CENTRO UNIVERSITÁRIO FEI
FABIO BINUESA

TÍTULO: Previsão de Mortalidade em Cirurgia Cardíaca Congênita utilizando Aprendizagem de Máquina.

São Bernardo do Campo
2020

FABIO BINUESA

TÍTULO: Previsão de Mortalidade em Cirurgia Cardíaca Congênita utilizando Aprendizagem de Máquina.

Dissertação apresentada ao Centro Universitário FEI, como parte dos requisitos necessários para obtenção do título de Mestre em Engenharia Mecânica. Orientado pelo Prof. João Chang Junior.

São Bernardo do Campo

2020

Binuesa, Fabio.
Previsão de Mortalidade em Cirurgia Cardíaca Congênita utilizando Aprendizagem de Máquina / Fabio Binuesa. São Bernardo do Campo, 2020.
113 p. : il.

Dissertação - Centro Universitário FEI.
Orientador: Prof. Dr. João Chang Junior.

1. doença cardíaca congênita. 2. machine learning. 3. modelo preditivo. 4. classificação. I. Chang Junior, João , orient. II. Título.

Elaborada pelo sistema de geração automática de ficha catalográfica da FEI com os dados fornecidos pelo(a) autor(a).

centro universitário FEI

APRESENTAÇÃO DE DISSERTAÇÃO
ATA DA BANCA EXAMINADORA

Programa de Pós-Graduação Stricto Sensu em Engenharia Mecânica

Mestrado

PGM-10

Aluno: Fabio Binuesa **Matrícula:** 218111-3

Título do Trabalho: Previsão de mortalidade após cirurgia cardíaca congênita utilizando aprendizagem de máquina.

Área de Concentração: Produção

Orientador: Prof. Dr. João Chang Junior

Data da realização da defesa: 19/02/2020

ORIGINAL ASSINADA

Avaliação da Banca Examinadora:

São Bernardo do Campo, / / .

MEMBROS DA BANCA EXAMINADORA	
Prof. Dr. João Chang Junior	Ass.: _____
Prof. Dr. Carlos Eduardo Thomaz	Ass.: _____
Prof. Dr. Luiz Fernando Caneo	Ass.: _____

A Banca Julgadora acima-assinada atribuiu ao aluno o seguinte resultado:

APROVADO ☒ REPROVADO ☐

VERSÃO FINAL DA DISSERTAÇÃO

APROVO A VERSÃO FINAL DA DISSERTAÇÃO EM QUE FORAM INCLUÍDAS AS RECOMENDAÇÕES DA BANCA EXAMINADORA

_____

Aprovação do Coordenador do Programa de Pós-graduação

_____

Prof. Dr. Rodrigo Magnabosco

Dedico esta pesquisa para minha esposa Claudia e minhas filhas, Erika e Leticia, que me apoiaram na conquista deste objetivo.

AGRADECIMENTOS

Agradeço imensamente a minha esposa Claudia pelo incentivo e apoio, além da dedicação e compreensão do objetivo.

Agradeço ao Doutor Alfredo Fernandes pela oportunidade de participar dessa pesquisa junto ao InCor.

Agradeço ao Doutor Luiz Fernando Caneo pela aceitação da pesquisa e orientações.

Agradeço a Doutora Aida Turquetto pelo apoio, acompanhamento e orientações.

Agradeço aos demais integrantes do Departamento de Cardiopatias Congênitas do InCor pelo apoio e informações fornecidas.

Agradeço ao Doutor João Chang Junior pela orientação e oportunidade de trabalhar nesse projeto.

Agradeço aos demais professores da Produção e da Elétrica do Centro Universitário FEI pelos ensinamentos.

“O que sabemos é uma gota; o que ignoramos é um oceano.”

Isaac Newton

RESUMO

A doença cardíaca congênita (*congenital heart diseases*) é a causa mais comum de grandes anomalias de mesmo gênero, responsável por quase um terço de todas as principais anomalias congênitas. Os defeitos cardíacos congênitos são condições graves e comuns com impacto significativo na morbidade, mortalidade e custos de saúde para crianças e adultos.

No tratamento de pacientes com cardiopatia congênita são raras as pesquisas relacionadas ao risco de mortalidade pré-cirúrgica. Este estudo tem como objetivo propor um modelo de previsão individual de risco de morte para cirurgia cardíaca de pacientes com cardiopatia congênita e auxiliar profissionais da saúde na compreensão de quais diagnósticos ou variáveis estão associadas ao risco de morte.

Utilização de técnicas de aprendizagem de máquina *(machine learning)* como ferramenta de apoio à tomada de decisão na área da medicina vem aumentando nos últimos anos. De posse das informações de cirurgias realizadas em pacientes com cardiopatia congênita extraídas do banco de dados ASSIST do InCor, foi possível treinar seis diferentes algoritmos de aprendizagem de máquina (*machine learning*) na previsão de risco de mortalidade pré cirúrgica e compreender quais variáveis impactam o risco de morte desses pacientes. Os algoritmos treinados neste estudo foram: *Multilayer Perceptron* (MLP), *Random Forest* (RF), *Extra Trees* (ET), *Stochastic Gradient Boosting* (SGB), *AdaBoost Classification* (ABC) e *Bagged Decision Trees* (BDT).

Para prever o risco de mortalidade do paciente, o modelo com melhor desempenho foi o *Random Forest* (RF) com ROC AUC (*area under the receiver operating characteristics)* de 90,2%, índices de AP (*average precision*) de 0,73 e índice de sensibilidade (*recall*) de 92,2%. Os algoritmos de aprendizagem de máquina (*machine learning*) podem auxiliar na compreensão dos riscos de mortalidade de pacientes com cardiopatia congênita quando submetida a uma intervenção cirúrgica cardíaca e permitem que especialistas clínicos compreendam melhor os riscos associados à intervenção cirúrgica, fornecendo informações para apoio à decisão dos profissionais de saúde, pacientes e seus familiares.

Palavras-chave: doença cardíaca congênita, *machine learning*, modelo preditivo, classificação.

ABSTRACT

Congenital heart disease is the most common cause of major anomalies of the same gender, accounting for almost a third of all major congenital anomalies. Congenital heart defects are serious and common conditions with a significant impact on morbidity, mortality and health costs for children and adults.

In the treatment of patients with congenital heart disease, research related to the risk of pre-surgical mortality is rare. This study aims to propose a model of individual risk of death prediction for cardiac surgery of patients with congenital heart disease and to assist health professionals in understanding which diagnoses or variables are associated with the risk of death.

The use of machine learning techniques as a tool to support decision making in the field of medicine has been increasing in recent years. With the information on surgeries performed on patients with congenital heart disease extracted from the ASSIST database of InCor, it was possible to train six different machine learning algorithms in predicting the risk of pre-surgical mortality and to understand which variables impact the risk death of these patients. The algorithms trained in this study were: Multilayer Perceptron (MLP), Random Forest (RF), Extra Trees (ET), Stochastic Gradient Boosting (SGB), AdaBoost Classification (ABC) and Bagged Decision Trees (BDT).

To predict the risk of patient mortality, the model with the best performance was the Random Forest (RF) with ROC AUC (area under the receiver's operating characteristics) of 90.2%, AP indexes (average precision) of 0, 73 and sensitivity index (recall) of 92.2%. The machine learning algorithm (machine learning) can assist in understanding the mortality risks of patients with congenital heart disease when undergoing cardiac surgery and using clinical drugs that understand the best risks associated with surgical interventions, providing information to support the decision. health professionals, patients and their families.

Keywords: congenital heart disease, machine learning, predictive model, classification.

LISTA DE ILUSTRAÇÕES

Figura 1 - Arquitetura Árvore de Decisão 24
Figura 2 - Método Bagging 28
Figura 3 - Comparação Bagging x Boosting 31
Figura 4 - Exemplo Arquitetura *Multilayer Perceptron* (MLP) 35
Figura 5 - Diagrama de uma rede neural artificial 36
Figura 6 - Diagrama de Métricas de Desempenho 39
Figura 7 - Etapas de construção do modelo de previsão 41
Figura 8 - Exemplo de *K-Fold Cross Validation* (K=5) 44
Figura 9 - Importância das Variáveis do *Random Forest* (RF) 50
Figura 10 - Curva de Validação do *Random Forest* (RF) 54
Figura 11 - Confusion Matrix *Random Forest* (RF) 55
Figura 12 - Curva ROC AUC do *Random Forest* (RF) 56
Figura 13 - Precision-Recall Curve do *Random Forest* (RF) 57
Figura 14 - Gráfico de Calibragem do *Random Forest* (RF) 58
Figura 15 - Curva de Validação do *Bagged Decision Trees* (BDT) 61
Figura 16 - Confusion Matrix do *Bagged Decision Trees* (BDT) 61
Figura 17 - Curva ROC AUC do *Bagged Decision Trees* (BDT) 62
Figura 18 - Precision-Recall Curve do *Bagged Decision Trees* (BDT) 63
Figura 19 - Curva Calibragem *Bagged Decision Trees* (BDT) 64
Figura 20 - Resultado balanceamento para *Stochastic Gradient Boosting* (SGB) 65
Figura 21 - Importância das Variáveis do *Stochastic Gradient Boosting* (SGB) 66
Figura 22 - Confusion Matrix do *Stochastic Gradient Boosting* (SGB) 69
Figura 23 - Curva ROC AUC do *Stochastic Gradient Boosting* (SGB) 70
Figura 24 - Precision-Recall Curve do *Stochastic Gradient Boosting* (SGB) 71
Figura 25 - Gráfico de Calibragem do *Stochastic Gradient Boosting* (SGB) 71
Figura 26 - Resultado balanceamento para *Extra Trees* (ET) 73
Figura 27 - Importância das Variáveis do *Extra Trees* (ET) 74
Figura 28 - Confusion Matrix do *Extra Trees* (ET) 77
Figura 29 - Curva ROC AUC do *Extra Trees* (ET) 78
Figura 30 - Precision-Recall Curve do *Extra Trees* (ET) 79
Figura 31 - Gráfico de Calibragem do *Extra Trees* (ET) 80
Figura 32 - Resultado balanceamento para *Adaboost Classification* (ABC) 81

Figura 33 - Importância das Variáveis do *Adaboost Classification* (ABC) 82
Figura 34 - Confusion Matrix do *Adaboost Classification* (ABC) 85
Figura 35 - Curva ROC AUC do *Adaboost Classification* (ABC) 86
Figura 36 - Precision-Recall Curve do *Adaboost Classification* (ABC).....87
Figura 37 - Gráfico de Calibragem do *Adaboost Classification* (ABC) 88
Figura 38 - Curva de Aprendizagem do *Multilayer Perceptron* (MLP).....92
Figura 39 - Curva de aprendizado com conjunto de treino balanceado 92
Figura 40 - Confusion Matrix do *Multilayer Perceptron* (MLP).....93
Figura 41 - Curva ROC AUC do *Multilayer Perceptron* (MLP).....94
Figura 42 - Precision-Recall Curves do *Multilayer Perceptron* (MLP) 95
Figura 43 - Gráficos de Dispersão das variáveis com maior % importância 99
Figura 44 - Histograma das variáveis com % importância 101

LISTA DE TABELAS

Tabela 1 – Tipos de Diagnósticos Cardiopatia Congênita 18
Tabela 2 – Estudo sobre nascimento de portadores de cardiopatia congênitas, 1990–2017 19
Tabela 3 - Estudo sobre mortalidade de portadores de cardiopatia congênitas, 1990–2017 21
Tabela 4 - Variáveis selecionadas para o Random Forest (RF) 51
Tabela 5 - Parâmetros utilizados no Random Forest (RF) 53
Tabela 6 - Métricas de Performance do Random Forest (RF) 56
Tabela 7 - Performance do Random Forest (RF) 59
Tabela 8 - Parâmetros utilizados no Bagged Decision Trees (BDT) 60
Tabela 9 - Métricas de Performance do Bagged Decision Trees (BDT) 62
Tabela 10 - Performance do Bagged Decision Trees (BDT) 64
Tabela 11 - Variáveis selecionadas para o Stochastic Gradient Boosting (SGB) 67
Tabela 12 - Parâmetros utilizados no Stochastic Gradient Boosting (SGB) 69
Tabela 13 - Métricas de Performance do Stochastic Gradient Boosting (SGB) 70
Tabela 14 - Performance do Stochastic Gradient Boosting (SGB) 72
Tabela 15 - Variáveis selecionadas para o Extra Trees (ET) 75
Tabela 16 - Parâmetros utilizados no Extra Trees (ET) 77
Tabela 17 - Métricas de Performance do Extra Trees (ET) 78
Tabela 18 - Performance do Extra Trees (ET) 80
Tabela 19 - Variáveis selecionadas para o Adaboost Classification (ABC) 83
Tabela 20 - Parâmetros utilizados no Adaboost Classification (ABC) 84
Tabela 21 - Métricas de Performance do Adaboost Classification (ABC) 86
Tabela 22 - Performance do Adaboost Classification (ABC) 88
Tabela 23 - Métricas de performance do Multilayer Perceptron (MLP) 94
Tabela 24 - Performance do *Multilayer Perceptron* (MLP) 95
Tabela 25 – Tabela Consolidada de Resultados de Performance 96
Tabela 26 - Tabela Consolidada da Análise do %importância 97

SUMÁRIO

1 INTRODUÇÃO 14
2 FUNDAMENTAÇÃO TEÓRICA 17
2.1 CARDIOPATIA CONGENITA 17
2.2 ALGORITMOS DE APRENDIZAGEM DE MÁQUINA 22
2.2.1 Árvore de decisão – Algoritmos *ensemble* 23
2.2.1.1 Bagged Decision Trees (BDT) 27
2.2.1.2 Random Forest (RF) 29
***2.2.1.3 Stochastic Gradient Boosting* (SGB) 30**
2.2.1.4 Extra Trees (ET) 32
2.2.1.5 AdaBoost Classification (ABC) 32
2.2.2 Multilayer Perceptron (MLP) 33
2.3 MÉTRICAS DE DESEMPENHO DO MODELO 37
3 MATERIAIS E MÉTODOS 41
3.1 CONJUNTO DE DADOS INCOR 42
3.2 NORMALIZAÇÃO OU PADRONIZAÇÃO DAS VARIÁVEIS 42
3.3 SEPARAÇÃO DO CONJUNTO DE DADOS EM TREINO E VALIDAÇÃO 43
3.4 BALANCEAMENTO DO CONJUNTO DE TREINAMENTO 45
3.5 TREINAMENTO E AJUSTES DO ALGORITMO 46
3.6 ANÁLISE PERFORMANCE MODELO 47
4 EXPERIMENTOS E RESULTADOS 48
4.1 RANDOM FOREST (RF) 49
4.2 BAGGED DECISION TREES (BDT) 59
4.3 STOCHASTIC GRADIENT BOOSTING (SGB) 65
4.4 EXTRA TREES (ET) 72
4.5 ADABOOST CLASSIFICATION (ABC) 81
4.6 MULTILAYER PERCEPTRON (MLP) 89
5 DISCUSSÃO E CONCLUSÃO 96
5.1. ANÁLISES 96
5.2. DISCUSSÃO 102
REFERÊNCIAS 105

1 INTRODUÇÃO

Os defeitos cardiovasculares congênitos, também conhecidos como cardiopatias congênitas (*congenital heart diseases*), são problemas estruturais que surgem na formação do coração ou dos principais vasos sanguíneos, com impacto significativo na morbidade, mortalidade e custos de saúde em crianças e adultos. Os defeitos variam em gravidade, desde minúsculos orifícios entre câmaras que são resolvidos naturalmente ou malformações que podem exigir múltiplos procedimentos cirúrgicos, sendo uma das principais causas de mortalidade perinatal e infantil (TENNANT et al., 2010). Alguns tipos de cardiopatias congênitas estão associados à diminuição da qualidade de vida, bem como déficits no funcionamento cognitivo e nos resultados do desenvolvimento neurológico (BENJAMIN et al., 2017).

A cardiopatia congênita, além de ser uma das mais comuns, é responsável por quase um terço de todas as principais anomalias congênitas (BENJAMIN et al., 2017)(LINDE et al., 2011). As estimativas relatadas de nascimento de portadores de cardiopatia congênita variam amplamente entre os estudos em todo o mundo, sendo que a estimativa de 9 por 1.000 nascidos vivos é geralmente aceita como a melhor aproximação e estima-se que nasçam anualmente mais de 1,35 milhões de crianças com alguma cardiopatia congênita (BENJAMIN et al., 2017)(LINDE et al., 2011).

O diagnóstico médico da cardiopatia congênita é complexo, pois há muitas variáveis a serem analisadas. Além disso, exige-se amplo conhecimento e muitos anos de experiência antes de realizar um diagnóstico adequado (RUIZ-FERNÁNDEZ et al., 2016). As malformações podem exigir vários procedimentos cirúrgicos, procedimentos estes que podem resultar em morte (BENJAMIN et al., 2017). Dependendo do subtipo de anomalia congênita, observa-se taxas de sobrevida de 20 anos, variando de 66,2% a 97,6%, mas há ainda uma escassez de conhecimento sobre sobrevivência, especialmente nos primeiros ano de vida (BENJAMIN et al., 2017)(TENNANT et al., 2010).

Para reduzir a mortalidade de pacientes com cardiopatia congênita, ferramentas de análise de risco de mortalidade têm fornecido apoio à tomada de decisão médica, evitando intervenções clínicas ou tratamentos ineficazes (LIN; HU; KONG, 2019). Existem alguns modelos de estratificação de risco para mortalidade e morbidade para crianças com cardiopatia congênita, exemplo o RACHS-1 (JENKINS et al., 2002)(JENKINS, 2004)(JENKINS; GAUVREAU, 2002), *Aristotles Basic*

Complexity (ABC) e *Aristotles Comprehensive Complexity* (ACC) (LACOUR-GAYET et al., 2004). Esses modelos foram desenvolvidos com base na experiência e consenso entre os especialistas neste campo, devido à falta de dados adequados naquele momento (HÖRER et al., 2016).

No entanto, as previsões desses modelos não são precisas o suficiente para análise individual, seja porque o desempenho de uma determinada pontuação de risco reflete a média de um grupo ou porque existam particularidades sociodemográficas que impactam a resposta do modelo (VILELA DE ABREU HAICKEL NINA et al., 2007)(MEJIA et al., 2018). Diante desse cenário, modelos de previsão de mortalidade individualizado tem sido propostos e vêm apresentando melhor desempenho em comparação ao sistema de pontuação de gravidade padrão (RUIZ-FERNÁNDEZ et al., 2016)(AWAD; BADER-EL-DEN; MCNICHOLAS, 2017)(LI et al., 2017).

Em pesquisas recentes sobre modelos de previsão de mortalidade individualizado, observamos que o uso de técnicas de aprendizagem de máquina *(machine learning)* como ferramenta de apoio à tomada de decisão na área da medicina vem aumentando nos últimos anos (RUIZ-FERNÁNDEZ et al., 2016). Essas técnicas têm demonstrado melhor performance em relação às técnicas tradicionais, tais como a regressão logística, (AWAD; BADER-EL-DEN; MCNICHOLAS, 2017)(XU et al., 2016)(AWAD et al., 2017), inclusive em estudos de previsão de mortalidade (AWAD; BADER-EL-DEN; MCNICHOLAS, 2017).

A proposta de modelos de risco de mortalidade de pacientes com cardiopatia congênita é promissora, uma vez que são raras as pesquisas sobre este assunto. Em pesquisas recentes encontra-se apenas o estudo de Ruiz Fernandez et al. (2016), em que é utilizado algoritmos de aprendizagem de máquina (*machine learning*) para classificar grupos de riscos de morte em cirurgia ou o estudo de Li et al. (2017), que propôs um modelo de rede neural artificial (RNA) para prever os riscos de cardiopatia congênita em gestantes. Ao buscar especificamente modelos de previsão individual de morte em cirurgia cardíaca de pacientes com cardiopatia congênita esse resultado é nulo. Esse fato abre espaço para propor um modelo de previsão de mortalidade para cirurgia cardíaca de pacientes com cardiopatia congênita, utilizando-se de técnicas de aprendizagem de máquina (*machine learning)* com dados de entrada pré-cirúrgico, permitindo que especialistas clínicos compreendam melhor os riscos associados a uma intervenção cirúrgica.

O modelo de previsão de morte em cirurgia cardíaca para pacientes com cardiopatia congênita é uma ferramenta essencial na tomada de decisão clínica, oferecendo suporte para os profissionais nos preparativos pré-cirúrgico e ampliando a compreensão das variáveis que influenciam o risco de morte. Esta afirmativa abre espaço para análises e constante avaliação de métodos que possam melhorar a acurácia das previsões e melhor compreensão das variáveis associadas ao risco de morte, e assim, fornecendo informações para apoio a decisão dos profissionais de saúde e familiares dos pacientes.

Diante desses fatos, este estudo tem os seguintes objetivos:

a) Propor um modelo de previsão individual de morte para cirurgia cardíaca de pacientes com cardiopatia congênita, permitindo que especialistas clínicos, pacientes e familiares compreendam melhor os riscos associados a uma possível intervenção cirúrgica cardíaca;

b) Entender quais diagnósticos ou variáveis impactam a probabilidade de mortalidade de um paciente com cardiopatia congênita quando submetida a uma intervenção cirúrgica cardíaca.

No próximo capitulo será apresentado informações sobre a cardiopatia congênita, um levantamento de trabalhos relacionados ao uso de algoritmos de aprendizagem de máquina (m*achine learning*) no diagnóstico médico, bem como diversos algoritmos de aprendizagem de máquina e as métricas de desempenho utilizadas para determinar o melhor modelo para o estudo proposto. Na sequência, capítulo 3, será abordado a metodologia utilizada na pesquisa, detalhando os dados utilizados, etapas de preparação do conjunto de dados e treinamento dos algoritmos de aprendizagem de máquina. No capítulo 4 serão apresentados os experimentos desenvolvidos e seus resultados. E por fim, o capítulo 5 apresentará a discussão final e sua conclusão.

2 FUNDAMENTAÇÃO TEÓRICA

Este capítulo apresenta um breve resumo sobre a cardiopatia congênita e traz um levantamento de pesquisas que utilizaram algoritmos de aprendizagem de máquina *(machine learning)* como ferramentas de apoio à decisão médica, como também um resumo sobre as métricas utilizadas na medição de desempenho dos algoritmos de aprendizagem de máquina com ênfase na área de medicina.

2.1 CARDIOPATIA CONGENITA

Os defeitos cardiovasculares congênitos, também conhecido como cardiopatia congênita surgem da formação anormal ou incompleta do coração e dos vasos sanguíneos. Estes defeitos variam em gravidade, desde pequenas anormalidades que não requerem tratamento a malformações complexas, incluindo porções ausentes, hipoplásicas ou atrésias do coração, válvulas ou vasos que podem exigir múltiplas cirurgias e intervenções, incluindo transplante cardíaco. Assim, há uma variabilidade significativa de defeitos com impacto significativo nos custos de morbidade, mortalidade e na assistência à saúde, tanto em crianças quanto em adultos. Alguns tipos de cardiopatia congênita impactam diretamente a qualidade de vida como também geram déficits no funcionamento cognitivo e neurológico (VIRANI et al., 2020).

São muitos os defeitos cardiovasculares congênitos, na tabela 1 são apresentados os tipos de cardiopatia congênita mapeados na base ASSIST do InCor.

Tabela 1 – Tipos de Diagnósticos Cardiopatia Congênita

Diagnósticos Cardiopatia Congênita
1. Cardiomyopathy
1. Cardiomyopathy dilated, restrictive, and hypertrophic
2. Cardiomyopathy, Endstage congenital heart disease
2. Conduit failure
2. Conduit failure
3. Cor Triatriatum
3. Cor Triatriatum
4. DOLV (Double Outlet Left Ventricle)
4. DOLV (Double Outlet Left Ventricle)
5. DORV (Double Outlet Right Ventricle)
1. DORV + AVSD (AV Canal)
2. DORV, IVS
3. DORV, Remote VSD (uncommitted VSD)
4. DORV, TGA type
5. DORV, TOF type
6. DORV, VSD type
6. Electrophysiological
1. Arrhythmia
2. Arrhythmia, Atrial
3. Arrhythmia, Heart block
4. Arrhythmia, Heart block, Acquired
5. Arrhythmia, Heart block, Congenital
6. Arrhythmia, Junctional
7. Arrhythmia, Pacemaker, Indication for replacement
8. Arrhythmia, Ventricular
7. Left Heart Lesions
1. Aortic insufficiency
2. Aortic insufficiency and aortic stenosis
3. Aortic stenosis, Subvalvar
4. Aortic stenosis, Supravalvar
5. Aortic stenosis, Valvar
6. Aortic valve atresia
7. Aortic valve, Other
8. Hypoplastic left heart syndrome (HLHS)
9. LV to aorta tunnel
10. Mitral regurgitation
11. Mitral regurgitation and mitral stenosis
12. Mitral stenosis
13. Mitral stenosis, Subvalvar
14. Mitral stenosis, Subvalvar, Parachute
15. Mitral stenosis, Supravalvar mitral ring
16. Mitral stenosis, Valvar
17. Mitral valve, Other
18. Shone's syndrome
19. Sinus of Valsalva aneurysm
8. Miscellaneous, Other
1. Aneurysm, Other
2. Aneurysm, Pulmonary Artery
3. Aneurysm, Ventricular, Left (including pseudoaneurysm)
4. Aneurysm, Ventricular, Right (including pseudoaneurysm)
5. Atrial Isomerism, Left
6. Atrial Isomerism, Right
7. Cardiac tumor
8. Cardiac, Other
9. Dextrocardia
10. Endocarditis
11. Hypoplastic LV
12. Hypoplastic RV
13. Levocardia
14. Meconium aspiration
15. Mediastinitis
16. Mesocardia
17. Miscellaneous, Other
18. Myocardial infarction
19. Normal heart
20. Peripheral vascular, Other
21. Persistent fetal circulation
22. Postoperative bleeding
23. Primary pulmonary hypertension
24. Prosthetic valve failure
25. Pulmonary AV fistula
26. Pulmonary embolism
27. Pulmonary vascular obstructive disease
28. Pulmonary vascular obstructive disease (Eisenmenger's)
29. Rheumatic heart disease
30. Situs inversus
31. Status post - Transplant, Heart
32. Status post - Transplant, Heart and lung
33. Status post - Transplant, lung(s)
34. Thoracic and/or mediastinal, Other
9. Pericardial Disease
1. Pericardial disease, Other
2. Pericardial effusion
3. Pericarditis
10. Pulmonary Venous Anomalies
1. Partial anomalous pulmonary venous connection (PAPVC)
2. Partial anomalous pulmonary venous connection (PAPVC), scimitar
3. Total anomalous pulmonary venous connection (TAPVC), Type 1 (supracardiac)
4. Total anomalous pulmonary venous connection (TAPVC), Type 2 (cardiac)
5. Total anomalous pulmonary venous connection (TAPVC), Type 3 (infracardiac)
6. Total anomalous pulmonary venous connection (TAPVC), Type 4 (mixed)
11. Pulmonary Venous Stenosis
1. Pulmonary venous stenosis
12. Right Heart Lesions
1. MAPCA(s) (major aortopulmonary collateral[s]) (without PAVSD)
2. Pulmonary atresia
3. Pulmonary atresia, IVS
4. Pulmonary atresia, VSD (Including TOF, PA)
5. Pulmonary atresia, VSDMAPCA (pseudotruncus)
6. Pulmonary insufficiency
7. Pulmonary insufficiency and pulmonary stenosis
8. Pulmonary valve, Other Other
9. DCRV
10. Pulmonary artery stenosis (hypoplasia), Main (trunk)
11. Pulmonary artery stenosis, Branch, Central (within the hilar bifurcation)
12. Pulmonary artery stenosis, Branch, Peripheral (at or beyond the hilar bifurcation)
13. Pulmonary artery, Discontinuous
14. Pulmonary stenosis, Subvalvar
15. Pulmonary stenosis, Valvar
16. TOF
17. TOF, Absent pulmonary valve
18. TOF, AVC (AVSD)
19. TOF, Pulmonary stenosis
20. Ebstein's anomaly
21. Tricuspid regurgitation and tricuspid stenosis
22. Tricuspid regurgitation, non-Ebstein's related
23. Tricuspid stenosis
24. Tricuspid valve, Other
13. Septal Defects
1. AP window (aortopulmonary window)
2. Pulmonary artery origin from ascending aorta (hemitruncus)
3. ASD, Common atrium (single atrium)
4. ASD, Coronary sinus
5. ASD, Secundum
6. ASD, Sinus venosus
7. PFO
8. AVC (AVSD), Complete (CAVSD)
9. AVC (AVSD), Intermediate (transitional)
10. AVC (AVSD), Partial (incomplete) (PAVSD) (ASD, primum)
11. Truncal valve insufficiency
12. Truncus arteriosus
13. Truncus arteriosus + Interrupted aortic arch
14. VSD, Multiple
15. VSD, Type 1 (Subarterial) (Supracristal) (Conal septal defect) (Infundibular)
16. VSD, Type 2(Perimembranous) (Paramembranous) (Conoventricular)
17. VSD, Type 3 (Inlet) (AV canal type)
18. VSD, Type 4 (Muscular) A VSD completely surrounded by muscle.
19. VSD, Type: Gerbode type (LV-RA communication
14. Shunt failure
1. Shunt failure
15. Single Ventricle
1. Single Ventricle + Total anomalous pulmonary venous connection (TAPVC)
2. Single ventricle, DILV
3. Single ventricle, DIRV
4. Single ventricle, Heterotaxia syndrome
5. Single ventricle, Mitral atresia
6. Single ventricle, Other
7. Single ventricle, Tricuspid atresia
8. Single ventricle, Unbalanced AV canal
16. Systemic Venous Anomalies
1. Systemic venous anomaly
2. Systemic venous obstruction
17. Thoracic and Mediastinal Disease
1. Diaphragm disease, Other
2. Diaphragm paralysis
3. Esophageal disease, Benign
4. Esophageal disease, Malignant
5. Lung disease, Benign
6. Lung disease, Malignant
7. Mediastinal disease
8. Mediastinal disease, Benign
9. Mediastinal disease, Malignant
10. Pectus
11. Chylothorax
12. Empyema
13. Pleural disease, Benign
14. Pleural disease, Malignant
15. Pleural effusion
16. Pneumothorax
17. Airway disease
18. Tracheal stenosis
18. Thoracic Arteries and Veins
1. Aortic aneurysm (including pseudoaneurysm)
2. Aortic dissection
3. Aortic arch hypoplasia
4. Coarctation of aorta
5. VSD + Aortic arch hypoplasia
6. VSD + Coarctation of aorta
7. Coronary artery anomaly, Aneurysm
8. Coronary artery anomaly, Anomalous aortic origin of coronary artery from aorta (AAOCA)
9. Coronary artery anomaly, Anomalous pulmonary origin (includes ALCAPA)
10. Coronary artery anomaly, Fistula
11. Coronary artery anomaly, Other
12. Interrupted aortic arch
13. Interrupted aortic arch + AP window (aortopulmonary window)
14. Interrupted aortic arch + VSD
15. Patent ductus arteriosus
16. Pulmonary artery sling
17. Vascular ring
19. Transposition of the Great Arteries
1. Congenitally corrected TGA
2. Congenitally corrected TGA, IVS
3. Congenitally corrected TGA, IVS-LVOTO
4. Congenitally corrected
5. TGA, VSD
6. Congenitally corrected TGA, VSD
7. Congenitally corrected TGA, VSD-LVOTO
8. TGA, IVS
9. TGA, IVS-LVOTO
10. TGA, VSD
11. TGA, VSD-LVOTO

Fonte: Base ASSIST do InCor

As estimativas relatadas de nascimento de crianças com cardiopatias congênitas variam amplamente entre os estudos em todo o mundo, sendo a estimativa de 9 por 1.000 nascidos vivos é geralmente aceita como a melhor aproximação (LINDE et al., 2011). A população com defeitos cardíacos congênitos tem crescido substancialmente nas últimas décadas, estima-se mais de 1,35 milhões de recém-nascidos com algum tipo de cardiopatia congênita a cada ano (VIRANI et al., 2020) (TENNANT et al., 2010).

Tabela 2 – Estudo sobre nascimento de portadores de cardiopatia congênitas, 1990–2017

	Prevalence at birth per 100 000 babies			Prevalence per 100 000 children younger than 1 year			Prevalence per 100 000 individuals (all ages)			Age-standardised prevalence rate per 100 000		
	1990	2017	Percent change, 1990–2017	1990	2017	Percent change, 1990–2017	1990	2017	Percent change, 1990–2017	1990	2017	Percent change, 1990–2017
Global	1716·1 (1519·6 to 1925·8)	1787·6 (1587·0 to 1999·2)	4·2% (2·8 to 5·6)*	1192·4 (1051·2 to 1353·9)	1233·1 (1092·3 to 1392·7)	3·4% (1·9 to 4·9)*	187·3 (169·8 to 206·7)	157·0 (143·4 to 171·8)	−16·2% (−17·4 to −14·9)*	167·5 (152·1 to 184·8)	170·6 (155·6 to 186·9)	1·9% (0·8 to 2·9)*
Low SDI	2517·0 (2212·7 to 2823·4)	2501·1 (2200·5 to 2807·6)	−0·6% (−2·6 to 1·2)	1613·0 (1425·4 to 1809·7)	1611·3 (1424·7 to 1812·2)	−0·1% (−2·0 to 1·7)	276·8 (250·4 to 305·1)	237·5 (215·5 to 260·9)	−14·2% (−15·5 to −12·9)*	184·3 (167·7 to 202·4)	185·1 (168·5 to 202·4)	0·4% (−1·0 to 1·9)
Low-middle SDI	1897·8 (1669·9 to 2128·2)	1867·1 (1648·8 to 2089·7)	−1·6% (−3·7 to 0·7)	1278·5 (1131·7 to 1442·7)	1273·8 (1126·3 to 1438·8)	−0·4% (−2·5 to 1·9)	217·3 (196·8 to 239·9)	182·7 (166·6 to 200·1)	−15·9% (−17·2 to −14·5)*	164·5 (149·2 to 180·5)	165·4 (151·0 to 180·9)	0·6% (−0·8 to 2·0)
Middle SDI	1384·5 (1217·9 to 1567·3)	1350·7 (1191·8 to 1531·5)	−2·4% (−4·3 to −0·8)*	1027·9 (894·1 to 1179·4)	1001·6 (879·9 to 1137·5)	−2·5% (−4·8 to −0·8)*	172·4 (155·2 to 191·5)	128·7 (116·8 to 141·5)	−25·4% (−26·9 to −23·9)*	150·3 (135·7 to 166·6)	152·4 (137·8 to 168·1)	1·4% (−0·1 to 2·7)
High-middle SDI	1207·1 (1052·7 to 1383·7)	1184·2 (1038·4 to 1342·2)	−1·9% (−4·2 to 0·5)	949·3 (816·3 to 1102·6)	937·4 (817·3 to 1076·3)	−1·3% (−3·8 to 1·3)	153·8 (136·1 to 169·9)	112·4 (102·1 to 123·6)	−25·9% (−27·9 to −24·1)*	151·0 (135·3 to 169·0)	154·9 (139·3 to 171·6)	2·6% (0·7 to 4·2)*
High SDI	1235·3 (1106·3 to 1381·0)	1255·5 (1129·6 to 1393·2)	1·6% (−0·2 to 3·6)	995·4 (878·7 to 1122·7)	1020·3 (910·6 to 1135·9)	2·5% (0·5 to 4·6)*	156·0 (144·0 to 170·0)	135·0 (125·1 to 145·4)	−13·5% (−14·8 to −12·0)*	193·3 (177·1 to 211·7)	196·4 (180·9 to 212·6)	1·6% (0·2 to 2·9)*
Central Europe, eastern Europe, and central Asia												
Central Asia	1582·5 (1387·5 to 1805·6)	1542·4 (1361·2 to 1751·5)	−2·5% (−7·0 to 2·2)	1215·2 (1057·5 to 1392·5)	1180·1 (1026·7 to 1362·9)	−2·9% (−7·6 to 2·1)	226·8 (202·5 to 252·7)	188·7 (168·9 to 210·4)	−16·8% (−19·8 to −14·0)*	187·1 (167·5 to 207·1)	187·2 (167·6 to 208·7)	0·1% (−3·4 to 3·5)
Central Europe	1221·5 (1053·9 to 1404·5)	1216·5 (1096·0 to 1350·8)	−0·4% (−6·0 to 5·4)	1007·2 (861·0 to 1167·5)	1014·6 (904·7 to 1130·5)	0·7% (−5·2 to 7·0)	148·7 (133·0 to 165·1)	118·1 (108·8 to 127·7)	−20·6% (−24·0 to −16·6)*	177·3 (157·5 to 198·3)	184·2 (168·4 to 200·4)	3·9% (−0·4 to 8·6)
Eastern Europe	1341·2 (1167·3 to 1542·7)	1314·2 (1146·0 to 1513·7)	−2·0% (−4·6 to 0·3)	1067·5 (918·4 to 1242·3)	1053·7 (907·5 to 1220·8)	−1·3% (−4·0 to 1·2)	146·0 (131·2 to 162·8)	122·4 (110·8 to 136·1)	−16·2% (−17·7 to −14·7)*	176·6 (157·8 to 197·1)	178·5 (158·8 to 200·0)	1·1% (−0·8 to 2·9)
High income												
Australasia	974·9 (854·8 to 1127·8)	1023·6 (891·6 to 1165·2)	5·0% (−3·1 to 14·1)	784·7 (680·8 to 910·5)	824·8 (715·0 to 949·8)	5·1% (−3·7 to 14·7)	132·5 (119·5 to 145·6)	121·9 (110·6 to 133·8)	−7·9% (−12·1 to −3·0)*	152·6 (136·9 to 168·9)	158·2 (142·2 to 174·6)	3·7% (−1·3 to 9·6)
High-income Asia Pacific	2116·9 (1853·4 to 2425·1)	2066·2 (1792·0 to 2378·9)	−2·4% (−5·5 to 1·1)	1700·2 (1461·8 to 1966·5)	1690·3 (1440·5 to 1954·9)	−0·6% (−4·0 to 3·1)	224·2 (202·6 to 247·1)	170·3 (154·9 to 186·5)	−24·1% (−25·7 to −22·4)*	290·4 (259·3 to 323·6)	294·0 (261·0 to 326·5)	1·2% (−1·1 to 3·8)
High-income North America	1232·2 (1091·7 to 1389·6)	1229·0 (1093·0 to 1382·1)	−0·3% (−3·5 to 3·3)	995·7 (872·4 to 1132·3)	995·8 (877·9 to 1128·3)	0·0% (−3·3 to 3·7)	163·2 (148·9 to 179·3)	144·6 (133·2 to 156·8)	−11·4% (−13·5 to −9·2)*	191·9 (174·3 to 211·4)	191·7 (175·7 to 209·7)	−0·1% (−2·3 to 2·4)
Southern Latin America	1711·8 (1501·9 to 1935·9)	1595·1 (1401·4 to 1806·9)	−6·8% (−12·9 to 0·0)	1226·0 (1080·5 to 1385·3)	1165·5 (1023·4 to 1328·5)	−4·9% (−11·2 to 2·0)	203·8 (185·7 to 223·9)	161·6 (146·9 to 177·9)	−20·7% (−24·3 to −17·1)*	199·5 (181·9 to 219·1)	187·5 (169·9 to 206·5)	−6·0% (−10·4 to −1·6)*
Western Europe	963·9 (891·7 to 1043·1)	1139·3 (1055·8 to 1226·5)	18·2% (15·8 to 20·6)*	773·9 (713·3 to 841·9)	920·9 (853·8 to 990·7)	19·0% (16·5 to 21·6)*	127·6 (119·8 to 136·3)	126·1 (118·5 to 134·5)	−1·1% (−2·8 to 0·6)	163·1 (153·3 to 174·0)	181·8 (170·7 to 193·6)	11·5% (9·6 to 13·5)*
Latin America and Caribbean												
Andean Latin America	1370·9 (1211·6 to 1536·3)	1326·8 (1166·8 to 1506·3)	−3·2% (−8·0 to 1·6)	964·0 (849·6 to 1083·3)	936·9 (831·4 to 1061·3)	−2·8% (−7·7 to 2·1)	187·5 (171·8 to 204·7)	152·0 (139·9 to 165·7)	−19·0% (−21·9 to −16·0)*	147·5 (135·3 to 160·9)	146·5 (135·0 to 159·6)	−0·7% (−4·2 to 2·8)
Caribbean	1445·5 (1271·8 to 1638·0)	1638·0 (1450·9 to 1854·8)	13·3% (7·8 to 19·2)*	1039·7 (917·2 to 1180·8)	1138·8 (1007·0 to 1299·6)	9·5% (4·4 to 14·9)*	180·1 (163·3 to 198·5)	150·1 (137·6 to 164·2)	−16·6% (−19·1 to −14·0)*	161·9 (147·3 to 178·2)	167·2 (152·8 to 183·1)	3·2% (0·3 to 6·3)*
Central Latin America	913·5 (800·1 to 1030·3)	894·1 (785·7 to 1003·6)	−2·1% (−4·4 to 0·2)	645·2 (564·1 to 735·9)	628·2 (554·4 to 709·4)	−2·6% (−5·0 to −0·2)*	135·6 (123·6 to 149·7)	106·0 (97·4 to 115·3)	−21·9% (−23·6 to −20·2)*	109·5 (100·1 to 119·9)	108·6 (99·7 to 118·2)	−0·8% (−2·5 to 0·8)
Tropical Latin America	1077·2 (939·7 to 1222·7)	983·0 (868·7 to 1103·0)	−8·8% (−11·0 to −6·1)*	752·2 (657·1 to 860·7)	688·5 (609·3 to 783·1)	−8·5% (−10·8 to −5·8)*	140·7 (127·1 to 155·7)	105·6 (96·9 to 114·7)	−25·0% (−27·1 to −22·7)*	123·6 (111·9 to 136·5)	122·8 (112·4 to 134·0)	−0·6% (−2·8 to 1·6)
North Africa and Middle East												
North Africa and Middle East	1251·3 (1105·6 to 1414·0)	1303·1 (1159·3 to 1464·2)	4·1% (1·0 to 7·1)*	903·2 (794·1 to 1032·7)	915·4 (805·9 to 1045·6)	1·4% (−1·8 to 4·4)	175·7 (157·8 to 195·4)	141·1 (127·6 to 155·9)	−19·7% (−21·7 to −17·9)*	132·1 (119·4 to 146·3)	135·0 (122·2 to 149·7)	2·2% (0·1 to 4·3)*
South Asia												
South Asia	2047·8 (1805·0 to 2295·6)	1987·3 (1751·0 to 2228·6)	−3·0% (−5·2 to −0·6)*	1299·6 (1152·0 to 1463·1)	1257·8 (1117·7 to 1414·8)	−3·2% (−5·2 to −1·1)*	215·8 (195·6 to 237·5)	165·4 (151·2 to 181·1)	−23·4% (−24·6 to −22·0)*	165·7 (150·7 to 181·3)	163·9 (149·8 to 179·3)	−1·1% (−2·5 to 0·4)
Southeast Asia, east Asia, and Oceania												
East Asia	1266·9 (1088·2 to 1473·5)	1243·7 (1077·8 to 1430·2)	−1·8% (−5·5 to 1·2)	979·1 (831·3 to 1156·5)	973·3 (838·8 to 1122·7)	−0·6% (−4·6 to 2·8)	145·9 (129·5 to 166·3)	100·7 (91·1 to 111·5)	−31·0% (−33·7 to −28·7)*	142·1 (125·9 to 161·8)	150·6 (134·9 to 168·0)	6·0% (2·7 to 8·8)*
Oceania	1905·4 (1664·4 to 2168·9)	2088·4 (1844·1 to 2380·3)	9·6% (2·3 to 16·8)*	1407·4 (1228·5 to 1622·1)	1521·0 (1331·0 to 1737·7)	8·1% (1·3 to 15·6)*	265·7 (237·6 to 297·7)	260·4 (234·6 to 290·4)	−2·0% (−6·6 to 3·2)	195·2 (175·4 to 217·3)	206·3 (186·3 to 229·2)	5·7% (1·0 to 11·0)*
Southeast Asia	1714·5 (1504·1 to 1944·3)	1686·7 (1477·1 to 1917·9)	−1·6% (−4·3 to 1·3)	1296·8 (1127·2 to 1475·9)	1280·8 (1120·3 to 1465·5)	−1·2% (−4·1 to 1·9)	224·1 (201·4 to 249·0)	168·8 (152·5 to 186·3)	−24·7% (−26·3 to −23·0)*	184·3 (165·9 to 204·4)	186·4 (168·0 to 206·4)	1·1% (−0·8 to 3·2)
Sub-Saharan Africa												
Central sub-Saharan Africa	2512·8 (2205·7 to 2833·7)	2659·8 (2299·4 to 3034·0)	5·8% (−0·0 to 12·9)	1708·1 (1497·0 to 1924·7)	1760·9 (1529·4 to 1994·7)	3·1% (−2·7 to 9·9)	307·3 (276·3 to 340·3)	284·0 (255·7 to 314·3)	−7·6% (−11·7 to −2·9)*	193·9 (175·4 to 213·7)	197·2 (178·2 to 217·8)	1·7% (−2·5 to 6·8)
Eastern sub-Saharan Africa	2855·5 (2504·8 to 3218·7)	2669·5 (2350·6 to 3010·3)	−6·5% (−8·6 to −4·4)*	1883·2 (1655·3 to 2127·7)	1788·0 (1577·6 to 2016·8)	−5·0% (−7·0 to −2·9)*	338·2 (304·7 to 373·4)	291·7 (263·1 to 320·9)	−13·7% (−15·2 to −12·2)*	208·6 (188·8 to 229·4)	202·2 (183·0 to 221·9)	−3·1% (−4·6 to −1·5)*
Southern sub-Saharan Africa	2005·5 (1764·9 to 2249·6)	2010·9 (1764·8 to 2257·8)	0·3% (−2·8 to 3·2)	1426·6 (1254·7 to 1615·7)	1429·2 (1257·8 to 1619·2)	0·2% (−2·9 to 3·1)	239·1 (215·5 to 264·2)	199·6 (180·3 to 220·4)	−16·5% (−18·5 to −14·7)*	187·1 (169·2 to 205·8)	187·3 (169·3 to 206·7)	0·1% (−2·1 to 2·3)
Western sub-Saharan Africa	2405·2 (2105·2 to 2721·1)	2321·3 (2033·9 to 2619·2)	−3·5% (−7·1 to 0·4)	1641·9 (1441·1 to 1866·0)	1595·4 (1401·8 to 1807·5)	−2·8% (−6·7 to 1·4)	288·8 (259·7 to 319·8)	266·5 (241·1 to 294·2)	−7·7% (−10·8 to −4·7)*	187·2 (169·2 to 205·9)	183·6 (166·9 to 201·5)	−1·9% (−5·0 to 1·0)

Data are estimate (95% uncertainty interval). SDI=Socio-demographic index. *Significant increase or decrease.

Table 2: Global Burden of Diseases, Injuries, and Risk Factors Study 2017 prevalence of congenital heart disease, 1990–2017

Fonte: *Table2: Global Burden of Diseases, Injuries, and Risk Factors Study 2017 congenital heart disease mortality, 1990–2017* (ZIMMERMAN et al., 2020)

Na tabela 2 é apresentado o estudo de Zimmerman (2020) sobre nascimento de crianças com cardiopatia congênita entre 1990 a 2017 em 195 países. Neste estudo verificamos que as taxas nascimento de crianças com cardiopatia congênita alteraram pouco no tempo ou nos grupos SDI (*Socio-demographic Index*). Aproximadamente 12 milhões de pessoas vivem com cardiopatia congênita em todo o mundo, observa-se neste estudo um aumento de 18,7% entre 1990 e 2017.

Os diagnósticos e tratamentos para doenças cardíacas congênitas melhoraram dramaticamente nos últimos 80 anos, estudos recentes mostram uma melhora substancial na sobrevivência nas regiões desenvolvidas do mundo, mas as mesmas taxas de sucesso ainda não são vistas nas regiões em desenvolvimento. No mundo, aproximadamente nove décimos das crianças nascidas com cardiopatia congênita vivem em locais com pouco ou nenhum tratamento, grupo mais propensos a morrer antes do quinto aniversário (ZIMMERMAN et al., 2020).

As cardiopatias congênitas causaram mais 260 mil mortes em todo o mundo em 2017, um declínio de 34,5% em relação a 1990, com aproximadamente 180 mil mortes entre bebês (com idade <1 anos). As taxas de mortalidade por doenças cardíacas congênitas diminuíram com o aumento do SDI (*Socio-demographic Index*), a maioria das mortes ocorreu em países com SDI baixo e médio-baixo (ZIMMERMAN et al., 2020). Estas são algumas estatísticas extraídas do estudo de Zimmerman (2020) e apresentadas na tabela 3.

Tabela 3 - Estudo sobre mortalidade de portadores de cardiopatia congênitas, 1990–2017

	Number of deaths (all ages)			Mortality per 100 000 children younger than 1 year			Age-standardised mortality per 100 000 individuals		
	1990	2017	Percentage change, 1990–2017	1990	2017	Percentage change, 1990–2017	1990	2017	Percentage change, 1990–2017
Global	398 580 (268 422 to 506 469)	261 247 (216 567 to 308 159)	-34·5% (-44·6 to -19·8)*	217·1 (146·8 to 277·5)	131·0 (106·5 to 155·3)	-39·7% (-49·1 to -27·1)*	6·3 (4·3 to 8·0)	3·9 (3·2 to 4·6)	-39·0% (-48·2 to -26·0)*
Low SDI	67 340 (25 251 to 109 078)	83 873 (56 785 to 110 218)	24·6% (-13·1 to 130·6)	179·5 (71·7 to 287·6)	167·9 (115·7 to 221·1)	-6·5% (-35·0 to 63·1)	5·5 (2·1 to 8·9)	4·9 (3·4 to 6·4)	-10·6% (-37·9 to 61·8)
Low-middle SDI	103 086 (60 630 to 140 879)	82 667 (64 667 to 104 233)	-19·8% (-37·2 to 8·1)	225·6 (135·8 to 330·2)	147·9 (110·5 to 189·4)	-34·5% (-48·3 to -14·6)*	6·8 (4·0 to 9·2)	4·4 (3·5 to 5·6)	-34·9% (-49·0 to -12·9)*
Middle SDI	127 734 (98 849 to 154 111)	58 512 (50 989 to 63 052)	-54·2% (-64·8 to -41·8)*	249·3 (190·1 to 300·6)	120·0 (102·8 to 132·4)	-51·8% (-63·5 to -38·5)*	7·0 (5·5 to 8·5)	3·5 (3·1 to 3·8)	-49·6% (-61·3 to -36·2)*
High-middle SDI	76 987 (61 961 to 89 280)	26 948 (24 814 to 29 373)	-65·0% (-71·0 to -54·8)*	262·8 (206·5 to 308·3)	99·9 (89·2 to 109·1)	-62·0% (-69·3 to -49·2)*	7·3 (5·8 to 8·4)	3·0 (2·7 to 3·7)	-59·3% (-66·5 to -47·4)*
High SDI	21 607 (17 568 to 23 083)	8500 (7653 to 9992)	-60·7% (-64·1 to -52·7)*	101·7 (78·2 to 110·8)	36·6 (32·8 to 41·7)	-64·0% (-68·0 to -54·5)*	3·1 (2·5 to 3·4)	1·2 (1·1 to 1·4)	-61·7% (-65·5 to -52·9)*
Central Europe, eastern Europe, and central Asia									
Central Asia	3834 (3183 to 4356)	3155 (2491 to 3815)	-17·7% (-34·2 to 0·7)	150·0 (125·3 to 170·0)	119·1 (94·2 to 148·1)	-20·6% (-36·2 to -2·7)*	4·2 (3·5 to 4·8)	3·5 (2·7 to 4·2)	-17·2% (-34·0 to 1·2)
Central Europe	4969 (3898 to 5568)	1111 (970 to 1247)	-77·6% (-80·8 to -70·1)*	212·6 (162·0 to 242·9)	59·4 (51·9 to 68·4)	-72·1% (-77·2 to -60·9)*	5·7 (4·4 to 6·4)	1·8 (1·6 to 2·0)	-68·9% (-73·7 to -57·5)*
Eastern Europe	7409 (6242 to 8297)	2751 (2390 to 3530)	-62·9% (-68·3 to -51·8)*	161·8 (136·7 to 182·1)	60·7 (52·0 to 77·7)	-62·5% (-68·8 to -50·0)*	4·7 (3·9 to 5·3)	2·0 (1·7 to 2·6)	-56·8% (-63·6 to -42·7)*
High income									
Australasia	374 (321 to 429)	197 (159 to 236)	-47·5% (-56·1 to -37·2)*	70·5 (60·3 to 81·6)	26·3 (19·8 to 33·9)	-62·8% (-72·3 to -51·0)*	2·3 (2·0 to 2·6)	0·9 (0·7 to 1·1)	-59·8% (-67·0 to -50·2)*
High-income Asia Pacific	3608 (2624 to 4060)	897 (792 to 1121)	-75·2% (-79·1 to -63·2)*	104·1 (73·3 to 120·3)	29·3 (25·4 to 35·1)	-71·9% (-77·0 to -57·4)*	3·3 (2·4 to 3·8)	1·0 (0·8 to 1·2)	-71·2% (-76·1 to -57·0)*
High-income North America	6430 (5446 to 7079)	3522 (3137 to 4311)	-45·2% (-49·3 to -36·0)*	84·1 (70·0 to 91·9)	37·8 (33·2 to 45·8)	-55·0% (-59·9 to -46·5)*	2·8 (2·3 to 3·0)	1·4 (1·2 to 1·7)	-51·0% (-55·0 to -42·4)*
Southern Latin America	2129 (1766 to 2450)	1259 (946 to 1509)	-40·9% (-54·6 to -24·7)*	159·3 (131·0 to 185·2)	85·6 (63·1 to 106·0)	-46·3% (-60·7 to -29·0)*	4·3 (3·6 to 5·0)	2·4 (1·8 to 2·9)	-43·8% (-57·2 to -28·2)*
Western Europe	7630 (5736 to 8323)	2539 (2138 to 2848)	-66·7% (-70·4 to -59·1)*	99·4 (64·5 to 113·0)	29·2 (22·3 to 33·2)	-70·6% (-74·5 to -61·4)*	3·0 (2·2 to 3·3)	1·0 (0·8 to 1·1)	-68·5% (-72·2 to -60·0)*
Latin America and Caribbean									
Andean Latin America	1891 (1359 to 2596)	2026 (1338 to 2593)	7·1% (-44·3 to 55·6)	115·0 (81·2 to 161·8)	108·1 (69·5 to 143·4)	-6·0% (-53·5 to 43·5)	3·4 (2·4 to 4·6)	3·1 (2·1 to 4·0)	-6·9% (-51·2 to 34·3)
Caribbean	2744 (1692 to 3463)	2120 (1442 to 2825)	-22·7% (-41·9 to -3·0)*	218·6 (138·6 to 281·3)	193·4 (129·5 to 270·3)	-11·5% (-35·5 to 14·4)	6·6 (4·1 to 8·3)	5·4 (3·7 to 7·2)	-17·9% (-38·6 to 3·2)
Central Latin America	10 767 (9594 to 14 534)	8965 (6154 to 10 775)	-16·7% (-52·9 to 1·9)	165·6 (146·8 to 222·6)	129·2 (85·3 to 150·7)	-22·0% (-58·5 to -1·9)*	4·8 (4·3 to 6·4)	3·7 (2·6 to 4·3)	-21·8% (-55·9 to -4·2)*
Tropical Latin America	19 295 (9542 to 26 614)	5962 (5095 to 6896)	-69·1% (-78·5 to -38·6)*	449·3 (209·9 to 625·3)	144·5 (118·1 to 165·1)	-67·8% (-78·1 to -32·1)*	11·3 (5·6 to 15·5)	3·7 (3·1 to 4·3)	-67·1% (-77·7 to -35·0)*
North Africa and Middle East									
North Africa and Middle East	60 031 (30 316 to 84 703)	35 710 (29 606 to 44 359)	-40·5% (-54·5% to -0·3%)*	450·8 (221·5 to 653·2)	211·7 (168·5 to 269·5)	-53·0% (-64·0% to -22·2%)*	11·8 (6·0 to 16·6)	5·8 (4·8 to 7·2)	-51·0% (-62·6% to -18·6%)*
South Asia									
South Asia	75 091 (38 351 to 109 094)	66 984 (50 484 to 81 590)	-10·8% (-38·2 to 34·5)	154·8 (76·6 to 223·5)	138·5 (95·5 to 176·6)	-10·6% (-38·3 to 29·4)	4·8 (2·5 to 7·0)	4·0 (3·0 to 4·9)	-17·0% (-42·1 to 21·0)
Southeast Asia, east Asia, and Oceania									
East Asia	93 878 (80 522 to 113 589)	31 885 (29 266 to 35 408)	-66·0% (-72·5 to -58·2)*	254·8 (217·6 to 312·0)	112·5 (103·0 to 125·7)	-55·9% (-66·0 to -46·2)*	7·4 (6·3 to 8·9)	3·4 (3·1 to 3·7)	-54·5% (-63·6 to -44·1)*
Oceania	775 (440 to 1071)	1307 (790 to 1793)	68·8% (23·8 to 126·7)*	244·2 (137·7 to 348·0)	226·4 (126·0 to 327·7)	-7·3% (-35·6 to 32·8)	8·3 (4·9 to 11·7)	7·7 (4·8 to 10·5)	-6·7% (-30·5 to 23·6)
Southeast Asia	42 680 (25 044 to 55 932)	23 976 (19 721 to 27 605)	-43·8% (-56·5 to -19·3)*	268·3 (151·0 to 352·8)	153·1 (120·2 to 181·0)	-42·9% (-55·3 to -17·6)*	7·3 (4·3 to 9·6)	4·4 (3·6 to 5·1)	-40·4% (-53·8 to -15·3)*
Sub-Saharan Africa									
Central sub-Saharan Africa	8152 (3377 to 14 002)	11 257 (7101 to 17 318)	38·1% (-0·3 to 128·0)	231·0 (112·1 to 388·3)	189·5 (121·2 to 288·3)	-17·9% (-43·0 to 30·0)	7·7 (3·2 to 13·1)	5·9 (3·8 to 9·0)	-23·4% (-44·7 to 26·7)
Eastern sub-Saharan Africa	25 818 (11 083 to 44 514)	26 998 (19 197 to 38 981)	4·6% (-21·1 to 97·0)	208·4 (100·9 to 344·1)	141·5 (99·7 to 203·7)	-32·1% (-47·8 to 16·3)	6·9 (3·1 to 11·9)	4·4 (3·1 to 6·3)	-36·9% (-52·6 to 17·2)
Southern sub-Saharan Africa	1568 (1301 to 1894)	1252 (1084 to 1444)	-20·1% (-36·4 to 4·2)	62·7 (51·0 to 78·2)	42·1 (35·0 to 50·5)	-32·8% (-48·2 to -10·3)*	2·4 (2·0 to 2·8)	1·5 (1·3 to 1·8)	-34·8% (-47·8 to -16·2)*
Western sub-Saharan Africa	19 507 (7079 to 30 605)	27 373 (15 756 to 39 149)	40·3% (7·1 to 121·7)*	164·6 (68·0 to 261·0)	121·1 (76·8 to 169·4)	-26·4% (-46·0 to 12·2)	5·5 (2·1 to 8·6)	4·0 (2·4 to 5·6)	-27·5% (-44·6 to 14·0)

Data in parentheses are 95% uncertainty intervals. SDI=Socio-demographic index. *Significant increase or decrease.

Table 1: Global Burden of Diseases, Injuries, and Risk Factors Study 2017 congenital heart disease mortality, 1990–2017

Fonte: *Table1: Global Burden of Diseases, Injuries, and Risk Factors Study 2017 congenital heart disease mortality, 1990–2017* (ZIMMERMAN et al., 2020)

2.2 ALGORITMOS DE APRENDIZAGEM DE MÁQUINA

O uso de algoritmos de aprendizagem de máquina *(machine learning),* como ferramenta de apoio à tomada de decisão na área da medicina, vêm aumentando nos últimos anos (RUIZ-FERNÁNDEZ et al., 2016). São encontrados estudos que utilizam os algoritmos de aprendizagem de máquina *(machine learning)* na previsão de mortalidade e no tempo de permanência na UTI (AWAD; BADER-EL-DEN; MCNICHOLAS, 2017).

Existem vários algoritmos de aprendizagem de máquina (*machine learning)* e cada um deles oferece benefícios diferentes dependendo da tarefa e dos tipos de conjunto de dados nos quais são usados. Os critérios utilizados para avaliar esses algoritmos são: precisão, complexidade computacional, robustez, escalabilidade, integração, compreensibilidade e estabilidade (RUIZ-FERNÁNDEZ et al., 2016). Alguns algoritmos possuem pontos fortes sobre outros em certos aspectos, como por exemplo, as Árvores de Decisão (DT), em que é possível medir a importância relativa de cada característica (*variável*) na previsão, ao contrário das Rede Neurais Artificiais (RNA), em que não conseguimos analisar a importância de cada variável em sua modelagem, no entanto, apresenta adaptabilidade, tolerância a falhas e boa classificação, mesmo quando os dados apresentam ruídos (XU et al., 2016). Na perspectiva computacional, os algoritmos de Árvores de Decisão (DT), *Random Forest* (RF), redes neurais artificiais, os métodos de Kernel, exemplo SVM (*Support Vector Machine*), podem manipular amostras de dados de grande porte e integrar o conhecimento de base na análise (DIETTERICH, 2000).

Dentre as diferentes técnicas de aprendizagem de máquina (*Machine Learning)*, aquelas relacionadas à classificação, como Redes Neurais Artificiais (RNA) ou Árvores de Decisão, são as mais utilizadas em pesquisas médicas e vêm demonstrando bom desempenho no diagnóstico clínico (MEJIA et al., 2018)(LI et al., 2017)(RUIZ-FERNÁNDEZ et al., 2016).

Observa-se em vários estudos que o performance de previsão dos algoritmos de aprendizagem de máquina (*machine learning*) vêm superando os métodos tradicionais (AWAD; BADER-EL-DEN; MCNICHOLAS, 2017)(XU et al., 2016)(AWAD et al., 2017). Nos estudos mais recentes no campo de análise de doenças cardíacas congênitas, observa-se o uso de alguns algoritmos de Redes Neurais Artificiais (RNA), como por exemplo *Multilayer Perceptron (MLP),* e de Árvores de Decisão (DT) (RUIZ-

FERNÁNDEZ et al., 2016)(LI et al., 2017). Ao ampliar a pesquisa para modelos de previsão de mortalidade, a variedade de algoritmos é ainda maior (AWAD; BADER-EL-DEN; MCNICHOLAS, 2017), são vários algoritmos utilizados nos estudos, destacando-se os algoritmos de Redes Neurais Artificiais (RNA) e algoritmos *Ensembles* (DIETTERICH, 2000) baseados em Árvores de Decisão (DT).

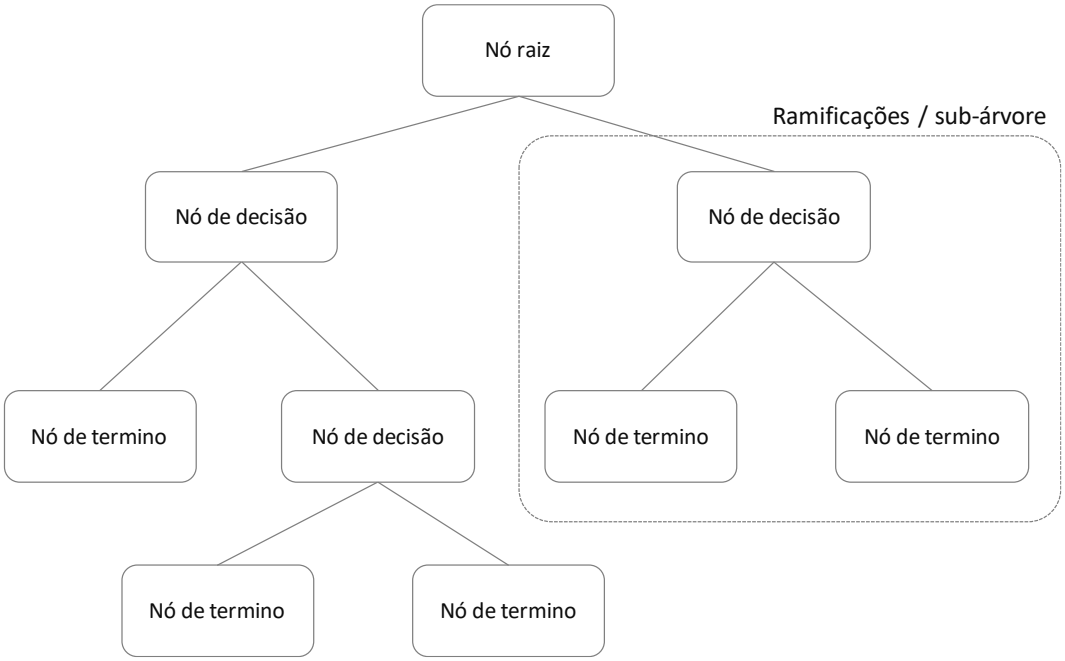
Como observado, não é possível concluir que determinado algoritmo seja capaz de compreender e responder assertivamente a todas as situações-problema. Para cada situação exige-se um estudo para identificar o melhor algoritmo a ser aplicado. Neste estudo são selecionados os algoritmos mais utilizados em aplicações médicas, algoritmos baseados em Redes Neurais Artificiais (RNA) e Árvores de Decisão (DT). O algoritmo de Redes Neurais Artificiais (RNA) selecionado neste estudo é o *Multilayer Perceptron* (MLP) e para as Árvores de Decisão (DT), foram selecionados os algoritmos *Ensembles* (DIETTERICH, 2000): *Random Forest* (RF), *Extra Trees* (ET), *Stochastic Gradient Boosting* (SGB), *AdaBoost Classification* (ABC) e *Bagged Decision Trees* (BDT).

2.2.1 Árvore de decisão – Algoritmos *ensemble*

O modelo de árvore de decisão é a base dos modelos *Ensemble* selecionados neste estudo, portanto entender o funcionamento de uma arvore de decisão ajudará na compreensão e especificidade de cada algoritmo selecionado.

Árvore de decisão é um algoritmo de aprendizagem supervisionada, amplamente utilizado em problemas de classificação (MITCHELL, 1997) e tornou-se base para os algoritmos *Ensembles* (DIETTERICH, 2000), exemplo *Random Forest* (RF). Este algoritmo funciona com variáveis categóricas e contínuas, como também é pouco sensível a ruídos nos dados, como pontos fora da curva (*outliers*) e valores faltantes (*missing values*). Sua arquitetura em forma de árvore facilita a compreensão dos resultados, sendo muito intuitiva e permite relacionar as hipóteses facilmente. Outra vantagem desse algoritmo é a sua capacidade de identificar as variáveis mais significativas e relações entre duas ou mais variáveis, auxiliando na exclusão ou geração de novas variáveis.

Figura 1 - Arquitetura Árvore de Decisão



Fonte: Adaptado de Mitchell (1997)

Na figura 1 observa-se um exemplo de arquitetura de uma árvore de decisão, composta por nó raiz, nó de decisão, nó de termino (ou folhas) e ramificações (sub-árvore). E abaixo os detalhes de cada uma dessas estruturas:

- Nó raiz: início da árvore, representa a população inteira ou amostra;
- Nó de decisão: nó que que é dividido em sub-nós adicionais;
- Nó de término (ou folha): os nós não divididos são chamados folha ou nó de Término;
- Ramificação (sub-árvore): uma sub-seção da árvore inteira é chamada de ramificação ou sub-árvore;
- Nó pai e nó filho: nó que é dividido é chamado de nó pai e os sub-nós são os nós filhos do nó pai.

A árvore de decisão segue uma abordagem *top-down* e *greedy* (guloso), ou seja, sua construção inicia-se no topo da árvore com todas as observações e sucessivamente é dividido em novos ramos no sentido inferior da árvore, apenas se preocupando com a divisão atual e não com divisões futuras, razão a qual é conhecido como *greedy algorithm* (algoritmo guloso).

Para definir a variável em cada nó e suas divisões, utiliza-se testes estatísticos que avaliam a capacidade de classificação. Estes testes estatísticos mudam de algoritmo para algoritmo, sendo os mais usuais o índice *Gini impurity* (impureza Gini) e *Information gain* (ganho de informação). O índice Gini é utilizado pelo algoritmo de geração de árvore conhecido como CART (*Classification And Regression Tree*), desenvolvido por Breiman. E o *Information gain* (ganho de informação) é utilizado pelos algoritmos de geração de árvore ID3 e C4.5, ambos desenvolvidos por Ross Quinlan.

O índice Gini é uma medida de desigualdade de dados, variando de 0 a 1, onde 1 significa desigualdade total dos dados e 0 igualdade total.

O índice *Gini* mede a impureza de *D*, de um conjunto de dados de treinamento através da função:

$$Gini(D) = 1 - \sum_{i=1}^{m} Pi^2 \quad (1)$$

onde p_i é a probabilidade da variável em *D* pertencer à classe C_i e é estimada por $|C_i,D|/|D|$. A soma é computada sobre *m* classes. Se os dados pertencem a uma mesma classe a impureza é 1 − 1 = 0 (os dados são puros).

Este índice considera divisões binárias para cada variável, onde são examinados todos os possíveis subconjuntos formados usando os valores conhecidos da variável para determinar a melhor divisão. Dada um conjunto de dados, este teste é satisfeito se o valor da variável estiver entre os valores listados no conjunto S_A. Se uma divisão binária de *A* particiona *D* em D_1 e D_2, o índice Gini de *D* dado o particionamento é

$$Gini_A(D) = \frac{|D_1|}{|D|} Gini(D_1) + \frac{|D_2|}{|D|} Gini(D_2) \quad (2)$$

O subconjunto que dá o mínimo índice *Gini* (menos impureza) para a variável é selecionado como o subconjunto de divisão. A redução de impureza da variável *A* que é utilizada na divisão é dado por:

$$\Delta Gini\ A = Gini\ (D) - Gini_A(D) \quad (3)$$

Assim, a variável que maximiza a redução de impureza é escolhida como para a divisão do conjunto. Ou seja, quanto mais puro for a partição criada por *A*, menor é o $Gini_A(D)$ e maior será o $\Delta GiniA$.

O *Information gain* (ganho de informação) é baseado no conceito de entropia. Essa função minimiza a informação necessária para classificar os dados das divisões resultantes e reflete a impureza ou a aleatoriedade destas divisões. Para se calcular a entropia de cada classe do conjunto de dados é utilizado a função:

$$Entropy(D) = -\sum_{i=1}^{m} p_i \log_2(p_i) \quad (4)$$

onde p_i é a probabilidade da classe C_i em *D*, determinada dividindo-se o número de dados da classe C_i em *D* por |*D*|.

Com o objetivo de dividir o conjunto de dados de *D* usando alguma variável *A*, com *v* valores distintos *a1, a2, ...av*. calculamos a quantidade de informação necessária para a classificação exata de cada variável utilizando a função:

$$Info_A(D) = \frac{|D_1|}{|D|} Entropy(D_1) + \frac{|D_2|}{|D|} Entropy(D_2) \ldots + \frac{|D_v|}{|D|} Entropy(D_v) \quad (5)$$

Por fim, o *Information gain* (ganho de informação) é definido como a diferença entre a informação necessária para particionar os dados considerando o conjunto original e a informação necessária para particionar os dados após o uso da variável *A* na divisão:

$$Gain(A) = Info(D) - Info_A(D) \quad (6)$$

Independente do algoritmo de geração de árvore utilizado, CART ou C4.5, as árvores de decisão comumente sofrem de sobreajuste (*overfitting*) devido a sua abordagem *top-down* e *greedy,* responsável por gerar árvores com grandes profundidades (muitas divisões) que decoram o conjunto de treino, impossibilitando a generalização em novos conjuntos de dados. Para minimizar este efeito, utilizam-se

critérios de parada, definindo a profundidade da árvore e/ou a quantidade máxima de folhas, técnica denominada como “poda”.

Em geral, uma única árvore de decisão pode sofrer de variância, mas ter um baixo viés. Diante deste comportamento, surgem os métodos *Ensemble* com a proposta de agrupar vários modelos preditivos com o objetivo de melhorar a precisão e a estabilidade do modelo (DIETTERICH, 2000).

Os métodos *Ensemble* mais conhecidos para arvores de decisão são *Bagging* (BREIMAN, 1996) e *Boosting* (FREUND; SCHAPIRE, 1999). *Bagging* é uma técnica usada para reduzir a variância das previsões, combina o resultado de vários classificadores, modelados em diferentes sub-amostras do mesmo conjunto de dados, base dos algoritmos *Random Forest* (RF), *Extra Trees* (ET) e *Bagged Decision Trees* (BDT). E o princípio do *Boosting* é converter várias aprendizagens fracas em uma aprendizagem forte, utilizado nos algoritmos *Stochastic Gradient Boosting* (SGB) e *AdaBoost Classification* (ABC).

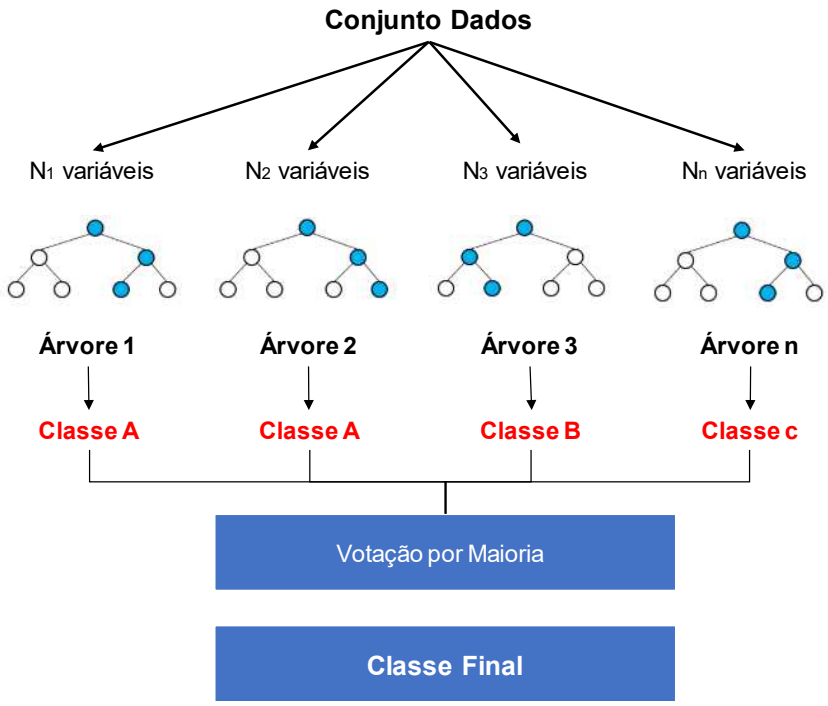
Nos próximos itens são detalhados os algoritmos *Ensemble* selecionados neste estudo.

2.2.1.1 Bagged Decision Trees (BDT)

Bagged Decision Trees é um algoritmo *Ensemble*, método que combina vários modelos para produzir um novo algoritmo com o objetivo de melhorar o desempenho preditivo em relação ao modelo individual (DIETTERICH, 2000). Este algoritmo combina o método de *Bagging* (BREIMAN, 1996) com modelos de Árvores de Decisão.

O método *Bagging* foi proposto por Breiman em 1996, baseado na amostragem *Bootstrap* dos dados originais. O *Bagging* consiste em gerar várias versões de um preditor e utilizá-las para obter um preditor agregado por votação de seus resultados (BREIMAN, 1996). E *Bootstrap* é um método de amostragem aleatória com reposição em que o número de objetos escolhidos é igual ao número inicial (EFRON, 1979). Testes em conjuntos de dados reais e simulados usando árvores de classificação e regressão mostram que o método *Bagging* pode proporcionar ganhos substanciais em precisão (BREIMAN, 2001) (DIETTERICH, 2000)(KOTSIANTIS; TSEKOURAS; PINTELAS, 2005).

Figura 2 - Método Bagging



Fonte: Adaptado de BREIMAN (1996)

A topologia de cada árvore é simples, composta por ramos, nós e folhas. Os nós contêm os atributos dos quais a função objetivo é dependente, e seus valores migram para as folhas através dos ramos. No processo de classificação de um novo caso é necessário percorrer a árvore através dos ramos até uma folha, passando por todos os valores de atributos (MITCHELL, 1997).

Uma única Árvore de Decisão geralmente se adapta aos dados dos quais está aprendendo, porque aprende apenas com um caminho de decisão, cenário com baixo viés, mas alta variabilidade. O algoritmo *Bagged Decision Trees* (BDT) se beneficia do baixo viés das Árvores de Decisão e a baixa variabilidade gerada pelo método *Bagged*. A baixa variabilidade é resultado da média das previsões das Árvores de Decisão geradas no método *Bagging* (HASTIE; TIBSHIRANI; FRIEDMAN, 2009).

2.2.1.2 Random Forest (RF)

O algoritmo *Random Forest* (RF) é muito popular entre os algoritmos de aprendizagem de máquina (*Machine Learning)* e pode ser utilizado para tarefas de classificação e regressão. Também é conhecido pela sua precisão, facilidade em treinar e ajustar seus parâmetros (HASTIE; TIBSHIRANI; FRIEDMAN, 2009).

Trata-se de um algoritmo de aprendizagem supervisionada gerado a partir da combinação (*Ensemble*) de preditores de Árvores de Decisão treinados com o método de *Bagging* (BREIMAN, 1996). Diferente do algoritmo *Bagged Decision Trees* (BDT), que utiliza os mesmos atributos (variáveis) para cada árvore gerada, no *Random Forest* (RF), cada árvore é construída com atributos (variáveis) totalmente aleatórios, gerando árvores totalmente distintas. Essa diferença no processo de construção das árvores gera menor erro de generalização e redução do tempo de processamento comparado a outros algoritmos baseados em árvores de decisão, segundo o estudo apresentado por Breiman em 2001.

No *Random Forest* (RF) cada árvore é construída utilizando-se de uma amostra de inicialização diferente dos dados originais. Cerca de um terço dos casos são deixados de fora da amostra de *Bootstrap* e não são utilizados na construção de novas árvore. Esses dados são usados para obter uma estimativa imparcial do erro de classificação à medida que as árvores são adicionadas à floresta e também para obter estimativas de importância de cada variável no modelo.

Na construção do *Random Forest* (RF) todas as árvores treinadas são combinadas por uma votação simples, utilizando-se do erro médio de todas as amostras. O uso da agregação de auto inicialização reduz o erro quadrático médio, diminuindo a variação do classificador treinado. A eficácia do *Bagging* reside no fato de que os algoritmos básicos (árvores de decisão) são treinados em várias amostras aleatórias e seus resultados podem variar bastante, enquanto seus erros são mutuamente compensados na votação (BREIMAN, 2001).

Sempre que uma divisão de um nó é feita na variável *n*, o critério de impureza gini para os dois nós descendentes é menor que o nó pai. A soma das diminuições de gini para cada variável individual em todas as árvores da floresta fornece uma importância variável rápida que geralmente é muito consistente com a medida de importância da permutação.

Breiman em seu estudo lista os seguintes benefícios em utilizar o *Random Forest* (RF) (BREIMAN, 2001):

- É insuperável em precisão entre os algoritmos atuais;
- Roda eficientemente em grandes bancos de dados;
- Pode lidar com milhares de variáveis de entrada sem a necessidade de exclusão;
- Fornece estimativas de quais variáveis são importantes na classificação;
- Gera uma estimativa interna imparcial do erro de generalização à medida que a construção da floresta progride;
- Possui um método eficaz para estimar dados ausentes garantindo precisão;
- Possui métodos para equilibrar erros em conjuntos de dados desequilibrados da população de classes;
- As florestas geradas podem ser salvas para uso futuro em outros dados;

Um grande problema de qualquer algoritmo de aprendizagem de máquina (*Machine Learning)* é o sobreajuste (*overfitting*), efeito que o *Random Forest* sofre em menor frequência, pois trabalha com subconjuntos aleatórios das características e constrói árvores menores a partir de tais subconjuntos. Em contrapartida, em algumas aplicações o grande número de árvores poderá tornar o algoritmo lento e inviabilizar a previsão em tempo real (HASTIE; TIBSHIRANI; FRIEDMAN, 2009).

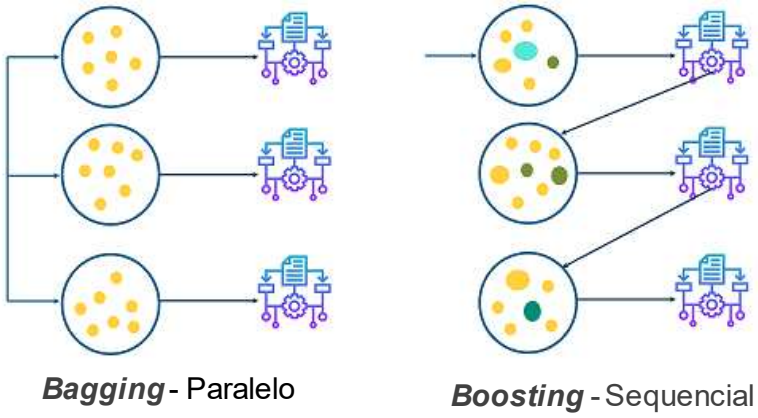
2.2.1.3 Stochastic Gradient Boosting (SGB)

O *Stochastic Gradient Boosting* (SGB) é um algoritmo *Ensemble* (DIETTERICH, 2000) que combina o método *Boosting* (FREUND; SCHAPIRE, 1999) com modelos de árvores de decisão. O princípio desse modelo é aprender com os erros e ajustá-los a cada nova geração de árvore, utilizando funções de perda gradual para minimizar os erros a cada etapa de aprendizado. (FRIEDMAN, 1999).

A abordagem por *Boosting* original foi proposta por Schapire em 1990, considerado por muitos pesquisadores um dos métodos de aprendizado mais poderoso introduzido nas últimas décadas (HASTIE; TIBSHIRANI; FRIEDMAN, 2009). Esse método foi originalmente projetado para problemas de classificação, mas

também pode ser aplicado para a regressão com ganhos. O método baseia-se na combinação dos resultados de muitos classificadores "fracos" no objetivo de produzir um poderoso "comitê" de classificadores. Nessa perspectiva, o *Boosting* tem uma semelhança com *Bagging* e outras abordagens baseadas no método *Ensemble* (DIETTERICH, 2000). Sua principal diferença em relação ao método *Bagging* é que a reamostragem é realizada para gerar aprendizado complementar e a importância do voto é ponderada com base no desempenho de cada modelo, em vez da atribuição de mesmo peso para todos os votos (QUINLAN, 2006).

Figura 3 - Comparação Bagging x Boosting



Fonte: Adaptado de DIETTERICH (2000)

A idéia geral do algoritmo *Stochastic Gradient Boosting* (SGB) é calcular uma sequência de árvores, em que cada árvore sucessiva é construída utilizando-se dos resíduos de previsão da árvore anterior e seu resultado é combinado ao longo do caminho.

Entre as vantagens do uso do *Stochastic Gradient Boosting* (SGB) podemos destacar:

- Suporta diferentes funções de erro;
- Trabalha bem com interações.

E suas desvantagens são:

- Propenso a excesso de ajuste;
- Requer ajuste cuidadoso de diferentes hiperparâmetros;
- Sensíveis a ruídos nos dados;

- O treinamento geralmente leva mais tempo porque as árvores são geradas seqüencialmente.

2.2.1.4 Extra Trees (ET)

Extra Trees (ET), também conhecido como *Extremely Randomized Tree* (Árvores Extremamente Aleatórias) são algoritmos de aprendizagem supervisionada gerados a partir da combinação (*Ensemble*) de preditores de árvores.

Esse algoritmo difere de outros algoritmos *Ensemble* baseados em árvore em seu processo de aleatorização e amostragem. Este algoritmo escolhe os pontos de cortes completamente aleatórios, diferente dos demais que utilizam as melhores divisões. E também utiliza toda a amostra de aprendizado, em vez de uma réplica de *bootstrap*, para cultivar as árvores.

As previsões das árvores são agregadas para produzir a previsão final, calculada a partir dos votos da maioria para o problema de classificação e média aritmética para os problemas de regressão. Do ponto de vista da variação de viés, a lógica por trás do algoritmo *Extra Trees* (ET) é que a randomização do ponto de corte combinada com a média do conjunto é capaz de reduzir fortemente sua variação quando comparada a outros métodos comumente utilizados (FRIEDMAN, 1999).

Em resumo, podemos dizer que o funcionamento do algoritmo *Extra Trees* (ET) se assemelha ao algoritmo *Random Forest* (RF), mas com a diferença de não inicializar observações e gerar as divisões dos nós de forma totalmente aleatória. Essas diferenças fazem que os algoritmos *Extra Trees* (ET) tenham menor variabilidade comparado aos demais algoritmos *Ensemble*.

2.2.1.5 AdaBoost Classification (ABC)

O algoritmo *AdaBoost Classification (Adaptive Boosting Classification)* foi apresentado por Freund e Schapire em 1997. Trata-se de um algoritmo de aprendizagem supervisionada gerado a partir da combinação (*Ensemble*) de preditores de árvores com o método *Boosting* (FREUND; SCHAPIRE; HILL, 1996),

utilizando o princípio de combinar regras relativamente fracas e imprecisas para criar uma regra altamente precisa. O algoritmo *AdaBoost Classification* (ABC) foi o primeiro algoritmo prático que utilizou o método *Boosting* e continua sendo um dos mais amplamente utilizados e estudados (FREUND; SCHAPIRE; HILL, 1996) (SCHAPIRE, 2013).

Como *Stochastic Gradient Boosting* (SGB), o *AdaBoost Classification* (ABC) converte um conjunto de aprendizagem fraco num único conjunto forte, mas se diferem na forma como geram os conjuntos de aprendizagem no processo de treinamento. No *AdaBoost Classification* (ABC) a distribuição da amostra é alterada modificando os pesos anexados a cada instância, os pesos das instâncias previstas incorretamente são aumentados e das instâncias previstas corretamente são reduzidas. Depois de treinado, o conjunto de aprendizado fraco é adicionado ao forte de acordo com seu desempenho, quanto mais alto o desempenho, maior a contribuição para o conjunto de aprendizagem mais forte.

No *Stochastic Gradient Boosting* (SGB) não ocorre a alteração na distribuição da amostra, em vez de treinar uma nova distribuição de amostra, o conjunto de aprendizagem treina os erros restantes do conjunto de aprendizagem forte. Diferente do *AdaBoost Classification* (ABC), a contribuição do conjunto de aprendizagem fraco para o forte é calculada utilizando um processo de otimização de descida de gradiente.

O algoritmo *AdaBoost Classification* (ABC) apresenta muitas vantagens, é rápido, simples e fácil de programar. Não possui muitos parâmetros de ajustes e não requer conhecimento prévio sobre o conjunto de aprendizagem fraca, tornando-o flexível com qualquer método. (FREUND; SCHAPIRE, 1997)

2.2.2 Multilayer Perceptron (MLP)

As redes *Multilayer Perceptron (MLP)* são caracterizadas pelas elevadas possibilidades de aplicações em diversos tipos de problemas relacionados com as mais diferentes áreas do conhecimento, sendo também consideradas uma das arquiteturas mais versáteis quanto à aplicabilidade. Entre essas potenciais aplicações, tem-se os seguintes destaques (SILVA; SPATTI; FLAUZINO, 2010):

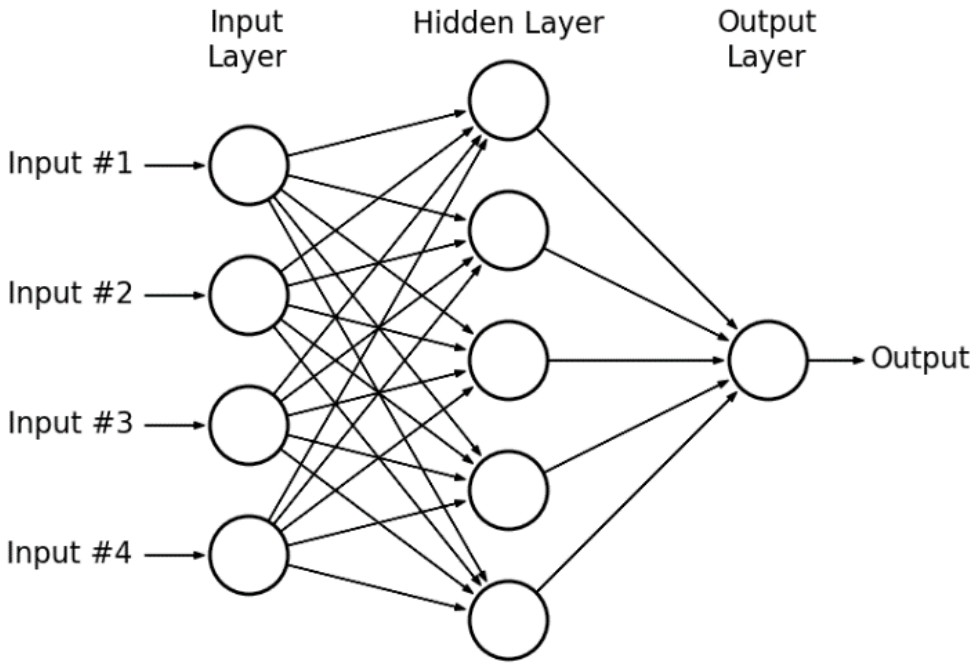
- Aproximação universal de funções;
- Reconhecimento de padrões;
- Identificação e controle de processos;
- Previsão de séries temporais;
- Otimização de sistemas.

As redes *Multilayer Perceptron* (MLP) têm sido aplicadas com sucesso para resolver diversos problemas difíceis, através do seu treinamento de forma supervisionada com um algoritmo muito popular conhecido como algoritmo de retropropagação de erro (*backpropagation*). Este algoritmo é baseado na regra de aprendizagem por correção do erro. Como tal, pode ser visto como uma generalização de um algoritmo de filtragem adaptativa igualmente popular: o onipresente algoritmo do mínimo quadrado médio (LMS) para o caso especial de um único neurônio linear (HAYKIN, 2001).

Essas redes são caracterizadas pela presença de pelo menos uma camada intermediária (escondida) de neurônios, situada entre a camada de entrada e a respectiva camada neural de saída. Consequentemente, as redes *Multilayer Perceptron* (MLP) possuem no mínimo duas camadas de neurônios, os quais estarão distribuídos entre as camadas intermediárias e a camada de saída (SILVA; SPATTI; FLAUZINO, 2010)

Sua arquitetura é formada com uma camada de entrada que recebe os valores das variáveis, uma ou mais camadas ocultas, local onde são realizados os processos de ajuste e a pôr fim a camada de saída. A quantidade de neurônios da entrada corresponde a quantidade de variáveis que serão utilizados pelo algoritmo. Em relação a camada oculta, não há regra, portanto, o número ideal deve ser estimado para cada problema, guiada pelo conhecimento e pela experimentação. E por fim a camada de saída será definida pela quantidade de respostas ou classificações do problema estudado (HASTIE; TIBSHIRANI; FRIEDMAN, 2009). A arquitetura de um *Multilayer Perceptron* (MLP) é apresentada na figura 4.

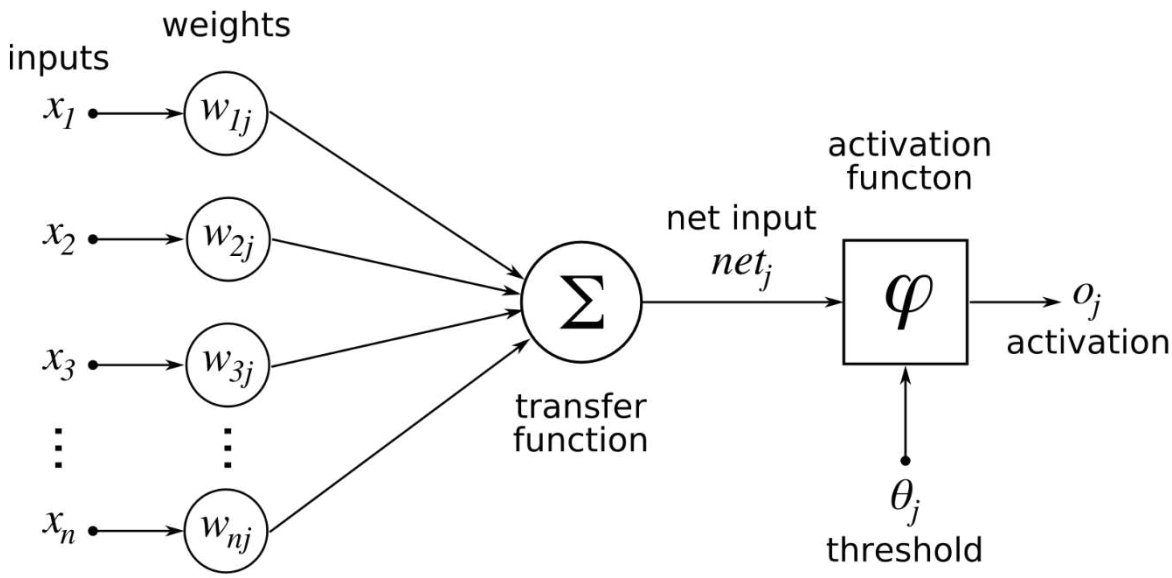
Figura 4 - Exemplo Arquitetura *Multilayer Perceptron* (MLP)



Fonte: Adaptado de HASTIE (2009)

A saída de uma rede neural é determinada pela função de ativação. Essas funções são anexadas a cada neurônio na rede, vide figura 5 e determinam se um neurônio deve ser ativado ou não, baseado nos dados de entrada. As funções podem calcular e aprender quase todas as funções e fornecer previsões ou classificações precisas. As funções de ativação podem ser binárias, lineares ou não lineares, sendo a última a mais utilizada, pela capacidade de aprender dados complexos, cenário mais próximo a realidade. Outra característica das funções não lineares é a possibilidade de retropropagação (*Backpropagation*), principal mecanismo de ajuste dos pesos de cada neurônio.

Figura 5 - Diagrama de uma rede neural artificial



Fonte: Adaptado de HASTIE (2009)

Entre as funções de ativação não lineares podemos listar a função Sigmoide, a Tangente Hiperbólica (TanH), ReLU (*Rectified Linear Unit*), *Leaky* ReLU, Softmax e Swish. A função Sigmoide normaliza a saída de cada neurônio entre 0 e 1, mas é impactada quando os dados de entradas possuem valores muito altos ou baixos, gerando baixa alteração nas previsões e consequentemente lentidão ou dificuldade no aprendizado. A função TanH facilita a modelagem de entradas com valores fortemente negativos, neutros e fortemente positivos, mas como a Sigmoide, também sofre com velocidade de treinamento.

A função ReLU é a mais utilizada nas redes neurais devido a sua capacidade de convergir rapidamente, mas pode ter dificuldades de aprendizado quando o gradiente da função se torna zero. Por conta dessa desvantagem da função ReLU, foram propostas variações da ReLU, exemplo a função *Leaky* ReLU que apresenta uma pequena inclinação positiva na área negativa, permitindo a retropropagação, mesmo para valores de entrada negativos.

Pode-se também destacar a função *Softmax*, com a capacidade de padronizar a saída em 0 ou 1, eficaz em tarefas de classificação, mas usado apenas na camada de saída para redes neurais que precisam classificar as entradas em várias categorias. E por fim pode-se listar a função *Swish*, recentemente proposta por pesquisadores do Google. Esta função é formulada como *x* vezes sigmoide *x* e vem

apresentado desempenho superior ao ReLU em experimentos de imagem com uso de redes profundas (*deep networks*) (RAMACHANDRAN; ZOPH; LE, 2017).

O super ajuste (*Overfitting*) dos dados de treinamento é uma questão importante para se acompanhar no aprendizado de uma rede *Multilayer Perceptron* (MLP). Quando ocorre observa-se redes que generalizam mal os novos dados, apesar do excelente desempenho dos dados de treinamento. Para minimizar o risco de *Overfitting*, saber quando parar o processo de descida de gradiente é fundamental, pois os métodos de validação cruzada são grandes alinhados para estimar o ponto de parada e evitar o *Overfitting* (MITCHELL, 1997).

2.3 MÉTRICAS DE DESEMPENHO DO MODELO

No processo de construção do modelo de aprendizagem de máquina (*Machine Learning)* é necessário avaliar o desempenho alinhada ao objetivo buscado. Existem várias métricas de avaliação da capacidade de erros e acertos do modelo, a escolha de quais métricas serão utilizadas leva em consideração a aplicação do modelo (objetivo do estudo), algoritmo utilizado e como o conjunto de dados estão estruturados (BADER-EL-DEN; TEITEI; ADDA, 2016).

Neste estudo o modelo terá o objetivo de prever com exatidão o risco de morte de crianças com cardiopatia congênita durante e pós cirurgia cardíaca, portanto trata-se de um problema de classificação binária. Em problemas de classificação binária tem se várias métricas de avaliação, as mais comuns em aprendizagem de máquina (*Machine Learning)* são Acurácia (*Accuracy*), Precisão (*Precision*), Sensibilidade (*Sensitivity* e *Recall*) e curva ROC (*Receiver Operating Characteristics*) AUC (*Area Under The Curve*), também escrita como AUROC (*Area Under the Receiver Operating Characteristics*).

Ao classificar um conjunto de dados balanceado, a métrica de Acurácia (*Accuracy*) resulta uma boa noção de desempenho do modelo, mas quando há um conjunto de dados desbalanceados, ou seja, quando um o número de instâncias pertencentes a uma classe excede o de qualquer outra classe, surgem alguns problemas na avaliação do desempenho. Por exemplo, em um problema de classificação binária, se a classe majoritária ultrapassa a classe minoritária em 9 para

1, e todas as instâncias são classificadas como a maioria, o classificador teria uma Acurácia de 90%, apesar de 0% da classe minoritária ser classificada corretamente, gerando falsa interpretação do desempenho do classificador proposto (AWAD; BADER-EL-DEN; MCNICHOLAS, 2017) (BADER-EL-DEN; TEITEI; ADDA, 2016).

Para evitar falsas interpretações sobre o desempenho do modelo, pode-se utilizar mais de uma métrica para avaliação de desempenho, tais como: Acurácia (*Accuracy*), Sensibilidade (*Sensitivity or Recall*), Especificidade (*Specificity*), Precisão (Precision), pontuação F1 (*F1 Score*), AUROC (*Area Under the Receiver Operating Characteristics*) e *Average precision* (AP), métricas comumente utilizadas na área de saúde.

A acurácia apresenta o percentual de acertos positivos do modelo, onde TP é a soma dos verdadeiros positivos (*True Positive*), TN é a soma dos verdadeiros negativos (*True Negative*), P é a soma dos positivos reais (*Positive*) e N é a soma do dos negativos reais (*Negative*). E é calculada com a função abaixo:

$$\text{Acurácia } (accuracy) = \frac{TP+TN}{P+N} \tag{7}$$

A sensibilidade é o número de exemplos classificados como verdadeiros positivos (TP) dividido pela soma de verdadeiros positivos (TP) e falsos negativos (FN). Pode ser interpretado como o percentual de cobertura do modelo, no nosso estudo, quanto maior esse índice, mais pacientes serão devidamente classificados como alto risco de morte e poderão receber atenção e cuidados especiais.

$$\text{Sensibilidade } (sensitivity\ or\ recall) = \frac{TP}{TP+FN} \tag{8}$$

A especificidade é o número de exemplos classificados como verdadeiros negativos (TN) dividido pela soma de verdadeiros negativos (TN) e falsos negativos (FN). Pode ser interpretado como o percentual de acertos de sobrevivência, quanto maior esse percentual, maior o número de pacientes classificados corretamente como sobreviventes.

$$\text{Especificidade } (specificity) = \frac{TN}{TN+FP} \tag{9}$$

A precisão é o número de exemplos classificados como verdadeiros positivos (TP) dividido pela soma entre verdadeiros positivos (TP) e falsos positivos (FP). Essa métrica retrata o percentual de acertos da previsão do algoritmo em relação ao conjunto de positivos (P), quanto maior esse índice menor os erros das predições realizadas.

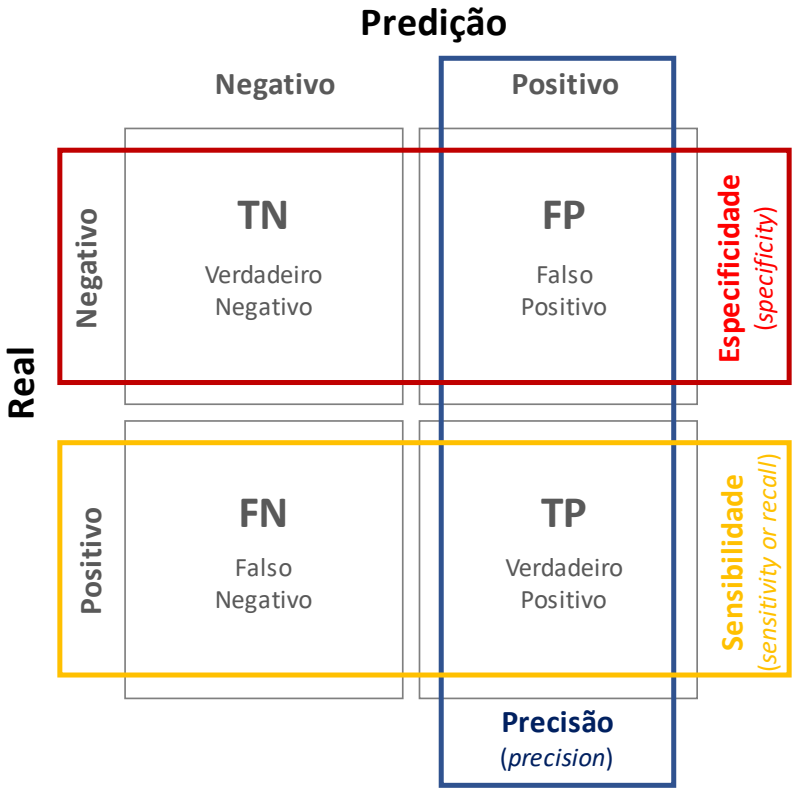
$$\text{Precisão } (precision) = \frac{TP}{TP+FP} \tag{10}$$

A pontuação F1 (*F1 Score*) é uma medida de equilíbrio entre precisão (*precision*) e sensibilidade (*recall*), calculada pela função abaixo:

$$F1 = 2 * \frac{Precision*Recall}{PRecision+Recall} \tag{11}$$

No diagrama da figura 6 podemos visualizar o racional das métricas de desempenho de forma sintética.

Figura 6 - Diagrama de Métricas de Desempenho



Fonte: Autor

Outra métrica comumente utilizada na avaliação de performance dos algoritmos de aprendizagem de máquina (*Machine Learning)* é a curva ROC (*Receiver Operating Characteristics)* AUC (*Area Under Curve*), também nomeada de AUROC (*Area Under the Receiver Operating Characteristics*). A curva é gerada plotando a taxa de verdadeiros positivos (TP) em relação à taxa de falsos positivos (FP) em várias configurações de limite. A área sob a curva também é calculada e representa a habilidade de acertos do modelo, índice nomeado de AUC (*Area Under Curve*), variando de 0,5 a 1,0, sendo que quanto mais próximo de 1,0 estiver esse índice, maior será a habilidade de acertos do modelo.

Segundo o levantamento feito por A. Awad, M. Bader-El-Den and J. McNicholas (2017) sobre modelos de previsão de mortalidade, o índice ROC AUC (ou AUROC) é amplamente utilizada na área da saúde para avaliação de desempenho dos modelos (AWAD; BADER-EL-DEN; MCNICHOLAS, 2017). Nos estudos mais recentes que utilizaram algoritmos de aprendizagem de máquina (*Machine Learning*) observamos o uso da curva ROC AUC (AUROC) como métrica de avaliação de performance (MEJIA et al., 2018)(LI et al., 2017)(XU et al., 2016)(AWAD et al., 2017)(CHURPEK et al., 2017)(ALGHAMDI et al., 2017).

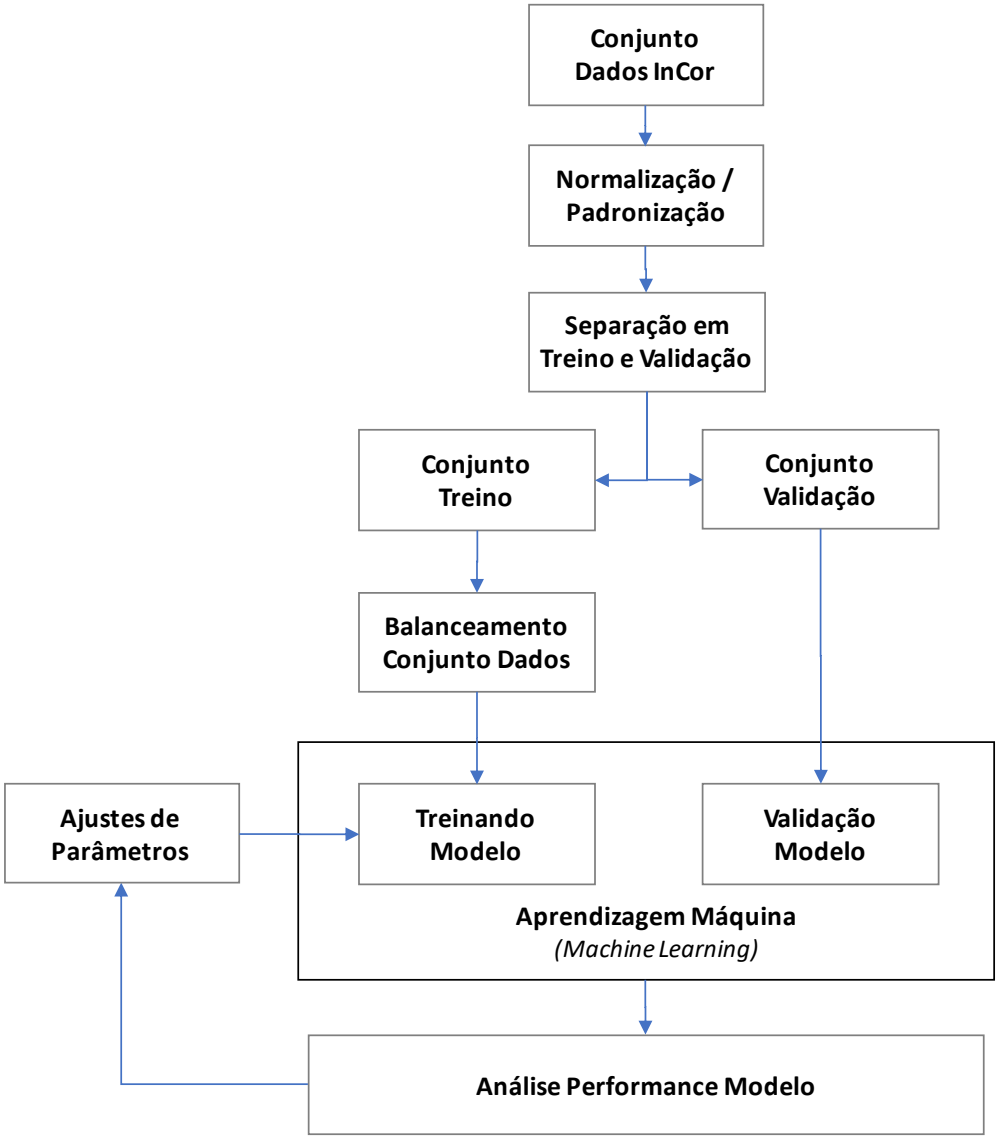
Apesar do índice ROC AUC (AUROC) ser amplamente utilizado na avaliação de desempenho de classificadores binários, o mesmo requer cuidado especial quando utilizado com conjuntos de dados desbalanceados (SAITO; REHMSMEIER, 2015)(DAVIS; GOADRICH, 2006). Os estudos de Saito (2015) e Davis (2006) alertam que deve-se tomar cuidado no uso da curva ROC AUC (AUROC) para conjuntos de dados desbalanceados e recomenda-se incluir a Curva de Precisão-Sensibilidade *(Precision-Recall Curves)* nas análises e tomada de decisão. A precisão (*Precision*) é uma medida da relevância dos resultados, enquanto a sensibilidade (*Recall*) é uma medida de quantos resultados verdadeiramente relevantes são retornados.

A Curva de Precisão-Sensibilidade *(Precision-Recall Curves)* é construída plotando em seus eixos o índice de precisão (P*recision*) e sensibilidade (R*ecall*). Um sistema com alta sensibilidade (*recall*), mas baixa precisão (*Precision*) retorna muitos resultados, mas com baixa assertividade nas previsões. Um sistema com alta precisão (*precision*), mas com baixa sensibilidade (*recall*), é exatamente o oposto, retornando muito poucos resultados, mas a maioria das previsões são corretas. Um sistema ideal com alta precisão (*precision*) e alta sensibilidade (*recall*), retornará muitos resultados com previsões corretas.

3 MATERIAIS E MÉTODOS

Os experimentos foram executados num notebook Intel® Core™ i7-7700HQ 2.80GHz, 16.0 GB de memória RAM, sob a plataforma Windows 10. E para a manipulação, análises e treinamento dos algoritmos utilizou-se o software Python 3.7.1 e as bibliotecas *Numpy, Pandas, Matplotlib, Seaborn, Scikit-Learn, Imblearn* e *PyTorch*. E no desenvolvimento do modelo de previsão de mortalidade foram seguidas as etapas listadas na figura 7.

Figura 7 - Etapas de construção do modelo de previsão



Fonte: Autor

No desenvolvimento do modelo de previsão foram executadas as seguintes etapas: preparação do conjunto de dados InCor, normalização ou padronização das variáveis, divisão dos dados em conjunto de treinamento e validação, balanceamento do conjunto de treinamento, treinamento e ajustes do algoritmo, e pôr fim, medição da performance de previsão do modelo. Cada uma dessas etapas será abordada nos próximos itens desse capítulo.

3.1 CONJUNTO DE DADOS INCOR

O conjunto de dados utilizado neste estudo foi fornecido pelo Instituto do Coração do Hospital das Clínicas (InCor), Faculdade de Medicina, Universidade de São Paulo, Brasil.

As informações utilizadas no estudo foram extraídas do banco de dados ASSIST do InCor e seus campos catalogados por um grupo de especialistas da área. O conjunto de dados contém o histórico de 2.240 cirurgias cardíaca realizadas em pacientes com cardiopatia no período de 2014 a 2018, e foi organizado em 87 campos (variáveis), distribuídos em quantitativas continuas e qualitativas categóricas, campos detalhados no apêndice A nesse estudo. A variável objetivo foi “Desfecho Final” tabulada em 0 ou 1, sendo que 0 para “Alta Hospitalar” e 1 para “Morte Hospitalar”.

Nesse estudo o uso de formulários de consentimento foi dispensado. O banco de dados não possui informações de identificação do paciente.

3.2 NORMALIZAÇÃO OU PADRONIZAÇÃO DAS VARIÁVEIS

Antes de iniciar o treinamento do algoritmo é preciso avaliar a necessidade da normalização ou padronização das variáveis, porque muitos algoritmos de aprendizado de máquina (*Machine Learning*) têm melhor desempenho ou convergem mais rapidamente quando os recursos estão em uma escala relativamente semelhante ou perto da distribuição normal, exemplo a regressão linear, regressão logística*, K-Nearest Neighbors algorithm* (KNN), redes neurais Artificiais (RNA), máquinas de vetores de suporte com funções do núcleo polarizado radialmente

(SVM), análise de componentes principais (PCA) e análise discriminante linear (LDA) (MITCHELL, 1997)(HAYKIN, 2001)(SILVA; SPATTI; FLAUZINO, 2010)(HASTIE; TIBSHIRANI; FRIEDMAN, 2009).

Na padronização os valores das variáveis resultam em uma média igual a 0 e um desvio padrão igual a 1 e seus valores nos intervalos de -1 a +1. E na normalização os valores das variáveis são calculados para um intervalo de 0 e 1, preservando a forma da distribuição original.

Na padronização aplicamos a função abaixo em cada variável:

$$Z = \frac{x-\mu}{\sigma} \tag{12}$$

Onde x é o valor original da variável, μ a média e σ o desvio padrão.

Para normalização aplica-se a função:

$$x_{normalizado} = \frac{x - x_{min}}{x_{max} - x_{min}} \tag{13}$$

Onde $x_{normalizado}$ é o valor original da variável, x é o valor original da variável, x_{max} valor máximo e x_{min} o valor mínimo da variável.

3.3 SEPARAÇÃO DO CONJUNTO DE DADOS EM TREINO E VALIDAÇÃO

Nesta etapa é separado o conjunto de dados para treinamento e teste (validação) do modelo de previsão. O conjunto de treinamento é usado para treinar o modelo e o conjunto de teste (validação) é usado para avaliar o desempenho do modelo. No entanto, essa abordagem pode levar a problemas de variância, ou seja, um problema de variância se refere ao cenário em que a precisão obtida em um teste é muito diferente da precisão obtida em outro conjunto de teste usando o mesmo algoritmo. Para evitar esse efeito, são utilizadas algumas técnicas que minimizam a

variância, garantindo melhor performance do modelo (BROWNE, 2001)(KOHAVI, 1995).

As técnicas mais utilizadas em aprendizado de máquina (*Machine Learning*) são *Hold-out* e *K-Fold Cross Validation*. A decisão do método depende do objetivo do modelo e do tamanho do conjunto de dados. O método *Hold-out* divide o conjunto de dados em duas partes, treinamento e teste, enquanto o método *K-Fold Cross Validation* divide o conjunto de dados em K partes de tamanho igual, também chamada de partições. O processo de treinamento é aplicado em todas as partições, com exceção de uma partição que é utilizada como conjunto de teste no processo de validação. Finalmente, a medida final de desempenho será a média de todos os testes de desempenho de todas as partições. A vantagem desse método é que todo o conjunto de dados é treinado e testado, reduzindo a variação do estimador escolhido. Isso garante uma previsão mais precisa e menos tendenciosidade do estimador de taxa positiva (KOHAVI, 1995)(BLAGUS; LUSA, 2015).

Neste estudo utilizou-se o método de *K-Fold Cross Validation* de 10 vezes (K=10), método utilizado em vários estudos relacionados à saúde e vêm demonstrado boa performance com baixa variabilidade entre a amostra de treino e teste (RUIZ-FERNÁNDEZ et al., 2016)(AWAD et al., 2017)(CHURPEK et al., 2017)(DESAUTELS et al., 2017)(COHEN et al., 2016)(PIRRACCHIO et al., 2015).

Figura 8 - Exemplo de *K-Fold Cross Validation* (K=5)

	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5
Splint 1	Treinamento	Treinamento	Treinamento	Treinamento	Validação
Splint 2	Treinamento	Treinamento	Treinamento	Validação	Treinamento
Splint 3	Treinamento	Treinamento	Validação	Treinamento	Treinamento
Splint 4	Treinamento	Validação	Treinamento	Treinamento	Treinamento
Splint 5	Validação	Treinamento	Treinamento	Treinamento	Treinamento

Fonte: Autor

Na figura 8 é apresentado um exemplo de *K-fold Cross Validation* de 5 vezes (K=5). O conjunto de dados é dividido aleatoriamente em 5 partes (*folds*) de mesmo tamanho, onde uma das partes (*fold*) é separada para validação e demais para treinamento do algoritmo. Esse processo é repetido 5 vezes (*splint*), sempre separando uma parte para validação e 4 partes para treinamento.

3.4 BALANCEAMENTO DO CONJUNTO DE TREINAMENTO

É comum observar nos conjuntos de dados grandes diferenças no percentual de representatividade nas classes estudadas, como por exemplo o conjunto de dados do estudo, onde observamos que 10,8% pacientes morrem após a cirurgia contra 89,2% que não. Quando as categorias de classificação não são representadas de forma igual, diz-se que o conjunto de dados é desbalanceado (CHAWLA et al., 2002) (PROVOST; FAWCETT, 2001).

Os algoritmos convencionais costumam ser tendenciosos para a classe majoritária porque suas funções de perda tentam otimizar quantidades como taxa de erro, não levando em consideração a distribuição de dados. Na pior das hipóteses, exemplos minoritários são tratados como discrepantes da classe majoritária e ignorados, fazendo que o modelo seja treinado apenas para identificar a classe majoritária, que nesse estudo é deixar de classificar um risco de morte de paciente.

O conjunto de dados do InCor é desbalanceado, ou seja, observa-se uma razão de 1:9 entre mortalidade e sobrevivência pós cirurgia (CHAWLA et al., 2002) (PROVOST; FAWCETT, 2001). Em alguns algoritmos, se esse efeito não for tratado poderá haver uma falsa interpretação de performance do modelo, algo que não é desejável se o objetivo do estudo é identificar e compreender o risco de morte, a classe minoritária do conjunto de dados estudado.

Para reduzir os impactos no desbalanceamento utilizam-se de métodos para sub-amostrar (*under-sampling*) e/ou sobre-amostrar (*over-sampling*) o conjunto de dados. As técnicas de sub-amostrar (*under-sampling*) consistem em reduzir a amostra da categoria de maior representatividade para aumentar a sensibilidade de um classificador para a classe minoritária, enquanto a técnica de sobre-amostrar (*over-sampling*) simplesmente aumenta a amostra da categoria minoritária utilizando

técnicas estatísticas para replicação das amostras minoritária, duplicando-as ou gerando novas amostras a partir das amostras reais (CHAWLA et al., 2002).

Existem vários métodos disponíveis para ajustar uma amostra desbalanceada, como exemplo a técnica chamada de SMOTE (*Synthetic Minority Over-sampling Technique*), que utiliza-se da combinação de técnicas de sub-amostrar (*under-sampling*) e sobre-amostrar (*over-sampling*) para alcançar um melhor desempenho do classificador do que apenas a sub-amostragem da classe majoritária. O SMOTE funciona criando amostras sintéticas da classe secundária, em vez de criar cópias. Esse método escolhe aleatoriamente um dos k vizinhos mais próximos e usa-o para criar uma nova observação similar, mas aleatoriamente ajustada (CHAWLA et al., 2002). A geração de novos exemplos fornece informações adicionais aos classificadores, melhorando a taxa de verdadeiros positivos no conjunto de testes. Embora isso também possa aumentar os falsos positivos, a expectativa é que o aumento nos verdadeiros positivos supere o aumento relativo de falsos positivos.

Observa-se a aplicação de métodos de sub-amostrar (*under-sampling*) e/ou sobre-amostrar (*over-sampling*) em alguns estudos na área de medicina (AWAD et al., 2017) (BERETTA et al., 2010) (ALGHAMDI et al., 2017), exemplo o estudo de Aya Awad e Mohamed Bader-El-Dena (2017) sobre previsão de mortalidade em UTI, o estudo de Manal Alghamdi e Mouaz Al-Mallah (2017) sobre previsão do diagnóstico de diabetes e o estudo de L. Beretta e A. Santaniello (2010) sobre previsão de mortalidade de pacientes com esclerose. Neste estudo serão testados os métodos de sub-amostrar (*under-sampling*) e sobre-amostrar (*over-sampling*) e será escolhido o método que experimentalmente obtiver melhor performance de previsão para o algoritmo estudado, sempre buscando a melhor performance de previsão de risco de morte.

3.5 TREINAMENTO E AJUSTES DO ALGORITMO

Uma vez preparado o conjunto de dados de treinamento, normalizado/padronizado e balanceado, pode-se iniciar a etapa de treino dos algoritmos de aprendizagem de máquina (*Machine Learning*) selecionados neste

estudo. Neste estudo foram treinados os seguintes algoritmos de aprendizagem de máquina (*Machine Learning*): *Random Forest* (RF), *Bagged Decision Trees* (BDT) e *Stochastic Gradient Boosting* (SGB), *Extra Trees* (ET), *AdaBoost Classification* (ABC), e *Multilayer Perceptron* (MLP). Dada a especificidade de cada algoritmo, os detalhes do treinamento e os ajustes utilizados em cada algoritmo serão apresentados no capítulo 4.

3.6 ANÁLISE PERFORMANCE MODELO

Por fim, em pose do conjunto de dados de teste (validação), separado na etapa de preparação, é avaliado a performance do modelo. Nesta etapa cada algoritmo de aprendizagem de máquinas (*Machine Learning*) tem sua performance medida nas seguintes métricas: acurácia (*accuracy*), sensibilidade (*recall* or *sensitivity*), especificidade (*specificity*), precisão (*precision*), ROC AUC (AUROC) e *Average precision* (AP). Os resultados de cada algoritmo treinado serão apresentados no capítulo 4.

4 EXPERIMENTOS E RESULTADOS

O experimento foi iniciado seguindo as etapas descritas na figura 6. Com o conjunto de dados normalizado separou-se as amostras de treino e teste (validação), utilizando o método *K-Fold Cross Validation* de 10 vezes (K=10), onde 90% da amostra foi separada para treinar os algoritmos de aprendizagem de máquina (*Machine Learning)* e o restante separado para a etapa de validação. Na sequência foi avaliado a necessidade de balanceamento do conjunto de dados, uma vez que temos um conjunto desbalanceado, razão de 1:9 entre mortalidade e sobrevivência pós cirurgia e buscamos maior previsibilidade de risco de morte. Quando necessário balanceamento, foi testado diferentes métodos de sub-amostrar (*under-sampling*) e/ou sobre-amostrar (*over-sampling*) para ajustar distribuição das categorias da amostra de treino.

Com as amostras de treino e teste (validação) separadas e ajustadas, iniciou-se a etapa de treino do algoritmo. No processo de treino dos algoritmos, foi testado várias configurações de parâmetros, sempre buscando minimizar os erros de generalização, sejam por *Overfitting* ou *Underfitting* (HAWKINS, 2004). *Overfiting* ocorre quando o algoritmo se adapta muito bem aos dados que está sendo treinado, mas perde a capacidade de replica-lo para um novo conjunto de dados. E *Underfitting* é quando o algoritmo não se adapta bem aos dados que foi treinado e possui baixa generalização para outros dados.

Se atentando as questões de generalização do algoritmo, foi maximizado os índices de ROC AUC (AUROC) e *Average precision* (AP), no objetivo de obter a melhor assertividade de previsão e minimizar o erro de não sinalizarmos um paciente com alto risco de mortalidade.

Como cada algoritmo possui suas especificidades, nos próximos tópicos serão detalhadas as etapas de preparação, treinamento e análise de performance de cada algoritmo selecionado neste estudo.

4.1 RANDOM FOREST (RF)

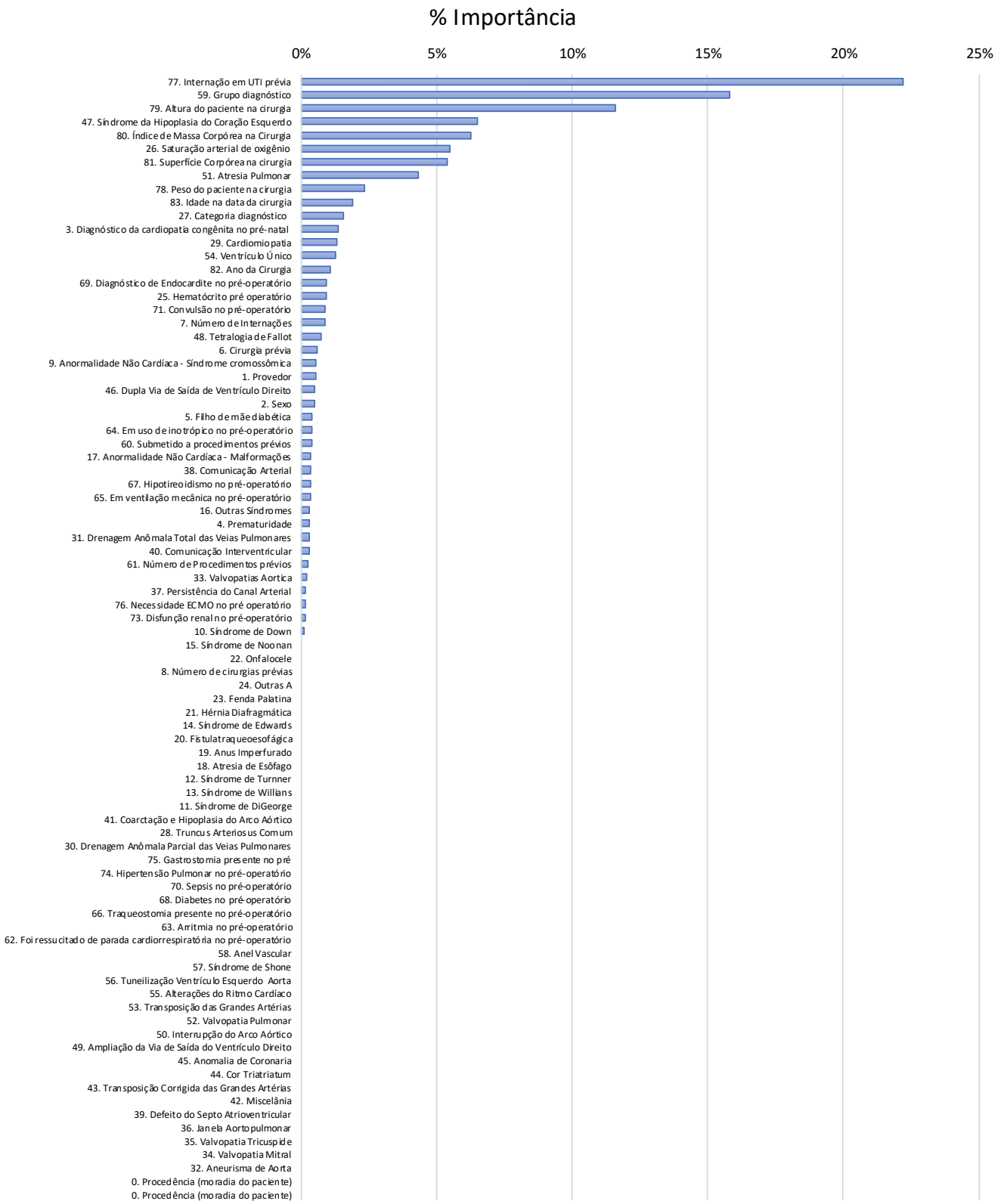
Para melhor entendimento do treinamento do algoritmo *Random Forest* (RF) a seção foi dividida nas seguintes etapas: seleção das variáveis relevantes para a previsão, ajustes dos parâmetros do modelo e análise dos resultados obtidos.

O algoritmo *Random Forest* (RF) possibilita classificar a importância das variáveis (BREIMAN, 2001), auxiliando na seleção das variáveis que contribuem na previsão. Isto é importante, porque ter muitas variáveis irrelevantes em seu conjunto de dados poderá diminuir a precisão do modelo, além de aumentar o tempo de treinamento e até mesmo gerar *Overfitting* ou *Underfitting.*

Baseado nessa premissa, o algoritmo *Random Forest* (RF) foi ajustado para gerar em sua saída uma classificação das variáveis do conjunto de treinamento, informação utilizada para eliminar as variáveis irrelevantes para a previsão. No gráfico 9 é apresentado o resultado dessa classificação.

No gráfico 9 temos no eixo Y o índice de importância de cada variável do conjunto de dados do InCor e no eixo X são apresentadas as variáveis ordenadas pelo índice de importância, de forma crescente da esquerda para a direita. Nesse gráfico observamos um número expressivo de variáveis com baixa importância para o *Random Forest* (RF), o próximo passo é selecionar as variáveis que utilizaremos no modelo. Para essa etapa, utilizamos a técnica chamada de RFE (*Recursive Feature Elimination*) (GUYON et al., 2002), que nos ajudará na seleção das variáveis importantes na previsão. O método RFE funciona removendo recursivamente as variáveis menos importantes e construindo um modelo com as variáveis que permanecem. Este procedimento é repetido recursivamente no conjunto podado, considerando conjuntos cada vez menores de variáveis, até que a precisão do modelo atinja o seu melhor índice. Esse método tem demonstrado performance superior em relação a outras técnicas em várias aplicações como observamos em estudos mais recentes (DARST; MALECKI; ENGELMAN, 2018)(ABDOH; ABO RIZKA; MAGHRABY, 2018).

Figura 9 - Importância das Variáveis do *Random Forest* (RF)



Fonte: Autor

A classificação da importância das variáveis utilizadas no *Random Forest* (RF) também nos traz uma informação importante para o estudo, conseguimos identificar quais variáveis ou diagnósticos influenciam na previsão de morte. Na tabela 4 apresentamos as variáveis selecionadas e o seu índice de importância.

Tabela 4 - Variáveis selecionadas para o *Random Forest* (RF)

Variáveis Selecionadas			
Feature	**Feature Selecionada**	**% importância**	**Descrição**
77	PRE_UTI_Previa_2sim_1nao_adm1_81	22,2%	Internação em UTI prévia
59	PRE_grupo_diag_61	15,8%	Grupo diagnóstico
79	PRE_estat_cm_adm1_83	11,6%	Altura do paciente na cirurgia
47	PRE_SHCE_49	6,5%	Síndrome da Hipoplasia do Coração Esquerdo
80	PRE_IMC_adm1_84	6,3%	Índice de Massa Corpórea na Cirurgia
26	PRE_sato2_preop_adm1_28	5,5%	Saturação arterial de oxigênio
81	PRE_SC_adm1_85	5,4%	Superfície Corpórea na cirurgia
51	PRE_Atresia_pulmonar_53	4,3%	Atresia Pulmonar
78	PRE_peso_kg_adm1_82	2,4%	Peso do paciente na cirurgia
27	PRE_Diagnostico_categoria1a12_29	1,6%	Categoria diagnóstico
29	PRE_CMP_31	1,3%	Cardiomiopatia
54	PRE_Ventriculo_unico_56	1,3%	Ventrículo Único
82	PRE_ANO_CC_ADM1_86	1,1%	Ano da Cirurgia
25	PRE_ht_preop_adm1_27	0,9%	Hematócrito pré operatório
7	PRE_intpre_9	0,9%	Número de Internações
48	PRE_T4F_50	0,7%	Tetralogia de Fallot
6	PRE_cirpre_8	0,6%	Cirurgia prévia
9	PRE_sindcrom_11	0,5%	Anormalidade Não Cardíaca - Síndrome cromossômica
5	PRE_maediab_7	0,4%	Filho de mãe diabética
60	PRE_proc_previos_2Sim_1Nao_63	0,4%	Submetido a procedimentos prévios
38	PRE_CIA_40	0,3%	Comunicação Arterial
65	PRE_Vent_mec_Pre_Adm1_68	0,3%	Em ventilação mecânica no pré-operatório
4	PRE_premat_6	0,3%	Prematuridade
40	PRE_CIV_42	0,3%	Comunicação Interventricular
61	PRE_proc_previos_01_64	0,2%	Número de Procedimentos prévios
33	PRE_Doenca_ValvaAO_35	0,2%	Valvopatias Aortica
73	PRE_Disf_renal_pre_Adm1_76	0,1%	Disfunção renal no pré-operatório
10	PRE_down_12	0,1%	Síndrome de Down
8	PRE_ncirpre_10	0,0%	Número de cirurgias prévias
11	PRE_digeorge_13	0,0%	Síndrome de DiGeorge
12	PRE_turnner_14	0,0%	Síndrome de Turnner
13	PRE_willians_15	0,0%	Síndrome de Willians
28	PRE_Truncus_30	0,0%	Truncus Arteriosus Comum
32	PRE_Aneurisma_Ao_34	0,0%	Aneurisma de Aorta
35	PRE_Doenca_ValvaTri_37	0,0%	Valvopatia Tricuspide
50	PRE_Interrupcao_ArcoAO_52	0,0%	Interrupção do Arco Aórtico

52	PRE_Doenca_ValvaPulm_54	0,0%	Valvopatia Pulmonar
53	PRE_TGA_55	0,0%	Transposição das Grandes Artérias
55	PRE_Eletrofisiologia_57	0,0%	Alterações do Ritmo Cardíaco
56	PRE_Tunel_VE_AO_58	0,0%	Tuneilização Ventrículo Esquerdo Aorta
57	PRE_Shone_59	0,0%	Síndrome de Shone
62	PRE_Ressuc_PCR_Pre_Adm1_65	0,0%	Foi ressucitado de parada cardiorrespiratória no pré-operatório

Fonte: Autor

Como observado na Tabela 4, as variáveis “Internação em UTI prévia” (22,2%), “Grupo diagnóstico” (15,8%) e “Altura do paciente na cirurgia” (11,6%) são as mais importantes no *Random Forest* (RF) e precisam ser melhor estudadas, uma vez que podem trazer informações que auxiliem em tratamentos prévios que reduzam a probabilidade de morte numa possível intervenção cirúrgica.

Na próxima etapa são definidos os parâmetros do algoritmo *Random Forest* (RF). Como em outros algoritmos, não temos uma receita pronta, cada parâmetro deve ser experimentado e avaliado para o objetivo desejado. Existem algumas técnicas que podem auxiliar na escolha do melhor parâmetro para o algoritmo utilizado, como as técnicas *GridSearchCV* e *RandomizedSearchCV* (BERGSTRA; BENGIO, 2012). Na técnica *GridSearchCV* fornecemos uma grade de valores para cada parâmetro que serão exaustivamente combinados até encontrar a combinação com o melhor índice de performance desejado. Enquanto que na técnica *RandomizedSearchCV* é fornecido um range para cada parâmetro e de forma aleatória são selecionados os índices que serão testados. O processo de seleção aleatória acontece até atingir uma quantidade de combinações pré-estabelecida, após esse número de combinações é apresentado o ranking de combinações com melhor índice de performance desejado. Na literatura encontramos exemplos de uso das duas técnicas, no estudo de Bengio (2012) observamos exemplos positivos para ambas as técnicas, mas todos os exemplos são de conjunto de dados com grande quantidade de observações e variáveis, cenários que o tempo de processamento computacional é importante. Neste estudo temos um conjunto de dados pequeno, tanto em observações como em variáveis, razão a qual é combinado os dois métodos com o objetivo de explorar o maior número de combinações. A primeira etapa é executar o *RandomizedSearchCV*, técnica que permitiu definir os parâmetros e seus índices com melhor performance. Em pose do resultado do *RandomizedSearchCV*, definisse a grade de parâmetros para *GridSearchCV.*

Os parâmetros do algoritmo *Random Forest* (RF) testados em ambas as técnicas foram:

- n_estimators = número de árvores geradas;
- max_features = número máximo de variáveis considerados para dividir um nó;
- max_depth = número máximo de níveis em cada Árvore de Decisão;
- min_samples_split = número mínimo de pontos de dados colocados em um nó antes do nó ser dividido;
- bootstrap = se utilizaremos o método *bootstrap* (com ou sem substituição) para amostrar;
- criterion = função de medição da qualidade de uma divisão. Os critérios suportados são "*gini*" para a impureza de Gini e "*entropy*" para o ganho de informações.

E os parâmetros e índices selecionados para o algoritmo *Random Forest* (RF) são apresentados na tabela 5.

Tabela 5 - Parâmetros utilizados no *Random Forest* (RF)

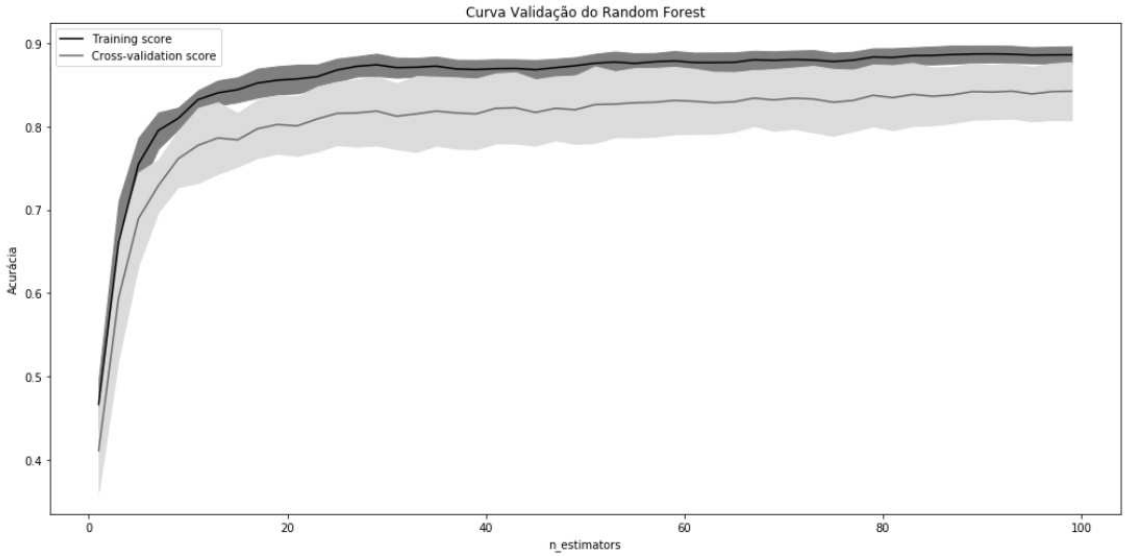
Parâmetro Algoritmo
Random Forest (RF) *Algoritmo Utilizado:* from imblearn.ensemble import BalancedRandomForestClassifier *Parâmetros utilizados*: bootstrap='False' random_state=1986 criterion='gini' max_depth=3 max_features=9 min_samples_split=3 n_estimators=25 n_jobs=-1

Fonte: Autor

Outra importante função do algoritmo *Random Forest* (RF) é a reamostragem da classe minoritária do conjunto de dado de treinamento, função que possibilita ajustar o desbalanceamento das classes enquanto o algoritmo é treinado, eliminando a necessidade de aplicarmos métodos de sub-amostrar (*under-sampling*) e/ou sobre-amostrar (*over-sampling*) no conjunto de dados antes de iniciar o treinamento. Essa reamostragem é aleatória e pode ser com ou sem reposição (*boostrap*).

Definido os parâmetros do *Random Forest* (RF) inicia-se o treinamento do algoritmo com o conjunto de dados de treino e na sequência valida-se o modelo utilizando o conjunto de dados de teste (validação), sempre buscando minimizar os erros de generalização, sejam por *Overfitting* ou *Underfitting* (HAWKINS, 2004). Na validação e análise de performance do modelo utiliza-se algumas análises gráficas apresentadas na sequência.

Figura 10 - Curva de Validação do *Random Forest* (RF)



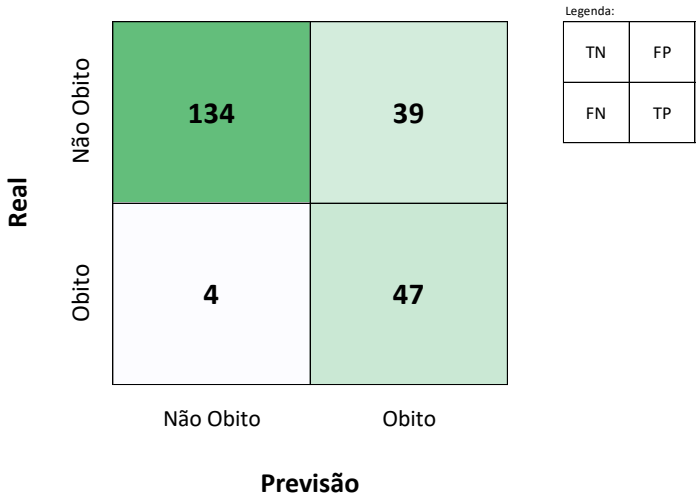
Fonte: Autor

Na figura 10 analisa-se a performance de acurácia (*Accuracy*) do conjunto de dados de treino (*Training Score*) e validação (*Cross-Validation Score*) em relação ao número de árvores geradas (*n_estimator*). Neste gráfico validamos o número de árvores que devemos gerar para obtemos o maior índice de acurácia, aproximadamente 25 árvores, acima dessa quantidade não observamos ganhos significativos no índice de acurácia. Adotando-se valores superiores a 25 observa-se o aumento do tempo de processamento e aumento do risco de *Overfitting* do modelo (BREIMAN, 2001). Outra informação importante apresentada no gráfico é o

comportamento uniforme das curvas, validando que o modelo treinado pode ser utilizado num novo conjunto de dados.

A próxima análise apresentada, figura 11, é a Matriz de Confusão (*confusion matrix*). Em uma Matriz de Confusão, cada coluna representa o número de classificações para cada classe predita, enquanto cada linha representa a quantidade de eventos reais de cada classificação.

Figura 11 - Confusion Matrix *Random Forest* (RF)



Fonte: Autor

Os dados da Matriz de Confusão são do conjunto de teste (validação), no qual pode-se verificar a quantidade de observações com acertos e erros de previsão, exemplo 4 observações de pacientes que morreram e não foi previsto no modelo, como as 47 demais observações que o modelo acertou. Essa matriz traz as informações que serão utilizadas para gerar os índices de acurácia (*accuracy*), sensibilidade (*sensitivity or Recall*), especificidade (*specificity*) e precisão (precision) do modelo, apresentados na tabela 6.

Tabela 6 - Métricas de Performance do *Random Forest* (RF)

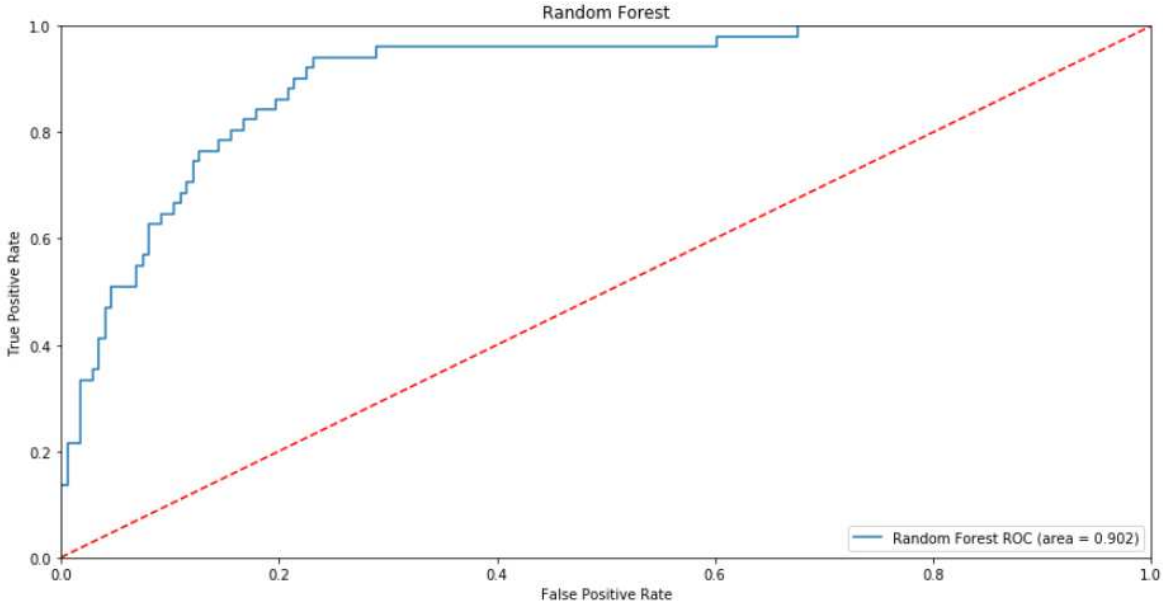
```
Random Forest
+-----------------+----------+
| Metric          |      Vlr |
+-----------------+----------+
| Accuracy        | 0.808036 |
| Recall          | 0.921569 |
| Precision       | 0.546512 |
| F1              | 0.686131 |
| AUC Probability | 0.901734 |
+-----------------+----------+
              precision    recall  f1-score   support

           0       0.97      0.77      0.86       173
           1       0.55      0.92      0.69        51

    accuracy                           0.81       224
   macro avg       0.76      0.85      0.77       224
weighted avg       0.87      0.81      0.82       224
```

Fonte: Autor

Na tabela 6 observa-se os índices de acurácia (*accuracy*), sensibilidade (*Recall)*, especificidade e precisão (*precision*) do modelo. Em resumo o modelo obteve uma acurácia de 80,8%, sensibilidade (*recall*) de 92,1% e precisão de 54,6%.

Outra métrica frequentemente utilizada na tomada de decisões médicas (HAJIAN-TILAKI, 2013) e tem sido cada vez mais adotada na comunidade de pesquisa de aprendizagem de máquina é a curva ROC AUC (AUROC) (FAWCETT, 2004), apresentada na figura 12.

Figura 12 - Curva ROC AUC do *Random Forest* (RF)

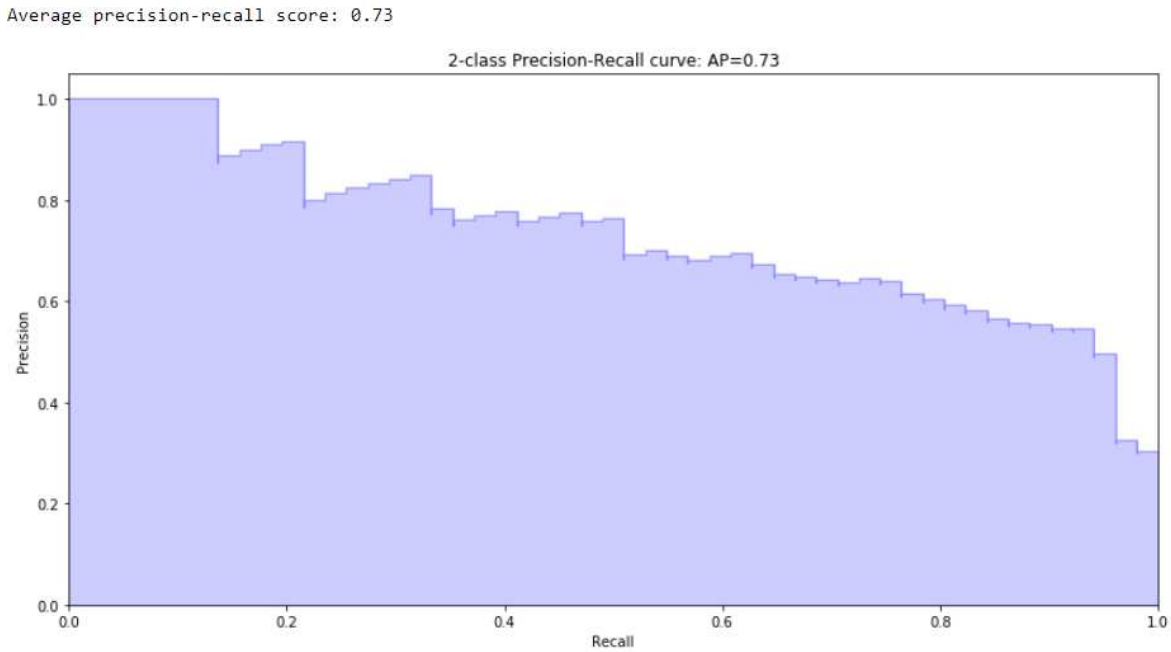


Fonte: Autor

No gráfico 12 observa-se a taxa de falsos positivos no eixo horizontal e a taxa de verdadeiros positivos no eixo vertical e a curva plotada é a curva ROC. A área sob a curva é chamada de AUC (*Area Under Curve*) e indica a habilidade de acertos do modelo, quanto mais esse índice se aproximar de 1, maior será a habilidade do modelo acertar suas previsões. O *Random Forest* (RF) atingiu um índice de 90,2% de ROC AUC, informação que utilizaremos para comparar com a performance dos demais modelos e também apoiar na decisão dos valores de corte para o objetivo desejado.

Apesar do índice ROC AUC (AUROC) ser amplamente utilizado na avaliação de desempenho de classificadores binários, o mesmo requer cuidado especial quando usado com conjuntos de dados desbalanceados (SAITO; REHMSMEIER, 2015)(DAVIS; GOADRICH, 2006). Os estudos de Saito (2015) e Davis (2006) alertam para os cuidados no uso da curva ROC AUC (AUROC) em conjuntos de dados desbalanceados, cenário muito comum na medicina, e recomenda-se incluir a Curva de Precisão-Sensibilidade *(Precision-Recall Curves)* nas análises e tomada de decisão. Seguindo a sugestão desses pesquisadores, é gerada a curva sugerida, apresentada na figura 13.

Figura 13 - Precision-Recall Curve do *Random Forest* (RF)

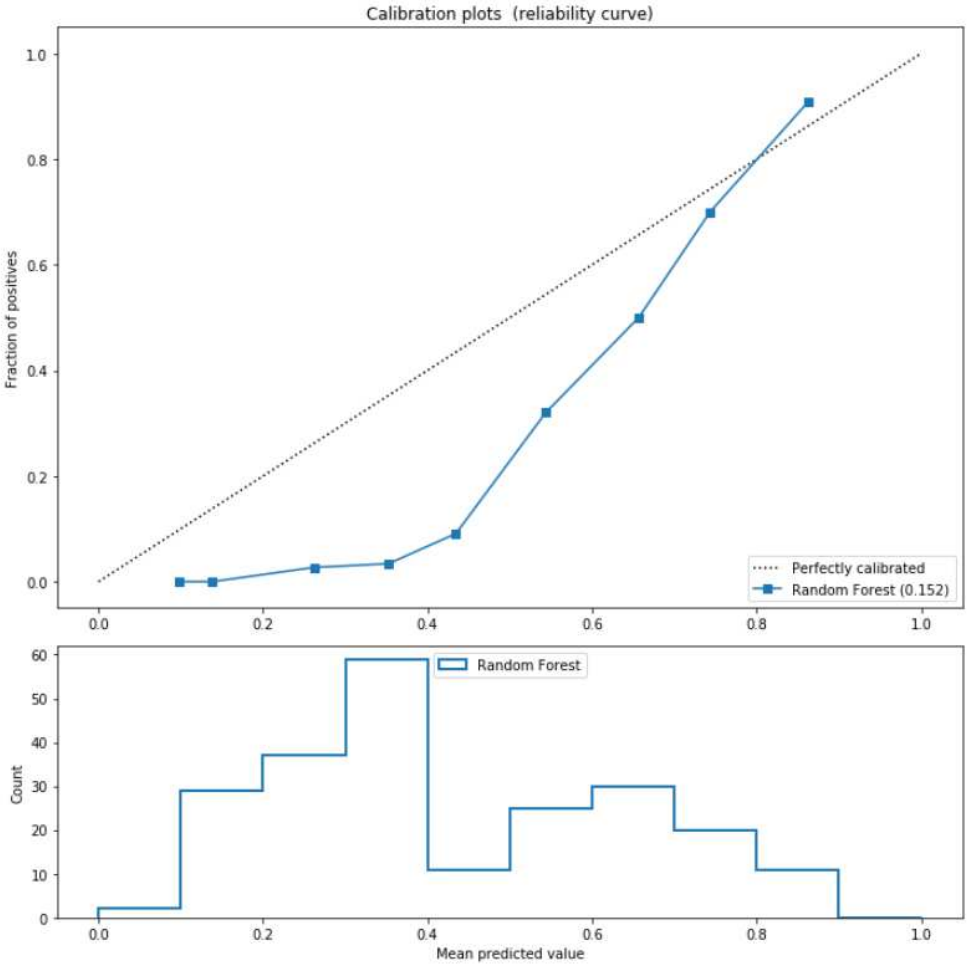


Fonte: Autor

Na curva Precisão-Sensibilidade (*Precision-Recall Curve)* pode-se verificar a correlação dos dois índices. Observa-se que quanto maior a sensibilidade (*recall*) menor será a precisão (*precision*) do modelo gerado, informação importante na definição do ponto de corte do modelo, uma vez que pode-se querer alta precisão em detrimento da sensibilidade ou o contrário, dependendo do objetivo do estudo. Também é calculado o índice *Average precision* (AP), que no *Random Forest* (RF) foi de 0,73.

Outra importante análise que gerado no estudo foi o gráfico de calibração das probabilidades geradas pelo *Random Forest* (RF), apresentado na figura 14. O algoritmo *Random Forest* (RF) por padrão classifica todas as observações com probabilidade acima de 0,5 como morte e abaixo desse índice como sobrevivência. Esse ponto de corte de classificação pode ser alterado, seja reduzindo-o ou o aumentando, dependendo do objetivo buscado.

Figura 14 - Gráfico de Calibragem do *Random Forest* (RF)



Fonte: Autor

O gráfico 14 traz duas informações, o percentual de acertos pela probabilidade gerada no *Random Forest* (RF) e a distribuição das observações por faixa de probabilidade. Essas informações, juntamente com a Curva Precisão-Sensibilidade (*Precision-Recall Curve)* auxiliam na tomada de decisão sobre o ponto de corte ótimo para o objetivo desejado.

Na tabela 7 são apresentadas todas as métricas de performance do algoritmo *Random Forest* (RF).

Tabela 7 - Performance do *Random Forest* (RF)

Algoritmos	**Acuracia** (Accuracy)	**Precisão** (Precision)	**Sensibilidade** (Recall)	**Especificidade** (Specificity)	**ROC AUC**	**AP**
Random Forest	80,8%	54,7%	92,2%	77,5%	90,2%	0,73

Fonte: Autor

4.2 BAGGED DECISION TREES (BDT)

O próximo algoritmo treinado é o *Bagged Decision Trees* (BDT), diferente do *Random Forest* (RF), esse algoritmo não possibilita gerar uma classificação de importância das variáveis e a técnica RFE de seleção de variáveis não poderá ser utilizada, portanto utilizaremos todas as variáveis no treinamento do algoritmo.

Outro ponto importante, como no *Random Forest* (RF) esse algoritmo permite reamostrar a classe minoritária do conjunto de dados, ajustando o desbalanceamento das classes e consequentemente eliminando a necessidade de aplicarmos métodos de sub-amostrar (*under-sampling*) e/ou sobre-amostrar (*over-sampling*) no conjunto de dados antes de iniciarmos o treinamento. Essa reamostragem é aleatória e pode ser com ou sem reposição (*booststrap*).

No ajuste dos parâmetros *Bagged Decision Trees* (BDT) utilizou-se a técnica *GridSearchCV* e os parâmetros testados foram:

- n_estimators = número de árvores geradas;

- bootstrap = se utilizaremos o método *bootstrap* (com ou sem substituição) para amostrar;
- bootstrap_features = se utilizaremos o método *bootstrap* (com ou sem substituição) na seleção das variáveis para construção das árvores;

Os parâmetros e índices selecionados para o algoritmo *Bagged Decision Trees* (BDT) são apresentados na tabela 8.

Tabela 8 - Parâmetros utilizados no *Bagged Decision Trees* (BDT)

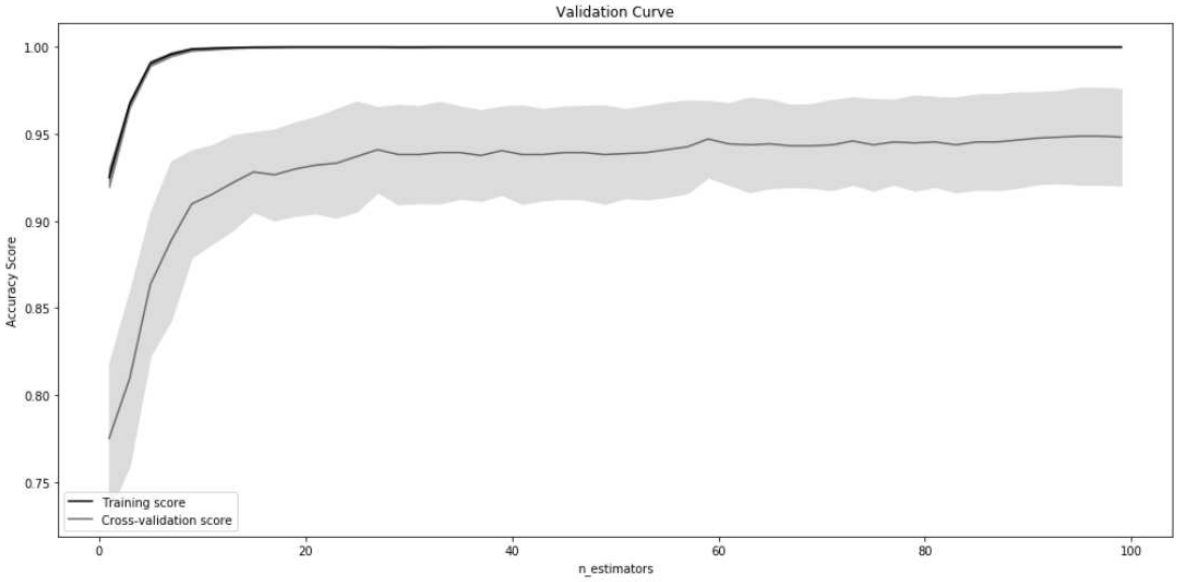
Parâmetros Algoritmos
Bagged Decision Trees (BDT) *Algoritmo utilizado:* from imblearn.ensemble import BalancedBaggingClassifier *Parâmetros utilizados:* base_estimator=DecisionTreeClassifier() n_estimators=75 bootstrap= True bootstrap_features= True random_state=7

Fonte: Autor

Definido os parâmetros do *Bagged Decision Trees* (BDT) inicia-se o treinamento do algoritmo com o conjunto de dados de treino e na sequência a validação do modelo utilizando-se o conjunto de dados de teste (validação).

A primeira análise é a curva de validação apresentada na figura 15. No gráfico observa-se o comportamento uniforme das curvas de treino (*Training Score*) e validação (*Cross-Validation Score*), validando a capacidade de generalização do modelo. Observa-se também que número ideal de árvores geradas (*n_estimator*) é de aproximadamente 75 árvores.

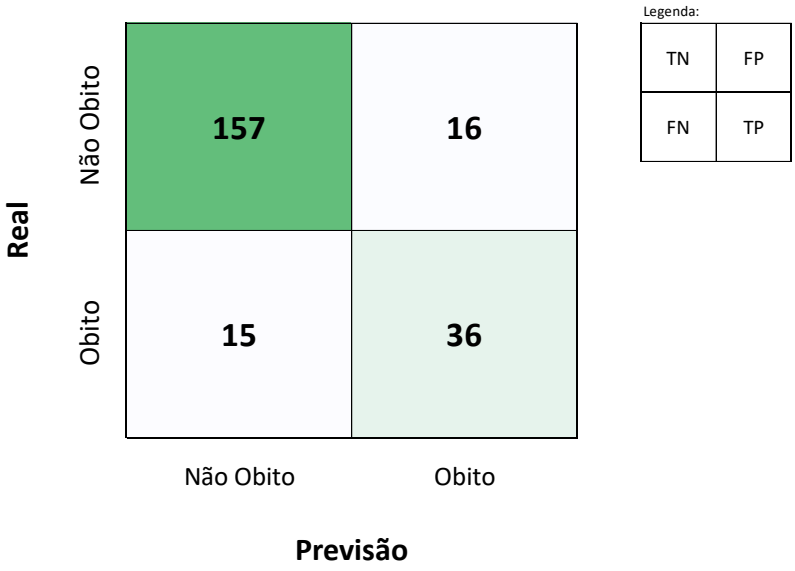
Figura 15 - Curva de Validação do *Bagged Decision Trees* (BDT)



Fonte: Autor

Na figura 16 é apresentado a Matriz de Confusão (*confusion matrix*) do *Bagged Decision Trees* (BDT). Nesta matriz observa-se que o modelo previu corretamente 36 mortes e 157 sobrevida, mas apontou erroneamente 15 mortes como sobrevida, e 16 sobrevida como morte.

Figura 16 - Confusion Matrix do *Bagged Decision Trees* (BDT)



Fonte: Autor

As informações da Matriz de Confusão geraram os índices de Acurácia (*Accuracy*), Sensibilidade (*Sensitivity or Recall*), Especificidade (*Specificity*) e Precisão (*Precision*) do modelo, apresentados na tabela 9.

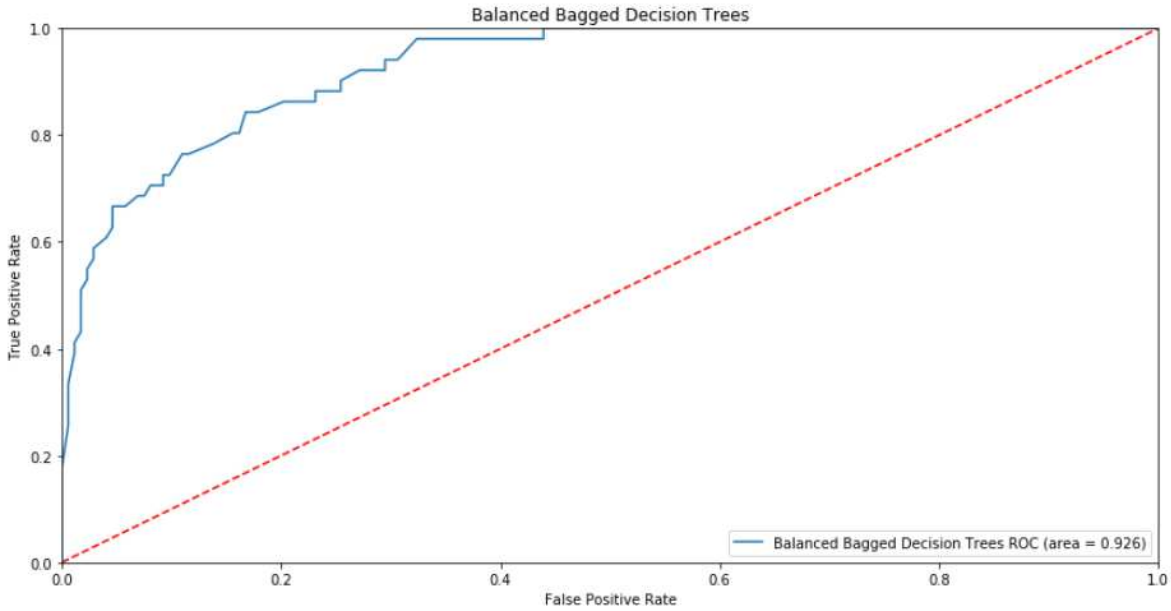
Tabela 9 - Métricas de Performance do *Bagged Decision Trees* (BDT)

```
Balanced Bagged Decision Trees
+-----------------+----------+
| Metric          |      Vlr |
+-----------------+----------+
| Accuracy        | 0.861607 |
| Recall          | 0.705882 |
| Precision       | 0.692308 |
| F1              | 0.699029 |
| AUC Probability | 0.925989 |
+-----------------+----------+
              precision    recall  f1-score   support

           0       0.91      0.91      0.91       173
           1       0.69      0.71      0.70        51

    accuracy                           0.86       224
   macro avg       0.80      0.81      0.80       224
weighted avg       0.86      0.86      0.86       224
```

Fonte: Autor

O modelo obteve uma acurácia de 86,1%, Sensibilidade (*Recall*) de 70,5% e Precisão 69,2%. Na figura 17 é apresentado o gráfico ROC (*Receiver Operating Characteristics*) (FAWCETT, 2004).

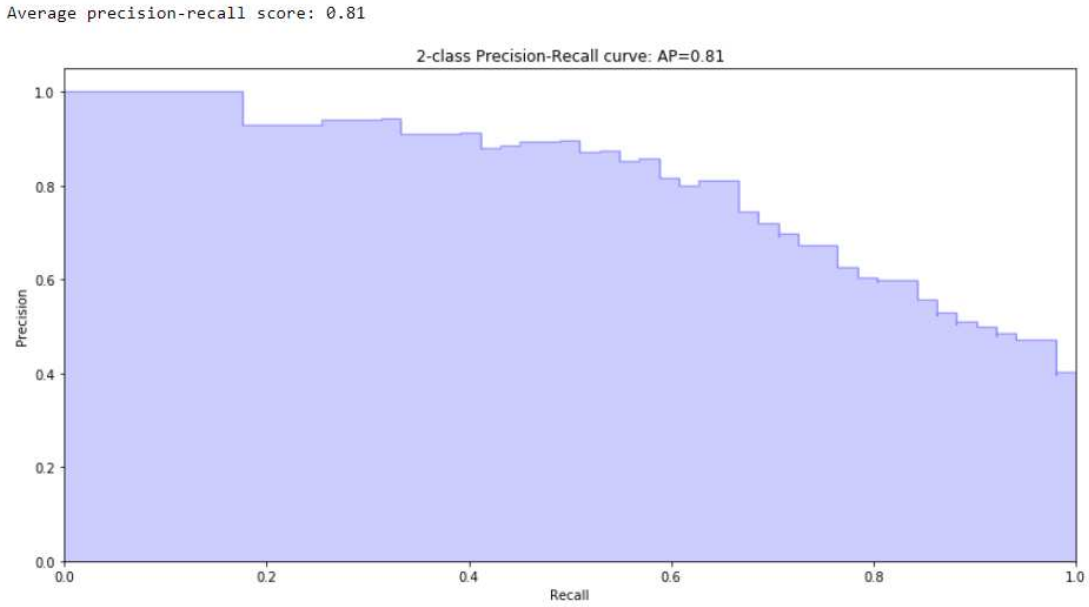
Figura 17 - Curva ROC AUC do *Bagged Decision Trees* (BDT)



Fonte: Autor

O *Bagged Decision Trees (BDT)* obteve um ROC AUC de 92,6%, informação que será utilizada na comparação de performance com os demais algoritmos estudados e também apoiará na decisão do ponto de corte do modelo. E no gráfico 18 é apresentado a Curva de Precisão-Sensibilidade *(Precision-Recall Curves).*

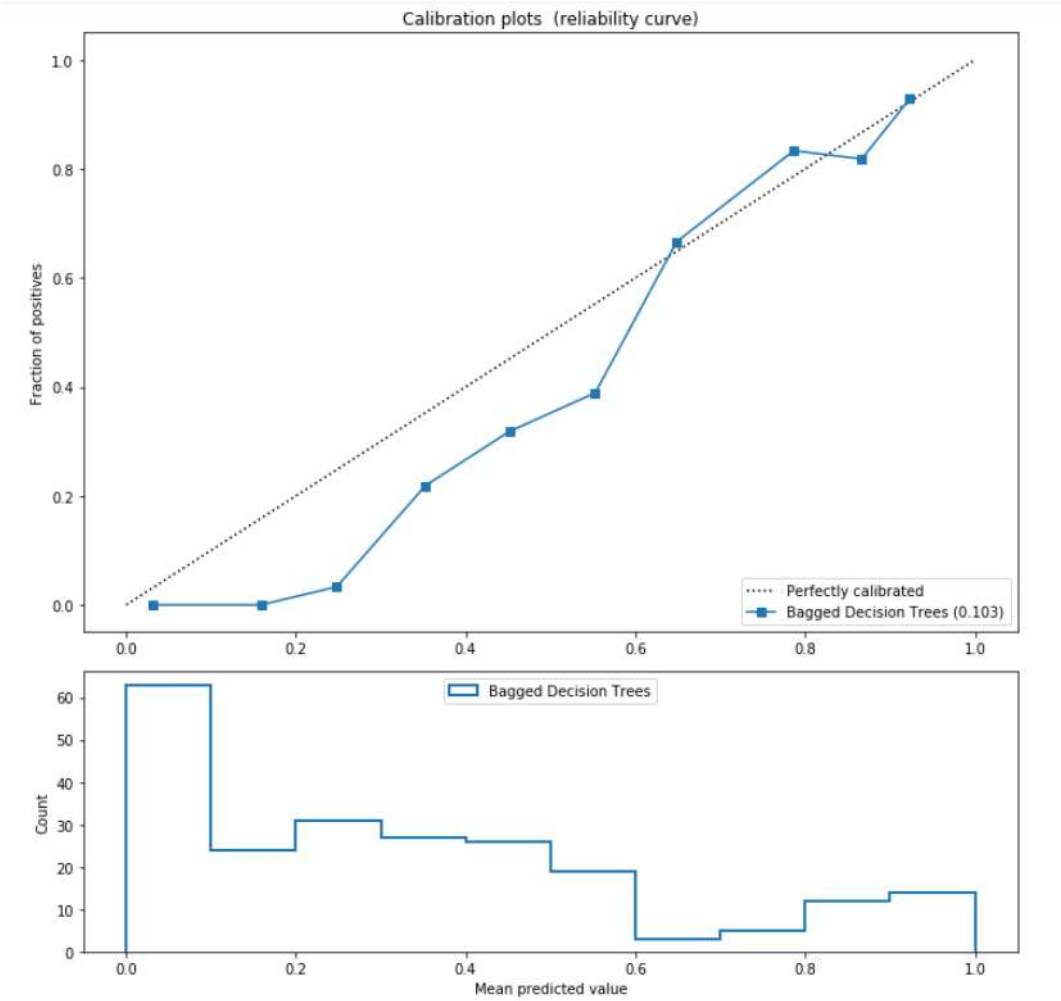
Figura 18 - Precision-Recall Curve do *Bagged Decision Trees* (BDT)



Fonte: Autor

O índice *Average precision* (AP) do *Bagged Decision Trees* (BDT) foi de 0,81. E na figura 19 é apresentado o gráfico de calibração das probabilidades.

Figura 19 - Curva Calibragem *Bagged Decision Trees* (BDT)



Fonte: Autor

Finaliza-se as análises do *Bagged Decision Trees (BDT)* e o resumo das métricas de performance são apresentadas na tabela 10.

Tabela 10 - Performance do *Bagged Decision Trees* (BDT)

Algoritmos	**Acuracia** (Accuracy)	**Precisão** (Precision)	**Sensibilidade** (Recall)	**Especificidade** (Specificity)	**ROC AUC**	**AP**
Bagged Decision Trees	86,2%	69,2%	70,6%	90,8%	92,6%	0,81

Fonte: Autor

Em resumo, não foi identificado efeitos de *Overfitting* ou *Underfitting* no treinamento do *Bagged Decision Trees* (BDT), habilitando-o para comparação de performance com os demais algoritmos utilizados nesse estudo. Apesar da simplicidade do treinamento desse algoritmo, os resultados foram muito próximos do *Random Forest* (RF), mas com a desvantagem de utilizar todas as variáveis do conjunto de treino, o que aumenta a complexidade de implementação, e a ausência de análise de importância das variáveis utilizadas, impossibilitando a compreensão dos efeitos que aumentam o risco de morte.

4.3 STOCHASTIC GRADIENT BOOSTING (SGB)

O treinamento do *Stochastic Gradient Boosting* (SGB) foi dividido nas seguintes etapas: balanceamento do conjunto de dados, seleção das variáveis relevantes para a previsão, ajustes dos parâmetros do modelo e análise dos resultados obtidos.

Como o conjunto de dados é desbalanceado, antes de iniciar o treinamento, é recomendável balancear as classes e assim garantir melhor qualidade nas previsões. No balanceamento foi aplicado uma técnica de sobre-amostragem (*over-sampling*) que gera cópias aleatórias com substituição das observações da classe minoritária e garanti o equilíbrio da representatividade entre as classes estudadas. As quantidades de observações de antes e depois de aplicada a técnica de sobre-amostragem (*over-sampling*) é apresentada na figura 20.

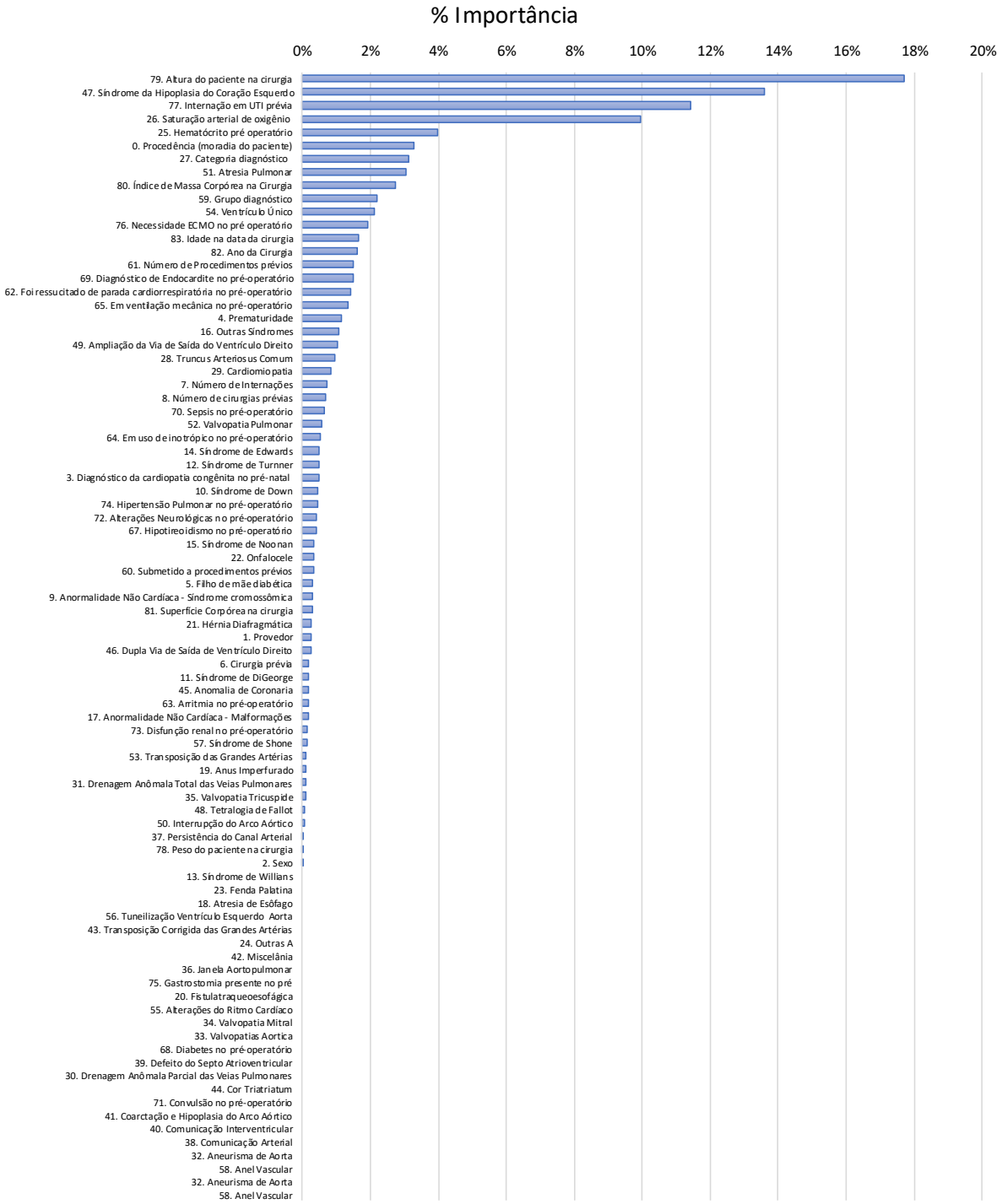
Figura 20 - Resultado balanceamento para *Stochastic Gradient Boosting* (SGB)

```
Qtde de amostras ANTES do rebalanceamento

(2024, 86) (2024,)
length of oversampled data is  2024
Number of no subscription in oversampled data 1832
Number of subscription 192
Proportion of no subscription data in oversampled data is  0.9051383399209486
Proportion of subscription data in oversampled data is  0.09486166007905138

Qtde de amostras DEPOIS do rebalanceamento

(3664, 86) (3664,)
length of oversampled data is  3664
Number of no subscription in oversampled data 1832
Number of subscription 1832
Proportion of no subscription data in oversampled data is  0.5
Proportion of subscription data in oversampled data is  0.5
```

Fonte: Autor

Como no *Random Forest* (RF), o algoritmo *Stochastic Gradient Boosting* (SGB) também possibilita a classificação da importância das variáveis, análise que auxilia na exclusão das variáveis irrelevantes para a previsão. No gráfico 21 é apresentado o resultado da classificação de importância das variáveis.

Figura 21 - Importância das Variáveis do *Stochastic Gradient Boosting* (SGB)



Fonte: Autor

A classificação da importância das variáveis utilizadas no *Stochastic Gradient Boosting* (SGB) também auxilia na identificação de quais variáveis ou diagnósticos influenciam na previsão de morte. Na tabela 11 são apresentados as variáveis selecionadas e o seu índice de importância.

Tabela 11 - Variáveis selecionadas para o *Stochastic Gradient Boosting* (SGB)

Variáveis Selecionadas			
Feature	**Variável Selecionada**	**% importância**	**Descrição**
79	PRE_estat_cm_adm1_83	17,7%	Altura do paciente na cirurgia
47	PRE_SHCE_49	13,6%	Síndrome da Hipoplasia do Coração Esquerdo
77	PRE_UTI_Previa_2sim_1nao_adm1_81	11,4%	Internação em UTI prévia
26	PRE_sato2_preop_adm1_28	10,0%	Saturação arterial de oxigênio
25	PRE_ht_preop_adm1_27	4,0%	Hematócrito pré operatório
0	PRE_proc_2	3,3%	Procedência (moradia do paciente)
27	PRE_Diagnostico_categoria1a12_29	3,1%	Categoria diagnóstico
51	PRE_Atresia_pulmonar_53	3,1%	Atresia Pulmonar
80	PRE_IMC_adm1_84	2,7%	Índice de Massa Corpórea na Cirurgia
59	PRE_grupo_diag_61	2,2%	Grupo diagnóstico
76	PRE_ECMO_VAD_pre_80	1,9%	Necessidade ECMO no pré operatório
83	PRE_IDADE_CIRURGIA_dias_adm1_90	1,6%	Idade na data da cirurgia
82	PRE_ANO_CC_ADM1_86	1,6%	Ano da Cirurgia
61	PRE_proc_previos_01_64	1,5%	Número de Procedimentos prévios
69	PRE_Endocardite_pre_Adm1_72	1,5%	Diagnóstico de Endocardite no pré-operatório
62	PRE_Ressuc_PCR_Pre_Adm1_65	1,4%	Ressuscitado de parada cardiorrespiratória no pré-operatório
65	PRE_Vent_mec_Pre_Adm1_68	1,4%	Em ventilação mecânica no pré-operatório
16	PRE_outcrom_18	1,1%	Outras Síndromes
49	PRE_RVOT_51	1,1%	Ampliação da Via de Saída do Ventrículo Direito
29	PRE_CMP_31	0,8%	Cardiomiopatia
7	PRE_intpre_9	0,7%	Número de Internações
8	PRE_ncirpre_10	0,7%	Número de cirurgias prévias
70	PRE_Sepsis_pre_Adm1_73	0,6%	Sepsis no pré-operatório
52	PRE_Doenca_ValvaPulm_54	0,6%	Valvopatia Pulmonar
3	PRE_prenat_5	0,5%	Diagnóstico da cardiopatia congênita no pré-natal
74	PRE_Hipert_Pulm_pre_Adm1_77	0,4%	Hipertensão Pulmonar no pré-operatório
72	PRE_Alt_Neuro_pre_adm1_75	0,4%	Alterações Neurológicas no pré-operatório
67	PRE_Hipotiroidismo_pre_Adm1_70	0,4%	Hipotireoidismo no pré-operatório
60	PRE_proc_previos_2Sim_1Nao_63	0,3%	Submetido a procedimentos prévios
5	PRE_maediab_7	0,3%	Filho de mãe diabética
9	PRE_sindcrom_11	0,3%	Anormalidade Não Cardíaca - Síndrome cromossômica
81	PRE_SC_adm1_85	0,3%	Superfície Corpórea na cirurgia
1	PRE_prov_3	0,3%	Provedor
6	PRE_cirpre_8	0,2%	Cirurgia prévia
11	PRE_digeorge_13	0,2%	Síndrome de DiGeorge

57	PRE_Shone_59	0,1%	Síndrome de Shone
48	PRE_T4F_50	0,1%	Tetralogia de Fallot
37	PRE_PCA_39	0,0%	Persistência do Canal Arterial
78	PRE_peso_kg_adm1_82	0,0%	Peso do paciente na cirurgia
38	PRE_CIA_40	0,0%	Comunicação Arterial
30	PRE_DAPVVPP_32	0,0%	Drenagem Anômala Parcial das Veias Pulmonares
34	PRE_Doenca_ValvaMi_36	0,0%	Valvopatia Mitral

Fonte: Autor

Observa-se na tabela 11 que as variáveis “Altura do paciente na cirurgia” (17,7%), “Síndrome da Hipoplasia do Coração Esquerdo” (13,6%), “Internação em UTI prévia” (11,4%) e “Saturação arterial de oxigênio” (10,0%) foram classificadas como as mais importantes para o algoritmo, essas 4 variáveis somadas respondem por 52,7% da importância classificada. Após classificar a importância das variáveis utiliza-se a técnica chamada de RFE (*Recursive Feature Elimination*) (GUYON et al., 2002) para selecionar as variáveis importantes na previsão que serão utilizadas no treinamento.

Com o conjunto de dados de treino balanceado e com apenas as variáveis selecionadas pela técnica RFE, inicia-se a etapa de ajustes dos parâmetros. Nesta etapa utilizou-se apenas a técnica *GridSearchCV,* uma vez que não havia muitos parâmetros de ajuste para este algoritmo. O parâmetro do algoritmo *Stochastic Gradient Boosting* (SGB) testado foi:

- n_estimators = número de árvores geradas;

Os parâmetros e índices selecionados para o algoritmo *Stochastic Gradient Boosting* (SGB) são apresentados na tabela 12.

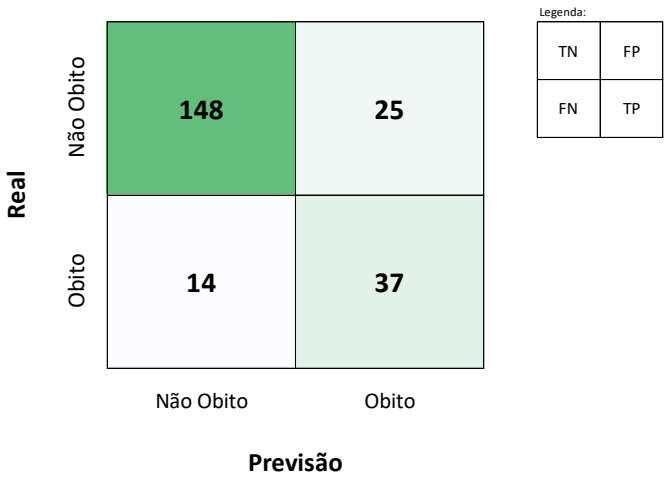
Tabela 12 - Parâmetros utilizados no *Stochastic Gradient Boosting* (SGB)

Parâmetros Algoritmos
Stochastic Gradient Boosting (SGB) *Algoritmo utilizado:* from sklearn.ensemble import GradientBoostingClassifier *Parâmetros utilizados:* n_estimators = 50 random_state = 7

Fonte: Autor

Definido os parâmetros do *Stochastic Gradient Boosting* (SGB) inicia-se o treinamento do algoritmo com o conjunto de dados de treino e na sequência validação da performance de previsão do modelo com o conjunto de dados de teste (validação). Os resultados de performance do conjunto de dados de validação são apresentados na Matriz de Confusão (*confusion matrix*), da figura 22.

Figura 22 - Confusion Matrix do *Stochastic Gradient Boosting* (SGB)



Fonte: Autor

As informações da Matriz de Confusão geraram os índices de acurácia (*accuracy*), sensibilidade (*sensitivity or Recall*), especificidade (*specificity*) e precisão (*precision*) do modelo, apresentados na tabela 13.

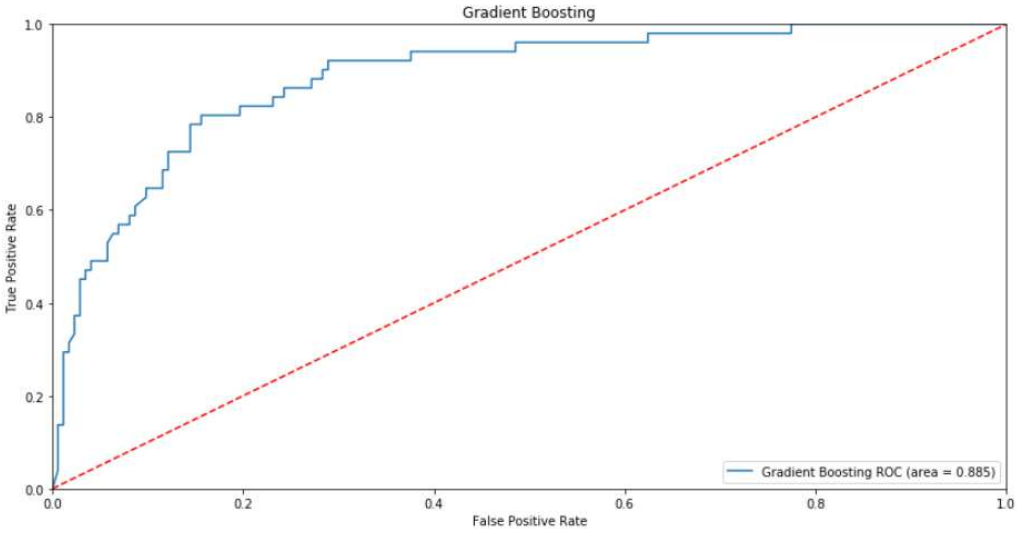
Tabela 13 - Métricas de Performance do *Stochastic Gradient Boosting* (SGB)

```
Gradient Boosting
+-----------------+----------+
| Metric          |      Vlr |
+-----------------+----------+
| Accuracy        | 0.825893 |
| Recall          | 0.725490 |
| Precision       | 0.596774 |
| F1              | 0.654867 |
| AUC Probability | 0.885413 |
+-----------------+----------+
              precision    recall  f1-score   support

           0       0.91      0.86      0.88       173
           1       0.60      0.73      0.65        51

    accuracy                           0.83       224
   macro avg       0.76      0.79      0.77       224
weighted avg       0.84      0.83      0.83       224
```

Fonte: Autor

O modelo obteve uma Acurácia (*Accuracy*) de 82,5%, Sensibilidade (*Recall*) de 72,5% e Precisão (*Precision*) 59,6%. Na figura 23 é apresentado o gráfico ROC AUC.

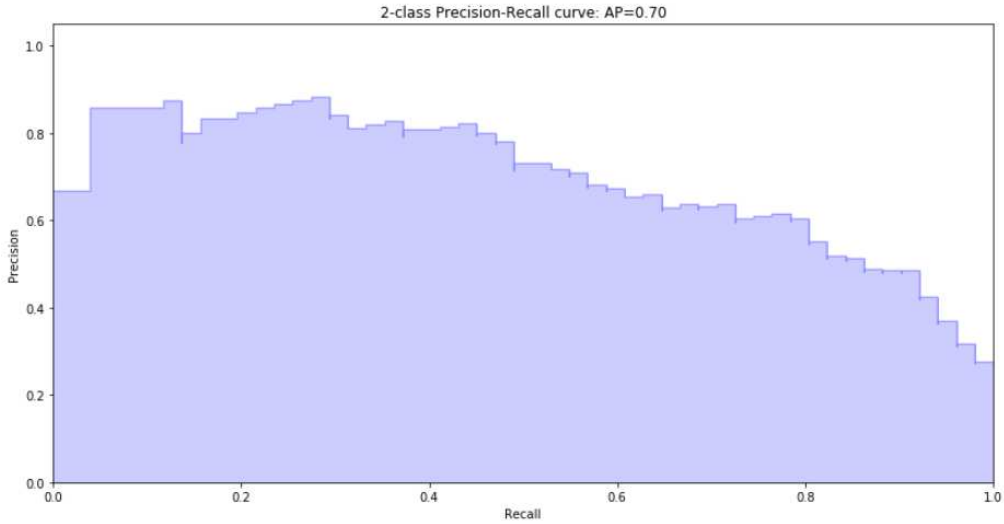
Figura 23 - Curva ROC AUC do *Stochastic Gradient Boosting* (SGB)



Fonte: Autor

O *Stochastic Gradient Boosting* (SGB) obteve um ROC AUC de 88,5%, informação que será utilizada na comparação de performance entre os algoritmos estudados e apoiar na definição do ponto de corte de previsão da morte. E no gráfico 24 é apresentado a Curva de Precisão-Sensibilidade *(Precision-Recall Curves).*

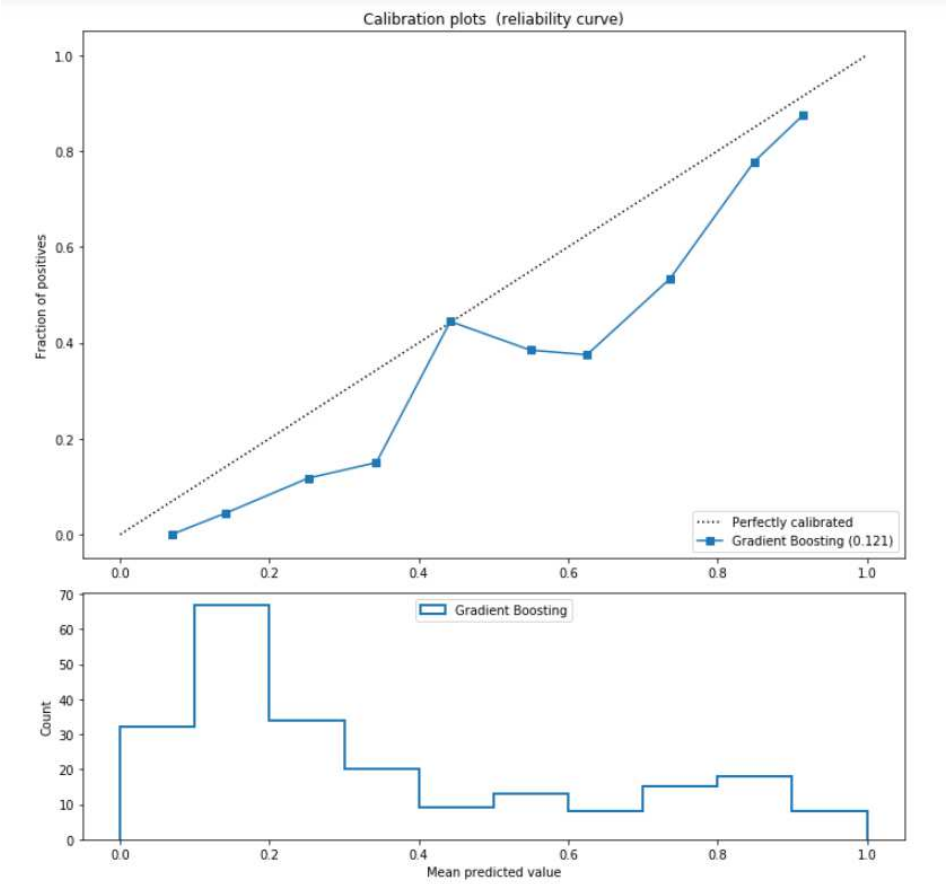
Figura 24 - Precision-Recall Curve do *Stochastic Gradient Boosting* (SGB)



Fonte: Autor

O índice *Average precision* (AP) do *Stochastic Gradient Boosting* (SGB) é de 0,70. E na figura 25 é apresentado o gráfico de calibração das probabilidades.

Figura 25 - Gráfico de Calibragem do *Stochastic Gradient Boosting* (SGB)



Fonte: Autor

O resumo das métricas de performance do *Stochastic Gradient Boosting* (SGB) são apresentadas na tabela 14.

Tabela 14 - Performance do *Stochastic Gradient Boosting* (SGB)

Algoritmos	**Acuracia** (Accuracy)	**Precisão** (Precision)	**Sensibilidade** (Recall)	**Especificidade** (Specificity)	**ROC AUC**	**AP**
Stochastic Gradient Boosting	82,6%	59,7%	72,5%	85,5%	88,5%	0,70

Fonte: Autor

Em resumo, não foi identificado efeitos de *Overfitting* ou *Underfitting* no treinamento do *Stochastic Gradient Boosting* (SGB), validando sua performance e habilitando-o para comparação de performance com os demais algoritmos utilizados neste estudo.

4.4 EXTRA TREES (ET)

Para o *Extra Trees* (ET) utilizou-se mesma divisão das etapas de treinamento do *Stochastic Gradient Boosting* (SGB): balanceamento do conjunto de dados, seleção das variáveis relevantes para a previsão, ajustes dos parâmetros do modelo e análise dos resultados obtidos.

Antes de iniciar o treinamento, precisa-se atentar para o desequilíbrio de classes existente no conjunto de dados de treino. Como o *Extra Trees* (ET) não possui função de ajuste para treino de conjuntos desbalanceados, como no *Random Forest* (RF) e no *Bagged Decision Trees* (BDT), foi realizado o balanceamento no conjunto de dados de treino antes de iniciar o treinamento. Para o balanceamento das classes foi adotada a técnica de sobre-amostragem (*over-sampling*) aleatória com substituição, mesma técnica utilizada no treinamento do *Stochastic Gradient Boosting* (SGB). As quantidades de observações de antes e depois de aplicada a técnica de sobre amostragem é apresentada na figura 26.

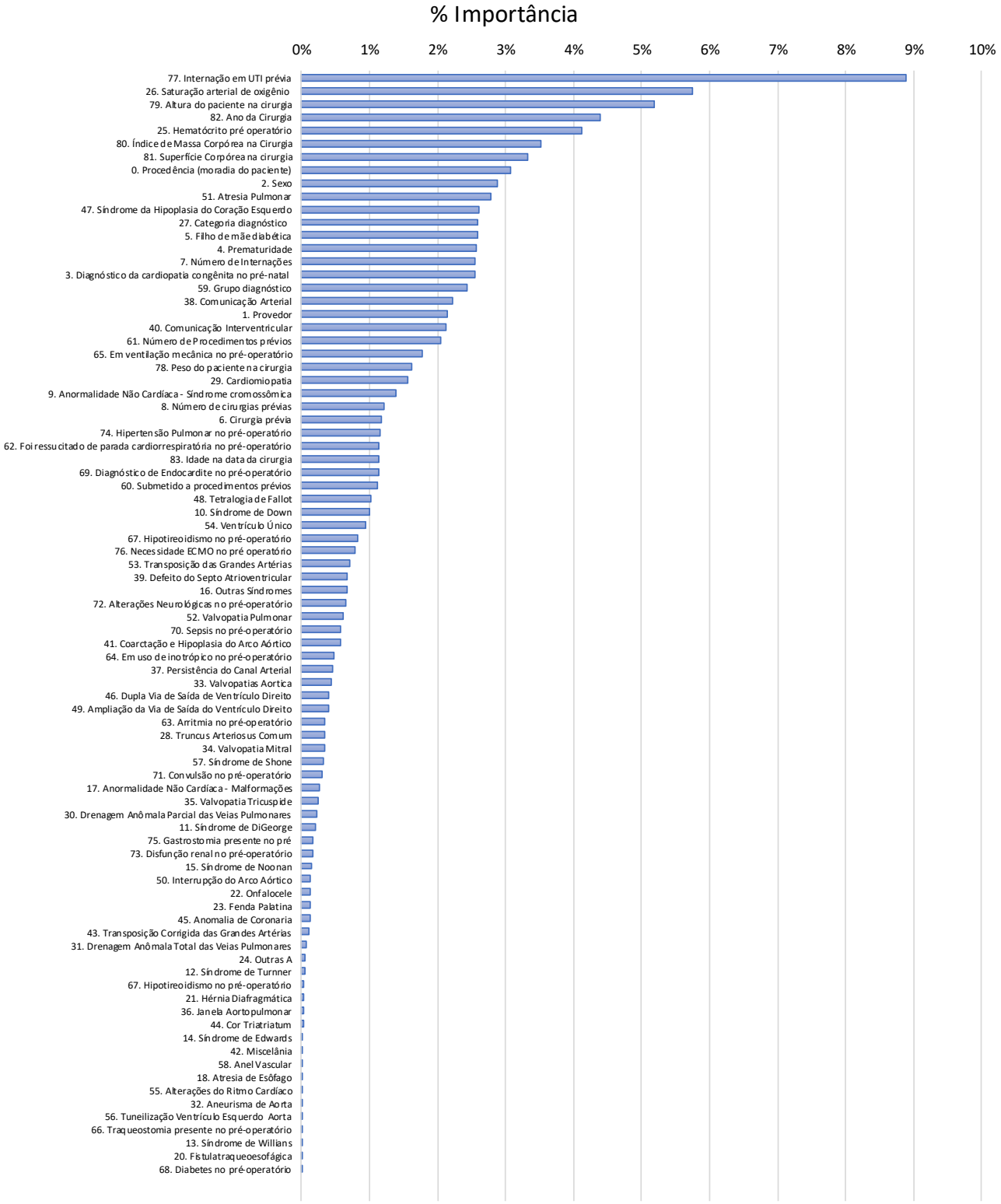
Figura 26 - Resultado balanceamento para *Extra Trees* (ET)

```
Qtde de amostras ANTES do rebalanceamento

(2024, 86) (2024,)
length of oversampled data is  2024
Number of no subscription in oversampled data 1832
Number of subscription 192
Proportion of no subscription data in oversampled data is  0.9051383399209486
Proportion of subscription data in oversampled data is  0.09486166007905138

Qtde de amostras DEPOIS do rebalanceamento

(3664, 86) (3664,)
length of oversampled data is  3664
Number of no subscription in oversampled data 1832
Number of subscription 1832
Proportion of no subscription data in oversampled data is  0.5
Proportion of subscription data in oversampled data is  0.5
```

Fonte: Autor

O algoritmo *Extra Trees* (ET) também possibilita classificar a importância das variáveis, análise que auxilia na exclusão das variáveis irrelevantes para a previsão. No gráfico 27 é apresentado o resultado desta classificação.

Figura 27 - Importância das Variáveis do *Extra Trees* (ET)



Fonte: Autor

Uma vez classificada a importância das variáveis é aplicada a técnica RFE (*Recursive Feature Elimination*) (GUYON et al., 2002) para selecionar as variáveis importantes na previsão. Na tabela 15 são apresentados as variáveis selecionadas e o seu índice de importância.

Tabela 15 - Variáveis selecionadas para o *Extra Trees* (ET)

Variáveis Selecionadas			
Feature	**Variável Selecionada**	**% importância**	**Descrição**
77	PRE_UTI_Previa_2sim_1nao_adm1_81	8,9%	Internação em UTI prévia
26	PRE_sato2_preop_adm1_28	5,7%	Saturação arterial de oxigênio
79	PRE_estat_cm_adm1_83	5,2%	Altura do paciente na cirurgia
82	PRE_ANO_CC_ADM1_86	4,4%	Ano da Cirurgia
25	PRE_ht_preop_adm1_27	4,1%	Hematócrito pré-operatório
80	PRE_IMC_adm1_84	3,5%	Índice de Massa Corpórea na Cirurgia
81	PRE_SC_adm1_85	3,3%	Superfície Corpórea na cirurgia
0	PRE_proc_2	3,1%	Procedência (moradia do paciente)
2	PRE_sexo_4	2,9%	Sexo
51	PRE_Atresia_pulmonar_53	2,8%	Atresia Pulmonar
47	PRE_SHCE_49	2,6%	Síndrome da Hipoplasia do Coração Esquerdo
27	PRE_Diagnostico_categoria1a12_29	2,6%	Categoria diagnóstico
5	PRE_maediab_7	2,6%	Filho de mãe diabética
4	PRE_premat_6	2,6%	Prematuridade
7	PRE_intpre_9	2,6%	Número de Internações
3	PRE_prenat_5	2,5%	Diagnóstico da cardiopatia congênita no pré-natal
59	PRE_grupo_diag_61	2,4%	Grupo diagnóstico
38	PRE_CIA_40	2,2%	Comunicação Arterial
1	PRE_prov_3	2,2%	Provedor
40	PRE_CIV_42	2,1%	Comunicação Interventricular
61	PRE_proc_previos_01_64	2,0%	Número de Procedimentos prévios
65	PRE_Vent_mec_Pre_Adm1_68	1,8%	Em ventilação mecânica no pré-operatório
78	PRE_peso_kg_adm1_82	1,6%	Peso do paciente na cirurgia
29	PRE_CMP_31	1,6%	Cardiomiopatia
9	PRE_sindcrom_11	1,4%	Anormalidade Não Cardíaca - Síndrome cromossômica
8	PRE_ncirpre_10	1,2%	Número de cirurgias prévias
6	PRE_cirpre_8	1,2%	Cirurgia prévia
74	PRE_Hipert_Pulm_pre_Adm1_77	1,1%	Hipertensão Pulmonar no pré-operatório
62	PRE_Ressuc_PCR_Pre_Adm1_65	1,1%	Foi ressuscitado de parada cardiorrespiratória no pré-operatório
83	PRE_IDADE_CIRURGIA_dias_adm1_90	1,1%	Idade na data da cirurgia
69	PRE_Endocardite_pre_Adm1_72	1,1%	Diagnóstico de Endocardite no pré-operatório
60	PRE_proc_previos_2Sim_1Nao_63	1,1%	Submetido a procedimentos prévios
48	PRE_T4F_50	1,0%	Tetralogia de Fallot
10	PRE_down_12	1,0%	Síndrome de Down
54	PRE_Ventriculo_unico_56	0,9%	Ventrículo Único
67	PRE_Hipotiroidismo_pre_Adm1_70	0,8%	Hipotireoidismo no pré-operatório
76	PRE_ECMO_VAD_pre_80	0,8%	Necessidade ECMO no pré-operatório

53	PRE_TGA_55	0,7%	Transposição das Grandes Artérias
39	PRE_DSAV_41	0,7%	Defeito do Septo Atrioventricular
16	PRE_outcrom_18	0,7%	Outras Síndromes
72	PRE_Alt_Neuro_pre_adm1_75	0,7%	Alterações Neurológicas no pré-operatório
52	PRE_Doenca_ValvaPulm_54	0,6%	Valvopatia Pulmonar

Fonte: Autor

Observa-se na tabela 15 que as variáveis “Internação em UTI prévia” (8,9%), “Saturação arterial de oxigênio” (5,7%) e “Altura do paciente na cirurgia” (5,2%) foram classificadas como as mais importantes deste algoritmo, somadas respondem por 19,8% da importância classificada.

Com o conjunto de dados de treino balanceado e com apenas as variáveis selecionadas pela técnica RFE, pode-se iniciar a etapa de ajustes dos parâmetros. Nessa etapa utiliza-se a técnica *GridSearchCV.* Os parâmetros do algoritmo *Extra Trees* (ET) testados foram:

- n_estimators = número de árvores geradas;
- max_features = número máximo de variáveis considerados para dividir um nó;
- max_depth = número máximo de níveis em cada Árvore de Decisão;
- min_samples_split = número mínimo de pontos de dados colocados em um nó antes do nó ser dividido;
- bootstrap = se utilizaremos o método *bootstrap* (com ou sem substituição) para amostrar;
- criterion = função de medição da qualidade de uma divisão. Os critérios suportados são "*gini*" para a impureza de Gini e "*entropy*" para o ganho de informações.

E os parâmetros e índices selecionados para o algoritmo *Extra Trees* (ET) são apresentados na tabela 16.

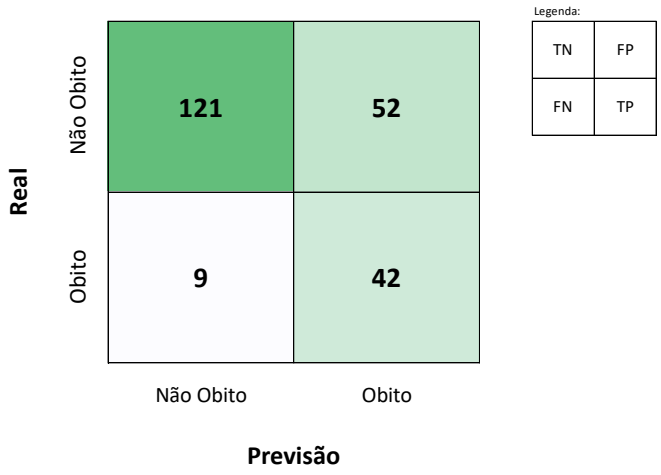
Tabela 16 - Parâmetros utilizados no *Extra Trees* (ET)

Parâmetros Algoritmos
Extra Trees (ET) *Algoritmo utilizado:* from sklearn.ensemble import ExtraTreesClassifier *Parâmetros utilizados:* bootstrap = False random_state=1986 criterion='entropy' max_depth=3 max_features=3 min_samples_split=3 n_estimators=25

Fonte: Autor

Definido os parâmetros do *Extra Trees* (ET) inicia-se o treinamento do algoritmo com o conjunto de dados de treino e na sequência a validação do algoritmo com o conjunto de dados de teste (validação). Os resultados de performance do conjunto de dados de validação são apresentados na Matriz de Confusão (*confusion matrix*) da figura 28.

Figura 28 - Confusion Matrix do *Extra Trees* (ET)



Fonte: Autor

As informações da Matriz de Confusão geraram os índices de Acurácia (*Accuracy*), Sensibilidade (*Sensitivity or Recall*), especificidade (*Specificity*) e Precisão (*Precision*) do modelo, apresentados na tabela 17.

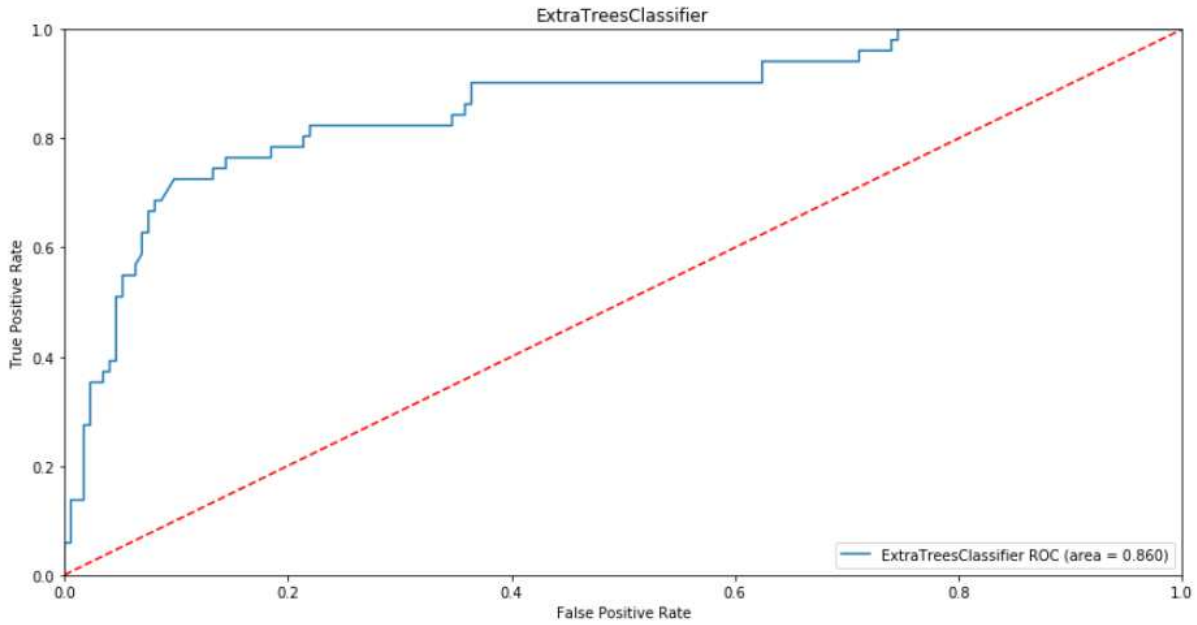
Tabela 17 - Métricas de Performance do *Extra Trees* (ET)

```
ExtraTreesClassifier
+-----------------+----------+
| Metric          |      Vlr |
+-----------------+----------+
| Accuracy        | 0.727679 |
| Recall          | 0.823529 |
| Precision       | 0.446809 |
| F1              | 0.579310 |
| AUC Probability | 0.859742 |
+-----------------+----------+
              precision    recall  f1-score   support

           0       0.93      0.70      0.80       173
           1       0.45      0.82      0.58        51

    accuracy                           0.73       224
   macro avg       0.69      0.76      0.69       224
weighted avg       0.82      0.73      0.75       224
```

Fonte: Autor

O modelo obteve uma acurácia de 72,7%, sensibilidade (*recall*) de 82,3% e precisão 44,7%. Na figura 29 é apresentado o gráfico ROC AUC.

Figura 29 - Curva ROC AUC do *Extra Trees* (ET)



Fonte: Autor

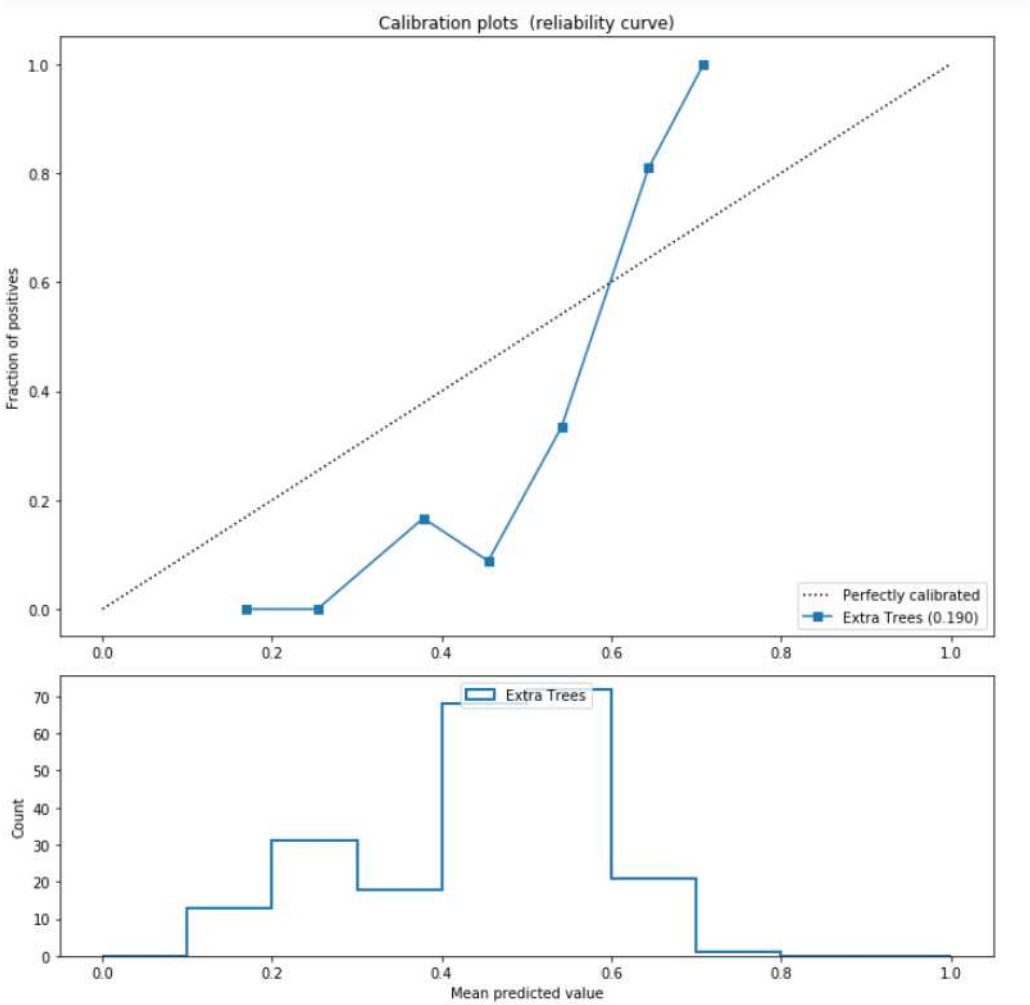
O *Extra Trees* (ET) obteve um ROC AUC de 86%, informação que será utilizada na comparação de performance entre os algoritmos estudados e apoiar na definição do ponto de corte de previsão da morte. E no gráfico 30 é apresentado a Curva de Precisão-Sensibilidade *(Precision-Recall Curves).*

Figura 30 - Precision-Recall Curve do *Extra Trees* (ET)

Fonte: Autor

O índice *Average precision* (AP) do *Extra Trees* (ET) é de 0,69. E na figura 31 é apresentado o gráfico de calibração das probabilidades.

Figura 31 - Gráfico de Calibragem do *Extra Trees* (ET)



Fonte: Autor

O resumo das métricas de performance do *Extra Trees* (ET) é apresentado na tabela 18.

Tabela 18 - Performance do *Extra Trees* (ET)

Algoritmos	**Acuracia** (Accuracy)	**Precisão** (Precision)	**Sensibilidade** (Recall)	**Especificidade** (Specificity)	**ROC AUC**	**AP**
Extra Trees	72,8%	44,7%	82,4%	69,9%	86,0%	0,69

Fonte: Autor

Não foi identificado efeitos de *Overfitting* ou *Underfitting* no treinamento do *Extra Trees* (ET), validando sua performance e habilitando-o para comparação de performance com os demais algoritmos utilizados nesse estudo.

4.5 ADABOOST CLASSIFICATION (ABC)

O treinamento do *Adaboost Classification* (ABC) será dividido nas seguintes etapas: balanceamento do conjunto de dados, seleção das variáveis relevantes para a previsão, ajustes dos parâmetros do modelo e análise dos resultados obtidos.

Antes de iniciar o treinamento, realizou-se o balanceamento do conjunto de dados como o realizado no treinamento do *Extra Trees* (ET) e *Stochastic Gradient Boosting* (SGB). Para o balanceamento do conjunto de dados utilizou-se novamente a técnica de sobre amostragem aleatória por substituição. O resultado é apresentado na figura 32.

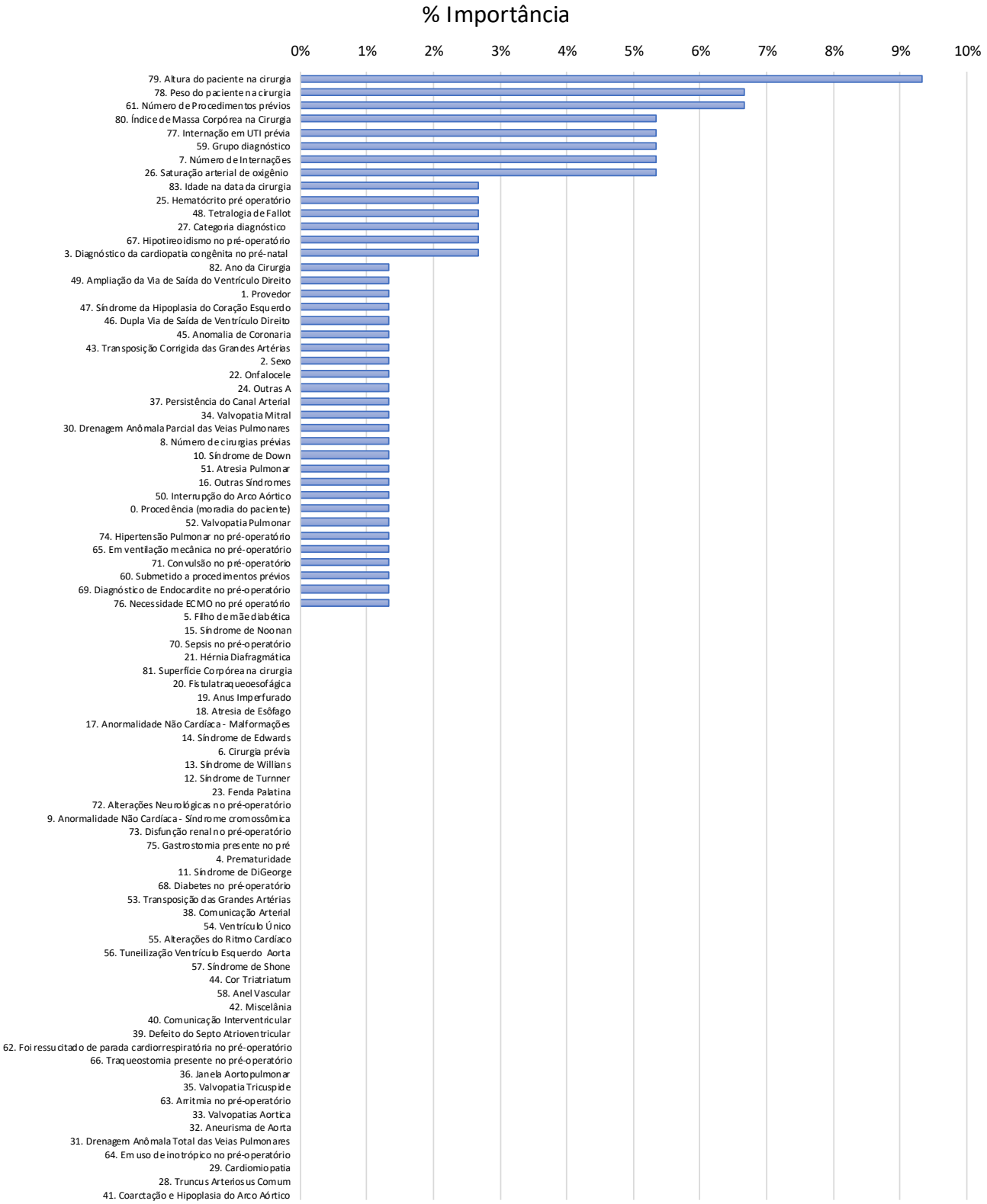
Figura 32 - Resultado balanceamento para *Adaboost Classification* (ABC)

```
Qtde de amostras ANTES do rebalanceamento

(2024, 86) (2024,)
length of oversampled data is  2024
Number of no subscription in oversampled data 1832
Number of subscription 192
Proportion of no subscription data in oversampled data is  0.9051383399209486
Proportion of subscription data in oversampled data is  0.09486166007905138

Qtde de amostras DEPOIS do rebalanceamento

(3664, 86) (3664,)
length of oversampled data is  3664
Number of no subscription in oversampled data 1832
Number of subscription 1832
Proportion of no subscription data in oversampled data is  0.5
Proportion of subscription data in oversampled data is  0.5
```

Fonte: Autor

Como os demais algoritmos *Ensemble*, o *Adaboost Classification* (ABC) também possibilita a classificação de importância das variáveis, análise que auxilia na exclusão das variáveis irrelevantes na previsão. No gráfico 33 é apresentado o resultado desta classificação.

Figura 33 - Importância das Variáveis do *Adaboost Classification* (ABC)



Fonte: Autor

Também utilizou-se a técnica RFE (*Recursive Feature Elimination*) (GUYON et al., 2002) para selecionar as variáveis para o treinamento do *Adaboost Classification* (ABC). Na tabela 19 são apresentados as variáveis selecionadas e o seu índice de importância.

Tabela 19 - Variáveis selecionadas para o *Adaboost Classification* (ABC)

Variáveis Selecionadas			
Feature	**Variável Selecionada**	**% importância**	**Descrição**
79	PRE_estat_cm_adm1_83	9,3%	Altura do paciente na cirurgia
61	PRE_proc_previos_01_64	6,7%	Número de Procedimentos prévios
78	PRE_peso_kg_adm1_82	6,7%	Peso do paciente na cirurgia
77	PRE_UTI_Previa_2sim_1nao_adm1_81	5,3%	Internação em UTI prévia
26	PRE_sato2_preop_adm1_28	5,3%	Saturação arterial de oxigênio
59	PRE_grupo_diag_61	5,3%	Grupo diagnóstico
7	PRE_intpre_9	5,3%	Número de Internações
80	PRE_IMC_adm1_84	5,3%	Índice de Massa Corpórea na Cirurgia
67	PRE_Hipotiroidismo_pre_Adm1_70	2,7%	Hipotireoidismo no pré-operatório
27	PRE_Diagnostico_categoria1a12_29	2,7%	Categoria diagnóstico
48	PRE_T4F_50	2,7%	Tetralogia de Fallot
83	PRE_IDADE_CIRURGIA_dias_adm1_90	2,7%	Idade na data da cirurgia
3	PRE_prenat_5	2,7%	Diagnóstico da cardiopatia congênita no pré-natal
25	PRE_ht_preop_adm1_27	2,7%	Hematócrito pré operatório
0	PRE_proc_2	1,3%	Procedência (moradia do paciente)
22	PRE_onfalocele_24	1,3%	Onfalocele
30	PRE_DAPVVPP_32	1,3%	Drenagem Anômala Parcial das Veias Pulmonares
37	PRE_PCA_39	1,3%	Persistência do Canal Arterial
47	PRE_SHCE_49	1,3%	Síndrome da Hipoplasia do Coração Esquerdo
51	PRE_Atresia_pulmonar_53	1,3%	Atresia Pulmonar
71	PRE_Convulsao_pre_Adm1_74	1,3%	Convulsão no pré-operatório
76	PRE_ECMO_VAD_pre_80	1,3%	Necessidade ECMO no pré operatório
82	PRE_ANO_CC_ADM1_86	1,3%	Ano da Cirurgia
1	PRE_prov_3	1,3%	Provedor
8	PRE_ncirpre_10	1,3%	Número de cirurgias prévias
24	PRE_Outanoanat_26	1,3%	Outras A
43	PRE_TCGA_45	1,3%	Transposição Corrigida das Grandes Artérias
52	PRE_Doenca_ValvaPulm_54	1,3%	Valvopatia Pulmonar
60	PRE_proc_previos_2Sim_1Nao_63	1,3%	Submetido a procedimentos prévios
65	PRE_Vent_mec_Pre_Adm1_68	1,3%	Em ventilação mecânica no pré-operatório
69	PRE_Endocardite_pre_Adm1_72	1,3%	Diagnóstico de Endocardite no pré-operatório
74	PRE_Hipert_Pulm_pre_Adm1_77	1,3%	Hipertensão Pulmonar no pré-operatório
2	PRE_sexo_4	1,3%	Sexo
10	PRE_down_12	1,3%	Síndrome de Down
34	PRE_Doenca_ValvaMi_36	1,3%	Valvopatia Mitral

45	PRE_Anomalia_Coronaria_47	1,3%	Anomalia de Coronaria
49	PRE_RVOT_51	1,3%	Ampliação da Via de Saída do Ventrículo Direito
46	PRE_DVSVD_48	1,3%	Dupla Via de Saída de Ventrículo Direito
50	PRE_Interrupcao_ArcoAO_52	1,3%	Interrupção do Arco Aórtico
16	PRE_outcrom_18	1,3%	Outras Síndromes
31	PRE_DATVVPP_33	0,0%	Drenagem Anômala Total das Veias Pulmonares
6	PRE_cirpre_8	0,0%	Cirurgia prévia

Fonte: Autor

Observa-se na tabela 19 que as variáveis “Altura do paciente na cirurgia” (9,3%), “Número de Procedimentos prévios” (6,7%), “Peso do paciente na cirurgia” (6,7%), “Internação em UTI prévia” (5,3%), “Saturação arterial de oxigênio” (5,3%) e “Grupo diagnóstico” (5,3%) foram classificadas como as mais importantes para o algoritmo, essas 6 variáveis somadas respondem por 38,7% da importância classificada.

Com o conjunto de dados de treino balanceado e com apenas as variáveis selecionadas pela técnica RFE, inicia-se a etapa de ajustes dos parâmetros. No ajuste dos parâmetros do algoritmo Adaboost Classification *(ABC)* utilizou-se a técnica *GridSearchCV*.

Na tabela 20 são apresentados os parâmetros e índices selecionados para o algoritmo *Adaboost Classification* (ABC).

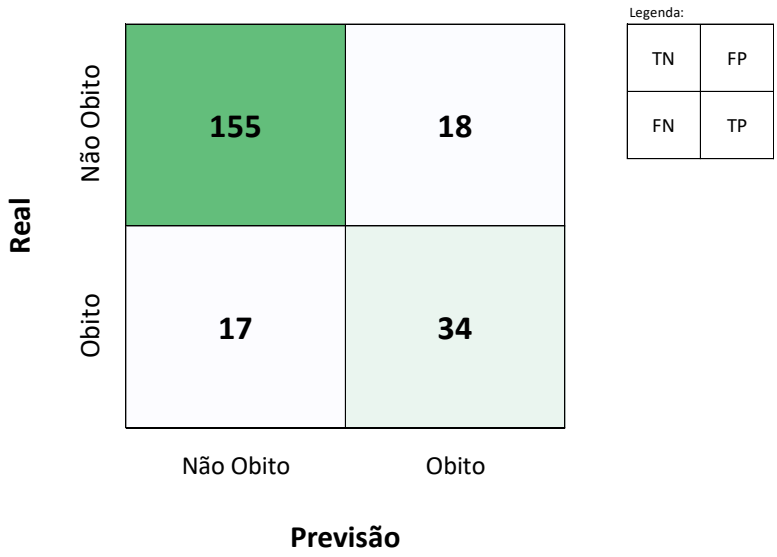
Tabela 20 - Parâmetros utilizados no *Adaboost Classification* (ABC)

Parâmetros Algoritmos
AdaBoost Classification (ABC) *Algoritmo utilizado:* from sklearn.ensemble import AdaBoostClassifier *Parâmetros utilizados:* n_estimators=200 random_state=7

Fonte: Autor

Definido os parâmetros do *Adaboost Classification* (ABC) inicia-se o treinamento do algoritmo com o conjunto de dados de treino e na sequência a validação com o conjunto de dados de teste (validação). Os resultados de performance do conjunto de dados de validação são apresentados na Matriz de Confusão (*confusion matrix*), da figura 34.

Figura 34 - Confusion Matrix do *Adaboost Classification* (ABC)



Fonte: Autor

Na Matriz de Confusão (*confusion matrix*) da figura 33 observa-se que o modelo previu corretamente 34 mortes e 155 sobrevida, mas apontou erroneamente 17 mortes como sobrevida, e 18 sobrevida como morte.

Na tabela 21 é apresentado o resultado de Acurácia (*Accuracy*), Sensibilidade (*Sensitivity or Recall*), Especificidade (*Specificity*), Precisão (*Precision*) e F1 do conjunto de treino e de validação.

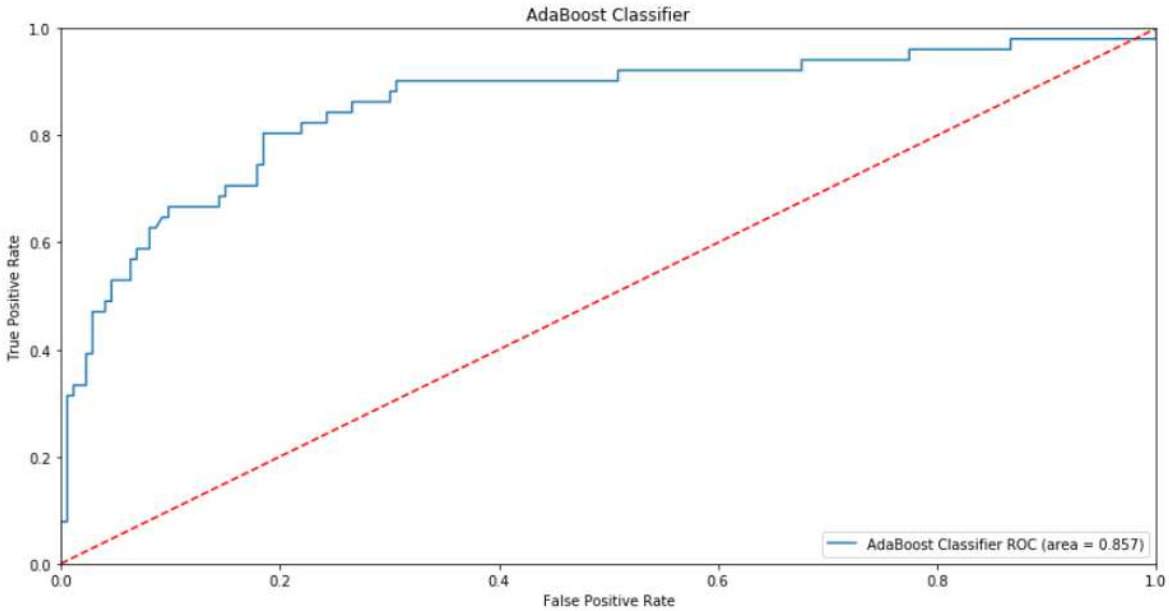
Tabela 21 - Métricas de Performance do *Adaboost Classification* (ABC)

```
AdaBoost Classifier
+-----------------+----------+
| Metric          |      Vlr |
+-----------------+----------+
| Accuracy        | 0.843750 |
| Recall          | 0.666667 |
| Precision       | 0.653846 |
| F1              | 0.660194 |
| AUC Probability | 0.856568 |
+-----------------+----------+
              precision    recall  f1-score   support

           0       0.90      0.90      0.90       173
           1       0.65      0.67      0.66        51

    accuracy                           0.84       224
   macro avg       0.78      0.78      0.78       224
weighted avg       0.84      0.84      0.84       224
```

Fonte: Autor

O modelo obteve uma acurácia de 84,3%, sensibilidade (*recall*) de 66,6%, precisão de 65,3%. Na figura 35 é apresentado o gráfico ROC (*Receiver Operating Characteristics*) (FAWCETT, 2004).

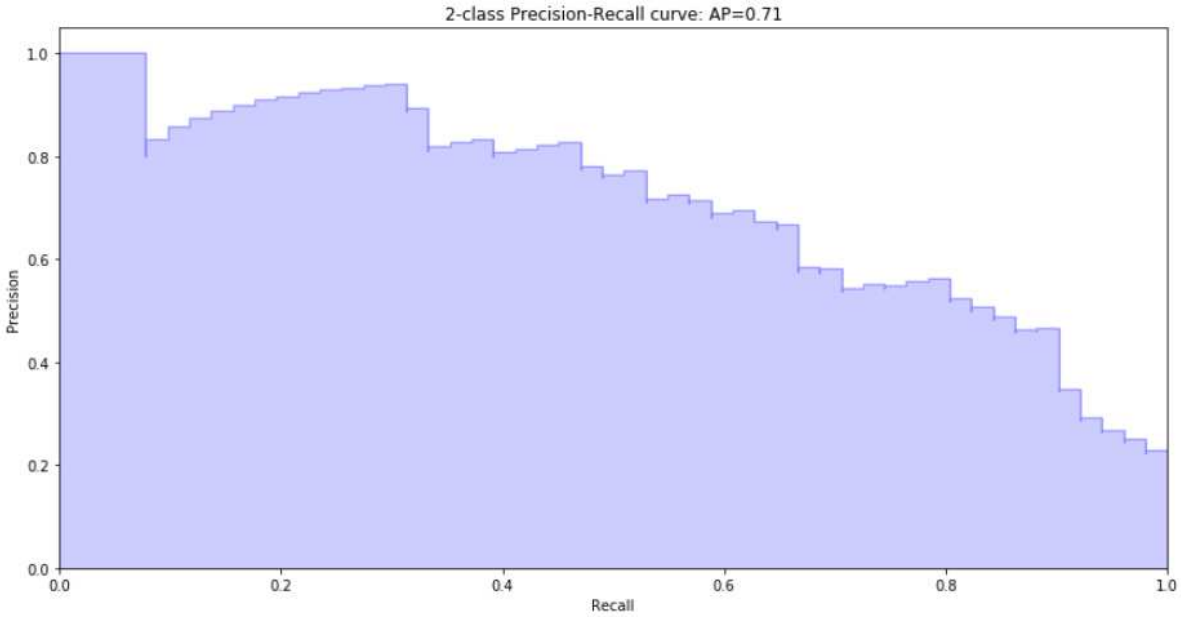
Figura 35 - Curva ROC AUC do *Adaboost Classification* (ABC)



Fonte: Autor

O *Adaboost Classification* (ABC) obteve um ROC AUC de 85,7%, informação que será utilizada na comparação de performance entre os algoritmos estudados e apoiar na definição do ponto de corte de previsão da morte. E no gráfico 36 é apresentado a Curva de Precisão-Sensibilidade *(Precision-Recall Curves).*

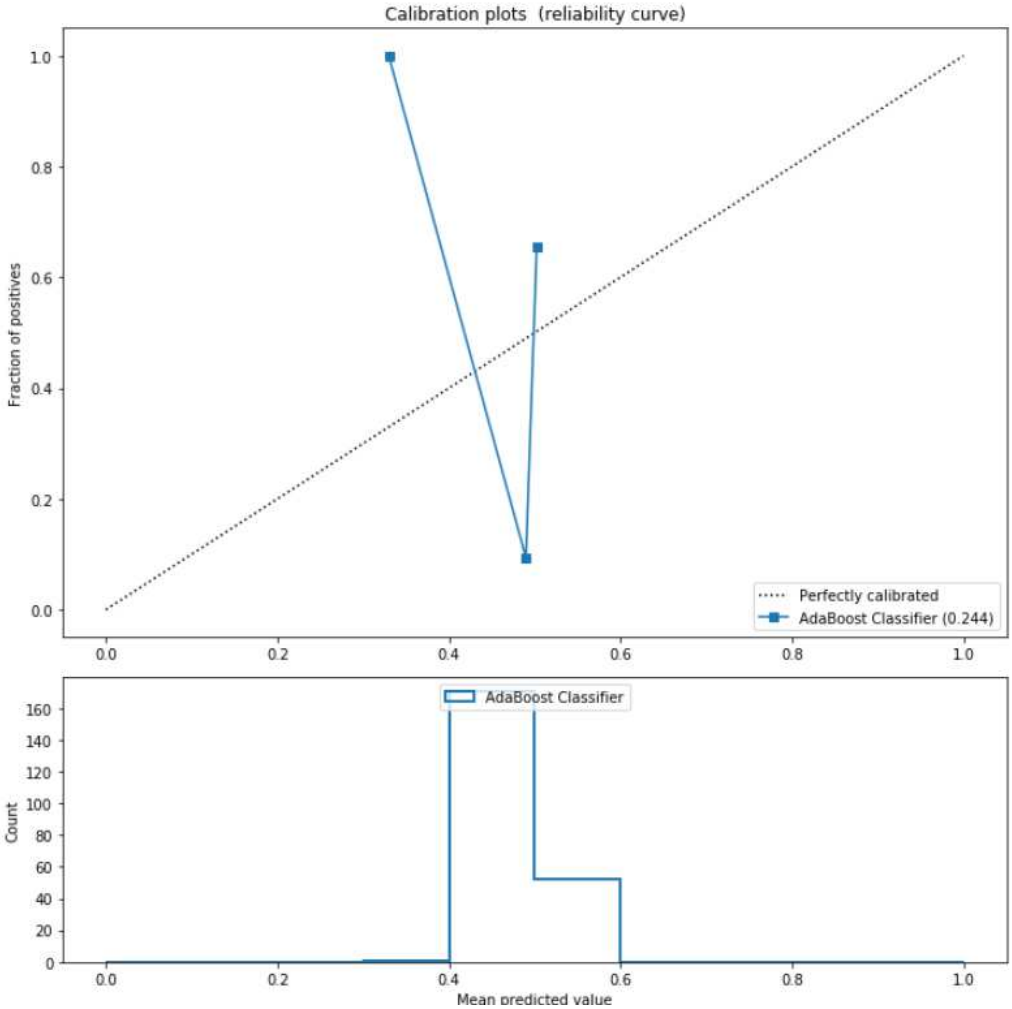
Figura 36 - Precision-Recall Curve do *Adaboost Classification* (ABC)



Fonte: Autor

O índice *Average precision* (AP) do *Adaboost Classification* (ABC) é de 72%. E na figura 37 é apresentado o gráfico de calibração das probabilidades.

Figura 37 - Gráfico de Calibragem do *Adaboost Classification* (ABC)



Fonte: Autor

O resumo das métricas de performance do *Adaboost Classification* (ABC) é apresentado na tabela 22.

Tabela 22 - Performance do *Adaboost Classification* (ABC)

Algoritmos	**Acuracia** (Accuracy)	**Precisão** (Precision)	**Sensibilidade** (Recall)	**Especificidade** (Specificity)	**ROC AUC**	**AP**
AdaBoost	84,4%	65,4%	66,7%	89,6%	85,7%	0,71

Fonte: Autor

Não foi identificado efeitos de *Overfitting* ou *Underfitting* no treinamento do *Adaboost Classification* (ABC), validando sua performance e habilitando-o para comparação de performance com os demais algoritmos utilizados nesse estudo.

4.6 MULTILAYER PERCEPTRON (MLP)

O último algoritmo testado no estudo foi a rede neural *Multilayer Perceptron* (MLP). Antes de iniciar o treinamento da rede neural, é necessário preparar o conjunto de dados (treinamento e validação), definir a arquitetura da rede e escolhe as funções de ativação, otimização e erro.

Na preparação do conjunto de dados de treinamento é importante recordar que a rede neural *Multilayer Perceptron* (MLP) é muito sensível a escala das variáveis do conjunto de dados, motivo que exigiu a normalização das informações, reduzindo ruídos na etapa de treinamento.

O segundo ponto de avaliação foi o desbalanceamento do conjunto de dados de treino. Como nos demais algoritmos, avaliou-se o uso de técnicas de sobre-amostragem (*over-sampling*) aleatória com substituição para equilibrar a quantidade de observações entre nossas duas classes estudadas (morte e sobrevida), mas observou-se ruído no treinamento e como resultado não obteve-se a mesma performance entre o conjunto de dados de treinamento e teste (validação), detalhes que serão explorados no decorrer desse capítulo.

Por fim, utilizou-se o método *K-Fold* de validação cruzada para minimizar o risco de *Overfitting. S*aber quando parar o processo de descida de gradiente é fundamental, os métodos de validação cruzada são grandes alinhados para estimar o ponto de parada e evitar o *Overfitting* (MITCHELL, 1997).

Após a preparação do conjunto de dados de treinamento e teste (validação), foi definido a arquitetura e funções da rede neural. A arquitetura de uma rede neural *Multilayer Perceptron* (MLP) é definida por uma camada de entrada, uma camada de saída e pelo menos uma camada intermediária (ou oculta) de neurônios. A quantidade de neurônios da camada de entrada foi de 86, igual à quantidade de variáveis do conjunto de dados estudado. Na camada de saída utilizou-se 1 neurônio para gerar o valor de probabilidade de morte. Em relação a camada oculta, não há regra, portanto, o número ideal de neurônios deve ser estimado para cada problema, guiada pelo conhecimento e pela experimentação (HASTIE; TIBSHIRANI; FRIEDMAN, 2009). Neste estudo, após vários experimentos, chegou-se a definição de 1 camada oculta com 21 neurônios.

Em resumo, utilizou-se uma rede neural de 3 camadas, sendo 1 camada de entrada com 86 neurônios, 1 camada intermediária (ou oculta) com 21 neurônios e 1 uma camada de saída com 1 neurônio. As camadas utilizadas foram do tipo densa (*dense*), ou seja, neurônios totalmente conectados.

Na definição das funções de ativação foi testada as funções não lineares Sigmoide, a Tangente Hiperbólica (TanH) e a ReLU (*Rectified Linear Unit*). Os testes foram realizados mantendo-se a arquitetura, alterando apenas a função de ativação. No estudo a função de ativação ReLU obteve melhor desempenho na previsão em relação as demais funções, confirmando o seu desempenho comparativo apresentado em outros estudos (GLOROT; BORDES; BENGIO, 2011). E por fim, utilizou-se a função de ativação Sigmoid na saída da rede, dessa forma o modelo gerou um valor entre 0 e 1, resultado que será utilizado para definir o ponto de corte da previsão de morte. Resumindo, a rede neural utilizou a função de ativação ReLU nas camadas de entrada e oculta e para a camada de saída utilizou a função Sigmoid.

Próximo passo foi selecionar a função de otimização da rede neural, responsável pela atualização dos pesos dos neurônios. Neste estudo foi testado as seguintes funções: *SGD* (descida de gradiente estocástico), *RMSprop* (método de otimização adaptativa da taxa de aprendizado proposto por Geoff Hinton) e *Adam* (Estimação de Momento Adaptativo). A função que gerou melhor previsão foi o otimizador *SGD.* Uma vez selecionada a função de otimização, definiu-se o peso inicial dos neurônios, inicializados com pequenos valores uniformemente aleatórios entre 0 e 0,05. Outro parâmetro necessário para a função de otimização é a taxa de aprendizagem, parâmetro que controla o valor que os pesos são atualizados. Utilizou-se o valor de 0,001 na taxa de aprendizagem, pequenos valores são recomendados para evitar o *Overfitting*.

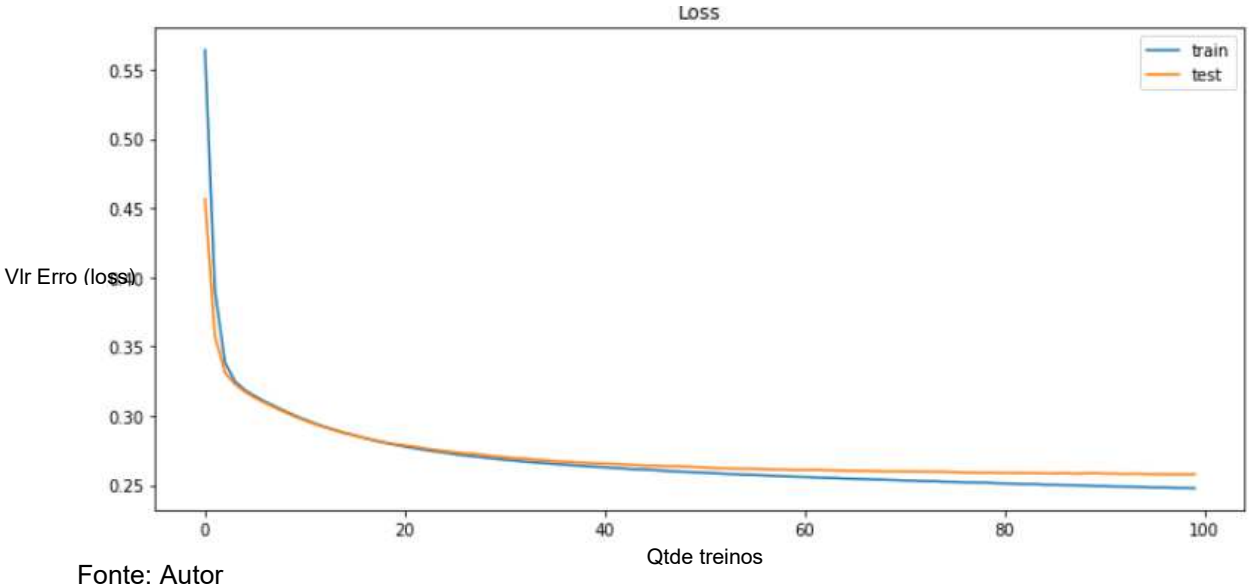
Na próxima etapa foi definido a função de correção de erro (ou perda), necessária para o processo de retropropagação (*backpropagation*), principal mecanismo de ajuste dos pesos de cada neurônio. Para problemas de classificação binária podemos utilizar a função de perda binária de entropia cruzada *(binary crossentropy*), função de perda de dobradiça (*Hinge Loss*) ou a função de perda de dobradiça ao quadrado (*Squared Hinge Loss).* Neste estudo utilizou-se a função de perda binária de entropia cruzada *(binary crossentropy*), uma vez que as funções de perda de dobradiça (*Hinge Loss*) e perda de dobradiça ao quadrado (*Squared Hinge Loss)* são utilizadas para previsão de valores entre -1 e 1, que não é o objetivo deste

estudo. Outro ponto importante para se destacar é que a função de perda binária de entropia cruzada *(binary crossentropy*) requer que a camada de saída seja configurada para um único neurônio e a função de ativação seja Sigmoid, pois assim obtém-se como resultado final a previsão de probabilidade de morte.

Com os conjuntos de dados de treino e teste (validação) ajustados, arquitetura e funções da rede neural definidas, inicia-se a etapa de treinamento da rede neural. No processo de treinamento são geradas as previsões e as mesmas são comparadas ao valor esperado e um erro é calculado. Esses valores de erros são retropropagados (*Backpropagation)* na rede, corrigindo os pesos dos neurônios. Esse processo pode ser realizado a cada amostra do conjunto de dados ou em lotes, ou seja, são calculados os erros de previsão de uma quantidade de amostra (lote) antes de executar a retropropagação (*Backpropagation)*. A definição do tamanho dos lotes é experimental e neste estudo foi definido lotes de 1 amostra, ou seja, houve ajuste dos pesos dos neurônios para cada linha do conjunto de dados. A quantidade de treinos também é definida experimentalmente, no estudo a quantidade de 100 treinos gerou a melhor previsão, evitando o *Overfitting* (MITCHELL, 1997).

Para analisar a performance do treinamento utilizou-se a análise gráfica das curvas de aprendizado apresentada na figura 38. As curvas de aprendizado auxiliam no diagnóstico de problemas de aprendizado, exemplo o *Underfitting* ou *Overfitting*, como também problemas na representatividade do conjunto de validação. No eixo horizontal (X) são apresentados a quantidade de treinos realizados e no eixo vertical (Y) temos o valor do erro (*loss*) medido naquele treino. Observa-se que no decorrer dos treinos o valor do erro (*loss*) reduz até que se estabilize. E são plotadas duas curvas, uma curva gerada a partir do conjunto de dados de treino (*train*) e a outra curva com as informações do conjunto de teste (*test*).

Figura 38 - Curva de Aprendizagem do *Multilayer Perceptron* (MLP)

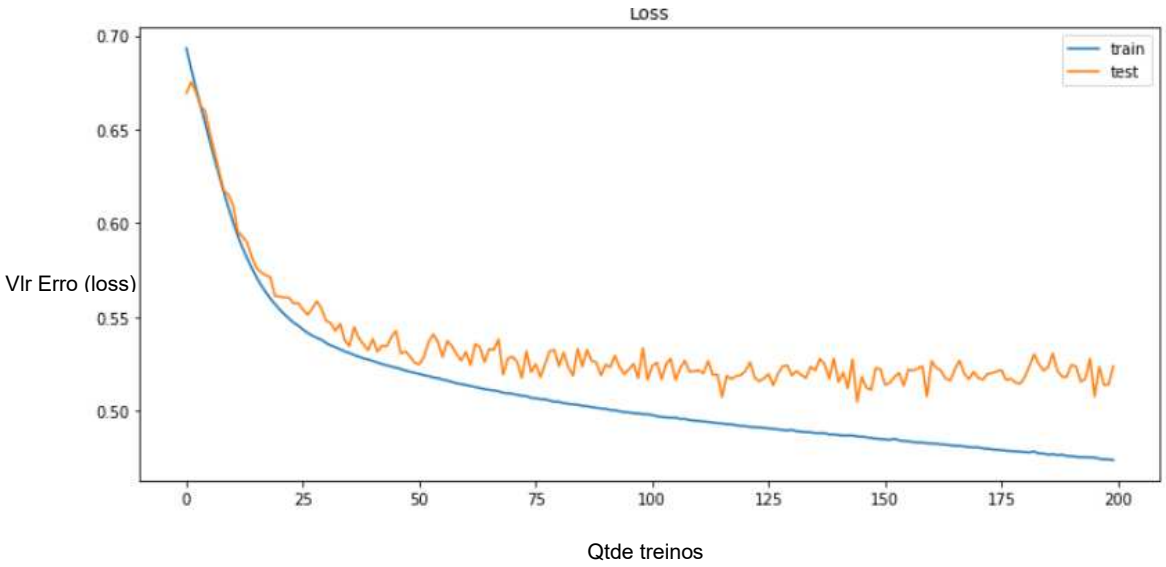


Fonte: Autor

No gráfico de aprendizado verifica-se que a curva *train* (conjunto de treino) e a curva *test* (conjunto de teste) possuem o mesmo comportamento e sua diferença de amplitude é mínima, confirmando que o algoritmo treinado pode ser generalizado. Outra informação relevante é o risco de *Underfitting* acima de 100 treino, região do gráfico que onde observa-se o início do distanciamento das curvas *train* e *test*.

E na figura 39 é apresentado uma curva de aprendizado de um conjunto de dados de treino balanceado com a técnica de sobre amostragem aleatória por substituição.

Figura 39 - Curva de aprendizado com conjunto de treino balanceado

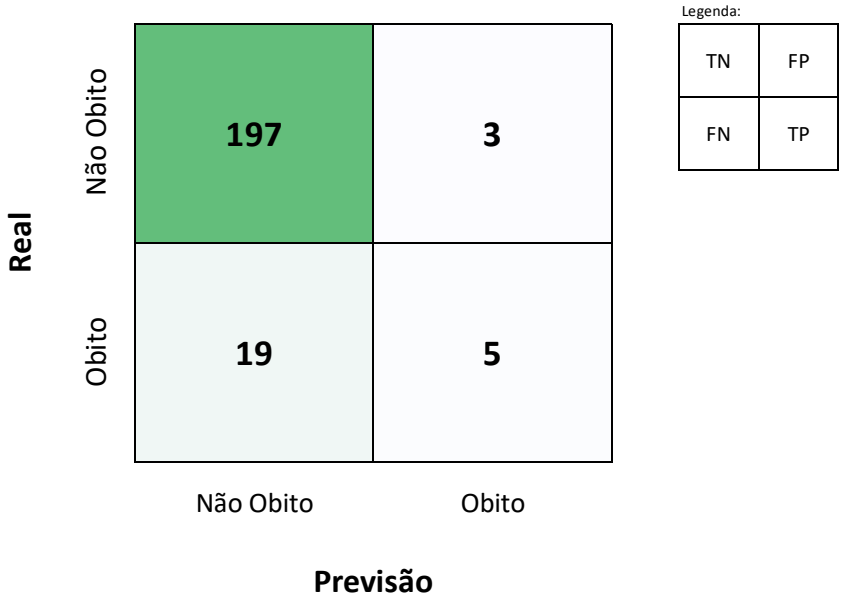


Fonte: Autor

Como foi relatado no início desse capítulo, no gráfico 39 observa-se que o balanceamento do conjunto de treino gerou instabilidade no treinamento na rede neural, confirmado com o comportamento da curva *test*.

Na figura 40 é apresentado a Matriz de Confusão (*confusion matrix*) do *Multilayer Perceptron* (MLP), primeira análise de performance do algoritmo.

Figura 40 - Confusion Matrix do *Multilayer Perceptron* (MLP)



Fonte: Autor

Na Matriz de Confusão (*confusion matrix*) da figura 39 observa-se que o modelo previu corretamente 5 mortes e 197 sobrevidas, mas apontou erroneamente 19 mortes como sobrevida, e 3 sobrevida como morte.

Na sequência é apresentado na tabela 23 o resultado de Acurácia (*Accuracy*), Sensibilidade (*Sensitivity or Recall*) e Precisão (*Precision*) do conjunto de treino e de teste (validação).

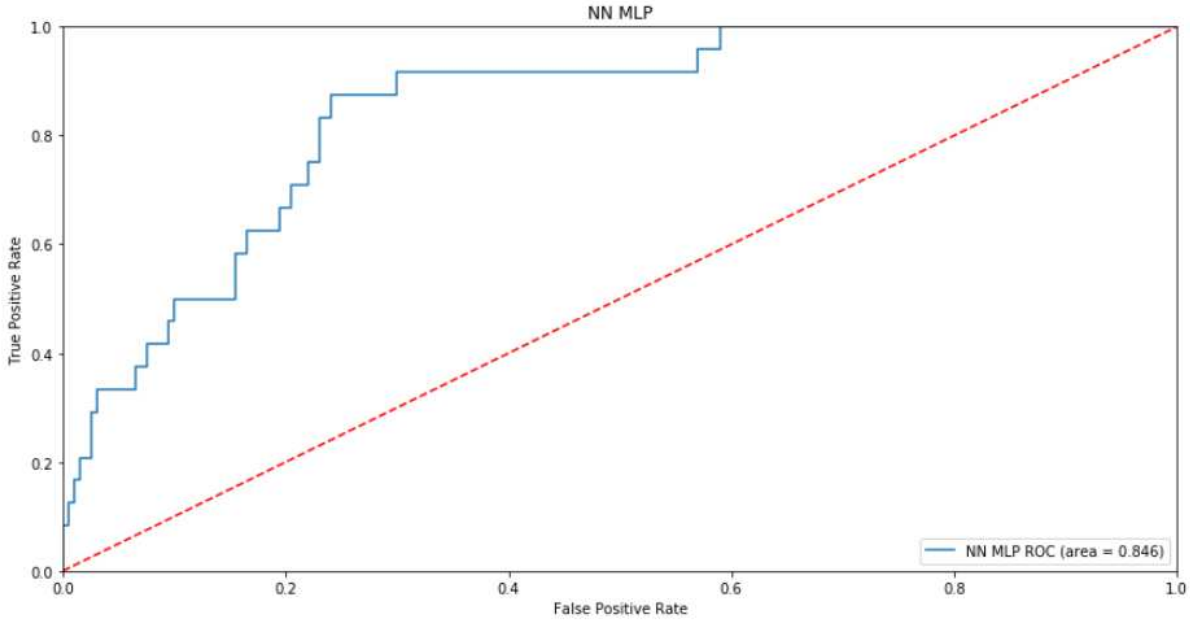
Tabela 23 - Métricas de performance do *Multilayer Perceptron* (MLP)

```
NN MLP
+-----------------+----------+
| Metric          |      Vlr |
+-----------------+----------+
| Accuracy        | 0.901786 |
| Recall          | 0.208333 |
| Precision       | 0.625000 |
| F1              | 0.312500 |
| AUC Probability | 0.845833 |
+-----------------+----------+
              precision    recall  f1-score   support

           0       0.91      0.98      0.95       200
           1       0.62      0.21      0.31        24

    accuracy                           0.90       224
   macro avg       0.77      0.60      0.63       224
weighted avg       0.88      0.90      0.88       224
```

Fonte: Autor

O modelo obteve uma acurácia de 90,1%, sensibilidade (*recall*) de 20,8%, precisão 62,5%. Na figura 41 é apresentado o gráfico ROC AUC (FAWCETT, 2004).

Figura 41 - Curva ROC AUC do *Multilayer Perceptron* (MLP)

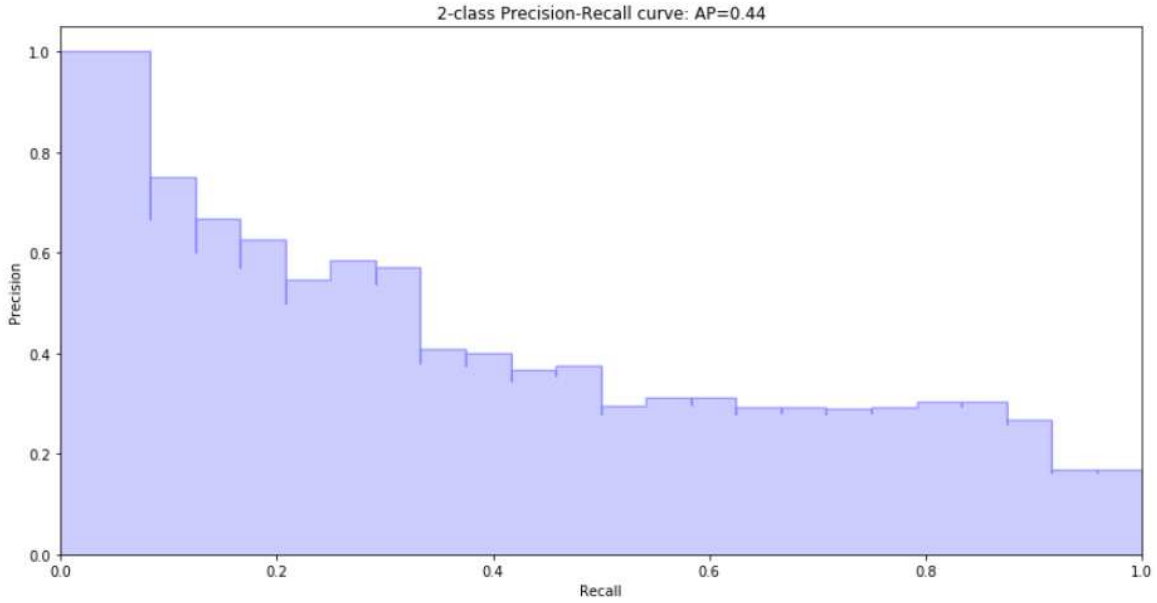


Fonte: Autor

A rede neural *Multilayer Perceptron* (MLP) obteve um ROC AUC de 84,6%, informação que será utilizada na comparação de performance entre os algoritmos

estudados e apoiar na definição do ponto de corte de previsão da morte. E no gráfico 42 é apresentado a Curva de Precisão-Sensibilidade *(Precision-Recall Curves).*

Figura 42 - Precision-Recall Curves do *Multilayer Perceptron* (MLP)



Fonte: Autor

O índice *Average precision* (AP) do *Multilayer Perceptron* (MLP) é de 0,44. O resumo das métricas de performance do *Multilayer Perceptron* (MLP) é apresentado na tabela 24.

Tabela 24 - Performance do *Multilayer Perceptron* (MLP)

Algoritmos	**Acuracia** (Accuracy)	**Precisão** (Precision)	**Sensibilidade** (Recall)	**Especificidade** (Specificity)	**ROC AUC**	**AP**
NN MLP	90,2%	62,5%	20,8%	98,5%	84,6%	0,44

Fonte: Autor

Em resumo, não se identificou efeitos de *Overfitting* ou *Underfitting* no treinamento da rede neural *Multilayer Perceptron* (MLP) validando sua performance e habilitando-o para comparação de performance com os demais algoritmos utilizados nesse estudo.

5 DISCUSSÃO E CONCLUSÃO

5.1. ANÁLISES

Nesse capítulo são discutidos os resultados dos algoritmos de aprendizagem de máquina testados neste estudo, como também as variáveis que impactam a previsão. Inicialmente são apresentadas na tabela 25 as métricas de performance de previsão de todos os algoritmos treinados neste estudo.

Tabela 25 – Tabela Consolidada de Resultados de Performance

Algoritmos	**Acuracia** (Accuracy)	**Precisão** (Precision)	**Sensibilidade** (Recall)	**Especificidade** (Specificity)	**ROC AUC**	**AP** (Average Precision)
Bagged Decision Trees	86,2%	69,2%	70,6%	90,8%	92,6%	0,81
Random Forest	80,8%	54,7%	92,2%	77,5%	90,2%	0,73
Extra Trees	72,8%	44,7%	82,4%	69,9%	86,0%	0,69
AdaBoost	84,4%	65,4%	66,7%	89,6%	85,7%	0,71
Stochastic Gradient Boosting	82,6%	59,7%	72,5%	85,5%	88,5%	0,70
NN MLP	90,2%	62,5%	20,8%	98,5%	84,6%	0,44

Fonte: Autor

Na tabela 25 observa-se que a rede neural *Multilayer Perceptron* (MLP) obteve os maiores índices de acurácia (*accuracy*) e especificidade (*specificity*) em relação aos demais algoritmos estudados, respectivamente 90,2% e 98,5%. Em contrapartida, obteve o menor índice de sensibilidade (*recall*) (20,8%), ROC AUC (84,6%) e AP (*average precision*) (0,44). Esses resultados demonstram que a rede neural *Multilayer Perceptron* (MLP) obteve melhor performance na acurácia das previsões de sobrevivência, fato reforçado com o índice de especificidade (*specificity*) de 98,5%, mas em contrapartida a sua capacidade de identificar pacientes com risco de morte é o mais baixo entre os modelos estudados, somente 20,8% do total de pacientes que morreram na cirurgia foram apontados como risco de morte pela rede neural.

De acordo com a literatura (SAITO; REHMSMEIER, 2015)(DAVIS; GOADRICH, 2006), o índice ROC AUC e AP (*average precision*) são bons parâmetros para análise de performance de modelos gerados a partir de conjuntos desbalanceados. Os algoritmos *Bagged Decision Trees* (BGT), *Random Forest* (RF) e *Stochastic Gradient Boosting* (SGB) se destacam com os maiores índices de ROC AUC entre os algoritmos estudados, respectivamente 92,6%, 90,2% e 88,5%. Também apresentam

os maiores índices de AP (*average precision*), 0,81 para o *Bagged Decision Trees* (BGT), 0,73 para o *Random Forest* (RF) e 0,70 para o *Stochastic Gradient Boosting* (SGB). Outro índice que dever ser considerado na tomada de decisão é o índice de sensibilidade (*recall*), onde observa-se o índice de 92,2% para o *Random Forest* (RF), maior índice entre os algoritmos estudados.

Alinhado ao objetivo de prever o risco de mortalidade pré cirúrgico, os algoritmos *Bagged Decision Trees* (BGT) e *Random Forest* (RF) se destacam no requisito performance. O algoritmo *Bagged Decision Trees* (BGT) demonstrou melhor performance para previsão da sobrevivência, especificidade (*specificity*) de 90,8%, sem abrir mão da capacidade de identificar risco de morte, índice de sensibilidade (recall) de 70,6%. Enquanto o *Random Forest* (RF) destacou-se na capacidade de identificar risco de morte, atingindo índice de sensibilidade (recall) de 92,2%.

Para suportar a discussão de como as variáveis impactam a previsão é apresentado o resultado da análise de importância dos algoritmos *Random Forest* (RF), *Stochastic Gradient Boosting* (SGB), *Extra Trees* (ET) e *Adaboost Classification* (ABC).

Tabela 26 - Tabela Consolidada da Análise do %importância

Tabela Consolidada - Análise de %importância						
			% importância			
Feature	**Variável Selecionada**	**Descrição**	Random Forest (RF)	Stochastic Gradient Boosting (SGB)	Extra Trees (ET)	AdaBoost Classification (ABC)
77	**PRE_UTI_Previa_2sim_1nao_adm1_81**	**Internação em UTI prévia**	**22,2%**	**11,4%**	**8,9%**	**5,3%**
59	**PRE_grupo_diag_61**	**Grupo diagnóstico**	**15,8%**	**2,2%**	**2,4%**	**5,3%**
79	**PRE_estat_cm_adm1_83**	**Altura do paciente na cirurgia**	**11,6%**	**17,7%**	**5,2%**	**9,3%**
47	**PRE_SHCE_49**	**Síndrome da Hipoplasia do Coração Esquerdo**	**6,5%**	**13,6%**	**2,6%**	**1,3%**
80	**PRE_IMC_adm1_84**	**Índice de Massa Corpórea na Cirurgia**	**6,3%**	**2,7%**	**3,5%**	**5,3%**
26	**PRE_sato2_preop_adm1_28**	**Saturação arterial de oxigênio**	**5,5%**	**10,0%**	**5,7%**	**5,3%**
81	PRE_SC_adm1_85	Superfície Corpórea na cirurgia	5,4%	0,3%	3,3%	0,0%
51	PRE_Atresia_pulmonar_53	Atresia Pulmonar	4,3%	3,1%	2,8%	1,3%
78	PRE_peso_kg_adm1_82	Peso do paciente na cirurgia	2,4%	0,0%	1,6%	6,7%
27	PRE_Diagnostico_categoria1a12_29	Categoria diagnóstico	1,6%	3,1%	2,6%	2,7%
29	PRE_CMP_31	Cardiomiopatia	1,3%	0,8%	1,6%	0,0%
54	PRE_Ventriculo_unico_56	Ventrículo Único	1,3%	0,0%	0,9%	0,0%
82	PRE_ANO_CC_ADM1_86	Ano da Cirurgia	1,1%	1,6%	4,4%	1,3%

25	PRE_ht_preop_adm1_27	Hematócrito pré operatório	0,9%	4,0%	4,1%	2,7%
7	PRE_intpre_9	Número de Internações	0,9%	0,7%	2,6%	5,3%
48	PRE_T4F_50	Tetralogia de Fallot	0,7%	0,1%	1,0%	2,7%
6	PRE_cirpre_8	Cirurgia prévia	0,6%	0,2%	1,2%	0,0%
9	PRE_sindcrom_11	Anormalidade Não Cardíaca - Síndrome cromossômica	0,5%	0,3%	1,4%	0,0%
5	PRE_maediab_7	Filho de mãe diabética	0,4%	0,3%	2,6%	0,0%
60	PRE_proc_previos_2Sim_1Nao_63	Submetido a procedimentos prévios	0,4%	0,3%	1,1%	1,3%
38	PRE_CIA_40	Comunicação Arterial	0,3%	0,0%	2,2%	0,0%
65	PRE_Vent_mec_Pre_Adm1_68	Em ventilação mecânica no pré-operatório	0,3%	1,4%	1,8%	1,3%
4	PRE_premat_6	Prematuridade	0,3%	0,0%	2,6%	0,0%
40	PRE_CIV_42	Comunicação Interventricular	0,3%	0,0%	2,1%	0,0%
61	PRE_proc_previos_01_64	Número de Procedimentos prévios	0,2%	1,5%	2,0%	6,7%
33	PRE_Doenca_ValvaAO_35	Valvopatias Aortica	0,2%	0,0%	0,0%	0,0%
73	PRE_Disf_renal_pre_Adm1_76	Disfunção renal no pré-operatório	0,1%	0,0%	0,0%	0,0%
10	PRE_down_12	Síndrome de Down	0,1%	0,0%	1,0%	1,3%
8	PRE_ncirpre_10	Número de cirurgias prévias	0,0%	0,7%	1,2%	1,3%
11	PRE_digeorge_13	Síndrome de DiGeorge	0,0%	0,2%	0,0%	0,0%
12	PRE_turnner_14	Síndrome de Turnner	0,0%	0,0%	0,0%	0,0%
13	PRE_willians_15	Síndrome de Willians	0,0%	0,0%	0,0%	0,0%
28	PRE_Truncus_30	Truncus Arteriosus Comum	0,0%	0,0%	0,0%	0,0%
32	PRE_Aneurisma_Ao_34	Aneurisma de Aorta	0,0%	0,0%	0,0%	0,0%
35	PRE_Doenca_ValvaTri_37	Valvopatia Tricuspide	0,0%	0,0%	0,0%	0,0%
50	PRE_Interrupcao_ArcoAO_52	Interrupção do Arco Aórtico	0,0%	0,0%	0,0%	1,3%
52	PRE_Doenca_ValvaPulm_54	Valvopatia Pulmonar	0,0%	0,6%	0,6%	1,3%
53	PRE_TGA_55	Transposição das Grandes Artérias	0,0%	0,0%	0,7%	0,0%
55	PRE_Eletrofisiologia_57	Alterações do Ritmo Cardíaco	0,0%	0,0%	0,0%	0,0%
56	PRE_Tunel_VE_AO_58	Tuneilização Ventrículo Esquerdo Aorta	0,0%	0,0%	0,0%	0,0%
57	PRE_Shone_59	Síndrome de Shone	0,0%	0,1%	0,0%	0,0%
62	PRE_Ressuc_PCR_Pre_Adm1_65	Ressuscitado de parada cardiorrespiratória no pré-operatório	0,0%	1,4%	1,1%	0,0%

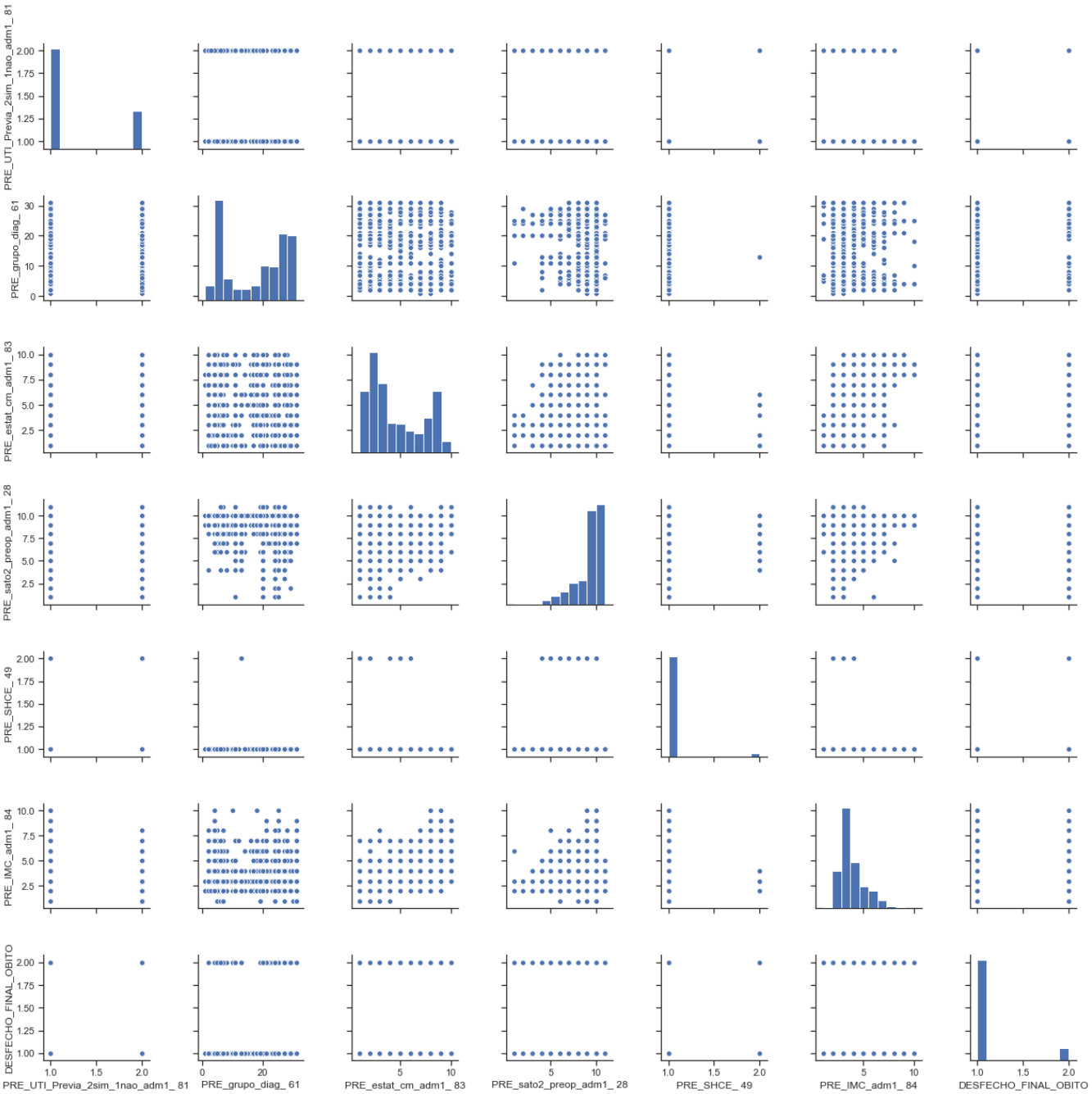
Fonte: Autor

Observa-se na tabela 26 que algumas variáveis são listadas como as mais importantes nos quatros algoritmos estudado, exemplo as variáveis PRE_UTI_Previa_2sim_1nao_adm1_81 (Internação em UTI prévia), PRE_grupo_diag_61 (Grupo diagnóstico), PRE_estat_cm_adm1_83 (Altura do paciente no momento da cirurgia), PRE_sato2_preop_adm1_28 (Saturação arterial de oxigênio), PRE_SHCE_49 (Síndrome da Hipoplasia do Coração Esquerdo) e PRE_IMC_adm1_84 (Índice de Massa Corpórea na Cirurgia). Estas variáveis somadas representam 67,8% do %importância do *Random Forest* (RF), 57,6% do

Stochastic Gradient Boosting (SGB), 28,4% do *Extra Trees* (ET) e 32,0% do *Adaboost Classification* (ABC).

Na figura 43 são apresentados gráficos de dispersão utilizando-se as seis variáveis com maior índice de importância para os algoritmos *Random Forest* (RF), *Stochastic Gradient Boosting* (SGB), *Extra Trees* (ET) e *Adaboost Classification* (ABC) e a variável resposta DESFECHO_FINAL_OBITO (Óbito).

Figura 43 - Gráficos de Dispersão das variáveis com maior % importância



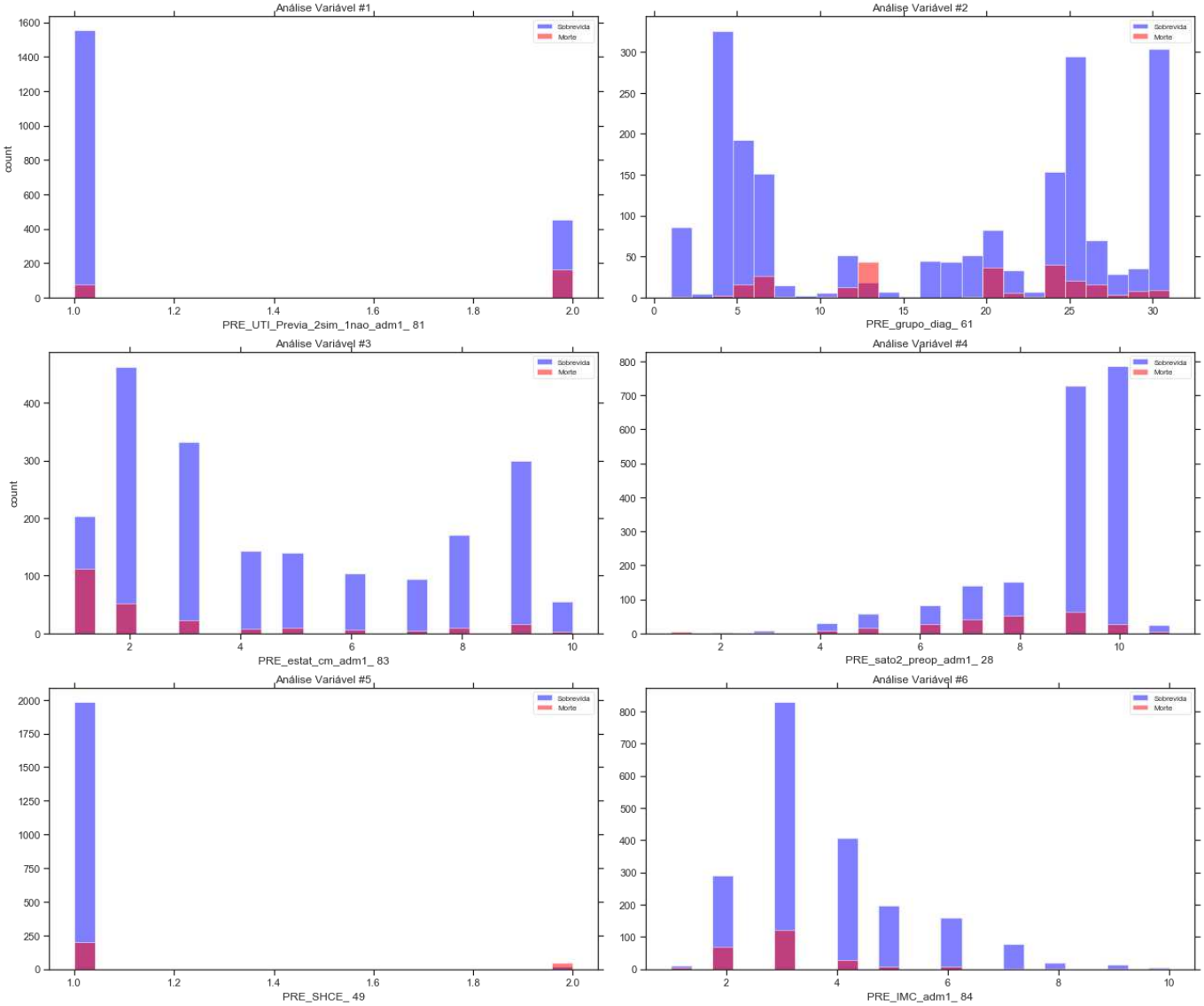
Fonte: Autor

Na análise dos gráficos da figura 43 não se observou correlação entre as variáveis com maior índice de importância para os algoritmos *Random Forest* (RF), *Stochastic Gradient Boosting* (SGB), *Extra Trees* (ET) e *Adaboost Classification* (ABC). E quando estas variáveis foram cruzadas com a variável resposta, DESFECHO_FINAL_OBITO (Óbito), a conclusão foi a mesma.

Na figura 43 são apresentados histogramas das seis variáveis com maior índice de importância para os algoritmos *Random Forest* (RF), *Stochastic Gradient Boosting* (SGB), *Extra Trees* (ET) e *Adaboost Classification* (ABC). São seis histogramas gerados a partir das variáveis PRE_grupo_diag_61 (Grupo diagnóstico), PRE_estat_cm_adm1_83 (Altura do paciente no momento da cirurgia), PRE_sato2_preop_adm1_28 (Saturação arterial de oxigênio), PRE_SHCE_49 (Síndrome da Hipoplasia do Coração Esquerdo), PRE_IMC_adm1_84 (Índice de Massa Corpórea na Cirurgia) e DESFECHO_FINAL_OBITO (Óbito). No eixo horizontal são plotados os índices da variável analisada e no eixo vertical a quantidade de observações.

Nos histogramas da figura 44 as colunas azuis correspondem a quantidade de observações de pacientes com o desfecho “sobrevida” e as colunas vermelhas a quantidade de observações com o desfecho “morte hospitalar”.

Figura 44 - Histograma das variáveis com % importância



Fonte: Autor

No histograma gerado a partir da variável PRE_grupo_diag_61 (Grupo diagnóstico) observa-se alta representatividade de “morte hospitalar” no índice 13, *Hypoplastic Left Heart Syndrome* (Síndrome da Hipoplasia do Coração Esquerdo), onde a quantidade de observações com “morte hospitalar” supera a quantidade de observações com “sobrevida”. No histograma da variável PRE_estat_cm_adm1_83 (Altura do paciente no momento da cirurgia) observa-se concentração de observações de “morte hospitalar” nas faixas 1 (até 51 cm) e 2 (de 52 a 66 cm), sendo que a representatividade de observações de “morte hospitalar” na faixa 1 (até 51 cm) é a maior entre as faixas analisadas. No histograma da variável PRE_UTI_Previa_2sim_1nao_adm1_81 (Internação em UTI prévia) observa-se maior

representatividade de “morte hospitalar” no índice 2, índice que sinaliza que o paciente foi internado na UTI antes da cirurgia. No histograma da variável PRE_sato2_preop_adm1_28 (Saturação arterial de oxigênio) observa-se que a representatividade de “morte hospitalar” reduz com aumento do índice. Para o histograma da variável PRE_SHCE_49 (Síndrome da Hipoplasia do Coração Esquerdo) observa-se que a existência dessa síndrome aumenta a representatividade de “morte hospitalar”, quase totalidade dos pacientes com essa síndrome tiveram o desfecho “morte hospitalar”. E por fim, no gráfico da variável PRE_IMC_adm1_84 (Índice de Massa Corpórea na Cirurgia) observa-se maior quantidade e representatividade de “morte hospitalar” nas menores faixas de IMC.

5.2. DISCUSSÃO

Um dos desafios da medicina é compreender melhor os riscos associados a uma intervenção cirúrgica, servindo de apoio aos profissionais da saúde nos preparativos pré cirúrgicos e ampliando a compreensão das variáveis que influenciam o risco de morte. No tratamento de pacientes com cardiopatia congênita são raras as pesquisas relacionadas ao risco de mortalidade pré cirurgia e as técnicas de aprendizagem de máquina *(machine learning)* estão cada vez mais presentes na área da medicina, substituindo os modelos tradicionais matemáticos e apoiando a tomada de decisão, fatos que direcionaram o objetivo deste estudo.

De posse das informações de cirurgias realizadas em pacientes com cardiopatia congênita extraídas do banco de dados ASSIST do InCor, foi possível desenvolver seis diferentes modelos de previsão de risco de mortalidade pré cirurgica e compreender quais variáveis impactam o risco de morte desses pacientes, permitindo que especialistas clínicos, pacientes e familiares compreendam melhor os riscos associados a uma possível intervenção cirúrgica cardíaca.

Dentre os seis algoritmos de aprendizado de máquina (*machine learning*) estudados, observa-se que os algoritmos *Bagged Decision Trees* (BGT) e *Random Forest* (RF) se destacaram na previsão da mortalidade em relação aos demais algoritmos estudados. Na decisão do algoritmo que cumprirá o objetivo proposto no estudo, foi considerado a complexidade de implementação de cada um deles e a possibilidade de compreensão das variáveis no risco de morte. Baseado nestes

parâmetros destaca-se o algoritmo *Random Forest* (RF), que possibilita a análise da importância das variáveis, função não presente no *Bagged Decision Trees* (BDT). Também observa-se que o algoritmo *Random Forest* (RF) não utilizou todas as variáveis do conjunto de dados na geração do modelo, reduzindo assim a complexidade de implementação. Por fim, pode-se destacar o *Random Forest* (RF) pela sua precisão, facilidade de treinamento e de ajustes (HASTIE; TIBSHIRANI; FRIEDMAN, 2009).

Diante destas análises, este estudo sugere o uso do *Random Forest* (RF) como modelo de previsão individual de morte para cirurgia cardíaca de pacientes com cardiopatia congênita. Os resultados de previsão do *Random Forest* (RF) corroboram que os algoritmos de aprendizagem de máquina (*machine learning*) podem auxiliar especialistas clínicos, pacientes e familiares na análise dos riscos associados a uma possível intervenção cirúrgica cardíaca.

Entender quais diagnósticos ou variáveis impactam a probabilidade de mortalidade de um paciente com cardiopatia congênita, quando submetido a uma intervenção cirúrgica cardíaca, permite que especialistas clínicos compreendam melhor os riscos associados a uma intervenção cirúrgica, fornecendo informações para apoio à decisão dos profissionais de saúde e familiares dos pacientes. Neste estudo observou-se que o diagnóstico da cardiopatia e outros dados, tais como: estatura do paciente (altura), índice de saturação arterial de oxigênio, existência da síndrome da hipoplasia do coração esquerdo, internação prévia na UTI e índice de massa corpórea (IMC), estão entre as seis variáveis mais importantes geradas na análise importância dos algoritmos *Random Forest* (RF), *Stochastic Gradient Boosting* (SGB), *Extra Trees* (ET) e *Adaboost Classification* (ABC). Estas variáveis somadas representam 67,8% de importância para o algoritmo *Random Forest* (RF), 57,6% para o *Stochastic Gradient Boosting (SGB)*, 28,4% para o *Extra Trees* (ET) e 32,0% para o *Adaboost Classification* (ABC).

Analisando as variáveis listadas como mais importantes, observou-se que a representatividade de “morte hospitalar” é maior nos pacientes com até 66 cm de altura e IMC abaixo de 13,0, e que a representatividade do percentual de “morte hospitalar” decai com o aumento do índice de saturação arterial de oxigênio, variáveis que podem ser foco de atuação e intervenção médica, a fim de amenizar o risco de morte.

Na internação prévia na UTI, observou-se que o percentual de mortes de pacientes com internação prévia antes da cirurgia é muito maior do que os pacientes que não necessitaram dessa intervenção. A análise aprofundada dos efeitos desta variável é oportuna na compreensão dos riscos de morte e pode ser alvo de estudos em pesquisas futuras.

Em relação ao grupo de diagnóstico, observa-se que algumas cardiopatias possuem maior percentual de morte do que outras, exemplo a síndrome da hipoplasia do coração esquerdo. Diante deste relato é oportuno direcionar estudos específicos para estes grupos e assim compreender quais variáveis possam direcionar ações mitigadoras do risco de morte.

Considerando os fatos acima discutidos, conclui-se que este estudo cumpriu seu objetivo de propor um modelo de previsão individual de morte para cirurgia cardíaca de pacientes com cardiopatia congênita e auxiliou na compreensão de quais diagnósticos ou variáveis impactam a probabilidade de mortalidade de portadores de cardiopatia congênita quando submetida a uma intervenção cirúrgica cardíaca. Desta forma, permite que especialistas clínicos, pacientes e familiares compreendam melhor os riscos associados a uma possível intervenção cirúrgica cardíaca.

Pesquisas futuras podem avaliar novos algoritmos de aprendizagem de máquina (*machine learning*) ou mesmo testar novas variáveis e diagnósticos, bem como a análise aprofundada dos efeitos dessas variáveis na compreensão dos riscos de morte.

REFERÊNCIAS

ABDOH, S. F.; ABO RIZKA, M.; MAGHRABY, F. A. Cervical cancer diagnosis using random forest classifier with SMOTE and feature reduction techniques. **IEEE Access**, v. 6, p. 59475–59485, 2018.

ALGHAMDI, M. et al. Predicting diabetes mellitus using SMOTE and ensemble machine learning approach: The Henry Ford ExercIse Testing (FIT) project. **Plos One**, v. 12, n. 7, p. e0179805, 2017.

AWAD, A. et al. Early hospital mortality prediction of intensive care unit patients using an ensemble learning approach. **International journal of medical informatics**, v. 108, n. October, p. 185–195, 2017.

AWAD, A.; BADER-EL-DEN, M.; MCNICHOLAS, J. Patient length of stay and mortality prediction: A survey. **Health Services Management Research**, v. 30, n. 2, p. 105–120, 2017.

BADER-EL-DEN, M.; TEITEI, E.; ADDA, M. Hierarchical classification for dealing with the Class imbalance problem. **Proceedings of the International Joint Conference on Neural Networks**, v. 2016- Octob, p. 3584–3591, 2016.

BENJAMIN, E. J. et al. **Heart Disease and Stroke Statistics'2017 Update: A Report from the American Heart AssociationCirculation**, 2017.

BERETTA, L. et al. Development of a five-year mortality model in systemic sclerosis patients by different analytical approaches. **Clinical and Experimental Rheumatology**, v. 28, n. 2 SUPPL. 58, 2010.

BERGSTRA, J.; BENGIO, Y. Random search for hyper-parameter optimization. **Journal of Machine Learning Research**, v. 13, p. 281–305, 2012.

BLAGUS, R.; LUSA, L. Joint use of over-and under-sampling techniques and cross-validation for the development and assessment of prediction models. **BMC Bioinformatics**, v. 16, n. 1, p. 1–10, 2015.

BREIMAN, L. Bagging Predictors. **Machine Learning**, v. 24, n. 421, p. 123–140, 1996.

BREIMAN, L. Random Forests. **Machine learning**, v. 45, p. 5–32, 2001.

BROWNE, M. W. Cross-Validation Methods. **Journal of Mathematical Psychology**, v. 44, n. 4, p. 108–132, 2001.

CHAWLA, N. V. et al. SMOTE: Synthetic Minority Over-sampling Technique. **Journal of Artificial Intelligence Research**, v. 16, n. 1, p. 321–357, 2002.

CHURPEK, M. M. et al. Multicenter Comparison of Machine Learning Methods and Conventional Regression for Predicting Clinical Deterioration on the Wards. **Crit Care Med**, v. 44, n. 2, p. 298–305, 2017.

COHEN, K. B. et al. Methodological Issues in Predicting Pediatric Epilepsy Surgery Candidates through Natural Language Processing and Machine Learning. **Biomedical Informatics Insights**, v. 8, p. BII.S38308, 2016.

DARST, B. F.; MALECKI, K. C.; ENGELMAN, C. D. Using recursive feature elimination in random forest to account for correlated variables in high dimensional data. **BMC Genetics**, v. 19, n. Suppl 1, p. 1–6, 2018.

DAVIS, J.; GOADRICH, M. **The Relationship Between Precision-Rcall and ROC Curves**. Madison: [s.n.].

DESAUTELS, T. et al. Prediction of early unplanned intensive care unit readmission in a UK tertiary care hospital: A cross-sectional machine learning approach. **BMJ Open**, v. 7, n. 9, p. 1–9, 2017.

DIETTERICH, T. G. Ensemble Methods in Machine Learning. **In Proc. 1st International Work- shop on Multiple Classifer Systems, LNCS**, v. Vol 1857, n. Springer-Verlag, p. 1–10, 2000.

EFRON, B. Bootstrap Methods: Another Look at the Jackknife. **The Annals of Statistics**, v. 7, n. 1, p. 1–26, 1979.

FAWCETT, T. ROC graphs: Notes and practical considerations for researchers. **Machine learning**, v. 31, n. 1, p. 1–38, 2004.

FREUND, Y.; SCHAPIRE, R. E. A decision-theoretic generalization of on-line learning and an application to boosting. **journal of computer and system sciences**, v. 55, p. 119–139, 1997.

FREUND, Y.; SCHAPIRE, R. E. A Short Introduction to Boosting. **Journal of Japanese Society for Artificial Intelligence**, v. 14, n. 5, p. 771–780, 1999.

FREUND, Y.; SCHAPIRE, R. E.; HILL, M. **Experiments with a New Boosting Algorithm**. [s.l: s.n.].

FRIEDMAN, J. H. Stochastic Gradient Boosting. **Computational statistics & data analysis**, v. 1, n. 3, p. 1–10, 1999.

GLOROT, X.; BORDES, A.; BENGIO, Y. **Deep Sparse Rectifier Neural Networks**. Proceedings of the fourteenth international conference on artificial intelligence and statistics. **Anais**...Montreal, Canada: 2011

GUYON, I. et al. Gene selection for cancer classification using Support Vector Machines. **Machine Learning**, v. 46, p. 389–422, 2002.

HAJIAN-TILAKI, K. Receiver operating characteristic (ROC) curve analysis for medical diagnostic test evaluation. **Caspian Journal of Internal Medicine**, v. 4, n. 2, p. 627–635, 2013.

HASTIE, T.; TIBSHIRANI, R.; FRIEDMAN, J. **The Elements of Statistical Learning**. Second Edi ed. [s.l.] Springer Series in Statistics, 2009.

HAWKINS, D. M. The Problem of Overfitting. **J. Chem. Inf. Comput. Sci.**, v. 44, p. 1–12, 2004.

HAYKIN, S. **Redes neurais: princípios e prática**. 2. ed. ed. Porto Alegre: Bookman, 2001.

HÖRER, J. et al. Mortality Following Congenital Heart Surgery in Adults Can Be Predicted Accurately by Combining Expert-Based and Evidence-Based Pediatric Risk Scores. **World Journal for Pediatric and Congenital Heart Surgery**, 2016.

JENKINS, K. J. et al. Consensus-based method for risk adjustment for surgery for congenital heart disease. **Journal of Thoracic and Cardiovascular Surgery**, 2002.

JENKINS, K. J. Risk adjustment for congenital heart surgery: The RACHS-1 method. **Pediatric Cardiac Surgery Annual**, v. 7, n. 1, p. 180–184, 2004.

JENKINS, K. J.; GAUVREAU, K. Center-specific differences in mortality: Preliminary analyses using the risk adjustment in congenital heart surgery (RACHS-1) method. **Journal of Thoracic and Cardiovascular Surgery**, v. 124, n. 1, p. 97–104, 2002.

KOHAVI, R. A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection. **International Joint Conference On Artificial Intelligence**, n. 0, p. 0–6, 1995.

KOTSIANTIS, S. B.; TSEKOURAS, G. E.; PINTELAS, P. E. Bagging model trees for classification problems. **Lecture Notes in Computer Science**, v. 3746 LNCS, n. January, p. 328–337, 2005.

LACOUR-GAYET, F. et al. The Aristotle score: A complexity-adjusted method to evaluate surgical results. **European Journal of Cardio-thoracic Surgery**, v. 25, n. 6, p. 911–924, 2004.

LI, H. et al. An artificial neural network prediction model of congenital heart disease based on risk factors A hospital-based case-control study. **Medicine (United States)**, 2017.

LIN, K.; HU, Y.; KONG, G. Predicting in-hospital mortality of patients with acute kidney injury in the ICU using random forest model. **International Journal of Medical Informatics**, v. 125, n. 38, p. 55–61, 2019.

LINDE, D. VAN DER et al. Birth Prevalence of Congenital Heart Disease Worldwide. **Israel Medical Association Journal**, v. Vol. 58, N, 2011.

MEJIA, O. A. V. et al. Predictive performance of six mortality risk scores and the development of a novel model in a prospective cohort of patients undergoing valve surgery secondary to rheumatic fever. **Journal Plos One**, p. 1–14, 2018.

MITCHELL, T. M. **Machine Learning**. [s.l.] McGraw-Hill Science/Engineering/Math, 1997.

PIRRACCHIO, R. et al. Mortality prediction in intensive care units with the Super ICU Learner Algorithm (SICULA): A population-based study. **The Lancet Respiratory Medicine**, v. 3, n. 1, p. 42–52, 2015.

PROVOST, F.; FAWCETT, T. Robust classification for imprecise environments. **Machine Learning**, v. 42, n. 3, p. 203–231, 2001.

QUINLAN, J. R. **Bagging, Boosting, and C4.5University of Sydney**, 2006.

RAMACHANDRAN, P.; ZOPH, B.; LE, Q. V. **Searching for Activation Functions**. [s.l: s.n.]. Disponível em: <http://arxiv.org/abs/1710.05941>.

RUIZ-FERNÁNDEZ, D. et al. Aid decision algorithms to estimate the risk in congenital heart surgery. **Computer Methods and Programs in Biomedicine**, 2016.

SAITO, T.; REHMSMEIER, M. The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. **PLoS ONE**, v. 10, n. 3, 2015.

SCHAPIRE, R. E. Explaining AdaBoost. In: **Empirical Inference**. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013. p. 37–52.

SILVA, I. N. DA; SPATTI, D. H.; FLAUZINO, R. A. **Redes neurais artificiais: para engenharia e ciências aplicadadas**. 2.ed. ed. São Paulo: Artliber, 2010.

TENNANT, P. W. G. et al. 20-year survival of children born with congenital anomalies: a population-based study. **Lancet 2010**, 2010.

VILELA DE ABREU HAICKEL NINA, R. et al. O escore de risco ajustado para cirurgia em cardiopatias congênitas (RACHS-1) pode ser aplicado em nosso meio? **Rev Bras Cir Cardiovasc**, v. 22, n. 4, p. 425–431, 2007.

VIRANI, S. S. et al. **Heart Disease and Stroke Statistics—2020 UpdateCirculation**. [s.l: s.n.]. Disponível em: <https://www.ahajournals.org/doi/10.1161/CIR.0000000000000757>.

XU, J. et al. **Data Mining on ICU Mortality Prediction Using Early Temporal Data: A Survey**. [s.l: s.n.]. v. 16

ZIMMERMAN, M. S. et al. Global, regional, and national burden of congenital heart disease, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. **The Lancet Child & Adolescent Health**, v. 4642, n. 19, 2020.

APÊNDICE A – VARIÁVEIS CONJUNTO DADOS INCOR

Feature	Variável	Descrição	Tipo Variável	Campo	Descrição campo
0	PRE_proc_2	Procedência (moradia do paciente)	qualitativas categóricas	1 - 28	1 - Ignorada 2 - Pará 3 - São Paulo - estado 4 - São Paulo - capital 5 - Espírito Santo 6 - Goiás 7 - Minas Gerais 8 - Rio Grande do Norte 9 - Tocantins 10 - Paraná 11 - Bahia 12 - Roraima 13 - Brasília 14 - Mato Grosso do Sul 15 - Pernambuco 16 - Paraíba 17 - Ceará 18 - Santa Catarina 19 - Amazonas 20 - Piauí 21 - Mato Grosso 22 - Rio de Janeiro 23 - Maranhão 24 - Rio Grande do Sul 25 - Amapá 26 - Alagoas 27 - Rondônia 28 - Acre
1	PRE_prov_3	Provedor	qualitativas categóricas	1 - 4	1 = SUS 2 = Particular 3 = Convênio 4 = Ignorado
2	PRE_sexo_4	Sexo	qualitativas categóricas	1 - 2	1 - Masculino 2 - Feminino
3	PRE_prenat_5	Diagnóstico da cardiopatia congênita no pré-natal	qualitativas categóricas	1 - 3	1 = não 2 = sim 3 = s/info
4	PRE_premat_6	Prematuridade	qualitativas categóricas	1 - 3	1 = não 2 = sim 3 = s/info
5	PRE_maediab_7	Filho de mãe diabética	qualitativas categóricas	1 - 3	1 = não 2 = sim 3 = s/info
6	PRE_cirpre_8	Cirurgia prévia	qualitativas categóricas	1 - 2	1 = não 2 = sim
7	PRE_intpre_9	Número de Internações	quantitativas continuas	1 - 12	1 a 12
8	PRE_ncirpre_10	Número de cirurgias prévias	quantitativas continuas	1 - 7	1 a 7
9	PRE_sindcrom_11	Anormalidade Não Cardíaca - Síndrome cromossômica	qualitativas categóricas	1 - 2	1 = não 2 = sim
10	PRE_down_12	Síndrome de Down	qualitativas categóricas	1 - 2	1 = não 2 = sim
11	PRE_digeorge_13	Síndrome de DiGeorge	qualitativas categóricas	1 - 2	1 = não 2 = sim
12	PRE_turnner_14	Síndrome de Turnner	qualitativas categóricas	1 - 2	1 = não 2 = sim
13	PRE_willians_15	Síndrome de Willians	qualitativas categóricas	1 - 2	1 = não 2 = sim
14	PRE_edwards_16	Síndrome de Edwards	qualitativas categóricas	1 - 2	1 = não 2 = sim
15	PRE_Noonan_17	Síndrome de Noonan	qualitativas categóricas	1 - 2	1 = não 2 = sim
16	PRE_outcrom_18	Outras Síndromes	qualitativas categóricas	1 - 2	1 = não 2 = sim

17	PRE_anoanat_19	Anormalidade Não Cardíaca - Malformações	qualitativas categóricas	1 - 2	1 = não 2 = sim
18	PRE_atresiaesofag_20	Atresia de Esôfago	qualitativas categóricas	1 - 2	1 = não 2 = sim
19	PRE_anusimperf_21	Anus Imperfurado	qualitativas categóricas	1 - 2	1 = não 2 = sim
20	PRE_Fistraqueo_22	Fistulatraqueoesofágica	qualitativas categóricas	1 - 2	1 = não 2 = sim
21	PRE_herniadiaf_23	Hérnia Diafragmática	qualitativas categóricas	1 - 2	1 = não 2 = sim
22	PRE_onfalocele_24	Onfalocele	qualitativas categóricas	1 - 2	1 = não 2 = sim
23	PRE_fendaPalat_25	Fenda Palatina	qualitativas categóricas	1 - 2	1 = não 2 = sim
24	PRE_Outanoanat_26	Outras A	qualitativas categóricas	1 - 2	1 = não 2 = sim
25	PRE_ht_preop_adm1_27	Hematócrito pré operatório	qualitativas categóricas	1 - 9	1. 21 a 26 2. 27 a 32 3. 33 a 38 4. 39 a 44 5. 45 a 50 6. 51 a 56 7. 57 a 62 8. 63 a 70 9 . s/info
26	PRE_sato2_preop_adm1_28	Saturação arterial de oxigênio	qualitativas categóricas	1 - 10	1. 55 a 60% 2. 61 a 64% 3. 65 a 67% 4. 70 a 73% 5. 74 a 78% 6. 79 a 82% 7. 83 a 87% 8. 88 a 92% 9. 93 a 96% 10. 97 a 100% 11. S/info
27	PRE_Diagnostico_categoria1a12_29	Categoria diagnóstico	qualitativas categóricas	1 - 12	1 = cardiomiopatia 2 = cor triatriatum 3 = DORV - double outlet right ventricle 4 = electrophysiological 5 = left heart lesions 6 = miscellaneous, other 7 = pulmonary venous anomalies 8 = right heart lesions 9 = septal defects 10 = single ventricle 11 = thoracic arteries and veins 12 = transportation of the great arteries
28	PRE_Truncus_30	Truncus Arteriosus Comum	qualitativas categóricas	1 - 2	1 = não 2 = sim
29	PRE_CMP_31	Cardiomiopatia	qualitativas categóricas	1 - 2	1 = não 2 = sim
30	PRE_DAPVVPP_32	Drenagem Anômala Parcial das Veias Pulmonares	qualitativas categóricas	1 - 2	1 = não 2 = sim
31	PRE_DATVVPP_33	Drenagem Anômala Total das Veias Pulmonares	qualitativas categóricas	1 - 2	1 = não 2 = sim
32	PRE_Aneurisma_Ao_34	Aneurisma de Aorta	qualitativas categóricas	1 - 2	1 = não 2 = sim
33	PRE_Doenca_ValvaAO_35	Valvopatias Aortica	qualitativas categóricas	1 - 2	1 = não 2 = sim
34	PRE_Doenca_ValvaMi_36	Valvopatia Mitral	qualitativas categóricas	1 - 2	1 = não 2 = sim
35	PRE_Doenca_ValvaTri_37	Valvopatia Tricuspide	qualitativas categóricas	1 - 2	1 = não 2 = sim

36	PRE_Janela_AO_Pulm_38	Janela Aortopulmonar	qualitativas categóricas	1 - 2	1 = não 2 = sim
37	PRE_PCA_39	Persistência do Canal Arterial	qualitativas categóricas	1 - 2	1 = não 2 = sim
38	PRE_CIA_40	Comunicação Arterial	qualitativas categóricas	1 - 2	1 = não 2 = sim
39	PRE_DSAV_41	Defeito do Septo Atrioventricular	qualitativas categóricas	1 - 2	1 = não 2 = sim
40	PRE_CIV_42	Comunicação Interventricular	qualitativas categóricas	1 - 2	1 = não 2 = sim
41	PRE_CoAo_HipoArcoAO_43	Coarctação e Hipoplasia do Arco Aórtico	qualitativas categóricas	1 - 2	1 = não 2 = sim
42	PRE_Miscelania_44	Miscelânia	qualitativas categóricas	1 - 2	1 = não 2 = sim
43	PRE_TCGA_45	Transposição Corrigida das Grandes Artérias	qualitativas categóricas	1 - 2	1 = não 2 = sim
44	PRE_CorTriatriatum_46	Cor Triatriatum	qualitativas categóricas	1 - 2	1 = não 2 = sim
45	PRE_Anomalia_Coronaria_47	Anomalia de Coronaria	qualitativas categóricas	1 - 2	1 = não 2 = sim
46	PRE_DVSVD_48	Dupla Via de Saída de Ventrículo Direito	qualitativas categóricas	1 - 2	1 = não 2 = sim
47	PRE_SHCE_49	Síndrome da Hipoplasia do Coração Esquerdo	qualitativas categóricas	1 - 2	1 = não 2 = sim
48	PRE_T4F_50	Tetralogia de Fallot	qualitativas categóricas	1 - 2	1 = não 2 = sim
49	PRE_RVOT_51	Ampliação da Via de Saída do Ventrículo Direito	qualitativas categóricas	1 - 2	1 = não 2 = sim
50	PRE_Interrupcao_ArcoAO_52	Interrupção do Arco Aórtico	qualitativas categóricas	1 - 2	1 = não 2 = sim
51	PRE_Atresia_pulmonar_53	Atresia Pulmonar	qualitativas categóricas	1 - 2	1 = não 2 = sim
52	PRE_Doenca_ValvaPulm_54	Valvopatia Pulmonar	qualitativas categóricas	1 - 2	1 = não 2 = sim
53	PRE_TGA_55	Transposição das Grandes Artérias	qualitativas categóricas	1 - 2	1 = não 2 = sim
54	PRE_Ventriculo_unico_56	Ventrículo Único	qualitativas categóricas	1 - 2	1 = não 2 = sim
55	PRE_Eletrofisiologia_57	Alterações do Ritmo Cardíaco	qualitativas categóricas	1 - 2	1 = não 2 = sim
56	PRE_Tunel_VE_AO_58	Tuneilização Ventrículo Esquerdo Aorta	qualitativas categóricas	1 - 2	1 = não 2 = sim
57	PRE_Shone_59	Síndrome de Shone	qualitativas categóricas	1 - 2	1 = não 2 = sim
58	PRE_Anel_Vascular_60	Anel Vascular	qualitativas categóricas	1 - 2	1 = não 2 = sim

59	PRE_grupo_diag_61	Grupo diagnóstico	qualitativas categóricas	1 - 31	1 - Aortic Aneurysm 2 - Aortic Valve Disease 3 - AP Window 4 - ASD 5 - AV Canal 6 - Cardiomyopathy 7 - Coarctation of Aorta and Aortic Archhypoplasia 8 - Congenitally Correted TGA 9 - Cor Triatriatum 10 - Coronary Artery Anomalies 11 - DORV 12 - Electrophysiological 13 - Hypoplastic Left Heart Syndrome 14 - Interrupted Arch 15 - LV to Aorta Tunnel 16 - Miscellaneous 17 - Mitral Valve Disease 18 - Partial Anomalous Pulmonary Venous Connection 19 - Patent Ductus Arteriosus 20 - Pulmonary Atresia 21 - Pulmonary Valve Disease 22 - RVOT Obstruction and/or Pulmonary Stenosis 23 - Shone's Syndrome 24 - Single Ventricle 25 - Tetralogy of Fallot 26 - Total Anomalous Pulmonary Venous Connection 27 - Transposition of the Great Arteries 28 - Tricuspid Valve Disease and Ebstein Anormaly 29 - Truncus Arteriosus 30 - Vascular Rings and Slings 31 - VSD
60	PRE_proc_previos_2Sim_1Nao_63	Submetido a procedimentos prévios	qualitativas categóricas	1 - 2	1 = não 2 = sim
61	PRE_proc_previos_01_64	Número de Procedimentos prévios	qualitativas categóricas	1 - 85	1 a 85
62	PRE_Ressuc_PCR_Pre_Adm1_65	Foi ressucitado de parada cardiorrespiratória no pré-operatório	qualitativas categóricas	1 - 2	1 = não 2 = sim
63	PRE_Arritmia_pre_Adm1_66	Arritmia no pré-operatório	qualitativas categóricas	1 - 2	1 = não 2 = sim
64	PRE_Inotropicos_pre_Adm1_67	Em uso de inotrópico no pré-operatório	qualitativas categóricas	1 - 2	1 = não 2 = sim
65	PRE_Vent_mec_Pre_Adm1_68	Em ventilação mecânica no pré-operatório	qualitativas categóricas	1 - 2	1 = não 2 = sim
66	PRE_Traqueo_pre_Adm1_69	Traqueostomia presente no pré-operatório	qualitativas categóricas	1 - 2	1 = não 2 = sim
67	PRE_Hipotiroidismo_pre_Adm1_70	Hipotireoidismo no pré-operatório	qualitativas categóricas	1 - 2	1 = não 2 = sim
68	PRE_Diabetes_Pre_Adm1_71	Diabetes no pré-operatório	qualitativas categóricas	1 - 2	1 = não 2 = sim
69	PRE_Endocardite_pre_Adm1_72	Diagnóstico de Endocardite no pré-operatório	qualitativas categóricas	1 - 2	1 = não 2 = sim
70	PRE_Sepsis_pre_Adm1_73	Sepsis no pré-operatório	qualitativas categóricas	1 - 2	1 = não 2 = sim
71	PRE_Convulsao_pre_Adm1_74	Convulsão no pré-operatório	qualitativas categóricas	1 - 2	1 = não 2 = sim
72	PRE_Alt_Neuro_pre_adm1_75	Alterações Neurológicas no pré-operatório	qualitativas categóricas	1 - 2	1 = não 2 = sim
73	PRE_Disf_renal_pre_Adm1_76	Disfunção renal no pré-operatório	qualitativas categóricas	1 - 3	1 = não 2 = sim 3 = s/info

74	PRE_Hipert_Pulm_pre_Adm1_77	Hipertensão Pulmonar no pré-operatório	qualitativas categóricas	1 - 3	1 = não 2 = sim 3 = s/info
75	PRE_Gastro_pre_Adm1_79	Gastrostomia presente no pré	qualitativas categóricas	1 - 3	1 = não 2 = sim 3 = s/info
76	PRE_ECMO_VAD_pre_80	Necessidade ECMO no pré operatório	qualitativas categóricas	1 - 3	1 = não 2 = sim 3 = s/info
77	PRE_UTI_Previa_2sim_1nao_adm1_81	Internação em UTI prévia	qualitativas categóricas	1 - 2	1 = não 2 = sim
78	PRE_peso_kg_adm1_82	Peso do paciente na cirurgia	qualitativas categóricas	1 - 10	1 = de 1,7 a 13,1 Kg 2 = de 13,2 a 24,6 Kg 3 = de 24,7 a 36,1 Kg 4 = de 36,2 a 47,6 Kg 5 = de 47,7 a 59,1 Kg 6 = de 59,2 a 70,6 Kg 7 = de 70,7 a 82,1 Kg 8 = de 82,2 a 93,6 Kg 9 = de 93,7 a 105,1 Kg 10 = acima de 105,1 Kg
79	PRE_estat_cm_adm1_83	Altura do paciente na cirurgia	qualitativas categóricas	1 - 9	1 - até 51 cm 2 - de 52 a 66 cm 3 - de 67 a 82 cm 4 - de 83 a 97 cm 5 - de 98 a 112 cm 6 - de 113 a 128 cm 7 - de 129 a 143 cm 8 - de 144 a 159 cm 9 - de 160 a 174 cm 10 - acima de 175 cm
80	PRE_IMC_adm1_84	Índice de Massa Corpórea na Cirurgia	qualitativas categóricas	1 - 10	1 - de 6,3 a 9,6 2 - de 9,7 a 13,0 3 - de 13,1 a 16,4 4 - de 16,5 a 19,8 5 - de 19,9 a 23,2 6 - de 23,3 a 26,6 7 - de 26,7 a 30,0 8 - de 30,1 a 33,4 9 - de 33,5 a 36,8 10 - de 36,9 a 39,1
81	PRE_SC_adm1_85	Superfície Corpórea na cirurgia	qualitativas categóricas	1 - 8	1 = 0,2 a 0,4 m2 2 = 0,5 a 0,7 m2 3 = 0,8 a 1,0 m2 4 = 1,1 a 1,3 m2 5 = 1,4 a 1,6 m2 6 = 1,7 a 1,9 m2 7 = 2,0 a 2,3 m2 8 = acima de 2,3 m2
82	PRE_ANO_CC_ADM1_86	Ano da Cirurgia	qualitativas categóricas	1 - 5	1 = 2014 2 = 2015 3 = 2016 4 = 2017 5 = 2018
83	PRE_IDADE_CIRURGIA_dias_adm1_90	Idade na data da cirurgia	quantitativas continuas	1 - 10	1 - de 0 a 2432 dias 2 - de 2433 a 4864 dias 3 - de 4864 a 7295 dias 4 - de 7296 a 9727 dias 5 - de 9728 a 12159 dias 6 - de 12160 a 14590 dias 7 - de 14591 a 17022 dias 8 - de 17023 a 19454 dias 9 - de 19455 a 21885 dias 10 - de 21886 a 24316 dias
84	DESFECHO_FINAL_OBITO	Óbito	qualitativas categóricas	1 - 2	1 = Alta hospitalar 2 = Óbito hospitalar