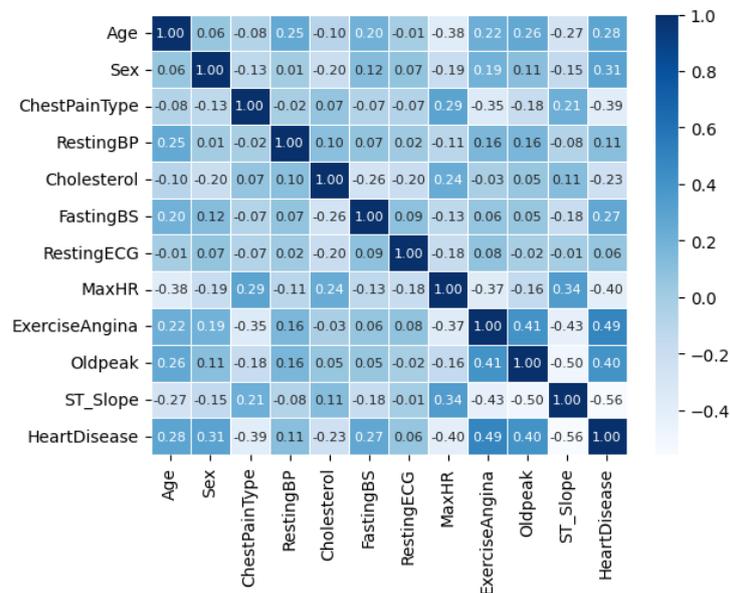
Fonte: Elaboração própria (2024)

3.2.1. Análise Exploratória de Dados

Nessa etapa foi observada a correlação entre as variáveis usando o gráfico *heatmap*. O *heatmap*, também conhecido como mapa de calor, é um gráfico utilizado para representar graficamente as correlações entre as variáveis independentes e a variável dependente. Na Figura 2, pode-se observar que quanto mais “azul escuro”, mais positivamente correlacionadas estão as variáveis e quanto mais “branco” elas forem, maior é a correlação negativa.

Houve uma correlação forte e negativa entre a variável-alvo, *Heart Disease*, e as variáveis *ST Slope*, *MaxHR* e *ChestPainType*, além da relação forte e positiva entre a variável *Exercise Angina* e a variável *Heart Disease*. Apesar da pouca diferença entre os registros nas classes alvo, foi escolhido realizar o balanceamento para avaliar o desempenho dos modelos com os dados balanceados.

Figura 2. Gráfico *Heatmap* mostrando a correlação entre as variáveis independentes e a variável dependente, *Heart Disease*



Fonte: Elaboração própria (2024)

3.3. Preparação dos Dados

3.3.1. Formatação dos dados

Nessa etapa foi realizada a transformação das variáveis categóricas *Sex*, *ChestPainType*, *RestingECG*, *Exercise Angina* e *ST Slope* em numéricas. Para isso, foi utilizada a função *LabelEncoder* do submódulo *preprocessing* da biblioteca de ML, *Scikit-learn*. Essa função tem o objetivo de codificar rótulos não numéricos (categóricos) em numéricos (Scikitlearn, 2024).

O balanceamento das classes foi feito utilizando a técnica de *oversampling*, que consiste em aumentar a quantidade de registros da classe com menor frequência, visando trazer um equilíbrio nas classes. A técnica *oversampling* utilizada foi o *Smote*, pois ela é uma forma tradicional de balancear os dados.

Essa técnica consiste na escolha de um dado aleatório e, a partir do dado escolhido, o algoritmo calcula os k-vizinhos mais próximos a esse dado. Após isso, o algoritmo seleciona aleatoriamente um vizinho do dado escolhido inicialmente. Após a seleção, é calculada a diferença entre esses dois dados e o valor obtido é multiplicado por um número aleatório entre 0 e 1.

O processo é repetido N/100 vezes. N é a quantidade de sobreamostragem necessária. O resultado dessa divisão é o valor dos vizinhos aleatórios mais próximos considerados pelo algoritmo para gerar novos dados sintéticos (Maklin, 2022).

Na etapa de divisão dos dados em treinamento e testes foi feita a divisão do conjunto de dados em 80% de dados de treinamento e 20% para o conjunto de testes. Os dados de treinamento são utilizados para treinar os modelos de classificação, enquanto os dados de teste são utilizados para medir a performance do modelo. Os dados de treinamento consistem em 734 linhas e 11 colunas, enquanto os dados de testes consistem em 184 linhas e 11 colunas.

Após a divisão do conjunto de dados, foi aplicado o *Standard Scaler* para normalizar os dados. A normalização é utilizada quando é preciso lidar com dados com diferentes dimensões. No conjunto de dados analisados, as variáveis estão em escalas diferentes, por exemplo, a variável pressão arterial em repouso é medida em milímetro de mercúrio (mmHg), enquanto a variável BS em jejum (glicemia em jejum) é medida por miligrama por decilitro (mg/dl), por isso foi utilizada a normalização para colocar as variáveis numéricas em uma escala comum.

3.4. Modelagem

Foram escolhidos cinco classificadores diferentes. A escolha deles foi feita baseada nos estudos e artigos citados no referencial teórico desse trabalho que utilizam classificadores para prever doenças cardíacas. Os modelos escolhidos foram os seguintes: *Logistic Regression*, *K-Nearest Neighbors* (KNN), *Multi-layer Perceptron* (MLP), *Extreme Gradient Boost* (XGBoost) e *Random Forest*.

Para fazer os ajustes dos hiperparâmetros dos modelos foi utilizado o *Grid Search CV*, pois ele faz uma busca em grade pelos melhores hiperparâmetros para um modelo. Após esse processo de busca, o *GridSearchCV* avalia os hiperparâmetros utilizando algum tipo de validação, no caso foi usada a validação cruzada (*k-fold* = 10).

A Tabela 3 mostra a faixa de valores usados em cada um dos parâmetros utilizados nos modelos. No modelo *Logistic Regression* foram utilizados diferentes valores para os parâmetros *C* e *max iter*. No KNN, o parâmetro utilizado foi o *n-neighbors*.

No modelo MLP, valores diversos foram testados para os parâmetros *activation*, *solver*, *alpha* e *learning rate*. O parâmetro *max iter* foi definido para 50 para diminuir o tempo de processamento feito pelo modelo neural. Para o *XGBoost*, os parâmetros utilizados foram o *max depth* e *n-estimators*. No *Random Forest*, diferentes números foram testados para os parâmetros *n-estimators*, *max depth* e *max features*.

Tabela 3. Valores utilizados nos hiperparâmetros dos modelos

Classificador	Parâmetros	Valores utilizados
<i>Logistic Regression</i>	<i>C</i>	[(-3) - 7]
	<i>max iter</i>	50,100
KNN	<i>n-neighbors</i>	[1 - 31]
MLP	<i>activation</i>	<i>identity, logistic, tanh, relu</i>
	<i>solver</i>	<i>lbfgs, sgd, adam</i>
	<i>alpha</i>	[(-5) - 10]
	<i>learning rate</i>	<i>adaptive, constant, invscaling</i>
	<i>max iter</i>	50
<i>XGBoost</i>	<i>max depth</i>	3, 5, 7, 9
	<i>n-estimators</i>	15, 50, 200
<i>Random Forest</i>	<i>n-estimators</i>	15, 50, 200
	<i>max depth</i>	3, 5, 7, 9
	<i>max features</i>	0.1, 1

Fonte: Elaboração própria (2024)

A lista abaixo mostra os melhores hiperparâmetros encontrados pelo *GridSearch*:

- *Logistic Regression*: *C*: 0.1, *max iter*: 50.
- KNN: *n-neighbors*: 11
- MLP: *activation*: *tanh*, *alpha*: 0.005994842503189409, *learning rate*: *constant*, *max iter*: 50, *solver*: *adam*.
- *XGBoost*: *max depth*: 3, *n-estimators*: 15
- *Random Forest*: *max depth*: 9, *max features*: 1, *n-estimators*: 200.

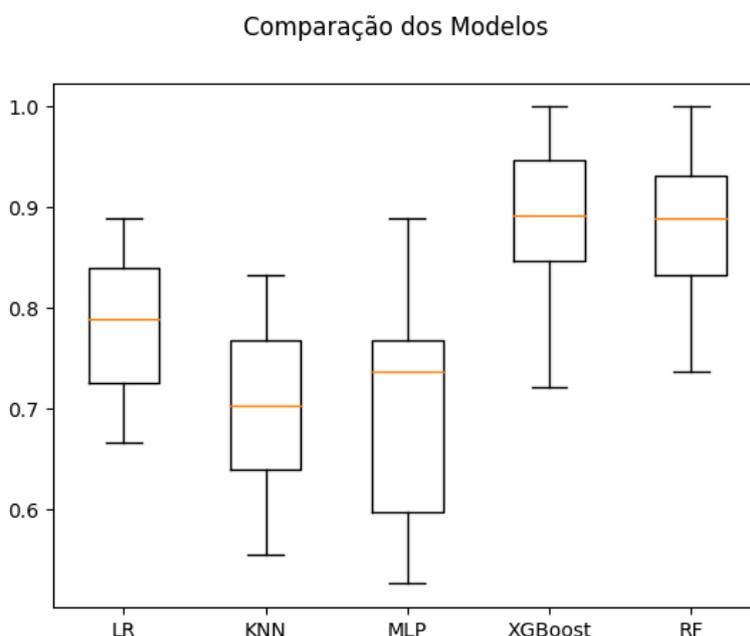
Após o processo de modelagem, foi avaliado e aplicado a XAI nos cinco modelos para entender quais foram as variáveis que mais influenciaram os classificadores ao prever doenças cardíacas.

4. Resultados obtidos

Nessa etapa são avaliados os resultados obtidos com os classificadores. As métricas escolhidas para avaliar os modelos foram a acurácia, precisão, *recall* e *f1-score*. A Figura 3 mostra a comparação entre a acurácia dos classificadores analisados no conjunto de dados de teste.

Pode-se observar que os modelos *Random Forest* e *XGBoost* obtiveram um desempenho melhor entre os modelos, obtendo uma acurácia de mais 85% com três execuções realizadas. Os modelos que apresentaram menor desempenho entre os classificadores foram o *Logistic Regression*, o KNN e o MLP.

Figura 3. Boxplot da acurácia dos modelos



Fonte: Elaboração própria (2024)

O modelo *Random Forest* obteve os melhores resultados, apresentando a maior acurácia entre os modelos avaliados com 90%. Ele também foi o modelo que errou menos nas previsões, a Tabela 4 abaixo mostra os valores obtidos com as métricas do modelo.

Tabela 4. Métricas de avaliação do *Random Forest*

Acurácia	Precisão	<i>Recall</i>	<i>F1-score</i>	Média	Desvio Padrão
90%	92,30%	89,71%	90,99%	0.8809	0.0814

Fonte: Elaboração própria (2024)

Outro modelo que obteve resultados satisfatórios entre os classificadores foi o *XGBoost*, com 87% de acurácia. A Tabela 5 abaixo mostra os resultados obtidos com as métricas.

Tabela 5. Métricas de avaliação do XGBoost

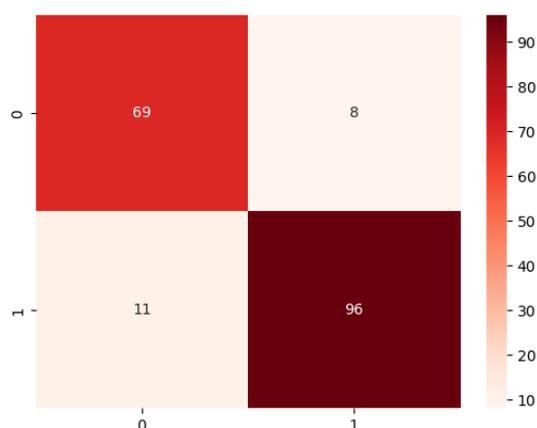
Acurácia	Precisão	Recall	F1-score	Média	Desvio Padrão
87%	91,08%	85,98%	88,46%	0.8856	0.0793

Fonte: Elaboração própria (2024)

Outra métrica utilizada para avaliar os modelos com os dados de teste foi a matriz de confusão. Essa métrica mostra como foram distribuídos os acertos e erros que o modelo teve em relação a cada classe de interesse.

No modelo *Random Forest*, a Figura 4 mostra as taxas de acertos e erros que as classes Normal (0) e Doença Cardíaca (1) obtiveram. Na Classe Doença cardíaca, o modelo conseguiu identificar bem a classe, porém errou 8 vezes ao classificá-la de forma errada como Normal (ausência da doença cardíaca). Houve 11 erros na Classe Normal, classificada incorretamente como Doença cardíaca.

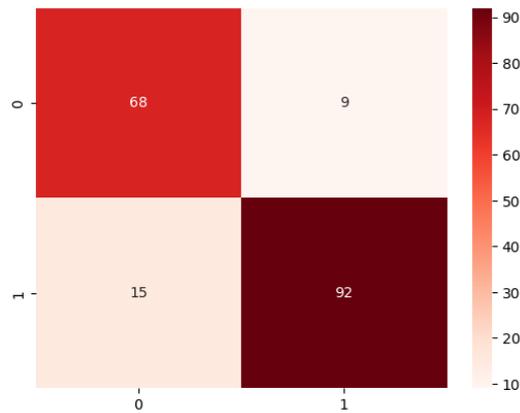
Figura 4. Matriz de Confusão do modelo *Random Forest*. As colunas da matriz representam as classes Normal (0) e Doença cardíaca (1)



Fonte: Elaboração própria (2024)

O modelo *XGBoost* teve as segundas menores taxas de erros entre os modelos. Na Figura 5, pode-se observar que na Classe Doença cardíaca, o modelo conseguiu identificar bem a classe da doença, porém errou 9 vezes ao classificar de forma errada como Normal. Na Classe Normal, o modelo fez uma boa identificação da classe, porém errou 15 vezes ao classificar incorretamente como Doença cardíaca.

Figura 5. Matriz de Confusão do modelo XGBoost



Fonte: Elaboração própria (2024)

O modelo que obteve o melhor desempenho entre os dois modelos com melhores resultados foi o *Random Forest* pelas altas taxas de precisão e *recall* que o modelo teve, além da menor taxa de erros entre os classificadores avaliados no conjunto de dados de teste.

Como esse trabalho usa dados para prever doenças cardíacas, a métrica *recall* e a baixa taxa de erros nas classes alvo foram outras métricas importantes consideradas nesse estudo, uma vez que valores altos no *recall* demonstram a capacidade do modelo de identificar corretamente a presença de uma doença.

Os resultados obtidos com os classificadores *Random Forest* e *XGBoost* foram semelhantes aos resultados obtidos na literatura. Na Tabela 1 pode-se observar que os estudos que usaram a mesma base de dados utilizada nesse trabalho obtiveram acurácias semelhantes ao modelo escolhido nesse estudo.

O estudo de Sun e Pan (2023) obteve os resultados de 88,52% na acurácia e obteve a taxa de falsos negativos de 8,82% com o *Logistic Regression* no conjunto de dados de teste. O estudo de Ahamad *et al.* (2023), utilizando o conjunto de teste, obteve acurácias de 87,91% com o modelo KNN utilizando a base *UCI Kaggle Cleveland*, além da acurácia de 99,03% com o modelo *XGBoost* usando a base *UCI Kaggle Cleveland, Hungary, Switzerland, and Long Beach V*. O classificador *Random Forest*, modelo melhor avaliado nesse trabalho, obteve resultados semelhantes aos estudos citados acima. Em relação ao conjunto de dados de teste, o modelo obteve uma acurácia de 90% e *recall* de 89,71%.

Após a avaliação das métricas dos classificadores, foi utilizado o algoritmo de IA Explicável, SHAP, o qual é um modelo global que avalia quais foram as variáveis que mais tiveram influência nas decisões tomadas pelos modelos. O gráfico *beeswarm* foi escolhido entre os gráficos de explicação que o SHAP possui por que ele “é projetado para exibir um resumo denso de informações de como os principais recursos em um conjunto de dados impactam a saída do modelo” (Lundberg, 2018). O eixo X do gráfico

representa o valor SHAP calculado em cada *feature*, enquanto o eixo Y representa a *feature*.

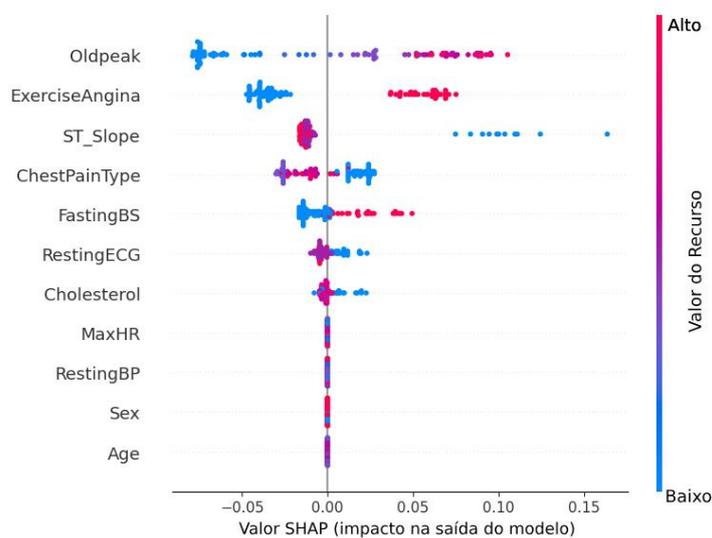
As cores nos pontos do gráfico mostram o valor que cada *feature* possui. A intensidade desse valor (baixo, intermediário e alto) pode ser observada em uma escala no lado direito do gráfico. A ordenação das características é feita por meio de um cálculo para verificar o grau de importância que elas possuem. O SHAP calcula o valor médio absoluto dos valores SHAP em cada *feature* e ordena as características de acordo com esse valor.

No modelo *Random Forest*, as variáveis *Oldpeak*, Angina induzida por exercício (*Exercise Angina*) e a Inclinação do segmento ST do pico do exercício (*St Slope*) foram as três características com maiores graus de importância.

A Figura 6 mostra que valores altos no *Oldpeak* (pontos vermelhos e roxos no lado direito do gráfico), produção maior de angina enquanto algum esforço físico for feito (Pontos vermelhos no lado direito do gráfico), além de valores mais baixos no *St Slope* indicam probabilidades maiores da presença de doença cardíaca.

As características menos relevantes apontadas pelo *Random Forest* foram a Frequência cardíaca máxima alcançada (*MaxHR*), Pressão arterial em repouso (*RestingBP*), Sexo e Idade, demonstrando assim que essas características não possuem um grau de importância relevante ao definir se um paciente tem doença cardíaca.

Figura 6. Variáveis mais importantes apontadas pelo SHAP no modelo *Random Forest* em ordem decrescente

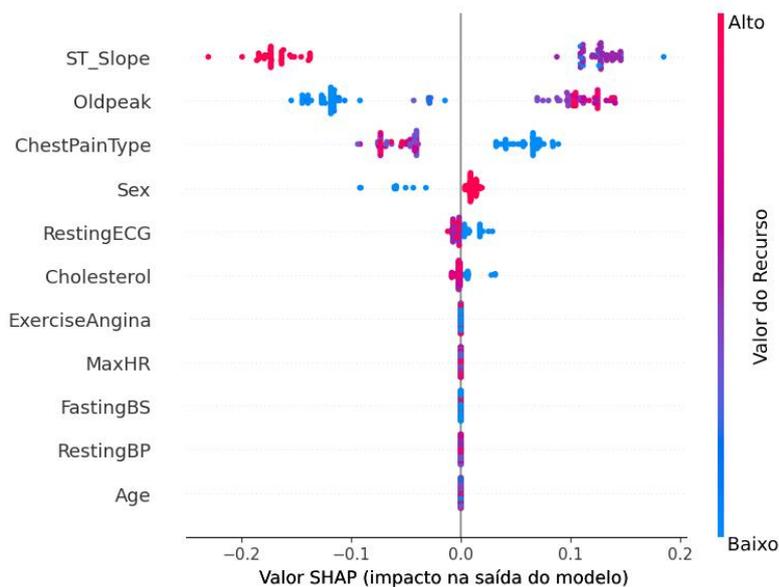


Fonte: Elaboração própria (2024)

No modelo *XGBoost*, as variáveis *ST-Slope*, *Oldpeak* e o tipo da dor no peito torácica (*ChestPainType*), respectivamente, foram os três principais fatores que mais influenciaram nas saídas do modelo. Na Figura 7, pode-se observar que valores mais

baixos no *ST-Slope*, valores mais altos no *Oldpeak* e ter um tipo de dor torácica (*ChestPainType*) são fatores importantes ao indicar a presença de doença cardíaca.

Figura 7. Variáveis mais importantes apontadas pelo SHAP no modelo *XGBoost* em ordem decrescente



Fonte: Elaboração própria (2024)

As características menos importantes apontadas pelo modelo foram a *Exercise Angina*, *MaxHR*, Glicemia em jejum (*FastingBS*), *RestingBP* e Idade, evidenciados pelos valores SHAP iguais a zero obtidos pelas características.

A variável *Exercise Angina* não foi considerada importante no *XGBoost*, assim como foi no modelo *Random Forest*, esse fator pode ser explicado pela a média dos valores SHAP obtidos pela variável. A variável teve uma média igual a zero, enquanto no modelo *Random Forest*, a variável obteve média de 0.04.

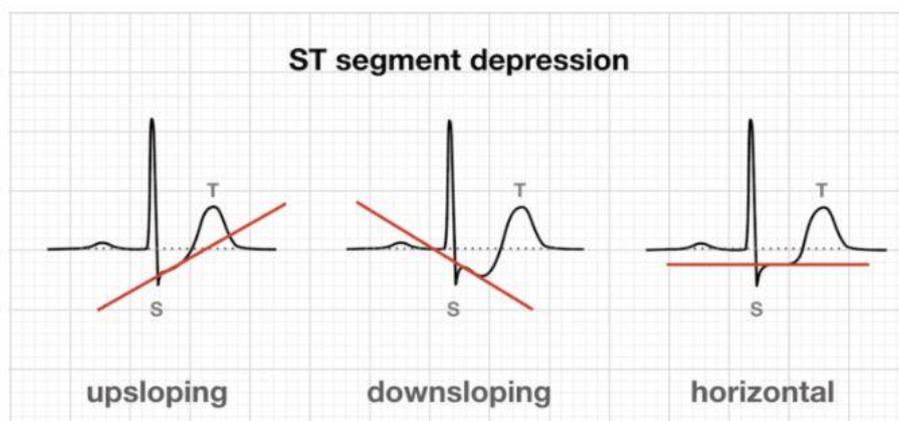
Esse resultado pode explicar por que a variável não foi considerada importante no modelo *XGBoost*. Assim como o modelo *Random Forest*, as características, frequência cardíaca máxima alcançada (*MaxHR*), Pressão arterial em repouso (*RestingBP*) e Idade foram as variáveis comuns apontadas pelos dois modelos que não tiveram um grau de importância significativa ao indicar a presença de doença cardíaca.

Ao aplicar o SHAP nos modelos pode-se observar que as variáveis que tiveram mais recorrência nos modelos foram essas: a Inclinação do segmento ST do pico do exercício (*ST Slope*) e o *Oldpeak* nos dois modelos, a Angina induzida por exercício (*Exercise Angina*) no modelo *Random Forest* e a variável Tipo de dor torácica (*ChestPainType*) no modelo *XGBoost*.

O *ST Slope* é uma variável medida a partir da análise do exame eletrocardiograma (ECG) durante um exercício físico feito por um paciente e realizado por médicos

cardiologistas. O desnível da depressão da inclinação do segmento ST do pico do exercício pode ser indício de doenças cardíacas. A Figura 8 mostra os tipos de morfologia da inclinação do segmento ST. Segundo Uchida, Moffa e Storti (2009), o desnível do segmento ST com padrões horizontal e descendente (*downsloping*) são critérios aceitos internacionalmente que indicam a presença de doenças cardíacas.

Figura 8. Tipos de morfologia na depressão da inclinação do segmento ST



Fonte: (Burns; Cadogan, 2024)

Os valores do *Oldpeak* são valores numéricos medidos na depressão do segmento ST induzida pelo exercício em relação ao repouso. A Angina induzida por exercício, outra variável indicada pela XAI nos modelos, se refere a um tipo de dor torácica causada quando há redução no transporte de oxigênio para o músculo cardíaco. Essa redução no oxigênio do sangue levado ao coração é causada quando há depósitos de gordura nas artérias, também chamada de aterosclerose. Essa condição ocorre durante esforços físicos ou quando há estresse emocional, levando o coração a trabalhar mais, necessitando de mais oxigênio (Sweis, 2024).

O tipo da dor no peito torácica, outra variável apontada pelos modelos escolhidos, também é um fator importante a ser considerado ao analisar se um paciente possui doença cardíaca, pois segundo a Sociedade Brasileira de Cardiologia, “[a] característica anginosa da dor torácica tem sido identificada como o dado com maior poder preditivo de doença coronariana aguda” (Volschan *et al.*, 2002, p. 4). Portanto, as variáveis apontadas pelos classificadores estão conforme os fatores de risco apontados pela comunidade médica para detectar a presença de doenças cardíacas em pacientes.

5. Conclusão

O objetivo desse trabalho era verificar como o uso de uma técnica de XAI pode ser empregada em classificadores ao prever doenças cardíacas. Esse trabalho mostrou como o uso da XAI foi positiva ao ser usada para criar camadas de explicação em modelos de aprendizagem de máquina tipo caixa-preta. Como o uso da técnica não é muito usado na literatura, principalmente nos casos de previsão de doenças cardíacas, o uso do SHAP se

mostrou um modelo eficaz ao apontar de forma global as características mais importantes consideradas pelos modelos ao classificar se uma pessoa possui doença cardíaca.

Os resultados mostraram que as características mais importantes apontadas pelo SHAP nos modelos com melhor desempenho, *Random Forest* e o *XGBoost*, foram o *Oldpeak*, *ST Slope*, *Exercise Angina* e *ChestPainType*. Elas foram consideradas relevantes ao indicar a presença de doenças cardíacas. Por outro lado, características como *MaxHR*, *FastingBS*, *RestingBP* e Idade não foram consideradas significativas pelos modelos nas previsões.

As variáveis *ST Slope* e *ChestPainType* apontadas pelo gráfico *heatmap* (ver Figura 2) como variáveis importantes e correlacionadas com a classe alvo *Heart Disease* também foram indicadas pelo SHAP como importantes nos modelos *Random Forest* e *XGBoost*. Como houve uma correlação entre as variáveis apontadas pelo *heatmap* e pelo SHAP, isso evidencia o quão relevantes essas variáveis são e como elas podem ser usadas ao prever doenças cardiovasculares.

As variáveis mais importantes consideradas pela XAI nos classificadores usados nesse trabalho estão conforme a medicina propõe, pois, entre as características analisadas pelos médicos ao verificar se um paciente possui ou não doença cardíaca, está a Inclinação do segmento ST do pico do exercício, medida observada nos exames de ECG e o tipo da dor no peito torácica do paciente, por exemplo.

As limitações do estudo compreendem o fato de, apesar das características apontadas pelo SHAP estarem entre os fatores que a comunidade médica aponta ao analisar se um paciente possui uma doença cardíaca, esse resultado não retira a importância que uma avaliação médica tem na construção de modelos e sistemas de saúde com o uso da Inteligência Artificial.

Como esse trabalho usou relativamente poucos modelos de aprendizagem de máquina, como trabalhos futuros, é sugerido usar mais modelos de classificação ao prever doenças cardíacas, assim como a utilização de uma base de dados com mais características clínicas, para verificar se características como o uso de cigarro e histórico familiar, por exemplo, podem ser consideradas importantes ao indicar a presença de doenças cardíacas.

Como o uso da XAI trouxe resultados positivos ao apontar as características mais relevantes ao indicar a presença de doença cardíaca, é sugerido utilizar outra técnica da XAI, como o ALE (*Accumulate Local Effects*), essa técnica, assim como o SHAP, gera explicações em modelos de aprendizagem de máquina de forma global. As contribuições que esse estudo propõe é mostrar como o uso da XAI pode ser utilizada nos modelos de aprendizagem de máquina ao prever doenças, principalmente doenças cardíacas. As explicações geradas pela XAI podem auxiliar cientistas de dados ao construir sistemas inteligentes de diagnóstico de doenças mais precisos, assim como auxiliar nas tomadas de decisão médicas.

Referências

AMERICAN HEART ASSOCIATION. **AI technology improved detection of heart disease during and after pregnancy**. 13 nov. 2023. Disponível em:

<https://newsroom.heart.org/news/ai-technology-improved-detection-of-heart-disease-during-and-after-pregnancy>. Acesso em: 24 out. 2024.

AHAMAD, Ghulab Nabi *et al.* Influence of optimal hyperparameters on the performance of machine learning algorithms for predicting heart disease. **Processes**, v. 11, n. 3, 1 mar. 2023. Disponível em: <https://www.mdpi.com/2227-9717/11/3/734>. Acesso em: 24 out. 2024.

BHATT, Chintan M. *et al.* Effective heart disease prediction using machine learning techniques. **Algorithms**, v. 16, n. 2, p. 88, 6 fev. 2023. Disponível em: <https://www.mdpi.com/1999-4893/16/2/88>. Acesso em: 24 out. 2024.

BRITISH HEART FOUNDATION. **Global Heart & Circulatory Diseases Factsheet**. [S. l.: s. n.], 2024. 12 p. Disponível em: <https://www.bhf.org.uk/-/media/files/for-professionals/research/heart-statistics/bhf-cvd-statistics-global-factsheet.pdf?rev=e61c05db17e9439a8c2e4720f6ca0a19&hash=6350DE1B2A19D939431D876311077C7B>. Acesso em: 24 out. 2024.

BURNS, Ed.; CADOGAN, Mike. **Myocardial Ischaemia**. 8 out. 2024. Disponível em: <https://litfl.com/myocardial-ischaemia-ecg-library/>. Acesso em: 24 out. 2024.

CARNEIRO, Victor Santos Carvalho; SORRENTINO, Taciano Amaral; ROCHA, Miriam Karla. **O impacto do acolhimento: Regressão logística explica os fatores externos à universidade que contribuem para reprovação em Geometria Analítica**. 2023. 20 p. TCC (Graduação) — Universidade Federal Rural do Semi-Árido, Mossoró, 2023. Disponível em: <https://repositorio.ufersa.edu.br/server/api/core/bitstreams/deac2dd4-def1-4e0d-b453-d728b2a72965/content>. Acesso em: 25 out. 2024.

CESARO, Juliana. **Avaliação de Discriminação em Aprendizagem de Máquina usando Técnicas de Interpretabilidade**. 2021. 81 p. Dissertação (Mestrado) — Universidade de São Paulo, São Paulo, 2021. Disponível em: <https://www.teses.usp.br/teses/disponiveis/3/3141/tde-31052021-114333/pt-br.php>. Acesso em: 25 out. 2024.

COSTA, Herbert da Silva. **Análise de explicabilidade de modelos de classificação aplicados à evasão na educação a distância**. 2023. 98 p. Dissertação (Mestrado) — Instituto Federal do Espírito Santo, Serra, 2023. Disponível em: https://repositorio.ifes.edu.br/bitstream/handle/123456789/3636/DISSERTAÇÃO_Análise_de_explicabilidade_de_modelos.pdf?sequence=1&isAllowed=y. Acesso em: 25 out. 2024.

DATTANI, Saloni *et al.* **Cardiovascular Diseases**. 2023. Disponível em: <https://ourworldindata.org/cardiovascular-diseases>. Acesso em: 25 out. 2024.

DESAI, Meha; SHAH, Manan. An anatomization on breast cancer detection and diagnosis employing multi-layer perceptron neural network (MLP) and Convolutional neural network (CNN). **Clinical eHealth**, v. 4, p. 1-11, 2021. Disponível em: <https://www.sciencedirect.com/science/article/pii/S2588914120300125>. Acesso em: 25 out. 2024.

DHALIWAL, Sukhpreet S.; NAHID, Abdullah-Al; ABBAS, Robert. Effective intrusion detection system using xgboost. **Information**, v. 9, n. 7, 21 jun. 2018. Disponível em: <https://www.mdpi.com/2078-2489/9/7/149>. Acesso em: 25 out. 2024.

FEDESORIANO. **Heart Failure Prediction Dataset**. 10 set. 2021. Disponível em: <https://www.kaggle.com/datasets/fedesoriano/heart-failure-prediction>. Acesso em: 25 out. 2024.

FERNANDES, Milena Seibert. **Inteligência Artificial Explicável aplicada a Aprendizado de Máquina: Um estudo para Identificar Estresse Ocupacional em Profissionais da Saúde**. 2022. 75 p. Monografia (Graduação) — Universidade Federal de Santa Catarina, Campus Araranguá, Araranguá, SC., 2022.

FERRAZ, Thomas Palmeira *et al.* Inteligência Artificial explicável para atenuar a falta de transparência e a legitimidade na moderação da Internet. **Estudos Avançados**, v. 38, n. 111, p. 381-405. Disponível em: <https://www.scielo.br/j/ea/a/KPMcWYkkqHy5ZK3zTFCBpFj/?lang=pt>. Acesso em: 25 out. 2024.

JONES, Lisa. **New AI technology for advanced heart attack prediction**. 4 set. 2019. Disponível em: <https://www.bhf.org.uk/what-we-do/news-from-the-bhf/news-archive/2019/september/new-ai-technology-for-advanced-heart-attack-prediction>. Acesso em: 25 out. 2024.

INC. **Evolução do infarto agudo do miocárdio**. Disponível em: https://inc.saude.gov.br/htm/dash_iam.html. Acesso em: 25 out. 2024.

JAGATI, Shiraz. **O problema da caixa preta da IA: desafios e soluções para um futuro transparente**. 5 maio 2023. Disponível em: <https://br.cointelegraph.com/news/ai-s-black-box-problem-challenges-and-solutions-for-a-transparent-future>. Acesso em: 25 out. 2024.

KUO, Chris. **Explain Your Model with the SHAP Values**. 13 set. 2019. Disponível em: <https://medium.com/dataman-in-ai/explain-your-model-with-the-shap-values-bc36aac4de3d>. Acesso em: 25 out. 2024.

LUNDBERG, Scott. **Beeswarm plot**. Disponível em: https://shap.readthedocs.io/en/latest/example_notebooks/api_examples/plots/beeswarm.html. Acesso em: 25 out. 2024.

LUNDBERG, Scott. **Welcome to the SHAP documentation**. Disponível em: <https://shap.readthedocs.io/en/latest/index.html>. Acesso em: 25 out. 2024.

MAHESH, Batta. Machine learning algorithms-a review. **International Journal of Science and Research (IJSR)**, v. 9, n. 1, p. 381—386, 2020.

MAKLIN, Cory. **Synthetic Minority Over-sampling Technique (SMOTE)**. 14 maio 2022. Disponível em: <https://medium.com/@corymaklin/synthetic-minority-over-sampling-technique-smote-7d419696b88c>. Acesso em: 25 out. 2024.

MOHSENI, Sina; ZAREI, Niloofar; RAGAN, Eric D. A multidisciplinary survey and framework for design and evaluation of explainable AI systems. **ACM Transactions on Interactive Intelligent Systems (TiiS)**, v. 11, n. 3-4, p. 1-45, 2021.

NHS. **Cardiovascular disease**. 22 abr. 2022. Disponível em: <https://www.nhs.uk/conditions/cardiovascular-disease/>. Acesso em: 26 out. 2024.

OPAS. **Doenças cardiovasculares**. Disponível em:

[https://www.paho.org/pt/topicos/doencas-](https://www.paho.org/pt/topicos/doencas-cardiovasculares#:~:text=A%20causa%20de%20ataques%20cardíacos)

[cardiovasculares#:~:text=A%20causa%20de%20ataques%20cardíacos](https://www.paho.org/pt/topicos/doencas-cardiovasculares#:~:text=A%20causa%20de%20ataques%20cardíacos). Acesso em: 25 out. 2024.

PAUDEL, Prakash *et al.* Unveiling Key Predictors for Early Heart Attack Detection using Machine Learning and Explainable AI Technique with LIME. *In: 10TH NSYSS 2023: 10TH INTERNATIONAL CONFERENCE ON NETWORKING, SYSTEMS AND SECURITY*, Khulna Bangladesh. **10th NSysS 2023: 10th International Conference on Networking, Systems and Security**. New York, NY, USA: ACM, 2023. Disponível em: <https://doi.org/10.1145/3629188.3629193>. Acesso em: 26 out. 2024.

PESSANHA, Cíntia. **Random Forest**: como funciona um dos algoritmos mais populares de ML. 20 nov. 2019. Disponível em:

[https://medium.com/cinthiabpessanha/random-forest-como-funciona-um-dos-](https://medium.com/cinthiabpessanha/random-forest-como-funciona-um-dos-algoritmos-mais-populares-de-ml-cc1b8a58b3b4)
[algoritmos-mais-populares-de-ml-cc1b8a58b3b4](https://medium.com/cinthiabpessanha/random-forest-como-funciona-um-dos-algoritmos-mais-populares-de-ml-cc1b8a58b3b4). Acesso em: 26 out. 2024.

RAMOS, Jorge Luís Cavalcanti *et al.* CRISP-EDM: uma proposta de adaptação do Modelo CRISP-DM para mineração de dados educacionais. *In: SIMPÓSIO BRASILEIRO DE INFORMÁTICA NA EDUCAÇÃO*, Brasil. **Simpósio Brasileiro de Informática na Educação**. [S. l.]: Sociedade Brasileira de Computação, 2020.

Disponível em: <https://doi.org/10.5753/cbie.sbie.2020.1092>. Acesso em: 26 out. 2024.

SAHU, Rekha *et al.* Prediction of Heart Attack and Death: Comparison Between 1 DCNN and Conventional ML Approaches. *In: 2023 1ST INTERNATIONAL CONFERENCE ON CIRCUITS, POWER AND INTELLIGENT SYSTEMS (CCPIS)*, 2023, Bhubaneswar, India. **2023 1st International Conference on Circuits, Power and Intelligent Systems (CCPIS)**. [S. l.]: IEEE, 2023. Disponível em:

<https://doi.org/10.1109/ccpis59145.2023.10291179>. Acesso em: 26 out. 2024.

SBC. **Cardiômetro - Morte por doenças cardiovasculares no Brasil**. Disponível em: <http://www.cardiometro.com.br/>. Acesso em: 26 out. 2024.

SCHIAFFINO, Matheus Costa. **Desenvolvimento de um método para classificação de comportamentos de ratos wistar utilizando o algoritmo de aprendizado supervisionado florestas aleatórias (random forests)**. 2020. 81 p. Dissertação (Mestrado) — Universidade Federal do Rio Grande do Sul, Porto Alegre, 2020.

SCIKITLEARN. **Label Encoder**. Disponível em: [https://scikit-](https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.LabelEncoder.html)

[learn.org/stable/modules/generated/sklearn.preprocessing.LabelEncoder.html](https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.LabelEncoder.html). Acesso em: 26 out. 2024.

SCIKITLEARN. **Random Forest Classifier**. Disponível em: [https://scikit-](https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier)

[learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier](https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier). Acesso em: 26 out. 2024.

SILVA, Larissa Camila Ferreira da. **Modelo de Transferência de Aprendizagem baseado em Regressão Linear Regularizada**. 2017. 30 p. Monografia (Graduação) — UNIVERSIDADE FEDERAL DE PERNAMBUCO, RECIFE, 2017.

SILVA, Vitória de Oliveira. **Detecção de fraudes na utilização de cartões usando a técnica de regressão logística**: uma aplicação com dados desbalanceados. 2022. 61 p.

Monografia (Graduação) — Universidade Estadual Paulista (Unesp), Presidente Prudente, 2022.

SUN, Huating; PAN, Jianan. Heart Disease Prediction Using Machine Learning Algorithms with Self-Measurable Physical Condition Indicators. **Journal of Data Analysis and Information Processing**, v. 11, n. 1, p. 1-10, 2023. Disponível em: <https://www.scirp.org/journal/paperinformation?paperid=122494>. Acesso em: 26 out. 2024.

SWEIS, Ranya N.; JIVAN, Arif. **Angina**. Fev. 2024. Disponível em: <https://www.msmanuals.com/pt/casa/distúrbios-do-coração-e-dos-vasos-sanguíneos/doença-arterial-coronariana/angina?ruleredirectid=762>. Acesso em: 26 out. 2024.

TORRES, Helainy. **Uma proposta do algoritmo KNN sobre uma perspectiva riemanniana para o problema de classificação de imagens**. 2022. 54 p. Dissertação (Mestrado) — Universidade Federal Rural do Rio de Janeiro, Seropédica, 2022.

UCHIDA, Augusto Hiroshi; MOFFA, Paulo; STORTI, Fernanda Coutinho. Aspectos básicos da ergometria contemporânea. **Revista de Medicina**, v. 88, n. 1, p. 1, 6 mar. 2009. Disponível em: <https://doi.org/10.11606/issn.1679-9836.v88i1p1-6>. Acesso em: 26 out. 2024.

VOLSCHAN, André *et al.* I diretriz de dor torácica na sala de emergência. **Arquivos Brasileiros de Cardiologia**, v. 79, p. 1-22, Aug 2002. Disponível em: <https://doi.org/10.1590/S0066-782X2002001700001>. Acesso em: 26 out. 2024.

WHO. **The top 10 causes of death**. 7 ago. 2024. Disponível em: <https://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death>. Acesso em: 26 out. 2024.

ZHOU, Zhi-Hua. **Machine learning**. Singapore: Springer Singapore, 2021. 459 p. *E-book*. ISBN 9789811519666. Disponível em: <https://books.google.com.br/books?id=ctM-EAAAQBAJ>. Acesso em: 27 out. 2024.