

UNIVERSIDADE DE SÃO PAULO

Instituto de Ciências Matemáticas e de Computação

**Aprendizado de máquina aplicado à análise de
imagens de ressonância magnética para detecção de
cardiomiopatias**

Guilherme Lisbôa da Silva

Monografia - MBA em Inteligência Artificial e Big Data

SERVIÇO DE PÓS-GRADUAÇÃO DO ICMC-USP

Data de Depósito:

Assinatura: _____

Guilherme Lisbôa da Silva

Aprendizado de máquina aplicado à análise de imagens de ressonância magnética para detecção de cardiomiopatias

Monografia apresentada ao Departamento de Ciências de Computação do Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo - ICMC/USP, como parte dos requisitos para obtenção do título de Especialista em Inteligência Artificial e Big Data.

Área de concentração: Inteligência Artificial

Orientador: Profa. Dra. Fátima de Lourdes dos Santos Nunes

Versão original

São Carlos

2023

AUTORIZO A REPRODUÇÃO E DIVULGAÇÃO TOTAL OU PARCIAL DESTE TRABALHO, POR QUALQUER MEIO CONVENCIONAL OU ELETRÔNICO PARA FINS DE ESTUDO E PESQUISA, DESDE QUE CITADA A FONTE.

Ficha catalográfica elaborada pela Biblioteca Prof. Achille Bassi, ICMC/USP, com os dados fornecidos pelo(a) autor(a)

S856m	da Silva, Guilherme Lisbôa Aprendizado de máquina aplicado à análise de imagens de ressonância magnética para detecção de cardiomiopatias / Guilherme Lisbôa da Silva ; Orientadora Fátima de Lourdes dos Santos Nunes. – São Carlos, 2023. 82 p. : il. (algumas color.) ; 30 cm. Monografia (MBA em Inteligência Artificial e Big Data) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, 2023. 1. LaTeX. 2. abnTeX. 3. Classe USPSC. 4. Editoração de texto. 5. Normalização da documentação. 6. Tese. 7. Dissertação. 8. Documentos (elaboração). 9. Documentos eletrônicos. I. Nunes, Fátima de Lourdes dos Santos Nunes, orient. II. Título.
-------	--

Guilherme Lisboa da Silva

**Aprendizado de máquina aplicado à análise de imagens de
ressonância magnética para detecção de cardiomiopatias**

Monograph presented to the Departamento de Ciências de Computação do Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo - ICMC/USP, as part of the requirements for obtaining the title of Specialist in Artificial Intelligence and Big Data.

Concentration area: Artificial Intelligence

Advisor: Profa. Dra. Fátima de Lourdes dos Santos Nunes

Original version

São Carlos

2023

Este trabalho é dedicado ao meu futuro, há tudo aquilo que posso me tornar.

AGRADECIMENTOS

Agradeço, primeiramente, a Deus.

Agradeço ao meu marido, Antonio Vitor Oliveira Mourão, que estava ao meu lado em todo esse processo, me dando forças e sonhando comigo.

Agradeço também a minhas amigas que me ajudaram em um momento que estava necessitado.

Por fim, a mim porque nunca desisti de nada que me propus a fazer.

“I got smarter, I got harder in the nick of time.”

Taylor Swift

RESUMO

da Silva, G. L. **Aprendizado de máquina aplicado à análise de imagens de ressonância magnética para detecção de cardiomiopatias**. 2023. 82p. Monografia (MBA em Inteligência Artificial e Big Data) - Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos, 2023.

As cardiomiopatias, doenças do músculo cardíaco, expõem um desafio do diagnóstico devido à sua variabilidade e complexidade. Apesar da ressonância magnética ser uma técnica promissora, há uma dificuldade no diagnóstico preciso, principalmente, em locais onde não há profissionais habilitados para sua análise. Este estudo abordou o treinamento dos modelos de aprendizagem de máquina clássicos, a partir de uma base de dados de imagens segmentadas de ressonância magnética, no entendimento em como técnicas de aprendizagem de máquinas podem auxiliar no diagnóstico de cardiomiopatias. Os resultados indicam que a combinação de imagens de ressonância e técnicas de aprendizagem de máquina podem auxiliar no diagnóstico, apesar de apresentar dificuldade em diferenciar cardiomiopatias dilatada de hipertróficas, apresentando valores de $F1_score$ acima de 0,70. A pesquisa visibiliza que as técnicas de aprendizado de máquina têm potencial para contribuir para o diagnóstico preciso e constituem uma base para futuras investigações e otimizações nesta área interdisciplinar da medicina e ciência da computação.

Palavras-chave: Aprendizagem de Máquina. Imagem de Ressonância Magnética. Modelos clássicos de aprendizagem de máquina.

ABSTRACT

da Silva, G. L. **Machine Learning applied to analysis of magnetic resonance imaging for detection of heart diseases**. 2023. 82p. Monograph (MBA in Artificial Intelligence and Big Data) - Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos, 2023.

Cardiomyopathy, diseases of the heart muscle, pose a diagnostic challenge due to their variability and complexity. Although magnetic resonance imaging is a promising technique, there is a difficulty in accurate diagnosis, especially in locations where there are no qualified professionals to analyze it. This study addressed the training of classic machine learning models, using a database of segmented magnetic resonance images, to understand how machine learning techniques can assist in the diagnosis of cardiomyopathy. The results indicate that the combination of magnetic resonance images and machine learning techniques can assist in the diagnosis, despite the difficulty in differentiating dilated of hypertrophic cardiomyopathy, showing F1_score values above 0.70. The research shows that machine learning techniques have the potential to contribute to accurate diagnosis and constitute a basis for future investigations and optimizations in this interdisciplinary area of medicine and computer science.

Keywords: Machine Learning. Magnetic Resonance Imaging. Classic machine learning models.

LISTA DE FIGURAS

Figura 1 – Camadas do coração	32
Figura 2 – Fluxo geral do processo executado no presente projeto de pesquisa. . .	43
Figura 3 – Interface do sistema <i>Segment</i> com um exame de Ressonância Magnética aberto.	44
Figura 4 – Ferramenta de segmentação semi-automática do ventrículo esquerdo. .	44
Figura 5 – Matriz de confusão do modelo LR com aplicação de SAC SFM e reali- zação de calibração	57
Figura 6 – Curva ROC do modelo LR com aplicação de SAC SFM e realização de calibração	58
Figura 7 – Matriz de confusão do modelo LR com aplicação de SAC SFM e reali- zação de calibração	58
Figura 8 – Curva ROC do modelo LR com aplicação de SAC SFM e realização de calibração	59
Figura 9 – Matriz de confusão do modelo LR com aplicação de SAC SFM e reali- zação de calibração	59
Figura 10 – Curva ROC do modelo LR com aplicação de SAC SFM e realização de calibração	60
Figura 11 – Matriz de confusão do modelo XGBOOST após calibração	64
Figura 12 – Curva ROC do modelo XGBOOST após calibração	64
Figura 13 – Matriz de confusão do modelo XGBOOST após calibração	65
Figura 14 – Curva ROC do modelo XGBOOST após calibração	65
Figura 15 – Matriz de confusão do modelo XGBOOST após calibração	66
Figura 16 – Curva ROC do modelo XGBOOST após calibração	66
Figura 17 – Matriz de confusão do modelo RF após calibração	71
Figura 18 – Curva ROC do modelo RF após calibração	71
Figura 19 – Matriz de confusão do modelo LR após calibração	73
Figura 20 – Curva ROC do modelo LR após calibração	75

LISTA DE TABELAS

Tabela 1 – Resultados da análise sem a calibração	52
Tabela 2 – Resultados da análise com a calibração	53
Tabela 3 – Métricas dos melhores modelos (CC = com calibração; SC = sem calibração)	53
Tabela 4 – Resultados da análise sem a calibração	60
Tabela 5 – Resultados da análise com a calibração	61
Tabela 6 – Métricas dos melhores modelos (CC = com calibração; SC = sem calibração)	61
Tabela 7 – Resultados da análise sem a calibração	67
Tabela 8 – Resultados da análise com a calibração	68
Tabela 9 – Métricas dos melhores modelos (CC = com calibração; SC = sem calibração)	68
Tabela 10 – Resultados da análise sem a calibração	72
Tabela 11 – Resultados da análise com a calibração	72
Tabela 12 – Métricas dos melhores modelos (CC = com calibração; SC = sem calibração)	73

LISTA DE QUADROS

Quadro 1 – Trabalhos correlatos obtidos na base de dados ACM e IEEE	39
Quadro 2 – Dados da análise da literatura (PP = Pré-Processamento; OT = Objetivo Técnico)	41
Quadro 3 – Descrição dos descritores cardíacos	46
Quadro 4 – Descritores gerados para a classificação dos pacientes por sexo e idade	47
Quadro 5 – Hiperparâmetros após a calibração do modelo LR original.	54
Quadro 6 – Hiperparâmetros após a calibração do modelo XGBOOST PCA	55
Quadro 7 – Hiperparâmetros após a calibração do modelo ADA SAC SKB	56
Quadro 8 – Hiperparâmetros após a calibração do modelo GBC SAC SKB	56
Quadro 9 – Hiperparâmetros após a calibração do modelo KNN SAC SKB	56
Quadro 10 – Hiperparâmetros após a calibração do modelo LR SAC SFM	57
Quadro 11 – Hiperparâmetros após a calibração do modelo DT SAC SFM	62
Quadro 12 – Hiperparâmetros após a calibração do modelo ADA SAC SKB	62
Quadro 13 – Hiperparâmetros após a calibração do modelo NB PCA	62
Quadro 14 – Hiperparâmetros após a calibração do modelo XGBOOST original . .	63
Quadro 15 – Hiperparâmetros após a calibração do modelo RF original	69
Quadro 16 – Hiperparâmetros após a calibração dos modelos	74

LISTA DE ABREVIATURAS E SIGLAS

AM	Aprendizado de Máquinas
OMC	Organização Mundial de Saúde
CAD	Computer-Aided Diagnosis
3D	Tridimensional
DU	Digital Universe
IDC	Internacional Data Corporation
EB	Exabytes
ZB	Zettabits
CC	Cardiopatía Congênita
DL	Deep Learning
ACM	ACM Digital Library
IEEE	IEEE Xplore Digital Library
SVM	Support Vector Machine
CNN	Convolutional Neural Network
RNA	Rede Neural Artificial
LR	Logistic Regression
RF	Random Forest
PCA	Principal Components Analysis
IRM	Imagem de ressonância Magnética
CMRI	Cardiac Magnetic Resonance Imaging
PCA	Principal Component Analysis

KNN	K Neighbors Classifier
DT	Decision Tree
XGBOOST	Extreme Gradient Boosting
ADA	Ada Boost
NB	Naive Bayes
RBFSVM	SVM - Radial Kernel
SAC	Seleção automática de características
SKB	SelectKBest
SFM	SelectFromModel
ADA	Ada Boost Classifier
GBC	Gradient Boosting Classifier
PP	Pré-Processament
OT	Objetivo Técnico
VP	Verdadeiro Positivo
FP	Falso Positivo
AUC	Area Under the Curve
VN	Verdadeiros negativos
FN	Falsos negativos
YOLO	You Only Look Once
LSTM	Long Short-Term Memory

SUMÁRIO

1	INTRODUÇÃO	27
1.1	Motivação e Justificativa	28
1.2	Objetivos	28
1.3	Organização do documento	29
2	FUNDAMENTAÇÃO TEÓRICA	31
2.1	Funcionamento do coração humano	31
2.2	Cardiomiopatias	32
2.3	Imagem de Ressonância Magnética	33
2.4	Aprendizado de máquina	33
2.4.1	Métodos de aprendizagem de máquina	34
2.4.2	Métricas de avaliação	36
3	TRABALHOS RELACIONADOS	39
3.1	Considerações finais do capítulo	41
4	MÉTODOS	43
4.1	Análise do conjunto de dados	43
4.1.1	Variação do ventrículo esquerdo	45
4.2	Seleção dos métodos e métricas	45
4.2.1	Biblioteca utilizada na pesquisa	46
4.2.2	Escolha da Métrica de Avaliação	47
4.3	Seleção automática de características	48
4.4	Fusão de características	49
4.5	Calibração	49
4.6	Coleta das métricas finais	50
4.7	Considerações finais do capítulo	50
5	RESULTADOS E DISCUSSÕES	51
5.1	Classificação Binária	52
5.1.1	Classificação sem segmentação de grupos	52

5.1.1.1	Análise do modelo com melhor desempenho	54
5.1.2	Classificação com a base de dados segmentada	58
5.1.2.1	Análise do modelo com melhor desempenho	61
5.1.3	Melhor modelo da classificação binária	64
5.2	Classificação Multiclass	67
5.2.1	Classificação sem a base de dados segmentada	67
5.2.1.1	Análise do modelo com melhor desempenho	70
5.2.2	Classificação com a base de dados segmentada	70
5.2.2.1	Análise do modelo com melhor desempenho	72
5.2.3	Melhor modelo de classificação multiclass	76
6	CONSIDERAÇÕES FINAIS	77
6.1	Trabalhos futuros	77
	Referências	79

1 INTRODUÇÃO

Aproximadamente 16,7 milhões de mortes por ano são causadas por doenças cardíacas, conforme a Organização Mundial da Saúde (OMS). Essa situação é ainda mais evidente nos países mais desenvolvidos, nos quais, complicações cardíacas são responsáveis por quase 88% das mortes (ORGANIZATION, 2023).

O coração é um órgão vital para o funcionamento do corpo humano (ALBERTS *et al.*, 2017). Um desequilíbrio no coração, mesmo que pequeno, pode ser fatal para pessoas com problemas cardiovasculares (YOON *et al.*, 2019). Devido a sensibilidade nesse órgão e sua importância para manutenção da vida, busca-se nas últimas décadas aprimorar os conhecimentos sobre o tema através de tecnologias não invasivas, como o uso de imagens por exemplo, ultrassonografia, raios-X, tomografia computadorizada e Imagem de Ressonância Magnética (IRM) (YOON *et al.*, 2019), visando contribuir com o diagnóstico de doenças cardiomiopatias.

A IRM usa a rotação de ímãs que pulsam uma corrente de radiofrequência através do corpo para ordenar os prótons sincronizados com a frequência da ressonância (WANG *et al.*, 2019). Com a ajuda da transformada de Fourier, é possível construir imagens tridimensionais (3D) de alta qualidade. Uma desvantagem do uso da IRM é o tempo que leva para escanear o corpo do paciente. Todavia, é um exame seguro e ideal para um diagnóstico preciso do coração (WANG *et al.*, 2019).

O avanço tecnológico na medicina impulsionou o armazenamento de dados biomédicos. Esse grande conjunto de dados impulsionou a criação de Sistemas de Diagnóstico Auxiliado por Computador (*Computer-Aided Diagnosis - CAD*), uma vez que a análise de dados biomédicos é um procedimento complexo e que demanda tempo substancial. Por isso, sistemas CAD estão se tornando cada vez mais populares para auxiliar médicos em suas tomadas de decisão (IBRAHIM *et al.*, 2019).

No que tange a assertividade dos dados biométricos, o uso de técnicas de Aprendizagem de Máquina (AM) na construção de CADs esta sendo explorando com o objetivo é auxiliar no diagnóstico de patologias, incluindo doenças cardiopatias (GIGER, 2018) Dessa forma, possibilita-se a criação de predições que auxiliem os médicos na precisão do

diagnóstico (LI; LI, 2022).

1.1 Motivação e Justificativa

Cotidianamente, um enorme volume de dados é gerado e armazenado ao redor do mundo, criando assim o termo *Digital Universe* (DU) que trata-se da quantidade de dados replicados, criados e consumidos anualmente (DASH *et al.*, 2019). Conforme a *Internacional Data Corporation* (IDC), em 2005, estimava-se um DU equivalente a 130 *exabytes* (EB). Em 2017, esse valor estendeu-se para 16 *zettabytes* (ZB) e, posteriormente, em 2020, foi aferido, aproximadamente, 40 ZB de acondicionamento de dados (DASH *et al.*, 2019). A evolução da tecnologia e, conseqüentemente, do armazenamento de dados, nos últimos anos também contribuiu para que a área da saúde pudesse beneficiar-se de novas possibilidades para analisar os dados gerados e aprimorar as investigações clínicas.

Dada a complexidade do diagnóstico médico, uma vez que os dados médicos são extremamente heterogêneos (IBRAHIM *et al.*, 2019), o processo se torna demorado e sujeito a algumas variáveis, como a experiência do médico, o conhecimento da enfermidade e a capacidade de distingui-la de outras patologias, contribuindo para um diagnóstico impreciso ou errôneo de uma condição clínica (LÓPEZ; BRAVO; HUERTA, 2009). Doenças como a cardiomiopatia hipertrófica, cardiopatia congênita, cardiopatia reumática, cardiopatia não compactada e outras cardiopatias são similares e de difícil identificação, exigindo uma grande gama de exames e um conhecimento específico da doença (MATTOS *et al.*, 2012). Ademais, a precoce e assertiva identificação de cardiopatias ampliam o desafio ao diagnosticar uma enfermidade.

A disponibilidade de bases de dados médicas favorece a construção de sistemas CAD, que têm como função principal auxiliar os médicos na identificação de patologias (KIM *et al.*, 2018). Assim sendo, este trabalho tem como principal propósito a realização de um estudo de modelos de AM para a classificação de cardiomiopatias a partir de IRM, a fim de identificar qual modelo apresenta o melhor desempenho com o intuito de auxiliar no processo de diagnóstico de enfermidades cardiopáticas.

1.2 Objetivos

O presente estudo visa desenvolver e validar um modelo de AM capaz de classificar características que foram extraídas de IRM para o diagnóstico de cardiomiopatias.

Para ser atingido o objetivo geral, este trabalho tem como objetivos específicos:

- Selecionar e adaptar técnicas de aprendizado de máquina supervisionado;
- Definir e implementar recursos de pré-processamento de dados;
- Selecionar e implementar métricas;
- Definir um modelo de AM eficiente a partir da combinação de técnicas de pré-processamento e algoritmos de classificação.

1.3 Organização do documento

A estrutura deste estudo é organizado conforme a descrição subsequente:

- O capítulo 2 apresenta a fundamentação teórica que possibilita agnição sobre o funcionamento do coração e suas cardiopatias, além dos conceitos das IRM como auxiliares no diagnóstico de cardiomiopatias.
- O capítulo 3 apresenta a pesquisa bibliográfica dos trabalhos relacionados ao estudo as técnicas de uso de IRM de AM para segmentar ou detectar doenças cardiomiopáticas.
- O capítulo 4 apresenta as descrições metodológicas utilizadas na seleção dos métodos e métricas, calibragem e coleta dos modelos aplicados para analisar a base de dados a ser utilizada.
- O capítulo 5 apresenta os resultados obtidos com as classificações binárias, multiclass.
- O capítulo 6 apresenta as considerações finais do trabalho e atribuições de trabalhos futuros.

2 FUNDAMENTAÇÃO TEÓRICA

Neste capítulo serão abordados conceitos relevantes para o entendimento da pesquisa. Inicialmente, na seção 2.1 será apresentado o funcionamento do coração. Na seção 2.2 será explicado quais são as cardiomiopatias, suas características, sintomas e tratamentos. Por fim, na seção 2.3, será discutido como os exames de IRM podem auxiliar no diagnóstico de cardiomiopatias.

2.1 Funcionamento do coração humano

O coração humano é um órgão vital, tendo como função o bombeamento do sangue pelos vasos sanguíneos (MARIEB; WILHELM; MALLATT, 2012). Revestido pela parede cardíaca encontram-se três camadas: o epicárdio, a mais externa; o miocárdio, a intermediária; e o endocárdio, a mais interna. O miocárdio é responsável pelos movimentos de contração e relaxamento cardíacos, que compõem o ciclo cardíaco e causam o bombeamento sanguíneo (HALL; GUYTON, 2020).

Para bombear sangue pelo corpo, o coração tem que se contrair e relaxar, formando o complexo ciclo cardíaco (HALL; GUYTON, 2020). Este comportamento tem dois momentos-chave: a sístole, quando o coração se aperta e libera o sangue, e a diástole, momento de descanso e reabastecimento de sangue no coração (HALL; GUYTON, 2020). Durante a sístole, os ventrículos se ativam, direcionando o sangue do ventrículo esquerdo para a principal artéria, a aorta. As válvulas que separam os átrios dos ventrículos se fecham, barrando a passagem do sangue para as câmaras superiores. Na fase da diástole, há o fechamento das válvulas que impedem o retorno de sangue aos ventrículos e, simultaneamente, as válvulas entre os átrios e ventrículos se abrem para uma nova carga sanguínea. Esse mecanismo é vital para o sangue continuar fluindo e garantindo que cada parte do corpo obtenha os recursos vitais de que precisa (MARIEB; WILHELM; MALLATT, 2012). A Figura 1 mostra uma representação gráfica das estruturas cardíacas.

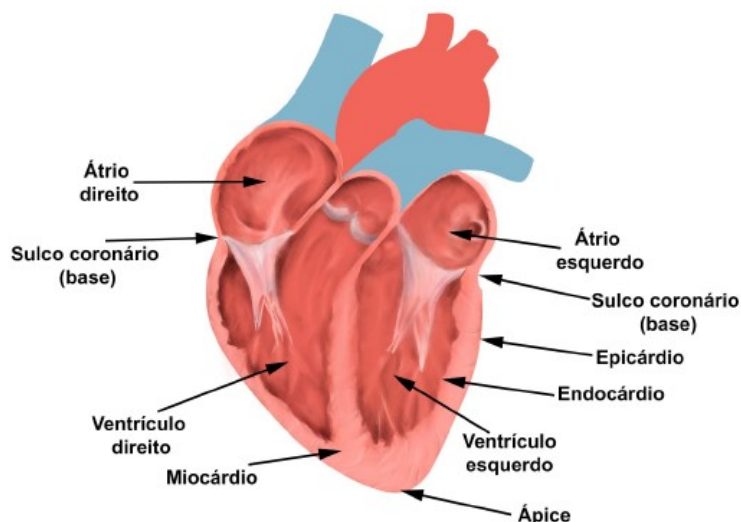


Figura 1 – Camadas do coração

Fonte: Ribeiro (2021)

2.2 Cardiomiopatias

As cardiomiopatias são doenças que afetam o músculo cardíaco, também conhecido como miocárdio (MAURIZI *et al.*, 2018). São classificadas em três tipos principais: cardiomiopatia dilatada, cardiomiopatia hipertrófica e cardiomiopatia restritiva (WEXLER *et al.*, 2009). A cardiomiopatia dilatada é marcada pelo aumento das paredes do coração, o que pode causar uma diminuição da capacidade de bombeamento de sangue (MERLO *et al.*, 2020). A cardiomiopatia hipertrófica é caracterizada através do espessamento do miocárdio, o que pode causar uma diminuição do tamanho das câmaras cardíacas (MERLO *et al.*, 2020). Por fim, a cardiomiopatia restritiva é representada pela rigidez do miocárdio, o que pode dificultar o enchimento adequado das câmaras cardíacas (WEXLER *et al.*, 2009).

As cardiomiopatias podem ser causadas por diversos fatores, incluindo problemas genéticos, infecções virais, exposição a toxinas e doenças autoimunes (MAURIZI *et al.*, 2018). Os sintomas podem incluir falta de ar, fadiga, inchaço nas pernas e nos tornozelos, palpitações e tonturas (MAURIZI *et al.*, 2018). O diagnóstico pode ser feito por meio de exames como eletrocardiograma, ecocardiograma e ressonância magnética cardíaca (MAURIZI *et al.*, 2018).

O tratamento pode envolver o uso de medicamentos para controlar os sintomas e melhorar a função cardíaca, assim como mudanças no estilo de vida, como uma dieta

saudável e a prática de exercícios físicos (MERLO *et al.*, 2020). Em alguns casos, pode ser necessária a realização de procedimentos cirúrgicos, como a colocação de um dispositivo de assistência ventricular ou um transplante cardíaco (MAURIZI *et al.*, 2018).

2.3 Imagem de Ressonância Magnética

A IRM é uma técnica de diagnóstico por imagem que utiliza um campo magnético e ondas de rádio para produzir imagens detalhadas do corpo humano sem o uso de radiação ionizante (GROVER *et al.*, 2015). A técnica faz com que os átomos de hidrogênio mude sua posição e, em seguida, voltem à sua posição original emitindo um sinal detectado pelo equipamento (DALE; BROWN; SEMELKA, 2015). Quando um campo magnético forte é aplicado, os átomos de hidrogênio são, momentaneamente, deslocados de sua posição de equilíbrio e, seguidamente, liberam energia sob a forma de ondas de rádio. Essas ondas são detectadas por um dispositivo chamado antena de radiofrequência e processadas por um computador para produzir imagens detalhadas dos tecidos e órgãos do corpo. A técnica de IRM é particularmente útil na detecção e diagnóstico de doenças neurológicas, cardíacas, ortopédicas e oncológicas (DALE; BROWN; SEMELKA, 2015).

Por tratar-se de uma técnica não invasiva que permite a obtenção de imagens detalhadas do coração e dos vasos sanguíneos, a IRM tem se mostrado uma ferramenta valiosa para o diagnóstico precoce e a avaliação da gravidade de cardiomiopatias (KONST *et al.*, 2023). Isso porque a IRM permite uma avaliação precisa da estrutura e da função do ventrículo esquerdo, o que é essencial para o diagnóstico e o monitoramento das cardiomiopatias (KONST *et al.*, 2023). Com a evolução da tecnologia e a melhoria das técnicas de imagem, a IRM se tornou uma ferramenta cada vez mais importante no diagnóstico e no tratamento das cardiomiopatias (KUMAR; BAGUR, 2015). É importante destacar, no entanto, que a interpretação adequada das imagens de IRM requer um conhecimento aprofundado da anatomia e da fisiologia cardíacas, bem como das características específicas de cada tipo de cardiomiopatia (MATTOS *et al.*, 2012).

2.4 Aprendizado de máquina

Aprendizado de Máquina (AM) é uma área que pode contribuir para melhoria de um software, a partir da análise de exemplos de seu comportamento desejado, em contrapartida, a um sistema previamente programado (PROVOST; FAWCETT, 2013). Um uso adequado

de AM é quando não se consegue fornecer especificações para o comportamento do programa, apenas exemplos em que este comportamento ocorre (CHATZILYGEROUDIS; HATZILYGEROUDIS; PERIKOS, 2021), como o reconhecimento de escrita, caracteres e fala. Um sinal que há necessidade de aplicar AM é quando uma tarefa pode ser executada, contudo, sem uma articulação precisa de como realizá-la (PROVOST; FAWCETT, 2013). Situações em que a tarefa muda simultaneamente ou é muito adaptável às alterações das variáveis de análise, também são exemplos de aplicabilidade de AM (CHATZILYGEROUDIS; HATZILYGEROUDIS; PERIKOS, 2021).

Ao considerar a presença de rotulação dos dados e a consequente forma de treinamento, pode-se classificar a AM, em três categorias: supervisionado, não-supervisionado e reforçado (CHATZILYGEROUDIS; HATZILYGEROUDIS; PERIKOS, 2021).

O AM supervisionado constitui uma categoria de técnicas em que um modelo é treinado a partir de dados rotulados previamente (PROVOST; FAWCETT, 2013). Com base na rotulação, o método de aprendizado pode fazer previsões precisas em dados futuros não vistos anteriormente (MOHRI; ROSTAMIZADEH; TALWALKAR, 2012).

O AM não-supervisionado estabelece técnicas de aprendizagem através da ausência de variável alvo ou rótulos associados aos dados de treinamento (MOHRI; ROSTAMIZADEH; TALWALKAR, 2012). Sendo tarefa do algoritmo encontrar padrões e estruturas nos dados (PROVOST; FAWCETT, 2013).

Por último, o AM por reforço refere-se a um conjunto de técnicas de AM em que um agente aprende a tomar ações em um ambiente para maximizar uma recompensa numérica (MOHRI; ROSTAMIZADEH; TALWALKAR, 2012). O agente não recebe exemplos rotulados como no aprendizado supervisionado, nem tem informações prévias sobre o ambiente como no aprendizado não supervisionado. À vista disso, ele aprende a tomar ações com base na interação com o ambiente e nas recompensas recebidas ou penalidades impostas (SALAKHUTDINOV, 2014).

2.4.1 Métodos de aprendizagem de máquina

Os principais métodos utilizados neste trabalho foram *Ada Boost Classifier* (ADA), *decision tree* (DT), *Gradient Boosting Classifier* (GBC), *K-Nearest Neighbors Algorithm* (KNN), *Logistic Regression* (LR), *Naive Bayes* (NB), *Support Vector Machines Radial Basis Function Kernel* (RBFSVM), *Random Forest* (RF), *eXtreme Gradient Boosting*

(XGBOOST) e *Support Vector Machine* (SVM).

- **ADA:** um método de *boosting* que combina múltiplos modelos fracos para criar um modelo forte. Ele ajusta iterativamente os pesos das observações com base nos erros das iterações anteriores (PROVOST; FAWCETT, 2013).
- **DT:** um método de aprendizado de máquina que utiliza uma estrutura de árvore para tomar decisões com base em regras. Cada nó representa um atributo e cada ramo representa uma decisão, levando a folhas que são classificações (PROVOST; FAWCETT, 2013).
- **GBC:** um método de aprendizado de máquina que utiliza *boosting* para otimizar a classificação por meio de árvores de regressão. Em cada estágio, as árvores de regressão se adaptam ao valor negativo da função de perda (AKERKAR; SAJJA, 2016).
- **KNN:** um método de aprendizado supervisionado que classifica um ponto de dados com base em como seus vizinhos são classificados (PROVOST; FAWCETT, 2013).
- **LR:** um método de classificação. Ele estima a probabilidade de uma instância pertencer a uma categoria específica usando uma função logística (AKERKAR; SAJJA, 2016).
- **NB:** um método de classificação baseado no teorema de Bayes, que assume independência entre os atributos (PROVOST; FAWCETT, 2013).
- **RBF SVM:** é um modelo SVM que utiliza a função de base *radial basis function Kernel* (RBF) como kernel. Ela mapeia os dados para um espaço de características de alta dimensão para encontrar o hiperplano de separação ótimo (AKERKAR; SAJJA, 2016).
- **RF:** Trata-se de um método de aprendizado de máquina baseado em conjunto que combina múltiplas árvores de decisão durante o treinamento. Pode ser usado tanto para classificação quanto para regressão (AKERKAR; SAJJA, 2016).
- **SVM:** um método de aprendizado supervisionado usado para classificação e regressão. Ele funciona encontrando o hiperplano que melhor separa as classes de dados em um espaço de alta dimensão (AKERKAR; SAJJA, 2016).

- **XGBOOST:** um método de aprendizado de máquina otimizado baseado em árvore de incremento (PROVOST; FAWCETT, 2013).

2.4.2 Métricas de avaliação

As métricas de AM são ferramentas vitais para avaliar a precisão e a eficácia de modelos preditivos (PROVOST; FAWCETT, 2013). Elas fornecem uma visão quantitativa do desempenho do modelo em comparação com a realidade, o que permite identificar áreas de melhoria e garantir que o modelo atenda aos requisitos do problema (AKERKAR; SAJJA, 2016). A classificação é medida entre a relação de verdadeiros positivos e falsos positivos. Alguns exemplos que clarificam os aspectos dessas métricas são a taxa de acertos do modelo, a proporção de verdadeiros positivos e a prevenção de falsos positivos (CHATZILYGEROUDIS; HATZILYGEROUDIS; PERIKOS, 2021).

As métricas utilizadas neste trabalho foram a acurácia (*accuracy*), precisão (*precision*), revocação (*recall*) e *F1_score*, curva AUC. Para o devido entendimento da fórmula de cada métrica é importante a compreensão do que é verdadeiro positivo e falso positivo.

- **Verdadeiro positivo (VP):** o número de casos em que o modelo previu a classe positiva corretamente. Em outras palavras, o modelo prevê ser positivo quando o valor real também é positivo (PROVOST; FAWCETT, 2013).
- **Falso positivo (FP):** o número de casos em que o modelo prevê, incorretamente, a classe positiva. Ocorre quando o valor real é negativo, mas o modelo espera que seja positivo (PROVOST; FAWCETT, 2013).
- **Verdadeiros negativos (VN):** o número de casos em que o modelo prevê, corretamente, que uma determinada instância é negativa, e de fato, ela é negativa no conjunto de dados real (PROVOST; FAWCETT, 2013).
- **Falsos negativos (FN):** o número de casos em que o modelo prevê, incorretamente, que uma determinada instância é negativa, quando, na verdade, ela é positiva no conjunto de dados real (PROVOST; FAWCETT, 2013).

Para fins matemáticos encontra-se na literatura a nomenclaturas VP, FP, VN e FN descritas (PROVOST; FAWCETT, 2013) como:

- TP : é o número de verdadeiros positivos.
- TN : é o número de verdadeiros negativos.
- FP : é o número de falsos positivos.
- FN : é o número de falsos negativos.

Do que tange as métricas utilizadas no desenvolvimento do projeto, apresenta-se, conceitualmente, as métricas empregadas.

A acurácia é uma métrica usada em estudos estatísticos e de AM para calcular a proporção de acertos em relação ao total de previsões que um modelo fez. Calcula-se através da divisão do número total de previsões pelo número de previsões corretas. Outrossim, a acurácia mostra a eficácia do modelo nas previsões, independentemente de o resultado ser positivo ou negativo (PROVOST; FAWCETT, 2013). É definida, matematicamente, como:

$$\text{Acurácia} = \frac{\text{Número de previsões corretas}}{\text{Número total de previsões}} = \frac{TP + TN}{TP + TN + FP + FN} \quad (2.1)$$

Em AM, a precisão é uma medida que afere a quantidade de identificações positivas corretamente feitas em comparação com o total de identificações positivas, incluindo falsos positivos (PROVOST; FAWCETT, 2013). É matematicamente definido como:

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

A proporção de positivos reais identificados corretamente é chamada de revocação, também denominada como sensibilidade ou taxa de verdadeiros positivos (PROVOST; FAWCETT, 2013). A fórmula que o define é:

$$\text{Revocação} = \frac{\text{Verdadeiro Positivo}}{\text{Verdadeiro Positivo} + \text{Falso Negativo}} = \frac{TP}{TP + FN} \quad (2.3)$$

O $F1_score$ é uma métrica de desempenho balanceado que combina precisão e revocação. Ele é a média harmônica entre precisão e revocação e dá igual importância a ambas. O $F1_score$, particularmente, é útil quando as classes são desequilibradas. Seu

valor varia entre 0 e 1, sendo 1 o desempenho perfeito e 0 o pior desempenho possível (PROVOST; FAWCETT, 2013). A fórmula para o $F1_score$ é:

$$F1 = 2 \times \frac{\text{precisão} \times \text{revocação}}{\text{precisão} + \text{revocação}} \quad (2.4)$$

Por fim, a curva AUC refere-se à “Área sob a Curva” do ROC, sendo ROC uma sigla para *Receiver Operating Characteristic* (PROVOST; FAWCETT, 2013). Há dois conceitos importantes:

- **Curva ROC:** é uma representação gráfica do desempenho de um classificador binário à medida que o limiar de decisão é variado. Ela traça a taxa de verdadeiros positivos (sensibilidade) contra a taxa de falsos positivos (especificidade) para diferentes pontos de corte. A curva ROC demonstra a capacidade de um modelo em distinguir entre as classes positiva e negativa (PROVOST; FAWCETT, 2013).
- **AUC (*Área Under the Curve*):** AUC é, literalmente, a área sob a curva ROC. Ela fornece uma medida agregada do desempenho do modelo em todos os possíveis limiares. AUC pode ter um valor entre 0 e 1 (PROVOST; FAWCETT, 2013), onde:
 - 0 indica que o modelo está classificando todas as observações de maneira inversa (PROVOST; FAWCETT, 2013).
 - 0,5 sugere que o modelo não tem capacidade discriminativa (equivalente a uma previsão aleatória) (PROVOST; FAWCETT, 2013).
 - 1 indica que o modelo tem um desempenho perfeito (PROVOST; FAWCETT, 2013).

3 TRABALHOS RELACIONADOS

A fim de identificar trabalhos relacionados, foram realizadas pesquisas com diferentes *strings* de busca nas bases *ACM Digital Library* e *IEEE Xplore Digital Library* (IEEE), conforme apresentado no Quadro 1. De acordo com o observado, foi constatado que existe uma predominância de trabalhos científicos com o uso de técnicas de ML convencionais para a análise de imagens de IRM com o intuito da realização de identificação de patologias cardíacas.

Índices	<i>String</i> de busca	Base de dados	Quantidade
1	"mri"AND "heart disease"AND "machine learning"AND NOT "deep learning"	<i>ACM Digital Library</i>	20
2	"MRI"AND "cardiomyopathy"AND "machine learning"	<i>IEEE Xplore Digital Library</i>	3
3	"cardiomyopathy"AND "machine learning"and "Magnetic Resonance Imaging"	<i>ACM Digital Library</i>	62
4	"cardiomyopathy"AND "machine learning"and "Magnetic Resonance Imaging"	<i>IEEE Xplore Digital Library</i>	2
5	"cardiomyopathy"AND "machine learning"	<i>ACM Digital Library</i>	57
6	"machine learning"AND "Magnetic Resonance Imaging"AND "heart disease"	<i>ACM Digital Library</i>	45
7	"machine learning"AND "Cardiac Magnetic Resonance Imaging"	<i>ACM Digital Library</i>	7
8	"machine learning"AND "Cardiac Magnetic Resonance"	<i>ACM Digital Library</i>	8
9	"machine learning"AND "Cardiac Magnetic Resonance"AND "cardiomyopathy"	<i>ACM Digital Library</i>	3
10	"machine learning"AND "Cardiac Magnetic Resonance"	<i>ACM Digital Library</i>	9
11	"machine learning"AND "Magnetic Resonance"AND "cardiomyopathy"	<i>ACM Digital Library</i>	11
12	"machine learning"AND "Magnetic Resonance"AND "cardiomyopathy"	<i>IEEE Xplore Digital Library</i>	3

Quadro 1 – Trabalhos correlatos obtidos na base de dados ACM e IEEE

Os artigos recuperados foram analisados e aqueles que estavam relacionados com o tema do presente trabalho foram lidos na íntegra (Quadro 2). Os estudos selecionados apresentam diferentes técnicas de AM aplicadas à análise de imagens de Cardiac Magnetic Resonance Imaging (CMRI) e à detecção de doenças cardíacas.

O primeiro grupo utiliza técnicas de segmentação para identificar estruturas cardíacas nas imagens de IRM. Dentre os artigos do Quadro 2 são identificados Nasr-Esfahani *et al.* (2018) e Ibrahim *et al.* (2019) que usam a rede CNN para segmentar o coração.

O segundo grupo de artigos utiliza técnicas de classificação para identificar diferentes doenças cardíacas. Os artigos de Xiao *et al.* (2020), Ramesh *et al.* (2021) e Chang *et al.* (2018) propõem a utilização de rede CNN para classificar diferentes cardiomiopatias, contudo Chang *et al.* (2018) antes de aplicar CNN utiliza You Only Look Once (YOLO) para obter a região de interesse de toda a sequência de IRM diastólica e sistólica. Já Xiao *et al.* (2020) usam *Long Short-Term Memory* (LSTM) além de CNN para auxiliar na classificação de cardiomiopatia dilatada, hipertrófica e normal.

Em seguida, há três artigos que propõem técnicas para estimar diferentes índices e volumes cardíacos a partir de imagens CMRI. Kim *et al.* (2018) propõe uma metodologia automatizada para estimar o volume do ventrículo esquerdo usando uma combinação de técnicas de segmentação e análise de forma. Para estimar o volume do ventrículo esquerdo, é usada a CNN. Seguindo a mesma tendência, o Zheng *et al.* (2019) utiliza CNN para estimar diferentes índices cardíacos a partir de imagens de ressonância magnética. Por fim, Margeta *et al.* (2017) propõe um método baseado na CNN para recuperar automaticamente informações de planos de aquisição cardíaca ausentes.

Em suma, os estudos analisados usam, na maioria, DL e poucos usam técnicas clássicas de AM. Dentre as técnicas usadas, foram citadas predominantemente CNN para a análise de imagens de CMRI e a detecção de doenças cardíacas. A segmentação, a classificação e a estimativa de volumes cardíacos são usadas em diversas combinações para lidar com os desafios específicos de cada aplicação.

O Quadro 2 mostra que a maioria dos artigos usa técnicas de pré-processamento, com destaque para a normalização espacial das imagens, normalização dos dados e intensidade. Poucos artigos usam aumento de dados.

Índice	Artigo	Categoria de técnicas	Tipo de base de dados	PP	OT
1	(NASR-ESFAHANI <i>et al.</i> , 2018)	Deep Learning	Não mencionado	Normalização e padronização dos dados; Aumento de dados; Normalização espacial em imagens; Suavização de ruídos em imagens	Segmentação do ventrículo esquerdo
2	(RAMESH <i>et al.</i> , 2021)	Deep Learning	híbrida	não especificado	Deteção de doenças cardíacas
3	(XIAO <i>et al.</i> , 2020)	Deep Learning	pública	não informa	reconhecimento de cardiomiopatia
4	(IBRAHIM <i>et al.</i> , 2019)	Deep learning	própria	normalização de intensidade; Suavização de ruídos em imagens	segmentação do ventrículo esquerdo
5	(CHANG <i>et al.</i> , 2018)	Deep Learning	pública	normalização e redimensionamento das imagens	segmentação do ventrículo esquerdo e classificação de cardiomiopatia
6	(KIM <i>et al.</i> , 2018)	Deep Learning	própria	Normalização de intensidade	segmentação do ventrículo esquerdo
7	(MARGETA <i>et al.</i> , 2017)	Deep Learning	própria	não especifica	estimativa de índices cardíacos
8	(ZHENG <i>et al.</i> , 2019)	Deep Learning	própria	não especifica	estimativa de índices cardíacos

Quadro 2 – Dados da análise da literatura (PP = Pré-Processamento; OT = Objetivo Técnico)

3.1 Considerações finais do capítulo

Com base na revisão realizada, observa-se que a aplicação de técnicas avançadas de DL desempenha um papel crucial na análise de imagens de CMRI. A CNN é a técnica

predominante desde segmentação de estruturas cardíacas até a detecção de diferentes cardiopatias. Outras abordagens, como YOLO e LSTM, também foram incorporadas para melhorar a precisão da análise.

Em resumo, o campo da análise de imagens cardíacas para a detecção de cardiopatias está avançando rapidamente, baseado na revisão realizada neste capítulo, com as técnicas de DL predominantemente as CNNs.

4 METODOLOGIA

Como foi descrito na seção 1.2, o objetivo deste projeto é desenvolver e validar um modelo de AM capaz de classificar características que foram extraídas de IRM para o diagnóstico de cardiomiopatias. Para ser possível atingir esse objetivo foi necessário 6 etapas descritas na Figura 2.

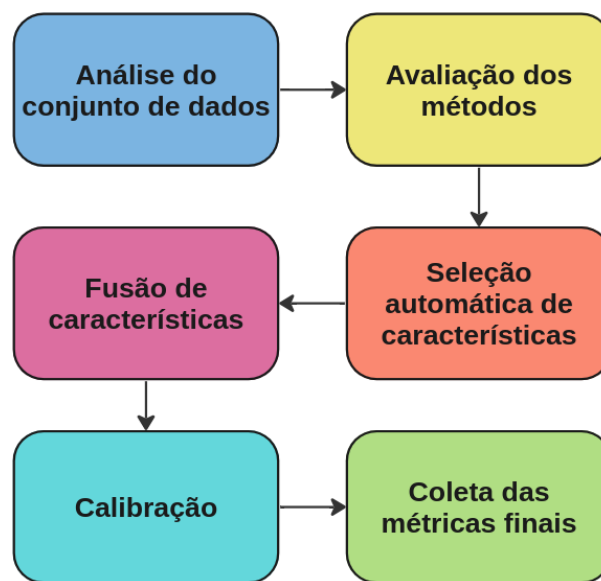


Figura 2 – Fluxo geral do processo executado no presente projeto de pesquisa.

Fonte: Elaborada pelo autor, 2023

4.1 Análise do conjunto de dados

Este trabalho utiliza-se da base de dados desenvolvida na pesquisa de Delmondes (2022) realizada no Laboratório de Aplicações de Informática em Saúde (LApIS) da Escola de Artes, Ciências e Humanidades(EACH) da USP, em colaboração com o Instituto do Coração (InCor). O laboratório teve acesso a uma base de IRM do InCor para validar suas técnicas. O estudo envolveu 384 exames de IRM Os especialistas classificaram esta base de dados com 113 pacientes com cardiomiopatia dilatada, 170 pacientes com cardiomiopatia hipertrófica e 101 pacientes com o coração normal.

Para a construção das bases de dados, utilizou-se o algoritmo de segmentação semi-automática do software *Segment*, *software Segment*, (TUFVESSON *et al.*, 2015), para extração de contornos do endocárdio e epicárdio do ventrículo esquerdo cardíaco. Dividido em duas etapas, a primeira foi a abertura dos arquivos contendo as imagens da Figura 3, seleciona posteriormente os *slices* inicial e final para segmentação, Figura 4. Os resultados mostram os contornos segmentados em cores diferentes.

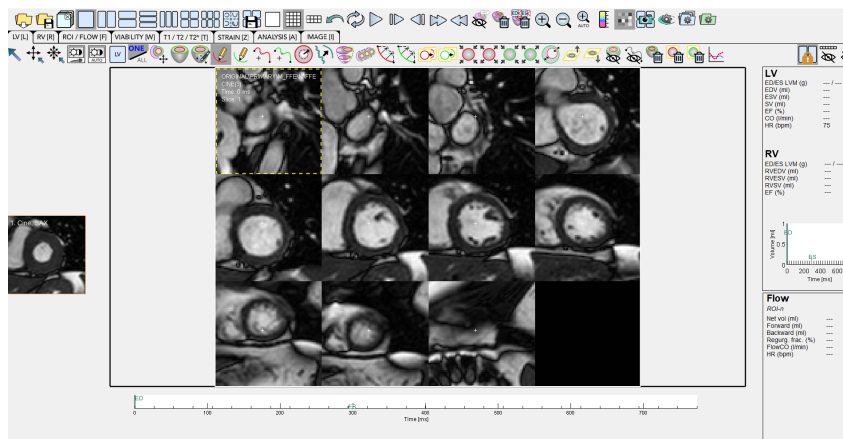


Figura 3 – Interface do sistema *Segment* com um exame de Ressonância Magnética aberto.

Fonte: Delmondes (2022)

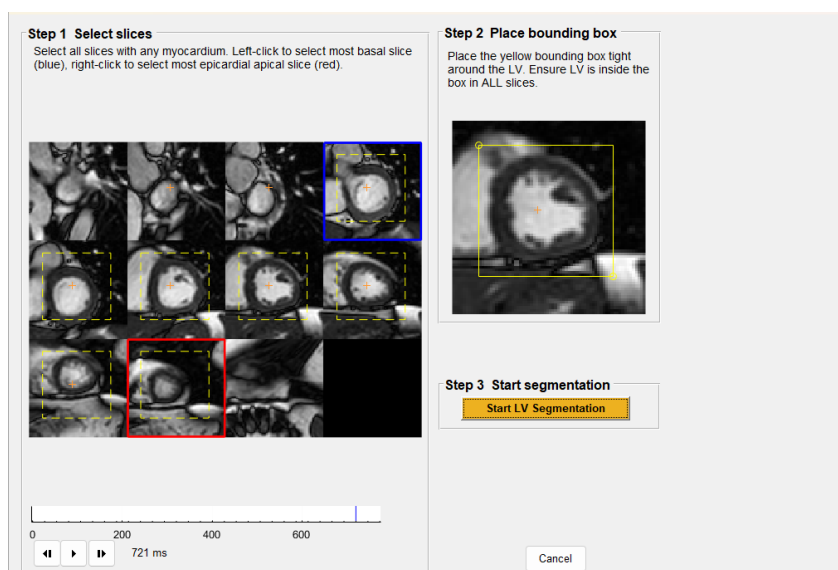


Figura 4 – Ferramenta de segmentação semi-automática do ventrículo esquerdo.

Fonte: Delmondes (2022)

O estudo aborda a criação de um vetor de 18 descritores, como consta do quadro 3, para avaliar exames de ressonância magnética do coração. Estes descritores são calculados por um algoritmo que avalia a distância entre pontos contornados do endocárdio e epicárdio em várias *slice* e momentos do ciclo cardíaco.

- **Descritores C1 a C6:** avaliam o movimento total do endocárdio durante o ciclo cardíaco ou em fases específicas, separadamente por *slice*. Eles são úteis para identificar cardiomiopatias que limitam o movimento do coração.
- **Descritores C7 a C15:** focam no movimento do endocárdio em uma seção específica do ventrículo esquerdo (basal, medial ou apical) durante todo o ciclo cardíaco ou em fases específicas. São úteis para identificar doenças que afetam seções específicas do coração.
- **Descritores C16 a C18:** são calculados com base na diferença de área entre o epicárdio e o endocárdio. Eles são úteis para diferenciar tipos de cardiomiopatias (hipertrófica e dilatada), que podem apresentar características similares nos outros descritores.

4.1.1 Variação do ventrículo esquerdo

A dimensão do ventrículo esquerdo pode apresentar variações com base no gênero e na idade dos indivíduos (KITZMAN; EDWARDS, 1990). Por isso, tornou-se essencial conduzir testes adicionais em conjuntos de dados segmentados para determinar se tais diferenças nas estruturas cardíacas impactam o desempenho do sistema. Assim, os participantes foram organizados em 8 categorias, conforme ilustrado no Quadro 4.

4.2 Seleção dos métodos e métricas

Após uma análise preliminar e compreensão aprofundada do conjunto de dados em questão, optou-se por empregar uma série de algoritmos tradicionais de aprendizagem de máquina para avaliar o desempenho de cada método em relação ao conjunto de dados. O objetivo dessa fase foi identificar o modelo que oferece o melhor desempenho em termos de precisão, velocidade e capacidade de generalização para a aplicação específica.

Descritor	Descrição
C1	Média do deslocamento total por <i>slice</i> no ciclo cardíaco completo.
C2	Desvio padrão do deslocamento total de cada <i>slice</i> no ciclo cardíaco completo.
C3	Média do deslocamento total por <i>slice</i> durante a sístole.
C4	Desvio padrão do deslocamento total por <i>slice</i> durante a sístole.
C5	Média do deslocamento total por <i>slice</i> durante a diástole.
C6	Desvio padrão do deslocamento total por <i>slice</i> durante a diástole.
C7	Deslocamento total do <i>slice</i> basal no ciclo cardíaco completo.
C8	Deslocamento total do <i>slice</i> medial no ciclo cardíaco completo.
C9	Deslocamento total do <i>slice</i> apical no ciclo cardíaco completo.
C10	Deslocamento total do <i>slice</i> basal durante a sístole.
C11	Deslocamento total do <i>slice</i> medial durante a sístole.
C12	Deslocamento total do <i>slice</i> apical durante a sístole.
C13	Deslocamento total do <i>slice</i> basal durante a diástole.
C14	Deslocamento total do <i>slice</i> medial durante a diástole.
C15	Deslocamento total do <i>slice</i> apical durante a diástole.
C16	Média entre todos os <i>slice</i> do volume do miocárdio no fim da diástole.
C17	Média entre todos os <i>slice</i> do volume do miocárdio no meio da sístole.
C18	Média entre todos os <i>slice</i> do volume do miocárdio no fim da sístole.

Quadro 3 – Descrição dos descritores cardíacos

Fonte: Delmondes (2022)

4.2.1 Biblioteca utilizada na pesquisa

Na metodologia deste estudo, foi empregada a biblioteca *PyCaret* (PyCaret Team, 2023) para simplificar e acelerar o processo de modelagem de aprendizado de máquina. *PyCaret* é uma biblioteca de automação de aprendizado de máquina de código aberto em Python que oferece um ambiente de ponta a ponta para desenvolver modelos de aprendizado de máquina de uma forma mais eficiente e menos complexa (PyCaret Team, 2023).

O *PyCaret* é uma biblioteca de automação do fluxo de trabalho de aprendizado de

Classificador	Sexo	intervalo de idade
G1	Masculino	≤ 20 anos
G2	Masculino	≥ 21 anos e ≤ 41 anos
G3	Masculino	≥ 42 anos e ≤ 61 anos
G4	Masculino	≥ 62 anos
G5	Feminino	≤ 20 anos
G6	Feminino	≥ 21 anos e ≤ 41 anos
G7	Feminino	≥ 42 anos e ≤ 61 anos
G8	Feminino	≥ 62 anos

Quadro 4 – Descritores gerados para a classificação dos pacientes por sexo e idade

Fonte: Delmondes (2022)

máquina construída sobre várias bibliotecas existentes, incluindo o *scikit-learn*. O *scikit-learn* fornece uma gama de algoritmos de aprendizado supervisionado e não supervisionado. Ao usar o *scikit-learn*, o *PyCaret* treina modelos, ajusta hiperparâmetros, avalia modelos dentre outros recursos.

PyCaret também oferece uma variedade de modelos e algoritmos pré-configurados que podem ser facilmente chamados para experimentação. Isso facilita a fase de seleção do modelo, no qual diferentes algoritmos são testados para determinar qual oferece o melhor desempenho para o problema em questão. No caso deste estudo, o foco estava em maximizar a métrica *F1_score*, uma métrica escolhida por sua relevância na identificação de doenças cardiovasculares.

Em resumo, a adoção de *PyCaret* na metodologia contribuiu para um fluxo de trabalho de modelagem simplificado e eficiente, permitindo que se concentrasse mais na interpretação dos resultados e na otimização do modelo, ao invés de se preocupar com aspectos técnicos e codificação.

4.2.2 Escolha da Métrica de Avaliação

Ao desenvolver modelos de aprendizado de máquina para identificar doenças cardiovasculares, a escolha da métrica de avaliação representa um passo crítico na metodologia do projeto. Dentre as várias métricas disponíveis, como acurácia, sensibilidade e especificidade, o *F1_score* é frequentemente apontado como a métrica mais adequada para esses casos específicos (SAEED; ONG, 2019).

O *F1_score* é uma média harmônica que considera tanto a precisão (a capacidade

do modelo de identificar verdadeiros positivos em relação a todos os casos positivos identificados) quanto o recall (a habilidade do modelo de identificar verdadeiros positivos em relação a todos os casos positivos reais) (BISHOP, 2022). Isso permite que o *F1_score* capture um equilíbrio entre os custos associados aos falsos positivos e os falsos negativos (SAEED; ONG, 2019).

Portanto, na metodologia adotada para este projeto, embora tenham sido obtidas todas as métricas clássicas empregadas em AM, *F1_score* foi escolhida para ser otimizada no processo de calibração, assim como a principal para comparar os modelos.

4.3 Seleção automática de características

Ao decorrer do presente trabalho, à etapa de seleção de características, tornou-se fundamental no desempenho e eficiência dos modelos de aprendizado de máquina. A seleção adequada de características, além de reduzir a dimensionalidade, também pode aumentar a precisão do modelo e reduzir o tempo de treinamento (BENKESSIRAT; BENBLIDIA, 2019).

Para garantir uma abordagem robusta à seleção de características, foram utilizadas duas técnicas na biblioteca *scikit-learn*: SKB (*SelectKBest*) e SFM (*SelectFromModel*).

A técnica *SelectKBest* seleciona as características a partir de testes estatísticos univariados (DEVELOPERS, 2023). A técnica avalia o impacto de cada atributo no resultado alvo, e as K características mais relevantes são selecionadas. O método proporciona flexibilidade, uma vez que possibilita que os pesquisadores decidam o número exato de características desejadas, possibilitando um controle apurado da dimensionalidade final dos dados (DEVELOPERS, 2023).

Por outra perspectiva, a técnica *SelectFromModel* é uma meta-transformador que seleciona características com base na relevância atribuída a um modelo base (DEVELOPERS, 2023). Em geral, um modelo é desenvolvido com base nos dados, e características consideradas menos relevantes são descartadas (DEVELOPERS, 2023). Essa técnica é, particularmente, vantajosa quando se lida com modelos que atribuem importância ou relevância às suas características, tais como modelos baseados em regulamentação.

Em síntese, a combinação dessas técnicas, obtidas por meio da biblioteca *scikit-learn*, possibilitou a seleção de um conjunto de características otimizadas, garantindo que o

modelo final fosse eficiente tanto em termos computacionais quanto em termos de previsão.

4.4 Fusão de características

O *Principal Component Analysis* (PCA) é uma técnica estatística que simplifica a complexidade de conjuntos de dados de grande tamanho, mantendo o máximo de informação possível. O PCA identifica os eixos principais da variabilidade, ou componentes principais, ao transformar os dados originais em um conjunto menor de combinações lineares. Essas partes capturam as características mais pertinentes do conjunto de dados e mostram as direções nas quais os dados variam mais (CAI; CAI; WEI, 2019).

A princípio, uma análise exploratória foi realizada para entender a estrutura e as relações entre as variáveis do conjunto de dados. Depois, o PCA foi aplicado usando a função específica da biblioteca *PyCaret*. A seleção do número de componentes principais a serem mantidos foi feita com base na variância explicada, um critério que indica quanta informação é retida ao se reduzir a dimensionalidade.

4.5 Calibração

Uma etapa significativa do desenvolvimento de modelos de aprendizado de máquina e inteligência artificial é a calibração de hiperparâmetros. Refere-se ao processo pelo qual os hiperparâmetros de um algoritmo são ajustados e otimizados para melhorar seu desempenho em um determinado conjunto de dados. Estes hiperparâmetros, que podem incluir parâmetros de regularização, número de camadas em redes neurais ou taxas de aprendizado (CHATZILYGEROUDIS; HATZILYGEROUDIS; PERIKOS, 2021). A precisão e a eficácia do modelo podem ser significativamente afetadas pela seleção adequada dos hiperparâmetros (CHATZILYGEROUDIS; HATZILYGEROUDIS; PERIKOS, 2021). Como resultado, é comum que pesquisadores e cientistas de dados usem técnicas como *grid seach* para examinar diversas combinações de hiperparâmetros e descobrir a configuração mais eficaz (CHATZILYGEROUDIS; HATZILYGEROUDIS; PERIKOS, 2021).

Neste trabalho foi escolhido o *F1_score* para auxiliar na descoberta dos hiperparâmetros que trarão o melhor desempenho para o modelo. Esta etapa de otimização visa encontrar o conjunto de hiperparâmetros que maximiza a capacidade do modelo de fazer previsões precisas em dados não vistos, sem cair no problema de *overfitting* ou *underfitting*.

A estratégia utilizada para a calibração foi a pesquisa em grade ou “*grid search*”, que consiste em testar uma série de combinações de valores e selecionar aqueles resultantes no melhor *F1_score*. Este processo foi automatizado para garantir eficiência e precisão, e as métricas de desempenho foram validadas mediante validação cruzada com a utilização do fold igual a 5 e uma divisão da base de dados de treino e teste na proporção de 70 para treino e 30 para teste, a base de dados foi estratificada em relação às classes, a fim de garantir a robustez do modelo. Após a calibração, o modelo otimizado foi avaliado em um conjunto de dados de teste separado para validar seu desempenho.

4.6 Coleta das métricas finais

Após executar todas as etapas, comparou-se os *F1_score* para cada modelo de aprendizado de máquina testado. Para as classificações binárias calculou-se a média e desvio padrão para cada métrica coletada, posteriormente analisou-se a matriz de confusão e a curva ROC para entender a forma que o modelo se comportou na classificação de cada classe. Esta fase significativa da avaliação objetiva-se a identificar qual dos métodos examinados apresenta o melhor desempenho, conforme os critérios rigorosos estabelecidos.

4.7 Considerações finais do capítulo

Neste capítulo, descreveu-se e analisou-se a metodologia empregada neste projeto de pesquisa. A essência deste trabalho reside em explorar e entender o comportamento dos modelos de AM na tarefa específica de classificação de cardiomiopatias. Para atingir o objetivo, seguiu-se rigorosamente a metodologia aqui delineada, garantindo uma abordagem sistemática e criteriosa. Este rigor metodológico permitiu uma avaliação mais precisa da eficácia dos modelos em questão. No capítulo 5, serão apresentados os achados e resultados alcançados ao longo deste estudo, evidenciando as contribuições práticas e teóricas do presente trabalho.

5 RESULTADOS E DISCUSSÕES

Neste capítulo, será apresentado e discutido os resultados obtidos após a implementação da metodologia proposta no capítulo 4. Na seção 5.1 será apresentado os resultados obtidos com a classificação binária. Na seção 5.2 é apresentado os resultados da classificação multiclass. A base de dados que será usada é a descrita na seção 4.1.

Após uma análise detalhada pode-se perceber que o ventrículo esquerdo varia considerando alguns parâmetros, como descrito na seção 4.1.1. Dessa forma, os modelos foram novamente treinados usando em conjunto os descritores dos quadros 3 e 4, no intuito de entender como o sexo biológico e a idade podem impactar na detecção de cardiomiopatias.

Durante este capítulo foi utilizado o $F1_score$ como métrica para elencar o método com o melhor desempenho. Na primeira rodada de testes cada base de dados foi processado quatro vezes, coletou-se o $F1_score$ de cada resultado dos modelos KNN, LR, XGBOOST, RF, NB, GBC (*Gradient Boosting Classifier*), DT, SVM, ADA (*Ada Boost Classifier*), RBFSVM foram classificados entre base de dados original, base de dados após ser aplicado PCA, após ser aplicado SAC (seleção automática de características) SFM e SKB. Como mencionado na seção 4.2.2, a métrica $F1_score$ foi escolhida para selecionar o modelo com o melhor desempenho.

Uma vez realizada a classificação binária, foi conduzida a classificação multiclasse, considerando as três classes definidas (Hipertrofico, Dilatado, Normal) e empregando os mesmos modelos utilizados na classificação binária.

Para todas as tabelas deste capítulo considerou-se as quatro bases de dados descritas a baixo.

- **Original:** considerou-se as 18 características descritas no quadro 3.
- **PCA:** base de dados original com a aplicação da técnica descrita na seção 4.4.
- **SAC SKB:** base de dados original após a aplicação de SAC SKB descrita na seção 4.3.

- **SAC SFM:** base de dados original após a aplicação de SAC SFM descrita na seção 4.3.

5.1 Classificação Binária

Essa seção é dividida em duas subseções, na subseção 5.1.1 consideraram-se apenas os descritores dispostos no quadro 3. Posteriormente, na subseção 5.1.2 consideraram-se os descritores dispostos nos quadros 3 e os grupos dispostos no quadro 4.

5.1.1 Classificação sem segmentação de grupos

Com o treinamento dos modelos para a classificação binários, tem-se a Tabela 1, na tentativa de melhorar os resultados obtidos foi realizada uma calibração de hiperparâmetros, os resultados obtidos após a calibração estão demonstrados na Tabela 2. Os hiperparâmetros após a calibração constam nos quadros 5, 6, 7, 8, 9 e 10.

Modelos	Original	PCA	SAC SKB	SAC SFM
ADA	$0,7 \pm 0,07$	$0,72 \pm 0,09$	$0,71 \pm 0,1$	$0,72 \pm 0,08$
NB	$0,66 \pm 0,12$	$0,66 \pm 0,05$	$0,67 \pm 0,08$	$0,68 \pm 0,09$
GBC	$0,75 \pm 0,09$	$0,72 \pm 0,08$	$0,71 \pm 0,1$	$0,78 \pm 0,07$
XGBOOST	$0,74 \pm 0,09$	$0,72 \pm 0,05$	$0,69 \pm 0,1$	$0,74 \pm 0,07$
KNN	$0,69 \pm 0,08$	$0,68 \pm 0,11$	$0,73 \pm 0,08$	$0,7 \pm 0,1$
LR	$0,76 \pm 0,1$	$0,71 \pm 0,09$	$0,72 \pm 0,14$	$0,81 \pm 0,1$
RBFSVM	$0,43 \pm 0,05$	$0,43 \pm 0,05$	$0,43 \pm 0,05$	$0,43 \pm 0,05$
SVM	$0,59 \pm 0,11$	$0,66 \pm 0,11$	$0,63 \pm 0,19$	$0,66 \pm 0,08$
RF	$0,75 \pm 0,09$	$0,71 \pm 0,09$	$0,71 \pm 0,09$	$0,77 \pm 0,06$
DT	$0,69 \pm 0,04$	$0,67 \pm 0,05$	$0,67 \pm 0,11$	$0,73 \pm 0,12$

Tabela 1 – Resultados da análise sem a calibração

A Tabela 3 apresenta métricas de avaliação para diferentes modelos de AM aplicados as quatro bases de dados utilizadas neste trabalho. Observa-se que o modelo de LR treinado nas bases “SAC SFM” e “original” demonstra um desempenho superior, com métricas consistentemente em todas as categorias em comparação aos outros modelos da Tabela 3, seja com calibração (CC) ou sem (SC). Enquanto isso, o modelo XGBOOST na base “PCA SC” apresenta valores inferiores em comparação com sua variante “PCA CC”. Além disso, os modelos KNN e ADA na base “SAC SKB” exibem métricas similares entre suas versões com e sem calibração. O modelo GBC em “CC” mostra desempenho similar em comparação com KNN e ADA de mesma base de dados. Por fim, o modelo LR conseguiu

Modelos	Original	PCA	SAC SKB	SAC SFM
ADA	$0,74 \pm 0,09$	$0,73 \pm 0,08$	$0,73 \pm 0,08$	$0,74 \pm 0,06$
NB	$0,66 \pm 0,12$	$0,66 \pm 0,05$	$0,67 \pm 0,08$	$0,68 \pm 0,09$
GBC	$0,75 \pm 0,09$	$0,72 \pm 0,07$	$0,73 \pm 0,08$	$0,78 \pm 0,07$
XGBOOST	$0,76 \pm 0,07$	$0,73 \pm 0,06$	$0,71 \pm 0,12$	$0,77 \pm 0,08$
KNN	$0,73 \pm 0,07$	$0,69 \pm 0,11$	$0,73 \pm 0,08$	$0,71 \pm 0,09$
LR	$0,77 \pm 0,09$	$0,71 \pm 0,09$	$0,73 \pm 0,14$	$0,81 \pm 0,1$
RBFSVM	$0,68 \pm 0,17$	$0,68 \pm 0,16$	$0,72 \pm 0,11$	$0,7 \pm 0,18$
SVM	$0,74 \pm 0,08$	$0,68 \pm 0,09$	$0,72 \pm 0,14$	$0,78 \pm 0,07$
RF	$0,75 \pm 0,07$	$0,72 \pm 0,09$	$0,73 \pm 0,1$	$0,77 \pm 0,06$
DT	$0,73 \pm 0,08$	$0,68 \pm 0,03$	$0,69 \pm 0,1$	$0,77 \pm 0,07$

Tabela 2 – Resultados da análise com a calibração

Modelo	Base de dados	acurácia	AUC	Revocação	Precisão	F1
LR	SAC SFM CC	$0,82 \pm 0,09$	$0,86 \pm 0,08$	$0,82 \pm 0,09$	$0,82 \pm 0,09$	$0,81 \pm 0,1$
LR	SAC SFM SC	$0,82 \pm 0,09$	$0,86 \pm 0,08$	$0,82 \pm 0,09$	$0,82 \pm 0,09$	$0,81 \pm 0,1$
XGBOOST	PCA CC	$0,73 \pm 0,06$	$0,76 \pm 0,07$	$0,73 \pm 0,06$	$0,73 \pm 0,06$	$0,73 \pm 0,06$
XGBOOST	PCA SC	$0,72 \pm 0,05$	$0,76 \pm 0,07$	$0,72 \pm 0,05$	$0,72 \pm 0,06$	$0,72 \pm 0,05$
KNN	SAC SKB SC	$0,73 \pm 0,07$	$0,74 \pm 0,11$	$0,73 \pm 0,07$	$0,73 \pm 0,08$	$0,73 \pm 0,08$
KNN	SAC SKB CC	$0,73 \pm 0,08$	$0,75 \pm 0,11$	$0,73 \pm 0,08$	$0,74 \pm 0,09$	$0,73 \pm 0,08$
ADA	SAC SKB CC	$0,73 \pm 0,08$	$0,77 \pm 0,11$	$0,73 \pm 0,08$	$0,73 \pm 0,08$	$0,73 \pm 0,08$
GBC	SAC SKB CC	$0,71 \pm 0,1$	$0,76 \pm 0,15$	$0,71 \pm 0,1$	$0,71 \pm 0,1$	$0,71 \pm 0,1$
LR	original SC	$0,76 \pm 0,09$	$0,81 \pm 0,1$	$0,76 \pm 0,09$	$0,77 \pm 0,1$	$0,76 \pm 0,1$
LR	original CC	$0,77 \pm 0,09$	$0,83 \pm 0,09$	$0,77 \pm 0,09$	$0,77 \pm 0,1$	$0,77 \pm 0,09$

Tabela 3 – Métricas dos melhores modelos (CC = com calibração; SC = sem calibração)

desempenhar superiormente, tanto com ou sem calibração, a todos os outros modelos em todas as métricas apresentadas na Tabela 3.

Segundo a Tabela 1 os modelos com melhor $F1_score$ são LR (base de dados original), XGBOOST (base de dados após a aplicação PCA), KNN (base de dados após

Classificação	Parâmetros e Valores
Normal;Hipertrófico	C=1, class_weight=None, dual=False, fit_intercept=True, intercept_scaling=1, l1_ratio=None, max_iter=1000, multi_class='auto', n_jobs=None, penalty='l2', random_state=800733, solver='saga', tol=0.0001, verbose=0, warm_start=False
Normal;Dilatado	C=0.1, class_weight=None, dual=False, fit_intercept=True, intercept_scaling=1, l1_ratio=None, max_iter=1000, multi_class='auto', n_jobs=None, penalty='l1', random_state=347594, solver='saga', tol=0.0001, verbose=0, warm_start=False
Hipertrófico;Dilatado	C=1, class_weight=None, dual=False, fit_intercept=True, intercept_scaling=1, l1_ratio=None, max_iter=1000, multi_class='auto', n_jobs=None, penalty='l1', random_state=28661, solver='liblinear', tol=0.0001, verbose=0, warm_start=False

Quadro 5 – Hiperparâmetros após a calibração do modelo LR original.

a aplicação SAC SKB) e LR (base de dados após a aplicação SAC SFM), comparando os resultados obtidos o modelo com maior *F1_score* é LR (SAC SFM) como descrito na Tabela 1, comparando com a base de dados original houve um aumento no valor aproximadamente de 6,58%.

Com a calibração dos modelos, o modelo LR ainda apresenta o maior *F1_score* na base de dados original (Tabela 2) com um aumento de *F1_score* no valor aproximadamente de 1,32%, tal *F1_score* foi alcançado com os hiperparâmetros demonstrados no quadro 10. A base de dados após a aplicação de PCA não houve uma alteração do modelo que apresenta o melhor desempenho, permaneceu sendo o XGBOOST (Tabela 2) com um aumento no *F1_score* no valor aproximadamente de 1,39% e hiperparâmetros que constam no quadro 6.

A análise cruzada da Tabela 1 e 2 não houve aumento do *F1_score* na base de dados após a aplicação de SAC SFM como também não houve alteração do melhor modelo, contudo a base de dados após a aplicação de SAC SKB o modelo com melhor desempenho deixou de ser KNN e passou a ser GBC, ADA e KNN o *F1_score* permaneceu no valor de 0,73 para todos os modelos, com os hiperparâmetros respectivamente nos quadros 8, 7 e 9.

5.1.1.1 Análise do modelo com melhor desempenho

A partir da comparação das Tabelas 1 e 2 o modelo que apresentou o melhor *F1_score* foi o modelo LR no valor 0,81 e desvio padrão $\pm 0,01$. A matriz de confusão

Classificação	Parâmetros e Valores
Normal;Hipertrófico	<pre> objective=binary:logistic; base_score=None; booster=gbtree; callbacks=None; colsample_bylevel=None; colsample_bynode=None; colsample_bytree=0.9; device=None; early_stopping_rounds=None; enable_categorical=False; eval_metric=None; feature_types=None; gamma=None; grow_policy=None; importance_type=None; interaction_constraints=None; learning_rate=0.5; max_bin=None; max_cat_threshold=None; max_cat_to_onehot=None; max_delta_step=None; max_depth=10; max_leaves=None; min_child_weight=2; missing=nan; monotone_constraints=None; multi_strategy=None; n_estimators=10; n_jobs=-1; num_parallel_tree=None; random_state=776509; reg_alpha=None; reg_lambda=None; sampling_method=None; scale_pos_weight=None; subsample=1.0; tree_method=auto; validate_parameters=None; verbosity=0 </pre>
Normal;Dilatado	<pre> objective=binary:logistic; base_score=None; booster=gbtree; callbacks=None; colsample_bylevel=None; colsample_bynode=None; colsample_bytree=None; device=None; early_stopping_rounds=None; enable_categorical=False; eval_metric=None; feature_types=None; gamma=None; grow_policy=None; importance_type=None; interaction_constraints=None; learning_rate=None; max_bin=None; max_cat_threshold=None; max_cat_to_onehot=None; max_delta_step=None; max_depth=None; max_leaves=None; min_child_weight=None; missing=nan; monotone_constraints=None; multi_strategy=None; n_estimators=None; n_jobs=-1; num_parallel_tree=None; random_state=364619; reg_alpha=None; reg_lambda=None; sampling_method=None; scale_pos_weight=None; subsample=None; tree_method=auto; validate_parameters=None; verbosity=0 </pre>
Hipertrófico;Dilatado	<pre> objective=binary:logistic; base_score=None; booster=gbtree; callbacks=None; colsample_bylevel=None; colsample_bynode=None; colsample_bytree=1; device=None; early_stopping_rounds=None; enable_categorical=False; eval_metric=None; feature_types=None; gamma=None; grow_policy=None; importance_type=None; interaction_constraints=None; learning_rate=0.01; max_bin=None; max_cat_threshold=None; max_cat_to_onehot=None; max_delta_step=None; max_depth=7; max_leaves=None; min_child_weight=3; missing=nan; monotone_constraints=None; multi_strategy=None; n_estimators=100; n_jobs=-1; num_parallel_tree=None; random_state=61431; reg_alpha=None; reg_lambda=None; sampling_method=None; scale_pos_weight=None; subsample=0.9; tree_method=auto; validate_parameters=None; verbosity=0 </pre>

Quadro 6 – Hiperparâmetros após a calibração do modelo XGBOOST PCA

(Figura 5) demonstra a dificuldade em diferenciar uma cardiomiopatia dilatado de uma hipertrófica. No entanto, no desempenho da distinção de cardiomiopatia hipertrófica de um coração normal (Figura 9) não há dificuldades em diferenciá-los, assim como um coração normal de um que apresenta cardiomiopatia dilatada (Figura 7).

Classificação	Parâmetros e Valores
Normal;Hipertrófico	algorithm='SAMME.R', base_estimator='deprecated', estimator=None, learning_rate=0.01, n_estimators=150, random_state=29857, class=nxh
Normal;Dilatado	algorithm='SAMME', base_estimator='deprecated', estimator=None, learning_rate=0.5, n_estimators=150, random_state=394043, class=nxd
Hipertrófico;Dilatado	algorithm='SAMME.R', base_estimator='deprecated', estimator=None, learning_rate=0.5, n_estimators=100, random_state=114357, class=hxd

Quadro 7 – Hiperparâmetros após a calibração do modelo ADA SAC SKB

Classificação	Parâmetros e Valores
Normal;Hipertrófico	algorithm='SAMME.R', base_estimator='deprecated', estimator=None, learning_rate=0.01, n_estimators=150, random_state=29857
Normal;Dilatado	algorithm='SAMME', base_estimator='deprecated', estimator=None, learning_rate=0.5, n_estimators=150, random_state=394043
Hipertrófico;Dilatado	algorithm='SAMME.R', base_estimator='deprecated', estimator=None, learning_rate=0.5, n_estimators=100, random_state=114357

Quadro 8 – Hiperparâmetros após a calibração do modelo GBC SAC SKB

Classificação	Parâmetros e Valores
Normal;Hipertrófico	algorithm='auto', leaf_size=30, metric='euclidean', metric_params=None, n_jobs=-1, n_neighbors=10, p=2, weights='distance'
Normal;Dilatado	algorithm='auto', leaf_size=30, metric='minkowski', metric_params=None, n_jobs=-1, n_neighbors=22, p=1, weights='uniform'
Hipertrófico;Dilatado	algorithm='auto', leaf_size=30, metric='euclidean', metric_params=None, n_jobs=-1, n_neighbors=24, p=2, weights='uniform'

Quadro 9 – Hiperparâmetros após a calibração do modelo KNN SAC SKB

Classificação	Parâmetros e Valores
Normal;Hipertrófico	C=1.0, class_weight=None, dual=False, fit_intercept=True, intercept_scaling=1, l1_ratio=None, max_iter=1000, multi_class='auto', n_jobs=None, penalty='l2', random_state=260995, solver='lbfgs', tol=0.0001, verbose=0, warm_start=False, class=nxh
Normal;Dilatado	C=1.0, class_weight=None, dual=False, fit_intercept=True, intercept_scaling=1, l1_ratio=None, max_iter=1000, multi_class='auto', n_jobs=None, penalty='l2', random_state=640002, solver='lbfgs', tol=0.0001, verbose=0, warm_start=False, class=nxd
Hipertrófico;Dilatado	C=1.0, class_weight=None, dual=False, fit_intercept=True, intercept_scaling=1, l1_ratio=None, max_iter=1000, multi_class='auto', n_jobs=None, penalty='l2', random_state=5827, solver='lbfgs', tol=0.0001, verbose=0, warm_start=False, class=hxd

Quadro 10 – Hiperparâmetros após a calibração do modelo LR SAC SFM

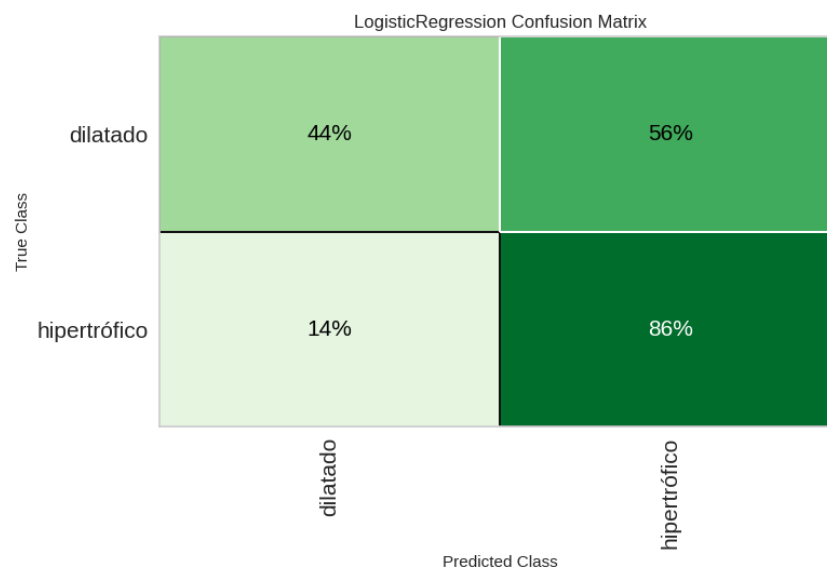


Figura 5 – Matriz de confusão do modelo LR com aplicação de SAC SFM e realização de calibração

O modelo LR tem um desempenho aceitável, contudo demonstra dificuldades em distinguir uma condição dilatada de uma hipertrófica (Figura 6) com um AUC de 0,75, de outra forma tem um AUC muito de 0,95 quando se trata de dilatado e normal (Figura 8) e por fim um AUC de 0,87 quando se trata de hipertrófico e normal (Figura 10). Quando analisado a média-micro e média-macro de todos as bases de dados, percebe-se um desempenho estável em todas as vezes que o modelo foi treinado.

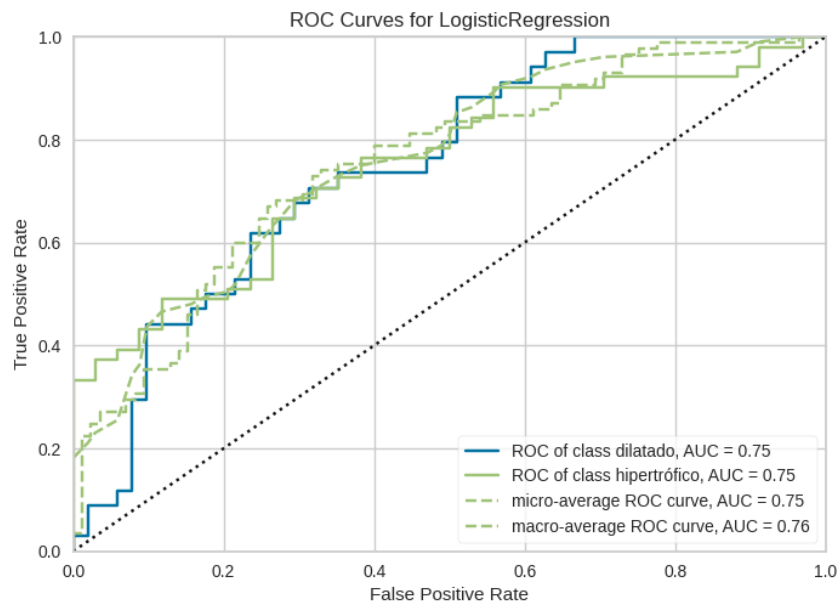


Figura 6 – Curva ROC do modelo LR com aplicação de SAC SFM e realização de calibração

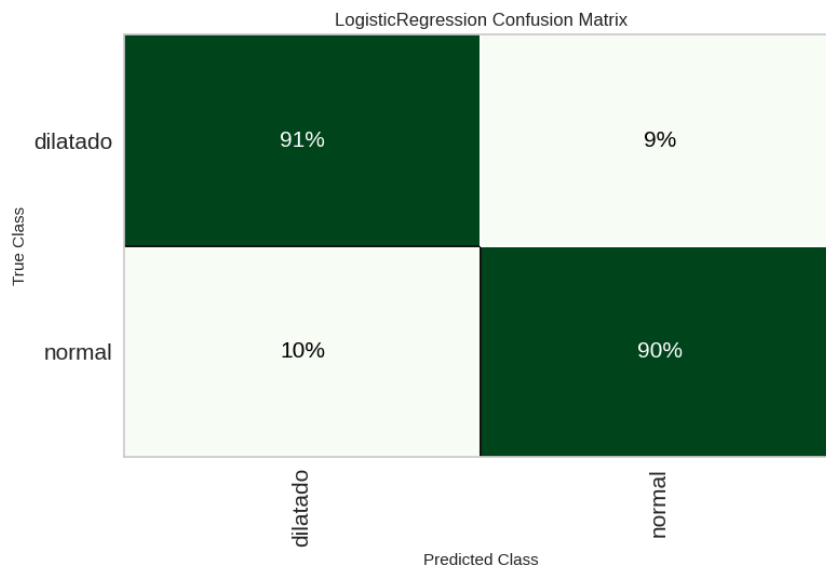


Figura 7 – Matriz de confusão do modelo LR com aplicação de SAC SFM e realização de calibração

5.1.2 Classificação com a base de dados segmentada

Ao decorrer desta seção será exposto os resultados obtidos considerando também os descritores no quadro 4, com a adição destas características foi realizado mais uma rodada de testes seguindo o mesmo método descrito na seção 5.1. Os resultados obtidos constam nas Tabelas 4 e 5. A Tabela 5 consta os modelos após a calibração e os melhores hiperparâmetros encontrados contam nos quadros 14, 13, 12 e 11.

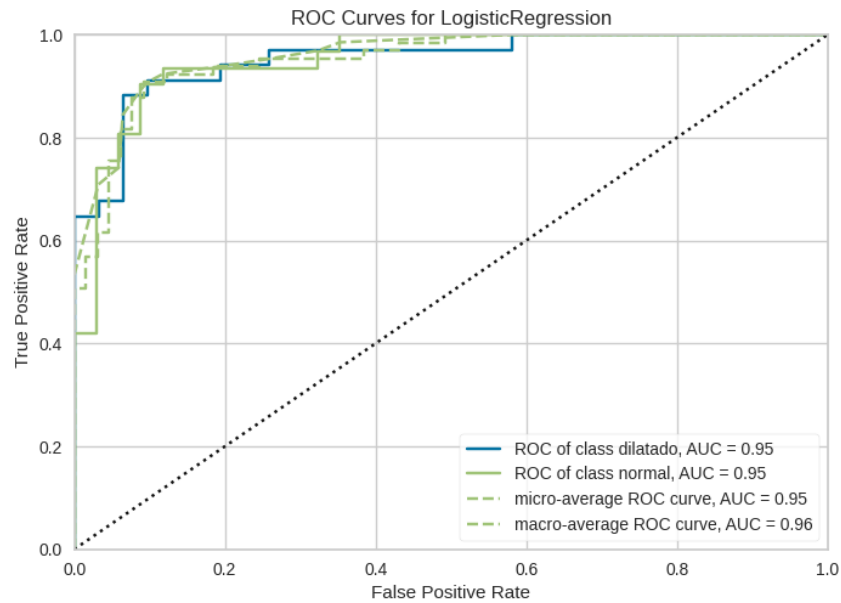


Figura 8 – Curva ROC do modelo LR com aplicação de SAC SFM e realização de calibração

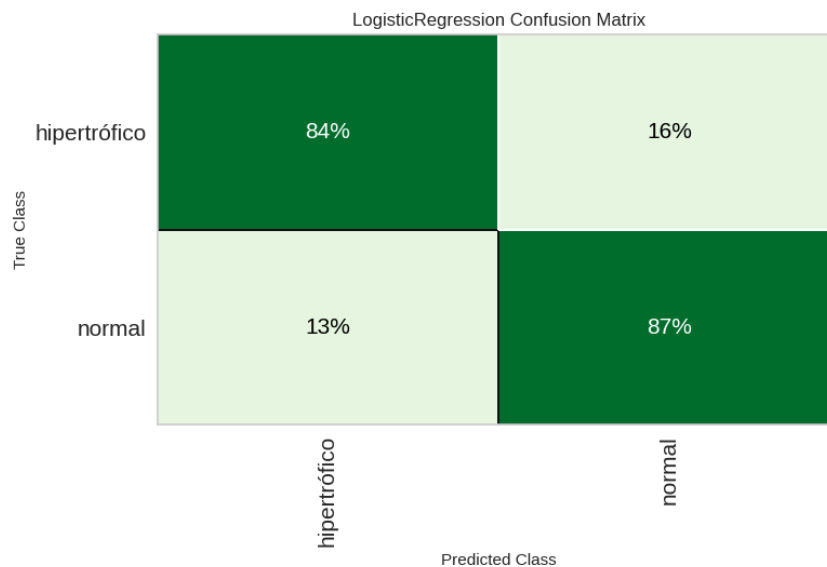


Figura 9 – Matriz de confusão do modelo LR com aplicação de SAC SFM e realização de calibração

Com a Tabela 4 os modelos com melhor desempenho foram XGBOOST (base de dados original), NB (base de dados após a aplicação de PCA), RF (base de dados após a aplicação de SAC SKB) e XGBOOST (base de dados após a aplicação de SAC SFM). Após a calibração não houve alteração dos melhores modelos na base de dados original nem na base de dados após a aplicação de PCA, contudo houve um aumento de 1,23% tanto para SAC SKB quanto SAC SFM.

A Tabela 6 apresenta as métricas dos melhores modelos de cada base de dados das

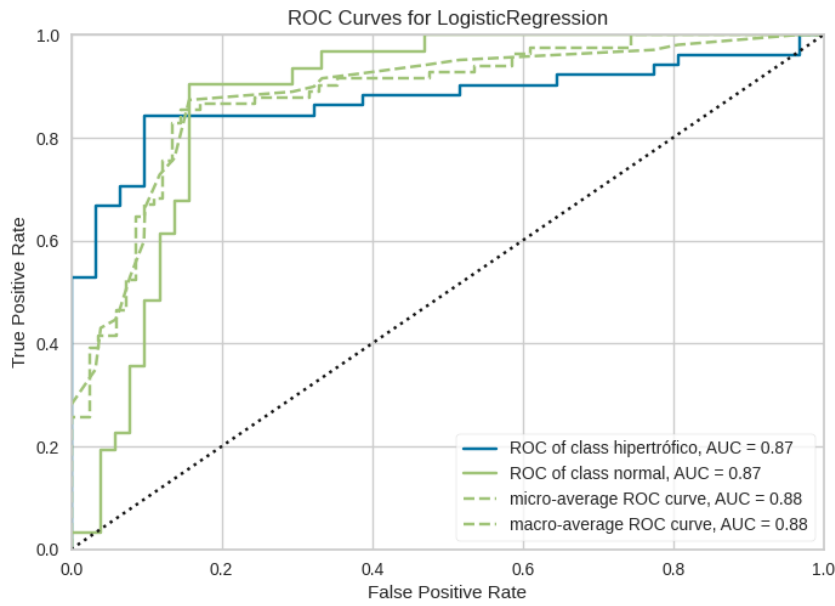


Figura 10 – Curva ROC do modelo LR com aplicação de SAC SFM e realização de calibração

Modelos	Original	PCA	SAC SKB	SAC SFM
ADA	0,81 $\pm$ 0,11	0,67 $\pm$ 0,1	0,74 $\pm$ 0,1	0,8 $\pm$ 0,12
NB	0,67 $\pm$ 0,1	0,76 $\pm$ 0,06	0,65 $\pm$ 0,1	0,7 $\pm$ 0,03
GBC	0,81 $\pm$ 0,13	0,71 $\pm$ 0,12	0,78 $\pm$ 0,11	0,78 $\pm$ 0,09
XGBOOST	0,83 $\pm$ 0,11	0,71 $\pm$ 0,09	0,81 $\pm$ 0,1	0,81 $\pm$ 0,11
KNN	0,65 $\pm$ 0,09	0,7 $\pm$ 0,08	0,7 $\pm$ 0,07	0,7 $\pm$ 0,09
LR	0,76 $\pm$ 0,05	0,71 $\pm$ 0,05	0,69 $\pm$ 0,13	0,75 $\pm$ 0,07
RBFSVM	0,43 $\pm$ 0,05	0,43 $\pm$ 0,05	0,43 $\pm$ 0,05	0,43 $\pm$ 0,05
SVM	0,65 $\pm$ 0,1	0,71 $\pm$ 0,06	0,62 $\pm$ 0,1	0,78 $\pm$ 0,09
RF	0,79 $\pm$ 0,12	0,74 $\pm$ 0,06	0,79 $\pm$ 0,11	0,78 $\pm$ 0,12
DT	0,75 $\pm$ 0,14	0,69 $\pm$ 0,05	0,77 $\pm$ 0,12	0,79 $\pm$ 0,09

Tabela 4 – Resultados da análise sem a calibração

Tabelas 4 e 5. O modelo XGBOOST apresenta um desempenho superior em comparação aos outros modelos descritos na Tabela 6, especialmente nas bases “original” e “SAC SKB”, independentemente da calibração. O modelo NB na base “PCA” tem um desempenho inferior em comparação aos demais modelos. Por sua vez, o modelo ADA, quando aplicado à base “SAC SFM”, demonstra métricas similares em comparação com as métricas do XGBOOST. Em suma, o XGBOOST se destaca como um modelo com desempenho consistente em várias bases de dados, enquanto os outros modelos exibem performances variadas dependendo da base e da presença de calibração.

Modelos	Original	PCA	SAC SKB	SAC SFM
ADA	$0,82 \pm 0,12$	$0,7 \pm 0,06$	$0,82 \pm 0,11$	$0,81 \pm 0,12$
NB	$0,67 \pm 0,11$	$0,76 \pm 0,06$	$0,65 \pm 0,1$	$0,72 \pm 0,03$
GBC	$0,81 \pm 0,13$	$0,75 \pm 0,07$	$0,8 \pm 0,13$	$0,78 \pm 0,09$
XGBOOST	$0,83 \pm 0,11$	$0,72 \pm 0,08$	$0,81 \pm 0,1$	$0,82 \pm 0,11$
KNN	$0,68 \pm 0,1$	$0,72 \pm 0,1$	$0,73 \pm 0,07$	$0,72 \pm 0,12$
LR	$0,79 \pm 0,05$	$0,72 \pm 0,06$	$0,72 \pm 0,09$	$0,77 \pm 0,08$
RBFSVM	$0,68 \pm 0,15$	$0,66 \pm 0,14$	$0,67 \pm 0,16$	$0,68 \pm 0,16$
SVM	$0,74 \pm 0,06$	$0,72 \pm 0,07$	$0,69 \pm 0,16$	$0,78 \pm 0,09$
RF	$0,79 \pm 0,12$	$0,74 \pm 0,06$	$0,79 \pm 0,11$	$0,8 \pm 0,11$
DT	$0,77 \pm 0,16$	$0,69 \pm 0,05$	$0,79 \pm 0,12$	$0,82 \pm 0,06$

Tabela 5 – Resultados da análise com a calibração

Modelo	Base de dados	acurácia	AUC	Revocação	Precisão	F1
XGBOOST	original SC	$0,82 \pm 0,11$	$0,86 \pm 0,13$	$0,82 \pm 0,11$	$0,83 \pm 0,11$	$0,83 \pm 0,11$
XGBOOST	original CC	$0,82 \pm 0,11$	$0,86 \pm 0,13$	$0,82 \pm 0,11$	$0,83 \pm 0,11$	$0,83 \pm 0,11$
NB	PCA SC	$0,72 \pm 0,12$	$0,78 \pm 0,1$	$0,72 \pm 0,12$	$0,76 \pm 0,09$	$0,72 \pm 0,12$
NB	PCA CC	$0,72 \pm 0,12$	$0,78 \pm 0,1$	$0,72 \pm 0,12$	$0,76 \pm 0,09$	$0,72 \pm 0,12$
XGBOOST	SAC SKB SC	$0,8 \pm 0,09$	$0,83 \pm 0,12$	$0,8 \pm 0,09$	$0,8 \pm 0,1$	$0,8 \pm 0,1$
ADA	SAC SKB CC	$0,82 \pm 0,1$	$0,84 \pm 0,14$	$0,82 \pm 0,1$	$0,83 \pm 0,11$	$0,82 \pm 0,11$
XGBOOST	SAC SFM SC	$0,82 \pm 0,1$	$0,86 \pm 0,11$	$0,82 \pm 0,1$	$0,82 \pm 0,11$	$0,81 \pm 0,11$
ADA	SAC SFM CC	$0,82 \pm 0,06$	$0,82 \pm 0,07$	$0,82 \pm 0,06$	$0,82 \pm 0,06$	$0,82 \pm 0,06$

Tabela 6 – Métricas dos melhores modelos (CC = com calibração; SC = sem calibração)

5.1.2.1 Análise do modelo com melhor desempenho

Analisando as Tabelas 4 e 5 o modelo que apresentou o melhor desempenho foi o XGBOOST (Tabela 5), que não teve aumento de desempenho após a calibração, com um $F1_score$ no valor de 0,83 e desvio padrão 0,11.

Apesar de apresentar um $F1_score$ no valor de 0,83, ainda assim o modelo XGBOOST tem uma dificuldade de diferenciar uma cardiomiopatia dilatada de uma hipertrófica (Figura 14), por outro lado, consegue diferenciar com satisfação um coração normal de

Classificação	Parâmetros e Valores
Normal;Hipertrófico	ccp_alpha=0.0, class_weight=None, criterion='gini', max_depth=5, max_features=None, max_leaf_nodes=30, min_impurity_decrease=0.0, min_samples_leaf=4, min_samples_split=10, min_weight_fraction_leaf=0.0, random_state=111258, splitter='best'
Normal;Dilatado	ccp_alpha=0.0, class_weight=None, criterion='gini', max_depth=None, max_features=None, max_leaf_nodes=None, min_impurity_decrease=0.0, min_samples_leaf=1, min_samples_split=2, min_weight_fraction_leaf=0.0, random_state=508836, splitter='best'
Hipertrófico;Dilatado	ccp_alpha=0.0, class_weight=None, criterion='gini', max_depth=5, max_features='log2', max_leaf_nodes=10, min_impurity_decrease=0.0, min_samples_leaf=4, min_samples_split=10, min_weight_fraction_leaf=0.0, random_state=81683, splitter='best'

Quadro 11 – Hiperparâmetros após a calibração do modelo DT SAC SFM

Classificação	Parâmetros e Valores
Normal; Hipertrófico	algorithm='SAMME', base_estimator='deprecated', estimator=None, learning_rate=0.5, n_estimators=100, random_state=702777
Normal;Dilatado	algorithm='SAMME.R', base_estimator='deprecated', estimator=None, learning_rate=0.05, n_estimators=150, random_state=398396
Hipertrófico; dilatado	algorithm='SAMME.R', base_estimator='deprecated', estimator=None, learning_rate=0.5, n_estimators=10, random_state=526913

Quadro 12 – Hiperparâmetros após a calibração do modelo ADA SAC SKB

Classificação	Parâmetros e Valores
Normal; Hipertrófico	priors=None, var_smoothing=1e-09
Normal;Dilatado	priors=None, var_smoothing=1e-09
Hipertrófico; dilatado	priors=None, var_smoothing=1e-09

Quadro 13 – Hiperparâmetros após a calibração do modelo NB PCA

um hipertrófico ou de um dilatado como demonstrado nas Figuras 15 e 11. Dentre todos os treinamentos do modelo XGBOOST o que teve o melhor desempenho foi o que está descrito na Figura 15 tendo uma taxa erro não relevante para o modelo.

No gráfico apresentado na Figura 16, indica uma alta taxa de verdadeiros positivos e uma baixa taxa de falsos positivos. Tal desempenho permanece ao analisar a Figura 12, apesar da pequena queda no desempenho. Já a Figura 14 apresenta a dificuldade do modelo em distinguir dilatado de hipertrófico, ficando um pouco superior a uma previsão aleatória.

Classificação	Parâmetros e Valores
Normal; Hipertrófico	binary:logistic ,base_score=None, booster=gbtree, callbacks=None, colsample_bylevel=None, colsample_bynode=None, colsample_bytree=0.9, device=None, early_stopping_rounds=None, enable_categorical=False, eval_metric=None, feature_types=None, gamma=None, grow_policy=None, importance_type=None, interaction_constraints=None, learning_rate=0.5, max_bin=None, max_cat_threshold=None, max_cat_to_onehot=None, max_delta_step=None, max_depth=10 max_leaves=None, min_child_weight=2, missing=nan, monotone_constraints=None, multi_strategy=None, n_estimators=10, n_jobs=-1, num_parallel_tree=None, random_state=776509, reg_alpha=None, reg_lambda=None, sampling_method=None, scale_pos_weight=None, subsample=1.0, tree_method=auto, validate_parameters=None, verbosity=0
Normal;Dilatado	binary:logistic ,base_score=None, booster=gbtree, callbacks=None, colsample_bylevel=None, colsample_bynode=None, colsample_bytree=0.9, device=None, early_stopping_rounds=None, enable_categorical=False, eval_metric=None, feature_types=None, gamma=None, grow_policy=None, importance_type=None, interaction_constraints=None, learning_rate=0.5, max_bin=None, max_cat_threshold=None, max_cat_to_onehot=None, max_delta_step=None, max_depth=10 max_leaves=None, min_child_weight=2, missing=nan, monotone_constraints=None, multi_strategy=None, n_estimators=10, n_jobs=-1, num_parallel_tree=None, random_state=776509, reg_alpha=None, reg_lambda=None, sampling_method=None, scale_pos_weight=None, subsample=1.0, tree_method=auto, validate_parameters=None, verbosity=0
Hipertrófico; dilatado	binary:logistic ,base_score=None, booster=gbtree, callbacks=None, colsample_bylevel=None, colsample_bynode=None, colsample_bytree=0.9, device=None, early_stopping_rounds=None, enable_categorical=False, eval_metric=None, feature_types=None, gamma=None, grow_policy=None, importance_type=None, interaction_constraints=None, learning_rate=0.5, max_bin=None, max_cat_threshold=None, max_cat_to_onehot=None, max_delta_step=None, max_depth=10 max_leaves=None, min_child_weight=2, missing=nan, monotone_constraints=None, multi_strategy=None, n_estimators=10, n_jobs=-1, num_parallel_tree=None, random_state=776509, reg_alpha=None, reg_lambda=None, sampling_method=None, scale_pos_weight=None, subsample=1.0, tree_method=auto, validate_parameters=None, verbosity=0

Quadro 14 – Hiperparâmetros após a calibração do modelo XGBOOST original

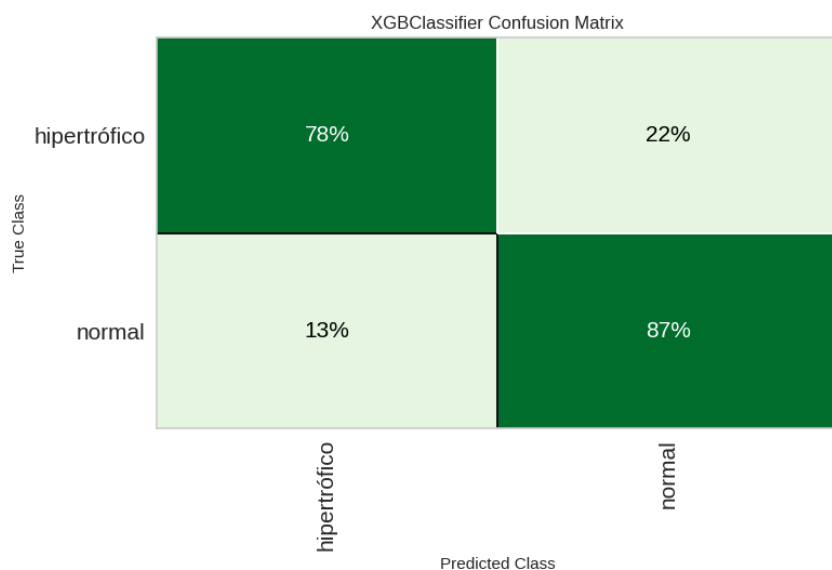


Figura 11 – Matriz de confusão do modelo XGBOOST após calibração

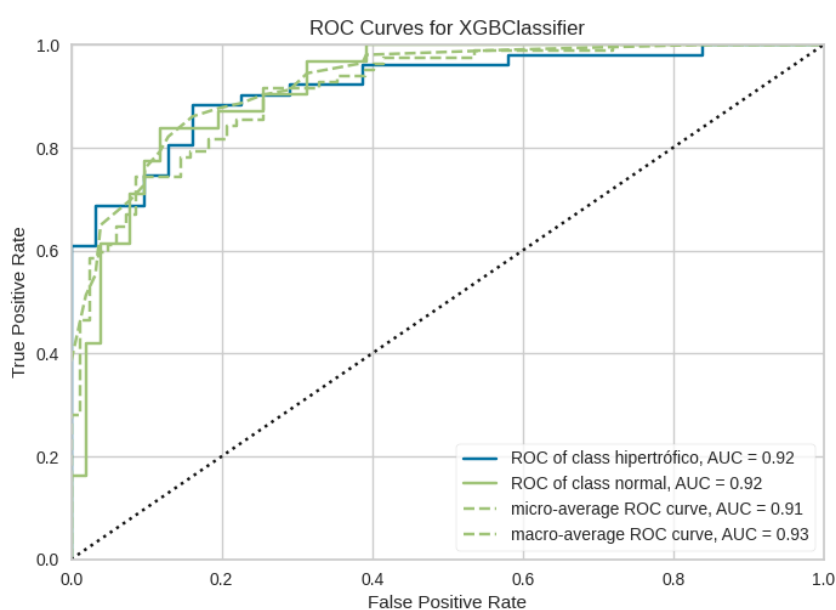


Figura 12 – Curva ROC do modelo XGBOOST após calibração

5.1.3 Melhor modelo da classificação binária

Dentre os modelos descritos na seção 5.1.1.1 e 5.1.2.1 o XGBOOST tem o melhor desempenho com um $F1_score$ no valor de 0,83 e desvio padrão 0,11, em relação ao modelo LR (seção 5.1.1.1) o modelo XGBOOST (seção 5.1.2.1) teve um aumento no valor de 2.47%

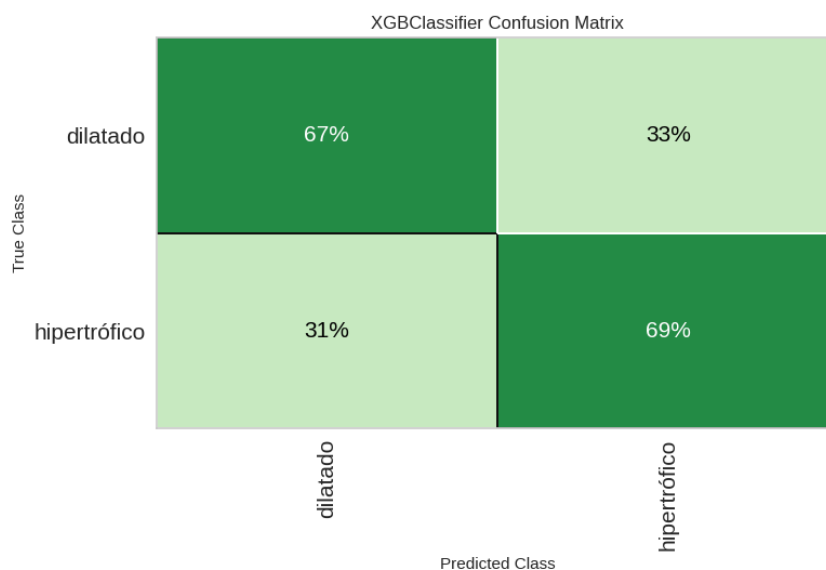


Figura 13 – Matriz de confusão do modelo XGBOOST após calibração

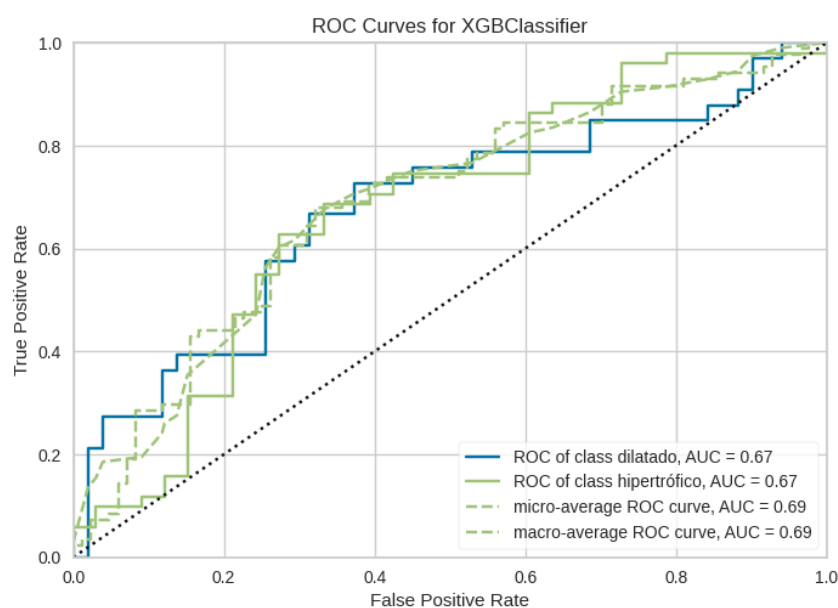


Figura 14 – Curva ROC do modelo XGBOOST após calibração

Ao comparar os dois modelos na classificação de dilatado e hipertrofico (Figuras 6 e 14), fica evidente que a LR supera o XGBOOST em termos de AUC para este conjunto de dados específico e para estas condições cardíacas. As curvas ROC do modelo de LR estão mais próximas do canto superior esquerdo, sugerindo uma proporção mais alta de acertos verdadeiros em comparação com os erros falsos positivos. O XGBOOST, embora ainda tenha algum poder discriminatório, não se saiu tão bem quanto a LR nesta tarefa específica.

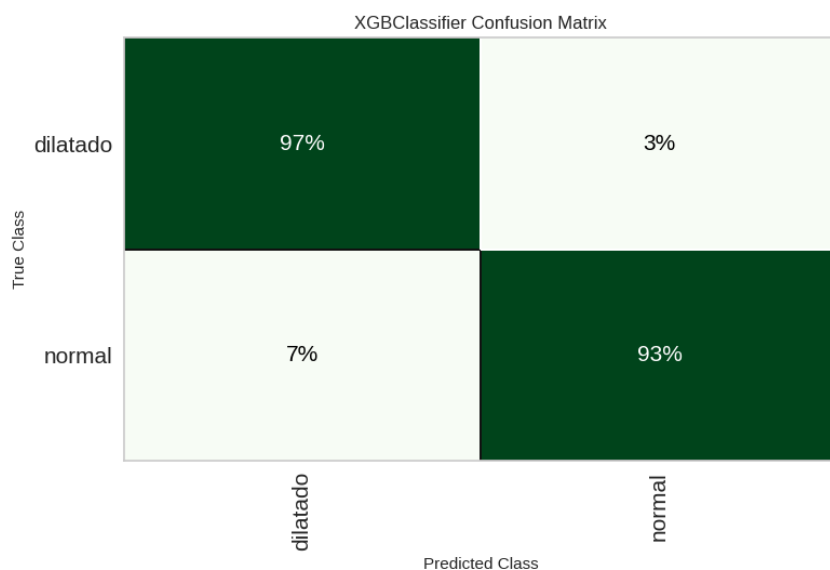


Figura 15 – Matriz de confusão do modelo XGBOOST após calibração

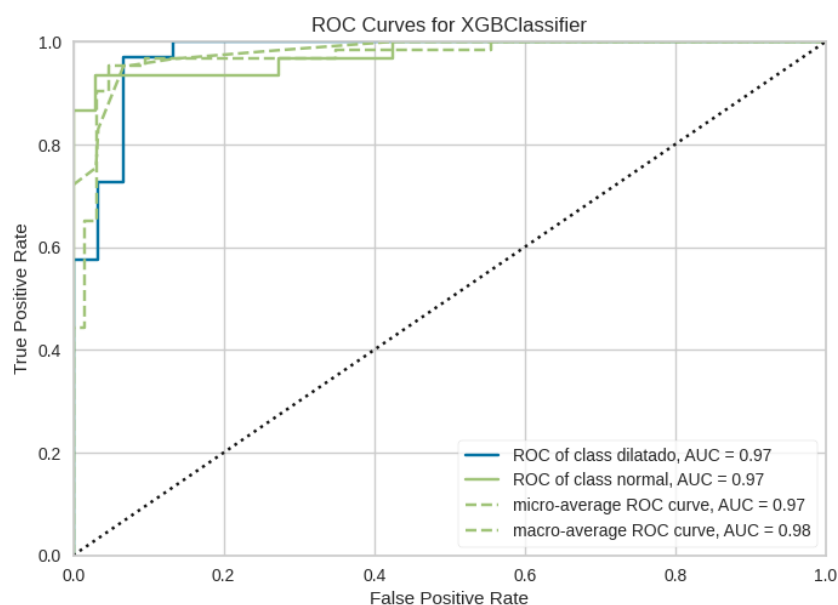


Figura 16 – Curva ROC do modelo XGBOOST após calibração

Contudo, tendo como base as Figuras 8 e 16, para a classificação de normal e dilatado, mostram que ambos possuem boas capacidades de classificação para esta tarefa específica. No entanto, o XGBOOST parece ter uma ligeira vantagem, ao possuir uma AUC ligeiramente superior em comparação com o LR. Isto sugere que o XGBOOST pode ter sensibilidade e especificidade ligeiramente maiores na diferenciação entre estados dilatados e normais.

Seguindo a mesma tendência ao analisar as Figuras 10 e 12, na classificação de

normal hipertrófico, é evidente que ambos oferecem um nível de classificação de alta qualidade para esta tarefa. Contudo, o XGBOOST exhibe uma vantagem, com AUCs ligeiramente mais elevadas em relação à LR. Isso sugere que o XGBOOST tem maior precisão, sensibilidade e especificidade na distinção entre corações hipertróficos e normais.

Uma comparação entre os modelos LR (seção 5.1.1.1) e XGBOOST (seção 5.1.2.1), o XGBOOST (seção 5.1.2.1) tem um desempenho melhor quando fazemos uma comparação geral, assim é o modelo com melhor desempenho da seção 5.1.

5.2 Classificação Multiclass

Após os resultados da classificação binárias (seção 5.1), a base de dados dos paciente com diagnóstico normal, hipertrófico, dilatado foram combinados e novamente submetido aos modelos, criado assim mais uma rodada de testes logo após os modelos foram calibrados (seção 5.2.1). Por fim na seção 5.2.2 usaram-se os descritores expostos no quadro 4 seguindo o mesmo processo na seção 5.2.1.

5.2.1 Classificação sem a base de dados segmentada

Os resultados obtidos nessa seção constam nas Tabelas 7 e 8. Após a calibração (Tabela 8) os hiperparâmetros de cada modelo constam no quadro 15.

Modelos	Original	PCA	SAC SKB	SAC SFM
ADA	0,48	0,57	0,54	0,53
NB	0,53	0,57	0,49	0,51
GBC	0,61	0,62	0,53	0,61
XGBOOST	0,64	0,61	0,56	0,55
KNN	0,62	0,53	0,56	0,51
LR	0,62	0,55	0,51	0,58
RBF SVM	0,27	0,27	0,27	0,27
SVM	0,58	0,46	0,35	0,65
RF	0,66	0,64	0,53	0,59
DT	0,59	0,53	0,54	0,56

Tabela 7 – Resultados da análise sem a calibração

Com os resultados da Tabela 7, o modelo RF apresenta o melhor desempenho tanto na base de dados original quanto após a aplicação de PCA, KNN na base de dados após a aplicação de SAC SKB e por fim SVM.

Modelos	Original	PCA	SAC SKB	SAC SFM
ADA	0,48	0,61	0,60	0,55
NB	0,53	0,57	0,49	0,51
GBC	0,63	0,64	0,56	0,61
XGBOOST	0,64	0,63	0,56	0,55
KNN	0,62	0,53	0,58	0,51
LR	0,65	0,55	0,51	0,58
RBFSVM	0,27	0,56	0,56	0,57
SVM	0,49	0,54	0,47	0,65
RF	0,66	0,64	0,56	0,59
DT	0,61	0,57	0,54	0,56

Tabela 8 – Resultados da análise com a calibração

Modelo	Base de dados	acurácia	AUC	Revocação	Precisão	F1
RF	original SC	0,66	0,79	0,66	0,67	0,66
RF	original CC	0,66	0,79	0,66	0,67	0,66
RF	PCA SC	0,65	0,76	0,65	0,66	0,64
RF	PCA CC	0,65	0,76	0,65	0,66	0,64
GBC	PCA CC	0,63	0,78	0,63	0,62	0,62
ADA	SAC SKB CC	0,61	0,71	0,61	0,61	0,60
KNN	SAC SKB SC	0,56	0,70	0,56	0,56	0,56
SVM	SAC SFM SC	0,65	0	0,65	0,65	0,65
SVM	SAC SFM CC	0,65	0	0,65	0,65	0,65

Tabela 9 – Métricas dos melhores modelos (CC = com calibração; SC = sem calibração)

A calibração fez com que na base de dados após a aplicação de PCA passou a ser GBC e RF e na base de dados SAC SKB passou a ser ADA, o modelo GBC houve um aumento no valor 3,23 %, já no modelo ADA houve um aumento de 11,11%, para SAC SFM o SVM permaneceu com maior $F1_score$.

A Tabela 9 apresenta as métricas dos melhores modelos de cada base de dados

Modelo	Base de dados	Parâmetros e Valores
RF	original	bootstrap=True, ccp_alpha=0.0, class_weight=None, criterion='gini', max_depth=None, max_features='sqrt', max_leaf_nodes=None, max_samples=None, min_impurity_decrease=0.0, min_samples_leaf=1, min_samples_split=2, min_weight_fraction_leaf=0.0, n_estimators=100, n_jobs=-1, oob_score=False, random_state=273153, verbose=0, warm_start=False
GBC	PCA	ccp_alpha=0.0, criterion='friedman_mse', init=None, learning_rate=0.5, loss='log_loss', max_depth=None, max_features='sqrt', max_leaf_nodes=None, min_impurity_decrease=0.0, min_samples_leaf=4, min_samples_split=5, min_weight_fraction_leaf=0.0, n_estimators=50, n_iter_no_change=None, random_state=830585, subsample=0.8, tol=0.0001, validation_fraction=0.1, verbose=0, warm_start=False
RF	PCA	ccp_alpha=0.0, criterion='friedman_mse', init=None, learning_rate=0.5, loss='log_loss', max_depth=None, max_features='sqrt', max_leaf_nodes=None, min_impurity_decrease=0.0, min_samples_leaf=4, min_samples_split=5, min_weight_fraction_leaf=0.0, n_estimators=50, n_iter_no_change=None, random_state=830585, subsample=0.8, tol=0.0001, validation_fraction=0.1, verbose=0, warm_start=False
ADA	SAC SKB	algorithm='SAMME', base_estimator='deprecated', estimator=None, learning_rate=1, n_estimators=50, random_state=765123
SVM	SAC SFM	alpha=0.1, average=True, class_weight=None, early_stopping=False, epsilon=1, eta0=0.01, fit_intercept=False, l1_ratio=0.3, learning_rate='adaptive', loss='hinge', max_iter=2000, n_iter_no_change=5, n_jobs=-1, penalty='l1', power_t=1, random_state=123, shuffle=True, tol=0.001, validation_fraction=0.1, verbose=0, warm_start=True

Quadro 15 – Hiperparâmetros após a calibração do modelo RF original

das Tabelas 7 e 8. O modelo RF mostra consistência em seu desempenho, com métricas semelhantes tanto na base “original” quanto na “PCA”, e pouca variação entre as configurações com e sem calibração. O modelo KNN na base “SAC SKB SC” possui métricas mais baixas em comparação com outros modelos.

5.2.1.1 Análise do modelo com melhor desempenho

A análise cruzada das Tabelas 7 e 8 destaca o RF como o modelo com melhor desempenho com um $F1_score$ de 0,66.

Ao analisar a matriz de confusão (Tabela 17), percebe-se que o modelo RF mostra um desempenho decente na identificação das três condições cardíacas. A classe hipertrófica e a classe normal apresentam taxas de acerto relativamente altas, com 71% e 74%, respectivamente. A classe dilatada, no entanto, é a mais desafiadora para o modelo, com uma taxa de acerto de 53%. O maior desafio aparente é distinguir entre dilatado e hipertrófico, como evidenciado pela taxa de classificação errada de 38%. Em termos gerais, enquanto o modelo mostra força em algumas áreas, há espaço para melhorias, especialmente na distinção entre corações dilatados e hipertróficos.

Para a classe dilatado, o AUC é de 0,78, indicando boa capacidade de discriminação entre corações dilatados e não dilatados (Figura 18). A classe hipertrófico possui um AUC de 0,74, mostrando uma precisão um pouco menor do que a classe dilatado. Já a classe normal apresenta o maior AUC de 0,88, sugerindo alta competência do modelo em identificar corações normais. Além disso, as métricas micro-average e macro-average com AUCs de 0,81 e 0,80, respectivamente, destacam a sólida desempenho geral do modelo.

5.2.2 Classificação com a base de dados segmentada

Com o resultado discutido na seção 5.2.1.1, é interessante analisar o uso dos descritores no quadro 4. Os hiperparâmetros obtidos constam no quadro 16.

A Tabela 12 apresenta métricas dos melhores modelos de cada dataset das Tabelas 10 e 11. O modelo de LR com a base de dados original e calibração demonstra o melhor

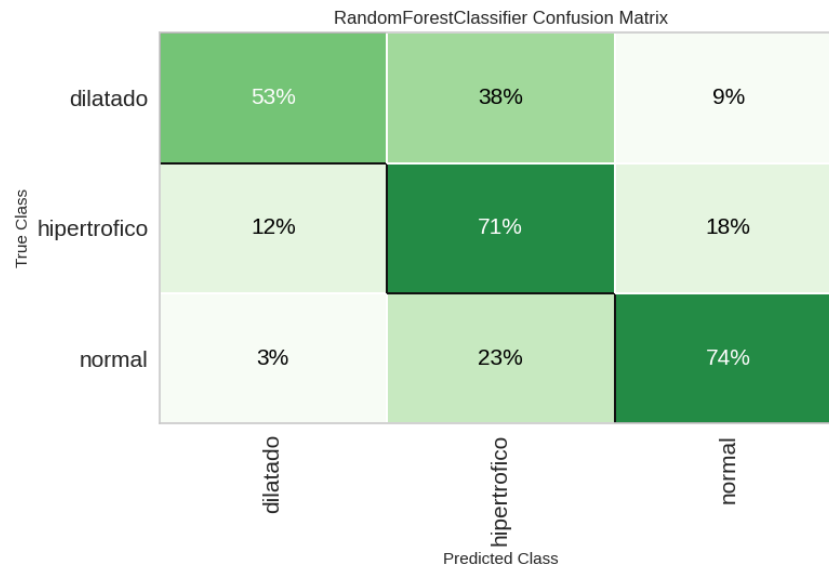


Figura 17 – Matriz de confusão do modelo RF após calibração

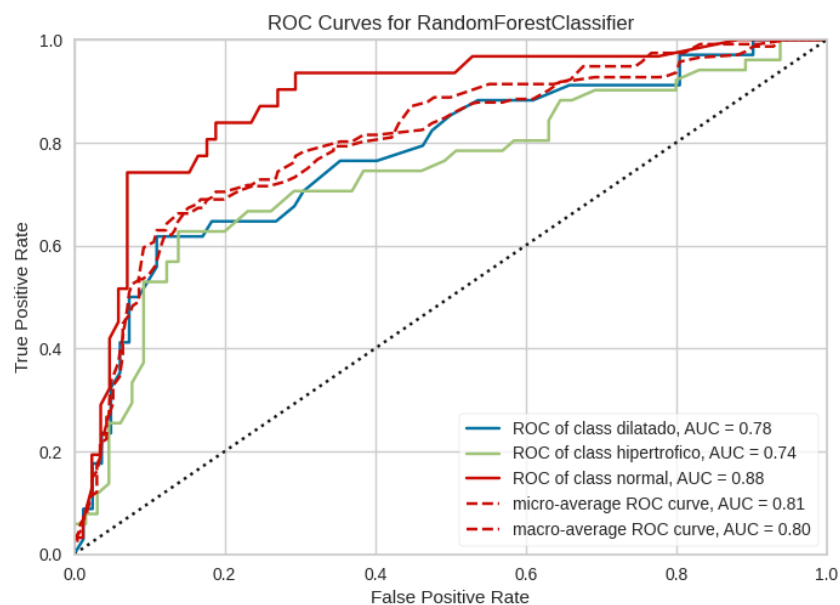


Figura 18 – Curva ROC do modelo RF após calibração

desempenho em termos de AUC 0,82 e também exibe alta acurácia, precisão, revocação e F1, todos com valor de 0,70. O XGBOOST, por sua vez, quando treinado na base de dados original sem calibração, apresenta um AUC de 0,81, sendo o segundo melhor em termos de AUC. Nota-se também que a maioria dos modelos apresenta métricas muito próximas, com variações pequenas, especialmente nas métricas de precisão, revocação e F1, sugerindo que os modelos possuem desempenhos similares para o conjunto de dados em questão.

Modelos	Original	PCA	SAC SKB	SAC SFM
ADA	0,54	0,56	0,59	0,57
NB	0,47	0,49	0,48	0,60
GBC	0,61	0,58	0,63	0,60
XGBOOST	0,65	0,57	0,62	0,68
KNN	0,52	0,53	0,45	0,54
LR	0,57	0,54	0,64	0,57
RBFSVM	0,28	0,28	0,28	0,28
SVM	0,45	0,50	0,26	0,46
RF	0,60	0,61	0,57	0,60
DT	0,52	0,49	0,60	0,53

Tabela 10 – Resultados da análise sem a calibração

Modelos	Original	PCA	SAC SKB	SAC SFM
ADA	0,60	0,56	0,64	0,60
NB	0,48	0,49	0,48	0,61
GBC	0,65	0,64	0,63	0,60
XGBOOST	0,65	0,57	0,62	0,68
KNN	0,52	0,54	0,53	0,54
LR	0,70	0,60	0,64	0,66
RBFSVM	0,53	0,55	0,55	0,53
SVM	0,59	0,51	0,59	0,55
RF	0,60	0,64	0,57	0,60
DT	0,58	0,50	0,58	0,53

Tabela 11 – Resultados da análise com a calibração

Os melhores modelos conforme a Tabela 10 são XGBOOST (base original), RF (base de dados após aplicar PCA), LR (base de dados após aplicar SAC SKB) e XGBOOST (base de dados após aplicar SAC SFM).

Com a calibração (Tabela 11) o modelo com melhor desempenho deixou de ser XGBOOST e passou a ser LR com um aumento de 7,69%. Já sobre PCA o modelo de melhor desempenho deixou de ser RF e passou a ser RF e GBC com um aumento de 4,92%. Com a base de dados após a aplicação de SAC SKB o modelo com o melhor desempenho deixou de ser LR e passou a ser LR e ADA.

5.2.2.1 Análise do modelo com melhor desempenho

Dentre os modelos descritos na seção 5.2.1.1 o com melhor desempenho foi LR (Tabela 11) com $F1_score$ de 0,70.

Modelo	Base de dados	acurácia	AUC	Revocação	Precisão	F1
XGBOOST	original SC	0,65	0,81	0,65	0,65	0,65
LR	original CC	0,70	0,82	0,70	0,70	0,70
RF	PCA SC	0,62	0,80	0,62	0,64	0,61
GBC	PCA CC	0,64	0,78	0,64	0,65	0,64
RF	PCA CC	0,62	0,80	0,62	0,64	0,61
LR	SAC SKB SC	0,65	0,79	0,6491	0,64	0,64
LR	SAC SKB CC	0,65	0,79	0,6491	0,64	0,64
ADA	SAC SKB CC	0,65	0,78	0,65	0,65	0,64
XGBOOST	SAC SFM SC	0,68	0,78	0,68	0,68	0,68
XGBOOST	SAC SFM CC	0,68	0,78	0,68	0,68	0,68

Tabela 12 – Métricas dos melhores modelos (CC = com calibração; SC = sem calibração)

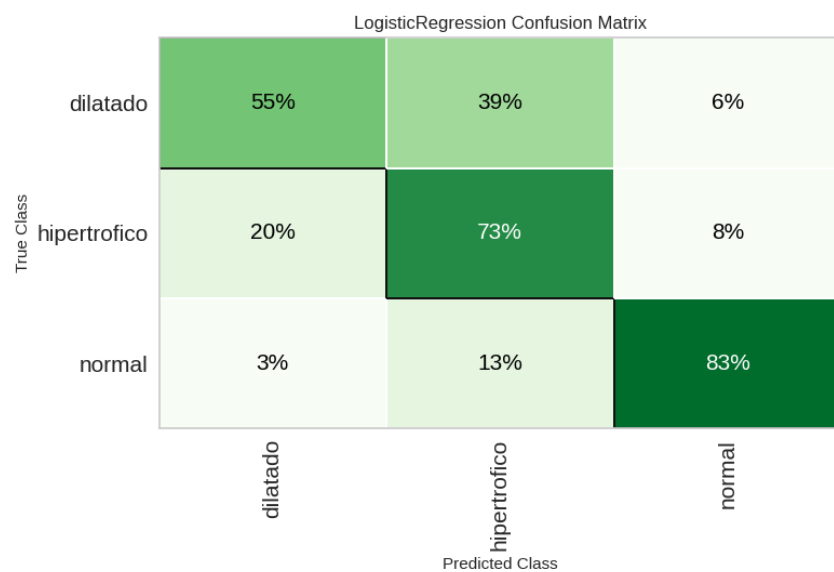


Figura 19 – Matriz de confusão do modelo LR após calibração

Modelo	Base de dados	Parâmetros e Valores
LR	original	<code>C=100, class_weight=None, dual=False, fit_intercept=True, intercept_scaling=1, l1_ratio=None, max_iter=1000, multi_class='auto', n_jobs=None, penalty='l1', random_state=888919, solver='liblinear', tol=0.0001, verbose=0, warm_start=False</code>
GBC	PCA	<code>ccp_alpha=0.0, criterion='friedman_mse', init=None, learning_rate=0.5, loss='log_loss', max_depth=7, max_features='sqrt', max_leaf_nodes=None, min_impurity_decrease=0.0, min_samples_leaf=4, min_samples_split=10, min_weight_fraction_leaf=0.0, n_estimators=10, n_iter_no_change=None, random_state=296840, subsample=0.9, tol=0.0001, validation_fraction=0.1, verbose=0, warm_start=False</code>
RF	PCA	<code>ccp_alpha=0.0, criterion='friedman_mse', init=None, learning_rate=0.5, loss='log_loss', max_depth=7, max_features='sqrt', max_leaf_nodes=None, min_impurity_decrease=0.0, min_samples_leaf=4, min_samples_split=10, min_weight_fraction_leaf=0.0, n_estimators=10, n_iter_no_change=None, random_state=296840, subsample=0.9, tol=0.0001, validation_fraction=0.1, verbose=0, warm_start=False</code>
ADA	SAC SKB	<code>algorithm='SAMME', base_estimator='deprecated', estimator=None, learning_rate=0.5, n_estimators=50, random_state=959782"</code>
LR	SAC SKB	<code>C=1.0, class_weight=None, dual=False, fit_intercept=True, intercept_scaling=1, l1_ratio=None, max_iter=1000, multi_class='auto', n_jobs=None, penalty='l2', random_state=959782, solver='lbfgs', tol=0.0001, verbose=0, warm_start=False</code>

Quadro 16 – Hiperparâmetros após a calibração dos modelos

A Figura 19 apresenta a desempenho do modelo LR na classificação de três condições

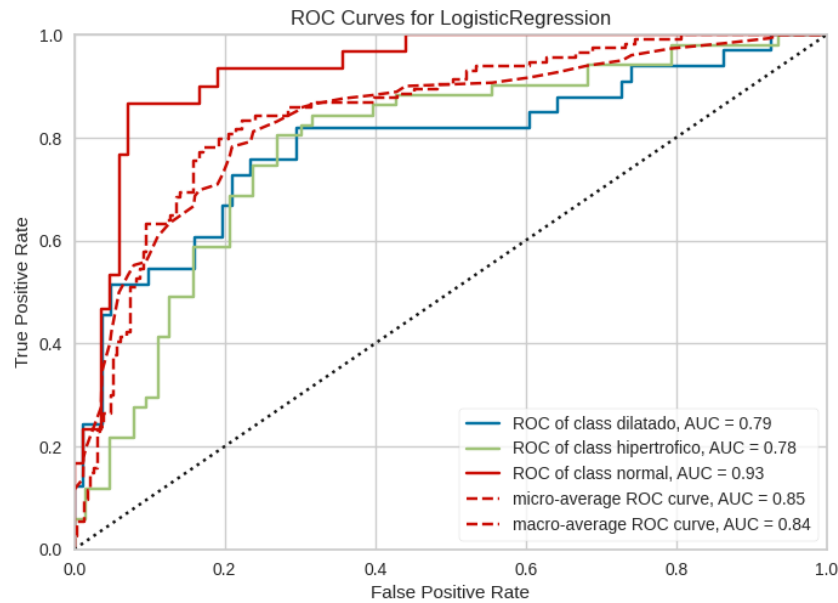


Figura 20 – Curva ROC do modelo LR após calibração

cardíacas. Para a classe dilatado, o modelo corretamente identificou 55% das amostras, mas classificou erroneamente 39% como hipertrófico e 6% como normal. Na classe hipertrófico, o modelo teve uma precisão de 73%, porém, confundiu 20% com dilatado e 8% com normal. A classe normal revelou a melhor precisão, com 83% das amostras corretamente classificadas, enquanto 3% foram identificadas como dilatado e 13% como hipertrófico.

Com a Figura 20 para a classe dilatado, o AUC foi de 0,79, indicando uma boa capacidade de discriminação. A classe hipertrófico apresentou um AUC ligeiramente menor, de 0,78. Já a classe normal destacou-se com um AUC de 0,93, evidenciando uma excelente capacidade do modelo em distinguir corações normais das outras condições. As métricas micro-average e macro-average, com AUCs de 0,85 e 0,84 respectivamente, sugerem que o modelo tem um desempenho geral consistente.

O modelo LR, ao analisar as Figuras 19 e 20, revela uma capacidade robusta de discriminar entre as três condições cardíacas, principalmente entre dilatado e hipertrófico, o modelo se destacou particularmente na identificação de corações normais. A combinação dessas métricas sugere que, enquanto há espaço para melhorias na classificação de dilatado e hipertrófico, o modelo é altamente preciso ao identificar a condição normal.

5.2.3 Melhor modelo de classificação multiclass

Ao cruzar os modelos descritos na seção 5.2.2.1 e 5.2.1.1 em geral, os dois modelos apresentaram desempenhos comparáveis, com diferenças sutis em cada categoria. O LR demonstrou uma ligeira vantagem na identificação da condição normal (Figura 19), enquanto o RF (Figura 17) foi um pouco mais preciso na classificação da condição hipertrófica.

Ao observar a classificação para a condição dilatado, os dois modelos mostram desempenhos comparáveis (Figuras 17 e 19). O RF possui um AUC de 0,78, enquanto o LR exibe um AUC ligeiramente superior, de 0,79. Esta marginal diferença sugere que ambos os modelos são quase igualmente eficazes na identificação de corações dilatados.

Para a condição hipertrófico, o RF apresenta um AUC de 0,74, enquanto o LR mostra um AUC de 0,78 (Figuras 17 e 19). Assim, o modelo LR tem vantagem ao identificar corações hipertróficos, com uma capacidade um pouco mais refinada de discriminação para esta condição específica.

No objetivo da identificação de estado normal, LR supera significativamente RF com uma excelente AUC de 0,93 em comparação com a AUC de RF de 0,88 (Figuras 17 e 19). Este desempenho indica que o modelo LR é particularmente bom em distinguir corações normais de outras doenças.

Além das categorias individuais, as métricas micro e macro-average também são úteis para avaliar o desempenho geral. O RF tem um AUC de 0,81 (micro) e 0,80 (macro), enquanto o LR mostra um desempenho superior com AUCs de 0,85 e 0,84, respectivamente.

Embora ambos os modelos demonstrem capacidades robustas de classificação para as condições cardíacas, o modelo LR, em geral, exibe um desempenho ligeiramente superior em comparação ao RF, particularmente na identificação de corações normais.

6 CONSIDERAÇÕES FINAIS

Este presente trabalho propôs entendimento de como os modelos de aprendizagem de máquina, desempenham para a identificação de cardiomiopatias a partir de imagens de ressonância magnéticas. A pesquisa bibliográfica mostrou não haver uma predominância de modelos de aprendizagem de máquina clássicos para esse diagnóstico (seção 2).

Foram treinados 10 modelos de AM e dentre eles o maior *F1_score* para a classificação multiclass foi 0,70 e para o binário foi 0,83. Observou-se que a classificação multiclass teve um desempenho inferior e isso se deu pela quantidade de dados usado para a realização do treinamento, para o desafio em questão deve-se obter mais dados para que o modelo multiclass consiga um melhor desempenho.

Com os resultados obtidos, o modelo XGBOOST apresentou um desempenho consistente e uma capacidade de distinguir as cardiomiopatias apresentadas neste trabalho de pesquisa.

6.1 Trabalhos futuros

Os resultados deste estudo ressaltam a complexidade de diferenciar cardiomiopatias dilatada e hipertrófica utilizando métodos clássicos de AM. Para abordar esta questão, os trabalhos futuros podem concentrar-se na exploração de modelos de aprendizado de máquina mais avançados ou no desenvolvimento de algoritmos híbridos que combinem várias técnicas para melhorar a classificação. Pode-se implementar técnicas de DL, como CNN ou RNN, já que tem uma capacidade maior de detectar padrões complexos nos dados.

Por fim, seria útil realizar um estudo mais abrangente envolvendo uma amostra maior de pacientes. Desta forma, os modelos clássicos podem apresentar um desempenho maior na diferenciação de cardiomiopatias hipertrófica e dilatada.

Além dos pontos a cima abordados, é importante tentar fazer a estratificação dos dados também considerando o sexo e a idade do paciente.

Ao enfrentar esses desafios, espera-se não apenas aprimorar as capacidades diagnósticas para condições cardíacas complexas, mas também contribuir para o campo da

inteligência artificial em medicina.

REFERÊNCIAS

- ACM Digital Library. **ACM Digital Library**. 2023. Available at: <https://dl.acm.org/>.
- AKERKAR, R.; SAJJA, P. S. **Intelligent Techniques for Data Science**. 1st. ed. [*S.l.: s.n.*]: Springer Publishing Company, Incorporated, 2016. ISBN 3319292056.
- ALBERTS, B. *et al.* **Fisiologia Humana: Uma abordagem integrada**. 6th. ed. [*S.l.: s.n.*]: Editora Artmed, 2017.
- BENKESSIRAT, A.; BENBLIDIA, N. Fundamentals of feature selection: An overview and comparison. *In: 2019 IEEE/ACS 16th International Conference on Computer Systems and Applications (AICCSA)*. [*S.l.: s.n.*], 2019. p. 1–6. Available at: <https://ieeexplore.ieee.org/document/9035281/>.
- BISHOP, N. Manipulation of machine learning algoirhtms. *In: Proceedings of the 21st International Conference on Autonomous Agents and Multiagent Systems*. Richland, SC: International Foundation for Autonomous Agents and Multiagent Systems, 2022. (AAMAS '22), p. 1833–1835. ISBN 9781450392136. Available at: <https://dl.acm.org/doi/10.5555/3535850.3536121>.
- CAI, H.; CAI, J.-F.; WEI, K. Accelerated alternating projections for robust principal component analysis. **J. Mach. Learn. Res.**, JMLR.org, v. 20, n. 1, p. 685–717, jan 2019. ISSN 1532-4435. Available at: <https://dl.acm.org/doi/10.5555/3322706.3322726>.
- CHANG, Y. *et al.* Automatic segmentation and cardiopathy classification in cardiac mri images based on deep neural networks. *In: 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. [*S.l.: s.n.*], 2018. p. 1020–1024. Available at: <https://ieeexplore.ieee.org/abstract/document/8461261>.
- CHATZILYGEROUDIS, K.; HATZILYGEROUDIS, I.; PERIKOS, I. Machine learning basics. *In: _____. Intelligent Computing for Interactive System Design: Statistics, Digital Signal Processing, and Machine Learning in Practice*. 1. ed. New York, NY, USA: Association for Computing Machinery, 2021. p. 143–193. ISBN 9781450390293. Available at: <https://doi.org/10.1145/3447404.3447414>.
- DALE, B.; BROWN, M.; SEMELKA, R. **MRI: Basic Principles and Applications**. Wiley, 2015. ISBN 9781119013037. Available at: <https://books.google.com.br/books?id=2D1UCgAAQBAJ>.
- DASH, S. *et al.* Big data in healthcare: management, analysis and future prospects. **Journal of Big Data**, v. 6, n. 1, p. 54, 2019. Available at: <https://doi.org/10.1186/s40537-019-0217-0>.
- DELMONDES, P. H. M. **Sistemas de auxílio ao diagnóstico de cardiomiopatias: uma abordagem baseada em descritores multi-slice e multi-frame**. 2022. Dissertação (Mestrado) — Escola de Artes, Ciências e Humanidades, University of São Paulo, São Paulo, 2022. Available at: <https://doi.org/10.11606/D.100.2022.tde-13022023-232135>.

DEVELOPERS scikit-learn. **Feature selection**. 2023. Acessado em: [data de acesso, e.g., 24 de setembro de 2023]. Available at: https://scikit-learn.org/stable/modules/feature_selection.html.

GIGER, M. L. Machine learning in medical imaging. **Journal of the American College of Radiology**, v. 15, n. 3, Part B, p. 512–520, 2018. ISSN 1546-1440. Data Science: Big Data Machine Learning and Artificial Intelligence. Available at: <https://www.sciencedirect.com/science/article/pii/S1546144017316733>.

GROVER, V. P. *et al.* Magnetic resonance imaging: Principles and techniques: Lessons for clinicians. **Journal of Clinical and Experimental Hepatology**, v. 5, n. 3, p. 246–255, 2015. ISSN 0973-6883. Available at: <https://www.sciencedirect.com/science/article/pii/S0973688315004156>.

HALL, J.; GUYTON, A. **Guyton and Hall Textbook of Medical Physiology**. 14th. ed. [S.l.: s.n.]: Elsevier Health Sciences, 2020.

IBRAHIM, S. M. *et al.* A study on heart segmentation using deep learning algorithm for mri scans. In: **2019 13th International Conference on Mathematics, Actuarial Science, Computer Science and Statistics (MACS)**. [S.l.: s.n.], 2019. p. 1–5. Available at: <https://ieeexplore.ieee.org/document/9024793>.

IEEE Xplore Digital Library. **IEEE Xplore Digital Library**. 2023. Available at: <https://ieeexplore.ieee.org>.

KIM, T.-H. *et al.* A novel fully automated cad system for left ventricle volume estimation. **Computers in biology and medicine**, Elsevier, v. 102, p. 203–215, 2018. Available at: <https://ieeexplore.ieee.org/document/8642754>.

KITZMAN, D. W.; EDWARDS, W. D. Age-Related Changes in the Anatomy of the Normal Human Heart. **Journal of Gerontology**, v. 45, n. 2, p. M33–M39, 03 1990. ISSN 0022-1422. Available at: <https://doi.org/10.1093/geronj/45.2.M33>.

KONST, R. *et al.* Prognostic value of cardiac magnetic resonance imaging in patients with a working diagnosis of minoca—an outcome study with up to 10 years of follow-up. **JACC: Cardiovascular Imaging**, v. 16, n. 2, p. 335–345, 2023. Available at: <https://doi.org/10.1161/CIRCIMAGING.122.014454>.

KUMAR, A.; BAGUR, R. Cardiac magnetic resonance in clinical cardiology. **World Journal of Cardiology**, United States, v. 7, n. 1, p. 6–9, 2015. ISSN 1949-8462. Available at: <https://doi.org/10.4330/wjc.v7.i1.6>.

LI, Y.; LI, Z. Heart disease prediction based on machine learning methods. Association for Computing Machinery, New York, NY, USA, p. 380–385, 2022. Available at: <https://doi.org/10.1145/3565291.3565352>.

LÓPEZ, F. M.-B.; BRAVO, R. M. A.; HUERTA, D. A. R. Echocardiographic features of non-compaction cardiomyopathy: missed and misdiagnosed disease. **SciELO - Scientific Electronic Library Online**, v. 1, n. 1, p. e22–e23, 2009.

MARGETA, J. *et al.* Fine-tuned convolutional neural nets for cardiac mri acquisition plane recognition. **Computer Methods in Biomechanics and Biomedical Engineering: Imaging e Visualization**, Taylor e Francis, v. 5, n. 5, p. 339–349, 2017. Available at: <https://doi.org/10.1080/21681163.2015.1061448>.

MARIEB, E. N.; WILHELM, P. B.; MALLATT, J. B. **Human Anatomy**. 6th. ed. Boston: Benjamin Cummings, 2012. ISBN 9780321753274.

MATTOS, B. P. e *et al.* O diagnóstico da obstrução da via de saída do ventrículo esquerdo na cardiomiopatia hipertrófica. **SciELO - Scientific Electronic Library Online**, v. 1, n. 1, p. 665–675, 2012.

MAURIZI, N. *et al.* Clinical and molecular aspects of cardiomyopathies: Emerging therapies and clinical trials. **Heart Fail Clin**, p. 161–178, 2018.

MERLO, M. *et al.* Treatment of cardiomyopathies: Where are we after the esc 2014/2015 guidelines? **Cardiovascular Drugs and Therapy**, Springer Science and Business Media LLC, v. 32, n. 8, p. 1125–1147, dez. 2020. Available at: <https://doi.org/10.1007/s10557-020-07016-7>.

MOHRI, M.; ROSTAMIZADEH, A.; TALWALKAR, A. **Foundations of Machine Learning**. [S.l.: s.n.]: The MIT Press, 2012. ISBN 026201825X.

NASR-ESFAHANI, M. *et al.* Left ventricle segmentation in cardiac mr images using fully convolutional network. In: **2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)**. [S.l.: s.n.], 2018. p. 1275–1278. Available at: <https://ieeexplore.ieee.org/abstract/document/8512536>.

ORGANIZATION, W. H. **World health statistics 2022: monitoring health for the SDGs, sustainable development goals**. 2023. Available at: <https://www.who.int/publications/i/item/9789240051157>.

PROVOST, F.; FAWCETT, T. **Data Science for Business: What You Need to Know about Data Mining and Data-Analytic Thinking**. 1th. ed. [S.l.: s.n.]: O'Reilly Media, 2013.

PyCaret Team. **PyCaret Documentation**. 2023. Acessado em: 30/08/2023. Available at: <https://pycaret.gitbook.io/docs/>.

RAMESH, M. *et al.* Machine learning based cardiac magnetic resonance imaging (cmri) for cardiac disease detection. In: **2021 Second International Conference on Smart Technologies in Computing, Electrical and Electronics (ICSTCEE)**. [S.l.: s.n.], 2021. p. 1–5. Available at: <https://ieeexplore.ieee.org/abstract/document/9708573>.

RIBEIRO, M. A. d. O. **Segmentação do ventrículo esquerdo em exames de ressonância magnética cardíaca com aprendizado profundo e modelos deformáveis contendo restrições de forma**. 2021. Dissertação (Mestrado) — Escola de Artes, Ciências e Humanidades, Universidade de São Paulo, São Paulo, 2021. Available at: <https://doi.org/10.11606/D.100.2021.tde-14122021-203305>.

SAEED, S.; ONG, H. C. A bi-objective hybrid algorithm for the classification of imbalanced noisy and borderline data sets. **Pattern Anal. Appl.**, Springer-Verlag, Berlin, Heidelberg, v. 22, n. 3, p. 979–998, aug 2019. ISSN 1433-7541. Available at: <https://doi.org/10.1007/s10044-018-0693-4>.

SALAKHUTDINOV, R. Deep learning. In: **Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining**. New York, NY, USA: Association for Computing Machinery, 2014. (KDD '14), p. 1973. ISBN 9781450329569. Available at: <https://doi.org/10.1145/2623330.2630809>.

TUFVESSON, J. *et al.* Validation and development of a new automatic algorithm for time resolved segmentation of the left ventricle in magnetic resonance imaging. **Journal of Cardiovascular Magnetic Resonance**, v. 17, n. 1, p. P68, 2015. ISSN 1532-429X. Available at: <https://doi.org/10.1186/1532-429X-17-S1-P68>.

WANG, L. *et al.* Use of magnetic resonance imaging combined with gene analysis for the diagnosis of fetal congenital heart disease. **BMC Medical Imaging**, v. 19, n. 1, p. 12, 2019. ISSN 1471-2342. Available at: <https://doi.org/10.1186/s12880-019-0314-8>.

WEXLER, R. K. *et al.* Cardiomyopathy: an overview. **American Family Physician**, American Academy of Family Physicians, v. 79, n. 9, p. 778–784, May 2009. ISSN 0002-838X. PMID: 20141097. Available at: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2999879/>.

XIAO, J. *et al.* Learning motion based auxiliary task for cardiomyopathy recognition with cardiac magnetic resonance images. *In: Proceedings of the 4th International Conference on Computer Science and Application Engineering*. New York, NY, USA: Association for Computing Machinery, 2020. (CSAE '20). ISBN 9781450377720. Available at: <https://doi.org/10.1145/3424978.3425122>.

YOON, Y. H. *et al.* Efficient b-mode ultrasound image reconstruction from sub-sampled rf data using deep learning. **IEEE Transactions on Medical Imaging**, v. 38, n. 2, p. 325–336, 2019. Available at: <https://ieeexplore.ieee.org/abstract/document/8432500>.

ZHENG, Y. *et al.* Direct multitype cardiac indices estimation via joint representation and regression learning. **IEEE Transactions on Medical Imaging**, IEEE, v. 39, n. 2, p. 311–321, 2019. Available at: <https://ieeexplore.ieee.org/abstract/document/7934404>.