

TargetEqualizer: Uma Abordagem para Lidar com o Desequilíbrio de Classes na Previsão de Distúrbios Cardiovasculares através da Classificação de Batimentos Cardíacos com Aprendizado de Máquina

Pedro Henrique do Nascimento Gomes¹ Erika Carlos Medeiros^{1*} Patrícia Cristina Moser¹
Jorge Cavalcanti Barbosa Fonsêca¹ Rômulo César Dias de Andrade¹
Fernando Ferreira de Carvalho^{1,2} Fernando Pontual de Souza Leão Junior¹

1. Universidade de Pernambuco, Caruaru, PE, Brasil

2. Cesar School, Recife, PE, Brasil

* E-mail do autor correspondente: erika.medeiros@upe.br

Resumo

Este artigo introduz o TargetEqualizer, uma solução para lidar com o desequilíbrio de classes em conjuntos de dados de doenças cardiovasculares, mantendo a especificidade das previsões. Utilizou-se o método em modelos de aprendizado de máquina para prever doenças cardíacas com dados de sons cardíacos capturados por estetoscópios, seguindo o *framework Cross-Industry Standard Process for Data Mining* em 585 arquivos de áudio distribuídos em cinco classes. Inicialmente, os sons cardíacos foram convertidos em espectrogramas. O TargetEqualizer, uma combinação de *oversampling* e *undersampling*, equilibrou a distribuição dos dados, aprimorando a acurácia dos modelos. Foram treinados os modelos de *Random Forest*, *Logistic Regression* e *Support Vector Machine* com o conjunto de dados balanceado, avaliando-os com acurácia, precisão, *recall* e *F1-Score*, demonstrando a eficácia da abordagem. Os resultados indicam alta acurácia, com *Random Forest* alcançando 90,47% e *Logistic Regression* e *Support Vector Machine* atingindo 87,30%. Isso compete efetivamente com estudos semelhantes e mantém a especificidade das previsões, ao contrário da prática comum de agrupar doenças cardíacas em duas classes, garantindo diagnósticos mais precisos.

Palavras-chave: Classificação de Batimentos Cardíacos, Distúrbios Cardiovasculares, Espectrograma, Modelos de Aprendizado de Máquina, Balanceamento de Classes.

DOI: 10.7176/RHSS/13-19-02

Publication date: December 31st 2023

1. Introdução

Doenças Cardiovasculares (DCV) representam uma crise global de saúde, sendo responsáveis por uma proporção alarmante de mortes em todo o mundo, correspondendo a aproximadamente 31% de todas as fatalidades no ano de 2020 (WHO, 2021). Essa categoria abrangente inclui uma variedade de condições, desde doença arterial coronariana até insuficiência cardíaca e acidente vascular cerebral. Essas condições frequentemente se manifestam como alterações estruturais ou funcionais no coração, desencadeando uma série de sintomas, como dor no peito, falta de ar e fadiga debilitante (Ren et al., 2022).

O diagnóstico precoce e preciso das DCV é fundamental para o tratamento eficaz e a prevenção de complicações graves, desempenhando um papel crucial na preservação da qualidade e longevidade das vidas dos pacientes. No entanto, diagnosticar DCV apresenta um desafio imponente e multifacetado que transcende as fronteiras da ciência médica. Este desafio é principalmente atribuído à natureza não específica de seus sintomas, que podem se entrelaçar de forma intrincada e sutil com os de várias outras condições médicas, muitas vezes camuflando a verdadeira patologia cardiovascular subjacente (Kim et al., 2022). Essa complexidade diagnóstica destaca a necessidade premente de abordagens diagnósticas inovadoras e altamente precisas, capazes de identificar os indicadores sutis, mas cruciais, das DCV no intrincado perfil médico dos pacientes. A ausculta com estetoscópio, uma técnica diagnóstica consagrada e não invasiva, está entre os métodos mais comuns utilizados para detectar DCV. Embora inestimável, a ausculta com estetoscópio possui subjetividade inerente e é suscetível a erros humanos, que podem influenciar a precisão do diagnóstico (Bao et al., 2022).

A prática da ausculta cardíaca com estetoscópio por profissionais de saúde há muito tempo é um pilar do diagnóstico de DCV. No entanto, esse método tradicional não está isento de limitações, principalmente decorrentes da subjetividade envolvida na interpretação dos sons cardíacos. A subjetividade na ausculta com estetoscópio surge de vários fatores. Em primeiro lugar, depende muito da experiência, treinamento e familiaridade individual do praticante com os sons cardíacos. Diferentes clínicos podem interpretar os mesmos sons de maneira diferente, levando a variações no diagnóstico. Além disso, fatores ambientais, como ruído ambiente em um ambiente clínico, podem interferir na ausculta precisa, tornando desafiador discernir nuances sutis nos sons cardíacos (Medeiros et al., 2023).

Entra em cena a Inteligência Artificial (IA), que promete mitigar esses desafios associados à subjetividade

na ausculta cardíaca. Algoritmos de IA podem ser treinados em vastos conjuntos de dados de sons cardíacos, abrangendo uma variedade de condições cardíacas e variações acústicas. Esses algoritmos possuem uma capacidade incomparável de discernir até mesmo os padrões e anomalias mais sutis dentro desses sons, padrões que podem escapar ao ouvido humano ou estar sujeitos à interpretação individual (Singh & Kumar, 2020).

A IA pode introduzir objetividade e consistência ao processo de diagnóstico, fornecendo avaliações padronizadas e baseadas em evidências dos sons cardíacos. Isso não apenas reduz o potencial de erro humano, mas também oferece aos profissionais de saúde uma camada adicional de confiança em seus diagnósticos. Além disso, sistemas de suporte à decisão impulsionados por IA podem fornecer feedback em tempo real aos clínicos durante a ausculta, alertando-os para possíveis anormalidades ou nuances que merecem investigação adicional. Essa combinação da expertise do clínico com insights impulsionados por IA aprimora a precisão geral do diagnóstico e garante que achados críticos não sejam negligenciados (Medeiros et al., 2023).

Neste estudo, foram utilizados dados coletados de um grupo diversificado de indivíduos por meio do aplicativo *iStethoscope Pro* para *iPhone* e conduziu-se um ensaio clínico em ambientes hospitalares, incorporando o uso do estetoscópio digital *DigiScope* (Jindal et al., 2021).

O objetivo geral deste estudo é desenvolver e explorar uma abordagem de balanceamento para uma base de dados multiclasse desbalanceada, a fim de otimizar o desempenho dos modelos de aprendizado de máquina (Kavitha et al., 2021) na previsão de doenças cardíacas com base em dados de ausculta de coração.

Para alcançar esse objetivo, estabeleceram-se os seguintes objetivos específicos:

1. Realizar o pré-processamento dos dados de ausculta cardíaca, que inclui a conversão dos dados de áudio em espectrogramas, bem como a normalização de valores, visando assegurar a qualidade dos dados de entrada para as etapas de modelagem;
2. Desenvolver e implementar uma estratégia de balanceamento multiclasse que leve em consideração a disparidade na distribuição das classes de doenças cardíacas na base de dados, com o intuito de otimizar o desempenho dos modelos de aprendizado de máquina na tarefa de prever essas doenças;
3. Testar a abordagem desenvolvida em modelos de aprendizado de máquina para previsão de doenças cardíacas com base em dados de ausculta cardíaca;
4. Avaliar e comparar o desempenho dos modelos utilizando métricas específicas, tais como acurácia, precisão, *recall* e *F1-Score*, com ênfase na capacidade dos modelos de efetuar previsões precisas e reduzir a confusão nas previsões das distintas classes de doenças cardíacas.

Além da introdução, este estudo é composto por quatro seções principais. A segunda seção, "Trabalhos Relacionados", compara artigos que fundamentem este trabalho. A terceira, "Metodologia", apresenta o framework e a metodologia detalhada utilizados neste estudo, incluindo compreensão dos dados, preparação dos dados, modelagem do *Random Forest* (RF) (Rigatti, 2017), do *Logistic Regression* (LR) (Das, 2021) e do *Support Vector Machine* (SVM) (Pisner & Schnyer, 2020), e procedimentos de treinamento, e finalmente na metodologia, a seção de Avaliação. A quarta seção, "Resultados Obtidos", analisa e interpreta os resultados obtidos a partir dos modelos de aprendizado de máquina. A quinta seção, "Conclusões", oferece um resumo das descobertas da pesquisa, limitações do estudo e propõe direções futuras para aprimorar o desempenho dos modelos. Por fim, o estudo inclui uma seção de "Referências".

2. Trabalhos Relacionados

Nesta seção de trabalhos relacionados, investigamos pesquisas e estudos que abordaram a previsão de doenças cardíacas por meio de diversas técnicas de aprendizado de máquina e mineração de dados. Esses trabalhos forneceram perspectivas valiosas para o desenvolvimento de modelos de previsão precisos e eficientes, com o objetivo de aprimorar o diagnóstico precoce e a gestão de condições relacionadas ao coração. Ao analisar esses trabalhos correlatos, podemos obter uma compreensão abrangente dos avanços nesse domínio e identificar as forças e limitações de diferentes abordagens para a previsão de doenças cardíacas.

O estudo realizado por Singh e Kumar (2020) ressalta que as doenças cardíacas se tornaram uma causa proeminente de morte em todo o mundo nos últimos anos. O diagnóstico preciso e oportuno tornou-se uma necessidade crítica para um gerenciamento eficaz. A mineração de dados emergiu como uma ferramenta potente na área da saúde, com pesquisadores utilizando várias técnicas de mineração de dados e aprendizado de máquina para analisar dados médicos complexos visando a previsão de doenças cardíacas. A pesquisa se concentra nos atributos das doenças cardíacas e introduz um modelo que emprega algoritmos de aprendizado supervisionado, incluindo SVM (precisão de 83%), *Decision Tree* (DT) (precisão de 79%) (Kotsiantis, 2013), LR (precisão de 78%) e *K-Nearest Neighbors* (KNN) (precisão de 87%) (Kramer, 2013). O conjunto de dados inclui 303 instâncias do banco de dados de Cleveland, contendo 76 atributos, mas apenas 14 foram selecionados para avaliar o desempenho dos algoritmos. O principal objetivo é estimar a probabilidade de desenvolvimento de doenças cardíacas, com KNN emergindo como o algoritmo com melhor desempenho nessa tarefa preditiva. No entanto, a abordagem de Singh e Kumar (2020) se concentra principalmente na análise de doenças cardíacas, categorizando pacientes em dois grupos amplos: aqueles com doenças cardíacas e aqueles sem. A abordagem

dos autores não considera a especificidade de várias doenças cardíacas, agrupando-as nessas duas categorias gerais. Essa simplificação excessiva limita a profundidade da análise para diferentes condições cardíacas, o que é vital para diagnósticos precisos e estratégias de tratamento específicas. Além disso, o estudo de Singh e Kumar (2020) se beneficia de um conjunto de dados com distribuições de classes equilibradas, facilitando a aplicação de algoritmos de aprendizado de máquina. Em contraste, muitos cenários do mundo real envolvem conjuntos de dados desequilibrados, onde certas condições cardíacas são raras em comparação com outras. Abordar esse desafio requer técnicas especializadas para garantir uma classificação equitativa e precisa de todas as condições cardíacas. Assim, há espaço para pesquisas que explorem mais a fundo as complexidades da classificação de doenças cardíacas, além do escopo do trabalho de Singh e Kumar (2020).

Jindal et al. (2021) apresentam que a pesquisa se concentra principalmente na previsão de doenças cardíacas com base em diversos atributos médicos. Eles pioneiramente desenvolveram um sistema de previsão de doenças cardíacas empregando algoritmos de aprendizado de máquina, como LR, KNN e RF. O modelo proposto apresenta uma precisão de 87,5%. A abordagem de Jindal et al. (2021) classifica principalmente os pacientes em duas categorias amplas: aqueles com doenças cardíacas e aqueles sem, sem especificar o tipo ou categoria específica de doença cardíaca. No entanto, uma lacuna em seu estudo é a ausência de uma explicação detalhada sobre como o balanceamento de classes foi abordado. Essa omissão dificultou a reprodutibilidade de seus experimentos, pois o desequilíbrio na distribuição de classes pode impactar significativamente o desempenho dos modelos de aprendizado de máquina.

Kavitha et al. (2021) introduziram uma abordagem de aprendizado de máquina voltada para prever doenças cardíacas utilizando o conjunto de dados de doenças cardíacas de Cleveland. Esta pesquisa utilizou técnicas de mineração de dados, incorporando regressão e classificação, e envolveu RF, DT e um modelo híbrido que combina ambos os algoritmos. O modelo híbrido alcançou uma taxa de precisão de 88,7% na previsão de doenças cardíacas. Além disso, eles projetaram uma interface amigável para previsões personalizadas. Kavitha et al. (2021) trabalharam com um conjunto de dados equilibrado com apenas duas classes: pacientes com doenças cardíacas e pacientes sem. Embora suas conquistas sejam significativas e contribuam para o campo de pesquisa em geral, a ausência de balanceamento de classes em sua abordagem levanta questões sobre a aplicabilidade do modelo em cenários onde o conjunto de dados exibe um desequilíbrio significativo entre as classes.

TR et al. (2022) destacam a afirmação da Organização Mundial da Saúde (OMS) de que aproximadamente 12 milhões de mortes são causadas por doenças cardíacas em todo o mundo. Essas doenças têm ramificações significativas para a indústria médica, especialmente em países com recursos limitados e acesso inadequado aos cuidados de saúde. A previsão precisa e oportuna de doenças cardíacas, embora desafiadora, receberam um impulso promissor dos avanços em técnicas de mineração de dados e aprendizado de máquina. Em seu estudo, uma variedade de algoritmos de aprendizado de máquina e de aprendizado profundo (DL) (LeCun et al., 2015) são aplicados a um conjunto de dados com 303 linhas e 14 atributos. Os modelos incluem SVM, DT, KNN, RF, LR e *Naive Bayes* (Webb, 2016), todos voltados para aprimorar a previsão de doenças cardíacas e reduzir as fatalidades associadas às condições cardiovasculares. O modelo SVM alcança uma precisão de 83,25%, enquanto o modelo DT atinge uma precisão de 83,89%. Além disso, o modelo KNN registra uma precisão de 86,45%, o modelo RF demonstra uma precisão de 88,35%, o modelo LR apresenta uma precisão de 84,22% e o modelo *Naive Bayes* com uma pontuação de precisão de 84,69%. Uma lacuna em seu estudo é a falta de detalhes específicos sobre as estratégias de balanceamento de classes. Essa lacuna surge porque o conjunto de dados deles tende a ser mais equilibrado quando considerando apenas duas classes: pacientes com doenças cardíacas e pacientes saudáveis. No entanto, essa omissão pode potencialmente prejudicar a aplicabilidade dos modelos propostos em contextos onde existe um desequilíbrio significativo de classes, um desafio que nosso trabalho aborda diretamente.

A revisão da literatura destaca que as doenças cardíacas constituem um desafio substancial para a saúde global, enfatizando a necessidade crítica de previsões precisas e oportunas para facilitar intervenções médicas eficazes. Embora investigações anteriores tenham explorado várias abordagens, incluindo o uso de modelos de aprendizado de máquina, como frequentemente são empregados nesta área, existem distinções ao estudo atual. Todos os estudos são baseados em conjuntos de dados com apenas duas classes, o que limita a especificidade das doenças em suas previsões.

3. Metodologia

Neste estudo, foi adotado o *framework Cross-Industry Standard Process for Data Mining* (CRISP-DM) (Schröer et al., 2021) para automatizar a classificação das gravações de sons cardíacos. A metodologia CRISP-DM abrange cinco fases principais: Compreensão do Negócio, Compreensão dos Dados, Preparação dos Dados, Modelagem e Avaliação. A fase de Compreensão do Negócio envolve o estudo do contexto, por meio da revisão de trabalhos relacionados, identificação de lacunas nas soluções já propostas, entendimento do contexto do problema e da relevância de uma solução. Isso foi feito na seção de introdução e fundamentação teórica.

a) Compreensão dos Dados:

O conjunto de dados de sons cardíacos incluiu 585 arquivos de áudio, sendo que cada arquivo foi rotulado em uma das cinco categorias distintas: *Normal*, *Murmur*, *Extrahls*, *Extrastole* e *Artifact*. Essas categorias encapsulam características específicas dos sons cardíacos, ocasionalmente acompanhadas por nuances de ruídos de fundo. A pesquisa começou com uma análise exploratória dos dados, com o objetivo de entender a distribuição do conjunto de dados, abordar desequilíbrios nas classes e identificar padrões subjacentes.

Um batimento cardíaco normal (*Normal*), mostrado na Figura 1.a, é caracterizado por um padrão de forma de onda consistente e rítmico. Começa com um pico acentuado e regular para cima, indicando o sinal elétrico que desencadeia a contração das câmaras do coração. Após esse pico, há uma breve linha reta representando uma pausa momentânea à medida que as câmaras do coração se enchem de sangue. Em seguida, ocorre outro pico acentuado, significando o sinal elétrico que faz com que os ventrículos se contraíam. Esse padrão regular se repete, criando uma forma de onda constante e uniformemente espaçada (Ullah et al., 2021).

Por outro lado, um sopro (*Murmur*), mostrado na Figura 1.b, apresenta uma forma de onda irregular. Ele inclui oscilações adicionais, frequentemente turbulentas, dentro dos picos regulares. Essas irregularidades na forma de onda refletem o fluxo sanguíneo anormal, resultando no som de sopro ou ruído turbulento geralmente associado a sopros. Essas irregularidades podem ocorrer entre os picos regulares ou sobrepor-los, dependendo das características específicas do sopro (Almanifi et al., 2022).

Um som cardíaco extra (*Extrahls*), mostrado na Figura 1.c, introduz um pico adicional inesperado na forma de onda. Esse pico adicional representa um som anormal originado no coração. Ele pode interromper o padrão normal de um batimento cardíaco (Dandu & Achanti, 2022).

A extra-sístole (*Extrastole*), mostrada na Figura 1.d, apresenta irregularidades na forma de onda que indicam batimentos cardíacos irregulares com batimentos extras ou pulos. Essas irregularidades são representadas por picos adicionais ou interrupções no padrão geralmente regular da forma de onda de um batimento cardíaco normal. Elas podem ocorrer em intervalos variados, refletindo a natureza irregular das extra-sístoles (Germanova et al., 2021). Finalmente, a classe de artefato (*Artifact*), mostrada na Figura 1.e, significa ruídos ambientais e não se refere a um tipo de batimento cardíaco.

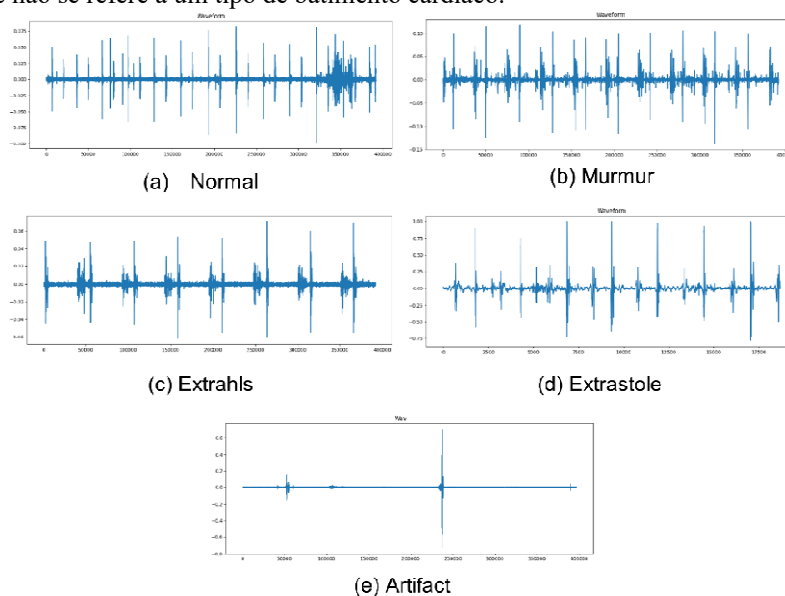


Figura 1. Formas de onda de batimentos cardíacos e classe de artefato

A distribuição das classes é a seguinte: a classe *Normal* contém 351 amostras, a classe *Murmur* consiste em 129 amostras, a classe *Extrastole* contém 46 amostras, a classe *Artifact* engloba 40 amostras e a classe *Extrahls* compreende 19 amostras. A Figura 2 mostra o número de amostras desbalanceadas.

Essa distribuição apresenta um desequilíbrio significativo entre as classes, o que pode introduzir viés nos modelos de aprendizado de máquina, favorecendo especialmente a classe majoritária *Normal*, que possui consideravelmente mais amostras. A razão principal para esse desequilíbrio é que condições cardíacas específicas, como murmúrios, extrassístoles e sons cardíacos extras, são relativamente menos comuns em comparação com a classe *Normal*. Para mitigar esse problema, pesquisadores em trabalhos relacionados optaram por agrupar todas as doenças cardíacas em uma única classe (doenças), sacrificando assim informações específicas sobre doenças e diagnósticos precisos.

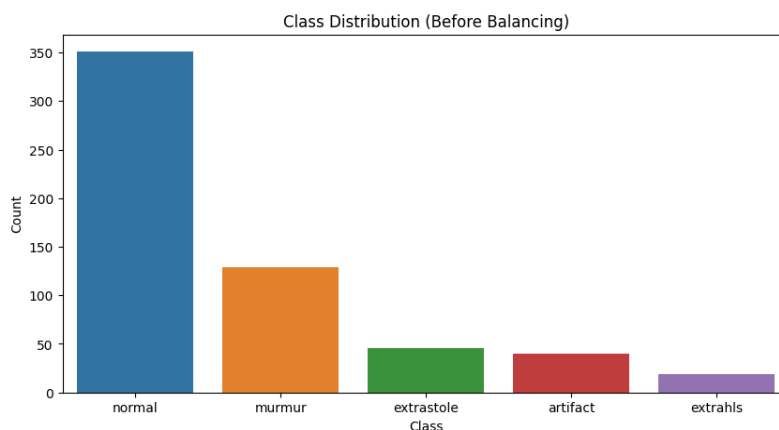


Figura 2. Classes Desbalanceadas

Na próxima seção, será introduzido o TargetEqualizer, uma abordagem de balanceamento de classes específica proposta neste estudo, projetada para enfrentar esses desafios. Esta abordagem combina técnicas de *oversampling* e *undersampling* para garantir uma distribuição equilibrada de amostras entre todas as classes, possibilitando previsões mais precisas e específicas para as doenças.

b) Preparação de Dados:

Usou-se espectrogramas para representar os áudios em representações visuais e conseqüentemente em valores numéricos. Para transformar os dados de áudio em espectrogramas, foi empregado um processo que captura a essência do som em uma representação visual. Essa transformação envolveu a conversão da forma de onda de áudio em um espectrograma mel, que mapeia essencialmente a intensidade de diferentes frequências ao longo do tempo. Esse espectrograma fornece uma visão detalhada de como o som se desenvolve, tornando-o adequado para análises mais aprofundadas (Ustubioglu et al., 2023). Os passos envolvidos incluíram a geração do espectrograma inicial, garantindo que tivesse um comprimento consistente, e então anexando-o junto com seu rótulo correspondente às listas respectivas. Esse processo permitiu trabalhar com dados de espectrograma, que são mais adequados para algoritmos de aprendizado de máquina reconhecerem e classificarem diferentes padrões de sons cardíacos.

Explorou-se a composição e o balanceamento dos dados de treinamento de sons cardíacos. É evidente a partir dessa distribuição, mostrada na Figura 2, que a classe *Normal* possui o maior número de amostras. No entanto, as outras classes tinham significativamente menos amostras, o que exigiu um manuseio cuidadoso para garantir o desempenho do modelo. Diferentemente da abordagem adotada no estudo de Medeiros et al. (2023), onde apenas SMOTE foi aplicado para equilibrar todas as classes, neste trabalho adotou-se uma abordagem aqui proposta, chamada TargetEqualizer. A Figura 3 mostra as etapas da abordagem TargetEqualizer realizadas no conjunto de treinamento. O método TargetEqualizer é uma abordagem projetada para abordar efetivamente o desequilíbrio de classes em conjuntos de dados, especialmente aqueles com múltiplas classes que representam doenças cardíacas específicas. O método consiste nas seguintes etapas detalhadas:

1. Calcular o Número de Amostras da Classe Majoritária ($N_{majority}$): Inicia-se identificando o número de amostras na classe majoritária, que geralmente tem um número significativamente maior de amostras do que as outras.
2. Calcular o Número Total de Amostras de 60% das Classes Minoritárias ($N_{minority}$): Determinar o número total de amostras dentro do subconjunto de 60% das classes minoritárias que representam condições cardíacas específicas e menos comuns. Nesse cenário com cinco classes, isso corresponde a três classes.
3. Calcular o Número Médio de Amostras nas Classes Minoritárias de 60% ($Mean_N_{minority}$): Para obter uma distribuição equilibrada de amostras entre as classes minoritárias, deve-se calcular a média aritmética do número de amostras nessas classes minoritárias, identificadas na etapa 3. Isso é alcançado somando-se o número de amostras em cada uma das classes minoritárias, identificado na etapa 3, e dividindo o total pelo número de classes. Isso garante que nenhuma condição cardíaca específica seja super representada ou sub-representada.
4. Determinar o Número Alvo de Amostras para todas as classes ($TARGET$): Calcular o número alvo de amostras que cada classe deve ter para criar um conjunto de dados equilibrado. Para fazer isso, deve-se subtrair o $Mean_N_{minority}$ (calculado na etapa 3) do número total de amostras na classe majoritária ($N_{majority}$). Essa diferença representa amostras excedentes, que são então divididas por 2. O valor de $TARGET$ resultante é sempre menor que $N_{majority}$ e maior que o número de amostras nas classes minoritárias de 60%.
5. Uma vez determinado o $TARGET$, o método aplica *RandomOverSampler* a todas as classes com um

número de amostras menor que *TARGET* para que tenham o número de amostras igual a *TARGET* e *RandomUnderSampler* a classes com um número de amostras maior que *TARGET*, para que tenham o número de amostras igual a *TARGET*. Essa abordagem abrangente de balanceamento garante que todas as classes sejam melhor representadas, contribuindo para melhorar a precisão dos modelos de aprendizado de máquina.

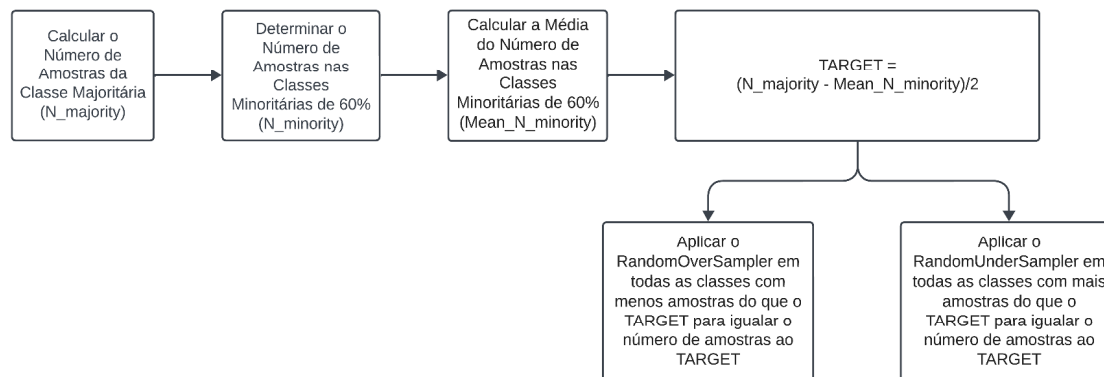


Figura 3. Etapas da Abordagem TargetEqualizer

Os benefícios de ter o *TARGET* menor que $N_{majority}$ e maior que o número de amostras nas 60% das classes minoritárias são duplos:

- Evita a predominância da classe majoritária: Ao garantir que o *TARGET* seja menor que $N_{majority}$, o método impede que a classe majoritária domine o conjunto de dados, mitigando preconceitos que podem afetar o desempenho dos modelos de aprendizagem de máquina. Uma vez determinado o *TARGET*, o método prossegue aplicando *RandomOverSampler* a todas as classes com um número de amostras menor que o *TARGET* e *RandomUnderSampler* a classes com um número de amostras maior que o *TARGET*. Essa abordagem abrangente de balanceamento assegura que todas as classes do conjunto de treino sejam adequadamente representadas, contribuindo para a melhoria da precisão dos modelos de aprendizagem de máquina.
- Aprimora a representação da classe minoritária: Ter o *TARGET* maior que o número de amostras nas 60% das classes minoritárias garante uma distribuição mais equitativa de amostras entre essas condições cardíacas específicas. Essa representação equilibrada permite que os modelos aprendam e façam previsões para doenças cardíacas raras de forma eficaz.

Para lidar com o desequilíbrio de classes no conjunto de dados, iniciou-se calculando o número alvo de amostras para cada classe na base de treino. Começando-se com a classe *Normal*, que tinha $N_{majority} = 281$ dados. Em seguida, observou-se que a média das amostras das classes de 60% com menos amostras (*Extrastole* tem 37 amostras, *Artifact* tem 32 amostras e *Extrahls* tem 15 amostras) é $N_{minority} = 84$. Em seguida, calculou-se o Número Médio de Amostras nas Classes Minoritárias de 60%, $Mean_N_minority = N_{minority}/3 = 1/3 = 28$. Para estabelecer o número alvo de amostras para todas as classes, subtraíu-se $Mean_N_minority$ (28) do número total ($N_{majority}$) de amostras na classe *Normal* ($N_{majority} - Mean_N_minority = 281 - 28$), o que resultou em 253. Para garantir uma representação equitativa entre todas as classes, dividimos esse número por 2, resultando no número alvo de 126 amostras por classe. Esse cálculo orientou a abordagem de preparação de dados, permitindo equilibrar eficazmente a distribuição de classes por meio de uma combinação de técnicas de *oversampling* e *undersampling*. Especificamente, utilizou-se o *RandomOverSampler* para as classes minoritárias (*Murmur*, *Extrahls*, *Extrastole* e *Artifact*) nos dados de treinamento, seguindo a meta de 126 amostras por classe. Posteriormente, foi aplicado o *RandomUnderSampler* à classe majoritária (*Normal*) nos dados de treinamento, para atingir o mesmo número alvo de amostras de 126.

Essa abordagem garantiu uma representação equilibrada entre todas as classes, facilitando o treinamento do modelo e melhorando o desempenho na tarefa de previsão de doenças cardíacas. Esse processo de preparação de dados visou abordar o desequilíbrio de classes, mantendo a integridade e representatividade do conjunto de dados, aprimorando, assim, a robustez de modelos preditivos. A Figura 4 mostra as classes já balanceadas, contendo 126 amostras.

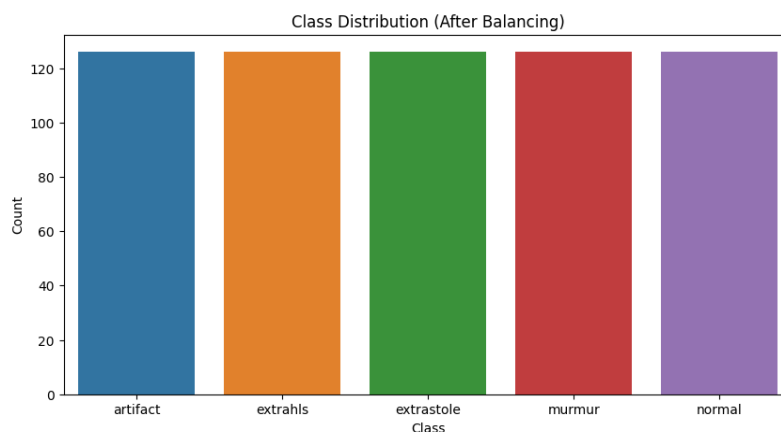


Figura 4. Classes Balanceadas

c) Modelagem:

Utilizou-se um conjunto de dados de treino balanceado com estratificação para garantir uma representação justa das diferentes classes de sons cardíacos. Essa abordagem é crucial para uma avaliação precisa do modelo e para a análise de desempenho. Em seguida, realizou-se o treinamento e a predição usando RF, LR e SVM com seus hiper parâmetros configurados com os valores padrão na biblioteca *scikit-learn* (Pedregosa et al., 2011).

d) Avaliação:

Os critérios de avaliação utilizados para avaliação dos modelos de aprendizagem de máquina acurácia, precisão, *recall* e o *F1-Score*. Além disso, foi utilizado matrizes de confusão para fornecer uma análise detalhada do desempenho de classificação dos modelos em diferentes patologias cardíacas, além de curvas *Receiver Operating Characteristic* (ROC) (Hanley, 2014), juntamente com o cálculo da *Area Under the Curve* (AUC) (Walter, 2005).

Na próxima seção, serão apresentados os resultados obtidos por esses modelos, discutindo as métricas de desempenho alcançadas no domínio da classificação de sons cardíacos.

4. Resultados Obtidos

Nesta seção, apresenta-se os resultados relacionados ao desempenho dos modelos de aprendizado de máquina, RF, LR e SVM, utilizados. O conjunto de dados de 585 amostras foi dividido em dados de treinamento e teste, na proporção de 80/20. Após a aplicação do TargetEqualizer aos dados de treinamento, os modelos foram treinados com dados equilibrados. Avalia-se métricas-chave como precisão, *recall* e *F1-Score*, além de matrizes de confusão e curvas ROC para obter o desempenho dos modelos em várias patologias cardíacas.

Na Tabela 1, o modelo RF apresentou resultados na classificação dos diferentes tipos de sons cardíacos. Demonstrou uma alta precisão na detecção da classe *Artifact*, bem como precisão e *recall* nas classes *Extrahls* e *Extrastole*, indicando uma grande capacidade de identificação dessas classes. No entanto, enfrentou desafios nas classes *Murmur* e *Normal*, com pontuações de precisão, *recall* e *F1-Score* ligeiramente inferiores. A acurácia global do modelo foi de 0,9047, refletindo um excelente desempenho global, com métricas médias em torno de 0,90 para precisão, *recall* e *F1-Score*, indicando um desempenho consistente e sólido na maioria das classes.

Tabela 1. Métricas do *Random Forest*

	Precisão	Recall	F1-Score	Suporte
Artifact	1,00	1,00	1,00	25
Extrahls	0,89	1,00	0,94	25
Extrastole	0,96	1,00	0,98	25
Murmur	0,86	0,73	0,79	26
Normal	0,80	0,80	0,80	25
Acurácia	0,9047			126
Macro Avg	0,90	0,91	0,90	126
Weighted Avg	0,90	0,90	0,90	126

A Figura 5 mostra a matriz de confusão do modelo RF. Nas classes *Artifact*, *Extrahls* e *Extrastole*, o modelo teve bons desempenhos, com 25 instâncias corretamente classificadas em cada uma delas, sem errar nenhuma amostra. Entretanto, na classe *Murmur*, houve duas instâncias classificadas incorretamente como falsos negativos para classe *Extrahls* e cinco como falsos positivos para classe *Normal*, indicando uma certa dificuldade em distinguir esses casos. Na classe *Normal*, o modelo classificou incorretamente uma instância como falso negativo para classe *Extrahls*, três como falsos negativos para classe *Murmur* e uma instância como falso negativo

para classe *Extrastole*.

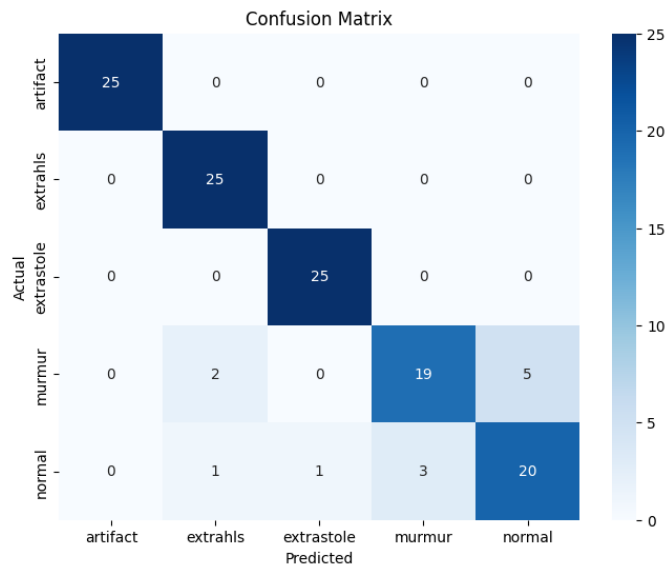


Figura 5. Matriz de Confusão do *Random Forest*

A Figura 6 demonstra as curvas ROC para o modelo RF, e observa-se que as classes *Artifact*, *Extrahls* e *Extrastole* atingiram uma pontuação AUC de 1,00, indicando a capacidade de discriminação para essas classes. Isso mostra que o modelo é capaz de distinguir entre as instâncias positivas e negativas para essas classes. Para as classes *Murmur* e *Normal*, as pontuações AUC de 0,97 e 0,98, respectivamente, indicam uma performance muito boa, sugerindo que o modelo tem uma alta capacidade de distinguir entre instâncias de murmúrios, normais e instâncias negativas.

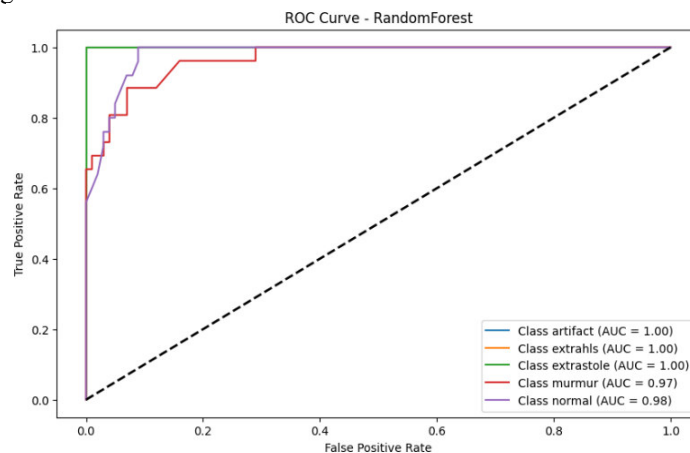


Figura 6. Curva ROC do *Random Forest*

A Tabela 2 apresenta as métricas do modelo LR. O modelo demonstra um desempenho sólido em várias classes de sons cardíacos. Ele alcança uma precisão perfeita (1,00) nas classes *Artifact*, mostrando sua habilidade para classificar corretamente todos os casos dessa classe. Nas classes *Extrahls* e *Extrastole*, o modelo exibe uma precisão de 0,89 e 0,86, respectivamente, indicando sua capacidade de identificar a maioria dos casos com alta precisão. Entretanto, nas classes *Murmur* e *Normal*, o modelo enfrenta desafios, com uma precisão de 0,74 e 0,88, respectivamente. Apesar disso, ele demonstra um *recall* razoável em todas as classes, variando de 0,60 a 1,00. O *F1-Score* geral do modelo é consistente, oscilando em torno de 0,87, indicando um desempenho equilibrado. A acurácia global é de 87,30%.

Tabela 2. Métricas de *Logistic Regression*

	Precisão	Recall	F1-Score	Suporte
Artifact	1,00	1,00	1,00	25
Extrahls	0,89	1,00	0,94	25
Extrastole	0,86	1,00	0,93	25
Murmur	0,74	0,77	0,75	26
Normal	0,88	0,60	0,71	25
Acurácia			0,8730	126
Macro Avg	0,88	0,87	0,87	126
Weighted Avg	0,87	0,87	0,87	126

A Figura 7 mostra a matriz de confusão do modelo LR. Na matriz de confusão, o modelo classificou corretamente os sons cardíacos de *Artifact*, *Extrahls* e *Extrastole* com alta precisão, conforme indicado pelos elementos diagonais correspondentes a essas classes. No entanto, a classe *Normal* está sendo erroneamente classificada como *Murmur*, enquanto a classe *Murmur* está sendo confundida com as outras classes.

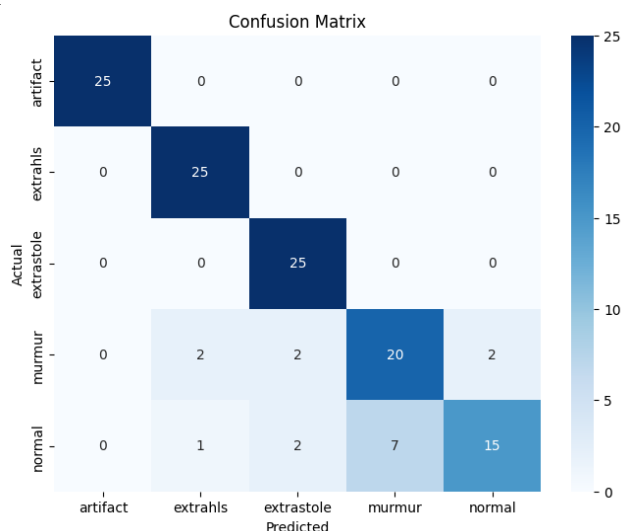


Figure 7. Matriz de confusão de *Logistic Regression*

A Figura 8 representa a curva ROC do modelo LR. As curvas ROC para o modelo indicam que ele possui capacidade de distinguir as classes *Artifact*, *Extrahls* e *Extrastole*, como evidenciado pelas pontuações AUC de 1,00, 0,99 e 0,99, respectivamente. Isso sugere uma alta taxa de verdadeiros positivos e uma baixa taxa de falsos positivos para essas classes. No entanto, nas classes *Murmur* e *Normal*, o modelo apresenta pontuações AUC um pouco mais baixas, indicando uma capacidade um pouco inferior de distinguir entre as instâncias positivas e negativas para essas classes. Apesar disso, as pontuações AUC ainda estão relativamente altas (0,93), sugerindo que o modelo tem um excelente desempenho em todas as classes.

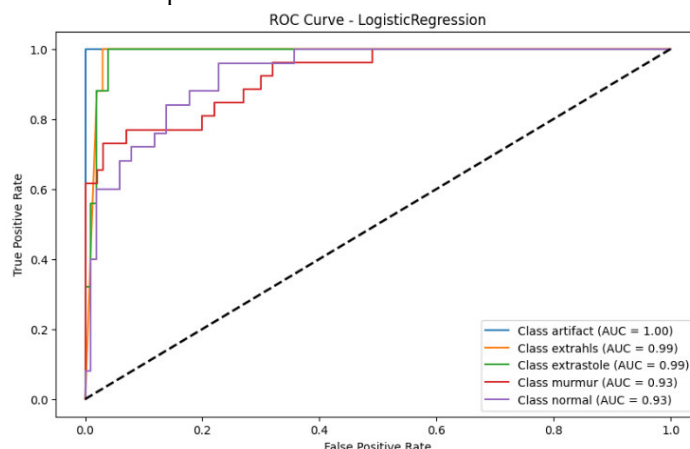


Figure 8. Curva ROC do *Logistic Regression*

A Tabela 3 faz referência às métricas do SVM. O modelo de SVM exibe um desempenho geral robusto, com precisões de 0,96, 0,93 e 0,78 nas classes de *Artifact*, *Extrahls* e *Extrastole*, respectivamente. Isso sugere uma capacidade sólida de identificar corretamente a maioria dos casos nessas classes. No entanto, na classe de

Murmur, o modelo enfrenta um desafio, com uma precisão de 0,83, indicando uma certa dificuldade em classificar com precisão esses casos. Na classe *Normal*, a precisão é de 0,89, mostrando ainda alto desempenho. A Figura 9 representa a matriz de confusão do SVM. Quanto à matriz de confusão, o modelo classifica corretamente todas as instâncias das classes de *Artifact*, *Extrahls* e *Extrastole*, conforme indicado pelos valores diagonais. No entanto, ocorrem confusões nas classes de *Murmur* e *Normal*. Duas instâncias de *Murmur* são classificadas incorretamente como *Extrahls*, três como *Extrastole* e duas como *Normal*. Além disso, quatro instâncias de som cardíaco *Normal* são erroneamente classificadas como *Extrahls* e quatro como *Murmur*.

Tabela 3. Métricas do *Support Vector Machine*

	Precisão	Recall	F1-Score	Suporte
Artifact	0,96	1,00	0,98	25
Extrahls	0,93	1,00	0,96	25
Extrastole	0,78	1,00	0,88	25
Murmur	0,83	0,73	0,78	26
Normal	0,89	0,64	0,74	25
Acurácia				0,8730
Macro Avg	0,88	0,87	0,87	126
Weighted Avg	0,88	0,87	0,87	126

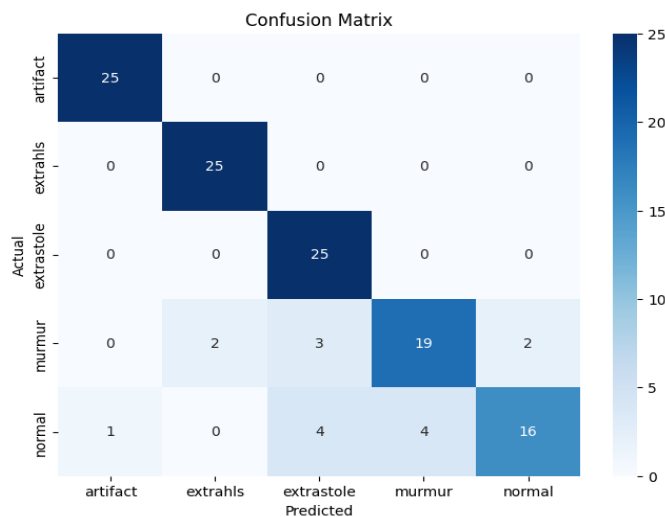


Figure 9. Matriz de Confusão do *Support Vector Machine*

A Figura 10 representa as curvas ROC do modelo SVM. As pontuações AUC para o modelo SVM indicam um excelente desempenho na diferenciação das classes *Artifact* e *Extrahls*, com AUC de 1,00 para ambas. Na classe *Extrahls*, o modelo também apresenta uma performance muito boa, com AUC de 0,99. Para as classes *Murmur* e *Normal*, embora as pontuações AUC sejam ligeiramente mais baixas (0,95 e 0,96, respectivamente), ainda estão em níveis bastante elevados, indicando uma capacidade sólida de discriminação entre as instâncias positivas e negativas dessas classes.

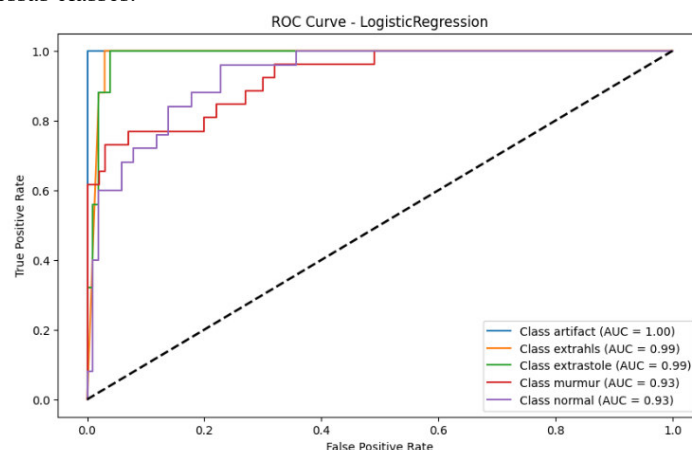


Figure 10. Curva ROC do *Support Vector Machine*

Em resumo, os modelos demonstraram desempenho sólido na classificação dos dados, com o RF alcançando a maior precisão, com uma acurácia de aproximadamente 90,5%. O LR e o SVM apresentaram resultados ligeiramente inferiores, ambas com uma acurácia de cerca de 87,3%. Estes resultados indicam que todos os modelos têm uma boa capacidade de fazer previsões precisas, com RF se destacando como o mais preciso entre os três. Além disso, os três modelos exibem desempenho competitivo na detecção de distúrbios cardíacos usando sons de estetoscópio, com métricas de precisão, *recall* e *F1-Score* indicando sua capacidade de fazer previsões precisas sobre patologias cardíacas. Isto representa um avanço no campo da IA para o diagnóstico de doenças cardíacas por meio da análise de áudio de estetoscópio. Esses modelos têm o potencial de impactar significativamente o domínio médico, oferecendo ferramentas eficientes e confiáveis para a detecção precoce, melhorando assim o cuidado ao paciente e a gestão global de condições relacionadas ao coração.

A análise estatística foi conduzida para três populações, cada uma composta por 30 amostras emparelhadas. O nível de significância familiar dos testes foi estabelecido em $p=0,050$. A hipótese nula (H_0) significa que todos os modelos se igualam no desempenho. Fazendo o teste de Friedman (Pereira et al., 2015) de normalidade das populações dos modelos, foi rejeitada a H_0 , uma vez que para as populações RF ($p=0,000$), LR ($p=0,003$) e SVM ($p=0,000$) apresentam $p \leq 0,05$, levando-nos a assumir que nem todas as populações são normais, ou seja, os modelos possuem diferenças estatísticas significativas em seus desempenhos.

O teste de Nemenyi pós-hoc (Herbold, 2020) foi empregado para inferir quais diferenças são significativas. Diferenças entre as populações foram consideradas significativas se a diferença na classificação média superasse a distância crítica $CD=0,605$ do teste de Nemenyi. Com base no teste de Nemenyi pós-hoc, inferimos a ausência de diferenças significativas dentro dos seguintes grupos: RF e LR; LR e SVM. Todas as outras diferenças foram consideradas significativas, como mostra a Figura 11.

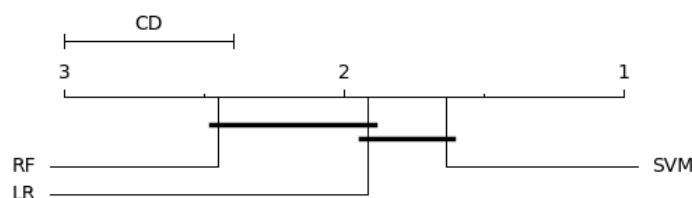


Figura 11. Testes estatísticos dos modelos

5. Conclusão

Neste estudo, usou-se dados de ausculta cardíaca, que incluiu a conversão dos dados de áudio em espectrogramas, bem como a normalização de valores, visando assegurar a qualidade dos dados de entrada para as etapas de modelagem, resolvendo o objetivo específico número um.

Manter a especificidade da doença apresenta desafios em termos de desequilíbrio de classes, pois algumas condições cardíacas são menos comuns do que outras. Para enfrentar esse problema, introduziu-se, neste estudo, a abordagem TargetEqualizer. Se destaca ao manter a especificidade de várias doenças cardíacas dentro do conjunto de dados. Esse recurso fundamental garante que os modelos de aprendizagem de máquina sejam treinados com dados que representam com precisão todo o espectro de condições cardíacas, aumentando assim a precisão diagnóstica.

O TargetEqualizer demonstra habilidade ao lidar com o desequilíbrio de classes, adotando uma abordagem equilibrada que evita tanto o *oversampling* excessivo, que poderia resultar na geração de muitas amostras sintéticas quando a classe majoritária possui um número significativamente maior de amostras, quanto o *undersampling* excessivo, que poderia levar à perda significativa de dados da classe majoritária. Essa estratégia equitativa visa preservar a representatividade de ambas as classes no conjunto de dados. Ao definir um número alvo de amostras intencionalmente menor do que a classe majoritária, ele mitiga eficazmente os vieses que, de outra forma, poderiam afetar o desempenho do modelo de aprendizagem de máquina. Essa abordagem garante uma representação equilibrada entre todas as classes, promovendo um treinamento de modelo equitativo e imparcial.

Além disso, a exigência da abordagem, com o número alvo de amostras sendo maior do que o número de amostras nas classes minoritárias, é fundamental para garantir uma distribuição mais justa de amostras entre condições cardíacas específicas, embora menos comuns. Essa representação equilibrada capacita os modelos de aprendizagem de máquina a aprender e fazer previsões para doenças cardíacas raras de forma eficaz, um aspecto crucial para um diagnóstico preciso.

Uma das contribuições mais significativas do TargetEqualizer é sua capacidade de melhorar a aplicabilidade de modelos de aprendizagem de máquina em cenários de saúde do mundo real. O desequilíbrio de classes é frequentemente encontrado em configurações clínicas, e o diagnóstico preciso de condições cardíacas específicas é imperativo. A abordagem do TargetEqualizer enfrenta esse desafio ao permitir o treinamento de

modelos que mantêm a especificidade da doença, ao mesmo tempo em que gerencia efetivamente o desequilíbrio de classes. Essa combinação aprimora a praticidade dos modelos, tornando-os altamente relevantes e aplicáveis em contextos de saúde onde o diagnóstico preciso e a especificidade são fundamentais. Uma limitação potencial é que o TargetEqualizer envolve tanto oversampling quanto undersampling, o que pode aumentar a complexidade computacional, especialmente em conjuntos de dados muito grandes.

O desenvolvimento do TargetEqualizer representa o desenvolvimento e implementação de uma estratégia de balanceamento multiclasse que leve em consideração a disparidade na distribuição das classes de doenças cardíacas na base de dados, com o intuito de otimizar o desempenho dos modelos de aprendizado de máquina na tarefa de prever essas doenças, o que resolve o objetivo específico número dois.

Para cumprir os objetivos específicos três e quatro, avaliou-se a aplicação do TargetEqualizer em modelos de aprendizagem de máquina com o objetivo de diagnosticar doenças cardíacas com alta precisão. Os modelos utilizados foram o RF, LR e SVM e foram treinados e testados em um conjunto de dados de sons cardíacos, onde cada amostra foi rotulada em uma das cinco categorias: *Normal*, *Murmur*, *Extrahls*, *Extrastole* e *Artifact*. O modelo RF, demonstrou uma acurácia de 90,5%, com LR e SVM não muito atrás, com 87,3%. Essas acurácias mostram o desempenho competitivo dos modelos na classificação precisa de várias condições cardíacas. Com ressalva que a pesquisa mantém a granularidade das doenças cardíacas, tornando os modelos mais aplicáveis a cenários do mundo real. Isso contrasta com a tendência comum observada em estudos relacionados, onde conjuntos de dados tendem a fundir várias condições cardíacas em duas classes (doença vs. saudável), sacrificando a especificidade da doença. Salienta-se que a acurácia obtida pelo RF de 90,5% foi maior que as acurácias apresentadas nos modelos de aprendizagem de máquina testados nos trabalhos relacionados.

Apesar dos resultados promissores, o trabalho tem limitações. Uma das principais limitações deste estudo é o tamanho relativamente pequeno do conjunto de dados. Embora equilibrado cuidadosamente usando a abordagem do TargetEqualizer, o tamanho do conjunto de dados ainda pode impactar a capacidade dos modelos de generalizar para uma população mais ampla. Para aumentar a robustez dos modelos, a aquisição de conjuntos de dados mais extensos deve ser considerada em futuros esforços de pesquisa. Outra limitação é a dependência de um único conjunto de dados. Embora tomadas medidas para equilibrar e diversificar o conjunto de dados, usar dados de várias fontes ou instituições médicas poderia melhorar ainda mais a representatividade dos modelos, levando em conta variações potenciais na coleta de dados e nas características demográficas dos pacientes. Os modelos foram treinados e avaliados em uma população específica, o que pode limitar sua generalização para diferentes grupos demográficos ou regiões geográficas. Futuros trabalhos devem explorar a adaptabilidade desses modelos em várias populações para garantir sua utilidade em configurações de saúde diversas. Embora os modelos de aprendizagem de máquina incluindo RF, LR e SVM, ofereçam desempenho competitivo, frequentemente carecem de interpretabilidade. Incorporar modelos interpretáveis no processo de diagnóstico poderia ajudar os profissionais de saúde a entender por que determinadas previsões são feitas, promovendo confiança nos diagnósticos assistidos por IA. Além disso, a avaliação atual dos modelos depende do processamento offline de dados. Para tornar esses modelos práticos em configurações clínicas, aplicativos em tempo real devem ser desenvolvidos, abordando desafios relacionados à baixa latência e integração perfeita com os sistemas de saúde existentes.

Essas limitações representam áreas onde futuras pesquisas podem concentrar seus esforços para refinar e expandir a aplicação de modelos de aprendizagem de máquina no diagnóstico de doenças cardíacas, contribuindo assim para aprimorar os resultados de saúde. No âmbito do avanço na detecção de distúrbios cardíacos por meio de aprendizagem de máquina, várias promissoras áreas de pesquisa emergem. Em primeiro lugar, a aquisição de conjuntos de dados mais amplos e diversificados, com um foco específico em incluir uma ampla variedade de condições cardíacas, é imperativa. Esses conjuntos de dados serviriam como uma base robusta para treinar modelos altamente precisos e adaptáveis. Além disso, explorar o potencial de arquiteturas avançadas de DL, como CNN e Redes Neurais Recorrentes (RNN) (Medsker & Jain, 1999), é a chave para liberar todo o poder do DL no campo da análise de sons cardíacos. Sistemas de diagnóstico em tempo real, capazes de fornecer avaliações rápidas e precisas de condições cardíacas durante exames de pacientes, representam outra fronteira a ser explorada, podendo potencialmente revolucionar práticas clínicas. Estudos de validação clínica, conduzidos em configurações de saúde reais e envolvendo a integração desses modelos nos sistemas de prontuários eletrônicos, forneceriam insights inestimáveis sobre seu desempenho no mundo real.

Propõe-se também explorar duas promissoras abordagens para aprimorar a eficácia do TargetEqualizer. Em primeiro lugar, investigar um método automatizado de seleção de valor alvo. Essa abordagem automatizada poderá utilizar análise estatística ou algoritmos de aprendizagem de máquina para determinar o valor alvo ótimo (*TARGET*) para o balanceamento de classes. Ao analisar as características do conjunto de dados, esse método pode ajustar de modo adaptável o valor alvo, garantindo alinhamento com as necessidades específicas do conjunto de dados e, assim, aumentando a precisão do modelo. Em segundo lugar, sugere-se explorar o ajuste dinâmico do alvo durante o treinamento do modelo. Essa abordagem dinâmica envolve ajustar continuamente o valor alvo à medida que o modelo aprende. Ao avaliar o progresso do aprendizado, é possível identificar

instâncias que são mais desafiadoras de classificar e atribuir a elas uma prioridade maior. Essa estratégia adaptativa garante que o modelo se concentre mais nos casos difíceis, o que pode levar a uma maior precisão e robustez à medida que o treinamento avança. Essas duas direções de pesquisa futura visam tornar o TargetEqualizer ainda mais versátil e eficaz ao enfrentar desafios de desequilíbrio de classes em aplicações de aprendizagem de máquina.

Em resumo, o estudo destaca o potencial da aprendizagem de máquina para auxiliar o diagnóstico cardíaco e destaca a eficácia de um conjunto de dados equilibrado no aprimoramento do desempenho do modelo. Essas descobertas abrem caminhos para futuras pesquisas no desenvolvimento de ferramentas impulsionadas por IA para a detecção precoce e o gerenciamento de doenças cardiovasculares, melhorando assim o cuidado ao paciente e os resultados no campo da cardiologia.

Referências

- Almanifi, O. R. A., Ab Nasir, A. F., Mohd Razman, M. A., Musa, R. M., & P.P. Abdul Majeed, A. (2022). Heartbeat murmurs detection in phonocardiogram recordings via transfer learning. *Alexandria Engineering Journal*, 61(12), 10995–11002. <https://doi.org/10.1016/j.aej.2022.04.031>
- Bao, X., Xu, Y., & Kamavuako, E. N. (2022). The Effect of Signal Duration on the Classification of Heart Sounds: A Deep Learning Approach. *Sensors*, 22(6), 2261. <https://doi.org/10.3390/s22062261>
- Dandu, N., & Achanti, N. (2022). Early Diagnosis of Heart Disease Via Google’s Auto ML. *Journal of Student Research*, 11(4). <https://doi.org/10.47611/jsrhs.v11i4.3240>
- Das, A. (2021). Logistic Regression. Em F. Maggino (Org.), *Encyclopedia of Quality of Life and Well-Being Research* (p. 1–2). Springer International Publishing. https://doi.org/10.1007/978-3-319-69909-7_1689-2
- Germanova, O. A., Germanov, V. A., Shchukin, Yu. V., Germanov, A. V., & Piskunov, M. V. (2021). Extrasystoles: Adverse effects of the first postextrasystolic contraction. *Bulletin of the Medical Institute “REAVIZ” (REHABILITATION, DOCTOR AND HEALTH)*, 6, 89–97. <https://doi.org/10.20340/vmi-rvz.2020.6.11>
- Hanley, J. A. (2014). Receiver operating characteristic (ROC) curves. *Wiley StatsRef: Statistics Reference Online*.
- Herbold, S. (2020). Autorank: A Python package for automated ranking of classifiers. *Journal of Open Source Software*, 5(48), 2173. <https://doi.org/10.21105/joss.02173>
- Jindal, H., Agrawal, S., Khera, R., Jain, R., & Nagrath, P. (2021). Heart disease prediction using machine learning algorithms. *IOP Conference Series: Materials Science and Engineering*, 1022(1), 012072. <https://doi.org/10.1088/1757-899X/1022/1/012072>
- Kavitha, M., Gnaneswar, G., Dinesh, R., Sai, Y. R., & Suraj, R. S. (2021). Heart Disease Prediction using Hybrid machine Learning Model. *2021 6th International Conference on Inventive Computation Technologies (ICICT)*, 1329–1333. <https://doi.org/10.1109/ICICT50816.2021.9358597>
- Kim, Y., Hyon, Y., Lee, S., Woo, S.-D., Ha, T., & Chung, C. (2022). The coming era of a new auscultation system for analyzing respiratory sounds. *BMC Pulmonary Medicine*, 22(1), 119. <https://doi.org/10.1186/s12890-022-01896-1>
- Kotsiantis, S. B. (2013). Decision trees: A recent overview. *Artificial Intelligence Review*, 39(4), 261–283. <https://doi.org/10.1007/s10462-011-9272-4>
- Kramer, O. (2013). K-Nearest Neighbors. Em *Dimensionality Reduction with Unsupervised Nearest Neighbors* (p. 13–23). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-38652-7_2
- LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436–444. <https://doi.org/10.1038/nature14539>
- Medeiros, E. C., Moser, P. C., Fonsêca, J. C. B., Andrade, R. C. D. de, & Carvalho, F. F. de. (2023). Predicting Cardiovascular Disorders Through Stethoscope Audio Using Convolutional Neural Network. *Research on Humanities and Social Sciences*, 13(14), 11.
- Medsker, L., & Jain, L. C. (1999). *Recurrent neural networks: Design and applications*. CRC press.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, É. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12(85), 2825–2830.
- Pereira, D. G., Afonso, A., & Medeiros, F. M. (2015). Overview of Friedman’s Test and Post-hoc Analysis. *Communications in Statistics - Simulation and Computation*, 44(10), 2636–2653. <https://doi.org/10.1080/03610918.2014.931971>
- Pisner, D. A., & Schnyer, D. M. (2020). Chapter 6—Support vector machine. Em A. Mechelli & S. Vieira (Orgs.), *Machine Learning* (p. 101–121). Academic Press. <https://doi.org/10.1016/B978-0-12-815739-8.00006-7>
- Ren, Z., Qian, K., Dong, F., Dai, Z., Nejdil, W., Yamamoto, Y., & Schuller, B. W. (2022). Deep attention-based

- neural networks for explainable heart sound classification. *Machine Learning with Applications*, 9, 100322. <https://doi.org/10.1016/j.mlwa.2022.100322>
- Rigatti, S. J. (2017). Random Forest. *Journal of Insurance Medicine*, 47(1), 31–39. <https://doi.org/10.17849/in-sm-47-01-31-39.1>
- Schröer, C., Kruse, F., & Gómez, J. M. (2021). A Systematic Literature Review on Applying CRISP-DM Process Model. *Procedia Computer Science*, 181, 526–534. <https://doi.org/10.1016/j.procs.2021.01.199>
- Singh, A., & Kumar, R. (2020). Heart Disease Prediction Using Machine Learning Algorithms. 2020 International Conference on Electrical and Electronics Engineering (ICE3), 452–457. <https://doi.org/10.1109/ICE348803.2020.9122958>
- Ullah, W., Siddique, I., Zulqarnain, R. M., Alam, M. M., Ahmad, I., & Raza, U. A. (2021). Classification of Arrhythmia in Heartbeat Detection Using Deep Learning. *Computational Intelligence and Neuroscience*, 2021, 1–13. <https://doi.org/10.1155/2021/2195922>
- Ustubioglu, A., Ustubioglu, B., & Ulutas, G. (2023). Mel spectrogram-based audio forgery detection using CNN. *Signal, Image and Video Processing*, 17(5), 2211–2219. <https://doi.org/10.1007/s11760-022-02436-4>
- Walter, S. D. (2005). The partial area under the summary ROC curve. *Statistics in medicine*, 24(13), 2025–2040.
- Webb, G. I. (2016). Naïve Bayes. Em C. Sammut & G. I. Webb (Orgs.), *Encyclopedia of Machine Learning and Data Mining* (p. 1–2). Springer US. https://doi.org/10.1007/978-1-4899-7502-7_581-1
- WHO. (2021). *WORLD HEALTH ORGANIZATION*. Cardiovascular Diseases. [https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-\(cvds\)](https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds))