

FUNDAÇÃO GETULIO VARGAS  
ESCOLA DE ADMINISTRAÇÃO DE EMPRESAS DE SÃO PAULO

MARCOS MORAES

**TOMADA DE DECISÃO NA PRIORIZAÇÃO DE PACIENTES EM FILA DE  
ESPERA CIRÚRGICA BASEADA EM APRENDIZADO DE MÁQUINA**

SÃO PAULO  
2021

MARCOS MORAES

**TOMADA DE DECISÃO NA PRIORIZAÇÃO DE PACIENTES EM FILA DE  
ESPERA CIRÚRGICA BASEADA EM APRENDIZADO DE MÁQUINA**

Trabalho Aplicado apresentado à Escola de Administração de Empresas de São Paulo da Fundação Getulio Vargas, como requisito para obtenção do título de Mestre em Gestão para a Competitividade.

Linha de Pesquisa: Gestão em Saúde

Orientador: Prof. Dr. Gonzalo Vecina Neto

Coorientador: Prof. Dr. Marco Antonio Gutierrez

SÃO PAULO  
2021

Moraes, Marcos.

Tomada de decisão na priorização de pacientes em fila de espera cirúrgica baseada em aprendizado de máquina / Marcos Moraes. - 2021.

102 f.

Orientador: Gonzalo Vecina Neto.

Co-orientador: Marco Antonio Gutierrez.

Dissertação (mestrado profissional MPGC) – Fundação Getulio Vargas, Escola de Administração de Empresas de São Paulo.

1. Aprendizado do computador. 2. Processo decisório. 3. Hospitais públicos - Listas de espera. 4. Cirurgia. 5. Teoria das filas. I. Vecina Neto, Gonzalo. II. Gutierrez, Marco Antonio. III. Dissertação (mestrado profissional MPGC) – Escola de Administração de Empresas de São Paulo. IV. Fundação Getulio Vargas. V. Título.

CDU 614.2

**MARCOS MORAES**

**TOMADA DE DECISÃO NA PRIORIZAÇÃO DE PACIENTES EM FILA DE  
ESPERA CIRÚRGICA BASEADA EM APRENDIZADO DE MÁQUINA**

Trabalho Aplicado apresentado à Escola de  
Administração de Empresas de São Paulo da  
Fundação Getulio Vargas, como requisito para  
obtenção do título de Mestre em Gestão para a  
Competitividade.

Linha de Pesquisa: Gestão em Saúde

Orientador: Prof. Dr. Gonzalo Vecina Neto

Coorientador: Prof. Dr. Marco Antonio Gutierrez

Data da Aprovação

\_\_\_\_/\_\_\_\_/\_\_\_\_

Banca Examinadora:

---

Prof. Dr. Gonzalo Vecina Neto  
FGV – EAESP

---

Prof. Dr. Fernando Lopes Alberto  
FGV – EAESP

---

Prof. Dr. Marco Antonio Gutierrez  
INCOR – HCFMUSP

---

Prof<sup>a</sup>. Dra. Jeane Mike Tsutsui  
Grupo Fleury

## **AGRADECIMENTOS**

Esta foi a última parte a ser escrita. Não foi por acaso. Queria ter passado por todo esse processo para, então, poder olhar para trás e refletir sobre ele com um pouco mais de clareza.

Meu pai e minha mãe, Joel e Gislaine, são pessoas que negaram as estatísticas. Fossem elas seguidas, seguramente, nem eles, nem eu teríamos feito ensino superior. Na infância, meu pai vendia mel, morava de favor, sem pai, nem mãe; minha mãe, na adolescência, ajudava como doméstica; formou-se no ensino superior com quase 50. Para mim, ter chegado até aqui, ter me formado médico pela Universidade de São Paulo e mestre pela Fundação Getulio Vargas concretiza uma vitória, deles e minha. Pai e mãe, muito obrigado. Tenho muito orgulho da jornada de vocês. Sei o tanto que abdicaram e trabalharam duro para que eu estivesse aqui. Obrigado também aos meus irmãos, Joel e Thiago, e à minha irmã, Caroline.

A maior parte deste trabalho foi escrita à noite e madrugada adentro. Em algumas dessas horas, tive o privilégio de ter a Malu dormindo em meus braços, momento em que aproveitava para ler. Maria e Malu, minhas filhas amadas, agradeço todos os dias pelo simples fato de vocês existirem. Que este trabalho possa frutificar e, de alguma forma, contribuir para a construção de um país melhor para todos nós. O mesmo vale para você Sofia, que é fundamental na minha vida.

Nesse caminho todo, tive o apoio dela, minha esposa e namorada, promotora das minhas melhores risadas e fortaleza das minhas piores piadas. Nath, agradecer apenas pelo suporte ao mestrado seria muito pouco; obrigado por trazer as melhores reflexões sobre a vida à mesa. Te Amo. A vida é muito mais alegre ao seu lado.

A vida foi muito generosa comigo, especialmente, em relação às pessoas com quem tive a oportunidade de conviver. Professores, Pablo Maria Alberto Pomerantzeff, Flávio Tarasoutchi e Wilson Mathias Júnior, este trabalho ficou muito melhor graças às observações cuidadosas e perspicazes que vocês fizeram. Muito obrigado pelo tempo que dedicaram às nossas discussões. Esses votos estendo também ao Sr. José Antonio Ramos Neto e equipe UIMH, especialmente André e Igor, por toda a ajuda que prestaram nas inúmeras extrações de dados que fizemos.

Tive um orientador e um coorientador, e isso foi ímpar. Prof. Gonzalo, você é uma pessoa para com quem terei uma eterna dívida intelectual. Do nosso jeito,

chegamos e concluímos essa jornada. Muito obrigado. Ter convivido com você no InCor e neste mestrado foi realmente especial.

Prof. Marco Gutierrez, obrigado por ter descoberto uma maneira de encaixar esta dissertação no dia a dia corrido do nosso hospital, no meio de uma pandemia. Reconheço e agradeço seu esforço. Todas nossas reuniões de *checkpoint*. Muito obrigado. Obrigado também por ter introduzido a Karina Zaccari Gagliotti neste projeto, nossa cientista de dados e com quem aprendi muito. Karina, meu eterno agradecimento a você.

Aos amigos e às amigas do mestrado. Esses dois anos foram transformadores para mim, e, seguramente, vocês são peças fundamentais disso. Conviver com vocês foi algo excepcional. Vida longa e sucesso a nós.

Finalmente, agradeço ao InCor, por ter possibilitado essa experiência. Tayar, obrigado por ter fomentado essa jornada. Guilherme do Espírito Santo Silva, você foi uma grande referência na decisão de fazer o mestrado, e à Prof<sup>a</sup>. Jeane e ao Prof. Fernando Alberto por terem aceitado participar da banca e terem contribuído para o trabalho. Muito obrigado!

O otimista é um tolo. O pessimista, um chato.  
Bom mesmo é ser um realista esperançoso.  
**Ariano Suassuna**

## RESUMO

Lidar com a presença de filas de espera para procedimentos cirúrgicos eletivos apresenta-se como uma realidade para os países com sistemas de saúde de acesso universal. O progresso tecnológico, o envelhecimento populacional e indivíduos cada vez mais conscientes da necessidade de cuidados são alguns dos fatores que pressionam essa demanda, não correspondida pela oferta finita de serviços de saúde. Nesse cenário, encontrar uma maneira que promova a melhor alocação possível dos recursos existentes e garanta a equidade no acesso é imperativo. Esse estudo apresenta modelos de aprendizado de máquina (*machine learning*) capazes de discriminar pacientes em fila de espera para cirurgia valvar que serão atendidos como urgentes caso não sejam priorizados enquanto aguardam pelo procedimento. A partir dos dados de 588 pacientes submetidos à cirurgia valvar no INCOR-HCFMUSP entre janeiro de 2010 e dezembro de 2019 e considerando 44 variáveis independentes (demográficas, laboratoriais, ecocardiográficas, medicamentos em uso, diagnósticos de comorbidades e de lesão anatômica valvar), foram construídos e testados modelos de regressão logística, árvore de decisão, florestas aleatórias (*random forests*) e K-vizinhos mais próximos que pudessem diferenciar pacientes eletivos dos de urgência. A amostra balanceada foi submetida ao pré-processamento de dados e utilizou-se a validação cruzada como técnica de reamostragem no treinamento dos algoritmos. Os modelos de florestas aleatórias e regressão logística foram os que apresentaram melhor desempenho, com sensibilidade, especificidade, acurácia, precisão e AUROC de 67,3%, 60,9%, 64,1%, 63,3%, 0,64 e 66,0%, 63,3%, 64,6%, 64,2% e 0,65, respectivamente. Esses resultados apontam para uma ferramenta de priorização de pacientes objetiva, em oposição ao método tradicional de priorização, manual e baseado em critérios pessoais, realizado pelo médico e cercado de imperfeições como a diferença de percepção intra e interobservadores, fadiga, disponibilidade de tempo e outros vieses associados à tomada de decisão cirúrgica. Com isso, concorre-se para uma solução mais inteligente, justa, ágil, escalável e equânime no momento de definir quem deve ser atendido primeiro e quem pode esperar.



**Palavras-chave:** Aprendizado de máquina. Tomada de decisão cirúrgica. Filas de espera. Priorização.

## ABSTRACT

Dealing with waiting lists for elective surgical procedures is a reality for countries with universal access health systems. Technological progress, population aging and individuals increasingly aware of the need for care are some of the factors that pressure this demand, which is not met by the finite supply of health services. In this scenario, it is essential to find a way to promote the best possible allocation of existing resources and ensure equity in access. This study presents machine learning models capable of discriminating patients on the waiting list for valve surgery who will be treated as urgent if they are not prioritized while waiting for the procedure. Based on data from 588 patients, derived from 44 independent variables (demographic, laboratory, echocardiographic, medication in use, diagnosis of comorbidities and anatomical valve lesion), who underwent valve surgery at INCOR-HCFMUSP between January 2010 and December 2019, algorithms of logistic regression, decision tree, random forests and K-nearest neighbors were constructed and tested, with the aim to differentiate elective from emergency patients. The balanced sample was submitted to data pre-processing and cross-validation was used as a resampling technique in algorithm training. Random forests and logistic regression showed the best performance, with sensitivity, specificity, accuracy, precision and AUROC of 67,3%, 60,9%, 64,1%, 63,3%, 0,64 and 66,0%, 63,3%, 64,6%, 64,2% and 0,65, respectively. These results point to an objective and transparent patient prioritization tool, as opposed to the traditional method of prioritization, manual and personalized, carried out by the physician and surrounded by imperfections such as the difference in intra- and inter-observer perception, fatigue, time availability and other biases associated with surgical decision-making. This contributes to a smarter, fairer, more equitable and scalable solution when defining who should be attended first and who can wait.

**Keywords:** Machine learning. Surgical decision making. Waiting lists. Prioritization.

## LISTA DE FIGURAS E ILUSTRAÇÕES

Figura 1 – Volume de dados criado, capturado, copiado e consumido ao redor do mundo entre 2010 e 2024. * Valor estimado.....	27
Figura 2 – Resumo das técnicas de inteligência artificial .....	31
Figura 3 – Abordagens ótima e subótima para tomada de decisão cirúrgica .....	32
Figura 4 – Desafios relacionados à adoção da inteligência artificial na saúde .....	37
Figura 5 – Fluxograma para definição da amostra de treinamento e avaliação dos modelos. ....	56
Figura 6 – Representação de uma árvore de decisão.....	59
Figura 7 – Algoritmo de Florestas Aleatórias. ....	60
Figura 8 – Algoritmo KNN. ....	61
Figura 9 – Divisão de dados e validação cruzada k-fold do estudo.....	63
Figura 10 – Estrutura de uma matriz de confusão utilizada para avaliar o desempenho de algoritmos de classificação.. ....	67
Figura 11 – Matrizes de confusão e respectivas curvas ROC para os algoritmos selecionados, utilizando-se a variável reoperação .....	78
Figura 12 – Matrizes de confusão e respectivas curvas ROC para os algoritmos selecionados, sem a variável reoperação .....	79

## LISTA DE TABELAS

Tabela 1 – Passo a passo do modelo de triagem AMIB/ABRAMEDE .....	25
Tabela 2 – Tipos de vieses na tomada de decisão cirúrgica .....	27
Tabela 3 – Variáveis independentes inicialmente consideradas para predição.....	42
Tabela 4 – Variáveis dependentes e independentes que compuseram os modelos de aprendizado de máquina .....	48
Tabela 5 – Distribuição das variáveis dependentes nas bases de treinamento, validação e teste utilizadas no processo de aprendizagem e avaliação de desempenho dos quatro algoritmos do estudo.....	60
Tabela 6 – Acurácia da classificação do tipo de cirurgia .....	69
Tabela 7 – Acurácia do registro da variável reoperação.....	65
Tabela 8 – Significância das variáveis identificadas na regressão logística do modelo I.....	66
Tabela 9 – Variáveis e significância identificadas na regressão logística da segunda rodada, aplicada sobre as 19 principais variáveis do modelo I.....	67
Tabela 10 – Resumo do desempenho dos quatro algoritmos nos modelos I (44 variáveis preditoras) e II (4 variáveis preditoras).....	73
Tabela 11 – Variáveis preditoras da árvore de decisão (modelo I).....	68
Tabela 12 – Variáveis preditoras das florestas aleatórias (modelo I) .....	74
Tabela 13 – Variáveis e significância identificadas na regressão logística ao se excluir a variável reoperação do modelo I.....	68
Tabela 14 – Resumo do desempenho dos quatro algoritmos nos modelos I-B (43 variáveis) e II-B (3 variáveis).....	76
Tabela 15 – Importância das variáveis no modelo de árvore de decisão sem a presença da variável reoperação (modelo I-B).....	69
Tabela 16 – Importância das variáveis no algoritmo de florestas aleatórias sem a presença da variável reoperação (modelo I-B).....	70

## LISTA DE ACRÔNIMOS E SIGLAS

ABRAMEDE – Associação Brasileira de Medicina de Emergência

ACC – Acurácia

AUROC – *Area Under the Receiver Operating Characteristic*

AM – Aprendizado de Máquina

AMIB – Associação de Medicina Intensiva Brasileira

ANCP – Academia Nacional de Cuidados Paliativos

CFM – Conselho Federal de Medicina

CID-10 – Classificação Estatística Internacional de Doenças e Problemas Relacionados à Saúde – Décima Revisão

FN – Falso Negativo

FP – Falso Positivo

FPP – Ferramentas de Priorização de Pacientes

IA – Inteligência Artificial

INCOR-HCFMUSP – Instituto do Coração do Hospital das Clínicas da Faculdade de Medicina da Universidade de São Paulo

IMC – Índice de Massa Corporal

INR – *International Normalized Ratio*

IPCA – Índice de Preços ao Consumidor Amplo

KNN – *K-Nearest Neighbors*

OCDE – Organização para Cooperação e Desenvolvimento Econômico

RES – Registros Eletrônicos de Saúde

ROC – *Receiver Operating Characteristic*

SI<sup>3</sup> – Sistema Integrado de Informações Institucionais

SBGG – Sociedade Brasileira de Geriatria e Gerontologia

SUS – Sistema Único de Saúde

TFN – Taxa de Falso Negativo

TFP – Taxa de Falso Positivo

TVN – Taxa de Verdadeiro Negativo

TVP – Taxa de Verdadeiro Positivo

UTI – Unidade de Terapia Intensiva

VN – Verdadeiro Negativo

VP – Verdadeiro Positivo

VPP – Valor Preditivo Positivo

WCWL – *Western Canada Waiting List Project*

## Sumário

<b>1. INTRODUÇÃO .....</b>	<b>14</b>
<b>2. REFERENCIAL TEÓRICO .....</b>	<b>17</b>
2.1 FILAS DE ESPERA E PRIORIZAÇÃO .....	17
2.2 <i>BIG DATA</i> E APRENDIZADO DE MÁQUINA .....	27
<b>3. OBJETIVOS .....</b>	<b>38</b>
3.1 OBJETIVO GERAL .....	38
3.2 OBJETIVOS ESPECÍFICOS .....	38
<b>4. MÉTODO.....</b>	<b>39</b>
4.1 CONSIDERAÇÕES GERAIS.....	39
4.2 LOCAL DA PESQUISA .....	39
4.3 CARDIOPATIAS VALVARES E FILA CIRÚRGICA NO INCOR-HCFMUSP .....	40
4.4 POPULAÇÃO E VARIÁVEIS.....	41
4.5 CRITÉRIOS DE INCLUSÃO.....	44
4.6 CRITÉRIOS DE EXCLUSÃO.....	44
4.7 PRÉ-PROCESSAMENTO DE DADOS.....	44
4.8 ALGORITMOS .....	56
4.8.1 REGRESSÃO LOGÍSTICA.....	57
4.8.2 ÁRVORES DE DECISÃO .....	58
4.8.3 FLORESTAS ALEATÓRIAS.....	59
4.8.4 K-VIZINHOS MAIS PRÓXIMOS .....	60
4.9 APRENDIZADO E ANÁLISE DE DESEMPENHO DOS MODELOS .....	62
<b>5. RESULTADOS .....</b>	<b>69</b>
5.1 VALIDAÇÃO DAS VARIÁVEIS.....	69
5.2 DESEMPENHO DOS MODELOS DE APRENDIZADO DE MÁQUINA.....	70
<b>6. DISCUSSÃO .....</b>	<b>81</b>
<b>7. CONCLUSÃO .....</b>	<b>89</b>
<b>ANEXO A – PARECER CONSUBSTANCIADO DO COMITÊ DE ÉTICA EM PESQUISA..</b>	<b>97</b>
<b>APÊNDICE I – CÁLCULO AMOSTRAL .....</b>	<b>100</b>
<b>APÊNDICE II – COMPARAÇÃO DE DESEMPENHO ENTRE OS ALGORITMOS DE REGRESSÃO LOGÍSTICA E FLORESTAS ALEATÓRIAS .....</b>	<b>101</b>

## 1. INTRODUÇÃO

A implementação de sistemas de saúde de acesso universal determinou, no Brasil e no mundo, inúmeros avanços em relação ao acesso e à utilização dos serviços de saúde. Como consequência, os resultados de saúde das populações assistidas melhoraram. No País, desde a operacionalização do Sistema Único de Saúde (SUS), em 1990, pela Lei nº 8.080 (BRASIL, 1990), que definiu seus atributos e seu funcionamento, houve aumento expressivo do acesso à água, ao saneamento básico e aos serviços de saúde. A reorganização dos serviços de saúde proposta pelo SUS, que passaram a ser regionalizados e hierarquizados, com a Estratégia Saúde da Família (Programa Saúde da Família, anteriormente) como porta de entrada do cidadão e atuando como coordenadora dos fluxos de pessoas e de informações, levou ao aumento expressivo da cobertura da população, especialmente em relação à atenção básica (VICTORA *et al.*, 2011). A expectativa de vida aumentou e houve declínio expressivo da mortalidade infantil. O acesso ao cuidado médico e bucal aumentou para toda a população. (CASTRO *et al.*, 2019). A mortalidade por doenças evitáveis caiu. (MALTA *et al.*, 2018).

Que pesem todos esses ganhos, alguns desafios surgiram e outros se intensificaram, como a presença de filas de espera para consultas de especialidades, exames e procedimentos cirúrgicos eletivos, problema compartilhado por diversos sistemas de saúde de acesso universal e entre muitos países-membros da Organização para Cooperação e Desenvolvimento Econômico (OCDE), especialmente naqueles cujo financiamento para tais procedimentos é público e a taxa de coparticipação dos usuários é zero ou muito baixa em relação ao pagamento; nestes casos, as barreiras financeiras ao acesso são eliminadas e o tempo de espera passa a figurar como um dos principais balanceadores entre a oferta e a demanda de procedimentos eletivos. (IVERSEN, 1993; SICILIANI; HURST, 2005; DÉRY *et al.*, 2020).

Conforme Siciliani, Moran e Borowitz (2014) apontam, nos primeiros anos da década 2000, diversos países europeus experimentaram queda proeminente nos tempos de espera para cirurgias eletivas, porém atingiram um platô e, em alguns casos, voltaram a apresentar aumento, como ocorreu com Portugal, Reino Unido, Espanha e Finlândia. Os autores apontam como causa, além da recessão



econômica instalada por volta de 2010, com diminuição na alocação de recursos para as políticas de controle dos tempos de espera, o aumento da demanda por procedimentos, que tem suplantado a elevação da oferta, fenômeno que se intensifica ano após ano e que está relacionado ao envelhecimento populacional e ao progresso tecnológico, que coloca à disposição da população novas possibilidades terapêuticas.

Hurst e Siciliani (2003) chamam a atenção para os avanços nas tecnologias cirúrgicas e anestésicas como motor para o aumento da demanda por cirurgias eletivas, visto que ampliam o alcance, a segurança e a eficácia dos procedimentos cirúrgicos e, por conseguinte, expandem a gama de condições de saúde tratáveis e as expectativas dos pacientes. No Brasil, além do aumento da disponibilidade de recursos tecnológicos e do aumento da expectativa de vida da população, a transição epidemiológica, indivíduos cada vez mais conscientes da necessidade de cuidados e o acesso facilitado a serviços e tratamentos também pressionam essa demanda crescente. (PEDROSO, MALIK, 2015).

Se, por um lado, a perspectiva é de ampliação da demanda, por outro, não há sinais de que a oferta, pelo menos de serviços públicos de saúde, aumentará nos próximos anos no Brasil, especialmente no contexto de austeridade fiscal instituído pela Emenda Constitucional nº 95 (BRASIL, 2016) que limita o gasto público federal anual, por 20 anos, ao valor do exercício anterior corrigido pela inflação, tendo como base o Índice de Preços ao Consumidor Amplo (IPCA). Por mais que essa restrição não seja direcionada especificamente para a saúde, isto é, trata da limitação do gasto público federal como um todo, alguns autores não veem motivos para desconsiderar o impacto dessa Emenda nesse setor, uma vez que, para aumentar o gasto público federal na saúde, seria necessário reduzir o gasto público em outras áreas, como educação, assistência social ou segurança pública. Para Rossi e Dweck (2016) e Castro *et al.* (2019), esse novo regime fiscal se apresenta como um risco real à expansão e à sustentabilidade do SUS, com grandes chances de promover desassistência e piorar os resultados de saúde do País.

Para além da austeridade fiscal no Brasil, a literatura indica que filas de espera parecem ser, pelo menos em sistemas universais de acesso à saúde, inevitáveis. Dado esse cenário de demanda maior que a oferta, alguns autores defendem que o foco deve então ser mudado: ao invés de mirar apenas a redução

dos tempos de espera para procedimentos, propõem que sejam também desenvolvidos sistemas mais inteligentes de priorização, isto é, capazes de diferenciar os casos mais graves, urgentes e com maior benefício esperado pelo tratamento, daqueles em que o paciente pode aguardar por mais tempo e suportar os custos da espera prolongada ou ainda daqueles em que o tratamento em si trará muito pouco benefício ao paciente. (GUTACKER, SICILIANI, COOKSON, 2016; SICILIANI, MORAN, BOROWITZ, 2014a).

As políticas e os métodos de priorização podem ser entendidos como uma estratégia de gerenciamento das filas de espera cujo intuito é promover a equidade e a alocação mais justa do recurso disponível, dado um cenário de restrição. É um meio de maximizar o bem-estar social. Quem pode esperar mais, espera mais; os mais necessitados e que terão maiores benefícios pelo uso do serviço passam na frente; pessoas com necessidades semelhantes aguardam tempos semelhantes. (DÉRY *et al.*, 2020; HADORN, 2000; TESTI *et al.*, 2008).

Embora a lógica de equidade e justiça na alocação de recursos soe ético e moral, implementar sistemas de priorização no dia a dia de médicos e gestores em saúde ainda esbarra numa série de desafios. Mcleod *et al.* (2004), em estudo realizado com médicos e gestores da Nova Zelândia, identificaram que apenas 19,5% dos cirurgiões concordavam que as ferramentas de priorização existentes na prática clínica, com critérios transparentes e objetivos, eram eficazes; para a maior parte deles, o julgamento clínico ainda era a melhor maneira de se fazer isso.

O contraponto a essa maneira individual e artesanal de priorização encontra respaldo na literatura, que mostra que o processo de decisão cirúrgica é repleto de vieses e influenciado por fatores, como cansaço e disponibilidade de tempo. (LOFTUS *et al.*, 2019). Nesse sentido, encontrar mecanismos de priorização com menor carga subjetiva, mais transparentes e que possam ser replicáveis, independente de dia e hora, de maneira exaustiva, mas sem comprometer seu desempenho, é auspicioso. Em partes, a inteligência artificial (IA) vem para endereçar várias dessas imperfeições.

## 2. REFERENCIAL TEÓRICO

### 2.1 FILAS DE ESPERA E PRIORIZAÇÃO

A presença em si de filas de espera para procedimentos eletivos não necessariamente configura um problema de saúde pública, visto que, em certa medida, sua existência é desejável, uma vez que reflete uma lógica econômica socialmente mais interessante. Conforme Cullis e Jones (1986) destacam, a fila de espera ótima para procedimentos eletivos não é aquela de tempo zero, pois, nessa intensidade, demandaria um sistema de suporte e a existência de equipamentos de saúde sempre de prontidão, situação que seria muito onerosa e, portanto, inviável.

No modelo de análise de filas de espera para procedimentos hospitalares proposto por Iversen (1993), dada uma capacidade produtiva definida, existe um ponto ótimo de espera que permite que a quantidade de procedimentos eletivos realizados seja máxima. Por um lado, nesse modelo de quantidade produtiva fixa, tempos de espera menores determinariam uma produção subótima, ociosidade do sistema e, por conseguinte, gastos desnecessários relacionados à manutenção da estrutura produtiva. Por outro lado, tempos de espera excessivos também determinariam uma produção de procedimentos eletivos abaixo da capacidade instalada, fruto de absenteísmo e cancelamento de procedimentos (WEST, MCKIBBIN, 1982) e do consumo de outros serviços de saúde da mesma instituição, reflexo da deterioração clínica apresentada pelos pacientes, os quais podem inclusive precisar de procedimentos de urgência e emergência.

Também, nesse contexto de espera excessiva, há desperdício frequente de consultas e exames que precisam ser repetidos. (IVERSEN, 1993 *apud* Soreid 1982). Cullis e Jones (1986) destacam outros efeitos e custos que podem se relacionar à espera, como disfunções psicológicas, absenteísmo no trabalho e diminuição da qualidade de vida das pessoas que aguardam pelo procedimento.

Para Siciliani e Hurst (2003), existe um ponto ótimo a ser perseguido, no qual o paciente pode aguardar um determinado período para o procedimento eletivo sem que tenha seu quadro clínico agravado a ponto de impor-lhe perda de utilidade ou óbito e no qual não são gerados custos adicionais de tratamento devido à demora (por aumento da complexidade cirúrgica ou por tempos de permanência de

internação maiores) nem por conta de perda de renda por incapacidade laboral ou por benefícios de transferências sociais (licença médica, por exemplo).

Embora não haja consenso sobre o que seja uma espera excessiva, vários dos países da OCDE utilizam de três a seis meses de espera como o tempo máximo aceitável, sendo essa espera definida como o período transcorrido entre a inserção do paciente na fila de espera para cirurgia pelo médico especialista e a execução do procedimento. (SICILIANI, MORAN, BOROWITZ, 2014b).

Hoogeboom *et al.* (2009) realizaram revisão sistemática da literatura para entenderem o impacto da espera por cirurgia de substituição total de quadril ou joelho na dor e na funcionalidade de pacientes com osteoartrose. Após analisarem os dados autorreferidos de 1.646 pacientes, concluíram que não houve piora nem da dor nem da função desses pacientes, num contexto de espera de até 180 dias para o procedimento. Já Nikolava, Harrison e Sutton (2015) utilizaram os dados de desfechos medidos pelo paciente (do inglês *patient reported outcome measures* – PROM) de pacientes submetidos a artroplastia de quadril (29.303), artroplastia de joelho (32.602), hernioplastia (22.889) ou retirada de varizes (9.184) na Inglaterra para avaliar o impacto do tempo de espera no resultado da cirurgia, sob a perspectiva da qualidade de vida do paciente.

Os autores concluíram que para artroplastia de quadril ou joelho a taxa de deterioração da qualidade de vida desses pacientes é de 0,1% por semana adicional de espera; para hernioplastia ou cirurgia para retirada de varizes, o tempo de espera não alterou o ganho de qualidade de vida pós-procedimento.

Mendes (2019) chama a atenção para o fracasso do modelo de regulação assistencial baseado exclusivamente na gestão da oferta dos serviços de saúde, caracterizado pelo foco no incremento indiscriminado, não racional e, principalmente, descoordenado da oferta, sem envolver mecanismos de racionalização da demanda (qualificação e estratificação, por exemplo) e desconsiderando os benefícios relacionados à implantação de sistemas logísticos de suporte (prontuário eletrônico integrado, transporte sanitário e infovia de central de regulação). Como saída, o autor propõe a adoção da regulação assistencial baseada na saúde da população, solução que seria mais adequada às condições crônicas de saúde atuais e que se caracteriza pela regionalização das unidades especializadas de atenção à saúde, gerenciadas com alta inteligência regulatória e cujas ações devem ser coordenadas.

Nessa mesma linha, a Comissão Intergestores Tripartite propõe que o planejamento das ações e dos serviços de saúde seja regional e integrado, e cabe ao Estado coordená-lo em articulação com os municípios e com a participação da União. Assim se busca a organização do sistema em redes de atenção à saúde, estruturação essa que seria mais eficiente no atendimento das demandas da população e que permitiria racionalizar os recursos existentes. (CONSELHO NACIONAL DE SECRETÁRIOS DE SAÚDE, 2018).

Especificamente no que tange à publicização das filas de espera, cabe destaque ao estado de Santa Catarina, que, por força de lei, desde 2017, deve publicar na internet as informações sobre a posição e previsão de atendimento nas listas de espera para serviços do SUS (consultas de especialidade, alguns tipos de exames, intervenções cirúrgicas e outros procedimentos), assim como a lista de pacientes agendados e atendidos. (SANTA CATARINA, 2017).

Que pesem essas propostas e o último exemplo, a realidade da quase totalidade de municípios e estados brasileiros aponta para um sistema de saúde fragmentado, carente de regulação do acesso e caracterizado pela existência de diversas filas de espera para um mesmo procedimento, distribuídas entre os diversos serviços de saúde que deveriam compor uma rede de atenção à saúde, mas que agem de forma isolada e descentralizada.

Essas características levaram o Conselho Nacional de Justiça a se posicionar diante da iminente escassez de leitos de UTI, ventiladores e demais insumos necessários ao enfrentamento da pandemia da COVID-19. A Nota Técnica nº 24, de maio de 2020, destaca a necessidade de articulação entre os três entes federativos com a finalidade de proporcionar melhores resultados de saúde à população, o que destaca a necessidade da centralização da gestão e da regulação única dos leitos de UTI disponíveis no País, públicos e privados. (CONSELHO NACIONAL DE JUSTIÇA, 2020).

A gestão em si de filas de espera não é tarefa simples. Ela encerra a necessidade de recursos tanto administrativos quanto assistenciais. Os recursos administrativos são os relacionados à inserção e ao acompanhamento dos pacientes em lista, à atualização de dados, ao processo de agendamento e à confirmação do procedimento, além do suporte ao paciente enquanto este aguarda em fila; também há os custos assistenciais gerados pela necessidade de repetição de exames e de

consultas enquanto o paciente aguarda, uma vez que seu quadro clínico pode mudar. (IVERSEN, 1993). Com isso, as listas de espera assumem caráter dinâmico e vivo, e inviabiliza que sejam regidas pela simples ordem de chegada dos pacientes. (HURST, SICILIANI, 2003).

Priorizar significa ponderar, pelo menos, gravidade, urgência, necessidade e benefício esperado. Como Hadorn (2000) explica, embora termos como gravidade e urgência, em alguns contextos, assumam significados muito próximos, é imprescindível que eles sejam diferenciados, assim como se deve adotar uma padronização dessa nomenclatura de modo a nos comunicarmos de maneira eficaz. Para tanto, neste trabalho será adotada a nomenclatura sugerida pelo *Western Canada Waiting List Project* (WCWL), no artigo *Setting priorities for waiting lists: defining our terms* (2000).

O WCWL foi uma iniciativa financiada pelo Governo Federal do Canadá, composta por 19 organizações parceiras, para desenvolver ferramentas que auxiliassem na gestão das filas de espera consideradas prioritárias no fim dos anos 90: cirurgia de catarata, cirurgia geral, prótese de quadril e joelho, ressonância magnética e serviços de saúde mental infantil. (HADORN, 2003). Conforme propõem, gravidade é o grau ou intensidade de sofrimento, limitação de atividades ou risco de morte prematura; urgência é sinônimo de necessidade e é definida como a gravidade somada à evolução natural da condição e ao benefício esperado com a intervenção; benefício esperado representa a sua magnitude e também sua probabilidade de ocorrer dado um caso específico, e é definido como a extensão em que os resultados desejados provavelmente excederão aqueles indesejados. (HADORN, 2000).

Cumpram também destacar o que são procedimentos eletivos, em oposição aos de urgência ou emergência. De acordo com Hurst e Siciliani (2003) e alinhado com o que normalmente se encontra na prática clínica, cirurgias de emergência são aquelas com risco iminente à vida e que devem ser executadas no exato momento em que o paciente se apresenta com o quadro clínico, como ocorre em casos de politraumatismo, ruptura de aorta ou de cordoalha mitral com insuficiência cardíaca aguda. Cirurgia de urgência designa aquela cujo procedimento se faz necessário em algum momento durante a internação, ou seja, antes da alta hospitalar, mas que não

necessariamente precisa ser executada de imediato; já a cirurgia eletiva é aquela em que o paciente pode aguardar em casa até sua execução, que será programada.

Como visto, dada a existência de filas de espera para procedimentos eletivos, a priorização surge como forma de identificar os pacientes de maior urgência e que terão maior benefício em serem atendidos primeiro. Trata-se de um processo de classificação, que gera uma ordem de atendimento a partir de um conjunto de critérios que, por sua vez, podem ser mais ou menos subjetivos e conferem maior ou menor transparência ao processo. (DÉRY *et al.*, 2020).

No Serviço Nacional de Saúde Inglês, o NHS (*National Health Service*), embora não exista uma política pública que defina critérios de priorização para cirurgias eletivas, isso é feito de modo isolado e não coordenado entre diferentes serviços. Cada um definindo seus critérios; em alguns casos, as listas de espera são atendidas unicamente pela ordem cronológica, isto é, o primeiro a chegar é o primeiro a ser atendido; já em outros, são considerados critérios de gravidade e urgência. (GUTACKER, SICILIANI, COOKSON, 2016). Estes autores, ao se debruçarem sobre as filas de espera do NHS para artroplastia de quadril e de joelho, constataram que os pacientes com maior gravidade esperam menos (a diferença dos tempos de espera entre os grupos de maior e menor gravidade é de 25% e 12% para artroplastia de quadril e de joelho, respectivamente). Eles também observaram que a priorização ocorre de forma mais intensa nos hospitais em que há esperas mais longas, logo sugere que, apesar de não haver uma política nacional para isso, médicos e gestores lançam mão da priorização à medida que as restrições e os tempos de espera se apresentam e se alargam.

A fim de trazer mais transparência para o processo de priorização, a literatura aponta para o uso de ferramentas de priorização de pacientes (FPP), as quais podem ser implementadas como parte de uma política nacional de priorização, a exemplo do que ocorre na Nova Zelândia desde a década de 90 (DEW *et al.*, 2005; “*Planned Care services*”, 2021), ou podem refletir uma ação regional, como aconteceu no *Western Canada Waiting List Project* e na Catalunha (ver adiante), podendo ainda retratar ações isoladas de serviços, como ocorrem na Espanha e no Reino Unido. (SOLANS-DOMÈNECH *et al.*, 2013).

Na revisão sistemática da literatura conduzida por Déry *et al.* (2020), a maior produção científica nesse assunto se encontra no Canadá, seguido por Espanha e

Nova Zelândia. Ferramentas de priorização de pacientes são instrumentos desenhados para auxiliar médicos e gestores na tomada de decisão sobre a classificação de pacientes em uma fila de espera, o que define quem deve ser visto antes, de modo explícito, transparente e, espera-se, mais justo. (DÉRY *et al.*, 2019, 2020).

Das 34 ferramentas estudadas pelos autores (descritas em 46 artigos), 50% eram instrumentos de pontuação que variavam de 0 a 100 a partir de um conjunto de critérios ponderados; já 8 delas (24%) usavam também um sistema de pontos, mas agrupavam os pacientes em categorias mais amplas (prioridade alta, média ou baixa, por exemplo). A média da quantidade de variáveis consideradas para a classificação dos pacientes nessas FPP foi de 7,6 (SD: 3,8), e variava entre 2 e 17. Outro ponto de destaque é que, para sua construção, participaram diferentes atores do cuidado, a saber: clínicos em 50% delas, médicos especialistas em 35%, médicos cirurgiões em 29%, médicos da família e comunidade 26%, gestores de saúde 24%, pacientes e pesquisadores em 15%.

Embora a maioria dessas FPP foram desenhadas para uso hospitalar (19/34), foi encontrado o uso para diferentes tipos de serviços além do hospital, como ambulatório de especialidades, clínica de reabilitação e atenção primária. No que tange às variáveis dos modelos, a principal se relacionava à independência/capacidade do paciente de desempenhar um papel, seguida pelas variáveis de limitação de funcionalidade, dor e sofrimento, probabilidade de recuperação da doença/progressão da doença/prognóstico, vantagens e benefícios da intervenção, ter ou ser cuidador, idade, nível de urgência e tempo na lista de espera. Para ferramentas específicas, voltadas para doenças específicas, também eram ponderados os sintomas relativos à doença em questão e a outras medidas padronizadas já existentes para avaliar a enfermidade.

Solans-Domènech *et al.* (2013) se dedicaram a desenvolver um sistema de pontuação universal de priorização de cirurgias eletivas que pudesse ser aplicado a todos os cidadãos da Catalunha, independente do tipo de cirurgia. Após revisarem a literatura e utilizarem as técnicas de grupo nominal e de Delphi para construir consenso, propuseram um sistema de pontuação centesimal, constituído por três dimensões (comprometimento clínico-funcional, benefício esperado e papel social), o que agregou um total de oito critérios com diferentes pesos máximos (gravidade da



doença (0 a 23 pontos), dor e outros sintomas principais (0 a 14 pontos), taxa de progressão da doença (0 a 15), dificuldade em realizar atividades da vida diária (0 a 14), probabilidade e grau de melhora esperado após a intervenção (0 a 12 pontos), necessidade de cuidador para auxiliar nas tarefas diárias (0 pontos para aqueles que não necessitam ou que necessitam e possuam um cuidador; 5 pontos para aqueles que necessitam, mas não possuem um cuidador), limitação para cuidar de seus dependentes (0 a 8 pontos) e limitações na capacidade de trabalhar, estudar ou procurar emprego (0 a 9 pontos)).

Embora a idade não tenha entrado como critério isolado, os autores advogam que ela foi representada por outras variáveis presentes, como probabilidade e grau de melhora esperado após a intervenção ou dificuldade para realizar as atividades de vida diária. Também, há de se observar que os maiores pesos foram atribuídos às variáveis clínico-funcionais e, os menores, a variáveis relacionadas ao papel social dos pacientes. Outro aspecto levantado por Solans-Domènech *et al.* (2013) diz respeito à facilidade de implementação que deve caracterizar essas FPP, dado que, quanto maior o número de profissionais necessários para a administração do instrumento e o registro da informação coletada, menor será a chance de efetivação dessa prática na vida diária das instituições e de seus profissionais.

No Brasil, ainda que não exista uma política nacional de priorização de procedimentos eletivos, existem ações que, para situações específicas, concorrem para a racionalização de recursos escassos. A Resolução CFM nº 2.156/2016 estabelece critérios de admissão e alta de unidades de terapia intensiva (UTI), que em muito se assemelham aos critérios e princípios discutidos até aqui. A Resolução postula que as admissões devem ter por base o diagnóstico e a necessidade do paciente, a disponibilidade de recursos na instituição de saúde, a oferta de leitos, o potencial benefício para o paciente em relação às intervenções que ocorrerão nesse ambiente controlado e de alta tecnologia, assim como deve levar em conta a condição clínica do paciente e seu prognóstico.

Também, em seus arts. 9º e 11, parágrafo único, destacam que a priorização deve ser feita de forma explícita e sob protocolos a serem divulgados ao corpo clínico do hospital e aos gestores do sistema de saúde, ao conferir, assim, transparência ao processo. (CONSELHO FEDERAL DE MEDICINA, 2016).

Essa resolução embasou o posicionamento conjunto da Associação de Medicina Intensiva Brasileira (AMIB), Associação Brasileira de Medicina de Emergência (ABRAMEDE), Sociedade Brasileira de Geriatria e Gerontologia (SBGG) e Academia Nacional de Cuidados Paliativos (ANCP), em abril de 2020, frente ao potencial esgotamento de recursos, como leitos de UTI, ventiladores mecânicos e profissionais de terapia intensiva por conta da pandemia por COVID-19. O posicionamento das sociedades não deixa dúvidas de que, caso necessário, a priorização deve ser feita a partir de protocolos claros, transparentes, tecnicamente fundamentados e embasados na ciência, nos princípios éticos praticados no Brasil e em conformidade com o arcabouço legal do País. Não menos importante, aconselham expressamente que esse processo não seja feito de modo subjetivo, inconsciente e sem o devido registro formal. (KRETZER *et al.*, 2020) . Conforme os autores destacam:

A utilização de um protocolo de maneira consistente pelas diversas instituições de saúde garante que um maior número de pacientes sejam igualmente sujeitos aos mesmos critérios cancelados pelas autoridades responsáveis tanto pelo zelo técnico-científico quanto o ético-legal do processo (KRETZER *et al.*, 2020, p. 3).

Na Tabela 1, observamos o passo a passo proposto pelo grupo, com destaque para a utilização de diversas escalas, a fim de tornar o processo menos subjetivo e obscuro possível.

Tabela 1 – Passo a passo do modelo de triagem AMIB/ABRAMEDE

Passos	Critérios	Pontuação				Total
		1	2	3	4	
1	Calcular SOFA (total: _____) e pontuar conforme estratificação ao lado	SOFA $\leq$ 8	SOFA 9 - 11	SOFA 12 - 14	SOFA > 14	
2	Tem comorbidades graves, com expectativa de sobrevida < que um ano?*	- - -	- - -	Sim	- - -	
3	Aplicar a ECOG e pontuar conforme a estratificação ao lado	0 - 1	2	3	4	
4	Calcular a pontuação total dos critérios 1 a 3					
5	Alocar o leito de UTI ou VM ao paciente com menor pontuação total desde que não tenha havido empate					
6	Em caso de empate utilizar os seguintes critérios hierarquicamente:					
6a	Menor score SOFA total					
6b	Julgamento clínico da equipe de triagem					

\*Fazer a avaliação preferencialmente através do SPICT-BR ou PIG-GSF

Fonte: Kretzer *et al.* (2020)

A constante preocupação em relação à transparência e à objetividade dos diferentes sistemas de priorização encontra lastro em trabalhos da literatura que analisaram o processo de decisão médica e o peso que cada parte interessada (médicos clínicos, cirurgiões, médicos do trabalho, médicos de família, pacientes e pessoas leigas (sem a doença)) atribui para cada dimensão ou variável envolvida no processo de priorização. Para uma mesma condição cirúrgica e dado nível de gravidade, Oudhoff *et al.* (2007) evidenciaram que cirurgiões atribuem menor prioridade para o caso quando comparados a médicos do trabalho, médicos de família, pessoas não doentes e pacientes. De modo geral, o nível de prioridade atribuído pelos doentes, pelas pessoas leigas e pelos médicos de família e comunidade se aproximou. Além dessa variação entre grupos, os autores chamaram a atenção para a grande variação existente intragrupos, isto é, mesmo dentro de um grupo específico (cirurgiões ou pacientes, por exemplo), existe um grau importante de discordância entre os pares no que tange à priorização de uma condição estabelecida. Dito de outra forma, quando se fala em prioridade, esta muda a depender do avaliador, do seu repertório e de seus valores.

Ao investigarem o julgamento clínico de médicos cardiologistas em relação à priorização e à urgência atribuídas a casos de cirurgia de revascularização do miocárdio, pesquisadores da Universidade de Queen's, na Irlanda do Norte,

apontaram que, se por um lado, existe consenso na inclusão de critérios maiores, como classe da angina (intensidade da dor no peito) ou lesão de tronco coronário esquerdo; por outro, alguns médicos consideram, em suas decisões de indicação cirúrgica, fatores como tabagismo e obesidade, variáveis estas que não estão relacionadas à melhor eficácia do procedimento cirúrgico quando comparado ao tratamento clínico. Para os autores, os médicos que dão prioridade a um grupo, em detrimento do outro (tabagista *versus* não tabagistas; ou obeso *versus* não obeso), fazem-no a partir de crenças próprias, o que ressalta, pois, como o julgamento do médico pode ser afetado pela percepção individual de cada profissional em relação a riscos envolvidos na doença coronariana. (KEE *et al.*, 1998).

Loftus *et al.* (2019) chamam a atenção para a complexidade envolvida no processo de decisão cirúrgica, cuja qualidade é influenciada pelos valores e pelas emoções do médico e do paciente, pela forma como se dá essa relação entre as partes, pela quantidade de tempo disponível para julgamento e deliberação, pela carga de trabalho sob a qual o médico está submetido no momento da decisão, assim como pela sua capacidade de raciocínio dedutivo hipotético, muitas vezes afetada por rotinas de privação de sono (COLEMAN *et al.*, 2019) e por outros vieses (Tabela 2).

Há de se destacar também que, não raro, no momento da decisão, os dados disponíveis são incompletos ou encontram-se fragmentados no prontuário eletrônico, o que impõe um processo manual e demorado de aquisição e construção de informações. Para Loftus *et al.* (2019) e Liu *et al.* (2019), todas essas dificuldades, restrições e potenciais danos podem ser melhor contornados com o uso de novas tecnologias associadas a *Big Data*, como modelos de aprendizado de máquina (*machine learning*), caracterizados, entre outros aspectos, por sua consistência (capacidade de gerar os mesmos resultados diante das mesmas variáveis) e sua inesgotável capacidade de fazer os mesmos cálculos, em diferentes cenários, sem prejuízo na análise, diferente do que ocorre com humanos, em que a fadiga e a pressão exercida pela escassez do tempo podem mudar seu desempenho. (OBERMEYER, EMANUEL, 2016).

Tabela 2 – Tipos de vieses na tomada de decisão cirúrgica

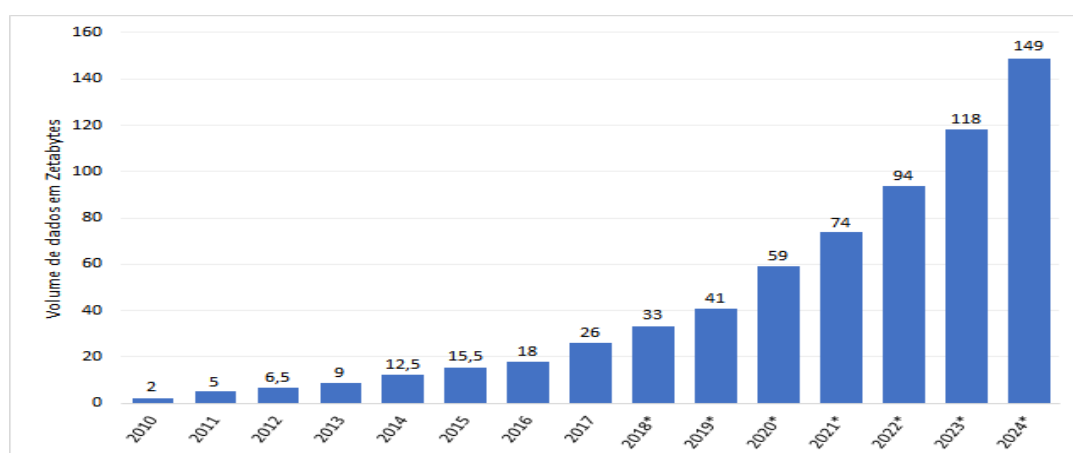
Tipo de Viés	Exemplos
Efeito de Enquadramento	O clínico apresenta o quadro do paciente a um cirurgião em um contexto diferente do que este teria percebido durante uma avaliação independente
Viés de excesso de confiança	O cirurgião não percebe como suas fraquezas e expectativas particulares afetam seu julgamento
Viés de Comissão	O cirurgião tende a agir mesmo quando a inação pode ser preferível, especialmente no contexto de viés de excesso de confiança
Viés de ancoragem	Os pacientes são informados dos resultados esperados usando-se dados agregados de populações de pacientes, sem ajustes para seu perfil de risco personalizado
Viés de recordação	Experiências recentes com uma determinada população de pacientes ou cirurgia afetam desproporcionalmente a tomada de decisão cirúrgica em relação a experiências mais antigas
Viés de confirmação	Os resultados são previstos usando crenças pessoais em vez de diretrizes baseadas em evidências

Fonte: Loftus *et al.* (2019)

## 2.2 BIG DATA E APRENDIZADO DE MÁQUINA

Segundo a empresa alemã Statista, especializada em dados de mercado e consumidores, a estimativa é a de que, em 2021, 74 zetabytes sejam capturados, criados e consumidos em todo o mundo (STATISTA, 2021) (Figura 1); isso representa um volume de dados 1.480% maior que o volume gerado em 2011, mas apenas metade do volume de dados previsto para 2024. Em termos comparativos, se assumirmos que a população mundial de 2021 é de 8 bilhões de pessoas, que cada byte equivallesse a 1 real e que fôssemos dividir todo esse dinheiro de forma igualitária, isso significaria dizer que, em 2021, cada cidadão do mundo teria o equivalente a 10 trilhões de reais.

Figura 1 – Volume de dados criado, capturado, copiado e consumido ao redor do mundo entre 2010 e 2024. \* Valor estimado



Fonte: Statista (2021)

No ambiente da saúde, isso não foi diferente. Segundo o estudo TIC Saúde 2019, 82% dos estabelecimentos de saúde do Brasil utilizaram algum sistema eletrônico para controlar as informações dos pacientes, um avanço de 9% em relação ao ano anterior. (NIC.BR, 2020).

Os registros eletrônicos de saúde (RES) são hoje um repositório riquíssimo de dados, com diagnósticos, história clínica, descrições e planos de tratamento, resultados de exames, medicações prescritas, imagens radiológicas e diversos outros registros gerados e capturados com os mais variados propósitos: cuidado assistencial, faturamento, auditoria, monitoramento da qualidade, entre outros. Redes sociais, como *Facebook*, *Instagram* e *Twitter*, se tornaram plataformas de compartilhamento de experiências de pacientes com condições crônicas, como diabetes e doenças do coração, com até 25% desses doentes interagindo e produzindo dados.

*Hardwares* de detecção estão cada vez mais avançados, apresentam-se em diferentes formatos (sensores vestíveis, implantáveis e ambientais) e permitem o monitoramento contínuo de parâmetros fisiológicos, como a pressão sanguínea e a glicemia dos pacientes. Na última década, os *smartphones*, inseridos nas vidas de mais de 1,75 bilhão de pessoas, assumiram um papel importante como dispositivo pessoal de registro de informações em saúde. (ANDREU-PEREZ *et al.*, 2015; HEMINGWAY *et al.*, 2018). Os desafios e as soluções relacionados à captura, ao armazenamento, à distribuição e ao processamento dessa grande massa de dados caracterizam o que hoje se chama de *Big Data*.

Apesar das diversas definições presentes na literatura, todas convergem para o entendimento de que *Big Data* se refere a uma quantidade de dados estruturados e não estruturados suficientemente grande a ponto de inviabilizar seu processamento e sua interpretação com as tecnologias tradicionais, demandando novos processos de captura, armazenamento, distribuição, processamento, visualização e interpretação. (ANDREU-PEREZ *et al.*, 2015; CHIAVEGATTO FILHO, 2015; HULSEN *et al.*, 2019).

Conforme Andreu-Perez *et al.* (2015) sugerem, a caracterização de *Big Data*, em saúde, engloba os “seis V’s do *Big Data*”, a saber:

- Volume: refere-se à quantidade crescente de dados que é gerada em diversos contextos, seja por meio de tecnologias associadas ao cuidado direto com o

paciente, seja por meio de sensores em vestimenta, *smartphones* ou pela genômica, proteômica, metabolômica, assim como por variáveis ambientais que podem ser capturadas e que impactam na saúde da população (nível de monóxido de carbono no ar, por exemplo).

- **Velocidade:** trata tanto da velocidade crescente com que os dados relacionados à saúde são gerados quanto da necessidade de um processamento igualmente rápido a fim de suportar a decisão clínica e de gestores de modo tempestivo.

- **Variedade:** estes diversos dados podem estar em bases estruturadas, semiestruturadas e não estruturadas (áudio, imagem, arquivo de vídeo); podem ser quantitativos (resultado de um exame de sangue, dados de um sensor) ou qualitativos (texto livre em que o profissional de saúde registra a evolução do paciente no prontuário eletrônico ou num *tweet* dessa mesma pessoa na rede social).

- **Veracidade:** confiabilidade dos dados obtidos. Pouco adianta muitos dados se estes forem pouco confiáveis. No contexto da saúde, essas informações podem ser geradas tanto por profissionais qualificados quanto por pessoas não treinadas, sem controle. Diferentes contextos, mesmo dentro de uma unidade de saúde, podem gerar dados mais ou menos confiáveis (medição ambulatorial vs. medição durante uma internação em uma UTI), sem contar abreviações e erros tipográficos.

- **Variabilidade:** refere-se à consistência dos dados ao longo do tempo e entre diferentes fontes. Conforme eles são transferidos e executados em diferentes aplicações, sua integridade pode ser corrompida caso não haja mecanismos de controle. No caso de textos livres, refere-se também aos diferentes significados que uma mesma palavra pode assumir de acordo com o contexto em que está inserida.

- **Valor:** a partir da análise e visualização desses dados, é imprescindível que estes se traduzam em melhorias para os pacientes, para as instituições, para os pesquisadores, enfim, para todos que compõem a cadeia de valor em saúde.

A possibilidade de explorar essa grande massa de dados e dela extrair novas informações abriu um vasto campo de oportunidades na prática clínica e de pesquisa. (HEMINGWAY *et al.*, 2018). Conforme Chiavegatto Filho (2015) destaca, muito do conhecimento científico e da prática médica ainda se alicerça sobre grandes médias. O autor cita o exemplo do estudo de Ruff *et al.* (2014), que

comparou o uso de novos anticoagulantes orais em relação ao uso de varfarina em pacientes com fibrilação atrial, e concluiu que as novas drogas, em relação à varfarina, reduzem em 19% o risco de acidente vascular cerebral ou eventos embólicos sistêmicos. Fato é que nenhum dos pacientes do estudo teve o risco diminuído em 19%; alguns tiveram o evento, outros simplesmente não o tiveram.

O *Big Data* e as suas ferramentas auxiliares aspiram conseguir resolver esse problema por meio da medicina de precisão, isto é, conseguindo identificar, entre todos os pacientes com fibrilação atrial, quais serão aqueles que, de fato, terão o benefício dos novos anticoagulantes orais e quais podem permanecer com varfarina, dado que os novos medicamentos não trarão o benefício esperado.

No campo da pesquisa, o *Big Data* também abre caminhos. Como ele apresenta novas capacidades de cruzar grande volume de dados e gerar *insights*, ele igualmente amplia a possibilidade de associações, outrora impensáveis. Diferente da pesquisa impulsionada por hipóteses, em que se busca validar ou recusar uma hipótese, muitas das pesquisas com *Big Data* serão geradoras de hipóteses, isto é, trarão à luz novas possibilidades de associação, as quais deverão ser confirmadas ou negadas pelo método científico tradicional. (HULSEN *et al.*, 2019).

Essa busca por tendências e associações não evidentes em conjuntos de dados menores concretizou-se à medida que a tecnologia de aprendizado de máquina (AM), um ramo da inteligência artificial, foi sendo aplicada e desenvolvida. Embora o uso de métodos de AM não seja recente na medicina, as últimas décadas vivenciaram pelo menos duas mudanças substanciais que permitiram seu avanço. A primeira, supramencionada, refere-se aos dados, que passaram a ser gerados e armazenados numa velocidade e quantidade nunca antes vista; a segunda foi o desenvolvimento de sistemas computacionais capazes de receber essa grande massa de dados, processá-la utilizando um maior número de operações matemáticas quando comparado às técnicas tradicionais e, com isso, estabelecer relações mais complexas entre dados e desfechos clínicos. (LIU *et al.*, 2019).

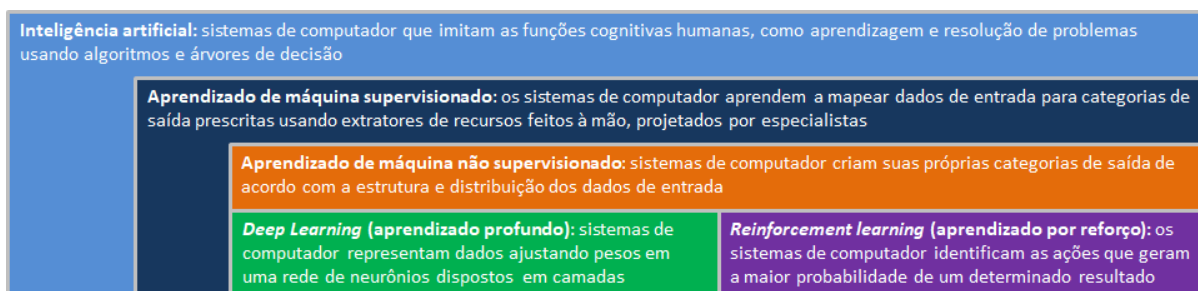
Diferente do que ocorre no método tradicional de programação de *software*, em que um conjunto de regras pré-definidas é organizado e as respostas são geradas a partir do processamento dos dados inseridos, os algoritmos de AM são capazes de criar suas próprias regras e perguntas a partir de um conjunto de dados



e rótulos (respostas), isto é, executam uma tarefa sem instruções explícitas. (OBERMEYER, EMANUEL, 2016; IBM BRASIL, 2021).

Conforme destacado na Figura 2, esses algoritmos podem contar ou não com a supervisão humana. Nos modelos supervisionados, compete ao homem definir as possíveis variáveis de interesse relacionadas a um problema e criar extratores que vão transformar esses dados brutos em representações desejadas (um vetor, por exemplo). Uma vez feito isso, o algoritmo cria suas próprias regras de correlação entre essas variáveis de entrada, associando-as a um resultado esperado (rótulo) e é capaz de classificar (rotular) novos dados. Já nos modelos não supervisionados, não existe uma relação previamente conhecida entre os dados de entrada e de saída (ou seja, não existem resultados/rótulos conhecidos), competindo ao algoritmo explorar matematicamente a estrutura e a distribuição dos dados de entrada e criar suas próprias categorias de saída. (MOONEY, PEJAVER, 2018; SHICKEL *et al.*, 2018; LIU *et al.*, 2019).

Figura 2 – Resumo das técnicas de inteligência artificial

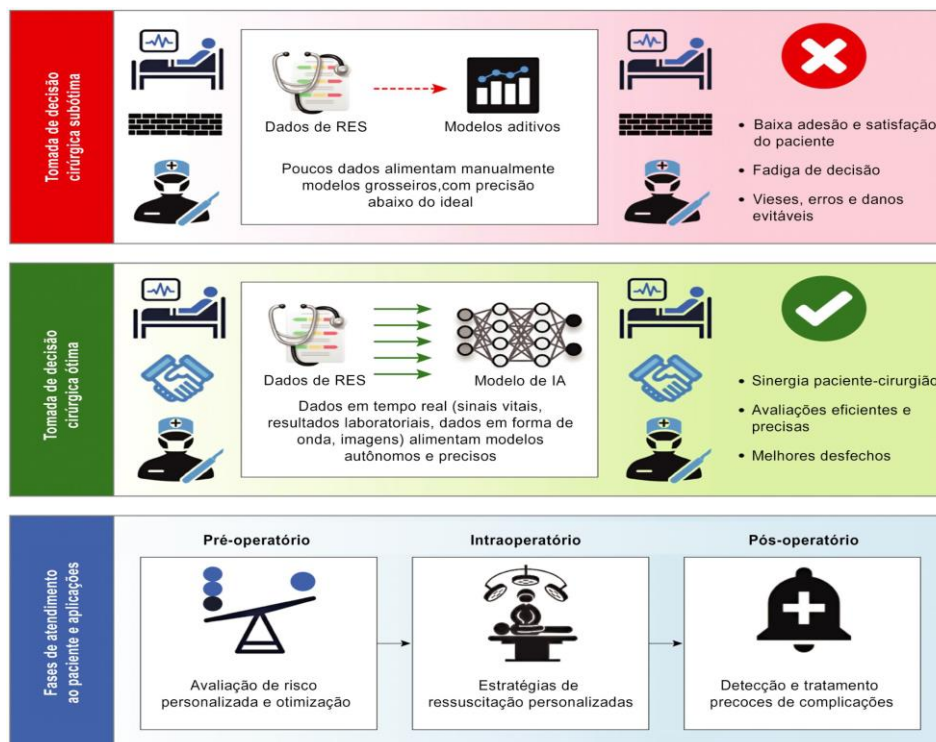


Fonte: Loftus *et al.* (2019)

A aplicação de AM na saúde é vasta e já demonstrou seu potencial de melhorar a forma como o cuidado é prestado e os sistemas de saúde são pensados e executados. Longe de serem uma panaceia, os algoritmos de AM surgem como ferramentas de apoio à decisão clínica e ao gerenciamento dos sistemas de saúde, cuja finalidade é aumentar o desempenho humano, não substituí-lo. (NGIAM, KHOR, 2019).

Especificamente em relação à tomada de decisão cirúrgica, eles despontam como uma forma de auxílio ao cirurgião nesse processo que normalmente é manual, individualizado, feito sob condições de restrição de tempo e repleto de incertezas e incompletude de dados (Figura 3). (LOFTUS *et al.*, 2019).

Figura 3 – Abordagens ótima e subótima para tomada de decisão cirúrgica



Fonte: Loftus *et al.* (2019)

Bihorac *et al.* (2019) integraram registros eletrônicos em saúde com bases de dados públicos e utilizaram o CEP da residência do paciente, por exemplo, para calcular a distância até o hospital e as características da vizinhança. A partir de 285 variáveis de 51.457 pacientes cirúrgicos, desenvolveram e validaram um algoritmo de AM capaz de prever, com grande acurácia, complicações infecciosas, insuficiência renal aguda, admissão em UTI por mais de 48hs, complicações cardiovasculares e neurológicas, sepse e tromboembolismo venoso, além da mortalidade prevista em 1, 3, 6, 12 e 24 meses após a cirurgia.

O algoritmo CURATE.AI, desenvolvido pela Universidade Nacional de Singapura em associação com a Universidade da Califórnia Los Angeles, auxilia a equipe clínica a otimizar a dosagem de antineoplásicos, especialmente no contexto de terapia combinada. Como resultado, a dose inicial de uma das drogas pode ser reduzida à metade, o que se traduz, na prática assistencial, numa terapia mais bem tolerada e efetiva, com menor custo e sem comprometer o efeito desejado: impedir a progressão da doença. (NGIAM, KHOR, 2019).

Hemingway *et al.* (2018) produziram uma revisão clínica sobre *Big Data* a partir de registros eletrônicos em saúde. Dentre os diversos estudos apontados,

chamam a atenção para um algoritmo de AM construído a partir de 46 variáveis (clínicas, laboratoriais, eletrocardiográficas e ecocardiográficas) que identificou dois subgrupos mutuamente excludentes de pacientes com insuficiência cardíaca de fração de ejeção preservada, com desfechos clínicos distintos, que até então eram entendidos como um grupo homogêneo.

Nam *et al.* (2018) utilizaram redes neurais para analisar imagens de radiografia de tórax de 34.000 pacientes, especificamente em busca de nódulos pulmonares malignos. O algoritmo foi mais preciso que 95% dos radiologistas que participaram do estudo. Na cardiologia, há mais de 40 anos, são usados algoritmos para realizar a leitura de eletrocardiogramas, entretanto, com os avanços recentes advindos das técnicas de AM, Strodthoff e Strodthoff (2018) conseguiram obter sensibilidade de 93% e especificidade de 90% na detecção de traçados de infarto agudo do miocárdio, desempenho comparável ao de cardiologistas. Em outro estudo retrospectivo, Zhang Jeffrey *et al.* (2018) lançaram mão de 14.035 ecocardiogramas para treinar e avaliar uma rede neural convolucional, capacitando-a a identificar as cinco principais janelas ecocardiográficas e a quantificar, entre outras variáveis, volume das câmaras cardíacas, massa ventricular e fração de ejeção. Como resultado, o algoritmo foi capaz de classificar achados, como cardiomiopatia hipertrófica, amiloidose cardíaca e hipertensão arterial com alta precisão.

Para além dos benefícios individuais, essas tecnologias de análise de grandes massas de dados se apresentam como ferramentas poderosas no auxílio à tomada de decisão pelos gestores. Pesquisadores da Faculdade de Saúde Pública da Universidade de São Paulo, a partir de dados do estudo Saúde Bem-estar e Envelhecimento (SABE), coorte que acompanha indivíduos com 60 anos ou mais residentes no município de São Paulo, desde 2000, conseguiram prever, com sucesso de 70%, os idosos que iriam morrer num período de cinco anos, isto é, o algoritmo acertou 83 dos 118 óbitos que, de fato, ocorreram. (SANTOS *et al.*, 2019).

Já Pan *et al.* (2017) compararam o desempenho do sistema de triagem do programa *Better Birth Outcomes* (melhores desfechos de natalidade, em tradução livre), do estado de Illinois, nos Estados Unidos, realizado por meio de triagem em papel, com o desempenho de algoritmos de AM. Nesse caso em especial, havia uma limitação administrativa, isto é, tratava-se de um programa para o qual existia um limite, por restrição financeira, de gestantes em seguimento (até 2.000 gestantes

ativas), ou seja, a efetividade do programa era função da sua capacidade de identificar corretamente gestantes e/ou recém-nascidos que realmente se beneficiariam do pré-natal de alto risco oferecido pelo programa.

Comparando-se os dois métodos de escolha de gestantes elegíveis ao programa, os algoritmos de aprendizagem de máquina tiveram desempenho 36% superior à triagem baseada em papel, o que equivale a dizer que de 100 a 170 mulheres a mais, gestantes de alto risco, seriam corretamente selecionadas e receberiam o serviço especializado caso a escolha fosse feita por IA. Isso evidencia o impacto positivo que essas novas tecnologias podem proporcionar aos sistemas de saúde como um todo.

Se por um lado o AM tem o potencial de revolucionar a maneira como pensamos e entregamos o cuidado; por outro, inúmeros também são os desafios a serem endereçados e transpostos para que esse potencial se realize. Ao mesmo tempo em que os hospitais são ricos em registros eletrônicos de saúde e demais dados dos pacientes, estes foram concebidos para atender demandas específicas da operação e, via de regra, encontram-se em sistemas distintos: faturamento, auditoria, gestão de risco, estoque e dispensação e assistência, entre outros.

O modo tradicional de armazenamento e transferência de dados, baseado na linguagem de consulta estruturada (do inglês *Structured Query Language* (SQL)), apresenta-se como uma estrutura lenta para o uso dos modelos de IA, o que demanda novas tecnologias, genericamente definidas como NoSQL. Criar um ambiente de tecnologia da informação em que esses dados convirjam e possam ser armazenados, manipulados, combinados, curados e analisados de modo rápido demanda esforço e visão estratégica das instituições. (HUANG *et al.*, 2015; HULSEN *et al.*, 2019).

Outra condição *sine qua non* para que esses modelos atinjam a finalidade desejada é a veracidade, a confiabilidade dos dados registrados e usados na confecção dos algoritmos de AM. Pouco valor tem uma grande massa de dados se ela não reflete a realidade clínica e o problema a que o algoritmo se propõe a resolver, dado que esses dados serão utilizados para treinar, para ensinar o sistema computacional.

O sucesso de um algoritmo supervisionado começa na precisão da rotulagem do desfecho de interesse; quanto maior for a imprecisão, piores serão os resultados

alcançados. (NGIAM, KHOR, 2019). Para além da veracidade, há de se destacar que, na prática clínica, que alimenta os registros eletrônicos em saúde, em oposição ao que ocorre em estudos clínicos, não é incomum que parte dos dados não sejam registrados, o que impõe desafios no momento da construção do algoritmo. (HEMINGWAY *et al.*, 2018).

Esse conjunto de técnicas que envolvem a seleção das variáveis, o seu entendimento, a preparação, a organização e a estruturação dos dados caracterizam a etapa de pré-processamento dos dados, passo crítico que antecede a confecção propriamente dita do modelo de AM. É nesse ponto que dados *outliers* são identificados e tratados, os dados são padronizados (assumem a mesma ordem de grandeza), a dimensão (número de variáveis) da base é analisada e, se necessário, reduzida. A própria ausência de um registro (*missing data*), a depender do contexto, pode agregar informação ao modelo, por exemplo, num algoritmo para predição de problemas cognitivos. A criticidade dessa etapa reside no fato de que falhas nesse estágio podem levar ao fenômeno de sobreajuste do algoritmo (do inglês, *overfitting*), situação em que o algoritmo, ao invés de entender a correlação entre as variáveis de entrada e o resultado esperado (rótulo) e criar suas regras, ele, na verdade, se ajusta a algum ruído ou erro aleatório detectado na base de dados. Na prática, é como se ele decorasse os dados atuais e os seus rótulos. O grande problema é que um algoritmo sobreajustado opera extremamente bem na base de treinamento, mas se mostra ineficaz e com baixo desempenho quando confrontado com novos dados. Ou seja, torna-se um algoritmo de predição que não faz boas previsões (baixa acurácia). (MOONEY, PEJAVER, 2018).

Finalmente, mas não menos importante, cabe destaque para duas questões relevantes: privacidade e integração dessas novas tecnologias à prática clínica. Em relação à privacidade, cabe lembrar que as instituições de saúde detêm a custódia dos dados, não sua propriedade. O proprietário do dado continua a ser o paciente e isso, por si só, pode impor limitações à adoção de tecnologias de IA. (HULSEN *et al.*, 2019). Mesmo no cenário de uso de dados anonimizados, Mooney e Pejaver (2018) chamam a atenção para o potencial fenômeno de divulgação dedutiva, isto é, situação em que a junção de duas ou mais bases de dados, com diferentes variáveis, leva à inferência e identificação do indivíduo que congrega todas aquelas características. Além disso, existe a possibilidade de divulgação inadvertida desses

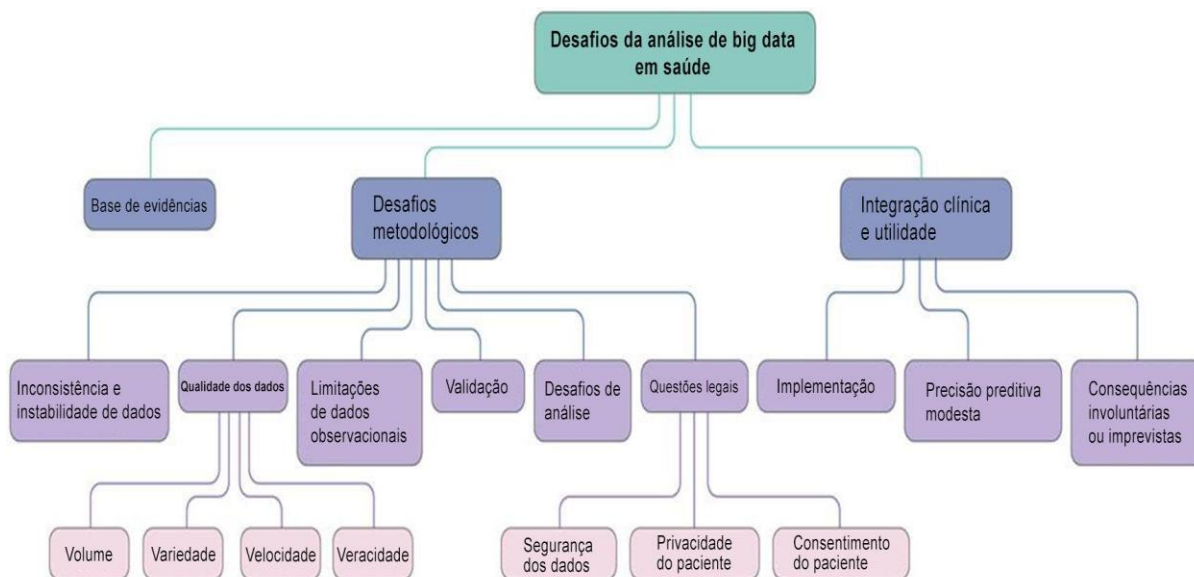
dados pelos pesquisadores e suas instituições, assim como o crime cibernético é uma realidade.

Em relação à aplicação do AM na vida real, ainda existe um caminho a ser percorrido (e que está sendo trilhado). Hoje a comunidade científica enxerga essas ferramentas como auxiliares à decisão humana. Isso, por si só, impõe a necessidade de médicos e demais tomadores de decisão de conhecerem minimamente essas novas tecnologias, apropriando-se de seu processo de confecção e validação, possíveis vieses e pontos fortes. (LIU *et al.*, 2019).

A incorporação dessas técnicas, no dia a dia dos trabalhadores, também dependerá de serem incorporadas mediante interfaces amigáveis aos usuários, isto é, se o processo de inicialização dessas aplicações for complexo e demandar que os trabalhadores parem suas rotinas para isso, menores serão as chances de serem, de fato, integradas às rotinas assistenciais e gerenciais. (HULSEN *et al.*, 2019). Mas antes disso, consoante com Topol (2019), um dos mais renomados pesquisadores de IA na saúde, é imprescindível que haja validação prospectiva desses algoritmos na vida real, mediante estudos aprofundados e passando pelo escrutínio da comunidade científica.

Na Figura 4, temos um resumo dos principais desafios relacionados à adoção da IA artificial na saúde.

Figura 4 – Desafios relacionados à adoção da inteligência artificial na saúde



Fonte: Rumsfeld, Joynt e Maddox (2016)

### **3. OBJETIVOS**

#### **3.1 OBJETIVO GERAL**

Propor um modelo de aprendizado de máquina capaz de identificar pacientes em fila de espera para procedimento cirúrgico eletivo que, caso não sejam atendidos de forma tempestiva, serão submetidos à cirurgia de urgência.

#### **3.2 OBJETIVOS ESPECÍFICOS**

- Determinar a acurácia, sensibilidade e especificidade do modelo proposto.
- Identificar os parâmetros do modelo de aprendizado de máquina proposto.



## 4. MÉTODO

### 4.1 CONSIDERAÇÕES GERAIS

Trata-se de pesquisa aplicada, quantitativa, experimental e retrospectiva, na qual foram aplicados algoritmos de aprendizado de máquina e buscou-se correlacionar a resposta de interesse  $y_i$  (realização de cirurgia cardíaca valvar de urgência ou eletiva) com os preditores  $x_i$  (Tabela 3).

A revisão bibliográfica foi a partir do *Google Scholar* e do motor de busca PubMed, utilizado para acessar a base de dados MEDLINE a partir das palavras-chave: *Waiting List and Prioritization*; *Waiting List and Elective Surgery and Prioritization*; *Big Data and Health*; *Machine Learning and Health*. Outros artigos foram selecionados a partir das referências bibliográficas citadas nos artigos encontrados nas primeiras buscas do *Google Scholar* e do PubMed. Não menos importante, foram fontes essenciais tanto de conhecimento quanto de referências dois trabalhos aplicados apresentados ao Mestrado Profissional, linha Saúde, da FGV-EAESP em anos anteriores: Predição da hospitalização em pacientes idosos no departamento de emergência: uma abordagem utilizando AM, de Fabricio Prospero Machado (2019) e Gerenciamento da fila de espera para cirurgia ginecológica em hospital municipal da Zona Sul de São Paulo: como garantir acesso e otimizar a utilização de recursos, de Mariana Granado Barbosa (2018).

O projeto foi avaliado e aprovado pelo Comitê de Ética em Pesquisa do Hospital das Clínicas da Faculdade de Medicina da Universidade de São Paulo. CAAE 39973920.1.0000.0068. Parecer nº 4.438.511. (Anexo A).

### 4.2 LOCAL DA PESQUISA

Embora sua concepção data da década de 1950, a oficialização do Instituto do Coração do Hospital das Clínicas da Faculdade de Medicina da Universidade de São Paulo (INCOR-HCFMUSP) ocorre, de fato, em 1963, pelo Decreto-Lei nº 42.817, então sob denominação de Instituto de Doenças Cardiopulmonares, mas já idealizado sob o tripé ensino, pesquisa e assistência em cardiologia, pneumologia e cirurgias torácica e cardíaca.

Trata-se de um hospital público, cujo modelo de gestão se baseia numa fundação de apoio, a Fundação Zerbini, responsável por garantir vantagens competitivas em relação à administração pública direta e indireta. Em relação a seu financiamento, 75% se dá por meio do Tesouro do estado de São Paulo e pelo Fundo Nacional de Saúde (faturamento da tabela SUS) e 25% por recursos provenientes do atendimento a pacientes de convênios e particulares.

O InCor iniciou o atendimento à população em janeiro de 1977, por meio de seu ambulatório. Em 2019, 42 anos depois, contava com 445 leitos instalados, 168 dos quais de terapia intensiva; executou 5.487 cirurgias, em 12.423 internações, além de outras 250.820 consultas médicas ambulatoriais. (INCOR, 2020). A quase totalidade das informações referentes aos pacientes é registrada no Sistema Integrado de Informações Institucionais, o SI<sup>3</sup>.

#### 4.3 CARDIOPATIAS VALVARES E FILA CIRÚRGICA NO INCOR-HCFMUSP

Nkomo *et al.* (2006) chamam a atenção para o fato de que, embora as doenças valvares moderadas e graves raramente sejam tratadas como um problema de saúde pública, sua prevalência é relevante (2,5% da população dos Estados Unidos da América) e aumenta com o envelhecimento populacional. No Brasil e na América Latina, em que a febre reumática ainda é realidade, a doença valvar reumática é estimada entre 1 e 7 para cada 1.000 crianças e pode chegar a algo entre 10 e 70 caso a pesquisa seja feita com o uso de ecocardiografia. A estenose aórtica aterosclerótica/degenerativa acomete entre 3% e 5% da população acima de 75 anos; 2% da população apresenta estenose aórtica por etiologia bicúspide. (TARASOUTCHI *et al.*, 2020).

Silva (2018) estudou a fila cirúrgica para tratamento de valvopatias no INCOR-HCFMUSP; em 2017, havia 1.297 pacientes, com um tempo de espera estimado para realização do procedimento cirúrgico de 36 meses. Após revisão e qualificação dessa lista, 1.117 pacientes foram mantidos. O autor aponta que, das 579 cirurgias valvares realizadas em 2017 na instituição, 68% foram realizadas em regime eletivo e 32% na vigência de urgência ou emergência. Essa mesma fila de espera, consultada pelo autor do presente trabalho em abril de 2021, apontou a presença de 1.252 pacientes, o que evidencia um aumento de 12% no período.

#### 4.4 POPULAÇÃO E VARIÁVEIS

Foram utilizados dados não identificados de pacientes submetidos à cirurgia cardíaca valvar entre janeiro de 2010 e dezembro de 2019 no INCOR-HCFMUSP, motivo pelo qual não foi necessária a aplicação do Termo de Consentimento Livre e Esclarecido.

As informações utilizadas neste estudo foram extraídas da base de dados do Sistema Integrado de Informações Institucionais (SI<sup>3</sup>) (FURUIE *et al.*, 2007). Nessa primeira extração, foram acessados e tabulados em Excel os dados de interesse de 5.191 pacientes, o que corresponde a um total de 5.401 procedimentos cirúrgicos, uma vez que parte dos pacientes foram operados mais de uma vez no período.

Conforme proposto na literatura, a escolha das variáveis de qualquer modelo de AM é um passo fundamental. Embora competirá ao algoritmo buscar as correlações entre elas, muitas destas não evidentes até então, alimentar o modelo com variáveis não representativas ou que não tenham uma fundamentação teórica para a associação pode gerar ruídos e prejudicar seu desempenho, especialmente por meio do *overfitting* por correlações espúrias. (CHIAVEGATTO FILHO, 2015).

Para endereçar essa necessidade, o pesquisador participou de reuniões científicas da Unidade Clínica de Valvopatias do INCOR-HCFMUSP, além de ter discutido a pesquisa e as possíveis variáveis do modelo com os diretores médicos tanto do grupo cirúrgico quanto clínico de cardiopatias valvares. Como resultado, foram selecionadas, na fase inicial, 60 variáveis independentes (Tabela 3).

Importante destacar que, como o algoritmo tem por objetivo prever os casos que serão admitidos como urgência ou eletivos, foi necessário adotar um critério temporal a partir do qual esses rótulos seriam considerados. Para tanto, nesta pesquisa, adotou-se o momento da “Indicação Cirúrgica”, isto é, a data de “Entrada na Fila”, ocasião em que o paciente é formalmente registrado no sistema de informações do INCOR-HCFMUSP, como o momento a partir do qual a predição seria feita. Com isso, para cada paciente, apenas os dados das variáveis anteriores à data de entrada na fila foram considerados.

Tabela 3 – Variáveis independentes inicialmente consideradas para predição

<b>Tipo de Variável</b>	<b>Variável</b>
Demográfica	Idade
	Sexo
Estado de Saúde	Índice de Massa Corporal (IMC)
Ecocardiográfica	Diâmetro do Átrio Esquerdo
	Diâmetro Diastólico do Ventrículo Esquerdo
	Diâmetro Sistólico do Ventrículo Esquerdo
	Volume Sistólico
	Fração de Ejeção
	Área Valvar
	Pressão Sistólica de Artéria Pulmonar
Laboratorial	Albumina sérica
	Ureia
	Creatinina
	Hemoglobina
	Plaquetas
	INR
	Hemoglobina Glicada
Medicação	Captopril
	Enalapril
	Isossorbida
	Furosemida
	Hidroclorotiazida
	Clortalidona
	Espironolactona
	Atenolol
	Carvedilol
	Metoprolol
	Diltiazem
	Verapamil

	Digoxina
	Amiodarona
	Ácido Acetilsalicílico
	Clopidogrel
	Varfarina
	Quantidade de Medicamentos
Comorbidade	Reoperação
	Insuficiência Coronariana
	Diabetes
	Fibrilação Atrial e Arritmia
	Tabagismo
	Doença Pulmonar Obstrutiva Crônica
	Tontura e/ou síncope
Lesão Anatômica	Estenose Mitral
	Insuficiência Mitral Reumática
	Estenose Mitral com Insuficiência
	Estenose Aórtica Reumática
	Insuficiência Aórtica Reumática
	Estenose Mitral com Insuficiência
	Estenose Tricúspide
	Insuficiência Tricúspide
	Estenose Tricúspide com Insuficiência
	Transtornos de Valvas Mitral e Aórtica
	Transtornos de Ambas as Valvas Mitral e Tricúspide
	Transtornos de Ambas as Valvas Aórtica e Tricúspide
	Estenose Tricúspide com Insuficiência
	Insuficiência (da Valva) Mitral
	Prolapso (da Valva) Mitral
	Estenose (da Valva) Aórtica
	Insuficiência (da Valva) Aórtica
	Estenose (da Valva) Aórtica com Insuficiência

Fonte: produzida pelo autor

#### 4.5 CRITÉRIOS DE INCLUSÃO

Pacientes submetidos à cirurgia valvar no INCOR-HCFMUSP entre janeiro de 2010 e dezembro de 2019, tanto de forma eletiva quanto de urgência ou emergência.

Ter seguimento ambulatorial no INCOR-HCFMUSP anterior à realização da cirurgia.

#### 4.6 CRITÉRIOS DE EXCLUSÃO

Pacientes que realizaram cirurgia valvar de urgência, emergência ou eletiva no INCOR-HCFMUSP, mas que não faziam seguimento ambulatorial na Instituição antes da data da indicação da cirurgia.

Pacientes que realizaram cirurgia valvar de urgência ou emergência por conta de endocardite infecciosa.

#### 4.7 PRÉ-PROCESSAMENTO DE DADOS

A base de dados inicialmente extraída (5.401 procedimentos cirúrgicos valvares, com 60 variáveis independentes e uma dependente (tipo de cirurgia: urgência, emergência ou eletiva)) foi submetida ao pré-processamento, uma das etapas mais críticas e importantes que antecedem a modelagem do algoritmo. (BENHAR, IDRI, FERNÁNDEZ-ALEMÁN, 2020).

Como mencionado anteriormente, erros nessa fase podem comprometer de modo significativo seu desempenho. O pré-processamento consiste em um conjunto de técnicas aplicadas aos dados brutos com o intuito de potencializar o desempenho dos modelos que serão construídos. (BATISTA, CHIAVEGATTO FILHO, 2019). Na pesquisa em questão, fizeram parte do pré-processamento:

1. Exclusão dos casos que não preenchiam os critérios de inclusão no estudo: foram excluídos 808 casos em que o paciente não fazia seguimento prévio no INCOR-HCFMUSP e 177 cuja motivação da cirurgia foi uma complicação aguda de endocardite infecciosa.

2. Foram retirados 2.035 procedimentos cujos pacientes não possuíam registro, até a data da indicação da cirurgia, de nenhum dos códigos da Classificação Estatística Internacional de Doenças e Problemas Relacionados com a Saúde (CID-10), adotados como representativos da lesão anatômica da doença valvar. Outros 35 pacientes foram retirados durante o processo de validação dos dados por inconsistências encontradas (sem a data de entrada na fila de espera para cirurgia que correspondesse ao procedimento cirúrgico realizado, por exemplo). Dessas etapas, obteve-se o conjunto de dados de desenvolvimento com 2.346 procedimentos cirúrgicos valvares.

3. Balanceamento da amostra: a literatura aponta que a precisão dos algoritmos de AM, especialmente os supervisionados, pode ser afetada pela prevalência dos rótulos nas bases que servirão para treinamento e teste dos modelos. (SHAMOUT, ZHU, CLIFTON, 2020). Uma base é dita desbalanceada quando a frequência de um dos rótulos é muito maior que a do outro; quando os resultados de interesse estão em quantidades aproximadas, é dita balanceada. A amostra do estudo refletiu o observado no dia a dia das cirurgias valvares do INCOR-HCFMUSP, isto é, um predomínio de cirurgias eletivas em relação às de urgência. A fim de balancear a amostra, inicialmente, identificaram-se 294 casos de cirurgia de urgência; em seguida, foram selecionados, aleatoriamente, outros 294 de cirurgias eletivas.

4. Cálculo amostral: para os 588 casos validados de uma população de 2.346 casos e considerando intervalo de confiança de 99%, a margem de erro calculada foi de 4,6%. Ao se considerar um intervalo de confiança de 95%, a margem de erro será de 3,5%. Detalhes do cálculo podem ser consultados no Apêndice I.

5. Validação do rótulo: como mencionado anteriormente, o sucesso de um algoritmo de AM supervisionado começa na precisão da rotulagem do desfecho de interesse; por conta disso, a variável dependente (tipo de cirurgia) foi validada. Neste estudo, dado o objetivo do algoritmo desenvolvido, adotou-se uma adaptação da nomenclatura clássica proposta por Hurst e Siciliani (2003). Foi considerada cirurgia eletiva aquela programada, na qual o paciente foi convocado para o procedimento; já as cirurgias de urgência e emergência foram consideradas correspondentes e, por isso, foram rotuladas como “urgência”. Outra variável validada foi a reoperação. Na extração da base inicial de dados, foram elencados os possíveis códigos que

pudessem indicar a presença de cirurgia cardíaca valvar prévia. Isso foi feito tanto a partir de códigos da CID 10 (Z95.2 – Presença de prótese de válvula cardíaca; T82.0 – Complicação mecânica de prótese valvular cardíaca; T82.6 – Infecção e reação inflamatórias devidas à prótese valvular cardíaca) quanto a partir de códigos utilizados no sistema de informação institucional (30200200; 30201210; 30201000; 030200200; 030201000; 030201200; 030201500).

6. Manuseio das variáveis: para conferir representatividade a algumas variáveis, é relativamente comum que elas tenham de ser trabalhadas previamente. (SHAMOUT, ZHU, CLIFTON, 2020). Na pesquisa em questão, hemoglobina, hemoglobina glicada, albumina e IMC, variáveis inicialmente presentes como quantitativas, foram transformadas em variáveis categóricas. Para cada uma destas, foi definido um ponto de corte a partir do qual se assumiu uma interpretação (Tabela 4).

7. Normalização: a diferença de magnitude entre as variáveis pode comprometer a estabilidade e o aprendizado do modelo. Como exemplo, na amostra selecionada, enquanto a contagem de plaquetas variou entre 59.000 e 663.000, a creatinina variou entre 0,29 e 10,22. Para lidar com isso, usou-se normalização com Z-escore com média zero e desvio padrão 1 para alinhar todas as escalas das variáveis preditoras, de forma a assumirem uma distribuição normal.

8. Outra característica de grandes bases de dados são os dados faltantes (do inglês, *missing data*). Conforme Hulsen *et al.* (2019) chamam a atenção, o registro de dados na prática assistencial é muito menos rigoroso do que em estudos de coorte. Mais ainda, não necessariamente todos os pacientes se apresentam clinicamente com as mesmas queixas, existe variabilidade intra- e inter-profissionais, e a adoção de protocolos clínico-cirúrgicos ainda é um desafio em muitas instituições de saúde. (SHAMOUT, ZHU, CLIFTON, 2020).

Em relação à amostra do presente estudo, dos 588 casos, 584 (99%) apresentavam dados de creatinina e hemoglobina; 538 (91%) dispunham dos principais dados ecocardiográficos (desempenho diastólico e sistólico de ventrículo esquerdo, volume sistólico e fração de ejeção); em 359 (61%) foi estimada a pressão sistólica de artéria pulmonar; e 156 (27%) apresentavam a medida de área valvar. Dos exames laboratoriais, as variáveis menos frequentes foram a dosagem de albumina (137 – 23%) e de hemoglobina glicada (169 – 29%).



Quanto às medicações, a mais frequente foi a furosemida (424 – 72%), seguida pelo enalapril (271 – 46%); as menos frequentes foram clortalidona (19 – 3%), verapamil (20 – 3%) e clopidogrel (11 – 2%).

Para os códigos da CID de comorbidades, excetuando-se a reoperação que foi validada, o mais frequente foi tontura e/ou síncope (41 – 7,0%), seguido por insuficiência coronariana (28 – 4,8%); os menos frequentes foram doença pulmonar obstrutiva crônica (8 – 1,4%) e tabagismo (15 – 2,6%). Para os códigos da CID de diagnóstico de lesão anatômica, os mais comuns foram estenose da valva aórtica (35,0%) e estenose mitral (29,6%); os menos registrados foram estenose tricúspide (3 – 0,5%) e transtornos de ambas as valvas aórtica e tricúspide (3 – 0,5%).

Para tratar os dados ecocardiográficos e laboratoriais faltantes, foram criados cinco subagrupamentos com base nos diagnósticos anatômicos da lesão valvar. O objetivo foi formar grupos relativamente homogêneos, cujos valores das variáveis pudessem ser estimados de forma mais fidedigna. Grupo 1: estenose aórtica reumática, transtornos de valvas mitral e aórtica, transtornos de ambas as valvas aórtica e tricúspide e estenose (da valva) aórtica; Grupo 2: estenose mitral e transtornos de ambas as valvas mitral e tricúspide; Grupo 3: insuficiência aórtica reumática e insuficiência (da valva) aórtica; Grupo 4: insuficiência mitral reumática, insuficiência (da valva) mitral e prolapso (da valva) mitral; e Grupo 5: estenose tricúspide e insuficiência tricúspide.

Se o dado faltante foi de um paciente com qualquer diagnóstico do Grupo 1, ele foi então estimado a partir da mediana encontrada em todos os pacientes com registro do dado faltante desse grupo. Por exemplo, se um paciente com diagnóstico de estenose aórtica reumática (Grupo 1) não dispunha de dado de hemoglobina, este valor foi estimado olhando-se para todos os pacientes que tinham pelo menos um dos quatro diagnósticos do Grupo 1 (estenose aórtica reumática, transtornos de valvas mitral e aórtica, transtornos de ambas as valvas aórtica e tricúspide e estenose (da valva) aórtica) e calculando-se a mediana do dado de hemoglobina para esse grupo. O valor encontrado foi o imputado para o paciente inicialmente sem o registro. Cabe destacar que, no caso das medicações e dos códigos da CID, a ausência de registro é interpretada como não existência da doença ou da necessidade de medicação, não como variável faltante.

9. Redução da Dimensionalidade do Conjunto de Dados: consiste na identificação e exclusão de variáveis altamente correlacionadas (SANTOS *et al.*, 2019), assim como ajustes que confirmam maior representatividade para os dados. Das 60 variáveis iniciais, foi identificada correlação acima de 60% entre diâmetro diastólico do ventrículo esquerdo, volume sistólico e diâmetro sistólico do ventrículo esquerdo, motivo pelo qual se optou por deixar apenas esta última no modelo. Também foi vista alta correlação entre fibrilação e arritmia e o uso de varfarina, razão pela qual fibrilação e arritmia foram excluídas.

Para as variáveis de medicação, dada a baixa frequência de algumas delas, optou-se por agrupá-las em classes, a saber: captopril e enalapril constituíram a classe IECA; furosemida, hidroclorotiazida, clortalidona e espironolactona, a classe diurético; atenolol, carvedilol e metoprolol a classe betabloqueador; diltiazem e verapamil a classe bloqueador de canal de cálcio; ácido acetilsalicílico e clopidogrel a classe antiagregante plaquetário. Raciocínio semelhante foi utilizado para algumas classificações de lesão anatômica: os casos cuja classificação correspondia à estenose e insuficiência, concomitantes, foram desmembrados e computados em estenose e insuficiência, separados. Por exemplo, a CID Estenose Mitral com Insuficiência (I05.2), foi transformada em Estenose Mitral (I05.0) e Insuficiência Mitral Reumática (I05.1).

A Tabela 4 apresenta as 44 variáveis independentes e a variável dependente (tipo de cirurgia), com suas respectivas frequências e demais informações pertinentes, antes do tratamento dos dados faltantes. Na figura 5 observa-se a aplicação dos critérios de inclusão e exclusão do estudo a partir da população inicial (5401 cirurgias), até a validação dos dados e a construção da amostra final (588 cirurgias).

Tabela 4 – Variáveis dependente e independentes que compuseram os modelos de aprendizado de máquina

Variável	Frequência na Base de Dados	Uso da Variável no Modelo	Valores Possíveis	Observação
Idade	100,0%	Quantitativa	Variados	Representa a idade do paciente em anos, no momento da indicação da cirurgia.
Sexo	100,0%	Categórica	Masculino ou Feminino	Representa do gênero do paciente
IMC	92,0%	Categórica	<30 kg/m <sup>2</sup> ; ≥ 30 kg/m <sup>2</sup>	Calculado a partir do peso e da altura do paciente, desde que registrados em sistema antes da indicação da cirurgia. IMC = Peso/Altura <sup>2</sup> ; peso em quilos e altura em metros.
Tipo de Cirurgia	100,0%*	Categórica	Eletiva ou Urgência	Representa o rótulo de interesse, a variável dependente; *essa variável foi validada pelo pesquisador.
Diâmetro do Átrio Esquerdo	91,3%	Quantitativa	Variados	Dado do último exame disponível em sistema antes da indicação da cirurgia.
Diâmetro Sistólico do Ventrículo Esquerdo	91,3%	Quantitativa	Variados	Dado do último exame disponível em sistema antes da indicação da cirurgia.
Fração de Ejeção (Teicholz)	91,7%	Quantitativa	Variados	Dado do último exame disponível em sistema antes da indicação da cirurgia.
Área Valvar	26,7%	Quantitativa	Variados	Dado do último exame disponível em sistema antes da indicação da cirurgia.
Pressão Sistólica de Artéria Pulmonar	61,1%	Quantitativa	Variados	Dado do último exame disponível em sistema antes da indicação da cirurgia.
Albumina	23,3%	Categórica	<3,3; ≥ 3,3	Dado do último exame disponível em sistema antes da indicação da cirurgia.

Variável	Frequência na Base de Dados	Uso da Variável no Modelo	Valores Possíveis	Observação
Ureia	94,9%	Quantitativa	Variados	Dado do último exame disponível em sistema antes da indicação da cirurgia.
Creatinina	99,3%	Quantitativa	Variados	Dado do último exame disponível em sistema antes da indicação da cirurgia.
Hemoglobina	99,3%	Categórica	<10g/dl; ≥ 10g/dl	Dado do último exame disponível em sistema antes da indicação da cirurgia.
Plaquetas	94,9%	Quantitativa	Variados	Dado do último exame disponível em sistema antes da indicação da cirurgia.
INR	94,9%	Quantitativa	Variados	Dado do último exame disponível em sistema antes da indicação da cirurgia.
Hemoglobina Glicada	28,7%	Categórica	<7%; ≥ 7%	Dado do último exame disponível em sistema antes da indicação da cirurgia.
IECA	48,5%	Categórica	Sim ou Não	A presença dessa variável foi considerada positiva quando havia registro de prescrição ambulatorial de Captopril e/ou Enalapril, em sistema, anterior à indicação da cirurgia.
Isossorbida	5,3%	Categórica	Sim ou Não	A presença dessa variável foi considerada positiva quando havia registro de prescrição ambulatorial dessa medicação, em sistema, anterior à indicação da cirurgia.
Diurético	84,9%	Categórica	Sim ou Não	A presença dessa variável foi considerada positiva quando havia registro de prescrição ambulatorial de Furosemida e/ou Hidroclorotiazida e/ou Clortalidona e/ou Espironolactona, em sistema, anterior à indicação da cirurgia.

Variável	Frequência na Base de Dados	Uso da Variável no Modelo	Valores Possíveis	Observação
Betabloqueador	57,1%	Categórica	Sim ou Não	A presença dessa variável foi considerada positiva quando havia registro de prescrição ambulatorial de Atenolol e/ou Carvedilol e/ou Metoprolol, em sistema, anterior à indicação da cirurgia.
Bloqueador de Canal de Cálcio	13,9%	Categórica	Sim ou Não	A presença dessa variável foi considerada positiva quando havia registro de prescrição ambulatorial de Diltiazem e/ou Verapamil, em sistema, anterior à indicação da cirurgia.
Digoxina	37,9%	Categórica	Sim ou Não	A presença dessa variável foi considerada positiva quando havia registro de prescrição ambulatorial dessa medicação, em sistema, anterior à indicação da cirurgia.
Amiodarona	10,9%	Categórica	Sim ou Não	A presença dessa variável foi considerada positiva quando havia registro de prescrição ambulatorial dessa medicação, em sistema, anterior à indicação da cirurgia.
Antiagregante Plaquetário	35,2%	Categórica	Sim ou Não	A presença dessa variável foi considerada positiva quando havia registro de prescrição ambulatorial de Ácido Acetilsalicílico e/ou Clopidogrel, em sistema, anterior à indicação da cirurgia.
Varfarina	35,4%	Categórica	Sim ou Não	A presença dessa variável foi considerada positiva quando havia registro de prescrição ambulatorial dessa medicação, em sistema, anterior à indicação da cirurgia.

Variável	Frequência na Base de Dados	Uso da Variável no Modelo	Valores Possíveis	Observação
Quantidade de Medicações	44,2%	Categórica	<4; ≥ 4	A presença dessa variável foi considerada positiva quando havia registro de prescrição ambulatorial em sistema, anterior à indicação da cirurgia, de pelo menos quatro das nove medicações/classes de medicações analisadas: IECA, ISOSSORBIDA, DIURÉTICO, BETABLOQUEADOR, BLOQUEADOR DE CANAL DE CÁLCIO, DIGOXINA, AMIODARONA, ANTIAGREGANTE PLAQUETÁRIO e VARFARINA.
Reoperação	100,0%*	Categórica	Sim ou Não	A presença dessa variável foi considerada positiva nos casos em que o paciente já havia sido submetido à procedimento cardíaco valvar antes da indicação da cirurgia, tanto no INCOR-HCFMUSP quanto em outras instituições; *essa variável foi validada pelo pesquisador.
Insuficiência Coronariana	4,8%	Categórica	Sim ou Não	A presença dessa variável foi considerada positiva quando havia registro em sistema, anterior à indicação da cirurgia, de pelo menos um dos seguintes códigos da CID10: I25; I25.0; I25.1; I25.2; I25.5; I25.6; I25.8; I25.9.
Diabetes	4,4%	Categórica	Sim ou Não	A presença dessa variável foi considerada positiva quando havia registro em sistema, anterior à indicação da cirurgia, de pelo menos um dos seguintes códigos da CID10: E10; E11; E12; E13; E14.
Tabagismo	2,6%	Categórica	Sim ou Não	A presença dessa variável foi considerada positiva quando havia registro em sistema, anterior à indicação da cirurgia, de pelo menos um dos seguintes códigos da CID10: Z72.0.

Variável	Frequência na Base de Dados	Uso da Variável no Modelo	Valores Possíveis	Observação
Doença Pulmonar Obstrutiva Crônica	1,4%	Categórica	Sim ou Não	A presença dessa variável foi considerada positiva quando havia registro em sistema, anterior à indicação da cirurgia, de pelo menos um dos seguintes códigos da CID10: J44; J44.0; J44.1; J44.8; J44.9.
Tontura e/ou síncope	7,0%	Categórica	Sim ou Não	A presença dessa variável foi considerada positiva quando havia registro em sistema, anterior à indicação da cirurgia, de pelo menos um dos seguintes códigos da CID10: R55; R42.
Estenose Mitral	29,6%	Categórica	Sim ou Não	A presença dessa variável foi considerada positiva quando havia registro em sistema, anterior à indicação da cirurgia, dos códigos da CID10: I05.0 e/ou I05.2 (Estenose Mitral com Insuficiência).
Insuficiência Mitral Reumática	23,0%	Categórica	Sim ou Não	A presença dessa variável foi considerada positiva quando havia registro em sistema, anterior à indicação da cirurgia, dos códigos da CID10: I05.1 e/ou I05.2 (Estenose Mitral com Insuficiência).
Estenose Aórtica Reumática	7,8%	Categórica	Sim ou Não	A presença dessa variável foi considerada positiva quando havia registro em sistema, anterior à indicação da cirurgia, dos códigos da CID10: I06.0 e/ou I06.2 (Estenose Aórtica Reumática com Insuficiência).
Insuficiência Aórtica Reumática	9,0%	Categórica	Sim ou Não	A presença dessa variável foi considerada positiva quando havia registro em sistema, anterior à indicação da cirurgia, dos códigos da CID10: I06.1 e/ou I06.2 (Estenose Aórtica Reumática com Insuficiência).

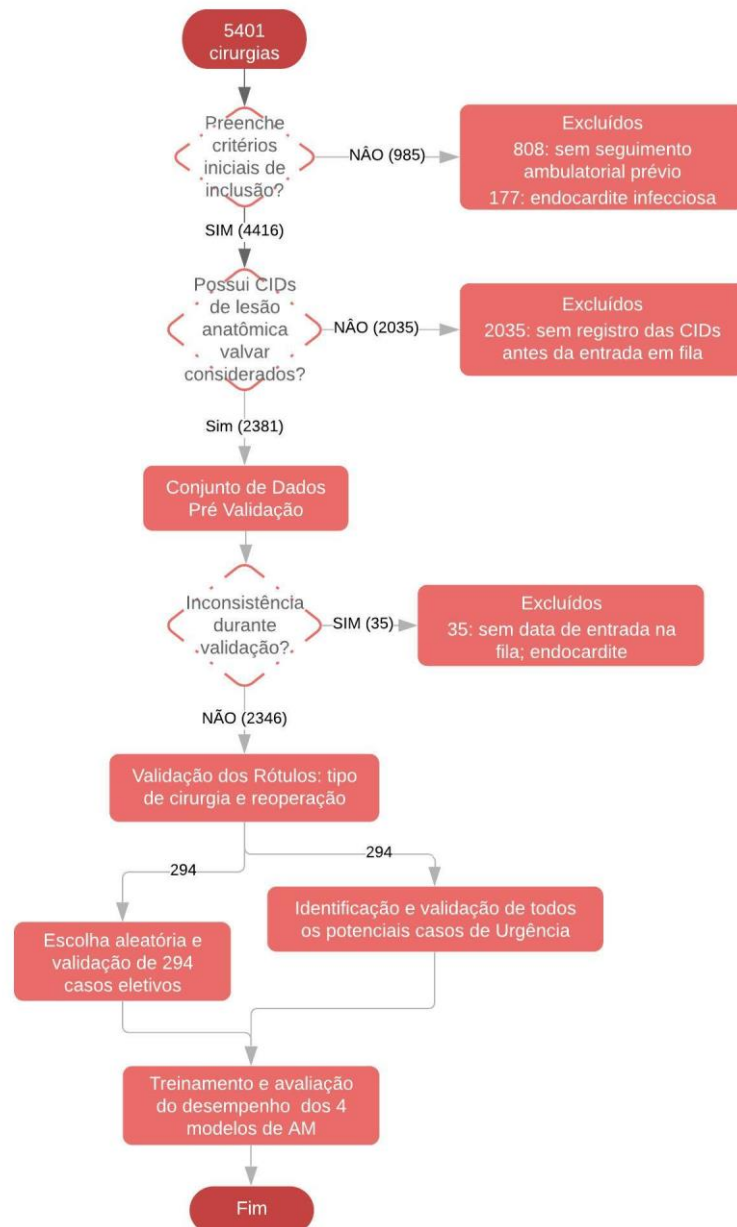
Variável	Frequência na Base de Dados	Uso da Variável no Modelo	Valores Possíveis	Observação
Estenose Tricúspide	0,5%	Categórica	Sim ou Não	A presença dessa variável foi considerada positiva quando havia registro em sistema, anterior à indicação da cirurgia, dos códigos da CID10: I07.0 e/ou I07.2 (Estenose Tricúspide com Insuficiência).
Insuficiência Tricúspide	6,1%	Categórica	Sim ou Não	A presença dessa variável foi considerada positiva quando havia registro em sistema, anterior à indicação da cirurgia, dos códigos da CID10: I07.1 e/ou I07.2 (Estenose Tricúspide com Insuficiência).
Transtornos de Valvas Mitral e Aórtica	3,7%	Categórica	Sim ou Não	A presença dessa variável foi considerada positiva quando havia registro em sistema, anterior à indicação da cirurgia, dos códigos da CID10: I08.0 e/ou I08.3 (Transtornos das Valvas Mitral, Aórtica e Tricúspide).
Transtornos de Ambas as Valvas Mitral e Tricúspide	0,7%	Categórica	Sim ou Não	A presença dessa variável foi considerada positiva quando havia registro em sistema, anterior à indicação da cirurgia, dos códigos da CID10: I08.1 e/ou I08.3 (Estenose Tricúspide com Insuficiência).
Transtornos de Ambas as Valvas Aórtica e Tricúspide	0,5%	Categórica	Sim ou Não	A presença dessa variável foi considerada positiva quando havia registro em sistema, anterior à indicação da cirurgia, dos códigos da CID10: I08.2 e/ou I08.3 (Estenose Tricúspide com Insuficiência).
Insuficiência (da Valva) Mitral	18,9%	Categórica	Sim ou Não	A presença dessa variável foi considerada positiva quando havia registro em sistema, anterior à indicação da cirurgia, do código da CID10: I34.0.
Prolapso (da Valva) Mitral	4,9%	Categórica	Sim ou Não	A presença dessa variável foi considerada positiva quando havia registro em sistema, anterior à indicação da cirurgia, do código da CID10: I34.1.



Variável	Frequência na Base de Dados	Uso da Variável no Modelo	Valores Possíveis	Observação
Estenose (da Valva) Aórtica	35,0%	Categórica	Sim ou Não	A presença dessa variável foi considerada positiva quando havia registro em sistema, anterior à indicação da cirurgia, dos códigos da CID10: I35.0 e/ou I35.2 (Estenose (da Valva) Aórtica com Insuficiência).
Insuficiência (da Valva) Aórtica	17,7%	Categórica	Sim ou Não	A presença dessa variável foi considerada positiva quando havia registro em sistema, anterior à indicação da cirurgia, dos códigos da CID10: I35.1 e/ou I35.2 (Estenose (da Valva) Aórtica com Insuficiência).

Fonte: produzida pelo autor

Figura 5 – Fluxograma para definição da amostra de treinamento e avaliação dos modelos.



Fonte: produzida pelo autor

#### 4.8 ALGORITMOS

Embora não exista uma regra de ouro para escolha dos modelos de aprendizado de máquina, a literatura chama a atenção para alguns aspectos: i) modelos mais complexos não necessariamente entregam melhores resultados; ii) o desempenho de um determinado algoritmo é função do problema que está sendo endereçado, ou seja, o algoritmo X pode ser ruim para o problema A, mas ótimo

para o problema B, motivo pelo qual é interessante que mais de um algoritmo seja utilizado em cada contexto; e iii) existe uma tendência a se utilizar, na saúde, modelos baseados em árvore, com bons resultados, especialmente os que fazem uso de técnicas de comitê de modelos (do inglês, *ensemble*). (JAMES *et al.*, 2014; BATISTA, CHIAVEGATTO FILHO, 2019; SANTOS *et al.*, 2019).

No estudo que se apresenta, foram utilizados quatro diferentes modelos classificatórios de AM: Regressão Logística, Árvores de Decisão, Florestas Aleatórias (do inglês *Random Forests*) e K-vizinhos mais próximos (do inglês, *K-Nearest Neighbors*, KNN).

#### 4.8.1 REGRESSÃO LOGÍSTICA

A regressão logística é tida como um dos modelos de classificação mais simples e é frequentemente utilizada como padrão de referência em relação a modelos mais sofisticados. (CHRISTODOULOU *et al.*, 2019). Como limitação, não é capaz de identificar correlações não lineares entre as variáveis de entrada, assim como assume independência entre elas. (SHAMOUT, ZHU, CLIFTON, 2020).

O objetivo desse modelo é estabelecer uma relação entre a probabilidade de ocorrência da variável de interesse (categórica) e as variáveis preditoras. No presente estudo, o evento de interesse é cirurgia de “urgência”, motivo pelo qual ela assume o valor 1 (classe positiva) e cirurgia “eletiva” assume o valor 0 (classe negativa).

De modo sintético, podemos escrever que a probabilidade de ocorrência da classe positiva (P) é dada por:

$P(y = 1|Z) = 1 / 1 + e^{(-Z)}$ , em que Z pode ser representado pela função:

$Z = \alpha + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_k X_{ik}$ , em que  $X_{i1}, X_{i2}, \dots, X_{ik}$  representam cada uma das variáveis preditoras,  $\alpha$  e  $\beta_1, \beta_2, \dots, \beta_k$  são os parâmetros estimados pelo modelo.

Isto posto, podemos dizer que compete ao modelo de regressão logística buscar a melhor combinação possível entre os coeficientes  $\alpha$  e  $\beta$ . Quanto melhor for essa combinação, melhor será o desempenho do algoritmo de aprendizado de máquina.

#### 4.8.2 ÁRVORES DE DECISÃO

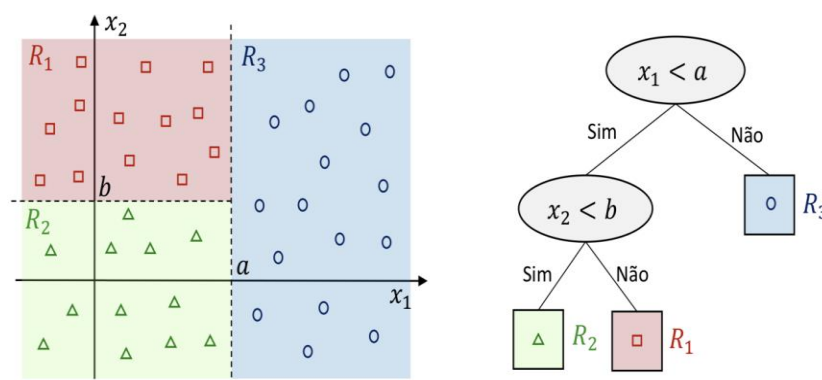
As árvores de decisão se assemelham muito às deliberações que tomamos no nosso cotidiano e na prática médica. Imaginemos uma situação em que a grande maioria das pessoas já vivenciou: entorse de tornozelo. Nem sempre, quando passamos por uma situação como essa, buscamos auxílio médico. E, normalmente, temos nossa saúde restabelecida mesmo sem a ajuda profissional.

Não é esdrúxulo imaginar que qualquer um diante de uma entorse faria algumas perguntas para decidir visitar ou não um médico: está doendo? Está roxo (equimose)? Está inchado (edema)? Limita a movimentação? Isso aconteceu com um jovem de 18 anos ou com uma senhora de 85 anos? A partir do conjunto de respostas e combinações possíveis, tomamos a decisão de ir ou não a um ortopedista. Embora possa parecer um problema um tanto quanto simples, pesquisadores desenvolveram um conjunto de regras de decisão (uma árvore de decisão) para prever fraturas de tornozelo em pacientes que comparecem à emergência dos hospitais com queixas como as descritas anteriormente, os critérios de Ottawa, auxiliando os médicos na tomada de decisão sobre quem deve ou não fazer radiografia de tornozelo. (STIELL *et al.*, 1992; LIU *et al.*, 2019).

Árvores de decisão são estruturas hierárquicas cujo objetivo é, a partir de um conjunto de dados iniciais, tomar sucessivas decisões que resultem na formação de grupos cada vez menores e mais homogêneos. Uma árvore é constituída por nós (que representam as variáveis do modelo), nos quais uma decisão deve ser tomada. A partir dessa decisão, um ramo (caminho) é aberto. Uma vez esgotados os nós e os ramos, chega-se a uma folha, que nada mais é do que o rótulo da classe predita (eletiva ou urgência, por exemplo). (BATISTA, CHIAVEGATTO FILHO, 2019; BENHAR, IDRI, FERNÁNDEZ-ALEMÁN, 2020).

Na Figura 6, temos uma representação de uma árvore de decisão utilizada para classificar um conjunto de dados.

Figura 6 – Representação de uma árvore de decisão. À direita observamos a árvore de decisão, em que  $x_1$  e  $x_2$  representam as variáveis preditoras do modelo. Conforme o valor encontrado, um ramo é seguido e pode-se levar a outro nó ou a um rótulo ( $R_1$ ,  $R_2$  ou  $R_3$ ). À esquerda temos uma ilustração gráfica dessa árvore.



Fonte: Mayrink (2016)

As principais vantagens das árvores de decisão são que apresentam alto grau de interpretação, sua implementação é relativamente simples e, no caso do AM, compete ao algoritmo decidir quais variáveis farão parte do modelo, isto é, ele é capaz de identificar e descartar aquelas que não trazem informações relevantes para o processo de decisão. Em relação às desvantagens, trata-se de um modelo com forte tendência ao sobreajuste, uma vez que ele é capaz de estruturar, na fase de treinamento, modelos altamente complexos, mas que ficam sensíveis a pequenas variações de dados.

#### 4.8.3 FLORESTAS ALEATÓRIAS

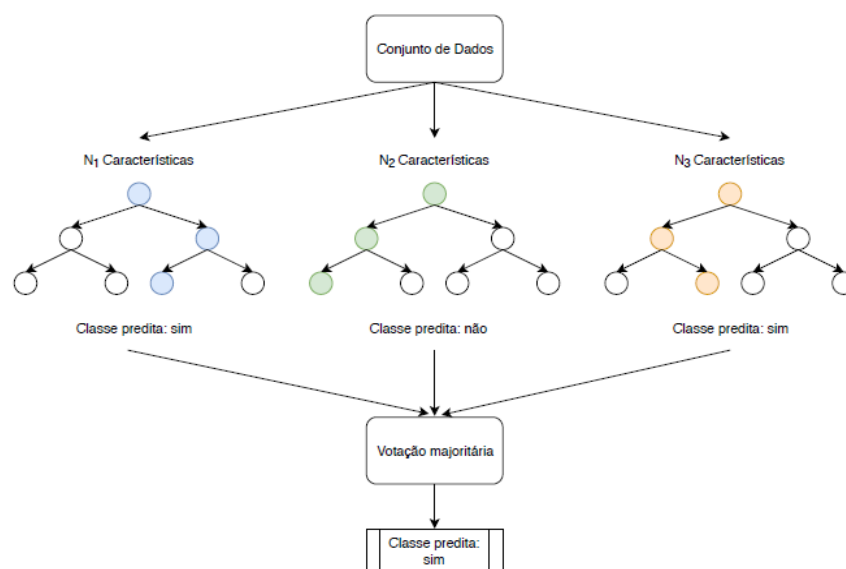
Os algoritmos de Florestas Aleatórias surgem como alternativa para endereçar as desvantagens das árvores de decisão, principalmente o sobreajuste (do inglês *overfitting*). Para tanto, nas Florestas Aleatórias, são utilizadas duas técnicas principais: *bagging* e *bootstrap*.

*Bagging* consiste em sortear aleatoriamente o conjunto de dados que fará parte do treinamento de cada árvore de decisão, expondo as árvores a diferentes conjuntos de dados. *Bootstrap* consiste em sortear as variáveis de cada nó. Com essas duas técnicas, espera-se reduzir o efeito que algumas variáveis “fortes” podem ter no desempenho do algoritmo, já que elas não estarão em várias das

árvores constituídas, ocorrendo o mesmo em relação às observações que não estarão 100% presentes em todos os treinamentos. (HASTIE, TIBSHIRANI, FRIEDMAN, 2008).

A Figura 7 ilustra um algoritmo de Florestas Aleatórias. Como se observa, diversas árvores de decisão são criadas, cada uma delas leva a um desfecho (rótulo). A reunião desse conjunto de árvores e respectivas predições constitui um comitê de predições (*ensemble*). O resultado final do algoritmo será dado pela votação da classe mais frequente.

Figura 7 – Algoritmo de Florestas Aleatórias. A partir de um conjunto de dados, diversas árvores de decisão são estruturadas. Tanto os subconjuntos de dados quanto as variáveis que entrarão em cada árvore são sorteados. Cada uma dessas árvores levará a um rótulo. Ao final, o resultado do algoritmo será dado pela votação da classe majoritária.



Fonte: Batista e Chiavegatto Filho (2019).

#### 4.8.4 K-VIZINHOS MAIS PRÓXIMOS

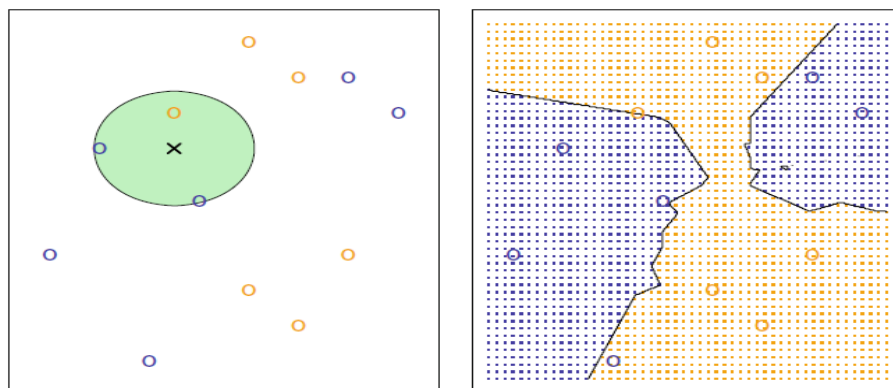
O algoritmo KNN é um modelo que se baseia em instâncias (exemplos, observações prévias), com bons resultados para situações em que os preditores e a resposta de interesse requerem um modelo mais flexível. Diferente dos outros três algoritmos vistos, em que os dados de treinamento são utilizados para construir uma função ajustada a seus preditores, no KNN, os dados de treinamento são armazenados, juntamente com seus rótulos, num plano n-dimensional. Quando uma

nova observação é apresentada ao algoritmo, ele compara a posição dessa nova observação (definida a partir de suas variáveis) com as K (número inteiro e ímpar) instâncias mais próximas da base de treinamento. A nova observação será classificada com o rótulo mais frequente entre os K rótulos mais próximos. (JAMES *et al.*, 2014; SANTOS, 2018).

A confecção desse algoritmo depende de duas definições essenciais. A primeira é o número K (1, 3, 5, ...) de vizinhos mais próximos que serão utilizados para constituir a vizinhança da nova observação, visto que K assume o papel de hiperparâmetro para ajuste do modelo: é o valor de K que determinará a relação viés-variância do algoritmo. A segunda é a medida de distância que será empregada para delimitar esse espaço n-dimensional em torno da nova observação: a distância euclidiana é um exemplo de medida frequentemente empregada para calcular a distância espacial entre um objeto e todos os outros. (BATISTA, CHIAVEGATTO FILHO, 2019).

Na Figura 8, temos um exemplo de algoritmo KNN.

Figura 8 – Algoritmo KNN. Nesse exemplo, temos seis observações em azul e seis observações em laranja, as quais ocupam um espaço bidimensional. A partir da definição de  $K = 3$ , na figura da esquerda, verificamos como ocorre a classificação de uma nova observação (x, em preto). As três observações mais próximas a x são identificadas, com seus respectivos rótulos (duas são azuis e uma é laranja). Nesse caso, a classificação será a do rótulo identificado em maior frequência, ou seja, x será classificado como azul. Na figura da direita, temos a representação da fronteira de decisão (linha preta), com  $K = 3$ , para todas possíveis posições de x, delimitando as áreas em que x será classificado como azul e as áreas em que será classificado como laranja.



Fonte: James *et al.* (2014).

#### 4.9 APRENDIZADO E ANÁLISE DE DESEMPENHO DOS MODELOS

Concluído o pré-processamento de dados e feita a escolha dos quatro algoritmos supramencionados, procedemos, então, com o aprendizado propriamente dito, a partir do qual foram gerados dados para atender a dois objetivos principais e distintos: i) avaliar o desempenho de cada algoritmo, identificando o melhor; e ii) uma vez definido o algoritmo final, estimar seu erro de generalização (erro de predição) diante a novas observações. (HASTIE, TIBSHIRANI, FRIEDMAN, 2008).

Em situações de abundância de dados, Hastie, Tibshirani e Fiedman (2008) propõem que o conjunto inicial de dados seja subdividido, randomicamente, em três partes: base de treinamento (50% dos dados, utilizada para ajustar o algoritmo), base de validação (25% dos dados, usada para avaliar o desempenho e o erro de predição do modelo construído) e base de teste (25% dos dados, utilizada para avaliar o erro de generalização do melhor modelo identificado).

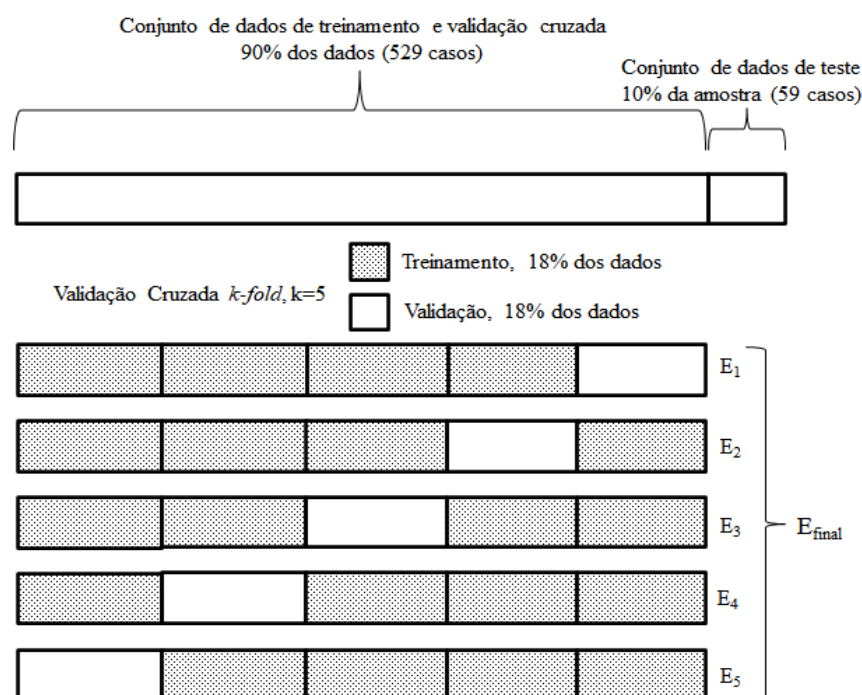
Embora o presente estudo não tenha sido edificado num contexto de escassez de dados, a quantidade de observações do conjunto de dados inicial (588) não representa abundância, especialmente quando se considera o campo de *Big Data* e aprendizado de máquina, motivo pelo qual se adotou a técnica de reamostragem conhecida como validação cruzada *k-fold*, uma das mais utilizadas em problemas de AM.

Nesse método, o conjunto inicial de dados é subdividido aleatoriamente em duas partes, uma que será utilizada para treinamento e validação cruzada e outra que ficará para o teste. A base de treinamento e validação cruzada é então subdividida em  $k$  partes iguais ( $k = 1, 2, 3, \dots$ ):  $k-1$  partes serão utilizadas para treinar o modelo e a parte restante será utilizada para fazer sua validação. Esse processo é repetido até que todas as  $k$  partes tenham sido utilizadas para treinamento e validação, e o desempenho final será dado pela média da execução de cada uma dessas iterações. (HASTIE, TIBSHIRANI, FRIEDMAN, 2008).

No presente estudo, 90% do conjunto de dados inicial foi destinado para treinamento e validação sob a técnica *k-fold*, e 10% ficou como base de teste. Adotamos  $k = 5$ . A Figura 9 ilustra esse processo.



Figura 9 – Divisão de dados e validação cruzada k-fold do estudo.  $E_1$ ,  $E_2$ ,  $E_3$ ,  $E_4$  e  $E_5$  representam o desempenho em cada uma das iterações do processo de validação cruzada do modelo.  $E_{final}$  representa o desempenho final, dado pela média do desempenho de cada iteração.



Fonte: produzida pelo autor

A frequência de cada variável independente nas bases de treinamento, validação e teste, após todo o processo de pré-processamento de dados, inclusive com o tratamento de dados faltantes, é apresentada na Tabela 5.

Tabela 5 – Distribuição das variáveis dependentes nas bases de treinamento, validação e teste utilizadas no processo de aprendizagem e avaliação de desempenho dos quatro algoritmos do estudo. DP: desvio padrão.

Variável Dependente	Treinamento	Validação	Teste
Idade média (dp)	57,1 (14,6)	55,6 (14)	55,1 (14,2)
Sexo			
Feminino n (%)	185 (43,7%)	53 (50%)	28 (47,5%)
Masculino n (%)	238 (56,3%)	53 (50%)	31 (52,5%)
IMC			
< 30 kg/m <sup>2</sup> n (%)	360 (85,1%)	85 (80,2%)	46 (78%)
>= 30 kg/m <sup>2</sup> n (%)	63 (14,9%)	21 (19,8%)	13 (22%)
Diâmetro do Átrio Esquerdo média (dp)	49,6 (8,5)	49,8 (8)	48,3 (6,9)
Diâmetro Sistólico do Ventrículo Esquerdo média (dp)	35,4 (6,3)	36,7 (6,4)	35,6 (6)
Fração de Ejeção (Teicholz) média (dp)	62,6 (5,1)	62 (5,5)	61,7 (5,2)

Variável Dependente	Treinamento	Validação	Teste
Área Valvar média (dp)	1,4 (0,3)	1,4 (0,3)	1,4 (0,3)
Pressão Sistólica de Artéria Pulmonar média (dp)	43,3 (10,7)	42,3 (11)	44,1 (11,9)
Albumina sérica < 3,3 n (%) ≥ 3,3 n (%)	142 (33,6%) 281 (66,4%)	39 (36,8%) 67 (63,2%)	23 (39%) 36 (61%)
Ureia média (dp)	40,8 (11,6)	40,4 (12)	38,8 (11,3)
Creatinina média (dp)	1 (0,2)	1 (0,2)	1 (0,2)
Hemoglobina < 10 g/dl n (%) ≥ 10 g/dl n (%)	35 (8,3%) 388 (91,7%)	8 (7,5%) 98 (92,5%)	6 (10,2%) 53 (89,8%)
Plaquetas média (dp)	208.370 (51.446,7)	216.523,6 (51.747,1)	211.118,6 (53.054,3)
INR média (dp)	1,3 (0,4)	1,2 (0,4)	1,2 (0,3)
Hemoglobina Glicada < 7% n (%) ≥ 7% n (%)	403 (95,3%) 20 (4,7%)	102 (96,2%) 4 (3,8%)	55 (93,2%) 4 (6,8%)
IECA Não n (%) Sim n (%)	219 (51,8%) 204 (48,2%)	54 (50,9%) 52 (49,1%)	30 (50,8%) 29 (49,2%)
Isossorbida Não n (%) Sim n (%)	401 (94,8%) 22 (5,2%)	101 (95,3%) 5 (4,7%)	55 (93,2%) 4 (6,8%)
Diurético Não n (%) Sim n (%)	63 (14,9%) 360 (85,1%)	17 (16%) 89 (84%)	9 (15,3%) 50 (84,7%)
Betabloqueador Não n (%) Sim n (%)	187 (44,2%) 236 (55,8%)	45 (42,5%) 61 (57,5%)	20 (33,9%) 39 (66,1%)
Bloqueador Canal de Cálcio Não n (%) Sim n (%)	360 (85,1%) 63 (14,9%)	94 (88,7%) 12 (11,3%)	52 (88,1%) 7 (11,9%)
Digoxina Não n (%) Sim n (%)	260 (61,5%) 163 (38,5%)	64 (60,4%) 42 (39,6%)	41 (69,5%) 18 (30,5%)
Amiodarona Não n (%) Sim n (%)	379 (89,6%) 44 (10,4%)	91 (85,8%) 15 (14,2%)	54 (91,5%) 5 (8,5%)
Antiagregante Plaquetário Não n (%) Sim n (%)	279 (66%) 144 (34%)	59 (55,7%) 47 (44,3%)	43 (72,9%) 16 (27,1%)
Varfarina Não n (%) Sim n (%)	273 (64,5%) 150 (35,5%)	69 (65,1%) 37 (34,9%)	38 (64,4%) 21 (35,6%)

<b>Variável Dependente</b>	<b>Treinamento</b>	<b>Validação</b>	<b>Teste</b>
Quantidade de medicações < 4 n (%) ≥ 4 n (%)	238 (56,3%) 185 (43,7%)	55 (51,9%) 51 (48,1%)	35 (59,3%) 24 (40,7%)
Reoperação Não n (%) Sim n (%)	253 (59,8%) 170 (40,2%)	59 (55,7%) 47 (44,3%)	35 (59,3%) 24 (40,7%)
Insuficiência Coronariana Não n (%) Sim n (%)	405 (95,7%) 18 (4,3%)	100 (94,3%) 6 (5,7%)	55 (93,2%) 4 (6,8%)
Diabetes Não n (%) Sim n (%)	404 (95,5%) 19 (4,5%)	99 (93,4%) 7 (6,6%)	59 (100%) 0 (0%)
Tabagismo Não n (%) Sim n (%)	415 (98,1%) 8 (1,9%)	100 (94,3%) 6 (5,7%)	58 (98,3%) 1 (1,7%)
Doença Pulmonar Obstrutiva Crônica Não n (%) Sim n (%)	416 (98,3%) 7 (1,7%)	105 (99,1%) 1 (0,9%)	59 (100%) 0 (0%)
Tontura e/ou síncope Não n (%) Sim n (%)	390 (92,2%) 33 (7,8%)	99 (93,4%) 7 (6,6%)	58 (98,3%) 1 (1,7%)
Estenose Mitral Não n (%) Sim n (%)	304 (71,9%) 119 (28,1%)	74 (69,8%) 32 (30,2%)	36 (61%) 23 (39%)
Insuficiência Mitral Reumática Não n (%) Sim n (%)	336 (79,4%) 87 (20,6%)	74 (69,8%) 32 (30,2%)	43 (72,9%) 16 (27,1%)
Estenose Aórtica Reumática Não n (%) Sim n (%)	388 (91,7%) 35 (8,3%)	98 (92,5%) 8 (7,5%)	56 (94,9%) 3 (5,1%)
Insuficiência Aórtica Reumática Não n (%) Sim n (%)	388 (91,7%) 35 (8,3%)	94 (88,7%) 12 (11,3%)	53 (89,8%) 6 (10,2%)
Estenose Tricúspide Não n (%) Sim n (%)	420 (99,3%) 3 (0,7%)	106 (100%) 0 (0%)	59 (100%) 0 (0%)
Insuficiência Tricúspide Não n (%) Sim n (%)	401 (94,8%) 22 (5,2%)	98 (92,5%) 8 (7,5%)	53 (89,8%) 6 (10,2%)
Transtornos de Valvas Mitral e Aórtica Não n (%) Sim n (%)	404 (95,5%) 19 (4,5%)	103 (97,2%) 3 (2,8%)	59 (100%) 0 (0%)

Variável Dependente	Treinamento	Validação	Teste
Transtornos de Ambas as Valvas Mitral e Tricúspide Não n (%) Sim n (%)	419 (99,1%) 4 (0,9%)	106 (100%) 0 (0%)	59 (100%) 0 (0%)
Transtornos de Ambas as Valvas Aórtica e Tricúspide Não n (%) Sim n (%)	420 (99,3%) 3 (0,7%)	106 (100%) 0 (0%)	59 (100%) 0 (0%)
Insuficiência (da Valva) Mitral Não n (%) Sim n (%)	347 (82%) 76 (18%)	81 (76,4%) 25 (23,6%)	49 (83,1%) 10 (16,9%)
Prolapso (da Valva) Mitral Não n (%) Sim n (%)	401 (94,8%) 22 (5,2%)	101 (95,3%) 5 (4,7%)	57 (96,6%) 2 (3,4%)
Estenose (da Valva) Aórtica Não n (%) Sim n (%)	269 (63,6%) 154 (36,4%)	73 (68,9%) 33 (31,1%)	40 (67,8%) 19 (32,2%)
Insuficiência (da Valva) Aórtica Não n (%) Sim n (%)	349 (82,5%) 74 (17,5%)	88 (83%) 18 (17%)	47 (79,7%) 12 (20,3%)

Fonte: produzida pelo autor.

No presente trabalho, os quatro algoritmos foram testados a fim de medir a capacidade de cada um deles de diferenciar as classes de interesse (urgência *versus* eletiva) diante de um conjunto de dados até então desconhecidos (base de teste).

Especificamente nos modelos de classificação, existem dois tipos de predição que ocorrem simultaneamente e se relacionam: a primeira é a probabilidade de uma determinada classe ocorrer (80% de chance dessa classe ser eletiva e 20% de ser urgência, por exemplo), e a segunda é a predição categórica, ou seja, apontar que uma determinada classe é urgência ou eletiva a partir do limiar (*threshold*, em inglês) de probabilidade de ocorrência estimada estabelecido (se a probabilidade de ocorrência da classe urgência for 60% (ou 40%, ou 90%), o algoritmo a classificará como urgência; se não, classificará como eletiva). (SANTOS, 2018).

Neste estudo, o limiar adotado foi 50%, valor comumente utilizado em problemas de AM. (JAMES *et al.*, 2014). Algumas métricas de discriminação são utilizadas a partir da comparação entre as respostas preditas (dadas pelo algoritmo) e as respostas observadas (dados da realidade). Via de regra esses dados são

apresentados na forma de uma matriz de confusão (Figura 10). (HASTIE, TIBSHIRANI, FRIEDMAN, 2008; BATISTA, CHIAVEGATTO FILHO, 2019; SHAMOUT, ZHU, CLIFTON, 2020):

Figura 10 – Estrutura de uma matriz de confusão utilizada para avaliar o desempenho de algoritmos de classificação. VP: verdadeiro positivo; FP: falso positivo; VN: verdadeiro negativo; FN: falso negativo.

		Resposta Predita	
		Negativa	Positiva
Resposta Observada	Negativa	Verdadeiro Negativo (VN)	Falso Positivo (FP)
	Positiva	Falso Negativo (FN)	Verdadeiro Positivo (VP)

Fonte: Batista e Chiavegatto Filho (2019)

- Sensibilidade (*recall*, taxa de verdadeiro positivo (TVP)): probabilidade de uma classe sabidamente positiva ser identificada como positiva pelo algoritmo. Cálculo:  $TVP = VP/(VP+FN)$ .

- Especificidade (taxa de verdadeiro negativo (TVN)): probabilidade de uma classe sabidamente negativa ser identificada como negativa pelo modelo. Cálculo:  $TVN = VN/(VN+FP)$ .

- Taxa de Falso Positivo (*fall-out*): probabilidade de uma classe sabidamente negativa ser identificada como positiva pelo modelo. Cálculo:  $TFP = FP/(FP+VN)$ .

- Taxa de Falso Negativo (*false omission rate*): probabilidade de uma classe sabidamente positiva ser identificada como negativa pelo modelo. Cálculo:  $TFN = FN/(VN+FN)$ .

- Precisão (valor preditivo positivo (VPP)): dado que o algoritmo classificou o resultado como positivo, qual é a probabilidade de ele realmente ser positivo. Cálculo:  $VPP = VP/(VP+FP)$ .

- Acurácia (ACC): proporção de resultados corretamente classificados, ou seja, porcentagem de casos negativos identificados como negativos e positivos identificados como positivos pelo algoritmo. Cálculo:  $ACC = (VP+VN)/(VP+VN+FP+FN)$ .

- F1-score: média harmônica entre precisão e sensibilidade. Cálculo:  $F1 = 2 \times (\text{Precisão} \times \text{Sensibilidade}) / (\text{Precisão} + \text{Sensibilidade})$ . A média harmônica é interessante de ser usada nesse caso por se tratarem de duas grandezas inversamente proporcionais. Em qualquer modelo de saúde, quanto maior a sensibilidade do teste, maior também será a taxa de falso positivo e menor será, por conseguinte, a precisão. Na prática, quanto mais próximo de 1 for o F1-score (modelo perfeito), mais o algoritmo estará equilibrado em relação à sensibilidade e à precisão.

Uma outra forma de visualizar o desempenho geral do modelo é por meio da curva *Receiver Operating Characteristic* (ROC). Essa curva é construída a partir dos diferentes limiares de probabilidade de ocorrência da resposta de interesse, para os quais estarão associados um respectivo valor de sensibilidade e de especificidade. A partir do cálculo da área sob a curva formada (*area under the receiver operating characteristic* (AUROC)), podemos estimar a capacidade discriminatória do algoritmo.

Uma AUROC igual a 0,5 mostra ausência de poder discriminatório do modelo. A partir de 0,5, quanto mais próximo de 1 for o valor da AUROC, melhor será seu desempenho classificatório. Para Hosmer e Lemeshow (2000), modelos classificatórios com AUROC entre 0,7 e 0,8 são considerados aceitáveis; entre 0,8 e 0,9 são excelentes preditores; acima de 0,9 são excepcionais, embora extremamente raros. Na mesma linha, Smith *et al.* (2008) sugerem que modelos com AUROC maior que 0,8 são bons classificadores; acima de 0,9, apresentam excelente capacidade discriminatória.

## 5. RESULTADOS

### 5.1 VALIDAÇÃO DAS VARIÁVEIS

Conforme descrito no método, duas variáveis passaram pelo processo de validação: tipo de cirurgia e reoperação. Em relação ao tipo de cirurgia, na base pré-validação, das 588 cirurgias, 502 (85,3%) estavam classificadas como eletiva, 69 (11,7%) como urgência e 17 (3%) como emergência; como, neste trabalho, cirurgia de urgência e emergência foram consideradas equivalentes, pode-se então dizer que 85,3% eram eletivas e 14,7% de urgência. A Tabela 6 apresenta o grau de concordância de registro entre o pré (registro em sistema de informação) e o pós validação (dado corrigido pelo autor). O erro mais comum foi o registro de cirurgia de urgência como eletiva em 226 casos (38,4% da amostra).

Tabela 6 – Acurácia da classificação do tipo de cirurgia

Pré-Validação	Pós-Validação	Frequência Absoluta	Frequência Relativa
ELETIVA	ELETIVA	276	46,9%
ELETIVA	URGÊNCIA	226	38,4%
URGÊNCIA	URGÊNCIA	60	10,2%
URGÊNCIA	ELETIVA	9	1,5%
EMERGÊNCIA	ELETIVA	9	1,5%
EMERGÊNCIA	URGÊNCIA	8	1,4%
	Total	588	100,0%

Fonte: produzida pelo autor

Em relação à variável reoperação (Tabela 7), dos 588 casos, 54 (9,2%) apresentavam o registro de pelo menos um dos códigos associados à reoperação descritos no método, sendo o registro mais comum o Z95.2, presente em 42 casos (7,1%). Desses 54 casos, 53, de fato, tratavam de casos de reoperação; um dos casos apresentava o registro Z95.2, mas não se tratava de reoperação. Chama a atenção que, em 188 casos (32%) da amostra, embora não houvesse registro de nenhum dos códigos supramencionados, os pacientes já haviam passado por cirurgia valvar previamente.

Tabela 7 – Acurácia do registro da variável reoperação

Pré-Validação	Pós-Validação	Frequência Absoluta	Frequência Relativa
SIM	SIM	53	9,0%
SIM	NÃO	1	0,2%
AUSÊNCIA DE REGISTRO	SIM	188	32,0%
AUSÊNCIA DE REGISTRO	NÃO	346	58,8%
Total	-	588	100,0%

Fonte: produzida pelo autor

## 5.2 DESEMPENHO DOS MODELOS DE APRENDIZADO DE MÁQUINA

A primeira rodada de avaliação de desempenho foi feita considerando-se todas as 44 variáveis preditoras (modelo I). A análise de significância de cada variável identificada no modelo de regressão logística dessa etapa pode ser visualizada na Tabela 8. A partir desses dados, foram selecionadas as 19 variáveis com menores p-valor; sobre elas aplicou-se novamente a regressão logística (segunda rodada), resultando em quatro variáveis com p-valor menor que 0.05, a saber: reoperação, insuficiência (da valva) aórtica (CID10 I351), idade e diâmetro do átrio esquerdo (Tabela 9, destacado em verde); essas quatro variáveis foram utilizadas para compor o modelo II.



Tabela 8 – Significância das variáveis identificadas na regressão logística do modelo I

	coef	std err	z	P> z	[0.025	0.975]
<hr/>						
IDADE_norm	0.2625	0.140	1.881	0.060	-0.011	0.536
PAC_SEXO_categ	-0.2973	0.285	-1.043	0.297	-0.856	0.262
IMC_categ	-0.0630	0.309	-0.204	0.838	-0.669	0.543
ECO_ATRIO_ESQUERDO_norm	0.2028	0.155	1.311	0.190	-0.100	0.506
ECO_DSVE_norm	-0.1356	0.145	-0.937	0.349	-0.419	0.148
ECO_FRACAO_EJECACAO_norm	-0.1310	0.134	-0.980	0.327	-0.393	0.131
ECO_AREA_VALVAR_norm	-0.1605	0.147	-1.094	0.274	-0.448	0.127
ECO_PSAP_norm	0.1062	0.133	0.798	0.425	-0.155	0.367
ALBUMINA_categ	0.3744	0.316	1.184	0.237	-0.246	0.994
UREIA_norm	0.0093	0.132	0.070	0.944	-0.249	0.268
CREATININA_norm	0.1881	0.138	1.366	0.172	-0.082	0.458
HEMOGLOBINA_categ	-0.5804	0.350	-1.657	0.098	-1.267	0.106
PLAQUETAS_norm	-0.1421	0.118	-1.209	0.227	-0.372	0.088
INR_norm	0.1042	0.139	0.752	0.452	-0.167	0.376
GLICADA_categ	0.1809	0.580	0.312	0.755	-0.957	1.318
REM_ISOSORBITIDA	0.4418	1.92e+06	2.31e-07	1.000	-3.75e+06	3.75e+06
REM_DICOXINA	-0.1806	1.92e+06	-9.42e-08	1.000	-3.75e+06	3.75e+06
REM_AMIODARONA	-0.1752	1.92e+06	-9.15e-08	1.000	-3.75e+06	3.75e+06
REM_VARFARINA	-0.1160	1.92e+06	-6.06e-08	1.000	-3.75e+06	3.75e+06
new_TECA	0.0322	1.92e+06	1.68e-08	1.000	-3.75e+06	3.75e+06
new_DURETICO	0.2549	1.92e+06	1.33e-07	1.000	-3.75e+06	3.75e+06
new_betabloq	0.0200	1.92e+06	1.05e-08	1.000	-3.75e+06	3.75e+06
new_bloq_calcio	0.2837	1.92e+06	1.48e-07	1.000	-3.75e+06	3.75e+06
new_antiagreg	-0.4749	1.92e+06	-2.48e-07	1.000	-3.75e+06	3.75e+06
qtde_medicacoes	0.0859	1.92e+06	4.49e-08	1.000	-3.75e+06	3.75e+06
REOPERACAO_categ	1.1047	0.302	3.664	0.000	0.514	1.696
INSUF_CORON	0.7407	0.530	1.397	0.162	-0.298	1.780
DIABETES	0.9985	0.676	1.477	0.140	-0.327	2.324
TABAGISMO	-0.6638	0.679	-0.978	0.328	-1.994	0.666
DPOC	-0.6388	0.930	-0.687	0.492	-2.461	1.183
TONT_SINCO	0.4745	0.498	0.953	0.341	-0.502	1.451
I050	-0.5328	0.387	-1.377	0.168	-1.291	0.225
I051	-0.1400	0.333	-0.421	0.674	-0.792	0.512
I060	0.3479	0.525	0.662	0.508	-0.682	1.378
I061	-0.7608	0.517	-1.473	0.141	-1.773	0.252
I070	-0.1289	1.848	-0.070	0.944	-3.751	3.494
I071	-0.7658	0.520	-1.474	0.141	-1.784	0.253
I080	-0.8030	0.607	-1.322	0.186	-1.993	0.387
I081	-1.3017	1.605	-0.811	0.417	-4.448	1.845
I082	0.1323	1.694	0.078	0.938	-3.188	3.452
I340	-0.3790	0.350	-1.083	0.279	-1.065	0.307
I341	0.4445	0.597	0.744	0.457	-0.726	1.615
I350	0.4887	0.328	1.489	0.136	-0.154	1.132
I351	-0.8327	0.330	-2.520	0.012	-1.480	-0.185

Fonte: produzida pelo autor

Tabela 9 – Variáveis e significância identificadas na regressão logística da segunda rodada, aplicada sobre as 19 principais variáveis do modelo I

Regressão Logística		
Variáveis	Coeficientes	Significância p-valor
REOPERACAO_categ	1,1606	-
I351	- 0,7605	0,016
IDADE_norm	0,2891	0,023
ECO_ATRIO_ESQUERDO_norm	0,2683	0,046
I071	- 0,7979	0,095
DIABETES	1,0038	0,101
I080	- 0,8203	0,142
I061	- 0,6702	0,143
CREATININA_norm	0,1847	0,143
HEMOGLOBINA_categ	- 0,3785	0,147
I350	0,3855	0,198
I050	- 0,4370	0,202
PLAQUETAS_norm	- 0,1349	0,229
ALBUMINA_categ	0,3168	0,239
INSUF_CORON	0,5438	0,277
I340	- 0,3078	0,338
ECO_AREA_VALVAR_norm	- 0,1274	0,344
ECO_FRACAO_EJECAO_norm	- 0,0918	0,409
PAC_SEXO_categ	- 0,1287	0,612

Fonte: produzida pelo autor

A Tabela 10 apresenta o desempenho dos quatro algoritmos quando estes foram expostos a todas as 44 variáveis preditoras (modelo I) e quando foram expostas apenas as quatro variáveis preditoras mais importantes (modelo II). O desempenho da regressão logística foi mais consistente no modelo II; já o KNN, a árvore de decisão e as florestas aleatórias tiveram melhor desempenho quando expostos a todas as variáveis (modelo I). As Tabelas 11 e 12 mostram as variáveis que compuseram a árvore de decisão e as florestas aleatórias no modelo I, com respectivas importâncias.

Tabela 10 – Resumo do desempenho dos quatro algoritmos nos modelos I (44 variáveis preditoras) e II (4 variáveis preditoras)

**Modelo I – Com todas as 44 variáveis**

Base	Técnica	Acurácia	Precisão	Sensibilidade	Especificidade	F1-Score
Treino	Regressão Logística	57,0%	57,1%	56,6%	57,0%	56,9%
Validação	Regressão Logística	58,3%	57,8%	61,7%	55,0%	59,7%
Teste	Regressão Logística	59,3%	58,1%	62,1%	57,0%	60,0%
Treino	KNN	65,3%	69,1%	55,6%	75,0%	61,6%
Validação	KNN	50,8%	50,7%	56,7%	47,0%	53,5%
Teste	KNN	62,7%	60,6%	69,0%	57,0%	64,5%
Treino	Árvore de Decisão	62,1%	63,9%	56,1%	68,0%	59,7%
Validação	Árvore de Decisão	62,5%	64,2%	56,7%	68,0%	60,2%
Teste	Árvore de Decisão	66,1%	64,5%	69,0%	63,0%	66,7%
Treino	Florestas Aleatórias	65,0%	64,5%	67,3%	63,0%	65,9%
Validação	Florestas Aleatórias	65,0%	64,5%	66,7%	63,0%	65,6%
Teste	Florestas Aleatórias	55,9%	54,1%	69,0%	43,0%	60,6%

**Modelo II – Com as quatro variáveis mais importantes**

Base	Técnica	Acurácia	Precisão	Sensibilidade	Especificidade	F1-Score
Treino	Regressão Logística	60,9%	61,0%	61,0%	61,0%	61,0%
Validação	Regressão Logística	59,2%	58,5%	63,3%	55,0%	60,8%
Teste	Regressão Logística	61,0%	60,0%	62,1%	60,0%	61,0%
Treino	KNN	66,5%	67,3%	64,4%	69,0%	65,8%
Validação	KNN	55,8%	54,2%	75,0%	37,0%	62,9%
Teste	KNN	47,5%	46,9%	51,7%	43,0%	49,2%
Treino	Árvore de Decisão	65,5%	66,5%	62,9%	68,0%	64,7%
Validação	Árvore de Decisão	55,8%	55,7%	56,7%	55,0%	56,2%
Teste	Árvore de Decisão	54,2%	53,1%	58,6%	50,0%	55,7%
Treino	Florestas Aleatórias	62,1%	59,8%	74,6%	50,0%	66,4%
Validação	Florestas Aleatórias	57,5%	55,4%	76,7%	38,0%	64,3%
Teste	Florestas Aleatórias	49,2%	48,8%	69,0%	30,0%	57,1%

Fonte: produzida pelo autor

Tabela 11 – Variáveis preditoras da árvore de decisão (modelo I)

Árvore de Decisão	
Variáveis	Importância da Variável
INR_norm	0,3654
ECO_PSAP_norm	0,3553
UREIA_norm	0,2793

Fonte: produzida pelo autor

Tabela 12 – Variáveis preditoras da florestas aleatórias (modelo I)

Florestas Aleatórias	
Variáveis	Importância da Variável
REOPERACAO_categ	0,3381
UREIA_norm	0,3255
I350	0,1856
CREATININA_norm	0,0932
INR_norm	0,0577

Fonte: produzida pelo autor.

A terceira rodada de avaliação de desempenho foi organizada ao se excluir a variável reoperação do modelo I (ou seja, todas as 44 variáveis menos reoperação). Aplicando-se novamente os passos descritos para seleção das principais variáveis do modelo de regressão logística, foram identificadas três variáveis finalistas, a saber: diâmetro do átrio esquerdo, insuficiência (da valva) aórtica (CID10 I351) e diabetes (Tabela 13, destacado em verde).

Tabela 13 – Variáveis e significância identificadas na regressão logística ao se excluir a variável reoperação do modelo I

Variáveis finalistas - Regressão Logística		
	coef	p-valor
ECO_ATRIO_ESQUERDO_norm	0,361	0,005
I351	- 0,714	0,018
DIABETES	1,154	0,052
PLAQUETAS_norm	- 0,190	0,077
CREATININA_norm	0,188	0,090
TONT_SINCO	0,721	0,104
I350	0,445	0,119
I340	- 0,450	0,120
IDADE_norm	0,184	0,128
I071	- 0,614	0,185
INSUF_CORON	0,597	0,222
HEMOGLOBINA_categ	- 0,230	0,240
ECO_FRACAO_EJECAO_norm	- 0,119	0,266
ALBUMINA_categ	0,209	0,379

Fonte: produzida pelo autor

As 43 variáveis constituíram o modelo I-B e as três variáveis identificadas, como principais na regressão logística da terceira rodada, formaram o modelo II-B. Os resultados dessa rodada são apresentados na Tabela 14. A regressão logística mostrou maior consistência no modelo II-B e os outros três algoritmos apresentaram maior consistência no modelo I-B.

Tabela 14 – Resumo do desempenho dos quatro algoritmos nos modelos I-B (43 variáveis) e II-B (3 variáveis)

**Modelo I – B – Com 43 variáveis (excluiu-se reoperação)**

Base	Técnica	Acurácia	Precisão	Sensibilidade	Especificidade	F1-Score
Treino	Regressão Logística	57,5%	57,2%	60,0%	60,0%	58,6%
Validação	Regressão Logística	61,7%	62,5%	58,3%	55,0%	60,3%
Teste	Regressão Logística	57,6%	57,7%	51,7%	67,0%	54,5%
Treino	KNN	62,8%	72,6%	41,5%	74,0%	52,8%
Validação	KNN	50,8%	51,1%	40,0%	52,0%	44,9%
Teste	KNN	57,6%	60,0%	41,4%	59,0%	49,0%
Treino	Árvore de Decisão	64,3%	65,4%	61,0%	68,0%	63,1%
Validação	Árvore de Decisão	52,5%	53,2%	41,7%	63,0%	46,7%
Teste	Árvore de Decisão	55,9%	55,6%	51,7%	60,0%	53,6%
Treino	Florestas Aleatórias	60,9%	61,0%	61,0%	61,0%	61,0%
Validação	Florestas Aleatórias	58,3%	57,6%	63,3%	53,0%	60,3%
Teste	Florestas Aleatórias	62,7%	60,0%	72,4%	53,0%	65,6%

**Modelo II – B – Com as 3 variáveis mais importantes (excluiu-se reoperação)**

Base	Técnica	Acurácia	Precisão	Sensibilidade	Especificidade	F1-Score
Treino	Regressão Logística	58,4%	58,5%	58,5%	58,0%	58,5%
Validação	Regressão Logística	55,0%	55,2%	53,3%	57,0%	54,2%
Teste	Regressão Logística	62,7%	63,0%	58,6%	67,0%	60,7%
Treino	KNN	67,5%	74,7%	53,2%	82,0%	62,1%
Validação	KNN	47,5%	46,9%	38,3%	57,0%	42,2%
Teste	KNN	45,8%	43,5%	34,5%	57,0%	38,5%
Treino	Árvore de Decisão	63,3%	61,5%	71,7%	55,0%	66,2%
Validação	Árvore de Decisão	53,3%	52,9%	60,0%	47,0%	56,2%
Teste	Árvore de Decisão	45,8%	46,2%	62,1%	30,0%	52,9%
Treino	Florestas Aleatórias	61,6%	61,0%	64,9%	58,0%	62,9%
Validação	Florestas Aleatórias	48,3%	48,1%	41,7%	55,0%	44,6%
Teste	Florestas Aleatórias	59,3%	58,1%	62,1%	57,0%	60,0%

Fonte: produzida pelo autor.

As Tabelas 15 e 16 mostram a importância das variáveis na árvore de decisão e nas florestas aleatórias no contexto em que a variável de reoperação foi removida.

Tabela 15 – Importância das variáveis no modelo de árvore de decisão sem a presença da variável reoperação (modelo I-B)

<b>Árvore de Decisão</b>	
<b>Variáveis</b>	<b>Importância da Variável</b>
INR_norm	0,3876
IDADE_norm	0,2954
CREATININA_norm	0,1567
ECO_FRACAO_EJCAO_norm	0,0845
UREIA_norm	0,0757

Fonte: produzida pelo autor

Tabela 16 – Importância das variáveis no algoritmo de florestas aleatórias sem a presença da variável reoperação (modelo I-B)

<b>Florestas Aleatórias</b>	
<b>Variáveis</b>	<b>Importância da Variável</b>
INR_norm	0,4331
CREATININA_norm	0,1921
PLAQUETAS_norm	0,1318
UREIA_norm	0,1210
new_betabloq	0,0658
ECO_AREA_VALVAR_norm	0,0315
ALBUMINA_categ	0,0223
REM_DIGOXINA	0,0024

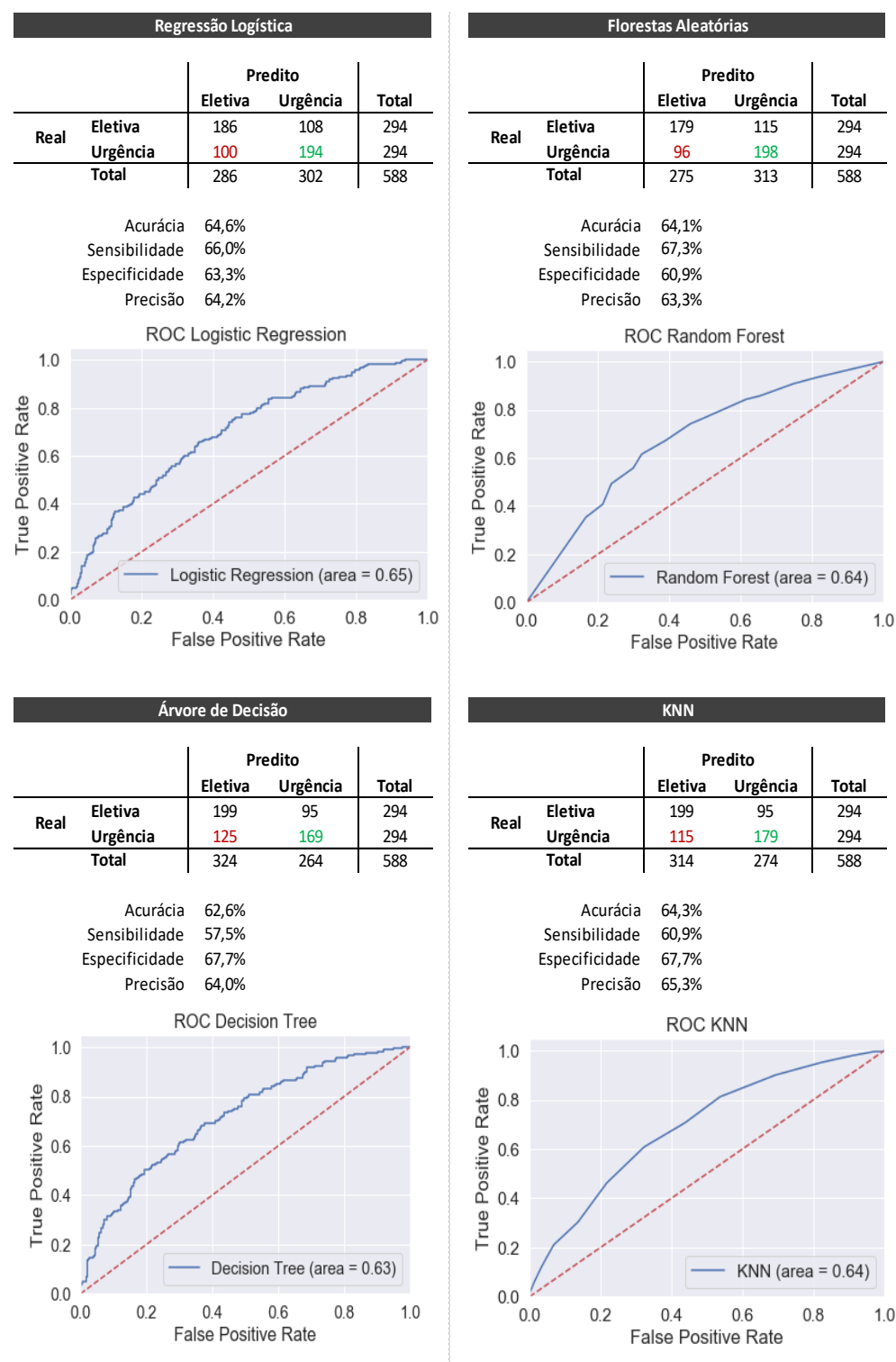
Fonte: produzida pelo autor

Comparando-se os modelos I, II, I-B e II-B, observamos maior desempenho dos quatro algoritmos quando a variável reoperação está presente (modelos I e II), visto que os dois modelos com maior consistência nas métricas são: florestas aleatórias do modelo I e regressão logística do modelo II.

Uma vez constituídos os algoritmos, foi confeccionada a matriz de confusão sobre os 588 casos da amostra do estudo, com as respectivas métricas de discriminação e curvas ROC.

A Figura 11 apresenta as matrizes de confusão e curvas ROC quando se utilizaram os modelos supramencionados com a variável reoperação, e a Figura 12 quando se usaram os modelos sem a variável reoperação.

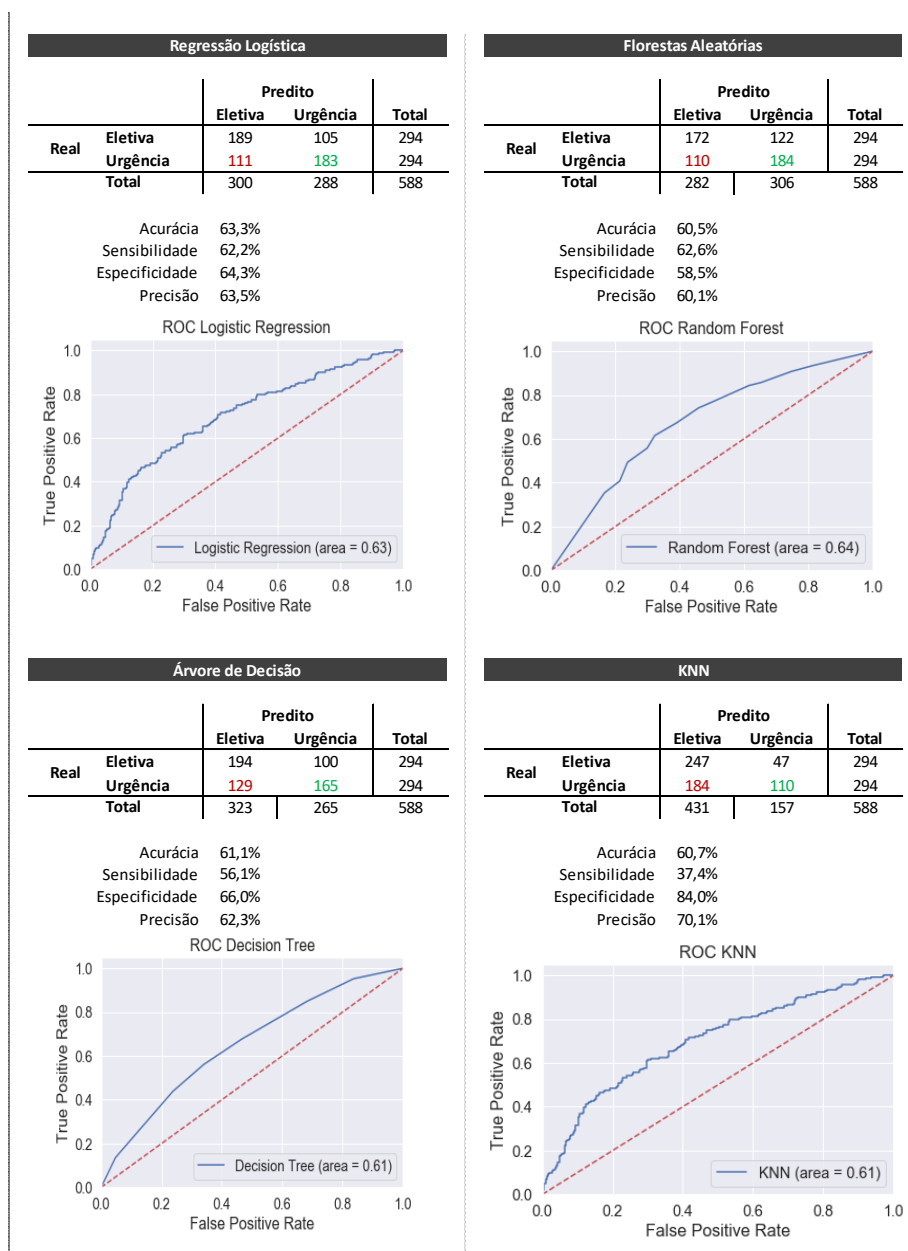
Figura 11 – Matrizes de confusão e respectivas curvas ROC para os algoritmos selecionados, utilizando-se a variável reoperação



Fonte: produzida pelo autor



Figura 12 – Matrizes de confusão e respectivas curvas ROC para os algoritmos selecionados, sem a variável reoperação



Fonte: produzida pelo autor

Finalmente, realizamos o teste de McNemar (DIETTERICH, 1998) para comparar se havia diferença de desempenho, com significância estatística, entre os algoritmos de regressão logística e florestas aleatórias, nos cenários com e sem a variável reoperação, dado que, na matriz de confusão, os resultados eram muito próximos.

A partir da formulação das hipóteses de teste e da construção das respectivas tabelas de contingência (Apêndice II), aplicamos o teste qui-quadrado e obtemos o

p-valor de 0,694, com 1 grau de liberdade, ao se comparar os resultados da regressão logística com o das florestas aleatórias no cenário com a variável reoperação. Já nesses algoritmos, sem a reoperação, o p-valor foi de 0,474. Isto posto, não se pode rejeitar a hipótese de nulidade, ou seja, os algoritmos de regressão logística e de florestas aleatórias, nos dois cenários, apresentam o mesmo desempenho. Eles são equivalentes em seu poder classificatório.

## 6. DISCUSSÃO

Raghupathi e Raghupathi (2014), Obermeyer e Emanuel (2016) e Ngiam e Khor (2019) destacam a importância da qualidade dos dados para que ferramentas de *Big Data*, como o aprendizado de máquina, possam ser aplicadas com sucesso. Grandes massas de dados, se não possuírem dados de qualidade (registros que, de fato, traduzam a realidade clínica dos pacientes), serão apenas isso: grandes massas de dados.

No estudo em questão, 41,4% dos dados de tipo de cirurgia estavam incorretos. Em relação à variável reoperação, apenas 22% dos casos em que os pacientes já haviam sido submetidos à cirurgia cardíaca prévia estavam apontados em campo estruturado, ou seja, em 78% dos casos, só foi possível descobrir isso a partir do processo de validação. Não fosse feito esse processo, 188 casos de reoperação teriam sido considerados como primeira cirurgia. A magnitude da variável reoperação é evidenciada no fato de os dois modelos com melhor desempenho (florestas aleatórias e regressão logística, modelos I e II respectivamente) apresentarem essa variável como a de maior significância; mais ainda, ao se excluir a variável reoperação (modelos I-B e II-B), observamos a queda na acurácia, sensibilidade e precisão dos modelos.

Há de se resgatar que, dos 4.416 casos potenciais iniciais, 2.035 precisaram ser excluídos por não possuírem registro, anterior à indicação da cirurgia, de nenhum dos códigos (CID-10) de lesão anatômica considerados como específicos, isto é, de maior valor diagnóstico. Esses achados vão ao encontro de Hemingway *et al.* (2018) que advertem que o registro das informações dos pacientes, na prática clínica, não segue o mesmo rigor que nos estudos clínicos, o que impõe desafios à construção de algoritmos de AM.

Para além de inteligência artificial, Raghupathi e Raghupathi (2014) advertem sobre a imprescindibilidade de dados de alta qualidade para o diagnóstico e tratamento dos pacientes assistidos, de modo a garantir maior coordenação do atendimento, evitar erros e reduzir custos. Os achados desta dissertação insuflam a necessidade da adoção e/ou intensificação, por parte do INCOR-HCFMUSP e de suas equipes, de processos de auditoria e curadoria dos registros eletrônicos de saúde, com vistas a produzir dados de maior qualidade, sejam para pesquisa, seja

para o ensino, seja para a assistência. Também, convidam a refletir se a estrutura dos prontuários eletrônicos dos pacientes utilizados hoje em dia fomenta a aplicação de soluções de IA ou é um obstáculo para tal.

Ainda, em relação à escolha das variáveis que compuseram este estudo, foi tomada a decisão estratégica de se selecionar apenas aquelas que estivessem registradas em campos estruturados no SI<sup>3</sup>, o sistema de informações da instituição. Também, não por acaso, foram escolhidas variáveis relativamente comuns na prática assistencial (sexo, idade, IMC, medicações, exames laboratoriais rotineiros, dados de ecocardiograma e diagnósticos passíveis de registro pela CID-10). Essas decisões encontram lastro em trabalhos na literatura, como o de Kruse *et al.* (2016), que apontam a maior facilidade de armazenagem, manipulação e análise de dados estruturados. A opção por se trabalhar com dados comuns da prática clínica visou a maior escalabilidade da solução proposta e a maior reprodutibilidade do método, como propõe Hulsen *et al.* (2019).

Por essas razões que variáveis como a classe funcional do paciente, segundo a *New York Heart Association*, não foram adotadas, tanto por esta não estar em campo estruturado no SI<sup>3</sup> quanto por apresentar grande variação ao longo do tratamento do paciente e ser muito sensível a medicações sintomáticas. O desejo de propor uma solução escalável também justifica o fato de se ter testado os quatro modelos com e sem a variável reoperação, uma vez que, ao longo do processo de validação, foi verificado que ela poderia ser um impeditivo para a aplicação da solução no mundo real, dada a fragilidade de seu registro.

No que tange ao desempenho dos modelos, apontado nas Figuras 11 e 12 (matrizes de confusão), é importante enfatizar que todos os 588 pacientes já eram elegíveis à cirurgia de troca valvar, ou seja, não era função do modelo identificar quem deveria ou não ser submetido à cirurgia. A necessidade do procedimento já era um fato. Isso posto, podemos dizer que os modelos deste estudo toleram bem (aceitam bem) os casos falso positivos (indicar cirurgia de urgência quando na verdade era eletiva), visto que, nesse contexto, não significaria submeter o paciente a um tratamento desnecessário, embora, na efetivação dessa antecipação de procedimento, isso possa significar o atraso no tratamento de um outro paciente que, de fato, deveria ser urgente.

Já o falso negativo (rotular o paciente urgente como eletivo) apresenta maior preocupação, uma vez que o tratamento estaria sendo retardado. A partir dessas pontuações, da acurácia, precisão e sensibilidade verificadas, além do teste de McNemar aplicado aos modelos de regressão logística e florestas aleatórias, esses dois algoritmos foram eleitos como os de melhor desempenho, com equivalência de poder classificatório. Em todos os cenários, esses modelos apresentaram-se muito próximos, embora em patamares diferentes a depender da presença ou ausência da reoperação, como visto anteriormente.

Esses resultados vão ao encontro dos achados de Christodoulou *et al.* (2019), cuja revisão sistemática mostrou equivalência entre regressão logística e modelos de predição clínica mais rebuscados (árvore de decisão, florestas aleatórias, rede neural artificial e *support vector machine*). Nos casos em que as comparações foram realizadas num contexto de baixo risco de viés (amostras balanceadas, processo de validação claro, mesmos preditores considerados, etc.), não houve superioridade de um algoritmo em relação ao outro. Isso também indica a qualidade do pré-processamento de dados realizado neste trabalho.

A análise da área sob a curva ROC (AUROC) dos modelos mostra valores que variam entre 0,63 (árvore de decisão) e 0,65 (regressão logística) quando a variável reoperação está presente e, na sua ausência, entre 0,61 (KNN) e 0,64 (florestas aleatórias). Esse achado reforça a importância da variável reoperação, como discutido anteriormente.

Para Hosmer e Lemeshow (2000), um algoritmo com uma AUROC entre 0,7 e 0,8 pode ser considerado um algoritmo com poder discriminatório aceitável; entre 0,8 e 0,9 apresenta excelente poder de classificação e, acima de 0,9, são modelos com poder discriminatório excepcionais. Embora a literatura, até o momento, tenha se debruçado pouco sobre a aprendizagem de máquina como ferramenta de priorização de pacientes, com vistas a fazer gestão populacional e garantir a melhor alocação de recursos escassos, Jing *et al.* (2020), a partir de dados de 26.781 pacientes com insuficiência cardíaca, compreendidos entre 2001 e 2019, treinaram e testaram três algoritmos capazes de prever morte global em 1 ano e, principalmente, de identificar o conjunto de pacientes com lacunas de cuidado (por exemplo, pacientes que deveriam ter tomado a vacina para a gripe, mas não a tomaram) com maior sobrevida esperada se a intervenção fosse feita. Em relação ao

desempenho desses algoritmos, destacamos que a AUROC foi melhorando conforme o tamanho da base de treinamento e de testes também aumentou, saindo de valores entre 0,57 e 0,66 e atingindo valores de 0,74 para regressão logística, 0,76 para florestas aleatórias e 0,77 para XGBoost.

Os modelos classificatórios apresentados nesta dissertação estão próximos de um desempenho satisfatório, embora ainda não tenham atingido esse patamar ( $AUROC \geq 0,7$ ). Esse achado aponta para, pelo menos, dois encaminhamentos. O primeiro é que ainda existem possibilidade e necessidade de ganho de desempenho, ou seja, a busca por melhores resultados deve continuar; por exemplo, novas variáveis podem ser consideradas (peptídeo natriurético do tipo B, volume do átrio esquerdo, entre outras), variáveis tratadas como categóricas podem ser tratadas como contínuas (e vice-versa), os pontos de corte podem ser alterados, o limiar de classificação dos modelos pode aumentar ou diminuir e outros métodos estatísticos podem ser utilizados no pré-processamento de dados.

A base de pacientes utilizada para o treinamento e teste também pode ser expandida, fato que pode levar ao aumento de desempenho evidenciado por Jing *et al.* (2020). O segundo encaminhamento, ou melhor, constatação, é que, embora ainda não esteja pronta, não se pode desconsiderar que esta dissertação apresenta uma ferramenta concreta e tangível, constituída a partir de dados de mundo real e de fácil captação, com poder discriminatório maior que 0,5. Temos uma ferramenta de priorização de pacientes e o objetivo geral deste trabalho aplicado foi atingido: Propor um modelo de aprendizado de máquina capaz de identificar pacientes em fila de espera para procedimento cirúrgico eletivo que, caso não sejam atendidos de forma tempestiva, serão submetidos à cirurgia de urgência.

A relevância de se ter um instrumento como esse, nos moldes em que foi construído, reside no fato de endereçar diversas imperfeições dos sistemas de saúde. Primeiro que confere transparência e objetividade ao processo de priorização de pacientes, em consonância com as melhores práticas sugeridas pela literatura no Brasil e no mundo (CONSELHO FEDERAL DE MEDICINA, 2016; KRETZER *et al.*, 2020; DÉRY *et al.* 2020). Como o processo será executado por um algoritmo, não será afetado por variações intra e interobservadores (OUDHOFF *et al.*, 2007), nem terá seu desempenho afetado por fatores, como cansaço e privação de sono

(COLEMAN *et al.*, 2019) ou por vieses relacionados ao processo de tomada de decisão cirúrgica (LOFTUS *et al.*, 2019).

No limite, trata-se de uma ferramenta que favorecerá o acesso ao sistema de saúde, uma vez que diferenciará aqueles que podem esperar daqueles que devem ser atendidos antes. Esse tipo de suporte urge no atual contexto que mescla a necessidade de enfrentamento da COVID-19 com a necessidade de retomada das atividades eletivas dos serviços de saúde, que ficaram paralisadas por meses, o que levou ao agravamento das filas de espera para procedimentos cardiológicos. (CESENA, 2021).

Ainda, o fato de essa ferramenta ter sido construída a partir de 44 variáveis permitiu a correlação entre variáveis e desfecho que, em contextos tradicionais, não seriam consideradas, como ocorreu com a ureia, segunda variável mais importante no modelo de florestas aleatórias, e ficou atrás apenas da reoperação. Essa possibilidade de se estabelecer novas correlações, inicialmente pouco valorizadas pela literatura, é uma das principais fortalezas da IA, como indicam Hulsén *et al.* (2019) e Liu *et al.* (2019).

Não menos importante, cumpre lembrar que a gestão de filas de espera demanda tanto recursos administrativos quanto assistenciais. À vista disso, automatizar parte desse processo pode ser entendido como algo positivo, pois permitirá, por exemplo, que o médico dedique mais tempo para o cuidado direto com o paciente e em tarefas que requeiram julgamento humano, intuição e empatia (NGIAM, KHOR, 2019), e deixe para a inteligência artificial fazer a tarefa repetitiva de constantemente analisar todos os pacientes em fila de espera que aguardam pelo procedimento.

Que pesem todos esses potenciais benefícios, não se pode deixar de lado outras tantas preocupações e desafios existentes no campo de *Big Data* e aprendizado de máquina como suporte à decisão médica. Se, por um lado, um algoritmo automatizado pode trazer os ganhos supramencionados; por outro, tem o potencial de induzir iatrogenia em massa, de lesar pacientes em série, razão pela qual seus resultados devem ser sempre confrontados com a realidade e analisados de forma crítica e contínua, submetidos a um processo de revisão por pares. (RUMSFELD, JOYNT, MADDOX, 2016; NGIAM, KHOR, 2019; HULSEN *et al.*, 2019; LIU *et al.*, 2019). Isso por si encerra outro desafio, que é de ter profissionais de

saúde, assistenciais e de gestão devidamente capacitados para executar essas tarefas. (TOPOL, 2019).

Conforme Hemingway *et al.* (2018) destacam, ainda poucos são os profissionais que tiveram treinamento formal em áreas correlatas a *Big Data* e aprendizado de máquina (informática, ciência de dados, programação computacional). Para esses autores, serão necessárias políticas nacionais de incentivo para endereçar essa lacuna.

Outro desafio que se impõe é de incorporar essas tecnologias à gestão e ao cuidado de maneira harmônica (HULSEN *et al.*, 2019). Por mais que a ferramenta deste estudo estivesse pronta e fosse segura, sua integração à rotina ainda demandaria o desenvolvimento de uma interface amigável ao usuário e, possivelmente, o desenho de novos processos de trabalho. Solans-Domènech *et al.* (2013) frisam que a efetivação desses instrumentos na vida diária das instituições e de seus profissionais é função da facilidade com que são aplicadas.

Pensando ainda na aplicação prática da ferramenta, há de se atentar para o fato de que toda solução de AM encerra em si uma limitação básica: o contexto em que ela foi produzida. Por serem desenvolvidas (treinadas, validadas e testadas) sob um conjunto específico de dados (fila cirúrgica valvar do INCOR-HCFMUSP, por exemplo), sua generalização para outras populações (fila cirúrgica valvar do estado de São Paulo) não é imediata. Conforme Ngiam e Khor (2019) enfatizam, essa extrapolação requer que os algoritmos sejam confeccionados com dados dessas populações de interesse. Assim, a solução apresentada neste trabalho, embora tenha potencial de ser útil em outros contextos, deve ser tratada com restrição quanto à sua universalização.

Não menos importante, convém lembrar de que existe um hiato entre a adoção de uma solução e os resultados que, de fato, decorrem dela. Para que algoritmos como os apresentados aqui sejam disruptivos, eles necessariamente precisam se traduzir em ganho na prática clínica e gerencial. Em contrário, serão apenas invenções. (HULSEN *et al.*, 2019). Esse é outro convite que floresce neste estudo, observar os benefícios que decorrem do uso de uma ferramenta de priorização como essa; se, de fato, menos pacientes serão admitidos como urgência; se isso diminuirá os custos e os tempos de permanência hospitalar, ampliando o acesso das pessoas aos recursos de saúde escassos.



Finalmente, cabe destacar algumas limitações importantes deste trabalho. Embora os modelos apresentados tenham logrado sucesso em classificar os pacientes em relação à admissão eletiva e de urgência, o espaço temporal envolvido nesse processo não foi definido, ou seja, até o limite desta dissertação o algoritmo diz quem será admitido como urgência, mas não define em quanto tempo, a partir da indicação da cirurgia, isso ocorrerá. Esse fato reforça a necessidade de novos estudos, no sentido de se construir um algoritmo que identifique um conjunto de pacientes que serão admitidos como urgência em um determinado espaço de tempo. Por exemplo, dos 1.252 pacientes atuais na fila de espera para procedimento cirúrgico valvar, o algoritmo seria capaz de apontar os 30 pacientes que apresentam 90% de chance de serem admitidos como urgência nos próximos 45 dias; ou os 150 doentes com 85% de chance de serem admitidos como urgência nos próximos 6 meses. Enfim, o autor entende que o fator temporal pode acrescentar informações relevantes no processo de priorização dos pacientes.

Outra decisão passível de questionamento foi a adoção dos últimos dados registrados no SI<sup>3</sup>, antes da indicação da cirurgia, como dados com valor presente. Na prática, isso significa dizer, por exemplo, que uma medicação prescrita, pela última vez, 24 meses antes da indicação da cirurgia, teve o mesmo valor que uma medicação prescrita 2 meses antes desse evento. No estudo em questão, não se optou por resgatar dados registrados apenas num determinado período antes da listagem dos pacientes em fila.

Também, há de se destacar que, excetuando-se a reoperação e o tipo de admissão, nenhuma das outras variáveis foi validada. Isso implica aceitar que parte dos dados poderiam ter sido melhorados (corrigidos). Entretanto, como o desejo do pesquisador era produzir algo passível de ser aplicado no mundo real, optou-se por não se alterar esses dados e deixou-os exatamente como se apresentavam nos campos estruturados do prontuário eletrônico do paciente.

Outra limitação foi a quantidade de casos que, de fato, entraram na construção do modelo – 588. Embora o cálculo amostral tenha apontado para uma amostra significativa, não se pode esquecer de que 2.035 casos foram excluídos por não possuírem registro da CID-10 específico para lesão anatômica do paciente. Ampliar a amostra do estudo pode ser uma solução interessante para se aumentar o

desempenho dos modelos, especialmente no contexto de *Big Data* e inteligência artificial, a exemplo do trabalho de Jing *et al.* (2020).

## 7. CONCLUSÃO

O envelhecimento populacional, o progresso tecnológico, os indivíduos cada vez mais conscientes da necessidade de cuidados, o acesso facilitado a serviços e tratamentos e o avanço nas tecnologias cirúrgicas e anestésicas pressionam a demanda por procedimentos cirúrgicos. Dada a oferta finita desses serviços, que não acompanhou a demanda nas últimas décadas, é fundamental que os sistemas de saúde e seus serviços lancem mão de mecanismos que garantam a melhor alocação dos recursos disponíveis. A existência de filas de espera em sistemas de saúde de acesso universal é um fato, em países em desenvolvimento e nos ricos também.

Uma das maneiras de destinar os recursos existentes de forma mais inteligente e justa é por meio da priorização. O presente estudo apresenta uma ferramenta de priorização de pacientes voltada para os que estão em fila cirúrgica valvar, que identifica aqueles que, caso não sejam priorizados, serão atendidos como urgência. Esperamos, com isso, promover um gerenciamento da fila de espera mais justo e equânime bem como garantir o acesso mais rápido àqueles que, de fato, necessitam.

Por ser um instrumento baseado em aprendizado de máquina, foi possível trabalhar com 44 variáveis preditoras potencialmente associadas ao desfecho “tipo de cirurgia” (eletiva ou urgência). Os algoritmos de regressão logística e florestas aleatórias apresentaram melhor poder discriminatório, visto que, para tal, este utilizou entre cinco e oito variáveis preditoras e aquele, entre três e quatro. Como é um método computacional, traz transparência e objetividade ao processo e garante que todos passem pelo mesmo escrutínio. Também protege os pacientes das diversas imperfeições inerentes ao processo manual de priorização: diferença de percepção intra e interobservadores, fadiga, disponibilidade de tempo e outros vieses associados à tomada de decisão cirúrgica. Esse é outro fator que concorre para promover um sistema de saúde mais equânime.

Para além dos benefícios esperados para os pacientes, acreditamos que a automatização do processo de priorização de pacientes em fila de espera também concorrerá para trazer eficiência administrativa e assistencial. Se implementado, o método tem a potencialidade de reduzir as horas gastas por médicos, clínicos e

cirurgiões, no processo de seleção de pacientes, o mesmo ocorrendo com as equipes administrativas que gerenciam essas filas.

Que pese o estado *hype* que a inteligência artificial e os algoritmos de aprendizado de máquina assumiram nos últimos anos, ainda existe um trabalho importante a ser feito de curadoria dos dados gerados ao longo da jornada dos pacientes em hospitais e outros serviços de saúde. Pouco adianta muitos dados se esses não forem de boa qualidade. Validar esses algoritmos diante da prática clínica também é um passo fundamental a ser dado, assim como a formação de profissionais habilitados a avaliar e dialogar com essas soluções é um desafio que se impõe.

Longe de ser uma solução que substituirá o médico e os demais profissionais envolvidos no processo de cuidado do paciente, os algoritmos propostos apresentam-se como uma ferramenta de auxílio à tomada de decisão, cuja aplicação tem por finalidade potencializar a capacidade humana de fazer escolhas inteligentes. Melhor que a máquina ou o homem tomando decisões, é o homem e a máquina decidindo em conjunto.

## REFERÊNCIAS

ANDREU-PEREZ, J. *et al.* Big Data for Health. **IEEE Journal of Biomedical and Health Informatics**, v. 19, n. 4, p. 1193–1208, jul. 2015.

BARBOSA, M. G. **Gerenciamento da fila de espera para cirurgia ginecológica em hospital municipal da Zona Sul de São Paulo**: como garantir acesso e otimizar a utilização de recursos. Dissertação (Mestrado) – [s.l.: s.n.].

BATISTA, A. F. DE M.; CHIAVEGATTO FILHO, A. D. P. Machine Learning aplicado à Saúde. In: **Minicursos do Simpósio Brasileiro de Computação Aplicada à Saúde 2019**. 19. ed. Porto Alegre: Sociedade Brasileira de Computação, 2019.

BENHAR, H.; IDRI, A.; FERNÁNDEZ-ALEMÁN, J. L. Data preprocessing for heart disease classification: A systematic literature review. **Computer Methods and Programs in Biomedicine**, v. 195, p. 105-635, 1 out. 2020.

BIHORAC, A. *et al.* MySurgeryRisk: Development and Validation of a Machine-learning Risk Algorithm for Major Complications and Death After Surgery. **Annals of Surgery**, v. 269, n. 4, p. 652–662, abr. 2019.

BRASIL. **Lei nº 8080**. Disponível em: <[http://www.planalto.gov.br/ccivil\\_03/leis/l8080.htm](http://www.planalto.gov.br/ccivil_03/leis/l8080.htm)>. Acesso em: 20 out. 2020.

BRASIL. **Emenda Constitucional nº 95**. Disponível em: <[http://www.planalto.gov.br/ccivil\\_03/constituicao/emendas/emc/emc95.htm](http://www.planalto.gov.br/ccivil_03/constituicao/emendas/emc/emc95.htm)>. Acesso em: 25 mar. 2021.

CASTRO, M. C. *et al.* Brazil's unified health system: the first 30 years and prospects for the future. **The Lancet**, v. 394, n. 10195, p. 345–356, 27 jul. 2019.

CESENA, F. H. Y. Eventos Cardiovasculares Evitáveis: um sério efeito colateral da Pandemia de COVID-19. **Arquivos Brasileiros de Cardiologia**, v. 116, p. 381–382, 23 abr. 2021.

CHIAVEGATTO FILHO, A. D. P. Uso de big data em saúde no Brasil: perspectivas para um futuro próximo. **Epidemiologia e Serviços de Saúde**, v. 24, n. 2, p. 325–332, jun. 2015.

CHRISTODOULOU, E. *et al.* A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models. **Journal of Clinical Epidemiology**, v. 110, p. 12–22, 1 jun. 2019.

COLEMAN, J. J. *et al.* To Sleep, Perchance to Dream: Acute and Chronic Sleep Deprivation in Acute Care Surgeons. **Journal of the American College of Surgeons**, v. 229, n. 2, p. 166–174, 1º ago. 2019.

CONSELHO FEDERAL DE MEDICINA. 2.156/2016. **Resolução CFM nº 2.156/2016**. 17 nov. 2016, Sec. 1, p. 138–139.

CONSELHO NACIONAL DE JUSTIÇA. **Nota Técnica nº 24**. 12 maio 2020.

CONSELHO NACIONAL DE SECRETÁRIOS DE SAÚDE. **Resolução CIT nº 37**. 22 março 2018.

CULLIS, J. G.; JONES, P. R. Rationing by Waiting Lists: An Implication. **The American Economic Review**, v. 76, n. 1, p. 250–256, 1986.

DÉRY, J. *et al.* Patient prioritization tools and their effectiveness in non-emergency healthcare services: a systematic review protocol. **Systematic Reviews**, v. 8, n. 1, p. 78, 30 mar. 2019.

DÉRY, J. *et al.* A systematic review of patient prioritization tools in non-emergency healthcare services. **Systematic Reviews**, v. 9, n. 1, p. 227, 6 out. 2020.

DEW, K. *et al.* Explicit rationing of elective services: implementing the New Zealand reforms. **Health Policy**, v. 74, n. 1, p. 1–12, 28 set. 2005.

DIETTERICH, T. G. Approximate Statistical Tests for Comparing Supervised Classification Learning Algorithms. **Neural Computation**, v. 10, n. 7, p. 1895–1923, 1º out. 1998.

FURUIE, S.S. *et al.*, "Managing Medical Images and Clinical Information: InCor's Experience," in IEEE Transactions on Information Technology in Biomedicine, vol. 11, no. 1, p. 17-24, Jan. 2007, doi: 10.1109/TITB.2006.879588.

GUTACKER, N.; SICILIANI, L.; COOKSON, R. Waiting time prioritisation: Evidence from England. **Social Science & Medicine**, v. 159, p. 140–151, 1 jun. 2016.

HADORN, D. Setting Priorities on Waiting Lists: Point-Count Systems as Linear Models. **Journal of Health Services Research & Policy**, v. 8, n. 1, p. 48–54, 1 jan. 2003.

HADORN, D. C. Setting priorities for waiting lists: defining our terms. **CMAJ: Canadian Medical Association Journal**, v. 163, n. 7, p. 857–860, 3 out. 2000.

HASTIE, T.; TIBSHIRANI, R.; FRIEDMAN, J. **The Elements of Statistical Learning: Data Mining, Inference, and Prediction**. 2. ed. [s.l.] Springer, 2008.

HEMINGWAY, H. *et al.* Big data from electronic health records for early and late translational cardiovascular research: challenges and potential. **European Heart Journal**, v. 39, n. 16, p. 1481–1495, 21 abr. 2018.

HOOGEBOOM, T. J. *et al.* The impact of waiting for total joint replacement on pain and functional status: a systematic review. **Osteoarthritis and Cartilage**, v. 17, n. 11, p. 1420–1427, 1 nov. 2009.

HOSMER, D. W.; LEMESHOW, S. **Applied Logistic Regression**. 2nd. ed. [s.l.] John Wiley & Sons, Inc., 2000.

HUANG, T. *et al.* Promises and Challenges of Big Data Computing in Health Sciences. **Big Data Research**, Special Issue on Computation, Business, and Health Science. v. 2, n. 1, p. 2–11, 1 mar. 2015.

HULSEN, T. *et al.* From Big Data to Precision Medicine. **Frontiers in Medicine**, v. 6, 2019.

HURST, J.; SICILIANI, L. Tackling Excessive Waiting Times for Elective Surgery: A Comparison of Policies in Twelve OECD Countries. **OECD Health Working Papers**, v. 6, 7 jul. 2003.

IBM BRASIL. **O que é Machine Learning e como utilizar?** IBM Brasil. Disponível em: <<https://www.ibm.com/br-pt/analytics/machine-learning>>. Acesso em: 21 mar. 2021.

INCOR. **Relatório Anual de Atividades 2019**. Unidade de Informações Médico Hospitalares, 24 jan. 2020. . Acesso em: 26 mar. 2021.

IVERSEN, T. A theory of hospital waiting lists. **Journal of Health Economics**, v. 12, n. 1, p. 55–71, 1 abr. 1993.

JAMES, G. *et al.* **An introduction to statistical learning: with applications in R**. [s.l.] Springer, 2014.

JING, L. *et al.* A Machine Learning Approach to Management of Heart Failure Populations. **JACC: Heart Failure**, v. 8, n. 7, p. 578–587, 1 jul. 2020.

KEE, F. *et al.* Urgency and priority for cardiac surgery: a clinical judgment analysis. **BMJ**, v. 316, n. 7135, p. 925–929, 1998.

KRETZER, L. *et al.* **Recomendações da AMIB (Associação de Medicina Intensiva Brasileira), ABRAMEDE (Associação Brasileira de Medicina de Emergência, SBGG (Sociedade Brasileira de Geriatria e Gerontologia) e ANCP (Academia Nacional de Cuidados Paliativos) de alocação de recursos em esgotamento durante a pandemia por COVID-19**. 1º maio 2020. Disponível em: <[https://www.amib.org.br/fileadmin/user\\_upload/amib/2020/abril/24/VJS01\\_maio\\_-\\_Versa\\_\\_o\\_2\\_-\\_Protocolo\\_AMIB\\_de\\_alocac\\_\\_a\\_\\_o\\_de\\_recursos\\_em\\_esgotamento\\_durante\\_a\\_pandemia\\_por\\_COVID.pdf](https://www.amib.org.br/fileadmin/user_upload/amib/2020/abril/24/VJS01_maio_-_Versa__o_2_-_Protocolo_AMIB_de_alocac__a__o_de_recursos_em_esgotamento_durante_a_pandemia_por_COVID.pdf)>. Acesso em: 13 nov. 2020.

KRUSE, C. S. *et al.* Challenges and Opportunities of Big Data in Health Care: A Systematic Review. **JMIR Medical Informatics**, v. 4, n. 4, p. e5359, 21 nov. 2016.

LIU, Y. *et al.* How to Read Articles That Use Machine Learning: Users' Guides to the Medical Literature. **JAMA**, v. 322, n. 18, p. 1806–1816, 12 nov. 2019.

LOFTUS, T. J. *et al.* Artificial Intelligence and Surgical Decision-making. **JAMA Surgery**, v. 155, n. 2, p. 148–158, dez. 2019.

MACHADO, F. P. **Predição da hospitalização de pacientes idosos no**

**departamento de emergência:** abordagem utilizando aprendizado de máquina. Dissertação (Mestrado)—[s.l: s.n.].

MALTA, D. C. *et al.* Mortes evitáveis no Sistema Único de Saúde na população brasileira, entre 5 e 69 anos, 2000 - 2013. **Revista Brasileira de Epidemiologia**, v. 21, 2018.

MAYRINK, V. T. DE M. **Avaliação do algoritmo Gradient Boosting em aplicações de previsão de carga elétrica a curto prazo.** Dissertação (Mestrado)—Brasil: Universidade Federal de Juiz de Fora (UFJF), 31 ago. 2016.

MCLEOD, D. *et al.* Use of, and attitudes to, clinical priority assessment criteria in elective surgery in New Zealand. **Journal of Health Services Research & Policy**, v. 9, n. 2, p. 91–99, 1 abr. 2004.

MENDES, E. V. **Desafios do SUS.** Brasília, DF: CONASS, 2019.

MOONEY, S. J.; PEJAVER, V. Big Data in Public Health: Terminology, Machine Learning, and Privacy. **Annual Review of Public Health**, v. 39, n. 1, p. 95–112, 2018.

NAM, J. G. *et al.* Development and Validation of Deep Learning–based Automatic Detection Algorithm for Malignant Pulmonary Nodules on Chest Radiographs. **Radiology**, v. 290, n. 1, p. 218–228, 25 set. 2018.

NGIAM, K. Y.; KHOR, I. W. Big data and machine learning algorithms for health-care delivery. **The Lancet Oncology**, v. 20, n. 5, p. e262–e273, 1 maio 2019.

NIC.BR. **Pesquisa sobre o uso das Tecnologias de Informação e Comunicação nos estabelecimentos de saúde brasileiros – TIC Saúde 2019.** Disponível em: <<https://nic.br/publicacao/pesquisa-sobre-o-uso-das-tecnologias-de-informacao-e-comunicacao-nos-estabelecimentos-de-saude-brasileiros-tic-saude-2019>>. Acesso em: 20 mar. 2021.

NIKOLOVA, S.; HARRISON, M.; SUTTON, M. The Impact of Waiting Time on Health Gains from Surgery: Evidence from a National Patient-reported Outcome Dataset. **Health Economics**, v. 25, n. 8, p. 955–968, 2015.

NKOMO, V. T. *et al.* Burden of valvular heart diseases: a population-based study. **The Lancet**, v. 368, n. 9540, p. 1005–1011, 16 set. 2006.

OBERMEYER, Z.; EMANUEL, E. J. Predicting the Future — Big Data, Machine Learning, and Clinical Medicine. **New England Journal of Medicine**, v. 375, n. 13, p. 1216–1219, 29 set. 2016.

LOUDHOFF, J. P. *et al.* Prioritising patients on surgical waiting lists: A conjoint analysis study on the priority judgements of patients, surgeons, occupational physicians, and general practitioners. **Social Science & Medicine**, v. 64, n. 9, p. 1863–1875, 1 maio 2007.



PAN, I. *et al.* Machine Learning for Social Services: A Study of Prenatal Case Management in Illinois. **American Journal of Public Health**, v. 107, n. 6, p. 938–944, 20 abr. 2017.

PEDROSO, M. C.; MALIK, A. M. As quatro dimensões competitivas da saúde. **Harvard Business Review Brasil**, v. 89, n. 3, p. 58–67, 2015. **Planned Care services**. Disponível em: <<https://www.health.govt.nz/our-work/hospitals-and-specialist-care/planned-care-services>>. Acesso em: 3 mai. 2021.

ROSSI, P.; DWECK, E. Impactos do novo regime fiscal na saúde e educação. **Cadernos de Saúde Pública**, v. 32, n. 12, 2016.

RUFF, C. T. *et al.* Comparison of the efficacy and safety of new oral anticoagulants with warfarin in patients with atrial fibrillation: a meta-analysis of randomised trials. **The Lancet**, v. 383, n. 9921, p. 955–962, 15 mar. 2014.

RUMSFELD, J. S.; JOYNT, K. E.; MADDOX, T. M. Big data analytics to improve cardiovascular care: promise and challenges. **Nature Reviews Cardiology**, v. 13, n. 6, p. 350–359, jun. 2016.

SANTA CATARINA. **Lei nº 17.066, de 11 de janeiro de 2017**. Disponível em: <[http://leis.alesc.sc.gov.br/html/2017/17066\\_2017\\_lei.html](http://leis.alesc.sc.gov.br/html/2017/17066_2017_lei.html)>. Acesso em: 3 mai. 2021.

SANTOS, H. G. DOS. **Comparação da performance de algoritmos de machine learning para a análise preditiva em saúde pública e medicina**. Tese (Doutorado)—[s.l.] Universidade de São Paulo, 28 set. 2018.

SANTOS, H. G. DOS *et al.* Machine learning para análises preditivas em saúde: exemplo de aplicação para predizer óbito em idosos de São Paulo, Brasil. **Cadernos de Saúde Pública**, v. 35, n. 7, 2019.

SHAMOUT, F.; ZHU, T.; CLIFTON, D. A. Machine Learning for Clinical Outcome Prediction. **IEEE Reviews in Biomedical Engineering**, p. 1–1, 2020.

SHICKEL, B. *et al.* Deep EHR: A Survey of Recent Advances in Deep Learning Techniques for Electronic Health Record (EHR) Analysis. **IEEE Journal of Biomedical and Health Informatics**, v. 22, n. 5, p. 1589–1604, set. 2018.

SICILIANI, L.; HURST, J. Tackling excessive waiting times for elective surgery: a comparative analysis of policies in 12 OECD countries. **Health Policy**, v. 72, n. 2, p. 201–215, 1º maio 2005.

SICILIANI, L.; MORAN, V.; BOROWITZ, M. Measuring and comparing health care waiting times in OECD countries. **Health Policy**, v. 118, n. 3, p. 292–303, 1º dez. 2014a.

SICILIANI, L.; MORAN, V.; BOROWITZ, M. Measuring and comparing health care waiting times in OECD countries. **Health Policy**, v. 118, n. 3, p. 292–303, 1º dez. 2014b.

SILVA, G. DO E. S. **A busca pelo cuidado baseado em valor em um hospital universitário**. Dissertação (Mestrado)—São Paulo: Escola de Administração de Empresas de São Paulo, 2018.

SMITH, G. B. *et al.* Review and performance evaluation of aggregate weighted ‘track and trigger’ systems. **Resuscitation**, v. 77, n. 2, p. 170–179, 1 maio 2008.

SOLANS-DOMÈNECH, M. *et al.* Developing a universal tool for the prioritization of patients waiting for elective surgery. **Health Policy**, v. 113, n. 1, p. 118–126, 1 nov. 2013.

STATISTA. **Total data volume worldwide 2010-2024**. Disponível em: <<https://www.statista.com/statistics/871513/worldwide-data-created/>>. Acesso em: 19 mar. 2021.

STIELL, I. G. *et al.* A study to develop clinical decision rules for the use of radiography in acute ankle injuries. **Annals of Emergency Medicine**, v. 21, n. 4, p. 384–390, 1 abr. 1992.

STRODTHOFF, N.; STRODTHOFF, C. Detecting and interpreting myocardial infarction using fully convolutional neural networks. **Physiological Measurement**, v. 40, n. 1, p. 015001, jan. 2019.

TARASOUTCHI, F. *et al.* Atualização das Diretrizes Brasileiras de Valvopatias – 2020. **Arquivos Brasileiros de Cardiologia**, v. 115, n. 4, p. 720–775, 13 out. 2020.

TESTI, A. *et al.* Prioritizing surgical waiting lists. **Journal of Evaluation in Clinical Practice**, v. 14, n. 1, p. 59–64, 2008.

TOPOL, E. J. High-performance medicine: the convergence of human and artificial intelligence. **Nature Medicine**, v. 25, n. 1, p. 44–56, jan. 2019a.

TOPOL, E. J. High-performance medicine: the convergence of human and artificial intelligence. **Nature Medicine**, v. 25, n. 1, p. 44–56, jan. 2019b.

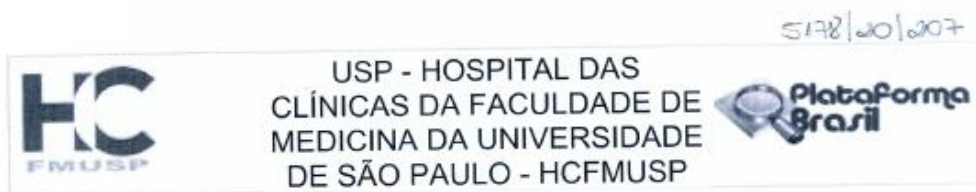
VICTORA, C. G. *et al.* Health conditions and health-policy innovations in Brazil: the way forward. **The Lancet**, v. 377, n. 9782, p. 2042–2053, 11 jun. 2011.

VINCENT, J.-L. *et al.* The SOFA (Sepsis-related Organ Failure Assessment) score to describe organ dysfunction/failure. **Intensive Care Medicine**, v. 22, n. 7, p. 707–710, 1 jul. 1996.

WEST, R. R.; MCKIBBIN, B. Shortening waiting lists in orthopaedic surgery outpatient clinics. **British Medical Journal (Clinical research ed.)**, v. 284, n. 6317, p. 728–730, 6 mar. 1982.

ZHANG JEFFREY *et al.* Fully Automated Echocardiogram Interpretation in Clinical Practice. **Circulation**, v. 138, n. 16, p. 1623–1635, 16 out. 2018.

## ANEXO A – PARECER CONSUBSTANCIADO DO COMITÊ DE ÉTICA EM PESQUISA



### PARECER CONSUBSTANCIADO DO CEP

#### DADOS DO PROJETO DE PESQUISA

**Título da Pesquisa:** Uso de aprendizagem de máquina como suporte à priorização de fila cirúrgica.

**Pesquisador:** MARCO ANTONIO GUTIERREZ

**Área Temática:**

**Versão:** 1

**CAAE:** 39973920.1.0000.0068

**Instituição Proponente:** Hospital das Clínicas da Faculdade de Medicina da USP

**Patrocinador Principal:** Financiamento Próprio

#### DADOS DO PARECER

**Número do Parecer:** 4.438.511

#### Apresentação do Projeto:

Uso de aprendizagem de máquina como suporte à priorização de fila cirúrgica. Trata-se de pesquisa aplicada, quantitativa, experimental e retrospectiva, na qual serão aplicados algoritmos de redes neurais artificiais (uma das técnicas de machine learning). Hipótese:

Modelos preditivos de machine learning são capazes de identificar pacientes em lista de espera para procedimento cirúrgico eletivo que

apresentarão descompensação clínica e necessitarão de cirurgia de urgência ou emergência, dado um determinado período de tempo.

#### Objetivo da Pesquisa:

**Objetivo Primário:**

Identificar os parâmetros e os hiperparâmetros para o modelo de machine learning proposto.

**Objetivo Secundário:**

Identificar a acurácia do modelo proposta para predição de cirurgia de urgência ou emergência em pacientes em fila de espera para cirurgia eletiva.

#### Avaliação dos Riscos e Benefícios:

**Riscos:**

Risco mínimo de divulgação dos dados dos pacientes, contudo o pesquisador se compromete a utilizá-los de forma anonimizada

**Benefícios:**

**Endereço:** Rua Ovídio Pires de Campos, 225 5º andar

**Bairro:** Cerqueira Cesar

**CEP:** 05.403-010

**UF:** SP

**Município:** SÃO PAULO

**Telefone:** (11)2661-7585

**Fax:** (11)2661-7585

**E-mail:** cappelq.adm@hc.fm.usp.br



USP - HOSPITAL DAS  
CLÍNICAS DA FACULDADE DE  
MEDICINA DA UNIVERSIDADE  
DE SÃO PAULO - HCFMUSP



Continuação do Parecer: 4.438.511

Auxiliar médicos e gestores na tomada de decisão em relação à priorização de pacientes em fila de espera.  
Produzir um mecanismo de priorização mais justo e equânime.  
Após validação da ferramenta, a mesma poderá ser utilizada para otimização do processo de gestão da fila de espera de cirurgia valvar.  
Durante o processo de construção do algoritmo, os pesquisadores poderão avaliar a qualidade dos dados registrados em sistema, fomentando a discussão sobre a necessidade de registros corretos.

**Comentários e Considerações sobre a Pesquisa:**

Adequada e de grande relevância pois a proposta do modelo poderá eventualmente ser ampliada para outras especialidades.

**Considerações sobre os Termos de apresentação obrigatória:**

Adequados.

**Recomendações:**

Aprovação sem restrições.

**Conclusões ou Pendências e Lista de Inadequações:**

Sem pendências ou inadequações.

**Considerações Finais a critério do CEP:**

Em conformidade com a Resolução CNS nº 466/12 – cabe ao pesquisador: a) desenvolver o projeto conforme delineado; b) elaborar e apresentar relatórios parciais e final; c) apresentar dados solicitados pelo CEP, a qualquer momento; d) manter em arquivo sob sua guarda, por 5 anos da pesquisa, contendo fichas individuais e todos os demais documentos recomendados pelo CEP; e) encaminhar os resultados para publicação, com os devidos créditos aos pesquisadores associados e ao pessoal técnico participante do projeto; f) justificar perante ao CEP interrupção do projeto ou a não publicação dos resultados.

**Este parecer foi elaborado baseado nos documentos abaixo relacionados:**

Tipo Documento	Arquivo	Postagem	Autor	Situação
Informações Básicas do Projeto	PB_INFORMAÇÕES_BÁSICAS_DO_P ROJETO_1657588.pdf	10/11/2020 12:30:36		Aceito
Outros	8_isencao_tcle.pdf	10/11/2020 12:30:24	Elaine Lagonegro Santana Martinho	Aceito

**Endereço:** Rua Ovídio Pires de Campos, 225 5º andar

**Bairro:** Cerqueira Cesar

**CEP:** 05.403-010

**UF:** SP

**Município:** SAO PAULO

**Telefone:** (11)2661-7585

**Fax:** (11)2661-7585

**E-mail:** cappesq.adm@hc.fm.usp.br



USP - HOSPITAL DAS  
CLÍNICAS DA FACULDADE DE  
MEDICINA DA UNIVERSIDADE  
DE SÃO PAULO - HCFMUSP



Continuação do Parecer: 4.438.511

Outros	8_ISENCAO_TCLE.docx	10/11/2020 12:30:11	Elaine Lagonegro Santana Martinho	Aceito
Orçamento	7_orcamento.pdf	10/11/2020 12:30:03	Elaine Lagonegro Santana Martinho	Aceito
Outros	7_orcamento.docx	10/11/2020 12:29:54	Elaine Lagonegro Santana Martinho	Aceito
Projeto Detalhado / Brochura Investigador	1_Protocolo_14_10_marco_moraes.doc X	10/11/2020 12:29:40	Elaine Lagonegro Santana Martinho	Aceito
Outros	4_anuEncia_FGV.pdf	10/11/2020 12:29:37	Elaine Lagonegro Santana Martinho	Aceito
Outros	4_ANUENCIA_FGV.doc	10/11/2020 12:29:19	Elaine Lagonegro Santana Martinho	Aceito
Outros	3_dados_digitais.pdf	10/11/2020 12:28:52	Elaine Lagonegro Santana Martinho	Aceito
Outros	3_dados_digitais.docx	10/11/2020 12:28:44	Elaine Lagonegro Santana Martinho	Aceito
Outros	2_cronograma.pdf	10/11/2020 12:28:35	Elaine Lagonegro Santana Martinho	Aceito
Outros	2_CRONOGRAMA.docx	10/11/2020 12:28:26	Elaine Lagonegro Santana Martinho	Aceito
Folha de Rosto	fr_pbassinada.pdf	10/11/2020 12:28:12	Elaine Lagonegro Santana Martinho	Aceito
Outros	SGP_ASSINADO.pdf	10/11/2020 12:28:07	Elaine Lagonegro Santana Martinho	Aceito

**Situação do Parecer:**

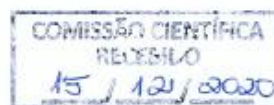
Aprovado

**Necessita Apreciação da CONEP:**

Não

SAO PAULO, 04 de Dezembro de 2020.

Assinado por:  
**ALFREDO JOSE MANSUR**  
(Coordenador(a))



Endereço: Rua Ovídio Pires de Campos, 225 5º andar  
Bairro: Cerqueira Cesar CEP: 05.403-010  
UF: SP Município: SAO PAULO  
Telefone: (11)2661-7585 Fax: (11)2661-7585 E-mail: cappelq.adm@hc.fm.usp.br



## APÊNDICE I – CÁLCULO AMOSTRAL

**Problema:** determinar a margem de erro quando sabemos o tamanho da amostra e da população.

Tamanho da população= 2.436

N = Tamanho da amostra = 588

Z= representa a variável aleatória normal padrão, ou seja,  $Z \sim N(0, 1)$

Z=1,645 para confiança de 90%

Z=1,96 para confiança 95%

Z=2,575 para confiança 99%

p = proporção da população =  $588/2436 = 0,251$

q =  $(1-p) = 0,749$

p\*q = 0,188

e = erro máximo da estimativa

- 1) Considerando **N=588**, intervalo de confiança **95%** e **Z=1,96**, a margem de erro será de **3,5%**:

$$e^2 = Z^2 * (p*q/N) = 0,00123 \rightarrow e=3,5\%$$

- 2) Considerando **N=588**, intervalo de confiança de **99%** e **Z=2,575**, a margem de será de **4,6%**:

$$e^2 = Z^2 * (p*q/N) = 0,0021 \rightarrow e=4,6\%$$

## APÊNDICE II – COMPARAÇÃO DE DESEMPENHO ENTRE OS ALGORITMOS DE REGRESSÃO LOGÍSTICA E FLORESTAS ALEATÓRIAS

**Problema:** determinar se a diferença de desempenho entre os algoritmos de regressão logística e florestas aleatórias são estatisticamente significantes.

**Hipóteses:**

- Hipótese nula (H0): os modelos possuem o mesmo desempenho.
- Hipótese alternativa (H1): há pelo menos uma diferença estatisticamente significativa entre os modelos aplicados.

**Fórmula de cálculo do qui-quadrado:**

$$\chi^2 = \frac{(|n_{01} - n_{10}| - 1)^2}{n_{01} + n_{10}}$$

**Cenário I (Regressão logística e florestas aleatórias com a variável reoperação)**

Tabela de Contingência:

		Regressão Logística	
		Correta	Incorreta
Random	Correta	299(sim/sim)	78 (sim/não)
Forest	Incorreta	84 (não/sim)	127 (não/não)

Cálculo do qui-quadrado com 1 grau de liberdade:

Estatística de teste =  $(\text{sim/não} - \text{não/sim})^2 / (\text{sim/não} + \text{não/sim})$

Estatística de teste com fator de correção =  $(|\text{sim/não} - \text{não/sim}| - 1)^2 / (\text{sim/não} + \text{não/sim})$

	Valor	p-valor
Estatística de teste qui-quadrado	0,222	0,637
Estatística de teste qui-quadrado com fator de correção	0,154	0,694

## Cenário II (Regressão logística e florestas aleatórias sem a variável reoperação)

Tabela de Contingência:

		Regressão Logística	
		Correta	Incorreta
Random Forest	Correta	244 (sim/sim)	112 (sim/não)
	Incorreta	124 (não/sim)	108 (não/não)

Cálculo do qui-quadrado com 1 grau de liberdade:

$$\text{Estatística de teste} = (\text{sim/não} - \text{não/sim})^2 / (\text{sim/não} + \text{não/sim})$$

$$\text{Estatística de teste com fator de correção} = (|\text{sim/não} - \text{não/sim}| - 1)^2 / (\text{sim/não} + \text{não/sim})$$

	Valor	p-valor
Estatística de teste qui-quadrado	0,610	0,435
Estatística de teste qui-quadrado com fator de correção	0,513	0,474