

UNIVERSIDADE FEDERAL FLUMINENSE

PEDRO GEMAL LANZIERI

**Estudo e modelagem do risco de morte por
insuficiência cardíaca em unidade de terapia
intensiva com base em aprendizado de máquina**

NITERÓI

2022

PEDRO GEMAL LANZIERI

**Estudo e modelagem do risco de morte por
insuficiência cardíaca em unidade de terapia
intensiva com base em aprendizado de máquina**

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Computação da Universidade Federal Fluminense como requisito parcial para a obtenção do Grau de Mestre em Computação. Área de concentração: Ciência da Computação

Orientador:
Flávio Luiz Seixas

NITERÓI

2022

Ficha catalográfica automática - SDC/BEE
Gerada com informações fornecidas pelo autor

L297e Lanzieri, Pedro Gemal
Estudo e modelagem do risco de morte por insuficiência cardíaca em unidade de terapia intensiva com base em aprendizado de máquina / Pedro Gemal Lanzieri. - 2022.
140 f.: il.

Orientador: Flávio Luiz Seixas.
Dissertação (mestrado)-Universidade Federal Fluminense, Instituto de Computação, Niterói, 2022.

1. Inteligência artificial. 2. Morte. 3. Insuficiência cardíaca. 4. Unidade de terapia intensiva. 5. Produção intelectual. I. Seixas, Flávio Luiz, orientador. II. Universidade Federal Fluminense. Instituto de Computação. III. Título.

CDD - XXX

Bibliotecário responsável: Debora do Nascimento - CRB7/6368

PEDRO GEMAL LANZIERI

Estudo e modelagem do risco de morte por insuficiência cardíaca em unidade de terapia intensiva com base em aprendizado de máquina

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Computação da Universidade Federal Fluminense como requisito parcial para a obtenção do Grau de Mestre em Computação. Área de concentração: Ciência da Computação

Aprovada em 08 de Dezembro de 2022.

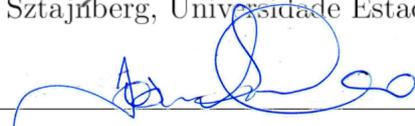
BANCA EXAMINADORA



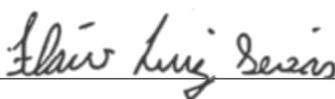
Prof. Ricardo Tavares de Carvalho, Universidade de São Paulo



Prof. Alexandre Sztajnberg, Universidade Estadual do Rio de Janeiro



Prof. José Witerbo Filho, Universidade Federal Fluminense



Prof. Flávio Luiz Seixas - Orientador, Universidade Federal Fluminense

Niterói

2022

Agradecimentos

Essa dissertação de mestrado é dedicada a todos que contribuíram, direta ou indiretamente, na minha trajetória educacional e profissional.

Agradeço a minha família, principalmente meu pai Pedro e minha mãe Cilene, por terem sempre incentivado e proporcionado tudo que fosse necessário para que eu pudesse receber uma educação de qualidade.

Uma vez na Universidade, agradeço ao meu orientador, Professor Doutor Flávio Luiz Seixas, por me acolher em seu grupo de pesquisa. Sou grato por seu apoio, sua paciência e disponibilidade para me direcionar. Obrigado pelo seu exemplo.

Agradeço também aos demais professores da pós-graduação, bem como demais professores com quem tive o privilégio de conhecer ao longo da minha história na Universidade Federal Fluminense. Estendo meus agradecimentos aos demais colaboradores da Universidade.

Sou grato ainda aos colegas de faculdade, de mestrado e de trabalho, com os quais pude crescer junto. A colaboração deles possibilitou a realização deste trabalho.

Meu agradecimento especial é para o meu núcleo familiar mais próximo, minha esposa Carolina e meus filhos Caio e Davi. O apoio incondicional de vocês foi fundamental para a concretização deste meu objetivo de vida. Muito obrigado.

Resumo

Doenças cardiovasculares (DCV) estão entre as principais causas de morte e, para melhorar os desfechos nesses casos, podem ser necessários cuidados intensivos específicos em unidades de terapia cardiointensiva (CICU). Os profissionais de CICU precisam tomar decisões rapidamente, e apoiados por ferramentas classificadoras (escores), cujos resultados são intervalos numéricos associados a um risco de morte do paciente. Os escores nem sempre apoiam a tomada de decisão precisamente, devido a diferenças entre as populações que embasaram a elaboração dessas ferramentas em relação à condição clínica de cada paciente. Por esse motivo, a estratificação de risco para pacientes graves é um desafio no contexto da CICU. Nesse sentido, é possível que os algoritmos baseados em aprendizado de máquina incrementem o desempenho de métodos tradicionais de predição baseados em estatística. O objetivo dessa pesquisa é desenvolver um sistema inteligente de apoio à decisão para os profissionais de saúde que auxilie na estratificação de risco de pacientes. A partir da base de prontuário eletrônico MIMIC-IV, identificamos um padrão de comportamento de um grupo de pacientes com diagnóstico de Insuficiência Cardíaca com maior taxa de mortalidade intra-hospitalar. Por isso, realizamos uma análise de sobrevivência para entender a proporção de mortalidade precoce e tardia, e projetamos uma sobrevida de 86% após sete dias, 70% após 14 dias e 50% após 28 dias. Por meio de um modelo de regressão de Cox, identificamos que a hiperpotassemia foi o fator que de maior efeito sobre a função de risco (*Hazard Ratio*, ou HR, de 1,22), seguido de aumento de creatinina (HR de 1,07). Esses resultados sugeriram a importância da disfunção renal na definição da chance de morte, mas que poderia ser causa ou consequência do diagnóstico de Insuficiência Cardíaca. Por esse motivo, realizamos um estudo de informação mútua para entender o grau de dependência entre as características dos pacientes e seus respectivos desfechos. O nível sanguíneo de creatinina apresentou maior associação com a mortalidade (3,66% de ganho de informação), seguido de débito urinário (2,95%), o que reforçou a associação cardio-renal e fundamentou a seleção das melhores variáveis para modelagem. Por fim, ajustamos um modelo de regressão logística (RL) multivariado para predição de mortalidade. Essa etapa introduziu inteligência aos fatores utilizados para determinar as condições de risco, em contraste com os escores tradicionais de mortalidade,

nos quais as variáveis de entrada são predefinidos e obrigatórias. A acurácia obtida pela RL foi de 78,4%, a área sob a curva ROC foi de 0,685 (valor- $p < 0,001$), e os valores de precisão, revocação e *F-score* foram, respectivamente, de 0,77, 0,80 e 0,76. Por ser um modelo explicável, identificamos que a variável mais associada com morte intra-hospitalar foi o *anion gap*, ou hiato aniônico (*Odds Ratio*, ou OR, de 1,08). Uma das principais contribuições dessa pesquisa foi a proposição de um modelo de decisão adaptável, o qual fornece uma resposta de risco estimado para um novo paciente com interpretações clínicas e sugestões.

Palavras-chave: Tomada de Decisão Clínica, Modelos de Predição, Prontuário eletrônico, Unidade de Terapia Intensiva Especializada.

Abstract

Cardiovascular diseases (CVD) are among the leading causes of death, and specific intensive care in cardio-intensive care units (CICU) may be required to improve outcomes in these cases. CICU professionals need to make decisions quickly, and they are supported by classification tools (scores), which results are numerical ranges associated with a patient's risk of death. However, scores only sometimes accurately support decision-making due to differences between the populations that backed the elaboration of these tools compared to each patient's clinical condition. For this reason, risk stratification for critically ill patients is challenging in the CICU context. In this sense, algorithms based on machine learning may improve the performance of traditional statistical-based prediction methods. This research aims to develop an intelligent decision support system for health professionals that helps in the risk stratification of patients. From the MIMIC-IV electronic medical record database, we identified a pattern of behavior in a group of patients diagnosed with heart failure with a higher in-hospital mortality rate. We then performed a survival analysis to understand the proportion of early and late mortality and projected survival of 86% after seven days, 70% after 14 days, and 50% after 28 days. Using a Cox regression model, we identified that hyperkalemia was the factor with the most significant effect on risk function (Hazard Ratio, or HR, of 1.22), followed by increased creatinine values (HR of 1.07). These results suggested the importance of renal dysfunction in defining the chance of death. However, it remained unclear whether this could be a cause or consequence of the diagnosis of heart failure. For this reason, we performed a mutual information study to understand the degree of dependence between patient characteristics and their respective outcomes. Serum bicarbonate level showed the most significant association with mortality (3.66% of information gain), followed by urinary output (2.95%). Those findings reinforced the cardiorenal association with death and supported the selection of the best variables for modeling. Finally, we fitted a multivariate logistic regression (LR) model to predict mortality. This step introduced intelligence to the factors used to determine risk conditions, in contrast to traditional mortality scores, in which the input variables are predefined and mandatory. The model's accuracy was 78.4%, the area under the ROC curve was 0.685 (p-value < 0.001), the precision was 0.77, the recall was

0.80, and F-score was 0.76. As it is an explainable model, we identified that the variable most associated with in-hospital death was the anion gap (Odds Ratio, or OR, of 1.08). One of the main contributions of this research was the proposition of an adaptable decision model, which provides an estimated risk response for a new patient with clinical interpretations and suggestions.

Keywords: Clinical Decision-Making Making, Prediction Models, Electronic Medical Records, Specialized Intensive Care Unit.

Lista de Figuras

1	Esquema do desenho do estudo.	19
2	Esquema didático representativo da base MIMIC-IV.	21
3	Etapas da construção da coorte de pacientes.	22
4	Representação da função sigmoide.	31
5	Etapas da revisão sistemática da literatura.	45
6	Comparação do erro absoluto médio avaliado para cada uma das técnicas de preenchimento de dados. MAE: Erro absoluto médio.	61
7	Distribuição do número de pacientes sobreviventes e morte intra-hospitalar ao longo do tempo de internação.	64
8	Curva de sobrevivência de Kaplan-Meier, definidas como a probabilidade de sobreviver em um determinado período. A área sombreada representa o IC.	68
9	Importância relativa calculada para o conjunto de variáveis analisado. . . .	71
10	Comparação entre o valor médio da área sob a curva ROC dos algoritmos de aprendizado de máquina supervisionado avaliados. A reta representa um modelo sem valor preditivo.	72
11	Acurácia (média, em verde, e desvio padrão, em laranja) do modelo de RL para diferentes números de variáveis preditoras no eixo X.	85
12	Classes de comentários.	89
13	Representação do arquivo de ajuda disponível no repositório.	131
14	Reprodutibilidade da pesquisa usando o notebook.	132
15	Documentação técnica do código principal do sistema.	133
16	Tela inicial do sistema.	134

17	Exemplo de preenchido do sistema com os requisitos mínimos.	135
18	Primeiro e segundo bloco de informações.	136
19	Terceiro e quarto bloco de informações.	137
20	Arquitetura do código desenvolvido conforme o padrão MVC	138
21	Estudos realizados com a base MIMIC-IV.	140

Lista de Tabelas

1	Artigos selecionados para leitura na íntegra ao final da RSL. RL: Regressão Logística; AD: Árvore de Decisão; KNN: K-Vizinhos mais Próximos; SVM: Máquina de Vetor de Suporte; RN: Rede Neural; RF: Floresta Aleatória; RB: Rede Bayesiana; XGB: Gradiente Extremo; N/A: Não se aplica.	44
2	Medidas de desempenho dos modelos de aprendizado de máquina avaliados.	73
3	Coefficientes (<i>exp</i>) e razão de chances (OR: <i>Odds ratio</i>) calculados a partir do modelo de RL. T-Hosp: Tempo de internação hospitalar; T-CICU: Tempo de internação na CICU; IC: Intervalo de confiança; *: IC não inclui o valor 1	74
4	Respostas em campo aberto fornecidas pelos avaliadores do sistema.	89
5	Principais algoritmos de aprendizado de máquina supervisionado considerados para uso nos estudos com a base MIMIC-IV.	120
6	Principais medidas de avaliação de desempenho para modelos de classificação. VP: Taxa de verdadeiro positivo; VN: Taxa de verdadeiro negativo; FP: Taxa de falso positivo; FN: Taxa de falso negativo; AUROC: área sob a curva <i>Receiver Operating Characteristics</i>	120
7	Dados selecionados da base MIMIC-IV.	122
8	Análise de dados <i>outliers</i> , com comparação entre os intervalos considerados normais para as variáveis do estudo, além da faixa de valores impossíveis. .	123
9	Perfil epidemiológico, clínico e laboratorial da população do estudo. DP: Desvio-padrão; T-CICU: Tempo de internação na CICU; T-H: Tempo de internação hospitalar; FC: frequência cardíaca; FR: frequência respiratória; PAS: pressão arterial sistólica; PAD: pressão arterial diastólica; PAM: pressão arterial média; SO2: oximetria de pulso; TAP: tempo de protrombina; PTT: tempo de tromboplastina; NT-ProBNP: fragmento N-terminal do peptídeo natriurético tipo B.	123

-
- 10 Tempo médio de sobrevivência de diferentes grupos populacionais baseados em comorbidades. Um mesmo paciente pode pertencer a mais de um grupo. *: *valor – p* menor que 0,05. 124
- 11 Riscos proporcionais pelo modelo de Regressão de Cox para o conjunto inicial de variáveis, que não incluiu exames laboratoriais. FC: frequência cardíaca; FR: frequência respiratória; PAS: pressão arterial sistólica; PAD: pressão arterial diastólica; PAM: pressão arterial média; SO2: oximetria de pulso. 124
- 12 Riscos proporcionais, com variáveis de exames laboratoriais, pelo modelo de Cox ajustado para idade, gênero e comorbidades. TAP: Tempo de Protrombina; PTT: Tempo de Tromboplastina. *: *valor – p* < 0,05. 124

Lista de Abreviaturas e Siglas

AD Árvore de Decisão

AUROC Área sob a curva ROC, do inglês, *Area Under the ROC Curve*

CICU Unidade de Terapia Coronariana, do inglês, *Coronary Intensive Care Unit*

CID Código Internacional de Doenças

CSV Valores Separados por Vírgula, do inglês, *Comma Separated Values*

DCV Doenças cardiovasculares

DM Diabetes Mellitus

DPOC Doença Pulmonar Obstrutiva Crônica

ECR Ensaio Clínico Randomizado

HAS Hipertensão Arterial Sistêmica

HR *Hazard rate*

ICFEN Insuficiência Cardíaca com Fração de Ejeção Normal

KM Kaplan-Meier

KNN K-Vizinhos mais Próximos, do inglês, *K-Nearest Neighbors*

MAR *Missing at Random*

MCAR *Missing Completely at Random*

MIMIC *Medical Information Mart for Intensive Care*

MIMIC-IV *Medical Information Mart for Intensive Care*, versão 3

MIMIC-IV *Medical Information Mart for Intensive Care*, versão 2

- MIMIC-IV** *Medical Information Mart for Intensive Care*, versão 4
- MNAR** *Missing Not at Random*
- MVC** *Model View Controller*
- MVS** Máquina de Vetor de Suporte
- NB** Naive Bayes
- NIH** *National Institute of Health*
- PEP** Prontuário Eletrônico do Paciente
- RB** Rede Bayesiana
- RD** Rede de Dependência, do inglês, *Dependency Network*
- RF** Floresta Aleatória, do inglês, *Random Forest*
- RL** Regressão Logística
- RN** Rede Neural
- ROC** *Receiver operating characteristic*
- RSL** Revisão Sistemática da Literatura
- SQL** Linguagem de Consulta Estruturada, do inglês, *Structured Query Language*
- SUS** Sistema Único de Saúde
- UTI** Unidade de Terapia Intensiva
- XGB** Gradiente Extremo, do inglês, *Extreme Gradient Boosting*

Sumário

1	Introdução	12
1.1	Definição do Problema	15
1.2	Objetivos	15
1.3	Organização da Dissertação	16
2	Materiais e Métodos	17
2.1	Desenho do Estudo	19
2.2	Base de Dados	20
2.2.1	População e Amostra	21
2.3	Procedimentos específicos	23
2.3.1	Análise Exploratória de Dados	23
2.3.2	Análise Estatística	24
2.3.3	Seleção de Características	25
2.3.4	Análise de Sobrevida	26
2.3.5	Aprendizado de Máquina	29
3	Fundamentos de Saúde Digital	34
3.1	Prontuário Eletrônico do Paciente	34
3.2	Classificação Internacional de Doenças	38
3.3	Unidade de Terapia Intensiva Cardíaca	38
3.4	Tomada de Decisão Clínica	39
4	Revisão da Literatura	41

4.1	Metodologia da Revisão Bibliográfica	41
4.2	Estudos de Predição de Desfechos em Saúde	46
4.3	Escores de Risco	49
4.4	Estudos com a Base MIMIC	52
5	Análises e Resultados	56
5.1	Pre-processamento	56
5.1.1	Tratamento dos Dados	58
5.2	Estudos realizados	63
5.2.1	Análise Exploratória de Dados	63
5.2.2	Análise Estatística	64
5.2.3	Análise de Sobrevivência	67
5.2.4	Seleção de Características	70
5.2.5	Aprendizado de Máquina	72
5.3	Discussão dos resultados	75
6	Apresentação do Sistema de Apoio à Decisão	79
6.1	Objetivo do Sistema de Apoio à Decisão	79
6.1.1	Análise do Sistema	80
6.1.2	Protótipo do Sistema	81
6.1.3	Projeto do Sistema	83
6.1.4	Ajuste do Modelo	84
6.2	Implementação do Sistema de Apoio à Decisão	86
6.3	Avaliação da Usabilidade do Sistema	87
6.3.1	Planejamento	87
6.3.2	Execução	88
6.3.3	Resultados da Avaliação	88

6.4	Implantação do Sistema	89
7	Conclusão	91
7.1	Contribuições	92
7.2	Trabalhos Futuros	93
	REFERÊNCIAS	95
	Apêndice A - Tabelas de informações complementares	120
	Apêndice B - Tabelas de análise	121
	Apêndice C - Questionário de validação do sistema	125
	Apêndice D - Documentos auxiliares do repositório	131
	Apêndice E - Documentos representativos do sistema	134
	Apêndice F - Figuras complementares aos estudos	139

1 Introdução

Doenças cardiovasculares (DCV) estão entre as principais causas de óbito no mundo. Para melhorar os desfechos nesses casos, podem ser necessários cuidados intensivos específicos (KASAOKA, 2017). O perfil de internações nas Unidade de Terapia Coronariana, do inglês, *Coronary Intensive Care Unit* (CICU) inclui pacientes com diferentes faixas etárias, múltiplas comorbidades, e causas de internação que podem ou não ser primariamente cardíacas (DAI *et al.*, 2020). A tendência de envelhecimento populacional no mundo representa um aumento na prevalência dessas doenças e da demanda por CICU (MORROW *et al.*, 2012). A taxa de diagnóstico de Insuficiência Cardíaca na população é considerável, com expectativa de que, em indivíduos com 55 anos, quase 1 em cada 3 desenvolverá insuficiência cardíaca durante o restante da vida (BLEUMINK *et al.*, 2004). Isso representa, atualmente, mais de 64 milhões de pacientes com Insuficiência Cardíaca (SAVARESE *et al.*, 2022).

Os profissionais que atuam em unidades de terapia intensiva frequentemente precisam tomar decisões rapidamente, baseados em seu julgamento clínico. Por este motivo, toda ferramenta capaz de auxiliar o entendimento de fatores que aumentam ou diminuem a criticidade de um quadro clínico poderá auxiliá-los na tomada de decisão, como ocorre com o uso dos escores (LIGHTHALL; VAZQUEZ-GUILLAMET, 2015). Faz parte da prática de cuidados em utilizar de dados de prontuário para caracterizar o estado de saúde do paciente e, com base nessas informações, prever resultados futuros como, por exemplo, o tempo previsto de internação ou a chance do paciente morrer durante a internação hospitalar (KANE *et al.*, 2007).

A tomada de decisão clínica é frequentemente apoiada pelo uso de ferramentas classificadoras que produzem escores, cujos resultados são intervalos numéricos associados a algum risco ao paciente. Os escores mais utilizados na prática da terapia intensiva, com a ressalva de que não foram necessariamente desenvolvidos especificamente para esta população, são: APACHE (KNAUS *et al.*, 1991), SOFA (VINCENT *et al.*, 1996), LODS (GALL, J., 1996), OASIS (JOHNSON; KRAMER; CLIFFORD, 2013) e SAPS II

(GALL, J. R. L., 1993). O *Mayo CICU Admission Risk Score* (M-CARS) foi proposto recentemente para esta finalidade, validado no contexto de CICUs, e segue em testes com populações maiores para avaliar calibração e concordância (CHICHAREON *et al.*, 2021). Nesse sentido, existem escores para populações gerais, para setores hospitalares e para doenças específicas, mas cujas sobreposições de contextos podem torná-los menos precisos.

É prática comum na medicina a estratificação de risco dos pacientes para gerenciar melhor as prioridades de cuidado (NAGAI; NAKAO; ANZAI, 2021), o que é facilitado pelo uso dos escores. Entretanto, as características diversas de pacientes com DCV torna desafiadora a sua classificação em grupos para estratificação de risco. Isso ocorre pela dificuldade de generalização de dados coletados de diferentes contextos, como o de ambulatórios, Unidade de Terapia Intensiva e CICUs, e também pela necessidade de reavaliar a aplicabilidade de marcadores previamente estudados em outros contextos. Um exemplo dessa situação seria o de um determinado exame laboratorial avaliado como um preditor de mortalidade na UTI, mas que não foi validado em CICU (KADOOKA *et al.*, 2021). Por essa razão, existe necessidade de pesquisa tanto do perfil de pacientes, quanto dos diferentes contextos em que eles procuram atendimento em saúde. O interesse nas pesquisas em CICU existe também porque, até a presente data, poucas ferramentas de predição de risco estão disponíveis para prática clínica nesse contexto (JENTZER *et al.*, 2019; TORRALBA *et al.*, 2020). Por este motivo, as sociedades médicas têm incentivado pesquisas para orientar condutas baseadas em evidências também em CICUs (BOHULA *et al.*, 2019).

A geração contínua de registros clínicos e laboratoriais em prontuário eletrônico vai ao encontro desse incentivo para realização de estudos de DCV no contexto das CICU. A estruturação desses dados na forma do prontuário eletrônico do paciente tem potencial de contribuir positivamente para as práticas assistenciais de saúde (AÑEL RODRÍGUEZ *et al.*, 2021). Ao analisar dados eletrônicos de experiências de muitos provedores com muitos pacientes, pode-se chegar mais próximo de uma resposta sobre qual seria, de fato, a melhor conduta para cada paciente (PINTO; LUO; CHURPEK, 2018). Diversos dados clínicos e de exames complementares já foram estudados como fatores de risco de mortalidade entre pacientes graves, e este conhecimento é usado para melhorar a gestão de recursos hospitalares físicos e humanos (BAKER; XIANG; ATKINSON, 2020). Entretanto, diferenças entre as populações podem alterar o efeito de determinados fatores, de forma que o mais confiável, apesar de difícil em termos práticos, seja definir o risco específico para cada população e em cada contexto de cuidado (ROSENBERG; WATTS, 2000). Além

disso, a avaliação dos dados de prontuário pode permitir a identificação de indivíduos de alto risco. A análise de sobrevivência pode ser útil, nestes casos, para analisar o tempo até o evento de interesse (óbito), de forma estimar o tempo de sobrevivência esperado, dada as observações clínicas de uma coorte de pacientes (SINGH; MUKHOPADHYAY, 2011). Dentre os fatores descritos, as diferenças populacionais e regionais merecem destaque porque tornam o cenário de pesquisa desafiador. Isso pode significar diversidade de características dos pacientes (perfil epidemiológico e de comorbidades), por diferenças nos processos de saúde (pacientes com diferentes complexidades sendo atendidos pela mesma unidade de saúde), ou mesmo por limitações dos recursos disponíveis em cada local (recursos materiais, humanos, e mesmo relacionados a disponibilidade de tecnologias). Nesses casos, por exemplo, comparar as taxas gerais de mortalidade entre unidades de pronto atendimento, um hospital universitário e um hospital privado poderão ser mais representativos das diferenças sociais do que de diferenças nos padrões ou na qualidade de atendimento (FUTOMA *et al.*, 2021; SINGH; MHASAWADE; CHUNARA, 2022).

Com os dados detalhados de pacientes e processos clínicos registrados em Prontuário Eletrônico do Paciente, é possível reduzir algumas das incertezas que permeiam os cuidados de saúde, as quais são agravadas pelas limitações de tempo e recursos para a tomada de decisão. Por esse motivo, existe uma oportunidade para os pesquisadores encontrarem evidências clínicas a partir dessa abundância de dados que já existem, com uso de técnicas aplicáveis a grandes volumes de dados, tais como as técnicas de aprendizado de máquina. O uso de modelos de aprendizado de máquina para predição de tempo de internação e mortalidade é uma oportunidade de pesquisa bastante explorada (BOROUJENI; YOUSEFI; ZURETTI, 2019; CHURPEK *et al.*, 2016; DONNELLY *et al.*, 2021; SOMANI *et al.*, 2020).

Por outro lado, pouco avanço foi feito em modelos que apliquem um determinado conhecimento em uma grande proporção de hospitais, devido a diversos desafios. Dentre eles, destacam-se: (1) a carência de dados de registro em formato de prontuário eletrônico, (2) a pouca interoperabilidade de dados e sistemas de PEP, (3) a necessidade de compreensão os conceitos técnicos envolvidos na análise dos resultados de um modelo, e (4) a necessidade lidar com dados desbalanceados, como frequentemente ocorre em desfechos como a mortalidade hospitalar (JANETT; YERACARIS, 2020; HOWE *et al.*, 2018; SARIPALLE; RUNYAN; RUSSELL, 2019). O desbalanceamento dos dados de treinamento é esperado porque um número proporcionalmente maior de pacientes tendem a sobreviver e ter como desfecho a alta hospitalar, e um número comparativamente menor morre durante a internação (HARERIMANA; KIM; JANG, 2021).

1.1 Definição do Problema

A estratificação de risco para pacientes graves é um desafio no contexto da Terapia Intensiva. Há uma variedade de ferramentas classificadoras que podem ser utilizados na CICU, porém seus resultados são limitados e imprecisos (PIRRACCHIO *et al.*, 2015). A melhoria da predição de risco em pacientes graves pode contribuir para orientação de uma oferta de cuidado mais adequada, ao apoiar decisões sobre tratamentos e gestão de recursos disponíveis (JENTZER *et al.*, 2019).

Algoritmos baseados em aprendizado de máquina podem incrementar o desempenho de métodos tradicionais de predição baseados em estatística. Paralelamente, o uso de dados de PEP segue o caminho da pesquisa clínica contemporânea, que tende para a utilização crescente de *big data* para melhorar a prestação de cuidados aos pacientes (CELI; MARK *et al.*, 2013). Dentre as bases de dados de saúde classificadas como de acesso aberto, o *Medical Information Mart for Intensive Care* (MIMIC) se destaca pela quantidade e qualidade dos seus registros. O MIMIC é uma base de dados de acesso público desenvolvido pelo *Laboratory for Computational Physiology*, do *Massachusetts Institute of Technology*, em parceria com o *Beth Israel Deaconess Medical Center* e a *Philips Healthcare*, com o apoio do Instituto Nacional de Imagem Biomédica e Bioinformática. Trata-se de informações relacionadas a pacientes internados de um complexo hospitalar terciário de Boston, Massachusetts, EUA (JOHNSON, A. *et al.*, 2020), que serão utilizadas nos estudos apresentados nesta pesquisa.

Nesta dissertação, a hipótese a ser testada é que o aprendizado de máquina pode auxiliar o julgamento dos profissionais de saúde na identificação de dados clínicos suficientes para estratificar os pacientes com Insuficiência Cardíaca com base no risco de morte durante sua internação na CICU.

1.2 Objetivos

O objetivo principal desta pesquisa é desenvolver um sistema de apoio à decisão para os profissionais de saúde para auxiliar na estratificação de risco de pacientes admitidos no hospital, dado um conjunto de dados clínicos do paciente.

Os objetivos secundários são:

1. Pesquisar, propor e validar modelos de aprendizado de máquina apropriados aos

processos de saúde presentes em uma CICU;

2. Analisar e projetar um sistema de apoio à decisão para os profissionais de saúde da CICU para ajudar na avaliação das condições de risco.

1.3 Organização da Dissertação

Os próximos capítulos estão organizados da seguinte forma. O Capítulo 2 descreve os processos de pesquisa aplicados, e apresenta o desenho do estudo, o mapeamento de uso dos dados e os procedimentos específicos de estudo. O Capítulo 3 apresenta a fundamentação teórica dos modelos de tomada de decisão aplicados à saúde, com os conceitos relacionados ao uso de dados abertos e anonimizados de PEP. O Capítulo 4 apresenta a revisão da literatura, com os trabalhos relacionados. O Capítulo 5 descreve os estudos realizados na base MIMIC-IV, com apresentação dos principais resultados obtidos. O Capítulo 6 apresenta a proposta do sistema e o uso do modelo preditivo por médicos. Por fim, o Capítulo 7 resume as principais contribuições da pesquisa e sugere possíveis trabalhos futuros.

2 Materiais e Métodos

Esta dissertação se encontra linha de pesquisa de Inteligência Artificial e faz parte do domínio da Ciência da Computação. A metodologia dessa pesquisa foi baseada nos fundamentos da *Design Science Research*, um processo de pesquisa originalmente proposto para a engenharia de produção que preza por eficiência e eficácia dos artefatos informacionais, bem como pelo rigor e pela produção de conhecimento científico reprodutível (LACERDA *et al.*, 2013; BLESSING; CHAKRABARTI, 2009).

Didaticamente, a metodologia proposta nesta pesquisa divide-se em seis estágios: (1) formulação do problema, (2) revisão bibliográfica, (3) coleta e exploração dos dados, (4) modelagem, (5) execução e validação dos estudos e (6) proposta de um uso prático ao modelo. Este capítulo descreve em detalhes os passos de pesquisa em cada uma dessas etapas.

O processo de desenvolvimento da questão de pesquisa envolve a definição de vários componentes do estudo, e também a seleção do tipo de estudo mais apropriado para utilizar esses componentes para produzir resultados válidos e confiáveis. A primeira etapa no processo de transformar uma questão clínica em questão de pesquisa é definir a amostra do estudo e o resultado de interesse. Nessa pesquisa, a população estudada foi selecionada a partir da base de dados de prontuário eletrônico de pacientes do MIMIC-IV, e o desfecho de interesse foi a morte intra-hospitalar. O passo seguinte é a escolha do desenho de estudo ideal conforme a questão e os recursos disponíveis. Nesse caso, o estudo de modelagem preditiva foi selecionado, de modo que fosse possível usar os dados disponíveis dos pacientes para determinar quais fatores contribuíram para o desfecho.

Na etapa seguinte, a Revisão Sistemática da Literatura (RSL) apresenta os fundamentos teóricos, o que possibilita a compreensão do estado atual de conhecimentos sobre o tema.

A obtenção e a exploração dos dados corresponde à terceira etapa da metodologia, e compreende uma análise exploratória de dados a partir de uma observação transversal,

retrospectiva e unicêntrica. Uma parte crítica desse processo é definir quais tipos de dados seriam necessários para responder à questão de pesquisa e, por isso, inclui desde o estudo dos dados brutos até a visualização e interpretação de dados e cálculos estatísticos. O método descritivo é usado para realizar análises de diversos aspectos da base de pacientes, com a seleção de características clínico-epidemiológicas que posteriormente servem para um procedimento experimental de modelagens.

A modelagem dos dados corresponde à etapa da metodologia em que os conhecimentos e as técnicas de engenharia de dados são aplicados (XING *et al.*, 2016). Nessa pesquisa, ela inclui diversas atividades: (i) a preparação dos dados, com tratamento de registros com valores ausentes, suavização de ruídos, identificação de *outliers* e remoção de dados duplicados; (ii) a integração dos dados, com associação das tabelas de dados de pacientes com dados de internação; (iii) a transformação dos dados, com técnicas de normalização e padronização; (iv) a avaliação sobre a necessidade de balanceamento de dados, reamostragem e redução de dados; e (v) a seleção das características, com avaliação da importância das características, remoção de multicolinearidade, análise de principais componentes e descarte de baixa variância.

A quinta etapa da metodologia consiste nos estudos com a base de dados, com ênfase na aplicação do aprendizado de máquina. Essa fase inclui as decisões sobre o tipo e os modelos de aprendizado de máquina, além do ajuste de hiperparâmetros, e da comparação entre diferentes algoritmos. Essa pesquisa compreendeu técnicas que não são classificadas como aprendizado de máquina, como a análise de sobrevivência e o estudo de entropia, mas que foram necessárias para refinar o entendimento da base de pacientes e das informações disponíveis.

Por fim, na etapa de estudo da aplicabilidade do modelo, o algoritmo de aprendizado de máquina com melhores características para servir como modelo preditivo do sistema é selecionado para compor uma ferramenta de apoio à tomada de decisão. Essa fase compreende, portanto, o desenvolvimento, a avaliação e a disponibilização do sistema para usuários.

A próxima subseção apresenta o desenho do estudo, o qual representa a lógica da sequência de análises realizadas na base de dados de pacientes.

2.1 Desenho do Estudo

O delineamento dessa pesquisa compreende três etapas: (1) descritiva, que inclui uma análise exploratória de dados a partir de uma observação transversal, retrospectiva e unicêntrica; (2) teste de hipóteses em uma coorte de pacientes da base descrita na etapa anterior, na qual existem diferenças entre as subpopulações com tem relevância estatística a ponto de fazerem parte de um conjunto de dados mínimo capaz de prever um desfecho; e (3) preditiva, que compreende o desenvolvimento de um sistema de apoio à decisão a partir de modelos treinados e validados com a base MIMIC-IV. Essas etapas são viabilizadas por uma série de estudos, representados no esquema da Figura 1.

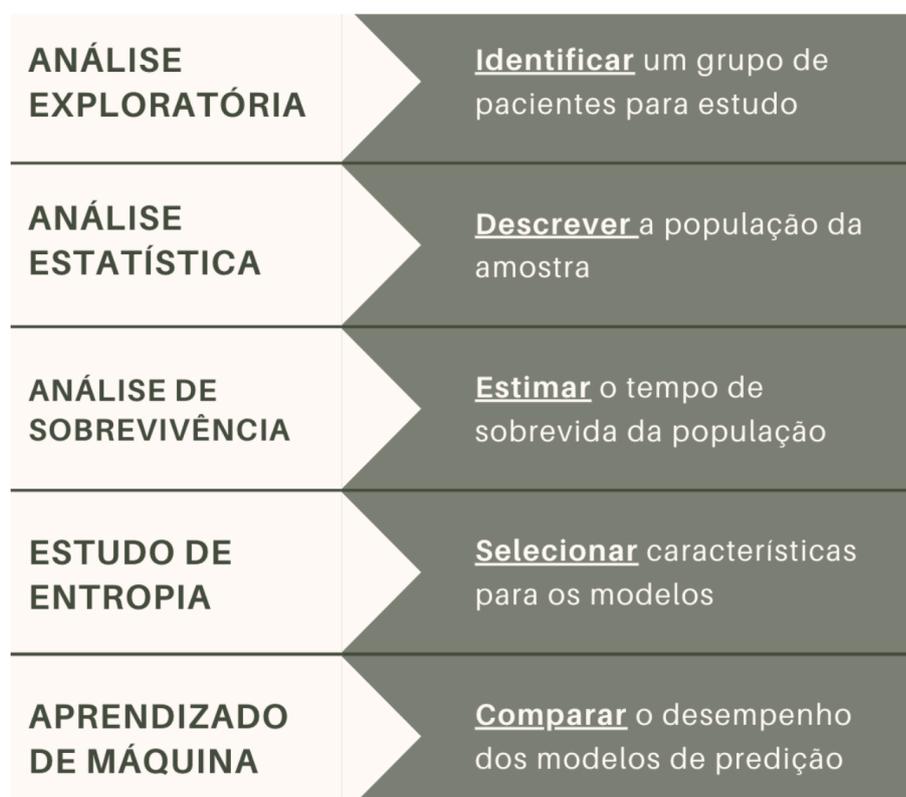


Figura 1: Esquema do desenho do estudo.

Esse desenho de estudo foi elaborado a partir da disponibilidade de uma base com significativa heterogeneidade de perfis de doença e de causas de internação. A partir da base de dados de prontuário eletrônico de pacientes MIMIC-IV, é identificado um padrão de comportamento de um grupo de pacientes com diagnóstico de Insuficiência Cardíaca, o que motiva uma busca por explicações e associações com a mortalidade dessa coorte. Uma análise de sobrevivência é realizada, e permite projetar curvas de sobrevida e identificar, por meio de um modelo de regressão de Cox, quais fatores mais se associaram a mortalidade dessa população. Um estudo de informação mútua possibilita entender o grau

de dependência entre as variáveis e a mortalidade e possibilitou a seleção das melhores características para modelagem. Por fim, os estudos com modelos de aprendizado de máquina supervisionado são utilizados com objetivo de predição de mortalidade.

Esta pesquisa não requer consentimento informado por apresentar risco mínimo para os pacientes. Os dados de identificação foram ocultados para garantia de privacidade, e a investigação está em acordo com os princípios descritos na Declaração de Helsinki¹([WORLD MEDICAL ASSOCIATION, 2013](#)). Após a conclusão do treinamento *online* do *National Institute of Health* (NIH) e do exame *Protecting Human Research Participants*, foi concedida permissão ao pesquisador para acesso aos dados da base MIMIC-IV. O projeto MIMIC foi aprovado pelo *Beth Israel Deaconess Medical Center* e pelo Conselho de Revisão Institucional do *Massachusetts Institute of Technology* (Cambridge, Massachusetts).

O acesso à MIMIC-IV é restrito a usuários autorizados por um credenciamento pelo portal [Physionet](#)² ([JOHNSON, A. et al., 2020](#)). É exigido o treinamento *CITI Data or Specimens only course* do *National Institutes of Health*, seguido da realização do exame *Protecting Human Research Participants*. Esse treinamento *online* abrange aspectos importantes da pesquisa com dados de participantes humanos. O autor dessa dissertação foi aprovado na qualificação com o número 45395590, e foi o único responsável pelo acesso aos arquivos brutos do banco de dados. A subseção a seguir apresenta os detalhes da base MIMIC-IV.

2.2 Base de Dados

O MIMIC-IV é um banco de dados disponível gratuitamente que compreende dados não identificados relacionados à saúde de mais de 70000 admissões em UTI. Trata-se de pacientes com 18 anos ou mais, internados no *Beth Israel Deaconess Medical Center*, entre 2008 e 2019, conforme detalhado na Figura 2.

A Tabela 7, disponível no Apêndice B, apresenta as *features* disponíveis para estudo na base MIMIC-IV. Na subseção a seguir são descritas as decisões para seleção de um grupo específico de pacientes e informações contidos na base MIMIC-IV.

¹Disponível em: <https://www.wma.net/policies-post/wma-declaration-of-helsinki-ethical-principles-for-medical-research-involving-human-subjects/>. Acesso em: 31 de dezembro de 2022.

²Disponível em: <https://physionet.org>. Acesso em: 31 de dezembro de 2022.

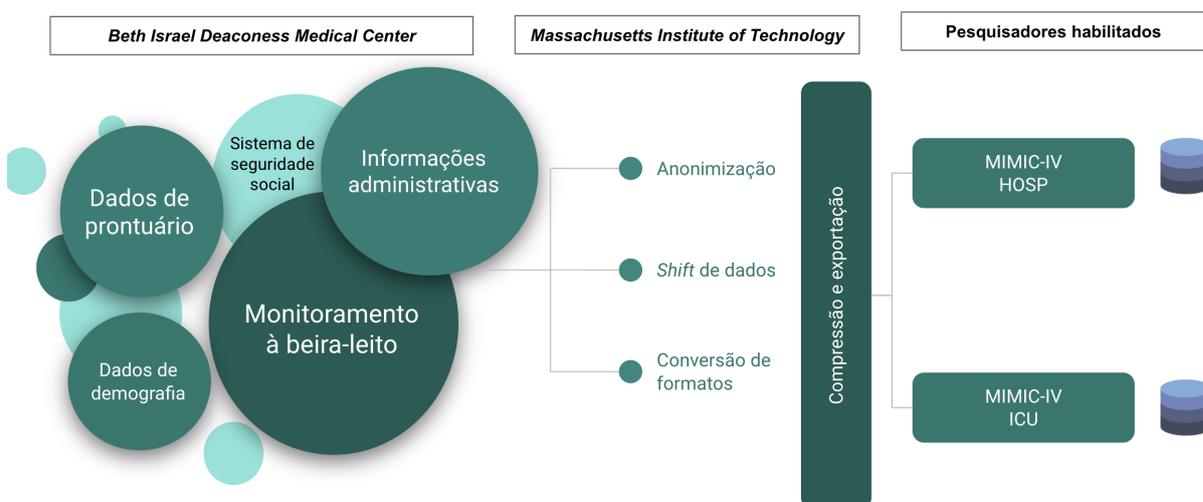


Figura 2: Esquema didático representativo da base MIMIC-IV.

2.2.1 População e Amostra

Um passo crítico para este estudo foi a seleção de uma coorte de pacientes apropriada para análise. A decisão sobre os critérios para construção da coorte do estudo resultou em um grupo de pacientes com Insuficiência Cardíaca internados na CICU. Dessa forma, torna-se possível treinar um modelo a partir de informações desse perfil de pacientes, o que vai ao encontro do objetivo de auxiliar na tomada de decisão de profissionais que assistem pacientes com Insuficiência Cardíaca internados na CICU. Para garantir uma amostra representativa, foram estabelecidos critérios de seleção dos pacientes do estudo.

Os critérios de inclusão foram:

1. Idade igual ou superior a 18 anos;
2. Admissões em unidade CICU;
3. Registro em prontuário de um ou mais código de diagnóstico (CID-9) representativos do diagnóstico de Insuficiência Cardíaca, para seleção de pacientes que tenham essa doença: CID-9 428.

Os critérios de exclusão foram:

1. Registros com informações inválidas ou incompletas para essa pesquisa, ou seja, sem medições de frequência cardíaca, ou procedimentos administrativos incompletos, sem registro de admissão ou alta da UTI;

2. Doadores de órgãos, os quais muitas vezes são registradas como “readmissões” para fins administrativos;
3. Internações com duração inferior a 4 horas, porque geralmente são pré-operatórias ou então para posterior transferência para outro setor do hospital.

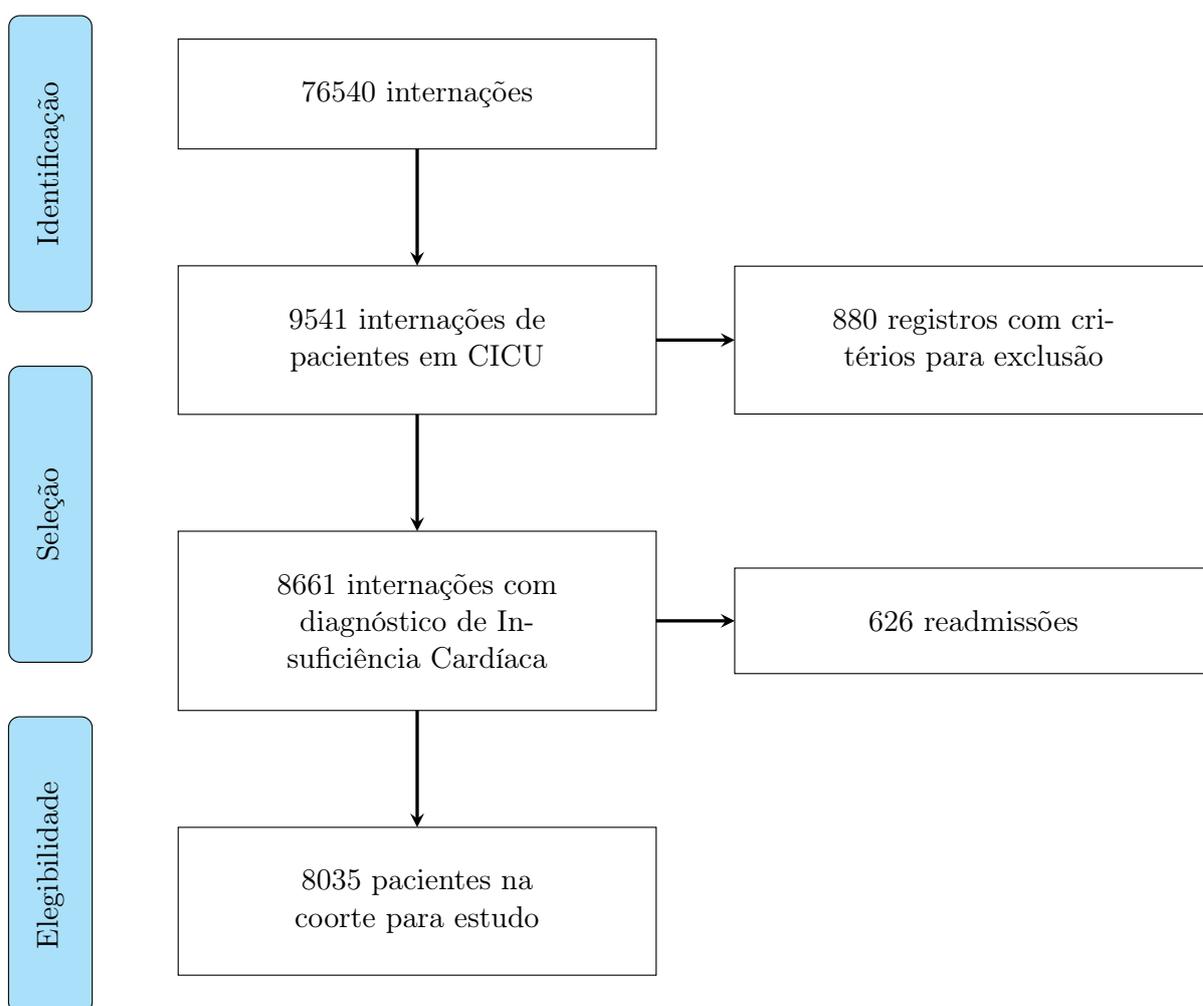


Figura 3: Etapas da construção da coorte de pacientes.

Foi necessária uma combinação de técnicas orientadas por dados e raciocínio clínico para ser possível filtrar esses pacientes. Uma vez que a base MIMIC-IV agrupa todos os pacientes de terapia intensiva, sem distinção do setor (por exemplo, cardíaco), as seguintes consultas precisaram ser realizadas para garantir os critérios de inclusão descritos no início deste capítulo:

- Internação em CICU;
- Primeira internação do paciente;

- Presença de registro CID-9 compatível com diagnóstico de Insuficiência Cardíaca;
- Presença da informação de ocorrência ou não de morte intra-hospitalar (resultado de interesse).

Uma vez definida o desenho do estudo e a população sobre a qual as análises são realizadas, a próxima subseção descreve os procedimentos técnicos específicos que serão aplicados nas análises apresentadas no Capítulo 5.

2.3 Procedimentos específicos

Nesta pesquisa, é utilizado o método descritivo para realizar análises de diversos aspectos da base de pacientes, com a seleção de características clínico-epidemiológicas que posteriormente serviram para um procedimento experimental de modelagens estatísticas e de aprendizado de máquina.

A Subseção 2.3.1 apresenta os fundamentos da análise exploratória de dados, os quais são aplicados, nesta pesquisa, para fornecer um entendimento geral das informações disponíveis. Na Subseção 2.3.2, são descritas as técnicas de análise estatística usadas nesta pesquisa para estudo da população e para comparação entre grupos de pacientes. A Subseção 2.3.4 apresenta os conceitos relacionados à análise de sobrevivência e à análise de riscos proporcionais. Na Subseção 2.3.3, é apresentado o racional para seleção de características da base de dados a partir do estudo de entropia e de ganho de informação, e a Subseção 2.3.5 descreve as técnicas de aprendizado de máquina que fazem parte do estudo preditivo apresentado no Capítulo 5.

O passo inicial de entendimento da base MIMIC-IV é concebido a partir de uma análise exploratória de dados, com objetivo de gerar visualizações da base e de identificar um padrão de doença para servir como objeto de estudo. A subseção a seguir apresenta os principais conceitos associados a esta etapa da pesquisa.

2.3.1 Análise Exploratória de Dados

A análise exploratória de dados é uma etapa essencial em muitos tipos de pesquisa, nos quais se destaca a análise de registros eletrônicos de saúde (KOMOROWSKI *et al.*, 2016). Em geral, ela é efetuada após a coleta e o pré-processamento dos dados.

Ela pode ser considerada um passo fundamental na geração de hipóteses de pesquisa,

principalmente se o pesquisador iniciar a exploração de um conjunto de dados em busca de informações esperadas e padrões, além da seleção de variáveis (FIFE; RODGERS, 2021). A visualização e a compreensão dos dados pode ser feita por métodos gráficos ou não gráficos, para examinar os dados para distribuição, valores discrepantes e anomalias, e assim direcionar o teste específico de hipóteses (OCONNOR *et al.*, 2020).

A exploração permite compreender melhor as características e possíveis problemas de um conjunto de dados. No caso dos dados de saúde, as abordagens podem ser univariadas (apenas uma variável, exposição ou desfecho) ou multi-variadas (várias variáveis de exposição isoladas ou com uma variável de desfecho) (KOMOROWSKI *et al.*, 2016).

De uma forma geral, a pesquisa com dados de saúde requer uma etapa de compreensão da estrutura do banco de dados, com o reconhecimento dos relacionamentos e das interdependências existentes. Esse conhecimento possibilita visualizar relações potenciais (direção e magnitude) entre as variáveis de exposição e desfecho, além de detectar possíveis anomalias e valores discrepantes. Isso possibilitará que sejam extraídas (ou derivadas) variáveis clinicamente relevantes (NASSIRI *et al.*, 2018).

Uma análise estatística será feita nesta dissertação, com ênfase nos métodos descritivos, conforme apresentado na subseção a seguir.

2.3.2 Análise Estatística

A análise estatística é uma etapa básica da metodologia de pesquisa na qual os dados são estudados e interpretados, a fim de identificar padrões e tendências (AGRESTI; HITCHCOCK, 2005). Ela é categorizada como parte da ciência matemática e pode fazer parte da base de alguns modelos de aprendizado de máquina, os quais são aplicações de estatística e probabilidade que capturam padrões nos dados por meio do uso de algoritmos computacionais (ROMEIJN, 2022).

Didaticamente, ela pode ser dividida em dois métodos: (1) estatística descritiva, que organiza e agrega informações, mas sem estabelecer relações ou conclusões; e (2) estatística de inferência, a partir da qual podem ser feitos testes de hipótese e conclusões a partir dos dados disponíveis (WARNE *et al.*, 2012). De forma geral, a maioria das análises produzidas na Subseção 5.2.2 são produto de técnicas descritivas, e possibilitam um melhor entendimento das características dos pacientes selecionados.

Uma vez definidos quais pacientes fazem parte da amostra populacional e quais são suas principais características, é pertinente avaliar quais dados são relevantes dentre todo

o universo de variáveis disponíveis, o que é discutido na subseção a seguir.

2.3.3 Seleção de Características

A entropia é uma medida da teoria da informação que mede a impureza ou incerteza em um grupo de observações. O ganho de informação calcula a redução na entropia da transformação de um conjunto de dados, comparando a entropia do conjunto de dados antes e depois. Ele é calculado a partir de uma divisão, em que são subtraídas as entropias ponderadas de cada ramo da entropia original. Quando usado para estimar a dependência estatística entre duas variáveis, para seleção de variáveis, é referido como Informação Mútua (PLUIM; MAINTZ; VIERGEVER, 2003).

A informação mútua entre duas variáveis aleatórias X e Y pode ser declarada formalmente como $I(X,Y) = H(X) - H(X|Y)$, na qual $I(X,Y)$ é a informação mútua para X e Y , $H(X)$ é a entropia para X , e $H(X|Y)$ é a entropia condicional para X dado Y (YUAN; KESAVAN, 1998).

O cálculo da entropia cruzada, para o caso de classificações binárias, onde o número de classes M é igual a 2, pode ser representado pela Expressão 2.1:

$$-(y \log(p) + (1 - y) \log(1 - p)) \quad (2.1)$$

Se $M > 2$, calculamos uma perda separada para cada rótulo de classe por observação e somamos o resultado, conforme apresentado na Expressão 2.2:

$$-\sum_{c=1}^M y_{o,c} \log(p_{o,c}) \quad (2.2)$$

Na Expressão 2.2, M representa número de classes, y é um indicador binário (0 ou 1) se o rótulo da classe c for a classificação correta para uma observação o , e p é a probabilidade prevista da observação para a classe c .

Na Subseção 5.2.4 será apresentado um estudo de entropia e ganho de informação a partir das variáveis disponíveis no conjunto de dados do MIMIC-IV.

Com o intuito de aprofundar o entendimento de como se comporta a população selecionada em termos de sobrevivência ao longo do tempo, a próxima subseção apresenta os principais fundamentos associados a estudos de análise de sobrevivência.

2.3.4 Análise de Sobrevivência

A análise de sobrevivência é um conjunto de processos estatísticos que estuda o tempo desde um instante inicial até a ocorrência de um evento. Em estudos de análise de sobrevivência, entende-se por evento como o objeto de interesse (que pode ser óbito, ou outro, como diagnóstico ou recuperação de uma doença). Já o termo tempo corresponde ao intervalo de tempo desde o início de um período de observação até a ocorrência do evento, ou então ao fim do estudo, ou então por perda da observação (GEORGE; SEALS; ABAN, 2014).

Os dados de sobrevivência são geralmente descritos e modelados em termos de duas probabilidades relacionadas: sobrevivência e risco (CLARK *et al.*, 2003). Eles podem ser usados para analisar características associadas a tempo até a ocorrência do evento de interesse (SCHOBER; VETTER, 2018).

Censura corresponde aos sujeitos da pesquisa em que se tem alguma informação, porém a observação é incompleta. De fato, dificilmente todos os indivíduos de uma coorte são acompanhados até a ocorrência do evento de interesse e, muitas vezes, não se conhece a data de início do acompanhamento da coorte. Nesses casos, o tempo de sobrevivência exato é desconhecido. De maneira simplificada, entende-se censura à direita quando se sabe que o tempo entre o início da observação e o evento é maior do que o tempo observado ($T > t^-$). Ela pode ser classificada como tipo I, se ocorrer o evento de interesse após o período predeterminado de observação, ou do tipo II, no caso de interrupção na observação. Já na censura à esquerda não se conhece o momento da ocorrência do evento, mas é possível afirmar que ocorreu antes do tempo observado ($T > t^+$). Censura intervalar, no que lhe concerne, representa situações onde o evento de censura ocorreu entre duas observações ($t^- < T < T^+$).

A função de sobrevivência $S(t)$ representa a probabilidade de um indivíduo sobreviver desde a origem do tempo até ao menos um tempo futuro especificado t . Isso representa que o evento de morte ainda não ocorreu no tempo t ou, equivalentemente, a probabilidade de um indivíduo sobreviver após o tempo t .

Sendo T um tempo de vida aleatório (possivelmente infinito, mas sempre não negativo) obtido da população em estudo, e t um tempo futuro específico, a função de sobrevivência $S(t)$ pode ser representada pela Equação 2.3:

$$S(t) = Pr(T > t) \quad (2.3)$$

Nessa modelagem, o efeito de um fator de exposição sobre o risco de um desfecho é estimado como uma probabilidade condicional (representada na Equação 2.3 por Pr).

O gráfico de $S(t)$ pelo tempo t é chamado de curva de sobrevivência.

A função de risco, $h(t)$, representa a probabilidade condicional de que um indivíduo que está sob observação em um momento t tenha um evento naquele momento. O risco representa a taxa instantânea do evento de interesse ocorrer no tempo t , para um indivíduo que já sobreviveu até o tempo t , independente ou não de já ter ocorrido anteriormente. A função de risco está representada pela Equação 2.4.

$$h(t) = \lim_{\varepsilon \rightarrow 0} \frac{Pr[(t \leq T < t + \varepsilon) | T \geq t]}{\varepsilon} \quad (2.4)$$

Existem diversos estimadores do modelo de sobrevivência. Os mais utilizados em estudos clínicos, e que serão aplicados nas análises descritas na Subseção 5.2.3 são: no caso de modelo não paramétrico, o estimador de Kaplan-Meier; em termos de modelo paramétrico, o estimador de Weibull; e o modelo de Cox, no caso de estimadores semi-paramétricos (BARAKAT *et al.*, 2019).

O método de Kaplan-Meier pode ser usado para estimar a curva de sobrevivência a partir dos tempos de sobrevida observados sem a suposição de uma distribuição de probabilidade. Ele pode ser representado graficamente como uma série de passos horizontais decrescentes que, com um tamanho de amostra grande o suficiente, se aproxima da verdadeira função de sobrevivência para aquela população.

O modelo de Weibull é uma distribuição de probabilidade contínua. É uma generalização da distribuição exponencial, porém considerado mais eficiente, pois possui um parâmetro adicional que permite ajustar diferentes formas para a função risco. Quando aplicado a análises de sobrevivência, ele considera taxas de falha proporcionais e um tempo de vida acelerado.

O modelo de riscos proporcionais de Cox é o mais utilizado na análise de sobrevivência com dados de saúde (EFRON, 1977). É análogo a um modelo de regressão múltipla, considerado flexível por ser formado por um elemento paramétrico e outro não-paramétrico. Ele permite que a diferença entre os tempos de sobrevida de grupos de pacientes seja testada enquanto incorpora co-variáveis dependentes do tempo.

O modelo de riscos proporcionais de Cox assume que o risco logarítmico de um indivíduo seja uma função linear de suas m co-variáveis $h_i, i \in \{1, \dots, m\}$, além de um risco

de base ao nível populacional que muda ao longo do tempo. Matematicamente, pode ser representado pela Equação 2.5:

$$h(t|x) = h_0(t) \exp \left(\sum_{i=1}^m h_i(x_i - \bar{x}_i) \right). \quad (2.5)$$

O termo riscos proporcionais refere-se à suposição de uma relação constante entre a variável dependente e os coeficientes de regressão. O exponencial desses coeficientes pode indicar o *odds ratio*. Considera-se que um número maior que 1 indica uma contribuição positiva da co-variável ao modelo de sobrevivência (aumenta ou reduz do tempo de sobrevida); se o coeficiente for menor que 1 e maior que 0, representa uma contribuição negativa; e se um número for igual ou próximo de 1, isso sugere que a co-variável é independente, o que pode indicar ausência de relação causal (ZHANG *et al.*, 2018).

Nas regressões de sobrevivência, é feita uma modelagem de regressão com co-variáveis (por exemplo, condições clínicas do paciente) em relação a outra variável — neste caso, tempo. O algoritmo de busca ajustará com base na maximização de uma função de máxima verossimilhança (HANCOCK *et al.*, 2014). Essa função permite avaliar o quanto os dados suportam cada valor possível de um determinado parâmetro estimado.

Nesse modelo, a variável dependente (resposta) é o risco (BEWICK; CHEEK; BALL, 2004). O teste de *log rank* pode ser usado para avaliar se há diferença entre os tempos de sobrevivência de diferentes grupos, embora não permita que outras variáveis explicativas sejam consideradas. É importante destacar que o teste de *log rank* e o modelo de risco proporcional de Cox assumem que a razão de risco é constante ao longo do tempo, o que pode ser considerado uma limitação (IN; LEE, 2019). A ideia subjacente é que o risco logarítmico de um indivíduo é uma função linear de suas co-variáveis, e que existe um risco de linha de base ao nível populacional que muda ao longo do tempo. O risco de um evento que ocorre durante uma observação nem sempre pode ser mantido constante, e a razão de risco não pode ser sempre a mesma. Este é o principal obstáculo para uma análise de dados clínicos usando um modelo de Cox (IN; LEE, 2019).

O índice de concordância avalia a capacidade do modelo em prever o tempo de sobrevivência relativo. Ele pode ser entendido como uma generalização da área sob a curva ROC que considera os dados censurados. Dessa forma, ele representa a avaliação global do poder de discriminação de um determinado modelo (GEORGE; SEALS; ABAN, 2014).

É importante destacar que, embora o modelo de sobrevivência modele diretamente o

risco, ele não se caracteriza como um modelo preditivo, ainda que seja possível estimar-se um risco a partir de uma função de *threshold*. Para ser possível realizar estudos com finalidade de predição, os modelos de aprendizado de máquina são mais adequados, conforme descrito na próxima subseção.

2.3.5 Aprendizado de Máquina

Os modelos de aprendizado de máquina e suas técnicas são amplamente usados em uma variedade de aplicações de previsão, detecção e mineração de dados.

O aprendizado de máquina aplicado à saúde pode auxiliar no entendimento de padrões de dados de saúde. A modelagem preditiva é uma etapa avançada do aprendizado de máquina na qual se espera obter informações a partir do algoritmo implementado (SHAILAJA; SEETHARAMULU; JABBAR, 2018).

As predições em saúde baseadas em informações clínicas são feitas a partir da análise de dados históricos e atuais, com a criação de um modelo para ajudar a prever resultados futuros. Na modelagem preditiva, os dados são coletados, um modelo estatístico é formulado, as previsões são feitas e o modelo é validado (ou revisado) à medida que dados adicionais se tornam disponíveis.

Os modelos de previsão podem ser definidos como qualquer combinação de dois ou mais preditores para estimar uma probabilidade individualizada de um desfecho específico em um período definido. Eles são cada vez mais comuns nas pesquisas clínicas e na prática médica, uma vez que podem apoiar decisões clínicas.

Existem diversas formas de categorizar os algoritmos de aprendizado de máquina. Esta subseção irá se limitar aos conhecimentos aplicados para a escolha dos modelos usados nos estudos do Capítulo 5.

O aprendizado de máquina é considerado supervisionado quando um modelo é treinado a partir de dados associadas a um resultado conhecido. O aprendizado não supervisionado não depende um rótulo predefinido e, no caso do aprendizado semi-supervisionado, os dados de treinamento contêm muito poucos exemplos rotulados e muitos exemplos não rotulados. No estudo da base MIMIC-IV, como a informação sobre o desfecho de cada paciente está presente, a opção realizada é pelo aprendizado supervisionado.

De uma forma geral, os modelos podem ser de classificação, quando sua saída esperada é um objeto categórico, ou de regressão, quando é esperado um valor como saída para o

modelo. Conforme será discutido no Capítulo 5, o estudo da característica morte intra-hospitalar foi entendido como um problema de classificação.

Outra forma de caracterizar os modelos é quanto à necessidade de parametrização: algoritmos paramétricos são mais adequados para problemas onde os dados de entrada são bem definidos e previsíveis, enquanto os não paramétricos são usualmente apropriados para casos onde os dados de entrada não são bem definidos, ou são muito complexos para serem modelados usando um algoritmo paramétrico. Conforme será discutido na análise da base MIMIC-IV no Capítulo 5, os algoritmos paramétricos serão mais apropriados para o estudo proposto.

A seguir, serão apresentados brevemente os algoritmos de aprendizado de máquina supervisionado mais utilizados em estudos de saúde, com ênfase para a Regressão Logística (RL), que será mais explorada nos estudos do Capítulo 5. A Tabela 5, disponível no Apêndice A, resume os principais critérios utilizados para seleção dos algoritmos para essa pesquisa.

A RL é um modelo capaz de aprender a relação matemática através uma função logística entre um conjunto de co-variáveis (variáveis preditoras, nesse caso) e uma variável de resultado binária (nessa pesquisa, será a mortalidade intra-hospitalar) (COX, David R, 1958). Depois que esse relacionamento é aprendido, o modelo pode fazer uma previsão para um novo caso, dados os valores do preditor desse novo paciente. A RL é amplamente utilizado na pesquisa em saúde graças à sua fácil interpretabilidade (SCHOBER; VETTER, 2021).

A RL permite estimar a probabilidade associada à ocorrência de um evento em relação a um conjunto de variáveis explanatórias. Nela, utiliza-se o método da máxima verossimilhança para estimar a probabilidade da variável dependente assumir um determinado valor em função dos valores conhecidos de outras variáveis, sendo que os resultados da análise ficam contidos em um intervalo de zero a um (NICK; CAMPBELL, 2007).

O racional matemático é de que a probabilidade de ocorrência de um evento pode ser estimada diretamente: No caso da variável dependente Y assumir apenas dois possíveis estados (1 ou 0) e haver um conjunto de n variáveis independentes $X_1, X_1, X_3, \dots, X_n$, o modelo de regressão logística pode ser definido pela Equação 2.6:

$$P(Y = 1) = \frac{1}{1 + e^{-f(x)}} \quad (2.6)$$

Na Equação 2.6, considera-se $f(x) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \dots + \beta_n X_n$. Nesse caso, os

coeficientes, apresentados como β , são estimados a partir do conjunto dados, pelo método da máxima verossimilhança, no qual se encontra uma combinação de coeficientes que maximiza a probabilidade da amostra ter sido observada. O exponencial de um valor positivo produz um coeficiente maior do que 1, enquanto um coeficiente negativo retorna um $\exp(\beta)$ menor do que 1; já um coeficiente de valor zero produz um $\exp(\beta)$ igual a 1, o que indica que a variável independente não afeta a chance de ocorrência da variável dependente. No caso de estimativas para discriminação entre dois grupos, como foi desenvolvido no Capítulo 5, na Subseção 5.2.5, a regra de classificação é que, se $P(Y = 1) > 0,5$, considera-se $Y=1$; já se $P(Y = 1) < 0,5$, considera-se $Y = 0$ (SANHARAWI; NAUDET, 2013). A função sigmoide da Figura 4 representa essa regra com seu formato em S.

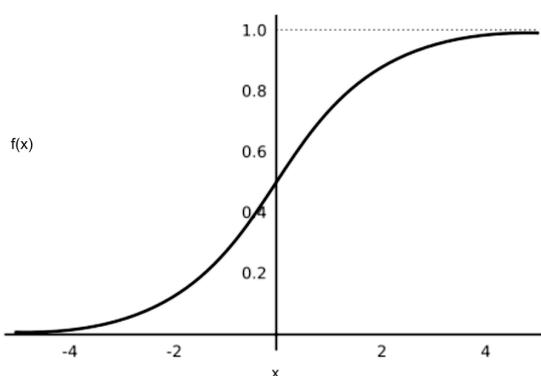


Figura 4: Representação da função sigmoide.

O algoritmo de Máquina de Vetor de Suporte (MVS) é semelhante à RL no sentido de que podem classificar (ou prever) um determinado caso em termos de resultado. Entretanto, como o MVS se baseia na teoria estatística e na minimização do risco estrutural, ele cria um limite de decisão ideal no espaço de dados onde as n dimensões são as covariáveis e todos os pontos de dados disponíveis são plotados. Dessa forma, o MVS define um limite de decisão que coloque tantos casos negativos quanto possível de um lado e os demais casos no outro lado do limite (CORTES; VAPNIK, 1995).

A Árvore de Decisão (AD) tem uma estrutura semelhante a uma árvore com nós de decisão em uma hierarquia. Cada nó de decisão leva a dois ramos que dependem do valor de uma variável específica, e assim ocorrem ramificações até um nó de folha terminal, o

qual está associado a um resultado específico. Com isso, a AD classifica uma instância baseada em variáveis decisórias conforme o caminho da raiz até as folhas. Os algoritmos de AD aprendem automaticamente uma estrutura de árvore de decisão ótima, dado um conjunto de dados. A Floresta Aleatória, do inglês, *Random Forest* (RF) é um algoritmo que se baseia em um conjunto de AD em que cada vetor aleatório é independente e identicamente distribuído. Nesse caso, cada árvore estima uma classificação (voto), e a classificação mais votada representará o resultado. Dessa forma, o erro de indução converge à medida que o número de árvores aumenta (QUINLAN, 1987).

A Rede Bayesiana (RB) utiliza teoria dos grafos, condições de Markov e distribuição de probabilidades para representar uma situação, suas variáveis e seus estados e, a partir dessas informações, realizar inferências. Matematicamente, a RB é uma representação de uma tabela de conjunção de probabilidades que, para representar a força da dependência, utiliza probabilidades associadas a cada grupo de nós pais-filhos na rede. Dessa forma, constrói-se um modelo que representa de forma simples as relações de causalidade das variáveis (PEARL, 1995). O Naive Bayes (NB) é um dos modelos probabilísticos bayesianos mais simples, baseado na suposição de independência condicional simplificadora. Nesse caso, a probabilidade final pode ser calculada a partir da probabilidade de cada classe ou grupo, e aquele com maior probabilidade será escolhido. Por essas características, ele depende que as classes sejam mutuamente exclusivas (LEWIS, 1998).

A Rede de Dependência, do inglês, *Dependency Network* (RD) é um modelo semelhante à cadeia de Markov, no qual cada vértice (nó) corresponde a uma variável aleatória e cada aresta representa as dependências entre as variáveis. Ao contrário da RB, a RD pode conter ciclos. Cada nó está associado a uma tabela de probabilidade condicional, que determina a realização da variável aleatória dado seus pais. A RD aproxima a distribuição conjunta a partir de um grupo de distribuições de probabilidade condicional aprendidas independentemente. Com isso, essa abordagem de aprendizado resulta em ganhos de eficiência significativos em relação aos modelos exatos (NEVILLE; JENSEN, 2007).

A Rede Neural (RN) é um modelo composto por unidades de processamento, conectadas por canais de comunicação que estão associados a determinado peso. As unidades fazem operações apenas sobre seus dados locais recebidos pelas suas conexões. O comportamento inteligente de uma RN decorre das interações entre as unidades de processamento da rede. Se o problema pode ser aprendido, um conjunto estável de pesos evolui até produzir boas respostas para a maioria das decisões ou predições da amostra. As camadas

são compostas de vários nós interconectados, semelhantes ao modelo biológico de uma rede neural ([RUMELHART; HINTON; WILLIAMS, 1986](#)).

O K-Vizinhos mais Próximos, do inglês, *K-Nearest Neighbors* (KNN) é um modelo no qual o aprendizado é baseado no grau de semelhança de um dado em relação a outro. Seu treinamento ocorre formado por vetores de N dimensões, e seu objetivo é classificar um dado novo a partir da comparação com K dados do conjunto de treinamento que estejam mais próximos no dado de entrada. Por esse motivo, para encontrar o K mais próximo vizinhos do novo dado, é necessário usar uma métrica de distância, que pode variar conforme o problema a ser modelado. Da mesma forma, a depender do tipo de dado estudado, a escolha do K pode variar ([COVER; HART, 1967](#)).

Ao treinar um modelo de aprendizado de máquina, diversas medidas podem ser utilizadas para avaliação de desempenho. A Tabela 6, disponível no Apêndice A, consolida as principais formas de cálculo para o caso de modelos de classificação, conforme será apresentado no Capítulo 5, com a ressalva de que a medida utilizada para determinação do melhor modelo depende de cada problema analisado ([VUJOVIC, 2021](#)).

Neste capítulo, foi descrita a metodologia da pesquisa dessa dissertação. Foi dada ênfase no delineamento do estudo, na fonte dos dados, na seleção de uma população dessa base de dados, e nas técnicas de análise aplicadas no Capítulo 5.

No próximo capítulo, será feita uma apresentação dos principais conceitos teóricos da saúde digital que embasaram o desenvolvimento desta pesquisa.

3 Fundamentos de Saúde Digital

Este capítulo apresenta os conceitos básicos, considerados pré-requisitos para se compreender e reproduzir a pesquisa desenvolvida.

Para melhor entendimento dos conceitos a partir dos quais a base MIMIC foi construída, é importante apresentar definições usualmente utilizadas em estudos de saúde. Por esse motivo, seguem-se quatro subseções: A Subseção 3.1 conceitua o PEP, a Subseção 3.2 define e exemplifica a Código Internacional de Doenças (CID), e a Subseção 3.3 contextualiza a CICU como uma especialização da UTI. Por fim, a Subseção 3.4 apresenta os fundamentos da tomada de decisão clínica.

A base MIMIC agrega dados relacionados a assistência à saúde. Por esse motivo, a Subseção 3.4 apresenta o conceito de tomada de decisão clínica e como ele se relaciona com a disponibilidade de informações de um paciente.

3.1 Prontuário Eletrônico do Paciente

A definição atual de PEP é de um sistema digital que pode receber e fornecer informações de cuidadores e pacientes. O PEP consegue processar de dados relevantes para análise e geração de relatórios que sirvam para boa comunicação, intervenções clínicas apropriadas, melhoria da qualidade de atendimento e segurança do paciente.

O PEP é um dos avanços mais importantes da tecnologia da informação no domínio da saúde e, quando implantado eficazmente, pode potencializar o trabalho de cuidadores, gestores e pesquisadores (HOLROYD-LEDUC *et al.*, 2011). Inicialmente desenvolvido e usado em instalações médicas acadêmicas, a maioria dos PEP institucionais foi substituída por PEP de grandes fornecedores (EVANS, R. S., 2016).

Foi demonstrado que a adoção do PEP ajuda a melhorar a confiabilidade e a satisfação na prestação de serviços de saúde, quando implementado adequadamente (HILLESTAD *et al.*, 2005). Por adequação compreendem-se apenas questões relacionadas a usabilidade do

sistema. Problemas consideráveis têm sido observados no uso de PEP, como resultado da falta de interoperabilidade entre os sistemas, e a ausência de padronização de interfaces entre esses sistemas. Esses fatores prejudicam a colaboração e a troca de informações sobre pacientes entre instituições e profissionais (JANETT; YERACARIS, 2020).

Uma etapa fundamental do processo de pesquisa é a decisão sobre quais tipos de dados serão necessários para responder à questão de pesquisa. Nesse sentido, existem desafios metodológicos inerentes à realização de pesquisas com PEP: (1) viés de seleção, (2) dados ausentes e (3) viés de confusão.

O viés de seleção (1) pode ser entendido como a situação em que população do estudo falha em representar a população real. Ele pode ocorrer, por exemplo, se todos os pacientes, incluindo os controles, já estiverem procurando atendimento médico em um sistema baseado em PEP. Um caso em que essa situação pode ocorrer é em estudos em que são comparados diferentes tipos de tratamento em que há opção por uma intervenção (cirúrgica, por exemplo) ou abordagem conservadora, já que o perfil dos pacientes que faz a opção por um ou outro tipo de tratamento pode ser diferente. Foi demonstrado que o uso de diferentes critérios de busca em uma base de dados de PEP pode alterar significativamente a população de pacientes selecionados para um estudo e, subsequente, alterar os desfechos e as conclusões da pesquisa (HRIPCSAK *et al.*, 2011). Embora seja virtualmente impossível controlá-lo, o viés de seleção pode ser significativamente reduzido quando ocorre randomização dos pacientes, além da adoção de estratégias para se reduzir as perdas de seguimento (HERNÁN *et al.*, 2008; HERNAN; HERNANDEZ-DIAZ; ROBINS, 2004).

Os dados ausentes (2) são frequentes no contexto da saúde, seja por causa intencional, em que não há razão de se inserir um registro (por exemplo, um paciente que não possuiu um dispositivo invasivo em veia central não fornece registros de pressão venosa central), ou por causa não intencional (como um profissional de saúde esquecer de anotar um dado em prontuário, por exemplo).

A depender da relação entre dados observados (disponíveis) e não observados (faltantes), três mecanismos de ausência podem ser identificados e serão descritos a seguir: *Missing Completely at Random* (MCAR), *Missing at Random* (MAR) e *Missing Not at Random* (MNAR).

No caso de dados faltantes completamente ao acaso, MCAR, as observações ausentes dependem das aferições observadas e não observadas; um exemplo prático seria o de um profissional de triagem que esquece ocasionalmente de aferir a temperatura de alguns

pacientes que chegam ao pronto-atendimento. Não há um mecanismo oculto relacionado a nenhuma variável e isso não depende de nenhuma característica dos pacientes.

Já no caso de dados faltando ao acaso, MAR, isso acontece em situações nas quais a probabilidade de um valor estar ausente está relacionada apenas aos dados observáveis. Um exemplo dessa situação ocorre quando pacientes não informam que usam anticoncepcionais quando são questionadas sobre uso de medicações contínuas, o que faz com que a taxa de resposta desse questionamento tenha uma dependência da variável gênero. Nesse caso, os dados observados estão estatisticamente relacionados com as variáveis ausentes.

Por fim, dados faltando não ao acaso, MNAR, são aqueles que não se encaixam nas categorias descritas anteriormente. Nessa categoria, os dados ausentes dependem dos valores ausentes e observados, por exemplo, pacientes sem diagnóstico de hipertensão estão menos propensos a terem sua pressão aferida com frequência. Nesse caso, os dados ausentes para a variável pressão arterial dependem parcialmente dos valores de pressão arterial do paciente.

A maioria dos softwares analíticos convencionais presumem a integridade da matriz de dados, o que gera a necessidade de se aplicarem estratégias de tratamento desses dados para poderem ser analisados. Foi demonstrado que pacientes com tempo prolongado e com maior complexidade no atendimento geralmente são aqueles com maior proporção de dados ausentes. Isso se explica pelo maior número de intervenções a que são submetidos, além da maior quantidade de profissionais envolvidos no cuidado desses pacientes (RUSANOV *et al.*, 2014). Nesses casos, pode ser útil aplicar métodos como exclusão de pares, substituição por média ou por regressão, ou uso de técnicas de modelagem para máxima verossimilhança e imputação múltipla (RASSENFOSSE; SELIGER, 2021). Essa situação será melhor exemplificada no Capítulo 5 desta dissertação, em que serão descritas as etapas aplicadas para tratamento dos dados ausentes no conjunto de dados estudado.

O viés de confusão (3) representa a falha em considerar apropriadamente uma variável adicional que influencia a variável dependente, mas também as variáveis independentes.

Devido à sua natureza observacional da maioria desses estudos com PEP, a causalidade não pode ser estabelecida. Por esse motivo, as extrapolações dos resultados a fim de criar generalizações e influenciar os padrões de prática devem ser feitos com muito cuidado.

Os bancos de dados de informações hospitalares foram inicialmente criados para fins administrativos. Por este motivo, muitas vezes não possuem a granularidade do nível

do paciente de dados clinicamente úteis, precisos e completos para abordar perguntas de pesquisa. Com o tempo, os bancos de dados tornaram-se mais abrangentes, dinâmicos e interconectados. Atualmente, a análise de dados de PEP pode informar as escolhas de sistemas de saúde em ampla escala, como em casos de farmacovigilância, ou decisões clínicas à beira do leito, como na escolha de um determinado antimicrobiano para o perfil do paciente e da instituição (HARRISON, 2012).

Grande parte das pesquisas que utilizaram grandes bancos de dados nos últimos anos se baseiam em registros de alta hospitalar e informações administrativas. Embora essas bases contenham muitos pacientes, frequentemente existe uma limitação dos atributos, além da falta de outros dados clínicos significativos que seriam necessários para responder a questões de pesquisa mais complexas (JHA, 2012).

Existem conjuntos de dados de saúde de acesso aberto que possibilitam a realização de pesquisas observacionais retrospectivas. Entretanto, o número é ainda pequeno, se comparado ao total de centros de saúde que já possuem seus dados em formato digital. Ainda que iniciativas de democratização dos dados sejam importantes e relevantes, é sempre necessário esforço considerável no mascaramento dos dados, para assegurar privacidade aos paciente (SAFRAN *et al.*, 2007). São alguns exemplos dessas bases: (1) MIMIC¹, que inclui dados de cerca de 70000 pacientes desde 2008 admitidos em UTI (JOHNSON, A. *et al.*, 2020); (2) o APACHE Outcomes², que inclui dados de aproximadamente 1 milhão de registros de 105 UTIs, correspondendo a 150.000 internações, catalogadas desde 2010 (ZIMMERMAN; KRAMER; MCNAIR *et al.*, 2006); e o (3) Phillips eICU³, que contém dados de mais de 2 milhões de internações (POLLARD *et al.*, 2018). As principais iniciativas de democratização de dados de saúde que favorecem pesquisas clínicas e cabem ser citadas são: a PCORnet (Rede Nacional de Pesquisa Clínica Centrada no Paciente), o Open NHS (Sistema Nacional de Saúde da Inglaterra), o VistA (Sistema de Arquitetura de Tecnologia de Informação em Saúde dos Veteranos) e o NSQUIP (Projeto Nacional de Melhoria da Qualidade Cirúrgica) (CELI; MARK *et al.*, 2013).

A tendência presente e futura para as pesquisas clínicas é a utilização cada vez mais frequente de *big data* para melhorar a prestação de cuidados aos pacientes (MEHTA; PANDIT, 2018). Os dados oriundos de PEP pode apoiar estudos observacionais, servir para triar pacientes em larga escala, e também para estudar e melhorar os resultados de

¹Disponível em: <https://archive.physionet.org/physiobank/database/mimicdb/>. Acesso em: 31 de dezembro de 2022.

²Disponível em: <https://apacheoutcomes.cernerworks.com/criticaloutcomes-home/>. Acesso em: 31 de dezembro de 2022.

³Disponível em: <https://eicu-crd.mit.edu/>. Acesso em: 31 de dezembro de 2022.

desfecho. Nesse contexto, a base MIMIC será apresentada em detalhes no Capítulo 2, pois foi usada nos estudos dessa pesquisa.

3.2 Classificação Internacional de Doenças

A Código Internacional de Doenças (CID) é um registro que agrega e organiza as mais diversas doenças e sintomas conhecidos em grupos ou categorias. Sua maior importância está em oferecer uma linguagem comum que permite aos profissionais de saúde compartilharem informações padronizadas em todo o mundo. A CID serve como uma base padronizada para identificar tendências e estatísticas de saúde ao nível global, com cerca de 55 mil códigos únicos para lesões, doenças e causas de morte (QUAN *et al.*, 2005).

Ela é mantida atualizada pela Organização Mundial de Saúde, e atualmente está na décima primeira versão, de 2019. As mudanças no conteúdo refletem conhecimentos e perspectivas sobre as doenças e suas causas que surgiram desde a primeira publicação (HARRISON *et al.*, 2021).

A base MIMIC utiliza a CID-9 (nona versão) e segue os padrões de nomenclatura norte-americanos. Os códigos da CID podem ser utilizados para agrupar pacientes conforme um ou mais diagnósticos, como foi detalhado na Subseção 2.2.1. O conhecimento da importância e da forma de categorização das doenças na CID foram necessários para selecionar pacientes elegíveis e construir a coorte de estudo, conforme descrito na Subseção 5.2.

3.3 Unidade de Terapia Intensiva Cardíaca

A UTI é um setor hospitalar dedicado ao suporte avançado de pacientes que necessitam de maiores cuidados e monitoramento 24 horas. A internação em CICU compreende pacientes que foram admitidos e tiveram alta da unidade de terapia intensiva cardiológica, independentemente do diagnóstico de admissão (LÜSEBRINK *et al.*, 2021).

O conceito de tempo de internação é caracterizado pela diferença de tempo, em dias, entre admissão e alta hospitalar. Já a mortalidade intra-hospitalar é definida pelo registro de óbito no prontuário em uma data contida no período de internação. Ambas as definições serão utilizadas para caracterização da coorte de pacientes descrita na Subseção 2.2.1.

3.4 Tomada de Decisão Clínica

A tomada de decisão clínica que melhor atende aos interesses dos pacientes se fundamenta na medicina baseada em evidências, no bom julgamento clínico e nas diretrizes médicas atualizadas. Tradicionalmente, a tomada de decisão clínica era compreendida como um processo racional e cognitivo. Recentemente, a inteligência emocional tem sido relacionada como ferramenta importante de apoio à tomada de decisão clínica (KOZLOWSKI *et al.*, 2017). A prática clínica requer um equilíbrio entre o que é ideal cientificamente e o que pode ser efetuado na prática médica (WHELEHAN; CONLON; RIDGWAY, 2020).

No entanto, uma porcentagem relativamente baixa de diretrizes clínicas e das decisões clínicas é baseada em Ensaio Clínico Randomizado (ECR) prospectivos bem desenhados (SARA *et al.*, 2021). Nesse sentido, bancos de dados de PEP podem ser uma opção para se realizar investigações em larga escala, mas com proporções de custo e dificuldades técnicas bastante inferiores ao da realização de ECRs. O uso de dados de PEP para realizar estudos observacionais retrospectivos pode, em alguns casos, ser suficiente para responder algumas perguntas de pesquisa sem necessidade de se realizar um ECR (MEHTA; KELLUM *et al.*, 2007; BROWER *et al.*, 2000). Existe uma oportunidade para as ferramentas de apoio à tomada de decisão clínica que se apoiam em aprendizado de máquina para prover informações mais confiáveis e ajustadas ao cenário onde o profissional se encontra (EMMANUEL, 2021). Além disso, pode-se fazer a avaliação de diagnósticos ou tratamentos para sub-populações selecionadas, bem como a da detecção de eventos adversos de medicamentos ou outras intervenções, com maior rapidez, através da análise dos dados de PEP (GHASSEMI; CELI; STONE, 2015).

O uso do PEP pode facilitar o desenvolvimento de sistemas de aprendizagem contínua, em que dados específicos do paciente conseguem ser registradas em um banco de dados populacional. Esses sistemas podem fornecer suporte de decisão em tempo real para pacientes individuais com base em dados de pacientes semelhantes em cenários semelhantes (CELI; ZIMOLZAK; STONE, 2014) e fundamentaram o desenvolvimento do sistema descrito no Capítulo 6.

Neste capítulo, foram apresentados os principais conceitos teóricos da saúde digital que embasaram o desenvolvimento desta pesquisa. É importante o conhecimento dessas definições para que as decisões descritas no Capítulo 5 possuam coerência e possam ser avaliados com rigor científico. Nesta pesquisa, além da análise de dados, será apresentado um modelo preditivo de classificação, na qual a variável de interesse é o óbito intra-

hospitalar.

No próximo capítulo será feita uma apresentação das principais publicações que serviram de orientação para esta dissertação.

4 Revisão da Literatura

Este capítulo consiste em uma breve exposição sobre trabalhos relacionados a esta pesquisa. A primeira Subseção 4.1 apresenta o uso da metodologia proposta por Petersen *et al.* (2008) para realização da Revisão Sistemática da Literatura (RSL). Após a apresentação do levantamento, da categorização e da seleção das publicações, as subseções seguintes descrevem os trabalhos que se relacionam com o contexto em que este estudo foi desenvolvido. Nesse grupo, incluem-se artigos atualizados sobre predição em saúde, a comparação com o uso de escores clínicos, além de outros estudos produzidos a partir da base de dados MIMIC. A Subseção 4.2 apresenta pesquisas relacionadas a estudos de predição, a Subseção 4.3 descreve o uso de escores na prática clínica, e a Subseção 4.4 resume as principais pesquisas realizadas com a base MIMIC.

4.1 Metodologia da Revisão Bibliográfica

A metodologia proposta por Petersen *et al.* (2008) foi utilizada para realização da Revisão Sistemática da Literatura (RSL), por ser particularmente adequada ao mapeamento sistemático de estudos em Engenharia de Software.

Para a seleção dos artigos relevantes para a pesquisa, a plataforma Parsifal¹ foi utilizada. Ela consiste em uma ferramenta projetada para a realização de revisões sistemáticas da literatura nessa área de conhecimento (KITCHENHAM; CHARTERS, 2007).

O mecanismo disponível para a avaliação da qualidade dos artigos selecionados pela Parsifal é constituído por três partes: (1) um conjunto de perguntas, (2) respostas pré-definidas e (3) notas de corte. Ao atribuir um peso para cada uma das respostas pré-definidas para as perguntas, é gerada uma pontuação final para cada um dos estudos selecionados. Dessa forma, ao aplicar a nota de corte, cada artigo avaliado acima da nota é incluído na RSL, enquanto aqueles que estiverem abaixo serão desconsiderados para as

¹Disponível em: <https://parsif.al>. Acesso em: 31 de dezembro de 2022.

etapas seguintes.

A coleta e categorização dos artigos seguiram, portanto, os seguintes passos: (1) uso de um termo de busca para filtragem das bases de dados selecionadas; (2) leitura de títulos e resumos dos trabalhos resultantes da etapa anterior, para classificá-los conforme critérios de inclusão ou exclusão; e (3) extração de informações dos artigos, pertinentes às perguntas propostas. A avaliação dos artigos foi feita pelo autor desta dissertação, com apoio do pesquisador orientador. O objetivo principal desta pesquisa foi avaliar se o aprendizado de máquina poderia auxiliar o julgamento clínico na identificação de um conjunto mínimo de dados clínicos capaz de identificar precocemente pacientes com Insuficiência Cardíaca e com alto risco para óbito intra-hospitalar.

Dessa forma, a RSL identificou estudos que utilizaram a base de pacientes MIMIC para estudar a internação e os desfechos de um determinado grupo de pacientes. Foram criadas perguntas de pesquisa no protocolo de avaliação dos estudos obtidos pela RSL quanto à pertinência ao objetivo desta dissertação:

- Quais estudos de análise de sobrevivência e estratificação de risco já foram realizados no banco de dados MIMIC-IV?
- Qual o perfil clínico dos pacientes de CICU no banco de dados MIMIC?
- Quais técnicas de aprendizado de máquina já foram avaliadas no banco de dados MIMIC?

A primeira etapa da RSL consistiu na determinação dos conceitos básicos que seriam explorados pela pesquisa. Com base no objetivo estabelecido, foram determinados os termos utilizados para a seleção para os artigos. A *string* de busca apresentada a seguir foi composta pela combinação de palavras-chave relativas a aprendizado de máquina, análise de sobrevivência e base de dados MIMIC. Essa *string* segue a estratégia PICO (COSTA SANTOS; MATTOS PIMENTA; NOBRE, 2007), que contempla os quatro componentes fundamentais da questão de pesquisa e da construção da pergunta para a busca bibliográfica de evidências (AKOBENG, 2005).

```
((("MIMIC"OR "MIMIC-II"OR "MIMIC-III"OR "MIMIC-IV") AND ("Survival  
Analysis"OR "Risk Stratification"OR "Machine Learning"OR "Artificial Intelligence"))  
AND ("Heart"OR "Cardiac"OR "CCU"OR "Cardiovascular"))
```

As bases de dados selecionadas para pesquisa foram escolhidas com base em sua relevância para publicações relacionadas às ciências médicas, em que se incluem Scopus, PubMed e Embase; e Ciências da computação, com inclusão da IEEE Digital Library (DYBA; DINGSOYR; HANSSSEN, 2007). O escopo desta pesquisa foi estabelecido a partir de critérios de inclusão e exclusão, os quais foram aplicados na seleção e classificação dos artigos. Os seguintes critérios de inclusão foram aplicados:

- Idioma inglês, espanhol ou português, para facilitar a análise mais detalhada dos artigos selecionados;
- Ano da publicação a partir de 2000, que corresponde ao ano da disponibilização da primeira versão do MIMIC;
- Delineamento experimental (ensaios clínicos, randomizados ou não) ou observacional (estudos de caso-controle ou de coorte), unicêntricos ou multicêntricos, por terem maior associação com os objetivos desta pesquisa;
- Uso de técnicas de aprendizado de máquina, análise de sobrevivência ou estratificação de risco, por se relacionarem com o escopo deste trabalho.

Em relação aos critérios de exclusão, foram definidos:

- Indisponibilidade da versão completa do artigo, por impossibilitar uma análise mais detalhada;
- Literatura cinzenta, pelo potencial risco de não confiabilidade;
- Artigos que não atendem a todos os critérios de inclusão.

A exploração bibliográfica avançada nos motores de busca exigiu a criação de diferentes estruturas de busca, as quais foram variações da *string* apresentada anteriormente. Uma vez testada a *string* para avaliar se houve retorno de resultados, foi feita a busca inicial nas bases de dados. Os resultados obtidos por cada base de dados foram: Scopus, com 242 selecionados e 6 aceitos; Pubmed, com 98 e 16, respectivamente; IEEE, com 57 e 7, respectivamente; e Embase, com 53 e 8, respectivamente. De todos os artigos selecionados, o ano com maior relevância de publicações foi 2021, que correspondeu a 75 artigos selecionados, seguido por 2022, com 40 artigos.

Após a identificação de 450 artigos por busca nas quatro bases de dados, foi possível realizar o primeiro refinamento, com objetivo de eliminar 101 artigos duplicados e outros dois que não possuíam informações completas quanto a título, autores, periódicos

e palavras-chave. Posteriormente, foi realizada uma segunda filtragem para avaliar sua elegibilidade a partir da leitura do título do artigo e de um sistema de pontuação relacionado às perguntas de pesquisa. Essa etapa resultou em 37 estudos selecionados, com as seguintes pontuações de um valor mínimo de zero e máximo de cinco pontos: 1 artigo recebeu nota 2; 9 artigos receberam nota 3; 15 artigos receberam nota 4; e 12 artigos receberam nota 5. Esses estudos foram incluídos para uma terceira fase de seleção, realizada através da leitura dos resumos e das palavras-chave dos artigos, para formar o grupo final de 19 artigos estudados na íntegra, e que coincidiu com os artigos que receberam notas 4 e 5. O fluxograma das etapas iterativas de seleção dos estudos foi elaborado a partir da metodologia Prisma e está apresentado na Figura 5 (PAGE *et al.*, 2021), e a Tabela 1 apresenta os artigos selecionados para leitura na íntegra.

Tabela 1: Artigos selecionados para leitura na íntegra ao final da RSL. RL: Regressão Logística; AD: Árvore de Decisão; KNN: K-Vizinhos mais Próximos; SVM: Máquina de Vetor de Suporte; RN: Rede Neural; RF: Floresta Aleatória; RB: Rede Bayesiana; XGB: Gradiente Extremo; N/A: Não se aplica.

Artigo	Número de pacientes	Modelos avaliados
Boroujeni, Yousefi e Zuretti (2019)	257	Agrupamento Hierárquico
Churpek <i>et al.</i> (2016)	269,999	RL, AD, KNN, SVM, RN
Donnelly <i>et al.</i> (2021)	2179	Regressão de Cox
Somani <i>et al.</i> (2020)	2864	N/A
Madden <i>et al.</i> (2020)	NA	N/A
Ye <i>et al.</i> (2021)	812	RL
Hansen, Mejldal e Nielsen (2020)	2130	Regressão de Cox
Atalla <i>et al.</i> (2020)	339	N/A
Zimmerman, Kramer, McNair <i>et al.</i> (2006)	131,618	N/A
Johnson, Kramer e Clifford (2013)	81,087	RL
Balkan, Essay e Subbian (2018)	135,68	RF
Celi, Galvin <i>et al.</i> (2012)	1400	RL, RB, RN
Lee, Maslove e Dubin (2015)	17,49	N/A
Lee e Maslove (2015)	17,152	RL, AD
Meyer <i>et al.</i> (1992)	651	N/A
Schulz <i>et al.</i> (2021)	245,711	N/A
Alistair E.W. Johnson <i>et al.</i> (2016)	53,423	N/A
Johnson, Pollard e Mark (2017)	N/A	RL, AD, XGB
Zimmerman, Kramer e Knaus (2013)	131,618	N/A

Esta subseção descreveu a metodologia de pesquisa utilizada para realização da Revisão Sistemática da Literatura (RSL). As etapas de seleção, refinamento e estudo da bibliografia foram apresentadas, e antecedem a discussão mais detalhada dos artigos mais relevantes selecionados. Os passos desta RSL podem ser reproduzidos com ou sem o auxílio da ferramenta Parsifal.

As próximas subseções irão apresentar os artigos selecionados ao final da RSL. Devido à similaridade dos temas, foram criados três grupos, apresentados em subseções distintas a seguir. A Subseção 4.2 apresenta as principais pesquisas que tiveram objetivo de predição de desfechos no contexto de pacientes internados. A Subseção 4.3 lista ferramentas clínicas usadas na prática e apresenta suas vantagens e limitações, para contrastá-las com as

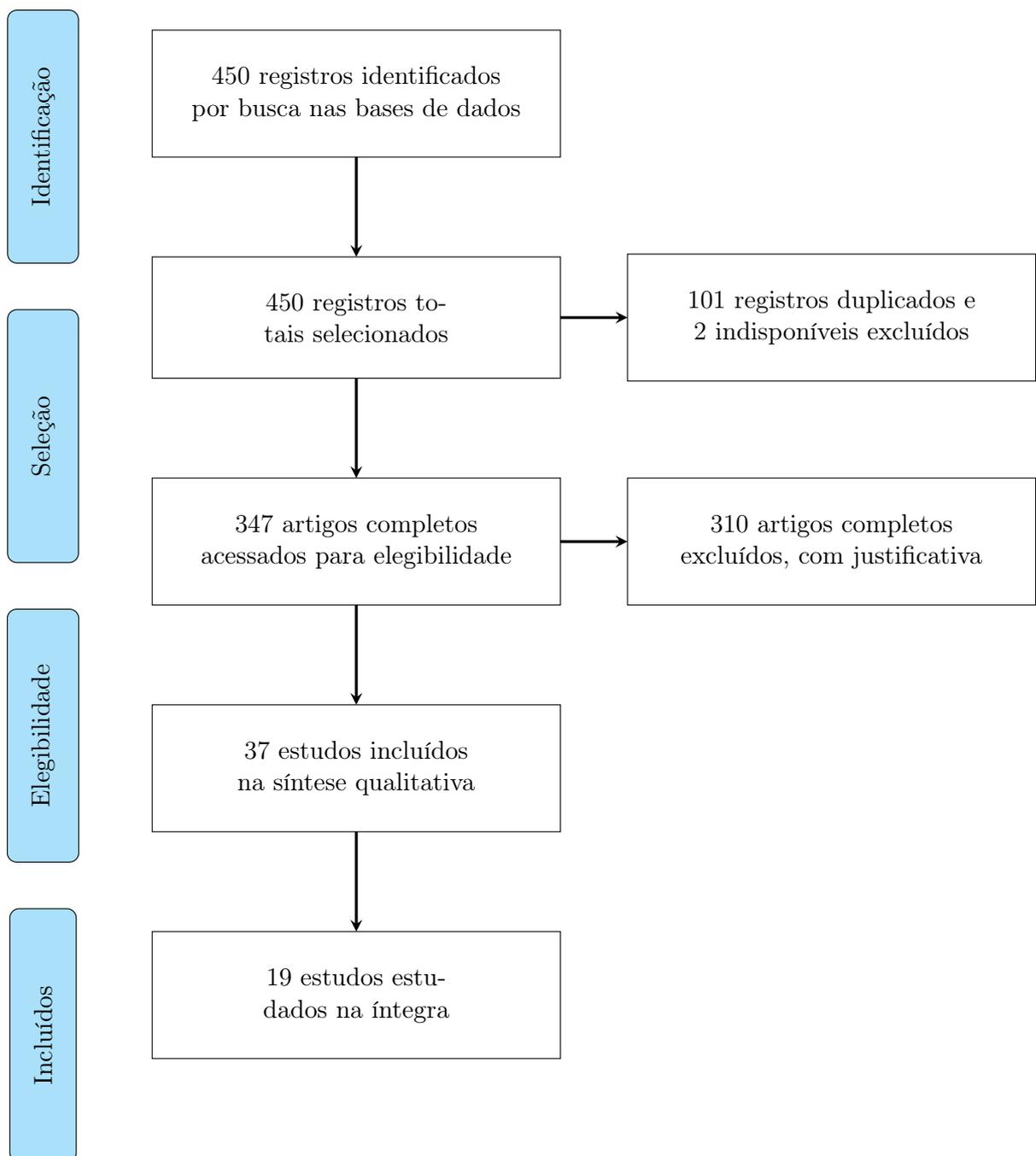


Figura 5: Etapas da revisão sistemática da literatura.

possibilidades que podem ser obtidas a partir de modelos de aprendizado de máquina. A Subseção 4.4 elenca as principais pesquisas já publicadas que tiveram como fonte de dados a base MIMIC, para apresentar as possibilidades de exploração consideradas nesta pesquisa.

Nos últimos anos, diversos estudos publicados aplicaram o conhecimento do aprendizado de máquina para solucionar problemas de saúde. Entretanto, pouco avanço foi feito em modelos que possam transferir aprendizados para o contexto brasileiro, em especial ao

SUS. A carência de publicações encontradas nas etapas iniciais da RSL reforça essa afirmação. A próxima subseção apresenta os principais estudos de predição de desfechos em saúde que foram utilizados para fundamentar essa dissertação. O estudo dessa categoria de artigos foi importante para fornecer uma visão geral sobre possibilidades e limitações em relação ao uso de dados para fins preditivos nesse contexto.

4.2 Estudos de Predição de Desfechos em Saúde

A implementação de modelos preditivos para readmissão e mortalidade é uma tarefa relativamente frequente em pesquisas de aprendizado de máquina. Existe um grande desafio em relação a obter a melhor acurácia do modelo e, paralelamente, em relação ao tempo necessário para treinar e validar os modelos. O cenário ideal seria algum modelo que fornecesse a estimativa precisa já na admissão do paciente, para organizar melhor os recursos de saúde em relação a cada atendimento. Entretanto, muitos preditores interessantes, como procedimentos realizados e registros de séries temporais de exames laboratoriais não estão disponíveis no primeiro dia de internação (BOROUJENI; YOUSEFI; ZURETTI, 2019; CHURPEK *et al.*, 2016). Um exemplo para essa situação seria o acompanhamento seriado de marcadores de necrose miocárdica, como a troponina. Um valor, isoladamente, tem menos relevância clínica do que a sua observação por intervalos de horas a dias. Outro exemplo, nesse mesmo contexto clínico, seria a disponibilidade de um exame de ecocardiograma apenas em dias úteis, o que faria com que um paciente internado durante o final de semana não tivesse essa informação disponível por até 48 horas de sua admissão. Esses exemplos práticos representam um desafio da pesquisa: escolher entre possuir, teoricamente, um melhor desempenho de previsão (considerando-se um modelo que poderia ser treinado em dados de outros dias subsequentes no período de hospitalização, o que representaria mais informações de entrada) em favor da disponibilidade rápida como, por exemplo, já prever um prognóstico para o paciente após apenas 12 ou 24 horas de internação. Essa questão pode se relacionar com melhores medidas de desempenho dos modelos, e também com intervalos de confiança com maior ou menor margem.

Donnelly *et al.* (2021) mostraram um estudo observacional, com uma coorte de 103 pacientes acompanhados por 6 meses no *Elias Emergency University Hospital*, que identificou seis variáveis independentes, e usou três delas (todas ecocardiográficas) para criar um escore de risco para readmissão. Foi avaliada a relação entre características de base com o desfecho primário de readmissão por Insuficiência Cardíaca e o desfecho secundário de mortalidade por todas as causas. Como resultado, foi criado o escore DEI para

estratificação de risco de pacientes com ICFEN.

[Somani *et al.* \(2020\)](#) avaliaram as características clínicas de pacientes com COVID-19 e considera a hipótese de que elas poderiam prever re-hospitalização em até 14 dias. Trata-se de um estudo observacional, com uma coorte de 2864 pacientes de cinco hospitais de Nova Iorque, Estados Unidos, acompanhados por dois meses. Os autores identificaram a correlação com HAS, DPOC, menor tempo de hospitalização e não uso de anticoagulação em dose plena com readmissão. Na população estudada, foi demonstrada que readmissões por COVID-19 não são frequentes, e que estão geralmente associadas a comorbidades já existentes.

[Madden *et al.* \(2020\)](#) estudaram se características clínicas e do contexto familiar e social de crianças e adolescentes se relacionam com reinternação por diagnóstico de distúrbio mental primário. Foi realizada uma RSL de oito bases de dados entre 1994 a 2018, com dois autores que revisaram independentemente 7903 artigos (sendo 30 desses selecionados para avaliação). A readmissão foi mais frequente em casos de sintomas mais graves, determinados diagnósticos (psicose, distúrbio afetivo e comportamento suicida), maior tempo de hospitalização e dinâmica familiar prejudicada. Esse estudo possibilitou estimular ensaios clínicos para melhorar o cuidado de saúde para esse público.

[Ye *et al.* \(2021\)](#) analisaram a hipótese de que seria possível dar alta mais precocemente para pacientes internados por COVID-19, desde que feito um monitoramento remoto, para evitar readmissão em até 14 dias, a partir do uso de técnicas de aprendizado de máquina. Os autores realizaram um estudo de intervenção em 409 pacientes (de um total de 812 altas hospitalares) que tiveram alta domiciliar, e acompanhados pelo programa de Cuidados Pós-Alta de um hospital de Nova Iorque, Estados Unidos. O Programa foi associado a redução da readmissão hospitalar e alta satisfação dos pacientes, e conseguiu estimular outras iniciativas semelhantes.

[Hansen, Mejldal e Nielsen \(2020\)](#) estudaram a relação entre abuso de consumo de álcool e readmissão hospitalar através de técnicas de aprendizado de máquina. Foi avaliada a hipótese de que seria possível prever readmissão de pacientes com Transtornos mentais e comportamentais devidos ao uso de álcool baseado em fatores clínicos e de seu contexto de vida. Realizou-se um estudo observacional de 2130 pacientes, acompanhados por dez anos, em hospital especializado em Odense, Dinamarca. Abandono precoce do tratamento, histórico de doença psiquiátrica, menor idade e maior tempo de tratamento foram associados a maior taxa de readmissão. O artigo sugere haver estudos semelhantes nestas condições e, também, conseguiu identificar variáveis que devem ser observadas com

mais atenção nessa população.

[Donnelly *et al.* \(2021\)](#) consideraram a hipótese de que características clínicas de pacientes com COVID-19 poderiam prever re-hospitalização ou morte em 60 dias. Foi realizado um estudo observacional, como uma coorte retrospectiva de 2179 pacientes de 132 hospitais de Veteranos dos Estados Unidos. Os autores demonstraram que as taxas de readmissão e de óbito em 60 dias foram menores do que comparados com pneumonia e Insuficiência Cardíaca; porém, nos 10 primeiros dias pós-alta, esses eventos foram mais frequentes nos pacientes com COVID-19, quando comparados aos outros diagnósticos. Os autores destacaram que o período de maior preocupação com complicações foi nos primeiros dias pós-alta em pacientes com diagnóstico de COVID-19.

[Atalla *et al.* \(2020\)](#) levantaram a discussão sobre características clínicas de pacientes com COVID-19 e se elas poderiam orientar o tempo ideal para alta hospitalar para evitar re-hospitalizações. Foi realizado um estudo observacional, como coorte retrospectiva, de 339 pacientes de um hospital de Rhode Island, Estados Unidos. Re-hospitalização foi associada com HAS, DM, DPOC, doença hepática, neoplasias e abuso de substâncias como drogas e álcool. Readmissões mais precoces (em menos de 12 dias) foram associadas com complicações da COVID-19, enquanto re-hospitalizações tardias foram geralmente associadas a outros fatores. Esse estudo recomendou uma estratificação de risco na alta dos pacientes com COVID-19 para tentar reduzir essas complicações precoces.

A análise desses estudos relacionados a predição de desfechos demonstrou que (1) grande volume de variáveis potenciais e (2) tamanho de amostra suficiente são duas condições importantes para o desenvolvimento de estudos de predição de desfechos. Esses conceitos serão importantes para a elaboração da pesquisa descrita no Capítulo 5, tendo em vista a ampla disponibilidade de variáveis na base MIMIC e a conseqüente necessidade de seleção daquelas que fossem mais relevantes para o objetivo desta pesquisa. Além disso, o estudo detalhado destes artigos também auxiliou na seleção de métricas utilizadas para avaliar o desempenho do modelo descrito no Capítulo 6.

Estudos de modelagem preditiva fundamentaram diversos escores de predição usados na prática clínica. Por esse motivo, uma explicação mais detalhada sobre o uso clínico dos escores está apresentado na Subseção 4.3.

4.3 Escores de Risco

Uma série de pontuações de gravidade da doença foram introduzidas na prática médica na UTI para prever resultados, incluindo o desfecho de morte, e são chamados de escores. Alguns dos mais relevantes serão apresentados a seguir: APACHE IV (KNAUS *et al.*, 1991), SOFA (VINCENT *et al.*, 1996), LODS (GALL, J., 1996), SAPS (GALL, J. R. L., 1993) e OASIS (JOHNSON; KRAMER; CLIFFORD, 2013). O APACHE (*Acute Physiology, Age, Chronic Health Evaluation*) é um sistema de pontuação que estima a gravidade de um paciente baseado em parâmetros clínicos. APACHE IV é a versão mais recente, publicada em 2006, e baseada em dados das primeiras 24 horas de internação de pacientes de UTI e um método de regressão logística multivariada para estimar a probabilidade de óbito hospitalar. É amplamente utilizado em UTIs e pode ser considerado um indicador de referência para as UTIs dos Estados Unidos, tendo em vista que foi desenvolvido a partir de dados de 45 hospitais deste país (ZIMMERMAN; KRAMER; MCNAIR *et al.*, 2006). O SOFA (*Sequential Organ Failure Assessment*) é um escore de gravidade de doença que fornece uma ferramenta de estimativa de mortalidade desenvolvida a partir de uma grande amostra de pacientes de UTIs de 40 países (VINCENT *et al.*, 1996). Foi publicado em 1996 como uma ferramenta de pesquisa para que grupos de pacientes pudessem ser classificados conforme o risco de óbito. Atualmente, é utilizada amplamente nas UTIs como um indicador calculado diariamente para acompanhar a evolução de pacientes críticos (FERREIRA, 2001). A avaliação sequencial da disfunção orgânica durante os primeiros dias de internação na UTI é um bom indicador de prognóstico (SHARMA *et al.*, 2013). O LODS (*Logistic Organ Dysfunction Score*) é um escore publicado em 1996 e cujo objetivo é avaliar seis sistemas fisiológicos (neurológico, cardiovascular, renal, pulmonar, hematológico e hepático) em termos da extensão da disfunção orgânica, com base em registros das primeiras 24 horas de internação em UTI (PELLATHY; PINSKY; HRAVNAK, 2021). Ele foi desenvolvido a partir de dados de UTIs de 12 países da Europa e América do Norte. A pontuação LODS é convertida em probabilidade de óbito, e considera tanto a gravidade relativa entre os sistemas orgânicos quanto o grau de gravidade dentro de um sistema orgânico (VINCENT *et al.*, 1996). O SAPS (*Simplified Acute Physiology Score*) é um escore desenvolvido a partir de dados de 137 UTIs clínicas ou cirúrgicas de 12 países. O SAPS II, versão em uso em muitas UTIs, foi publicado em 1993, e é usado até hoje porque fornece uma estimativa do risco de morte baseado em regressão logística sem a necessidade de se especificar um diagnóstico primário. Já o SAPS III, publicado em 2005, requer a informação de diagnóstico, e possibilitou calcular estimativas de óbito na UTI e óbito durante a internação hospitalar (GODINJAK, 2016). O

OASIS (*Oxford Acute Severity of Illness Score*) é um novo escore de gravidade publicado em 2013. Embora menos difundido na prática clínica, tem relevância para esta pesquisa porque foi desenvolvido a partir de dados da base MIMIC-IV (JOHNSON; KRAMER; CLIFFORD, 2013).

O cálculo dos escores de gravidade tradicionais, como previamente apresentados, pode ser trabalhoso em termos de quantidade de variáveis necessárias para o cálculo de probabilidade possa ser feito. Nesse sentido, o OASIS se propõe a exigir menos informações, e dados de fácil disponibilidade (JOHNSON; KRAMER; CLIFFORD, 2013). Apesar de o uso do escore OASIS na admissão ser suficiente para predição de desfechos clínicos a curto prazo de pacientes com sepse, ele se mostrou menos acurado do que o SAPS II para predição de mortalidade em geral na UTI. Diversos estudos avaliaram os impactos do uso de escores de risco para apoio a tomada de decisão clínica. Balkan, Essay e Subbian (2018) discutiram sobre a dificuldade de predição de risco com as ferramentas disponíveis (escores), por serem frequentemente imprecisas. A hipótese testada é de que o uso de inteligência artificial poderia compensar as limitações gerais subjacentes dos instrumentos de pontuação e, com isso, demonstrar a utilidade do estudo de bancos de dados médicos prontamente disponíveis (prontuários eletrônicos). Os autores utilizaram um banco de dados (*teleICU*) que permitiu a avaliação em grande escala das previsões dos escores APACHE IV e IVa, e compararam os valores previstos com os resultados reais registrados nos prontuários dos pacientes. Além disso, realizaram uma exploração preliminar de modelos preditivos para mortalidade do paciente e tempo de permanência no hospital e na UTI. Os autores observaram uma melhora no desempenho nos modelos recentemente desenvolvidos e treinados nas variáveis de entrada do APACHE. Por fim, os autores destacaram a utilidade do uso de bancos de dados de *teleICU* para o desenvolvimento e avaliação de modelos preditivos.

Os pacientes internados na CICU potencialmente possuem doenças ou lesões graves e com alto risco de óbito. Esse risco pode ser estimado com base na gravidade da doença a partir de critérios fisiológicos, clínicos e demográficos. As estimativas de risco de mortalidade são, no entanto, efetuadas com dados agregados de grandes grupos de pacientes, os quais, muitas vezes, não são homogêneos. Por este motivo, sua validade no contexto de qualquer atendimento de um determinado paciente não pode ser garantida. Em geral, as predições empregam modelos estáticos de tamanho único que funcionam principalmente para o paciente com características mais comuns (CELI; GALVIN *et al.*, 2012). Essa situação pode ser contornada através de uma estimativa mais personalizada do risco de mortalidade, como apresentado por Lee e Maslove (2015) e Lee, Maslove e Dubin (2015).

Esses textos abordaram a estimativa personalizada do risco de mortalidade, que pode mitigar essa questão. A hipótese de que a previsão de mortalidade usando escores baseada na severidade de doença tende a ter um desempenho inferior quando aplicada a casos que diferem da mistura de casos das coortes de derivação originais é explorada. É proposta uma adaptação desses escores para cada grupo populacional, com base nos resultados obtidos em melhor área sob curva ROC e curva precisão-revocação para o escore SAPS.

Lee, Maslove e Dubin (2015) reforçaram a importância de personalizar dinamicamente o suporte à decisão identificando e analisando pacientes semelhantes, de maneira análoga à recomendação personalizada de produtos no comércio eletrônico.

Meyer *et al.* (1992) apresentaram que, dentre os pacientes previstos pelo julgamento clínico e pelo escore APACHE II como provável óbito, mais de 40% sobreviveram. Com isso, foi concluído que nenhum método é confiável para prever com certeza a mortalidade de pacientes críticos. Foi discutida a questão de quais são as variáveis mais relevantes como preditoras de mortalidade, além da preocupação de como evitar a confusão que existe entre interpretar uma probabilidade estimada de mortalidade e a predição de se um determinado paciente irá sobreviver ou falecer.

Esses estudos relataram as vantagens e as limitações do uso de escores como preditores de mortalidade hospitalar, e a maioria dos escores disponíveis hoje são comparáveis em termos de previsão de resultados. Os modelos preditivos devem, no entanto, ser atualizados periodicamente para refletir as mudanças na prática médica e nas condições clínicas. Essas informações são importantes para o desenho do estudo apresentado no Capítulo 6, uma vez que houve a preocupação sobre a atualização dos dados de treino com base no incremento dos dados de entrada. Além disso, foi importante notar que, na maioria dos escores descritos, os métodos de cálculo se basearam em modelos de regressão, o que vai de encontro à hipótese de que o aprendizado de máquina poderia auxiliar o julgamento clínico. No Capítulo 5 é apresentado um modelo de aprendizado de máquina ajustado para os dados da base MIMIC-IV, o que contrasta com os escores tradicionais apresentados nesta subseção, nos quais os coeficientes e fatores são definidos e imutáveis, ainda que amplamente aceitos na literatura científica. Alguns desses escores dependem da presença de todas as informações para serem calculados, o que constitui uma potencial limitação de uso, principalmente para as primeiras horas de admissão hospitalar, momento em que há, em teoria, menos informações coletadas sobre um determinado paciente, quando comparado a uma internação com mais horas ou dias (SCHULZ *et al.*, 2021).

Por esse motivo, a Subseção 4.4 apresenta, a seguir, uma visão geral da base MIMIC,

além dos principais estudos já desenvolvidos a partir desse conjunto de dados. Esse conhecimento é importante para melhor entendimento das decisões de pesquisa apresentadas no Capítulo 5 em relação à seleção dos pacientes e das análises realizadas.

4.4 Estudos com a Base MIMIC

A primeira versão do MIMIC foi publicada no início dos anos 2000, financiado pelo *National Institute of Health* (NIH). Desde então, foram feitas atualizações em que se incluíram mais informações de pacientes, além de melhorias na arquitetura e na granularidade das informações. A sua relevância em termos de publicação científica se inicia com a versão MIMIC-IV, com cerca de 150 artigos atualmente indexados na base Pubmed. A MIMIC-IV é a base atualmente mais estudada, com mais de 500 artigos, e a MIMIC-IV é a mais recente, com menos de 100 artigos publicados. Essas contagens foram registradas em agosto de 2022 e, possivelmente, crescerão em número ao longo do tempo.

Os dados atualmente incluem sinais vitais, exames laboratoriais, anotações de prontuário, administração de medicamentos, códigos de procedimento e de diagnóstico, relatórios de imagem, registros de internação hospitalar e dados de alta, dentre outras informações. As variáveis incluídas são agrupadas em dois tipos básicos: (1) dados clínicos e (2) formas de onda de monitor.

O MIMIC se apresenta com as vantagens de possuir variáveis atualmente não disponíveis em outras bases, além da janela de mais de 10 anos de dados coletados e que, uma vez garantido o acesso, ser possível desenvolver diversas pesquisas sem necessidade de um novo credenciamento. Por essas razões, é amplamente utilizado por pesquisadores em todo o mundo, seja para pesquisas em informática clínica, epidemiologia ou aplicações de aprendizado de máquina. Alguns casos de uso interessantes nos quais o MIMIC foi utilizado foram nos cursos de aprendizado de máquina do *Massachusetts Institute of Technology*, Stanford, Columbia e Universidade do Texas, além de diversas pesquisas clínicas.

[Alistair E.W. Johnson *et al.* \(2016\)](#) descrevem o desenvolvimento e os resultados obtidos com a base de dados MIMIC. São descritas várias dificuldades relacionadas à tecnologia e a saúde, em que se destacam a dificuldade de adesão dos profissionais e dos hospitais aos prontuários eletrônicos, a questão da interoperabilidade de sistemas e a dificuldade de se reproduzir os estudos publicados. Em seguida, são descritas as características gerais dos pacientes presentes na base de dados, as classes de dados, o desenvolvimento da base (a origem dos dados a partir de diferentes sistemas e dificuldade

de agregação dos dados), os processos de anonimização e de distribuição dos códigos produzidos, e o modelo de dados (estrutura das tabelas). É abordada também a verificação técnica, que compreende os mecanismos de melhoria contínua que eles adotam, com ênfase para o registro de ocorrências diretamente no repositório de códigos oficial do projeto, do qual participam criadores e outros pesquisadores. Além disso, é descrito o processo de credenciamento para obter acesso à base, bem como exemplos de uso (situações em que o MIMIC já foi ou continua sendo usado, como, por exemplo, em aulas, *datatons* e pesquisas científicas). Por fim, é reforçada a importância da pesquisa científica colaborativa, da contribuição com melhorias ao repositório e da criação de padrões de modelos de dados, para serem reaproveitados por iniciativas semelhantes ao MIMIC.

Johnson, Pollard e Mark (2017) discutem que a reprodutibilidade é uma das premissas do método científico, mas que tem sido menos exigido como um requisito para se conseguir uma publicação atualmente. Os autores relatam que esta dificuldade pode ser devido à explicação incompleta (ou mesmo com conceitos que abrem margem para mais de uma interpretação), ou por não disponibilização da base de código ou do conjunto de dados usado em um determinado estudo. É discutido também que, em alguns casos, os autores dos artigos declaram que podem disponibilizar mediante contato, desde que julguem apropriado o seu compartilhamento, o que possibilita subjetividade na avaliação. Por outro lado, existe um interesse em que mais pesquisadores usem o MIMIC, razão pela qual este artigo apresentar a base MIMIC-IV como uma opção de acesso livre para realização de estudos científicos, além de mapear e reproduzir os resultados de estudos que utilizam a base de dados MIMIC-IV para predição de mortalidade. Para isso, foram selecionados 328 artigos publicados entre 2015 a 2017, com o critério de inclusão de que o tema central fosse de predição de mortalidade. Foi aplicado um processo de pesquisa em que foi selecionada a coorte de pacientes conforme os critérios descritos em cada estudo, seguido da seleção das variáveis, com subsequente aplicação em 2 modelos: Gradiente Extremo, do inglês, *Extreme Gradient Boosting* (XGB), classificado como um modelo não linear, e Regressão Logística (RL), considerado um modelo linear. Em seguida, os resultados obtidos foram comparados com os que foram reportados pelos 28 estudos selecionados através da área sob a curva ROC. Como resultados, ficou claro que os próprios criadores do MIMIC tiveram dificuldades em reproduzir os estudos que usaram a própria base. Na maioria das coortes reproduzidas, o número de pacientes incluídos foi maior do que o relatado pelo estudo original. A proporção de óbitos reportada diferiu em 20 estudos, e a área sob a curva ROC diferiu em todos os estudos. Independentemente do modelo usado no estudo original, o modelo de RL foi igual ou superior em 64% dos casos, e o de XGB

foi igual ou superior em 82% do que reportado nos artigos. Como fatores de dificuldade para reprodução dos estudos, os principais foram: poucos detalhes sobre os critérios para seleção das variáveis e dos pacientes; heterogeneidade na forma de apresentar os resultados dos estudos; e falta de compartilhamento dos códigos usados nas pesquisas. Por fim, os autores disponibilizaram um repositório de código com todos os passos da pesquisa, além de ter sido discutida a importância de não serem omitidos passos e achados nas pesquisas científicas.

O estudo sobre a reprodutibilidade dos estudos foi fundamental para se alcançar o objetivo desta dissertação, especialmente os passos iniciais de modelagem dos dados da base MIMIC, que será detalhado no Capítulo 5. Ao entender como outras pesquisas foram desenvolvidas utilizando esta mesma base, foi possível reconhecer e tratar dados confusos, tendenciosos e ausentes. Além disso, houve uma preocupação adicional em se garantir que todos os códigos produzidos e apresentados no Capítulo 6 efetivamente estiverem acessíveis e utilizáveis pela comunidade científica.

Apesar de a base MIMIC conter dados de várias UTIs, a amostra ainda provém de um único centro hospitalar, mesmo em sua versão mais recente. Esse fator pode ser um limitador quanto à generalização dos resultados de pesquisas realizadas com esta base, como a apresentada nesta dissertação. Entretanto, o estrato de pacientes incluídos na coorte MIMIC pode ser considerado representativo da população geral de pacientes de terapia intensiva, tendo em vista o perfil de comorbidades e a taxa de mortalidade hospitalar destes pacientes, semelhante à relatada para pacientes de UTI de outros centros em diferentes países no mesmo período (ZIMMERMAN; KRAMER; KNAUS, 2013).

Nesta subseção foram apresentados os artigos publicados que tiveram como objeto de estudo a base MIMIC e que foram mais relevantes para o desenvolvimento desta pesquisa. A partir da sua análise na íntegra, foi possível identificar oportunidades de estudo que apoiaram a hipótese de experimentar técnicas de aprendizado de máquina com esta base de pacientes. Além disso, ficou claro que estudos observacionais podem ter vieses de confusão não identificados e que influenciam os efeitos da exposição primária, o que reforça a necessidade de transparência da pesquisa e avaliação das limitações do estudo.

Neste capítulo, foram apresentados os principais estudos publicados que serviram como embasamento para esta dissertação. A escolha pela análise detalhada de estudos de predição de desfechos teve como principal objetivo entender quais seriam algumas das técnicas estatísticas e de aprendizado de máquina para se experimentar com a base de pacientes disponível. Já o estudo dos escores de risco foi importante para dar entendimento

sobre variáveis clínicas e epidemiológicas que estão presentes em PEP e que poderiam ser utilizadas nas estimativas de mortalidade hospitalar, o que foi posteriormente utilizado na elaboração do sistema descrito no Capítulo 6. Por fim, foi dada ênfase sobre artigos que exploraram a base de dados MIMIC para que fosse feita a modelagem correta dos dados e para que as informações disponíveis pudessem ser interpretadas de maneira correta, ainda a maioria das publicações disponíveis correspondessem a versões anteriores àquela que será estudada nesta dissertação. A análise desses estudos sobre a base MIMIC confirmou que ela seria uma boa fonte de dados para estudo, tendo em vista a confiabilidade das informações, a existência de uma boa documentação e a adequação para uso em estudos de modelagem preditiva. No próximo capítulo serão apresentados os resultados dos estudos realizados na base de pacientes MIMIC-IV.

5 Análises e Resultados

Este capítulo apresenta as análises e os estudos realizados com a base de dados MIMIC e os principais resultados obtidos. Para essa etapa, foi fundamental a leitura dos artigos que estabeleceram os principais conceitos médicos utilizados na interpretação das informações da base de dados, além de diretrizes técnicas sobre o uso apropriado das informações disponíveis na base MIMIC, conforme descrito no Capítulo 2.

Na Subseção 5.1, são descritas a consistência dos dados dos pacientes selecionados e as etapas de pre-processamento e organização dos dados. A Subseção 5.1.1 detalha as decisões técnicas para tratamento dos dados: as etapas desenvolvidas para dados ausentes, as decisões relacionadas a dados discrepantes, os tratamentos dos dados categóricos, e as estratégias adotadas com dados contínuos. A Subseção 5.2 apresenta as análises realizadas, que se iniciam pela Subseção 5.2.1, que apresenta a análise exploratória de dados. Na Subseção 5.2.2, é descrita a análise estatística. A Subseção 5.2.3 apresenta os estudos relacionados à análise de sobrevivência e à análise de riscos proporcionais. Na Subseção 5.2.4, é apresentado o estudo de seleção de características da base de dados, e a Subseção 5.2.5 descreve os estudos de aprendizado de máquina. Por fim, a Subseção 5.3 apresenta os principais resultados obtidos a partir dos estudos descritos, bem como é feita uma análise de como estes resultados estão conectados com o objetivo da dissertação.

O primeiro procedimento dessa fase da pesquisa foi o entendimento geral das informações dos pacientes disponibilizada na base de dados. Para isso, a Subseção 5.1 descreve as ações de pré-processamento, que resultou no mapeamento de diversos itens que seriam avaliados quanto à necessidade de tratamento dos dados. Na Subseção 5.1.1, são discutidos os casos em que houve, efetivamente, necessidade de tratamento de dados.

5.1 Pre-processamento

Na etapa de pre-processamento, foram avaliadas a consistência dos registros (com mensuração de dados faltantes) e a relevância para utilização no estudo.

O conjunto de dados necessitou de transformações básicas para uso neste estudo. Os dados brutos são obtidos do *Physionet* em formato Valores Separados por Vírgula, do inglês, *Comma Separated Values* (CSV) e, portanto, foi necessária uma etapa de modelagem das tabelas para que o conjunto de dados fosse representativo da jornada de um paciente, da admissão à alta hospitalar. Para esta finalidade, foi criado um banco de dados em PostgreSQL versão 14 contendo um esquema principal, que inclui todas as tabelas criadas a partir da importação dos arquivos brutos, além de um esquema auxiliar contendo visualizações materializadas e tabelas com dados agregados. A modelagem priorizou agregar as informações de cada internação hospitalar, de modo que um único paciente poderia ter diversos registros, cada um deles representativo de um atendimento, o qual poderia ou não significar uma internação, ou transferência de setor hospitalar. Dessa forma, todos os dados clínicos referentes a cada internação poderiam ser acompanhados temporalmente, independente do setor hospitalar em que foram capturados, além de ser possível compará-los em relação a outra internação do mesmo paciente, ou a dados clínicos de outros pacientes.

Para utilização nos notebooks, foram gerados novos arquivos em formato CSV que continham apenas o estrato da base de interesse. Os dados selecionados estão descritos na Tabela 7, disponível no Apêndice B. As variáveis demográficas, clínicas e bioquímicas foram consideradas *features* para os modelos experimentados na Subseção 5.2. Em relação aos tipos de dados, *int64* e *float64* representam dados numéricos inteiros e fracionários, respectivamente. O termo *object* indica uma estrutura de dado que contém caracteres.

A seleção das variáveis para estudo foi fundamentada nos seguintes critérios: (1) revisão de literatura, (2) conhecimento médico e (3) disponibilidade na prática clínica. A seguir, cada um dos critérios será detalhado.

A revisão da literatura apresentada na Subseção 4.4 foi a etapa inicial para fundamentar a seleção de variáveis. Nessa fase, foi possível ter uma visão geral de quais dados estariam disponíveis na base MIMIC, além de quais variáveis dessa base já foram utilizadas em outros estudos. Dessa forma, foi possível evitar explorar dados que, apesar de aparentemente disponíveis, não tinham informação contida (ou seja, havia um registro nas tabelas, porém este era sempre nulo), como foi o caso da Troponina I.

O conhecimento médico foi o segundo critério para filtragem das variáveis. Nessa etapa, foi avaliada a coerência entre as associações de informações, para evitar a análise de dados redundantes ou diretamente ligados. Um exemplo dessa situação foi a não inclusão de dados de Creatinina de pacientes que estivessem sendo submetidos a terapia

renal substitutiva durante a internação, já que seriam valores artificialmente baixos e que não representariam o grau de disfunção renal.

Por fim, a avaliação da disponibilidade das variáveis na prática clínica serviu para analisar o percentual de pacientes em que cada dado laboratorial estava presente, além de ponderar se cada exame é disponível em outros centros hospitalares e se tem uma correspondência para o contexto do SUS.

Uma vez definidos quais dados fariam parte do conjunto de informações pertinentes a essa pesquisa, o passo seguinte foi a modelagem dos dados, conforme descrito na Subseção 5.1.1 a seguir.

Para facilitar a modelagem apropriada e a identificação de fatores de confusão em relação aos dados da base MIMIC-IV, foi construída uma representação mais eficaz do conjunto de dados, com menos dimensões e com rótulos mais descritivos, e sem comprometer a integridade dos dados originais.

Para isso, o primeiro passo foi a inserção dos conjuntos de arquivos CSV em uma base de dados PostgreSQL (VOHRA, 2016), com um processo de engenharia de dados que seguiu as recomendações da documentação oficial do MIMIC (JOHNSON, A. *et al.*, 2020).

Em seguida, uma série de consultas em Linguagem de Consulta Estruturada, do inglês, *Structured Query Language* (SQL) foram feitas nessa base reconstruída e estão documentadas no [repositório de códigos desta pesquisa](#)¹. Inicialmente, o objetivo foi extrair as variáveis de interesse sem redundância nas informações. Assim, o conjunto de dados foi remodelado de forma que a consistência fosse garantida em três níveis: (1) cada variável fosse representada por uma coluna, (2) cada observação fosse representada por uma linha de registro e (3) cada valor fosse apresentado em uma célula.

Esse conjunto de dados exportado do MIMIC-IV foi utilizado nas etapas subsequentes de tratamento de dados que estão descritas na Subseção 5.1.1 a seguir.

5.1.1 Tratamento dos Dados

Como a maioria dos modelos estatísticos opera apenas em observações completas de variáveis de exposição e resultado, foi necessário lidar com dados incompletos ou ausentes. Esta subseção descreve o procedimento de tratamento dos dados da base MIMIC-IV, com

¹Disponível em: <https://github.com/ML4Healthcare/mimicivresearch>. Acesso em 31 de dezembro de 2022.

a apresentação das possíveis causas para dados ausentes e a descrição descreve os estudos realizados para se definir a melhor estratégia para tratá-los. Além disso, serão discutidas as abordagens sobre dados *outliers*, dados categóricos e dados numéricos contínuos.

No Capítulo 3, foram apresentados os principais conceitos teóricos relacionados aos dados ausentes. O entendimento dos tipos de ausência de dados, as técnicas para identificação da proporção de dados ausentes e as abordagens para como lidar com os dados ausentes foram as três etapas necessárias para tratamento dos dados ausentes e estão apresentadas nos próximos parágrafos.

Em relação aos tipos de ausência de dados, foram identificadas três razões básicas: (1) variáveis que foram medidas, mas não inseridas no sistema, (2) falhas temporárias na medida da variável, ou (3) variável não é medida porque não está relacionada com a condição do paciente.

Para identificar a proporção de dados ausentes, foi feito um cálculo de porcentagem de registros presentes ou ausentes para cada variável, através da comparação do percentual de pacientes com e sem a informação, e para cada paciente, por meio da avaliação de quantas variáveis cada paciente possuía sem valor válido. Isso foi necessário para a decisão sobre quais variáveis seriam consideradas candidatas para remoção ou imputação de dados, bem como para definir quais pacientes estariam elegíveis para as etapas seguintes deste estudo. O resultado destes cálculos está apresentado na coluna “Ausentes” da Tabela 7, disponível no Apêndice B, e demonstra que a maioria dos dados básicos dos pacientes estava presente, e que as ausências de dados nesse subconjunto de informações brutas correspondiam a valores iguais ou inferiores a 8,15%.

O conjunto de dados foi então submetido a um estudo com técnicas padrões para resolução de dados ausentes (ALJUAID; SASI, 2016). Os métodos foram avaliados considerando-se o conjunto de dados de interesse, os motivos da falta e a proporção de dados faltantes, para resolver a ausência de dados com o mínimo de polarização possível. Por este motivo, técnicas baseadas em exclusão, como análise de caso completo ou análise de caso disponível, ou métodos de imputação única, como substituição de valores ausentes por média ou mediana, foram descartados. Optou-se por estudar a imputação baseada em modelos, na qual um modelo preditivo é criado para estimar valores que substituirão os dados ausentes. Nesse caso, o conjunto de dados é dividido em dois subconjuntos: um sem valores ausentes para a variável em avaliação (usada para treinar o modelo) e um contendo valores ausentes, para realizar as estimativas.

O desempenho dos modelos preditivos foi testado no conjunto de dados completado

por vários métodos de imputação, da seguinte forma: a partir de uma amostra do conjunto de dados sem dados ausentes, foram introduzidas proporções crescentes de dados perdidos aleatoriamente, com início de 5% e incrementos de 5%, até o total de 50% de ausência. Com isso, os dados ausentes foram reconstruídos usando os métodos mais comuns (Regressão Linear, Regressão Estocástica, Imputação Múltipla, Vizinhos mais próximos). Por fim, foi calculado o Erro Médio Quadrado entre os dados reconstruídos e os originais, para cada método e cada proporção de dados perdidos.

O método com melhor desempenho no nível de dados ausentes foi o algoritmo de análise iterativa de componentes principais (*missMDA*). A Figura 6 indica que as melhores estratégias para imputação dos dados ausentes seriam aquelas baseadas em modelos. Nesse caso, o conjunto de dados foi dividido em dois subconjuntos: um subconjunto sem valores ausentes para a variável em avaliação (usada para treinar o modelo) e um subconjunto contendo valores ausentes, para serem estimados. A técnica com menor média de erro absoluto, independente se MCAR, MAR ou MNAR, foi o *missMDA Regularized*².

No algoritmo *missMDA*, o primeiro passo consiste em imputar valores ausentes com valores que seguem uma distribuição gaussiana com média e desvio padrão calculados a partir dos valores observados. Em seguida, é executada uma análise de componentes principais no conjunto de dados completo. No passo seguinte, os valores ausentes são imputados com base nas fórmulas de reconstrução de ordem de número de componentes. Essas etapas de estimação dos parâmetros através da análise dos componentes principais e imputação dos valores faltantes usando a matriz ajustada (regularizada) são iteradas até que todos os dados ausentes sejam preenchidos (FRANCOIS HUSSON, 2013; NUNES; KLÜCK; FACHEL, 2010).

Ao final dessas etapas, os registros nos quais houvesse qualquer valor ausente foram preenchidos pelo método de análise de componentes principais conforme o algoritmo *missMDA*, que considera as semelhanças entre as observações e a relação entre as variáveis (ABDI; WILLIAMS, 2010). Com isso, o conjunto de dados resultante passou a não possuir dados ausentes.

Após esta etapa, foi feita a busca de dados *outliers*, cuja definição adotada para *outlier* foi de *Z-score* maior ou igual a três desvios-padrão, nos casos nos quais a média e o desvio padrão pudessem ser estimados ou estivessem disponíveis no conhecimento do domínio (SHIFFLER, 1988). Na área médica, as principais fontes de *outliers* são:

²Disponível em: <https://cran.r-project.org/web/packages/missMDA/index.html>. Acesso em: 31 de dezembro de 2022.

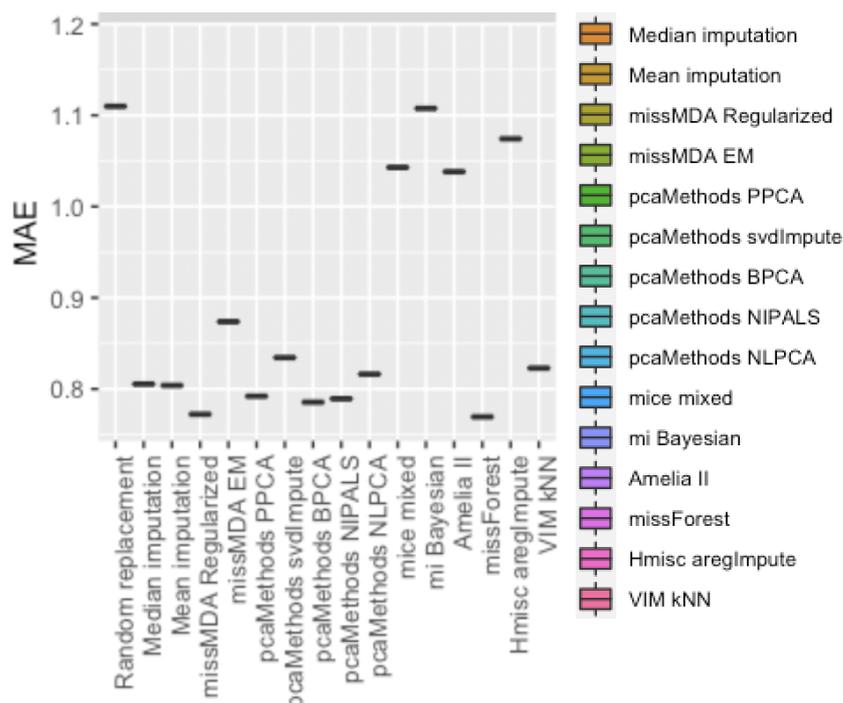


Figura 6: Comparação do erro absoluto médio avaliado para cada uma das técnicas de preenchimento de dados. MAE: Erro absoluto médio.

(1) erros humanos de registro; (2) problemas técnicos de equipamentos; (3) anomalias decorrentes de comportamentos específicos de um determinado paciente; e (4) variação natural de resultados entre os pacientes.

Nesta pesquisa, foram feitas duas abordagens para tratamento dos *outliers*: (1) uma técnica de análise de conhecimento especializado, que determinou valores máximos e mínimos para definição dos intervalos normais, críticos e possíveis; seguido por (2) uma pesquisa ativa na base para dados *outliers* a partir da análise de valores extremos acima de 3 desvios-padrão da média (DONG; PENG, 2013).

O resultado desta análise está representado na Tabela 8, disponível no Apêndice B, e indica que o conjunto de dados não continha valores impossíveis como, por exemplo, um valor negativo de frequência cardíaca, e que os valores extremos presentes se deviam, provavelmente, a variação biológica, e não a erros de mensuração ou de registro. Por esse motivo, os *outliers* presentes nessa base de dados contém informações úteis em seus valores extremos, e a exclusão automática resultaria na perda dessas informações. Conseqüentemente, não houve necessidade de eliminação de registros classificados como *outliers*.

Em relação aos dados categóricos, seu tratamento foi feito com a técnica de *one hot encoding* por dois principais motivos: eram dados não ordinais, ou seja, sem uma ordem

natural, e serem poucas ocorrências (BASKORO; SUNINDYO, 2019). Neste caso, as variáveis qualitativas foram representadas de forma binária. Isso foi feito com a criação de colunas adicionais no banco de dados, com valores possíveis de 0 ou 1, conforme indica ausência ou presença de um valor. Por exemplo, houve criação de uma coluna binária para indicar gênero masculino e outra para representar gênero feminino, que substituiu a informação previamente fornecida coluna categórica de gênero. Essa etapa teve utilidade futura para padronizar os dados como tipo numérico para serem utilizados nos modelos de aprendizado de máquina. Tratamento semelhante foi aplicado em outras variáveis categóricas, como etnia e local da alta hospitalar (POTDAR; PARDAWALA; PAI, 2017).

Baseado em outros estudos que agruparam pacientes com características clínicas semelhantes, foram também criadas categorias para o agrupamento de informações nominais, conforme a seguinte escala: 18 a 24 anos, 25 a 49 anos, 50 a 64 anos, 65 a 80 anos e acima de 80 anos (BAE *et al.*, 2020; DAMASKOS *et al.*, 2021).

Os dados quantitativos podem ser classificados em dois grupos: (1) discretos, em que se incluem os números inteiros, ou (2) contínuos, que podem apresentar um número variável de valores após a vírgula. Como exemplo, o número de vezes que um paciente é admitido em um hospital é discreto (um paciente não pode ser admitido 1,7 vezes), enquanto o seu peso corporal é um dado contínuo, pois pode assumir qualquer valor em uma faixa (OSTLUND *et al.*, 2011).

Por convenção, as estratégias adotadas para tratamento de dados numéricos seguiram quatro princípios: (1) foram realizados arredondamentos para duas casas decimais no caso de variáveis quantitativas contínuas de maior precisão como, por exemplo, valores de temperatura e peso corporal (SIEST *et al.*, 2012), (2) para construção de gráficos e imagens, as quantidades, quando presentes, são apresentadas conforme a informação original da base de dados (nesse caso, o número de pacientes, por exemplo, é sempre representado de maneira exata), (3) no caso de tabelas com porcentagens, a regra é arredondar para uma casa decimal, e (4) durante a descrição em texto, a representação é com dois dígitos significativos (diferentes de zero) (COLE, 2015). Essa abordagem foi utilizada para garantir um limiar de precisão dos resultados de cálculos estatísticos (OZARDA, 2016).

Todas essas etapas de tratamento dos dados, incluindo as decisões sobre dados ausentes, dados discretos, dados categóricos e dados contínuos, foram consideradas pré-requisitos para o início do estudo. A Subseção 5.2, a seguir, apresenta os estudos realizados sobre esse novo conjunto de dados.

5.2 Estudos realizados

Esta subseção descreve os estudos realizados a partir do conjunto de dados pré-processado e tratado, e está dividida em cinco subseções. A Subseção 5.2.1 apresenta os passos da análise exploratória de dados. A Subseção 5.2.2 descreve a análise estatística descritiva e inferencial. Na Subseção 5.2.3, é descrita a análise de sobrevivência, o porquê de ter sido feita, e suas principais conclusões, além de detalhar a análise de riscos proporcionais a partir do modelo de regressão de Cox. Na Subseção 5.2.4 são apresentados o estudo de entropia e ganho de informação. Por fim, a Subseção 5.2.5 detalha os estudos de modelagem preditiva.

5.2.1 Análise Exploratória de Dados

O primeiro estudo realizado sobre a base MIMIC-IV foi a análise exploratória de dados, cujo objetivo foi o de examinar os dados previamente à aplicação de qualquer técnica estatística. O produto esperado desse estudo foi o de se obterem representações gráficas e tabelas que seriam úteis para consolidar padrões, tendência e qualidade nos dados e, com isso, auxiliar na identificação de um grupo de pacientes para estudos subsequentes.

Das 76540 admissões registradas em UTI, 9541 delas foram em CICU, e dessas, foram excluídas 626 readmissões. 1795 registros foram tratados devido à ausência de uma ou mais das variáveis de interesse conforme a técnica descrita na Subseção 5.1.1. Um total de 8035 pacientes foram incluídos no estudo, com média de idade de 72,1 anos e proporção 57% do sexo masculino. A taxa de mortalidade intra-hospitalar foi de 17,5%.

Nos pacientes que faleceram durante a internação, houve maiores médias de idade, tempo de internação em CICU mais longo e registros mais elevados de frequência cardíaca e respiratória, todos com $valor - p < 0,001$ (Tabela 9, disponível no Apêndice B). Não houve diferença estatisticamente significativa para níveis de glicemia e tempo de internação hospitalar, as quais poderiam ser alterações também possíveis em pacientes que foram a óbito (HOSOI; OZONE; HAMANO, 2020). Quando comparados os dois grupos em relação ao sexo, também não houve diferença com significância estatística, o que pode sugerir uma importância maior para outros dados dos pacientes, como será estudado na Subseção 5.2.4.

O tempo médio de permanência na CICU foi de 2,9 dias (desvio padrão: 0,2). Quando estratificados por alta e morte intra-hospitalar, observaram-se médias de 2,5 e 3,6 dias de internação neste setor, respectivamente. Conforme esperado, existiu uma maior proporção

de registros à esquerda nos gráficos de tempo de internação hospitalar, o que indica que a maioria das internações tem um tempo mais curto, conforme representado na Figura 7. Esse é um comportamento esperado para o perfil de pacientes de unidades de terapia intensiva (KRAMER, 2017; VERBURG *et al.*, 2017).

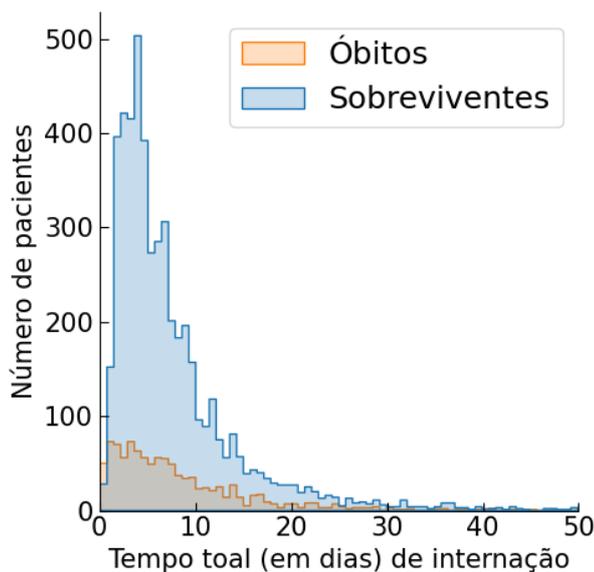


Figura 7: Distribuição do número de pacientes sobreviventes e morte intra-hospitalar ao longo do tempo de internação.

A incidência de readmissões foi de 7,42%. A mortalidade foi maior no grupo de pacientes em que ocorreu uma ou mais admissões (18,61% no primeiro grupo, e 28,73% no segundo), o que reitera o achado de outros estudos (HOSEIN *et al.*, 2014; ROJAS *et al.*, 2020).

O estudo subsequente à análise exploratória de dados foi a análise estatística, cujo objetivo foi o de explorar a relação dos dados com a amostra populacional selecionada. Esse estudo está apresentada na próxima subseção.

5.2.2 Análise Estatística

O segundo estudo realizado sobre o conjunto de dados tratado foi uma análise estatística descritiva e inferencial. A descrição da amostra de pacientes estudada foi feita com medidas de frequência para as variáveis qualitativas e média e desvio-padrão para as variáveis contínuas.

Foram considerados dois grupos de pacientes internados em CICU: pacientes sobreviventes (e que, conseqüentemente, tiveram alta hospitalar naquela internação) e pacientes

que faleceram durante a internação hospitalar. Eles foram comparados através do Teste-T de duas amostras para variáveis contínuas e teste de Qui-quadrado para variáveis categóricas (BOX, 1987).

O Teste-T de duas amostras foi escolhido porque se tratavam de amostras independentes (no caso, pacientes que sobreviveram e pacientes que faleceram) em que não eram previamente conhecidos as médias e os desvios padrão das variáveis. Já o teste Qui-quadrado foi utilizado porque havia independência entre as observações e variáveis.

A correlação pelo método de *Pearson* foi utilizada para avaliar a presença de relações lineares entre as variáveis (PEARSON, R. K., 2018). Esta técnica foi escolhida porque: (1) o valor obtido pelo seu cálculo é baseado em métodos de covariância, o que se aplica a este cenário de diversidade de variáveis epidemiológicas, fisiológicas e laboratoriais; (2) a amostra disponível era suficientemente grande para fornecer estimativas confiáveis; (3) as variáveis avaliadas eram contínuas; e (4) a distribuição das variáveis era semelhante à curva normal (PEARSON, K., 1920). Foram adotados os seguintes valores de classificação para relação linear, propostos por (EVANS, R. H., 1996): 0,0-0,19: muito fraca; 0,20-0,39: fraca; 0,40-0,59: moderada; 0,60-0,79: forte; 0,8-1,0: muito forte (MUKAKA, 2012).

O valor de um sinal vital foi considerado um *outlier* se estiver além dos intervalos fisiológicos plausíveis com base no conhecimento clínico (OZARDA *et al.*, 2018; SIEST *et al.*, 2012).

A Tabela 9, disponível no Apêndice B, apresenta as principais características da população, comparando os grupos que tiveram ou não óbito intra-hospitalar. Sua análise demonstra que, conforme esperado, os valores correspondentes a mediana de idade foram inferiores no grupo de sobreviventes. Da mesma forma, sinais vitais como a frequência cardíaca apresentaram valor médio mais próximo da faixa considerada normal neste grupo de pacientes. Por outro lado, foram observados valores de glicemia maiores no grupo de pacientes que sobreviveram, o que pode corresponder a maior frequência de diagnósticos de diabetes nessa população, hiperglicemia relacionada ao estresse, ou também a outros fatores não discriminados nesta análise (MAMADJANOV *et al.*, 2021). Em relação aos exames laboratoriais, é importante destacar a média de valores de creatinina, superior no grupo de pacientes que faleceram, bem como de potássio, que pode estar relacionado a piora da função renal, o que remete a uma associação frequente entre piora da função cardíaca e renal (RONCO; BELLASI; LULLO, 2018). Conforme esperado, valores de Troponina T e NT-proBNP foram superiores no grupo de pacientes de morte intra-hospitalar, o que reforça sua associação com pior prognóstico cardiovascular (GAGGIN;

JANUZZI, 2013). Essas informações reforçam que a coorte de pacientes é representativa de uma população com comorbidades cardiovasculares, tendo em vista as anormalidades laboratoriais descritas que remetem ao acometimento cardíaco, renal e endócrino.

Em relação aos estudos de correlação, não foi encontrada correlação forte entre nenhuma das variáveis selecionadas junto à coorte de pacientes. Isso sugere não haver um grau forte de dependência e poderão ser utilizadas para os modelos experimentais. Caso houvesse algum resultado de correlação superior a 0,8, seria feita a remoção de uma delas para que não influenciassem nas etapas futuras de análise (SMILDE *et al.*, 2008).

Conforme esperado, os maiores índices de correlação foram entre variáveis medidas em conjunto, relacionados à aferição da pressão arterial (0,48 para PAS e PAM, 0,59 entre PAD e PAM e 0,48 entre PAS e PAD). Isso era esperado porque, exceto em situações clínicas menos frequentes em que há pressão arterial divergente, os comportamentos de pressão tendem a seguir determinados padrões já conhecidos (ALLEN; KHAN, 2021).

A correlação entre tempo de internação hospitalar e tempo de internação em CICU foi de 0,42. Isso pode sugerir que pacientes que necessitam de cuidados intensivos poderão ter uma internação hospitalar mais prolongada, na qual se inclui o período fora da unidade, mas ainda no hospital. A literatura reforça essa hipótese, que pode estar associada com as causas da internação (KURTZ *et al.*, 2022).

Já em relação às comorbidades, a correlação entre doença renal e diabetes mellitus com complicações crônicas obteve o maior índice: 0,37. A associação dessas doenças é uma condição conhecida e pode apresentar prevalências de 30 a 50% na população (KOYE *et al.*, 2018; BRAUNWALD, 2019), o que está na faixa identificada nessa coorte de pacientes.

Ao contrário do que poderia ser esperado, já que uma das principais causas de Insuficiência Cardíaca na população é isquêmica (ZIAEIAN; FONAROW, 2016), a correlação entre Insuficiência Cardíaca e Infarto do Miocárdio com mortalidade foi de apenas 0,07 nestes pacientes. Apesar de a isquemia miocárdica ser a etiologia mais comum de Insuficiência Cardíaca (CLELAND; MCGOWAN, 1999), é possível que esse valor baixo de correlação seja porque não tenha sido a causa da internação estudada, mas apenas uma comorbidade do paciente registrada em prontuário.

O produto combinado das análises exploratória e estatística foi um entendimento aprofundado da amostra populacional, das suas características e do seu comportamento ao longo das internações, bem como das limitações relacionadas à qualidade dos dados

disponíveis. A análise de sobrevivência, descrita na próxima subseção, foi o estudo subsequente, e teve como objetivo descrever o tempo de sobrevivência da população e os efeitos das variáveis na sobrevivência. Com isso, seriam obtidas mais informações para embasar os estudos futuros de aprendizado de máquina com objetivo de predição de mortalidade nessa base de pacientes.

5.2.3 Análise de Sobrevivência

Após a compreensão das características gerais da base MIMIC-IV, da seleção de uma amostra e da análise estatística de seus principais atributos, o passo seguinte foi executar uma análise de sobrevivência sobre essa população selecionada. O resultado esperado ao final dessa etapa foi de se obter estimativas como, por exemplo, o tempo médio de vida após a primeira internação registrada na CICU, bem como a taxa de sobrevida esperada após um determinado período. Essas informações são relevantes para auxiliar na avaliação nos resultados de predição descritos na Subseção 5.2.5.

Essa análise de sobrevivência foi feita na coorte de pacientes descrita na Subseção 2.2.1 usando o método de Kaplan-Meier (KM), com teste de *Log-rank* para estabelecer comparações entre os grupos de óbito e sobrevida (BEWICK; CHEEK; BALL, 2004). As curvas de KM são, em teoria, uma boa escolha para estudo dos dados até o tempo do evento de interesse (neste caso, morte) antes de prosseguir com qualquer modelo estatístico mais complexo, e por isso foram utilizadas nessa análise (KAPLAN; MEIER, 1958). Os indivíduos que não morreram durante a internação foram rotulados como censurados à direita.

Ao final do período de acompanhamento, a sobrevida global estimada pelo método de KM sem estratificação foi de 86% após 7 dias, 70% após 14 dias e 50% com 28 dias de hospitalização. Essa informação evidencia que, nesta população estudada, metade dos pacientes que permanece internados por um mês na CICU irá falecer. A curva de sobrevivência global, representada na Subseção 8, sugere que, após um período inicial de declínio, a probabilidade de sobrevivência tende a atingir um valor estável por volta de 75 dias.

A sobrevida global mediana foi de 28 dias (IC 95%: 26–30 dias). Quando agrupados por gêneros, os valores foram de 27 (IC 95%: 25–31) para o sexo masculino e 26 (IC 95%: 22–31) para o feminino (*valor* – $p < 0,005$).

Quando segmentados conforme a presença das comorbidades descritas na Subse-

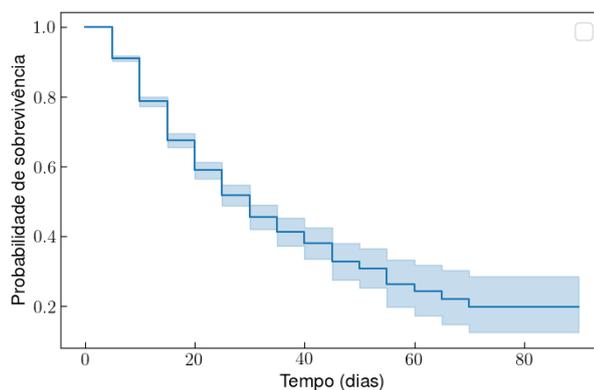


Figura 8: Curva de sobrevivência de Kaplan-Meier, definidas como a probabilidade de sobreviver em um determinado período. A área sombreada representa o IC.

ção 2.2.1, a probabilidade de sobrevivência tendeu a diminuir nos grupos com a doença, em todos os cenários avaliados, conforme reforça a literatura (GUIDET *et al.*, 2020). Os tempos médios de sobrevivência por comorbidades de interesse estão apresentados na Tabela 10, disponível no Apêndice B, sendo que o menor tempo de sobrevida média foi observado em pacientes com diagnóstico de Doença Arterial Periférica. Importante ressaltar que os diagnósticos não são excludentes e, frequentemente, um mesmo paciente possui mais de um, ou até todos os quatro diagnósticos reportados. A associação destas doenças está descrita na literatura médica (VIRANI *et al.*, 2020).

A partir do entendimento da curva de sobrevivência dessa população, o próximo passo foi explorar quais fatores mais explicariam esse comportamento. Para esse objetivo, foi feito um estudo de regressão de sobrevivência. A abordagem mais popular para estudar dados de saúde relacionados a mortalidade é o Modelo de Riscos Proporcionais de Cox (COX, D. R., 1972). A predição do risco e a trajetória de pacientes em terapia intensiva usando modelos de sobrevivência está bem estabelecida na literatura (HUG, 2007; HUG; SZOLOVITS, 2009; GHASSEMI; NAUMANN *et al.*, 2014). Por ser flexível e simples de implementar, a análise de sobrevivência pelo Modelo de Cox foi escolhida como método inicial para este estudo. O racional proposto é o de que o risco logarítmico de um indivíduo é uma função linear de suas co-variáveis, adicionado a um risco basal ao nível populacional.

Quando realizada a análise de riscos proporcionais na coorte, os pacientes que apresentaram menor risco de óbito na CICU foram aqueles que apresentavam as seguintes características: menor idade e frequências respiratória e cardíaca mais baixas (mas nos valores considerados normais). Todas essas características são condizentes com um perfil de saúde com menos propensão a doenças, de uma maneira geral, e também a menor morbidade cardiovascular (NORTH; SINCLAIR, 2012; LEONE; LANDINI, 2021).

A Tabela 11, disponível no Apêndice B, apresenta as variáveis utilizadas e os coeficientes obtidos pelo modelo de Cox. Os valores de HR são mostrados para a análise de riscos proporcionais, com os limites inferior e superior do IC de 95%. A interpretação dos resultados indica, por exemplo, que um aumento de uma unidade no valor da variável idade fará com que o risco da linha de base aumente por um fator de 1,02, o que representa 2% de aumento. No modelo de risco proporcional de Cox, um risco maior significa mais risco de ocorrência do evento (GEORGE; SEALS; ABAN, 2014). Nesse sentido, a variável temperatura corporal foi inversamente associada a mortalidade intra-hospitalar pelo modelo de Cox com significância estatística. Neste caso, o incremento no valor das variáveis quantitativas reduziu o risco da linha de base. A interpretação clínica dessa informação é que, nesta coorte de pacientes, aqueles que tinham um valor de temperatura corporal próximo ao normal (ou em níveis ligeiramente superiores à faixa normal) tinham maior chance de sobrevivência. Dessa forma, pacientes com alterações de sinais vitais não necessariamente teriam um pior prognóstico. Por se tratar de um cenário plausível, tende a ser interpretado como uma característica da população estudada, mais do que um resultado a ser interpretado.

Para que o modelo pudesse ser considerado adequado aos dados, foi necessário avaliar se a suposição de taxas de falha proporcionais foi atendida. Isso foi feito a partir da obtenção do coeficiente de correlação de Pearson e os resíduos padronizados de Schoenfeld e os tempos para cada co-variável, em que valores de coeficiente próximos de zero mostrariam não haver evidências para a rejeição da suposição de riscos proporcionais (CLARK *et al.*, 2003). Neste caso, as seguintes co-variáveis não apresentaram significância estatística: leucócitos, bicarbonato, *anion gap*, glicose, temperatura, débito urinário e NT-proBNP.

Por esse motivo, foi também realizado o teste paramétrico de Weibull, no qual a variável idade se apresentou como fator que reduziu o risco basal de morte intra-hospitalar (HR: 0,98). Neste caso, o incremento de 1 ano representa um aumento de chance de morte em um fator de $exp(1,03)$, aproximadamente, 3%. Clinicamente, isso significaria que, embora os pacientes mais velhos tenham uma chance maior de óbito, outras variáveis podem influenciar mais nesse desfecho, para esta população. As variáveis que representaram um HR inferior a 1 com significância estatística foram, além da idade, diagnóstico de infarto do miocárdio e maior frequência respiratória na admissão.

Quando se comparou o desempenho dos modelos semi-paramétrico e paramétrico (Cox e Weibull, respectivamente), os valores obtidos pelo critério de informação de Akaike (PORTET, 2020) foram de 19155 e 11622, o que indica que o modelo de Weibull teve

melhor desempenho. Já pelo Índice de Concordância (LONGATO; VETTORETTI; CAMILLO, 2020), foram obtidos os valores 0,656 e 0,657, respectivamente. Esses resultados indicam que, apesar da pouca diferença, o modelo de Weibull efetua uma estimativa mais apropriada.

Buscando-se ampliar os resultados, uma regressão em etapas foi realizada para encontrar outras variáveis de interesse que fossem amplamente disponíveis nesse subgrupo de pacientes e, simultaneamente, que tivesse relevância fisiológica e estatística. Os dados foram processado pelo modelo de Cox com ajuste para idade e sexo.

Os resultados estão apresentados na Tabela 12, disponível no Apêndice B. O risco mais significativo foi o de níveis de potássio séricos aumentados como um fator independente de risco para mortalidade intra-hospitalar (HR = 1,22, IC 95%: 1,14-1,30). Creatinina, hematócrito e leucometria também se revelaram variáveis representativas de risco de morte, porém com menor contribuição para o modelo. Esses resultados remetem à importância da função renal neste grupo de pacientes e será discutido em detalhes na Subseção 5.3.

Com estes resultados, foi possível representar graficamente as curvas à medida que se alterava o valor de uma única variável, o que permite entender o impacto da variação dos níveis de potássio em diferentes cenários de admissão de um paciente. Foi considerado hipopotassemia para valores de potássio iguais ou inferiores a 3,5 mEq/L, e hiperpotassemia quando o nível sérico fosse igual ou superior a 5,0 mEq/L (VARGA *et al.*, 2019).

Ainda para se entender o grau de dependência entre as variáveis e a mortalidade nesse grupo de pacientes, foi realizado um estudo de informação mútua, conforme apresentado na subseção a seguir.

5.2.4 Seleção de Características

A teoria da informação é reconhecida como um conjunto de métodos que pode ser usado para a seleção de *features*, pois o poder de previsão decorre da correlação (informação mútua) entre características e rótulos (TSALATSANIS; HOZO; DJULBEGOVIC, 2020). O cálculo da entropia de uma variável e da informação mútua entre duas variáveis são ferramentas analíticas valiosas em pesquisa, pois podem ser aplicados à maioria dos tipos de dados e são independentes do modelo (ZHANG, 2017). O resultado esperado para esse estudo foi de identificar interações não lineares entre as variáveis apresentadas na Subseção 5.1 (ZBILI; RAMA, 2021).

Com esse objetivo, foi feito um estudo da entropia através de métodos não para-

métricos baseados na estimativa de entropia de distâncias de k vizinhos mais próximos (SINGH, H. *et al.*, 2003). A escolha desta técnica foi devido ao seu desempenho na estimativa multivariada de dados com alta dimensionalidade (CHARZYNSKA; GAMBIN, 2015).

O objetivo dessa análise foi entender o grau de dependência estatística entre as variáveis pré-selecionadas conforme os critérios descritos na Subseção 5.1. As métricas obtidas foram independente dos valores absolutos dos rótulos e representam, em ordem decrescente, o grau de dependência em relação à variável de mortalidade intra-hospitalar.

A observação dos resultados, em ordem decrescente, demonstra que, em relação a dados clínicos e laboratoriais, os valores de creatinina tiveram maior representatividade em relação à mortalidade, fornecendo 3,66% do ganho de informação, seguido de débito urinário, com 2,95% e *anion gap*, com 2,93%. Esse resultado é significativo porque demonstra que informações de fácil obtenção, seja por exames laboratoriais de sangue, no caso da creatinina, seja por medidas a beira do leito, como o débito urinário, foram as mais representativas nesse estudo.

A Figura 9 apresenta os valores relativos de ganho de informação calculados para as variáveis incluídas nessa análise. Valores negativos estão representados como 0, porque a informação mútua verdadeira não pode ser negativa (ZHENG *et al.*, 2006). Esse resultado foi obtido com as seguintes variáveis: glicose, hematócrito, pressão arterial diastólica e média, peso corporal, Troponina T e frequência cardíaca. O significado clínico, neste caso, é que essas informações, ainda que sejam importantes para o cuidado integral do paciente, não foram proveitosas, para esse tipo de análise de relação não linear, para indicar associação com mortalidade intra-hospitalar.

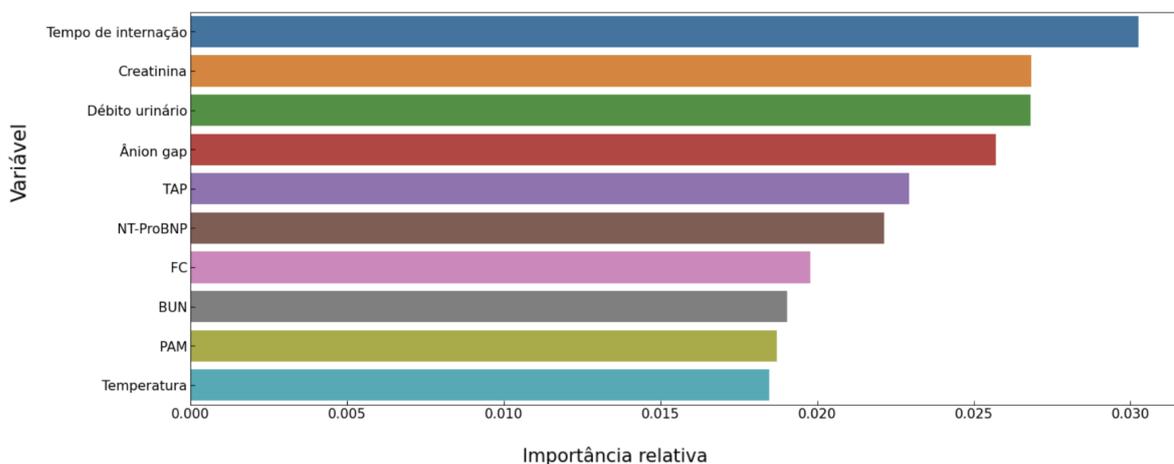


Figura 9: Importância relativa calculada para o conjunto de variáveis analisado.

O estudo da informação mútua possibilitou o entendimento de relações não lineares entre as variáveis disponíveis para a população estudada em relação à mortalidade intra-hospitalar. Com base nesse conhecimento e nos estudos realizados anteriormente, o próximo passo foi o de aplicar algoritmos de aprendizado de máquina com objetivo de predição da mortalidade desses pacientes. Os passos do estudo estão descritos na próxima subseção.

5.2.5 Aprendizado de Máquina

O entendimento do comportamento de sobrevivência dessa população, além da importância e da dependência entre as suas características, motivou a hipótese de que seria possível experimentar algoritmos de modelagem preditiva para a variável binária de interesse: óbito ou alta hospitalar. A literatura discutida na Subseção 4.4 direcionou essa etapa do estudo, em que foi comparado o desempenho de predição de mortalidade intra-hospitalar a partir de *features* extraídas da base MIMIC-IV por um algoritmo de aprendizado de máquina. A Figura 21, disponível no Apêndice F, apresenta os passos decisórios para seleção dos algoritmos de aprendizado de máquina para estudo.

Embora a maioria dos modelos avaliados apresentem AUROC próximos e que, em geral, representam bom desempenho preditivo, o fato de que RL, MVS e NB possuírem um valor semelhante pode indicar que seu desempenho pode depender mais das variáveis preditoras selecionadas, em relação ao modelo preditivo. A Tabela 2 apresenta as principais medidas de desempenho para todos os modelos avaliados nessa análise.

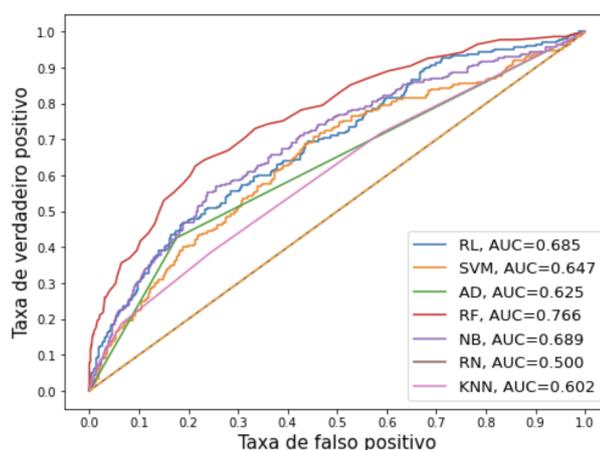


Figura 10: Comparação entre o valor médio da área sob a curva ROC dos algoritmos de aprendizado de máquina supervisionado avaliados. A reta representa um modelo sem valor preditivo.

A Figura 10 representa a métrica de desempenho preditivo dos principais modelos

Tabela 2: Medidas de desempenho dos modelos de aprendizado de máquina avaliados.

Algoritmo	Medida	Valor
Regressão Logística (RL)	Acurácia	0,78
	Precisão	0,59
	<i>F-score</i>	0,30
	AUROC	0,68
Máquina de Vetor de Suporte (SVM)	Acurácia	0,78
	Precisão	0,55
	<i>F-score</i>	0,60
	AUROC	0,64
Árvore de Decisão (AD)	Acurácia	0,72
	Precisão	0,37
	<i>F-score</i>	0,37
	AUROC	0,62
<i>Random Forest</i> (RF)	Acurácia	0,81
	Precisão	0,80
	<i>F-score</i>	0,30
	AUROC	0,75
Naive Bayes (NB)	Acurácia	0,27
	Precisão	0,23
	<i>F-score</i>	0,35
	AUROC	0,68
Rede Neural (RN)	Acurácia	0,78
	Precisão	0,60
	<i>F-score</i>	0,58
	AUROC	0,50
<i>K-Nearest Neighbors</i> (KNN)	Acurácia	0,75
	Precisão	0,36
	<i>F-score</i>	0,20
	AUROC	0,60

avaliados. A variável a ser predita (Y) foi definida como mortalidade intra-hospitalar. Foi usada a validação cruzada de 10 vezes para obter estimativas de desempenho do modelo, e a avaliação de desempenho foi feita com base na AUROC.

O algoritmo escolhido para implantação no sistema foi a RL, motivado por três fatores: (1) ser considerado um modelo explicável, no qual o exponencial de um parâmetro é a razão de chances, (2) já ter sido validado para predição de mortalidade em outras populações derivadas da base MIMIC (CHEN; GUESTRIN, 2016; PEDREGOSA *et al.*, 2012) e ser amplamente utilizada em estudos na área da saúde (WANG *et al.*, 2019), e (3) não requerer suposições sobre distribuições de classes. Por esse motivo, mesmo que a literatura indique haver superioridade de outros algoritmos para essa tarefa de classificação (JOHNSON; POLLARD; MARK, 2017), o fato de ser inteligível foi decisivo para a escolha desse

modelo.

As principais vantagens do uso da RL que aplicam a esse cenário de estudo são: (1) A facilidade de interpretação probabilística dos resultados, (2) a possibilidade de regularização do algoritmo para evitar *overfitting* e (3) a facilidade de atualização do modelo facilmente com novos dados por gradiente descendente estocástico. Por outro lado, essa escolha teve a desvantagem de que a RL tende a ter um desempenho inferior quando há limites de decisão múltiplos ou não lineares (TOLLES; MEURER, 2016). Mesmo nos casos de utilização de modelos explicáveis, como no caso da RL, foi importante decidir qual o intuito do ajuste do modelo, uma vez que o processo de modelagem difere: caso o intuito seja de inferência, as escolhas feitas durante o processo de modelagem são pautadas em medidas que avaliam a relação de explicação entre as variáveis. Por outro lado, caso o objetivo seja de predição, as escolhas no processo de modelagem são orientadas por medidas de desempenho preditivo, conforme será apresentado nos próximos parágrafos.

A Tabela 3 apresenta, a seguir, os coeficientes obtidos pelo modelo de RL e as razões de chance correspondentes.

Tabela 3: Coeficientes (*exp*) e razão de chances (OR: *Odds ratio*) calculados a partir do modelo de RL. T-Hosp: Tempo de internação hospitalar; T-CICU: Tempo de internação na CICU; IC: Intervalo de confiança; *: IC não inclui o valor 1

Variável	Coefficiente	OR (IC)
T-CICU*	0,1133	1,12 (1,09-1,16)
Anion gap*	0,0729	1,08 (1,04-1,11)
INR	0,0711	1,07 (0,83-1,38)
Temperatura corporal	0,0406	1,04 (0,93-1,17)
Idade*	0,0188	1,02 (1,01-1,03)
FR*	0,0218	1,02 (1,01-1,04)
Bicarbonato	0,0070	1,01 (0,98-1,04)
Leucometria*	0,0121	1,01 (1,00-1,02)
Potássio	0,0143	1,01 (0,91-1,13)
FC*	0,0056	1,01 (1,00-1,01)
BUN*	0,0120	1,01 (1,01-1,02)
Débito urinário em 24 horas	-0,0002	1,00 (1,00-1,00)
TAP	-0,0015	1,00 (0,97-1,02)
NT-proBNP	0,0005	1,00 (1,00-1,00)
PAS	-0,0046	1,00 (0,99-1,00)
Ck-mb	0,0014	1,00 (1,00-1,00)
Plaquetas	-0,0008	1,00 (1,00-1,00)
Peso corporal	0,0039	1,00 (1,00-1,01)
PTT	0,0027	1,00 (1,00-1,01)
Glicemia	-0,003	1,00 (1,00-1,00)
Troponina T	-0,0001	1,00 (1,00-1,00)
PAM	-0,0003	1,00 (1,00-1,00)
TAP	-0,0015	1,00 (0,97-1,02)
PAD	-0,0021	1,00 (0,99-1,00)
SpO2	-0,0074	0,99 (0,95-1,04)
Cloro	-0,0167	0,98 (0,95-1,02)
Sódio	-0,0168	0,98 (0,95-1,02)
Creatinina	-0,0177	0,98 (0,90-1,07)
T-Hosp*	-0,0416	0,96 (0,94-0,97)
Cálcio*	-0,1789	0,84 (0,72-0,97)
Hematócrito*	-0,4267	0,65 (0,55-0,77)

Como a RL é um tipo de modelo explicável, foi possível verificar que as variáveis mais

relevantes e com maiores chances de morte intra-hospitalar com significância estatística são o *anion gap* (*Odds Ratio* (OR): 1,08) e o tempo de permanência na CICU (OR: 1,12), enquanto hematócrito (OR: 0,65), cálcio (OR: 0,84) e tempo de internação hospitalar (OR: 0,96) apresentaram-se como fatores de proteção para morte intra-hospitalar. A interpretação clínica, nesse caso, é que existe uma relação inversa entre tempo de internação e mortalidade, que pode ser entendida clinicamente como: se o paciente estiver em estado crítico, é provável que a morte ocorra mais cedo; se houver uma permanência hospitalar mais prolongada ou uma permanência mais curta na CICU, o modelo espera que o paciente sobreviva. Esses resultados convergem com a curva de sobrevivência discutida na Subseção 5.2.3 e representada na Figura 8.

A acurácia média obtida no modelo de RL foi de 78.4%, e a área sob a curva ROC obtida após a validação cruzada foi de 0,685, com significância estatística (*valor - p* < 0.001) pelo Teste T pareado 5×2 . A escolha desse método de avaliação foi motivada para evitar deficiências do método do teste T pareado quando empregado para avaliação de algoritmos de classificação (DIETTERICH, 1998). Os valores de precisão, revocação e *F-score* para RL, respectivamente, de 0,77, 0,80 e 0,76. Apesar de não ter tido a maior AUROC dentre os algoritmos de aprendizado de máquina supervisionado considerados, a RL é utilizada no sistema descrito no Capítulo 6, por ser relativamente fácil de implementar, de interpretar, ser rápido e possui um relativo bom desempenho. Seu uso possibilita interpretar os parâmetros envolvidos do modelo e testar hipóteses para entender a relação entre as co-variáveis e a variável resposta, uma vez que é possível obter coeficientes e *valor - p*.

O estudo de predição foi a última etapa do estudo, e permitiu consolidar o entendimento do comportamento desse estrato de pacientes da base MIMIC-IV. A partir da descrição do perfil dos pacientes e suas características gerais, foi estudado o comportamento projetado de sobrevivência, e foram analisados os graus de contribuição de cada característica em relação ao desfecho de morte intra-hospitalar, até o estudo de modelos que pudessem prever esse desfecho. A próxima subseção apresenta e discute os principais resultados desse conjunto de estudos.

5.3 Discussão dos resultados

O estudo desta coorte de pacientes de CICU permitiu sugerir as associações entre variáveis clínicas e maior risco de morte intra-hospitalar.

Os resultados indicaram que a análise de sobrevivência feita a partir de dados de

prontuário eletrônico é possível quando os dados são tratados adequadamente, e pode destacar fatores, tanto clínicas quanto relacionadas aos indicadores de saúde, que representam maior risco de morte intra-hospitalar.

No modelo de Cox não existem suposições sobre a distribuição de probabilidade do risco. Neste caso, o risco logarítmico de um paciente foi calculado a partir de uma função linear de suas co-variáveis, acrescido de um risco basal de uma determinada população que muda com o tempo. Para o objetivo deste estudo, o modelo de Cox ajustado foi suficiente para evidenciar a importância dos níveis de potássio como fator de maior relação com óbito destes pacientes, como mostrado na Tabela 12, disponível no Apêndice B. Valores de creatinina, hematócrito e leucometria também tiveram contribuição. É importante ressaltar que variáveis classicamente associadas com maior risco de morte não estavam disponíveis nessa base como, por exemplo, informações de ecocardiograma e de classe funcional dos pacientes (JIN, 2018) e, por essa razão, não estão presentes nessa discussão.

Os distúrbios do potássio são prevalentes em pacientes críticos, e também são encontrados frequentemente associados a comorbidades cardíacas. No caso da hiperpotassemia, as causas podem ser multifatoriais, e as principais causas neste contexto são: (1) secundárias a distúrbios hormonais, diabetes mellitus ou doença renal crônica; (2) relacionadas a medicamentos (diuréticos, inibidores do sistema renina-angiotensina-aldosterona, antagonistas do receptor de aldosterona, anti-inflamatórios não esteroidais ou heparina); e (3) por ingestão excessiva de potássio, seja alimentar ou suplementar (SARWAR *et al.*, 2016). Em muitos casos, mesmo a hiperpotassemia sendo uma consequência ou uma manifestação de outro distúrbio, necessitará de intervenção médica para evitar desfechos desfavoráveis. Os achados deste artigo vão ao encontro do que foi demonstrado por Brueske *et al.* (2019) em outra coorte de pacientes de CICU.

Conforme apresentado na Tabela 12, disponível no Apêndice B, a creatinina também esteve relacionada a menor tempo de sobrevivência na população deste estudo, com uma razão de risco de 1,07, apesar de ter tido uma contribuição mais discreta que o potássio. Estudos anteriores já mostravam que pequenos aumentos na creatinina estão independentemente associados ao aumento da mortalidade (THOMAS *et al.*, 2015; LASSNIGG, 2004; LASSNIGG *et al.*, 2008). Esta alteração laboratorial é um ponto-chave das definições atuais de Lesão Renal Aguda (LRA). Já foi demonstrado que a LRA se associou independentemente a maiores taxas de incidência e progressão de doença renal crônica, bem como eventos subsequentes de insuficiência cardíaca e óbito entre os sobreviventes de uma hospitalização recente. (IKIZLER *et al.*, 2021). A avaliação da creatina neste con-

texto é complexa: ela pode representar uma perda de função renal prévia, concomitante e/ou consequente à doença cardíaca.

O aumento de hematócrito apresentou relação com risco aumentado para morte intra-hospitalar, como mostrado pela análise do coeficiente de razão de risco do modelo de regressão de Cox na Tabela 12, disponível no Apêndice B. Essa alteração laboratorial é relativamente frequente na população de CICU, e pode apresentar relação com comorbidades (como a Doença Pulmonar Obstrutiva Crônica), causa medicamentosa (hemoconcentração pelo uso de diuréticos), além de resposta inflamatória. Distúrbios das hemácias podem influenciar em maior risco de trombose arterial e venosa, além de influenciar na estabilidade de trombos já existentes (BYRNES; WOLBERG, 2017).

Na Tabela 12, disponível no Apêndice B, o aumento de leucócitos representou uma razão de risco de 1,02. A leucocitose é uma das alterações laboratoriais mais comuns na medicina e, nos pacientes estudados, contribuiu com significância estatística para maior risco de morte intra-hospitalar. Ela pode ter diversas causas, sendo as principais infecciosas, inflamatórias e desordens hematológicas (CERNY; ROSMARIN, 2012). Em relação à fisiopatologia cardiovascular, o aumento total de leucócitos é um fator de risco para aterosclerose, e se relaciona com maior incidência de doenças coronarianas e acidente vascular cerebral isquêmico. Foi demonstrado que a aterosclerose é uma doença multifatorial, com um fundo inflamatório básico (HAYBAR; PEZESHKI; SAKI, 2019). Algumas condições clínicas desses pacientes, como a desidratação por uso de diuréticos, podem causar tanto leucocitose quanto aumento de hematócrito.

Apesar do valor agregado com as variáveis de exames laboratoriais, ainda que nesta população estudada houvesse diferença estatisticamente significativa nos níveis de NT-ProBNP entre os grupos de sobrevivente e morte intra-hospitalar, como mostrado na Tabela 9, disponível no Apêndice B, esta variável não se apresentou como forte preditora pelo modelo de regressão. Já no caso da Troponina T, o valor médio entre os dois grupos foi equivalente e, da mesma forma, não houve impacto na mortalidade pelo modelo apresentado. Esses dois achados podem se dever ao fato de que uma parcela significativa dos pacientes não apresentava a medida deste exame laboratorial nas primeiras horas de internação (apenas 15% e 50% para NT-ProBNP e Troponina T, respectivamente).

Os resultados apresentados reforçaram que o prontuário eletrônico contém informações relevantes para compreender a progressão das doenças e identificar direções futuras para pesquisas. Entretanto, mesmo quando a distribuição de uma variável ocorre conforme um processo fisiológico conhecido, pode haver uma medição seletiva, resultando em

uma representação distributiva não fisiológica do que deveria ser uma variável fisiológica. Por esse motivo, a escolha do modelo estatístico para estudo das variáveis é um passo fundamental e, quando uma tecnologia de modelagem mais simples é desejada, a regressão de Cox continua sendo uma boa escolha.

Apesar do tamanho significativo de amostra e dados granulares, este estudo tem diversas limitações que são inerentes a todos os estudos de coorte retrospectivos: dados ausentes, um possível viés de calibração, e a incapacidade de ajustar todos os possíveis fatores de confusão relevantes (MARRARO; SPADA, 2021). A perda de dados, nesse contexto, pode ocorrer porque o paciente procura atendimento em outra instituição, ou devido à falta de padrões no preenchimento das informações. Além disso, os resultados de risco relativo estimados pelo modelo de RL foram relativamente baixos, quando comparados com fatores de risco usualmente calculados em estudos semelhantes (SCHNOHR, 2002; GRUNDY *et al.*, 1999). Possivelmente, caso fosse utilizada uma base com mais pacientes e com variáveis relacionadas a exames clínicos e laboratoriais cardiológicos, os resultados poderiam ser mais significativos. Houve ainda uma limitação adicional neste estudo que foi a indisponibilidade das anotações de prontuário da base MIMIC-IV. Por este motivo, informações relevantes como métricas obtidas pelo Ecodoplercardiograma transtorácico não puderam ser incluídas no estudo. Por fim, considera-se que a não utilização de outra base como teste foi também uma limitação inerente a este estudo.

Neste capítulo foram descritos os principais resultados obtidos a partir da exploração da base de pacientes MIMIC-IV. Esse banco possui dados de nível operacional sobre designações de leitos e transferências de serviços, bem como diagnósticos CID-9 e vários indicadores clínicos. O entendimento da organização dos dados brutos, do relacionamento entre eles e dos tratamentos a que os dados da base MIMIC-IV foram submetidos foram fundamentais para que o uso e a interpretação dos dados disponíveis fosse feita adequadamente.

Após a seleção de um grupo específico de pacientes desta base, foram realizadas análises estatísticas, modelagens de sobrevivência e de predição. Isso serviu como base para construção de conhecimento e também para codificação de um sistema que usa como dados de treinamento as informações da base MIMIC-IV.

No próximo capítulo será apresentado o sistema que disponibiliza dinamicamente o resultado desses modelos para apoio à tomada de decisão clínica.

6 Apresentação do Sistema de Apoio à Decisão

Este capítulo apresenta o sistema desenvolvido para processar dados clínicos de um paciente e para fornecer uma resposta de risco estimado. A Subseção 6.1 corresponde à etapa de análise do projeto, em que são descritos os objetivos gerais deste sistema de apoio à decisão, a definição de pronto e os critérios de aceite, bem como o caso de uso esperado. A Subseção 6.2 trata da implementação do sistema, na qual são descritas as etapas de concepção da arquitetura do sistema, desenvolvimento de código e hospedagem em nuvem. Em seguida, a Subseção 6.3 apresenta as etapas de avaliação do sistema, na qual são comparados os resultados obtidos pelo sistema com implementações cruas do código em um novo notebook. Por fim, a Subseção 6.4 descreve o plano de disponibilização do sistema em ambiente de produção para se tornar acessível ao usuário profissional de saúde.

6.1 Objetivo do Sistema de Apoio à Decisão

Esta subseção apresenta o objetivo pretendido com o desenvolvimento de um sistema de apoio à decisão a partir dos dados explorados da base MIMIC-IV. A Subseção 6.1.1 descreve como o sistema seria inserido na rotina do profissional de saúde que trabalha na CICU e faz uma demonstração de como os requisitos técnicos apresentam os comportamentos funcionais esperados. Na Subseção 6.1.2, são apresentados os protótipos de telas e de fluxos de uso do sistema. A Subseção 6.1.3 apresenta as principais decisões técnicas feitas para a concepção do sistema, em que se inclui a escolha da linguagem de programação, da arquitetura de código e os principais *frameworks*. Por fim, a Subseção 6.1.4 trata do conceito de adaptabilidade do sistema, que funciona inteligentemente a partir das informações disponibilizadas.

Tendo em vista o desafio de integrar o usuário não técnico (no caso, o profissional de saúde) a um sistema de apoio à tomada de decisão, existe a oportunidade de se criarem sistemas online com esta finalidade. Além disso, a possibilidade de tornar mais tangível

os resultados descritos no Capítulo 5 impulsionou essa etapa da pesquisa.

O principal objetivo do sistema proposto é funcionar como um recurso de apoio à tomada de decisão integrado na rotina dos profissionais que atuam em CICU. O modelo de predição de mortalidade intra-hospitalar ajustado a partir de dados do MIMIC-IV constituirá a base do sistema. Ao fornecer resultados em tempo real a partir do conjunto de informações fornecidas pelo usuário final, o sistema possibilitará um entendimento, na prática, de como a modelagem preditiva pode ser aplicada em situações do dia-a-dia desse profissional.

6.1.1 Análise do Sistema

O sistema foi concebido para ter como público alvo principal o médico. Entretanto, outros profissionais de saúde, bem como epidemiologistas e estatísticos podem também se beneficiar da ferramenta, principalmente para fins de pesquisa.

O sistema foi projetado para ser integrado na rotina do dia a dia do médico, como uma ferramenta de apoio à tomada de decisão. A principal motivação para utilizá-lo seria para aumentar o grau de suspeição clínica sobre a evolução de um paciente, principalmente para direcionar atenção a casos cujo prognóstico seja indicado como menos favorável. No limite, espera-se que o sistema evite com que um caso grave passe despercebido pela avaliação médica, o que pode contribuir positivamente para o cuidado desse paciente.

O profissional de saúde, diante de um paciente internado em CICU, precisa encontrar argumentos para uma tomada de decisão. Isso pode ocorrer diversas vezes em um mesmo dia, conforme as condições clínicas do paciente se modificarem. Além disso, existe uma variabilidade de informações disponíveis conforme o tempo de internação, a condição do paciente e sua capacidade de ser submetido a um determinado exame, o tempo necessário para se obter os resultados e, ainda, a disponibilidade técnica e profissional inerente a cada serviço de saúde. Isso configura uma potencial limitação das informações disponíveis, que faz parte do processo de decisão clínica e que precisa ser considerado na concepção desse sistema.

O médico intensivista pode usar o sistema da seguinte forma: no momento da elaboração da evolução médica, em que são feitos os registros em prontuário e a verificação de exames realizados, o médico detém todas as informações exigidas pelo sistema para fornecer uma análise. Dessa forma, pode ser oportuno utilizar essas novas informações para atualizar o sistema e obter, diariamente, uma nova interpretação do estado atual do

paciente.

6.1.2 Protótipo do Sistema

Para usar o sistema, conforme apresentado na Figura 16, disponível no Apêndice E, o primeiro passo é fazer a leitura das instruções de uso. Isso está indicado a partir de uma caixa de texto com instruções sobre como inserir informações no sistema. Em seguida, o usuário deve selecionar um mínimo de três características para inserir no sistema. Esse número de características pode ser incrementado ou diminuído em usos subsequentes, mas sempre respeitando o mínimo de três informações.

A cada variável selecionada a partir de uma lista de opções, um novo campo de informação estará disponível para ser editado pelo usuário. Este campo sempre irá respeitar unidades de medida e limites fisiológicos mínimos e máximos para cada variável, bem como irá exibir uma opção de ajuda indicada por um símbolo de ponto de interrogação.

Uma vez preenchidos todos os valores selecionados, o usuário deve clicar em uma caixa de aceite em relação aos termos de condições do uso do sistema. Caso estas condições sejam satisfeitas, um botão de submissão estará habilitado e, uma vez pressionado, fará com que o sistema processe as informações e exiba novas caixas de texto com as informações analisadas. Caso alguma das condições descritas acima não for atingida, é exibida uma mensagem específica informativa ao usuário. Esse cenário está apresentado na Figura 17, disponível no Apêndice E.

As respostas fornecidas pelo sistema, após a submissão dos dados pelo usuário, são organizadas em quatro blocos e são visíveis ao término do processamento, assim como uma opção para limpar as informações preenchidas e reiniciar o processo.

Conforme apresentado na Figura 18, disponível no Apêndice E, primeiro bloco de informações, “Informações Clínicas”, exibe uma predição sobre o grau de certeza que um paciente que possua as condições informadas pelo sistema irá falecer durante a sua internação hospitalar. Essa predição utiliza um modelo de RL que recebe como dados de treino as informações da coorte de pacientes utilizada nos estudos do Capítulo 5.

O segundo bloco, ainda na Figura 18, apresenta o resultado de cálculos estatísticos que embasam a interpretação clínica: (1) a acurácia do modelo preditivo escolhido, (2) o valor de perda residual obtido pela entropia cruzada, e (3) o valor do ganho de informação obtida pelo cálculo de entropia, com a indicação de qual variável, dentre as fornecidas pelo usuário, possui maior entropia e, conseqüentemente, maior influência sobre o desfecho.

A Figura 19, disponível no Apêndice E, apresenta o terceiro bloco de informações, que contém recomendações, baseadas no cálculo de informação mútua, sobre qual a importância relativa de cada variável para o objeto do estudo (no caso, a mortalidade intrahospitalar). Sendo assim, o sistema informa, a partir do conhecimento existente nas variáveis da coorte de pacientes, qual seria a variável que o usuário poderia acrescentar como uma nova informação no sistema e que mais poderia aumentar a precisão dos resultados informados. Além disso, o sistema exibe também uma segunda e uma terceira opção, para cobrir casos em que uma ou mais destas não esteja disponível pelo usuário.

Por fim, o quarto bloco, apresentado na Figura 19, oferece uma listagem das referências bibliográficas utilizadas para embasar a escolha das técnicas estatísticas e de aprendizado de máquina que suportam esse sistema.

Com base nas informações fornecidas pelo sistema, o profissional poderá decidir pela intervenção mais apropriada para cada situação: (1) para priorizar recursos hospitalares para um determinado paciente identificado pela ferramenta como tendo pior prognóstico; (2) para acompanhar com maior frequência um determinado dado clínico sinalizado pelo sistema como sendo de maior importância; (3) para incluir, em sua avaliação clínica, o acompanhamento de uma variável antes não avaliada, mas percebida pelo sistema como sendo de alta entropia e, por isso, maior associação com o desfecho; e (4) para eliminar, ou reduzir em frequência, o acompanhamento de uma determinada variável que seja menos relevante para um determinado perfil de paciente. Em todas essas situações, o objetivo comum acaba por permitir que o médico passe a dispor de mais tempo e atenção para os casos ou situações que são mais críticas, o que se relaciona principalmente com o contexto das unidades de terapia intensiva.

É importante ressaltar que, uma vez que o sistema foi treinado com base em uma coorte específica de pacientes internados em CICU, qualquer generalização dos resultados, ou transposição para uso em outros cenários, não é recomendada. Sendo assim, este sistema pode ser considerado um passo importante de pesquisa, mas que não deve ser disponível para uso pelo público em qualquer outro cenário diferente do centro hospitalar que serviu como referência.

Para que o sistema se comporte conforme descrito, são considerados requisitos gerais:

- Idioma inglês;
- Funcionamento online;
- Uso livre sem necessidade de identificação do usuário.

No que lhe concerne, dado que o usuário acessa o sistema, é esperado que, a partir dos dados de entrada (fornecidos pelo usuário), o sistema exiba os seguintes resultados:

- Rótulo (classificação) da variável desfecho considerada;
- Valor predito para a variável;
- Valor da entropia;
- Cálculo da informação, obtido pela fórmula negativa do logaritmo do valor predito;
- Identificação, dentre as variáveis inseridas no modelo, daquela com maior influência sobre o modelo;
- Sugestão, a partir de estudos prévios com as variáveis disponíveis na base de dados, de qual seria a variável com maior influência que poderia ser adicionada ao conjunto de dados de entrada;
- Sugestão de referências bibliográficas.

Como exemplo prático de caso de uso, foram utilizados dados fictícios, porém nos padrões considerados normais e aceitos pelo sistema, para que pudesse ser reproduzido. Conforme demonstrado na Figura 17, foram inseridas as seguintes informações: frequência cardíaca de 80 batimentos por minuto, pressão arterial diastólica de 150 milímetros de mercúrio, e creatinina sérica de 2,12 mg/dL. Com isso, os resultados foram calculados e apresentados conforme as Figuras 18 e Figura 19, disponíveis no Apêndice D.

A caracterização técnica e funcional do sistema foi um pré-requisito para que o seu desenvolvimento pudesse ser iniciado. Durante a elaboração do código, os cenários descritos foram reproduzidos para que se garantisse um alinhamento entre o sistema planejado e o sistema efetivamente desenvolvido. A próxima subseção detalha o projeto de implementação do sistema.

6.1.3 Projeto do Sistema

Para desenvolvimento, foi configurado um ambiente virtual local em Python versão 3.7 (VAN ROSSUM; DRAKE JR, 1995). A aplicação foi desenvolvida com base em *frameworks* de uso aberto e frequentes na linguagem Python (VAN ROSSUM; DRAKE JR, 1995). O repositório do código desenvolvido está hospedado em servidor privado na nuvem.

O *dataset* é importado na inicialização da aplicação e fica disponível em cache para execução do código. Na primeira execução do sistema, é feita uma referência ao arquivo que contém a base de dados usada como treino e teste do modelo preditivo. Esse arquivo permanece em *cache* e, a cada nova execução do sistema, o modelo preditivo é re-treinado e re-testado usando a mesma base de dados, porém somente com as variáveis indicadas pelo usuário que interagiu com o sistema.

O sistema foi desenvolvido seguindo uma arquitetura de projeto *Model View Controller* (MVC) (BURBECK, 1992). Essa escolha foi feita para aumentar a manutenibilidade e reutilização dos códigos desenvolvidos, com a separação de camadas que na redução de acoplamento e promove o aumento de coesão nas classes do projeto. A comunicação entre interfaces e regras de negócios é definida através de um controlador, e é a existência deste controlador que possibilita a separação entre as camadas. Um esquema dessa arquitetura está representado na Figura 20.

A camada de Modelo do sistema possui a lógica da aplicação, e é responsável pelo carregamento do banco de dados, cacheamento das informações e realização das operações matemáticas relacionadas à execução do modelo preditivo.

O usuário interage com a interface gráfica do sistema através da camada de Visualização. A interface gráfica repassa as informações para a camada Controladora, e esta interage com a camada de modelo, que executa as regras de negócios do sistema.

O *framework* Streamlit foi utilizado para gerar a interface gráfica interativa a partir do código desenvolvido.

A análise do código foi efetuada através da biblioteca *Pylint*. Com ela, foram realizadas buscas por erros de sintaxe, falhas de codificação e construções inapropriadas.

Uma premissa importante é que, a cada inserção de um conjunto de dados pelo usuário, o modelo deve ser treinado com base nas variáveis escolhidas. Caso um novo conjunto seja fornecido, é necessário haver novamente o treinamento do modelo, tendo em vista que a composição das variáveis pode ser alterada. Por isso, a próxima subseção detalhará a adaptabilidade do sistema.

6.1.4 Ajuste do Modelo

Para atender à condição de que a quantidade de informações sobre cada paciente é variável e potencialmente limitada, o sistema foi concebido para que conseguisse trabalhar com

essa variabilidade.

O número mínimo de variáveis para execução do modelo é três. Esse valor foi obtido a partir do método de eliminação recursiva de características, seguido do cálculo da acurácia para cada um dos casos, desde uma variável até as 31 disponíveis no conjunto de dados utilizado. A média e o desvio padrão da acurácia dos modelos com diferentes combinações de características foi comparada, o que revelou que o limiar de 0,7 ocorre no aumento de 2 para 3 e se mantém na mesma casa decimal. A partir de quatro características, o desempenho é melhor, sendo o mínimo recomendável (mas não obrigatório). Essa comparação está representada na Figura 11.

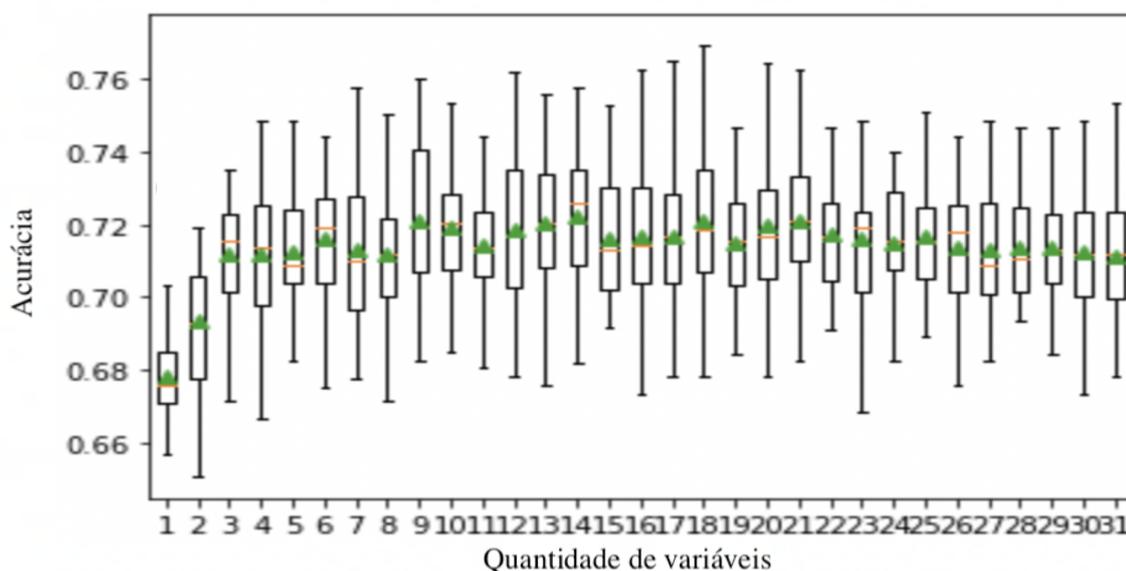


Figura 11: Acurácia (média, em verde, e desvio padrão, em laranja) do modelo de RL para diferentes números de variáveis preditoras no eixo X.

Cada vez que o sistema é utilizado com novas informações, ou seja, a cada vez que o profissional de saúde insere novos dados no sistema, todas as etapas de ajuste do modelo de aprendizado de máquina são executadas novamente, com a consideração de quais variáveis foram informadas pelo usuário. Essa etapa fornece adaptabilidade ao sistema, uma vez que são treinados n modelos de RL com todas as combinações de quantidade e variedade de *features*. Essa abordagem introduziu inteligência aos coeficientes e fatores utilizados para determinar as condições de risco, em contraste com os escores tradicionais de mortalidade, nos quais os coeficientes e fatores são definidos e estáticos, embora amplamente aceitos na literatura científica. Esse é um diferencial em relação a abordagens usuais de predição de risco que embasaram a maioria dos escores clínicos atualmente em uso (KNAUS *et al.*, 1991; VINCENT *et al.*, 1996; GALL, J., 1996; GALL, J. R. L., 1993).

6.2 Implementação do Sistema de Apoio à Decisão

Quando é feita uma submissão das informações pelo usuário, o sistema inicializa um *pipeline* de dados que contém as seguintes etapas: (1) transformação dos dados para adequação da escala dos dados, (2) seleção de características, e (3) execução do modelo preditivo com base nas características informadas.

O modelo linear para predição usado foi a RL, com base nos achados de estudos feitos com a base MIMIC (JOHNSON; POLLARD; MARK, 2017) e nos estudos conduzidos na Subseção 5.2.5. A recomendação de variáveis para o usuário do sistema considera a informação mútua, que avalia o grau de dependência entre duas variáveis (ROSS, 2014). Essa função depende da estimativa de entropia pela técnica dos k-vizinhos mais próximos.

Para avaliação do modelo, foram calculados os valores de acurácia do modelo de RL seguindo a técnica de validação cruzada em *K-folds*. Neste caso, cada observação integra um conjunto de validação exatamente uma vez e o conjunto de treinamento $k - 1$ vezes. Isso foi feito para minimizar o viés de seleção, pois a maioria dos dados são usados para treinamento, além de reduzir significativamente a variância, pois a maioria dos dados também está sendo usada no subconjunto de validação. O método de validação cruzada em *K-folds* e o valor atribuído a k foi de 10 foram escolhidos porque a literatura sugere bom desempenho desse método de avaliação em dados reais (KOHAVI, 2001; EFRON, 1983, 2000). O subconjunto de treinamento e o subconjunto de validação foi de 9/10 e 1/10, respectivamente, da base de pacientes. A escolha por essa proporção foi baseada na prática usual de modelos de aprendizado de máquina (HENDERSON; KAUWE; SPARKS, 2021; SINGH, V. *et al.*, 2021). Por fim, foram realizadas validações do modelo com diferentes quantidades e composições de variáveis. Os valores variaram entre 78 e 80% dentre as diferentes combinações possíveis.

Ao final, os resultados são apresentados na tela do sistema, seguidos de explicações em formato de texto e referências bibliográficas que suportam os conceitos utilizados para interpretação dos valores.

Após a disponibilização do sistema, o passo seguinte foi coletar opinião de profissionais de saúde que tivessem características do público-alvo desse sistema: médicos brasileiros que tivessem experiência de trabalho em CICU. A próxima subseção descreve o processo de avaliação do sistema por esses usuários.

6.3 Avaliação da Usabilidade do Sistema

A avaliação de usabilidade do sistema foi realizada em três etapas. Na Subseção 6.3.1 é apresentado o planejamento dessa etapa, em que se inclui a seleção da ferramenta de pesquisa e o fluxo de coleta das opiniões. A Subseção 6.3.2 descreve as etapas de construção do formulário e a interação com os usuários. Por fim, a Subseção 6.3.3 consolida os principais resultados obtidos a partir das interações dos profissionais com o sistema.

6.3.1 Planejamento

Durante o planejamento da avaliação do sistema, definiu-se o instrumento de coleta, o público-alvo e os passos para obtenção das respostas.

A Escala de Usabilidade do Sistema (*System Usability Scale*) foi escolhida por se tratar de uma técnica amplamente utilizada e validada para quantificar a usabilidade de produtos de software e hardware (BROOKE, 1996). Ela propõe um questionário de 10 itens (5 questões com afirmações positivas e 5 questões com afirmações negativas), no qual cada pergunta oferece como opções de resposta uma escala *Likert* que varia de “concordo totalmente” a “discordo totalmente” (MAYHEW, 1999). Essa escala é considerada equilibrada, com pontuações que variam de 0 a 100, e a pontuação média de 68 é uma útil como ponto de corte, no qual 50% dos softwares avaliados ficam abaixo e acima dele (SAURO; LEWIS, 2016).

Para seleção dos avaliadores do sistema, foi definido que o sistema seria distribuído para uso monitorado mediante o fornecimento de um *link* de acesso restrito. O perfil dos avaliadores seria de médicos que tivessem experiência de trabalho em CICU, sem outras restrições.

O processo de coleta das respostas foi planejado da seguinte forma: após uma breve introdução sobre a pesquisa através de um texto de apresentação, o *link* foi fornecido, sem nenhum detalhamento prévio sobre o funcionamento do sistema. Adicionalmente, foi dado acesso a um formulário¹ que continha, além da escala SUS, as seguintes seções: (1) uma guia de explicação geral sobre a pesquisa e sobre o formulário, com necessidade de aceite para um termos de aceite em relação ao compartilhamento das respostas; (2) perguntas gerais de caracterização dos usuários: idade, tempo de prática médica e especialidade médica; (3) as dez perguntas referentes ao SUS; e (4) duas perguntas de resposta aberta em relação a qual a característica que mais gostou no sistema, e se haveriam sugestões ou

¹Disponível em: <https://forms.gle/Qd5S5Zz5ciS6xBnb6>. Acesso em: 31 de dezembro de 2022.

críticas a serem feitas em relação ao sistema. Em razão do sistema ter sido desenvolvido em inglês, foi feita a opção pelo questionário nesse mesmo idioma. O Apêndice D apresenta o formulário utilizado.

6.3.2 Execução

Uma vez definido o desenho da avaliação, a execução consistiu na coleta das respostas. Foram feitas comunicações através de mensagem de texto para dez potenciais participantes da avaliação. A escolha desse número foi fundamentada nos seguintes critérios: facilitar o cálculo de porcentagens e capacidade de prova do conceito em relação à viabilidade de execução dessa avaliação, com possibilidade de ampliação desse número em rodadas de pesquisa futuras. Todas as respostas foram obtidas sem identificação dos participantes e registradas em uma planilha. Os questionários foram enviados para 10 pessoas, que tiveram o prazo de duas semanas para responder às perguntas.

6.3.3 Resultados da Avaliação

A maioria dos usuários tinha como perfil demográfico a idade entre 30 e 40 anos, com cerca de 10 anos de experiência médica, sendo a especialidade Medicina Interna a mais frequente (quatro respostas).

O escore médio calculado foi de 80. Quando convertido a escala de percentil, o resultado foi próximo de 85. Esse valor sugere que o sistema obteve uma classificação objetiva de excelência para esse pequeno grupo de avaliadores.

Em relação aos comentários abertos obtidos na quarta parte do questionário, não houve um padrão definido de respostas. Todavia, a descrição que pareceu mais interessar aos usuários foi a do sistema fornecer recomendações ou sugestões. Já em relação a críticas e sugestões, houve uma necessidade reportada por todos os usuários de simplificação das explicações e dos resultados apresentados. Além disso, metade dos usuários que responderam manifestaram interesse por algum tipo de integração entre o sistema e o PEP. As respostas livres estão compiladas em grupos de assunto descritos na Tabela 4 e demonstram um predomínio de comentários relacionados a melhorias propostas ao sistema, quando comparado a observações sobre disfunções ou erros apresentados.

O comentário apresentado na Tabela 4 em relação a maior grau de precisão nos resultados é considerado um item crítico em termos de *feedback*. Por um lado, foi possível perceber que existe um interesse para que o sistema forneça mais detalhes de previsão.

Tabela 4: Respostas em campo aberto fornecidas pelos avaliadores do sistema.

Classe de comentário	Frequência de solicitação
Simplificar as explicações estatísticas e orientá-las para o público médico	100%
Fornecer uma estimativa mais precisa e com intervalo de confiança	100%
Possibilitar cópia dos resultados para o prontuário eletrônico	50%
Incluir mais interpretações clínicas dos resultados	30%
Calcular risco previsto para parada cardiorrespiratória	10%
Fornecer opção para outras unidades laboratoriais	10%

Por outro, deve-se ponderar que essa informação é resultado de um modelo e que, isoladamente, não deve guiar uma decisão clínica. A Figura 12 consolida os comentários em classes, e reforça que, apesar dos avaliadores utilizarem o sistema e proporem mais funcionalidades, as informações fornecidas por ele ainda são consideradas complexas em termos de entendimento de uso.

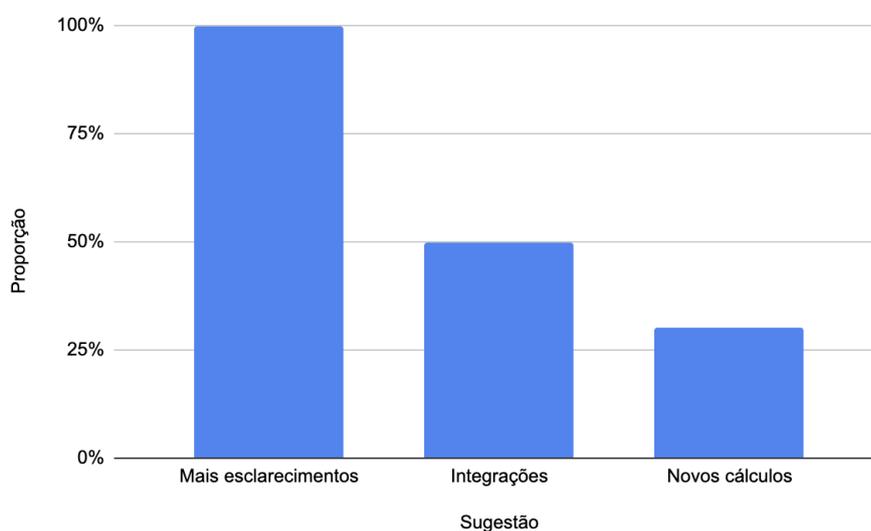


Figura 12: Classes de comentários.

Após a realização da avaliação do sistema, o passo final foi publicá-lo, de forma que pudesse ser acessado online por outros profissionais de saúde. Os passos para essa distribuição estão descritos na próxima subseção.

6.4 Implantação do Sistema

O sistema está disponível para uso [neste link](#)². Para distribuição na nuvem foi utilizado o Streamlit, um *framework* desenvolvido em Python (VAN ROSSUM; DRAKE JR, 1995), que permite a criação de aplicações em nuvem a partir de modelos de aprendizado de máquina.

²Disponível em: https://share.streamlit.io/pedrogemal/mimicivsandbox/main/model/prediction_full.py. Acesso em: 31 de dezembro de 2022.

Para favorecer a reprodutibilidade do estudo, foram construídos *notebooks* colaborativos que permitem que a documentação e o código sejam combinados em uma análise reproduzível (KLUYVER *et al.*, 2016). Em razão do inglês ser o idioma original da base MIMIC, além de ser também a escolha da maioria das publicações relacionadas, foi optado pelo desenvolvimento da interface nesse idioma. Além disso, no sistema de controle de versão *Git* foram criados arquivos complementares: (1) um arquivo de ajuda (*readme*), representado na Figura 13, disponível no Apêndice D, o qual provê uma introdução ao pacote de arquivos; (2) um arquivo de licença, que apresenta os termos de uso, modificação e redistribuição dos arquivos do repositório; e (3) um arquivo de documentação técnica do sistema, representado na Figura 15, disponível no Apêndice D.

Neste capítulo foram apresentados os detalhes técnicos do desenvolvimento, da avaliação e da publicação da aplicação desenvolvida. Foram descritas as premissas para sua elaboração, a concepção da arquitetura do software, o desenvolvimento dos códigos e a disponibilização do sistema para os usuários.

Esta etapa da pesquisa foi importante para apoiar a hipótese de que os modelos de aprendizado de máquina podem servir de auxílio à tomada de decisão clínica.

O sistema desenvolvido foi uma forma de apresentar ao usuário profissional de saúde os cálculos produzidos pelo modelo detalhado no Capítulo 5. A adição de uma interpretação clínica que adota uma linguagem mais próxima da área médica teve como intenção reduzir a barreira de adoção desta nova tecnologia na rotina de trabalho desses profissionais.

A apresentação de referências bibliográficas, em conjunto com os resultados matemáticos, foi feita para permitir que o próprio profissional pudesse verificar e validar as conclusões propostas pelo sistema. Além disso, reforça a ideia de que o sistema não se propõe a substituir a função do profissional de saúde, mas sim de apoiar a sua tomada de decisão, conforme propõe a hipótese desta dissertação.

No próximo capítulo serão apresentadas as principais contribuições e conclusões desta dissertação.

7 Conclusão

O estudo de uma coorte norte-americana de pacientes de CICU permitiu sugerir as associações entre variáveis clínicas e maior risco de morte intra-hospitalar. Os resultados indicaram que o aprendizado de máquina, combinada com dados de prontuário eletrônico, pode ser útil para orientar estudos com objetivos de inferência e de predição. Esse cenário é viável se os dados forem tratados adequadamente, e pode conseguir destacar fatores, tanto clínicos quanto relacionados aos indicadores de saúde, que representem maior risco de morte intra-hospitalar.

No grupo de pacientes analisado, os idosos representam um número significativo de internações em CICU. Isso pode se associar a um perfil diferente de comorbidades que se associem a maior risco de morte, e que podem ser exploradas em estudos futuros. A análise de sobrevivência de outros grupos de pacientes da base MIMIC-IV também é uma oportunidade de pesquisa que pode ser mais bem explorada. Novos estudos poderão também determinar modelos ideais com um conjunto maior de variáveis, que incluam exames laboratoriais, comorbidades, medicamentos e outras características potencialmente relevantes do paciente.

Apesar do acesso livre ao MIMIC-IV, ainda é difícil para os pesquisadores clínicos obterem uma visão geral dos dados e, posteriormente, aprofundarem seus questionamentos e suas análises. Esse cenário pode se modificar caso possuam uma compreensão de linguagens de programação e arquiteturas de banco de dados, além da disponibilidade de acesso a recursos computacionais.

As particularidades técnicas encontradas ao longo desta pesquisa revelam que devem ser traçadas linhas claras de responsabilidade para o acesso aos dados. Muitas vezes, é esperado, em pesquisas clínicas, que os dados sejam acessíveis para esclarecer a extensão das informações disponíveis para qualquer participante do grupo de pesquisa. Entretanto, em razão das restrições de acesso exigidas pelos mantenedores do MIMIC, essa situação não é possível, o que evidencia que existe uma desconexão entre a coleta de dados e

a pesquisa médica. Essa experiência reforça que uma etapa limitante nessas pesquisas corresponde à capacidade de reunir os dados em coortes e poder analisar janelas de tempo mais longas e com grandes volumes de dados de pacientes. Apesar de serem atividades relativamente comuns para profissionais capacitados em análise de dados, elas dependem de conhecimentos clínicos detidos pelos profissionais de saúde.

Em relação aos objetivos propostos para esta pesquisa, considera-se que o objetivo principal foi atingido, porque o protótipo descrito no Capítulo 6, uma vez embasado por pesquisas adicionais sobre a generalização de indicadores de risco conhecidos entre pacientes críticos internados em UTI, poderá servir como apoio à tomada de decisão dos profissionais de CICU. Já em relação aos objetivos secundários, considera-se que foram parcialmente atingidos, porque, ainda que tenha havido a proposição de diferentes algoritmos de aprendizado de máquina, esse estudo foi estritamente teórico e não envolveu interações em cenários reais com profissionais de saúde.

7.1 Contribuições

Uma das principais contribuições desta pesquisa foi a descrição de um processo de uso de uma base aberta de saúde para estudos com aprendizado de máquina, com intenção de ser reproduzível em outras bases de dados. Esse processo pode ser caracterizado por três estágios: (1) a descrição detalhada do processamento de dados de saúde, com objetivo de permitir a reprodução desta e outras pesquisas na base MIMIC-IV e mitigar dúvidas em relação a conceitos clínicos aplicados a uma base de dados, (2) o estudo de modelos de aprendizado de máquina de fácil entendimento e aplicabilidade em dados de PEP, e (3) a identificação das variáveis mais sensíveis para a predição de mortalidade, o que, no caso apresentado, correspondeu aos níveis de potássio e, mais amplamente, da função renal neste grupo de pacientes, com objetivo de identificar precocemente indícios de desfecho desfavorável. Destaca-se, portanto, o protótipo desenvolvido, que pode servir como guia para estudos futuros.

O desenvolvimento desta pesquisa gerou as seguintes contribuições secundárias:

1. Exploração da base MIMIC-IV. Por ser uma base ainda com relativamente poucas publicações, a criação de um repositório público no *GitHub* com os códigos desenvolvidos possibilita a reprodutibilidade da pesquisa desenvolvida, e também permite a extensão para novos estudos. Parte dos códigos também foi incorporada ao repositório oficial do *Massachusetts Institute of Technology*, como contribuição ao projeto

MIMIC.

2. Disponibilização do código, dos resultados e do sistema de decisão clínica, que podem ser usados como uma ferramenta de apoio à tomada de decisão em unidades de terapia intensiva cardíaca vinculadas ou não ao Sistema Único de Saúde (SUS).
3. Aplicação de modelo de sobrevivência paramétrico usando base de treinamento de alta dimensionalidade. Essa metodologia pode ser replicado em outros cenários para fins de estudo, incluindo hospitais vinculados ao SUS.
4. Levantamento de premissas para trabalhos futuros que transfiram o conhecimento, dado que o modelo já é usado com sucesso em outras infraestruturas de saúde, com objetivo ser replicado no SUS.

7.2 Trabalhos Futuros

Este trabalho propôs um modelo de apoio à decisão capaz de exibir o risco do paciente admitido no hospital baseado em um conjunto de dados clínicos observados. Para validar externamente o modelo proposto, é necessário aplicá-lo em outros conjuntos de dados, sejam da mesma instituição, ou relacionados a outros setores do hospital, ou ainda pertencentes a outros centros hospitalares. É possível também utilizar, como dados de entrada para o modelo, informações de outros perfis de pacientes, que não os cardiológicos, e comparar os desempenhos por diferentes medidas. Dessa forma, o sistema desenvolvido nesta pesquisa poderá servir como precursor de um novo escore clínico, que poderá apoiar a decisão clínica, assim como foi detalhado na Subseção 4.3. A natureza dinâmica dos dados de PEP representa um potencial para atualizar dados de treinamento de modelos conforme os dados clínicos mais recentes se tornam disponíveis. Teoricamente, isso possibilita a criação de sistemas de pontuação igualmente dinâmico, capazes de gerar previsões mais precisas. Dessa forma, essas ferramentas poderão conseguir aproveitar dados atuais, que provavelmente são mais representativos, além de incluir dados antigos, que podem ser menos relevantes, mas que agregam maior poder estatístico.

Com a intenção de aplicar os conhecimentos obtidos neste estudo no SUS, um trabalho futuro possível seria o de comparar diferentes medidas de desempenho do modelo de RL a partir de novas variáveis de entrada. Essa abordagem pode permitir que, ao analisar os coeficientes criticamente, seja possível criar diferentes cenários de estudo de sensibilidade (por exemplo, em situações em que nem todos os atributos estejam disponíveis, quais

as melhores combinações de variáveis, dentre aquelas possíveis). Esse estudo pode ser relevante caso se considere a heterogeneidade de disponibilidade de serviços e recursos no Brasil. Outra possibilidade a ser estudada é a de compreender melhor os dados mais sensíveis ou críticos em relação à morte intra-hospitalar. Nesse sentido, é possível tentar identificar, dados os valores de entrada para o modelo, qual seria outro valor de variável que iria aumentar ou reduzir a probabilidade de sobrevida de um paciente.

O estudo de diferentes algoritmos de aprendizado de máquina, notadamente os supervisionados, pode trazer informações adicionais em termos de aplicabilidade prática nas pesquisas médicas. Como as configurações de parâmetro padrão foram usadas para os algoritmos avaliados, outra próxima etapa possível seria a de investigar como que a alteração dos hiperparâmetros afeta o desempenho preditivo.

Todos esses estudos futuros, uma vez que busquem aplicabilidade na prática clínica, podem contribuir favoravelmente na melhor gestão de recursos de saúde, com uso mais eficiente e maior atenção aos casos de pacientes que apresentem maior complexidade. Uma vantagem de se usar apenas dados clínicos e demográficos é que eles estão mais rotineiramente disponíveis e, por isso, modelos baseados neles podem ser usados mais amplamente.

O sistema desenvolvido neste trabalho pode ser integrado a prontuários eletrônicos do SUS, para funcionar como um sistema de apoio à decisão pela estratificação de risco. Nesse caso, os pacientes que apresentarem um risco elevado, dadas as observações clínicas, podem receber um tratamento diferenciado. Futuramente, será possível associar os resultados da estratificação de risco a sugestões de conduta clínica na infraestrutura de saúde, com objetivo de melhorar os cuidados de saúde.

REFERÊNCIAS

- ABDI, Hervé; WILLIAMS, Lynne J. Principal component analysis. **Wiley Interdisciplinary Reviews: Computational Statistics**, Wiley, v. 2, n. 4, p. 433–459, jun. 2010. DOI: [10.1002/wics.101](https://doi.org/10.1002/wics.101). Disponível em: [10.1002/wics.101](https://doi.org/10.1002/wics.101).
- AGRESTI, Alan; HITCHCOCK, David B. Bayesian inference for categorical data analysis. **Statistical Methods and Applications**, Springer Science e Business Media LLC, v. 14, n. 3, p. 297–330, dez. 2005. DOI: [10.1007/s10260-005-0121-y](https://doi.org/10.1007/s10260-005-0121-y). Disponível em: [10.1007/s10260-005-0121-y](https://doi.org/10.1007/s10260-005-0121-y).
- AKOBENG, A K. Principles of evidence based medicine. **Archives of Disease in Childhood**, BMJ, v. 90, n. 8, p. 837–840, ago. 2005. DOI: [10.1136/adc.2005.071761](https://doi.org/10.1136/adc.2005.071761).
- ALJUAID, Tahani; SASI, Sreela. Proper imputation techniques for missing values in data sets. In: 2016 International Conference on Data Science and Engineering (ICDSE). [S. l.: s. n.], 2016. p. 1–5. DOI: [10.1109/ICDSE.2016.7823957](https://doi.org/10.1109/ICDSE.2016.7823957).
- ALLEN, Norrina B; KHAN, Sadiya S. Blood pressure trajectories across the life course. en. **Am. J. Hypertens.**, Oxford University Press (OUP), v. 34, n. 3, p. 234–241, abr. 2021.
- AÑEL RODRÍGUEZ, Rosa María *et al.* Historia clínica y receta electrónica: riesgos y beneficios detectados desde su implantación. Diseño, despliegue y usos seguros. **Atención Primaria**, Elsevier BV, v. 53, p. 102220, dez. 2021. Seguridad del Paciente en Atención Primaria. ISSN 0212-6567. DOI: <https://doi.org/10.1016/j.aprim.2021.102220>.
- ATALLA, Eleftheria *et al.* Readmissions among patients with COVID-19. **International Journal of Clinical Practice**, Hindawi Limited, v. 75, n. 3, out. 2020. DOI: [10.1111/ijcp.13700](https://doi.org/10.1111/ijcp.13700). Disponível em: [10.1111/ijcp.13700](https://doi.org/10.1111/ijcp.13700).
- BAE, SungA *et al.* Impact of cardiovascular disease and risk factors on fatal outcomes in patients with COVID-19 according to age: a systematic review and meta-analysis. **Heart**, BMJ, v. 107, n. 5, p. 373–380, dez. 2020. DOI:

[10.1136/heartjnl-2020-317901](https://doi.org/10.1136/heartjnl-2020-317901). Disponível em:

<<https://doi.org/10.1136/heartjnl-2020-317901>>.

BAKER, Stephanie; XIANG, Wei; ATKINSON, Ian. Continuous and automatic mortality risk prediction using vital signs in the intensive care unit: a hybrid neural network approach. **Scientific Reports**, Springer Science e Business Media LLC, v. 10, n. 1, dez. 2020. DOI: [10.1038/s41598-020-78184-7](https://doi.org/10.1038/s41598-020-78184-7). Disponível em:

<<https://doi.org/10.1038/s41598-020-78184-7>>.

BALKAN, Baran; ESSAY, Patrick; SUBBIAN, Vignesh. Evaluating ICU Clinical Severity Scoring Systems and Machine Learning Applications: APACHE IV/IVa Case Study. In: 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). [S. l.]: IEEE, jul. 2018. DOI:

[10.1109/embc.2018.8513324](https://doi.org/10.1109/embc.2018.8513324). Disponível em:

<<https://doi.org/10.1109/embc.2018.8513324>>.

BARAKAT, Ahmed *et al.* Understanding survival analysis: actuarial life tables and the Kaplan–Meier plot. **British Journal of Hospital Medicine**, Mark Allen Group, v. 80, n. 11, p. 642–646, nov. 2019. DOI: [10.12968/hmed.2019.80.11.642](https://doi.org/10.12968/hmed.2019.80.11.642). Disponível em: <<https://doi.org/10.12968/hmed.2019.80.11.642>>.

BASKORO, Sangaji; SUNINDYO, Wikan Danar. Predicting Issue Handling Process using Case Attributes and Categorical Variable Encoding Techniques. In: 2019 International Conference on Data and Software Engineering (ICoDSE). [S. l.]: IEEE, nov. 2019. DOI: [10.1109/icodse48700.2019.9092617](https://doi.org/10.1109/icodse48700.2019.9092617). Disponível em:

<<https://doi.org/10.1109/icodse48700.2019.9092617>>.

BEWICK, Viv; CHEEK, Liz; BALL, Jonathan. **Critical Care**, Springer Science e Business Media LLC, v. 8, n. 5, p. 389, 2004. DOI: [10.1186/cc2955](https://doi.org/10.1186/cc2955). Disponível em:

<<https://doi.org/10.1186/cc2955>>.

BLESSING, Lucienne T. M.; CHAKRABARTI, Amaresh. **DRM, a Design Research Methodology**. 1st. [S. l.]: Springer Publishing Company, Incorporated, 2009. ISBN 1848825862.

BLEUMINK, G *et al.* **European Heart Journal**, Oxford University Press (OUP), v. 25, n. 18, p. 1614–1619, set. 2004. DOI: [10.1016/j.ehj.2004.06.038](https://doi.org/10.1016/j.ehj.2004.06.038). Disponível em: <<https://doi.org/10.1016/j.ehj.2004.06.038>>.

- BOHULA, Erin A. *et al.* Demographics, Care Patterns, and Outcomes of Patients Admitted to Cardiac Intensive Care Units. **JAMA Cardiology**, American Medical Association (AMA), v. 4, n. 9, p. 928, set. 2019. DOI: [10.1001/jamacardio.2019.2467](https://doi.org/10.1001/jamacardio.2019.2467). Disponível em: <https://doi.org/10.1001/jamacardio.2019.2467>.
- BOROUJENI, Amir Momeni; YOUSEFI, Elham; ZURETTI, Alejandro. Time-Series Analysis of Laboratory Values in the Context of Long-Term Hospitalized Patient Mortality. **American Journal of Clinical Pathology**, Oxford University Press (OUP), v. 151, n. 5, p. 452–460, jan. 2019. DOI: [10.1093/ajcp/aqy163](https://doi.org/10.1093/ajcp/aqy163).
- BOX, Joan Fisher. Guinness, Gosset, Fisher, and Small Samples. **Statistical Science**, Institute of Mathematical Statistics, v. 2, n. 1, fev. 1987. DOI: [10.1214/ss/1177013437](https://doi.org/10.1214/ss/1177013437). Disponível em: <https://doi.org/10.1214/ss/1177013437>.
- BRAUNWALD, Eugene. Diabetes, heart failure, and renal dysfunction: The vicious circles. en. **Prog. Cardiovasc. Dis.**, Elsevier BV, v. 62, n. 4, p. 298–302, jul. 2019.
- BROOKE, John. SUS: A Quick and Dirty Usability Scale. Taylor e Francis, p. 207–212, jun. 1996. DOI: [10.1201/9781498710411-35](https://doi.org/10.1201/9781498710411-35). Disponível em: <https://doi.org/10.1201/9781498710411-35>.
- BROWER, Roy G. *et al.* Ventilation with Lower Tidal Volumes as Compared with Traditional Tidal Volumes for Acute Lung Injury and the Acute Respiratory Distress Syndrome. **New England Journal of Medicine**, Massachusetts Medical Society, v. 342, n. 18, p. 1301–1308, mai. 2000. DOI: [10.1056/nejm200005043421801](https://doi.org/10.1056/nejm200005043421801). Disponível em: <https://doi.org/10.1056/nejm200005043421801>.
- BRUESKE, Benjamin *et al.* Hyperkalemia Is Associated With Increased Mortality Among Unselected Cardiac Intensive Care Unit Patients. **Journal of the American Heart Association**, Ovid Technologies (Wolters Kluwer Health), v. 8, n. 7, abr. 2019. DOI: [10.1161/jaha.118.011814](https://doi.org/10.1161/jaha.118.011814). Disponível em: <https://doi.org/10.1161/jaha.118.011814>.
- BURBECK, Steve. Applications programming in smalltalk-80: how to use model-view-controller (mvc), jan. 1992.
- BYRNES, James R.; WOLBERG, Alisa S. Red blood cells in thrombosis. **Blood**, American Society of Hematology, v. 130, n. 16, p. 1795–1799, out. 2017. DOI: [10.1182/blood-2017-03-745349](https://doi.org/10.1182/blood-2017-03-745349). Disponível em: <https://doi.org/10.1182/blood-2017-03-745349>.

CELI, Leo Anthony; GALVIN, Sean *et al.* A Database-driven Decision Support System: Customized Mortality Prediction. **Journal of Personalized Medicine**, MDPI AG, v. 2, n. 4, p. 138–148, set. 2012. DOI: [10.3390/jpm2040138](https://doi.org/10.3390/jpm2040138).

CELI, Leo Anthony; MARK, Roger G. *et al.* Big Data in the Intensive Care Unit. Closing the Data Loop. **American Journal of Respiratory and Critical Care Medicine**, American Thoracic Society, v. 187, n. 11, p. 1157–1160, jun. 2013. DOI: [10.1164/rccm.201212-2311ed](https://doi.org/10.1164/rccm.201212-2311ed). Disponível em: <https://doi.org/10.1164/rccm.201212-2311ed>.

CELI, Leo Anthony; ZIMOLZAK, Andrew J; STONE, David J. Dynamic Clinical Data Mining: Search Engine-Based Decision Support. **JMIR Medical Informatics**, JMIR Publications Inc., v. 2, n. 1, e13, jun. 2014. DOI: [10.2196/medinform.3110](https://doi.org/10.2196/medinform.3110). Disponível em: <https://doi.org/10.2196/medinform.3110>.

CERNY, Jan; ROSMARIN, Alan G. Why Does My Patient Have Leukocytosis? **Hematology/Oncology Clinics of North America**, Elsevier BV, v. 26, n. 2, p. 303–319, abr. 2012. DOI: [10.1016/j.hoc.2012.01.001](https://doi.org/10.1016/j.hoc.2012.01.001). Disponível em: <https://doi.org/10.1016/j.hoc.2012.01.001>.

CHARZYNSKA, Agata; GAMBIN, Anna. Improvement of the k-nn entropy estimator with applications in systems biology. **Entropy (Basel)**, MDPI AG, v. 18, n. 1, p. 13, dez. 2015.

CHEN, Tianqi; GUESTRIN, Carlos. XGBoost. In: PROCEEDINGS of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. [S. l.]: ACM, ago. 2016. DOI: [10.1145/2939672.2939785](https://doi.org/10.1145/2939672.2939785). Disponível em: <https://doi.org/10.1145/2939672.2939785>.

CHICHAREON, Ply *et al.* Predicting mortality in cardiac care unit patients: external validation of the Mayo cardiac intensive care unit admission risk score. **European Heart Journal. Acute Cardiovascular Care**, Oxford University Press (OUP), v. 10, n. 9, p. 1065–1073, ago. 2021. DOI: [10.1093/ehjacc/zuab070](https://doi.org/10.1093/ehjacc/zuab070). Disponível em: <https://doi.org/10.1093/ehjacc/zuab070>.

CHURPEK, Matthew M. *et al.* Multicenter Comparison of Machine Learning Methods and Conventional Regression for Predicting Clinical Deterioration on the Wards. **Critical Care Medicine**, Ovid Technologies (Wolters Kluwer Health), v. 44, n. 2, p. 368–374, fev. 2016. DOI: [10.1097/ccm.0000000000001571](https://doi.org/10.1097/ccm.0000000000001571).

- CLARK, T G *et al.* Survival Analysis Part I: Basic concepts and first analyses. **British Journal of Cancer**, Springer Science e Business Media LLC, v. 89, n. 2, p. 232–238, jul. 2003. DOI: [10.1038/sj.bjc.6601118](https://doi.org/10.1038/sj.bjc.6601118). Disponível em: [10.1038/sj.bjc.6601118](https://doi.org/10.1038/sj.bjc.6601118).
- CLELAND, John G.F.; MCGOWAN, James. Heart Failure due to Ischaemic Heart Disease: Epidemiology, Pathophysiology and Progression. **Journal of Cardiovascular Pharmacology**, Ovid Technologies (Wolters Kluwer Health), v. 33, s17–s29, jun. 1999. DOI: [10.1097/00005344-199906003-00003](https://doi.org/10.1097/00005344-199906003-00003). Disponível em: [10.1097/00005344-199906003-00003](https://doi.org/10.1097/00005344-199906003-00003).
- COLE, T J. Too many digits: the presentation of numerical data. **Archives of Disease in Childhood**, BMJ, v. 100, n. 7, p. 608–609, abr. 2015. DOI: [10.1136/archdischild-2014-307149](https://doi.org/10.1136/archdischild-2014-307149). Disponível em: [10.1136/archdischild-2014-307149](https://doi.org/10.1136/archdischild-2014-307149).
- CORTES, Corinna; VAPNIK, Vladimir. Support-vector networks. **Machine learning**, Springer, v. 20, n. 3, p. 273–297, 1995.
- COSTA SANTOS, Cristina Mamédio da; MATTOS PIMENTA, Cibele Andrucio de; NOBRE, Moacyr Roberto Cuce. The PICO strategy for the research question construction and evidence search. **Revista Latino-Americana de Enfermagem**, FapUNIFESP (SciELO), v. 15, n. 3, p. 508–511, jun. 2007. DOI: [10.1590/s0104-11692007000300023](https://doi.org/10.1590/s0104-11692007000300023).
- COVER, T.; HART, P. Nearest neighbor pattern classification. **IEEE Transactions on Information Theory**, Institute of Electrical e Electronics Engineers (IEEE), v. 13, n. 1, p. 21–27, jan. 1967. DOI: [10.1109/tit.1967.1053964](https://doi.org/10.1109/tit.1967.1053964). Disponível em: [10.1109/tit.1967.1053964](https://doi.org/10.1109/tit.1967.1053964).
- COX, D. R. Regression Models and Life-Tables. **Journal of the Royal Statistical Society: Series B (Methodological)**, Wiley, v. 34, n. 2, p. 187–202, jan. 1972. DOI: [10.1111/j.2517-6161.1972.tb00899.x](https://doi.org/10.1111/j.2517-6161.1972.tb00899.x). Disponível em: [10.1111/j.2517-6161.1972.tb00899.x](https://doi.org/10.1111/j.2517-6161.1972.tb00899.x).
- COX, David R. The regression analysis of binary sequences. **Journal of the Royal Statistical Society: Series B (Methodological)**, Wiley Online Library, v. 20, n. 2, p. 215–232, 1958.

- DAI, Zheng *et al.* Analysis of adult disease characteristics and mortality on MIMIC-III. Edição: Omid Beiki. **PLOS ONE**, Public Library of Science (PLoS), v. 15, n. 4, e0232176, abr. 2020. DOI: [10.1371/journal.pone.0232176](https://doi.org/10.1371/journal.pone.0232176). Disponível em: [10.1371/journal.pone.0232176](https://doi.org/10.1371/journal.pone.0232176).
- DAMASKOS, Christos *et al.* Assessing Cardiovascular Risk in Patients with Diabetes: An Update. **Current Cardiology Reviews**, Bentham Science Publishers Ltd., v. 16, n. 4, p. 266–274, jul. 2021. DOI: [10.2174/1573403x15666191111123622](https://doi.org/10.2174/1573403x15666191111123622). Disponível em: [10.2174/1573403x15666191111123622](https://doi.org/10.2174/1573403x15666191111123622).
- DIETTERICH, Thomas G. Approximate Statistical Tests for Comparing Supervised Classification Learning Algorithms. **Neural Computation**, MIT Press - Journals, v. 10, n. 7, p. 1895–1923, out. 1998. DOI: [10.1162/089976698300017197](https://doi.org/10.1162/089976698300017197). Disponível em: [10.1162/089976698300017197](https://doi.org/10.1162/089976698300017197).
- DONG, Yiran; PENG, Chao-Ying Joanne. Principled missing data methods for researchers. **SpringerPlus**, Springer Science e Business Media LLC, v. 2, n. 1, mai. 2013. DOI: [10.1186/2193-1801-2-222](https://doi.org/10.1186/2193-1801-2-222). Disponível em: [10.1186/2193-1801-2-222](https://doi.org/10.1186/2193-1801-2-222).
- DONNELLY, John P. *et al.* Readmission and Death After Initial Hospital Discharge Among Patients With COVID-19 in a Large Multihospital System. **JAMA**, American Medical Association (AMA), v. 325, n. 3, p. 304, jan. 2021. DOI: [10.1001/jama.2020.21465](https://doi.org/10.1001/jama.2020.21465). Disponível em: [10.1001/jama.2020.21465](https://doi.org/10.1001/jama.2020.21465).
- DYBA, Tore; DINGSOYR, Torgeir; HANSSEN, Geir K. Applying Systematic Reviews to Diverse Study Types: An Experience Report. In: FIRST International Symposium on Empirical Software Engineering and Measurement (ESEM 2007). [S. l.: s. n.], 2007. p. 225–234. DOI: [10.1109/ESEM.2007.59](https://doi.org/10.1109/ESEM.2007.59).
- EFRON, Bradley. Estimating the Error Rate of a Prediction Rule: Improvement on Cross-Validation. **Journal of the American Statistical Association**, Taylor e Francis, v. 78, n. 382, p. 316–331, 1983. DOI: [10.1080/01621459.1983.10477973](https://doi.org/10.1080/01621459.1983.10477973).
- _____. The Bootstrap and Modern Statistics. **Journal of the American Statistical Association**, Taylor e Francis, v. 95, n. 452, p. 1293–1296, 2000. DOI: [10.1080/01621459.2000.10474333](https://doi.org/10.1080/01621459.2000.10474333).

- EFRON, Bradley. The Efficiency of Cox's Likelihood Function for Censored Data. **Journal of the American Statistical Association**, Informa UK Limited, v. 72, n. 359, p. 557–565, set. 1977. DOI: [10.1080/01621459.1977.10480613](https://doi.org/10.1080/01621459.1977.10480613). Disponível em: <https://doi.org/10.1080/01621459.1977.10480613>.
- EMMANUEL, Anton. Assisting clinical decision making: tools, rules and teams. **Clinical Medicine**, Royal College of Physicians, v. 21, n. 2, p. 81–81, mar. 2021. DOI: [10.7861/clinmed.ed.21.2.1](https://doi.org/10.7861/clinmed.ed.21.2.1). Disponível em: <https://doi.org/10.7861/clinmed.ed.21.2.1>.
- EVANS, R. S. Electronic Health Records: Then, Now, and in the Future. **Yearbook of Medical Informatics**, Georg Thieme Verlag KG, v. 25, S 01, s48–s61, ago. 2016. DOI: [10.15265/iys-2016-s006](https://doi.org/10.15265/iys-2016-s006). Disponível em: <https://doi.org/10.15265/iys-2016-s006>.
- EVANS, Richard H. An Analysis of Criterion Variable Reliability in Conjoint Analysis. **Perceptual and Motor Skills**, SAGE Publications, v. 82, n. 3, p. 988–990, jun. 1996. DOI: [10.2466/pms.1996.82.3.988](https://doi.org/10.2466/pms.1996.82.3.988). Disponível em: <https://doi.org/10.2466/pms.1996.82.3.988>.
- FERREIRA, Flavio Lopes. Serial Evaluation of the SOFA Score to Predict Outcome in Critically Ill Patients. **JAMA**, American Medical Association (AMA), v. 286, n. 14, p. 1754, out. 2001. DOI: [10.1001/jama.286.14.1754](https://doi.org/10.1001/jama.286.14.1754). Disponível em: <https://doi.org/10.1001/jama.286.14.1754>.
- FIFE, Dustin A.; RODGERS, Joseph Lee. Understanding the exploratory/confirmatory data analysis continuum: Moving beyond the “replication crisis”. **American Psychologist**, American Psychological Association (APA), nov. 2021. DOI: [10.1037/amp0000886](https://doi.org/10.1037/amp0000886). Disponível em: <https://doi.org/10.1037/amp0000886>.
- FRANCOIS HUSSON, Julie Josse. **Handling missing values with/in multivariate data analysis (principal component methods)**. [S. l.], 2013. R package version 1.7 — For new features, see the 'Changelog' file (in the package source).
- FUTOMA, Joseph *et al.* Generalization in Clinical Prediction Models: The Blessing and Curse of Measurement Indicator Variables. **Critical Care Explorations**, Ovid Technologies (Wolters Kluwer Health), v. 3, n. 7, e0453, jun. 2021. DOI: [10.1097/cce.0000000000000453](https://doi.org/10.1097/cce.0000000000000453). Disponível em: <https://doi.org/10.1097/cce.0000000000000453>.

- GAGGIN, Hanna K.; JANUZZI, James L. Biomarkers and diagnostics in heart failure. **Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease**, Elsevier BV, v. 1832, n. 12, p. 2442–2450, dez. 2013. DOI: [10.1016/j.bbadis.2012.12.014](https://doi.org/10.1016/j.bbadis.2012.12.014). Disponível em: <https://doi.org/10.1016/j.bbadis.2012.12.014>.
- GALL, J. R. Le. A new Simplified Acute Physiology Score (SAPS II) based on a European/North American multicenter study. **JAMA: The Journal of the American Medical Association**, American Medical Association (AMA), v. 270, n. 24, p. 2957–2963, dez. 1993. DOI: [10.1001/jama.270.24.2957](https://doi.org/10.1001/jama.270.24.2957). Disponível em: <https://doi.org/10.1001/jama.270.24.2957>.
- GALL, Jean. The Logistic Organ Dysfunction System. **JAMA**, American Medical Association (AMA), v. 276, n. 10, p. 802, set. 1996. DOI: [10.1001/jama.1996.03540100046027](https://doi.org/10.1001/jama.1996.03540100046027). Disponível em: <https://doi.org/10.1001/jama.1996.03540100046027>.
- GEORGE, Brandon; SEALS, Samantha; ABAN, Inmaculada. Survival analysis and regression models. **Journal of Nuclear Cardiology**, Springer Science e Business Media LLC, v. 21, n. 4, p. 686–694, mai. 2014. DOI: [10.1007/s12350-014-9908-2](https://doi.org/10.1007/s12350-014-9908-2). Disponível em: <https://doi.org/10.1007/s12350-014-9908-2>.
- GHASSEMI, Marzyeh; CELI, Leo Anthony; STONE, David J. State of the art review: the data revolution in critical care. **Critical Care**, Springer Science e Business Media LLC, v. 19, n. 1, dez. 2015. DOI: [10.1186/s13054-015-0801-4](https://doi.org/10.1186/s13054-015-0801-4). Disponível em: <https://doi.org/10.1186/s13054-015-0801-4>.
- GHASSEMI, Marzyeh; NAUMANN, Tristan *et al.* Unfolding physiological state. In: PROCEEDINGS of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining. [S. l.]: ACM, ago. 2014. DOI: [10.1145/2623330.2623742](https://doi.org/10.1145/2623330.2623742). Disponível em: <https://doi.org/10.1145/2623330.2623742>.
- GODINJAK, Amina Godinjak. Predictive value of SAPS II and APACHE II scoring systems for patient outcome in medical intensive care unit. **Acta Medica Academica**, Academy of Sciences, Arts of Bosnia e Herzegovina, v. 45, n. 2, p. 89–95, dez. 2016. DOI: [10.5644/ama2006-124.165](https://doi.org/10.5644/ama2006-124.165). Disponível em: <https://doi.org/10.5644/ama2006-124.165>.
- GRUNDY, Scott M. *et al.* Assessment of Cardiovascular Risk by Use of Multiple-Risk-Factor Assessment Equations. **Circulation**, Ovid Technologies (Wolters Kluwer Health), v. 100, n. 13, p. 1481–1492, set. 1999. DOI:

[10.1161/01.cir.100.13.1481](https://doi.org/10.1161/01.cir.100.13.1481). Disponível em:

<<https://doi.org/10.1161/01.cir.100.13.1481>>.

GUIDET, Bertrand *et al.* The contribution of frailty, cognition, activity of daily life and comorbidities on outcome in acutely admitted patients over 80 years in European ICUs: the VIP2 study. en. **Intensive Care Med.**, Springer Science e Business Media LLC, v. 46, n. 1, p. 57–69, jan. 2020.

HANCOCK, Mark J. *et al.* A guide to survival analysis for manual therapy clinicians and researchers. **Manual Therapy**, Elsevier BV, v. 19, n. 6, p. 511–516, dez. 2014.

DOI: [10.1016/j.math.2013.08.007](https://doi.org/10.1016/j.math.2013.08.007). Disponível em:

<<https://doi.org/10.1016/j.math.2013.08.007>>.

HANSEN, Emmanuel Mangkornkaew; MEJLDAL, Anna; NIELSEN, Anette Sjøgaard. Predictors of Readmission Following Outpatient Treatment for Alcohol Use Disorder. **Alcohol and Alcoholism**, Oxford University Press (OUP), v. 55, n. 3, p. 291–298, mar. 2020. DOI: [10.1093/alcalc/agaa018](https://doi.org/10.1093/alcalc/agaa018). Disponível em:

<<https://doi.org/10.1093/alcalc/agaa018>>.

HARERIMANA, Gaspard; KIM, Jong Wook; JANG, Beakcheol. A deep attention model to forecast the Length Of Stay and the in-hospital mortality right on admission from ICD codes and demographic data. **Journal of Biomedical Informatics**, v. 118, p. 103778, 2021. ISSN 1532-0464. DOI:

<https://doi.org/10.1016/j.jbi.2021.103778>. Disponível em:

<<https://www.sciencedirect.com/science/article/pii/S1532046421001076>>.

HARRISON, Charlotte. GlaxoSmithKline opens the door on clinical data sharing. **Nature Reviews Drug Discovery**, Springer Science e Business Media LLC, v. 11, n. 12, p. 891–892, nov. 2012. DOI: [10.1038/nrd3907](https://doi.org/10.1038/nrd3907). Disponível em:

<<https://doi.org/10.1038/nrd3907>>.

HARRISON, James E. *et al.* ICD-11: an international classification of diseases for the twenty-first century. **BMC Medical Informatics and Decision Making**, Springer Science e Business Media LLC, v. 21, S6, nov. 2021. DOI:

[10.1186/s12911-021-01534-6](https://doi.org/10.1186/s12911-021-01534-6). Disponível em:

<<https://doi.org/10.1186/s12911-021-01534-6>>.

HAYBAR, Habib; PEZESHKI, Seyed Mohammad Sadegh; SAKI, Najmaldin.

Evaluation of complete blood count parameters in cardiovascular diseases: An early indicator of prognosis? **Experimental and Molecular Pathology**, Elsevier BV,

v. 110, p. 104267, out. 2019. DOI: [10.1016/j.yexmp.2019.104267](https://doi.org/10.1016/j.yexmp.2019.104267). Disponível em: <https://doi.org/10.1016/j.yexmp.2019.104267>.

HENDERSON, Ashley N.; KAUWE, Steven K.; SPARKS, Taylor D. Benchmark datasets incorporating diverse tasks, sample sizes, material systems, and data heterogeneity for materials informatics. **Data in Brief**, Elsevier BV, v. 37, p. 107262, ago. 2021. DOI: [10.1016/j.dib.2021.107262](https://doi.org/10.1016/j.dib.2021.107262). Disponível em: <https://doi.org/10.1016/j.dib.2021.107262>.

HERNAN, Miguel A; HERNANDEZ-DIAZ, Sonia; ROBINS, James M. A structural approach to selection bias. en. **Epidemiology**, Ovid Technologies (Wolters Kluwer Health), v. 15, n. 5, p. 615–625, set. 2004.

HERNÁN, Miguel A. *et al.* Observational Studies Analyzed Like Randomized Experiments. **Epidemiology**, Ovid Technologies (Wolters Kluwer Health), v. 19, n. 6, p. 766–779, nov. 2008. DOI: [10.1097/ede.0b013e3181875e61](https://doi.org/10.1097/ede.0b013e3181875e61). Disponível em: <https://doi.org/10.1097/ede.0b013e3181875e61>.

HILLESTAD, Richard *et al.* Can Electronic Medical Record Systems Transform Health Care? Potential Health Benefits, Savings, And Costs. **Health Affairs**, Health Affairs (Project Hope), v. 24, n. 5, p. 1103–1117, set. 2005. DOI: [10.1377/hlthaff.24.5.1103](https://doi.org/10.1377/hlthaff.24.5.1103). Disponível em: <https://doi.org/10.1377/hlthaff.24.5.1103>.

HOLROYD-LEDUC, Jayna M *et al.* The impact of the electronic medical record on structure, process, and outcomes within primary care: a systematic review of the evidence: Figure 1. **Journal of the American Medical Informatics Association**, Oxford University Press (OUP), v. 18, n. 6, p. 732–737, nov. 2011. DOI: [10.1136/amiajnl-2010-000019](https://doi.org/10.1136/amiajnl-2010-000019). Disponível em: <https://doi.org/10.1136/amiajnl-2010-000019>.

HOSEIN, F Shaun *et al.* A meta-analysis to derive literature-based benchmarks for readmission and hospital mortality after patient discharge from intensive care. en. **Crit. Care**, Springer Science e Business Media LLC, v. 18, n. 6, p. 715, dez. 2014.

HOSOI, Takahiro; OZONE, Sachiko; HAMANO, Jun. Variations in vital signs at the end of life in non-cancer patients: a retrospective study. en. **Ann. Palliat. Med.**, AME Publishing Company, v. 9, n. 5, p. 2678–2683, set. 2020.

HOWE, Jessica L. *et al.* Electronic Health Record Usability Issues and Potential Contribution to Patient Harm. **JAMA**, American Medical Association (AMA), v. 319, n. 12, p. 1276, mar. 2018. DOI: [10.1001/jama.2018.1171](https://doi.org/10.1001/jama.2018.1171). Disponível em: <https://doi.org/10.1001/jama.2018.1171>.

HRIPCSAK, George *et al.* Bias Associated with Mining Electronic Health Records. **Journal of Biomedical Discovery and Collaboration**, University of Illinois Libraries, v. 6, p. 48–52, 2011. DOI: [10.5210/disco.v6i0.3581](https://doi.org/10.5210/disco.v6i0.3581). Disponível em: <https://doi.org/10.5210/disco.v6i0.3581>.

HUG, Caleb. Predicting the Risk and Trajectory of Intensive Care Patients Using Survival Models, ago. 2007.

HUG, Caleb; SZOLOVITS, Peter. ICU Acuity: Real-time Models versus Daily Models. **AMIA ... Annual Symposium proceedings / AMIA Symposium. AMIA Symposium**, v. 2009, p. 260–4, nov. 2009.

IKIZLER, T. Alp *et al.* A prospective cohort study of acute kidney injury and kidney outcomes, cardiovascular events, and death. **Kidney International**, Elsevier BV, v. 99, n. 2, p. 456–465, fev. 2021. DOI: [10.1016/j.kint.2020.06.032](https://doi.org/10.1016/j.kint.2020.06.032). Disponível em: <https://doi.org/10.1016/j.kint.2020.06.032>.

IN, Junyong; LEE, Dong Kyu. Survival analysis: part II applied clinical data analysis. **Korean Journal of Anesthesiology**, The Korean Society of Anesthesiologists, v. 72, n. 5, p. 441–457, out. 2019. DOI: [10.4097/kja.19183](https://doi.org/10.4097/kja.19183). Disponível em: <https://doi.org/10.4097/kja.19183>.

JANETT, Robert S.; YERACARIS, Peter Pano. Electronic Medical Records in the American Health System: challenges and lessons learned. **Ciencia e Saude Coletiva**, FapUNIFESP (SciELO), v. 25, n. 4, p. 1293–1304, abr. 2020. DOI: [10.1590/1413-81232020254.28922019](https://doi.org/10.1590/1413-81232020254.28922019). Disponível em: <https://doi.org/10.1590/1413-81232020254.28922019>.

JENTZER, Jacob C. *et al.* Changes in comorbidities, diagnoses, therapies and outcomes in a contemporary cardiac intensive care unit population. **American Heart Journal**, Elsevier BV, v. 215, p. 12–19, set. 2019. DOI: [10.1016/j.ahj.2019.05.012](https://doi.org/10.1016/j.ahj.2019.05.012). Disponível em: <https://doi.org/10.1016/j.ahj.2019.05.012>.

JHA, Ashish K. Sharing Clinical Data Electronically. **JAMA**, American Medical Association (AMA), v. 307, n. 16, p. 1695, abr. 2012. DOI: [10.1001/jama.2012.525](https://doi.org/10.1001/jama.2012.525). Disponível em: <https://doi.org/10.1001/jama.2012.525>.

JIN, Jill. Risk Assessment for Cardiovascular Disease With Nontraditional Risk Factors. **JAMA**, v. 320, n. 3, p. 316–316, jul. 2018. ISSN 0098-7484. DOI: [10.1001/jama.2018.9122](https://doi.org/10.1001/jama.2018.9122). Disponível em: <https://doi.org/10.1001/jama.2018.9122>.

JOHNSON, Alistair *et al.* **MIMIC-IV**. [S. l.]: PhysioNet, 2020. DOI:

[10.13026/A3WN-HQ05](https://doi.org/10.13026/A3WN-HQ05). Disponível em:

[<https://physionet.org/content/mimiciv/0.4/>](https://physionet.org/content/mimiciv/0.4/).

JOHNSON, Alistair E. W.; KRAMER, Andrew A.; CLIFFORD, Gari D. A New Severity of Illness Scale Using a Subset of Acute Physiology and Chronic Health Evaluation Data Elements Shows Comparable Predictive Accuracy. **Critical Care Medicine**, Ovid Technologies (Wolters Kluwer Health), v. 41, n. 7, p. 1711–1718, jul. 2013. DOI: [10.1097/ccm.0b013e31828a24fe](https://doi.org/10.1097/ccm.0b013e31828a24fe). Disponível em:

[<https://doi.org/10.1097/ccm.0b013e31828a24fe>](https://doi.org/10.1097/ccm.0b013e31828a24fe).

JOHNSON, Alistair E. W.; POLLARD, Tom J.; MARK, Roger G. Reproducibility in critical care: a mortality prediction case study. In_____. **Proceedings of the 2nd Machine Learning for Healthcare Conference**. [S. l.]: PMLR, ago. 2017. v. 68. (Proceedings of Machine Learning Research), p. 361–376. Disponível em:

[<https://proceedings.mlr.press/v68/johnson17a.html>](https://proceedings.mlr.press/v68/johnson17a.html).

JOHNSON, Alistair E.W. *et al.* MIMIC-III, a freely accessible critical care database. **Scientific Data**, Springer Science e Business Media LLC, v. 3, n. 1, mai. 2016. DOI: [10.1038/sdata.2016.35](https://doi.org/10.1038/sdata.2016.35). Disponível em:

[<https://doi.org/10.1038/sdata.2016.35>](https://doi.org/10.1038/sdata.2016.35).

KADOOKA, Kosuke *et al.* Non-cardiovascular disorders in a contemporary cardiovascular intensive care unit in Japan. **Journal of Cardiology**, Elsevier BV, v. 78, n. 2, p. 166–171, ago. 2021. DOI: [10.1016/j.jjcc.2021.03.002](https://doi.org/10.1016/j.jjcc.2021.03.002). Disponível em:

[<https://doi.org/10.1016/j.jjcc.2021.03.002>](https://doi.org/10.1016/j.jjcc.2021.03.002).

KANE, Robert L. *et al.* The Association of Registered Nurse Staffing Levels and Patient Outcomes. **Medical Care**, Ovid Technologies (Wolters Kluwer Health), v. 45, n. 12, p. 1195–1204, dez. 2007. DOI: [10.1097/mlr.0b013e3181468ca3](https://doi.org/10.1097/mlr.0b013e3181468ca3). Disponível em:

[<https://doi.org/10.1097/mlr.0b013e3181468ca3>](https://doi.org/10.1097/mlr.0b013e3181468ca3).

KAPLAN, E. L.; MEIER, Paul. Nonparametric Estimation from Incomplete Observations. **Journal of the American Statistical Association**, Informa UK Limited, v. 53, n. 282, p. 457–481, jun. 1958. DOI: [10.1080/01621459.1958.10501452](https://doi.org/10.1080/01621459.1958.10501452). Disponível em: [<https://doi.org/10.1080/01621459.1958.10501452>](https://doi.org/10.1080/01621459.1958.10501452).

KASAOKA, Shunji. Evolved role of the cardiovascular intensive care unit (CICU). **Journal of Intensive Care**, Springer Science e Business Media LLC, v. 5, n. 1, dez. 2017. DOI: [10.1186/s40560-017-0271-7](https://doi.org/10.1186/s40560-017-0271-7). Disponível em:

[<https://doi.org/10.1186/s40560-017-0271-7>](https://doi.org/10.1186/s40560-017-0271-7).

- KITCHENHAM, B.; CHARTERS, S. **Guidelines for performing Systematic Literature Reviews in Software Engineering**. [S. l.: s. n.], 2007.
- KLUYVER, Thomas *et al.* Jupyter Notebooks – a publishing format for reproducible computational workflows. In_____. **Positioning and Power in Academic Publishing: Players, Agents and Agendas**. [S. l.: s. n.], 2016. IOS Press, p. 87–90.
- KNAUS, William A. *et al.* The APACHE III Prognostic System. **Chest**, Elsevier BV, v. 100, n. 6, p. 1619–1636, dez. 1991. DOI: [10.1378/chest.100.6.1619](https://doi.org/10.1378/chest.100.6.1619). Disponível em: <https://doi.org/10.1378/chest.100.6.1619>.
- KOHAVI, Ron. A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection. v. 14, mar. 2001.
- KOMOROWSKI, Matthieu *et al.* Exploratory Data Analysis. In: SECONDARY Analysis of Electronic Health Records. [S. l.]: Springer International Publishing, 2016. p. 185–203. DOI: [10.1007/978-3-319-43742-2_15](https://doi.org/10.1007/978-3-319-43742-2_15). Disponível em: https://doi.org/10.1007/978-3-319-43742-2_15.
- KOYE, Digsu N *et al.* The global epidemiology of diabetes and kidney disease. en. **Adv. Chronic Kidney Dis.**, Elsevier BV, v. 25, n. 2, p. 121–132, mar. 2018.
- KOZLOWSKI, Desirée *et al.* The role of emotion in clinical decision making: an integrative literature review. **BMC Medical Education**, Springer Science e Business Media LLC, v. 17, n. 1, dez. 2017. DOI: [10.1186/s12909-017-1089-7](https://doi.org/10.1186/s12909-017-1089-7). Disponível em: <https://doi.org/10.1186/s12909-017-1089-7>.
- KRAMER, Andrew A. Are ICU length of stay predictions worthwhile? en. **Crit. Care Med.**, Ovid Technologies (Wolters Kluwer Health), v. 45, n. 2, p. 379–380, fev. 2017.
- KURTZ, Pedro *et al.* Hospital length of stay and 30-day mortality prediction in stroke: A machine learning analysis of 17,000 ICU admissions in Brazil. en. **Neurocrit. Care**, Springer Science e Business Media LLC, v. 37, Suppl 2, p. 313–321, ago. 2022.
- LACERDA, Daniel Pacheco *et al.* Design Science Research: método de pesquisa para a engenharia de produção. **Gestão e Produção**, FapUNIFESP (SciELO), v. 20, n. 4, p. 741–761, nov. 2013. DOI: [10.1590/s0104-530x2013005000014](https://doi.org/10.1590/s0104-530x2013005000014). Disponível em: <https://doi.org/10.1590/s0104-530x2013005000014>.
- LASSNIGG, A. Minimal Changes of Serum Creatinine Predict Prognosis in Patients after Cardiothoracic Surgery: A Prospective Cohort Study. **Journal of the American Society of Nephrology**, American Society of Nephrology (ASN), v. 15, n. 6,

- p. 1597–1605, jun. 2004. DOI: [10.1097/01.asn.0000130340.93930.dd](https://doi.org/10.1097/01.asn.0000130340.93930.dd). Disponível em: <https://doi.org/10.1097/01.asn.0000130340.93930.dd>.
- LASSNIGG, Andrea *et al.* Impact of minimal increases in serum creatinine on outcome in patients after cardiothoracic surgery: Do we have to revise current definitions of acute renal failure? **Critical Care Medicine**, Ovid Technologies (Wolters Kluwer Health), v. 36, n. 4, p. 1129–1137, abr. 2008. DOI: [10.1097/ccm.0b013e318169181a](https://doi.org/10.1097/ccm.0b013e318169181a). Disponível em: <https://doi.org/10.1097/ccm.0b013e318169181a>.
- LEE, Joon; MASLOVE, David M. Customization of a Severity of Illness Score Using Local Electronic Medical Record Data. **Journal of Intensive Care Medicine**, SAGE Publications, v. 32, n. 1, p. 38–47, mai. 2015. DOI: [10.1177/0885066615585951](https://doi.org/10.1177/0885066615585951). Disponível em: <https://doi.org/10.1177/0885066615585951>.
- LEE, Joon; MASLOVE, David M.; DUBIN, Joel A. Personalized Mortality Prediction Driven by Electronic Medical Data and a Patient Similarity Metric. Edição: Frank Emmert-Streib. **PLOS ONE**, Public Library of Science (PLoS), v. 10, n. 5, e0127428, mai. 2015. DOI: [10.1371/journal.pone.0127428](https://doi.org/10.1371/journal.pone.0127428). Disponível em: <https://doi.org/10.1371/journal.pone.0127428>.
- LEONE, Aurelio; LANDINI, Luigi. Modifying Cardiovascular Risk Factors: Bases to Improve Diagnostic and Instrumental Approaches. **Current Pharmaceutical Design**, Bentham Science Publishers Ltd., v. 27, n. 16, p. 1869–1870, mai. 2021. DOI: [10.2174/138161282716210430081537](https://doi.org/10.2174/138161282716210430081537). Disponível em: <https://doi.org/10.2174/138161282716210430081537>.
- LEWIS, David D. Naive (Bayes) at forty: The independence assumption in information retrieval. In: **MACHINE Learning: ECML-98**. [S. l.]: Springer Berlin Heidelberg, 1998. p. 4–15. DOI: [10.1007/bfb0026666](https://doi.org/10.1007/bfb0026666). Disponível em: <https://doi.org/10.1007/bfb0026666>.
- LIGHTHALL, Geoffrey K.; VAZQUEZ-GUILLAMET, Cristina. Understanding Decision Making in Critical Care. eng. **Clinical medicine & research**, Marshfield Clinic, v. 13, n. 3-4, p. 156–168, dez. 2015. ISSN 1554-6179. DOI: [10.3121/cmr.2015.1289](https://doi.org/10.3121/cmr.2015.1289). Disponível em: <https://doi.org/10.3121/cmr.2015.1289>.
- LONGATO, Enrico; VETTORETTI, Martina; CAMILLO, Barbara Di. A practical perspective on the concordance index for the evaluation and selection of prognostic time-to-event models. **Journal of Biomedical Informatics**, Elsevier BV, v. 108, p. 103496, ago. 2020. DOI: [10.1016/j.jbi.2020.103496](https://doi.org/10.1016/j.jbi.2020.103496). Disponível em: <https://doi.org/10.1016/j.jbi.2020.103496>.

- LÜSEBRINK, Enzo *et al.* New challenges in cardiac intensive care units. **Clinical Research in Cardiology**, Springer Science e Business Media LLC, v. 110, n. 9, p. 1369–1379, mai. 2021. DOI: [10.1007/s00392-021-01869-0](https://doi.org/10.1007/s00392-021-01869-0). Disponível em: [10.1007/s00392-021-01869-0](https://doi.org/10.1007/s00392-021-01869-0).
- MADDEN, Abaigael *et al.* Factors associated with psychiatric readmission of children and adolescents in the U.S.: A systematic review of the literature. **General Hospital Psychiatry**, Elsevier BV, v. 65, p. 33–42, jul. 2020. DOI: [10.1016/j.genhosppsych.2020.05.004](https://doi.org/10.1016/j.genhosppsych.2020.05.004). Disponível em: [10.1016/j.genhosppsych.2020.05.004](https://doi.org/10.1016/j.genhosppsych.2020.05.004).
- MAMADJANOV, Temur *et al.* Admission glucose level and short-term mortality in older patients with acute myocardial infarction: results from the KORA Myocardial Infarction Registry. **BMJ Open**, BMJ, v. 11, n. 6, e046641, jun. 2021. DOI: [10.1136/bmjopen-2020-046641](https://doi.org/10.1136/bmjopen-2020-046641). Disponível em: [10.1136/bmjopen-2020-046641](https://doi.org/10.1136/bmjopen-2020-046641).
- MARRARO, Giuseppe A.; SPADA, Claudio. Weaning and the Suitability of Retrospective Cohort Studies. **Critical Care Medicine**, Ovid Technologies (Wolters Kluwer Health), v. 49, n. 2, p. 369–372, jan. 2021. DOI: [10.1097/ccm.0000000000004798](https://doi.org/10.1097/ccm.0000000000004798). Disponível em: [10.1097/ccm.0000000000004798](https://doi.org/10.1097/ccm.0000000000004798).
- MAYHEW, Deborah J. The usability engineering lifecycle. In. DOI: [10.1145/632716.632805](https://doi.org/10.1145/632716.632805). Disponível em: [10.1145/632716.632805](https://doi.org/10.1145/632716.632805).
- MEHTA, Nishita; PANDIT, Anil. Concurrence of big data analytics and healthcare: A systematic review. **International Journal of Medical Informatics**, Elsevier BV, v. 114, p. 57–65, jun. 2018. DOI: [10.1016/j.ijmedinf.2018.03.013](https://doi.org/10.1016/j.ijmedinf.2018.03.013). Disponível em: [10.1016/j.ijmedinf.2018.03.013](https://doi.org/10.1016/j.ijmedinf.2018.03.013).
- MEHTA, Ravindra L; KELLUM, John A *et al.* Acute Kidney Injury Network: report of an initiative to improve outcomes in acute kidney injury. **Critical Care**, Springer Science e Business Media LLC, v. 11, n. 2, r31, 2007. DOI: [10.1186/cc5713](https://doi.org/10.1186/cc5713). Disponível em: [10.1186/cc5713](https://doi.org/10.1186/cc5713).
- MEYER, Anthony A. *et al.* Prospective Comparison of Clinical Judgment and APACHE II Score in Predicting the Outcome in Critically Ill Surgical Patients. **The Journal of Trauma: Injury, Infection, and Critical Care**, Ovid Technologies (Wolters Kluwer

Health), v. 32, n. 6, p. 747–754, jun. 1992. DOI: [10.1097/00005373-199206000-00013](https://doi.org/10.1097/00005373-199206000-00013). Disponível em: <https://doi.org/10.1097/00005373-199206000-00013>.

MORROW, David A. *et al.* Evolution of Critical Care Cardiology: Transformation of the Cardiovascular Intensive Care Unit and the Emerging Need for New Medical Staffing and Training Models. **Circulation**, Ovid Technologies (Wolters Kluwer Health), v. 126, n. 11, p. 1408–1428, set. 2012. DOI: [10.1161/cir.0b013e31826890b0](https://doi.org/10.1161/cir.0b013e31826890b0). Disponível em: <https://doi.org/10.1161/cir.0b013e31826890b0>.

MUKAKA, Mavuto. Statistics Corner: A guide to appropriate use of Correlation coefficient in medical research. **Malawi medical journal : the journal of Medical Association of Malawi**, v. 24, p. 69–71, set. 2012.

NAGAI, Toshiyuki; NAKAO, Motoki; ANZAI, Toshihisa. Risk Stratification Towards Precision Medicine in Heart Failure: Current Progress and Future Perspectives. **Circulation Journal**, Japanese Circulation Society, v. 85, n. 5, p. 576–583, abr. 2021. DOI: [10.1253/circj.cj-20-1299](https://doi.org/10.1253/circj.cj-20-1299). Disponível em: <https://doi.org/10.1253/circj.cj-20-1299>.

NASSIRI, Vahid *et al.* On using multiple imputation for exploratory factor analysis of incomplete data. **Behavior Research Methods**, Springer Science e Business Media LLC, v. 50, n. 2, p. 501–517, fev. 2018. DOI: [10.3758/s13428-017-1013-4](https://doi.org/10.3758/s13428-017-1013-4). Disponível em: <https://doi.org/10.3758/s13428-017-1013-4>.

NEVILLE, Jennifer; JENSEN, David. Relational Dependency Networks. **J. Mach. Learn. Res.**, JMLR.org, v. 8, p. 653–692, mai. 2007. ISSN 1532-4435.

NICK, Todd G.; CAMPBELL, Kathleen M. Logistic Regression. In: TOPICS in Biostatistics. [S. l.]: Humana Press, 2007. p. 273–301. DOI: [10.1007/978-1-59745-530-5_14](https://doi.org/10.1007/978-1-59745-530-5_14). Disponível em: https://doi.org/10.1007/978-1-59745-530-5_14.

NORTH, Brian J.; SINCLAIR, David A. The Intersection Between Aging and Cardiovascular Disease. **Circulation Research**, Ovid Technologies (Wolters Kluwer Health), v. 110, n. 8, p. 1097–1108, abr. 2012. DOI: [10.1161/circresaha.111.246876](https://doi.org/10.1161/circresaha.111.246876). Disponível em: <https://doi.org/10.1161/circresaha.111.246876>.

NUNES, Luciana Neves; KLÜCK, Mariza Machado; FACHEL, Jandyra Maria Guimarães. Comparação de métodos de imputação única e múltipla usando como exemplo um modelo de risco para mortalidade cirúrgica. **Revista Brasileira de Epidemiologia**, FapUNIFESP (SciELO), v. 13, n. 4, p. 596–606, dez.

2010. DOI: [10.1590/s1415-790x2010000400005](https://doi.org/10.1590/s1415-790x2010000400005). Disponível em: [10.1590/s1415-790x2010000400005](https://doi.org/10.1590/s1415-790x2010000400005).
- OCONNOR, Siobhan *et al.* Data visualization in health care: The Florence effect. **Journal of Advanced Nursing**, Wiley, v. 76, n. 7, p. 1488–1490, mar. 2020. DOI: [10.1111/jan.14334](https://doi.org/10.1111/jan.14334). Disponível em: [10.1111/jan.14334](https://doi.org/10.1111/jan.14334).
- OSTLUND, Ulrika *et al.* Combining qualitative and quantitative research within mixed method research designs: A methodological review. **International Journal of Nursing Studies**, Elsevier BV, v. 48, n. 3, p. 369–383, mar. 2011. DOI: [10.1016/j.ijnurstu.2010.10.005](https://doi.org/10.1016/j.ijnurstu.2010.10.005). Disponível em: [10.1016/j.ijnurstu.2010.10.005](https://doi.org/10.1016/j.ijnurstu.2010.10.005).
- OZARDA, Yesim. Reference intervals: current status, recent developments and future considerations. **Biochemia Medica**, Croatian Society for Medical Biochemistry e Laboratory Medicine, p. 5–16, 2016. DOI: [10.11613/bm.2016.001](https://doi.org/10.11613/bm.2016.001). Disponível em: [10.11613/bm.2016.001](https://doi.org/10.11613/bm.2016.001).
- OZARDA, Yesim *et al.* Distinguishing reference intervals and clinical decision limits – A review by the IFCC Committee on Reference Intervals and Decision Limits. **Critical Reviews in Clinical Laboratory Sciences**, Informa UK Limited, v. 55, n. 6, p. 420–431, jul. 2018. DOI: [10.1080/10408363.2018.1482256](https://doi.org/10.1080/10408363.2018.1482256). Disponível em: [10.1080/10408363.2018.1482256](https://doi.org/10.1080/10408363.2018.1482256).
- PAGE, Matthew J *et al.* PRISMA 2020 explanation and elaboration: updated guidance and exemplars for reporting systematic reviews. **BMJ**, BMJ, p. 160, mar. 2021. DOI: [10.1136/bmj.n160](https://doi.org/10.1136/bmj.n160). Disponível em: [10.1136/bmj.n160](https://doi.org/10.1136/bmj.n160).
- PEARL, Judea. From Bayesian Networks to Causal Networks. In: **MATHEMATICAL Models for Handling Partial Knowledge in Artificial Intelligence**. [S. l.]: Springer US, 1995. p. 157–182. DOI: [10.1007/978-1-4899-1424-8_9](https://doi.org/10.1007/978-1-4899-1424-8_9). Disponível em: [10.1007/978-1-4899-1424-8_9](https://doi.org/10.1007/978-1-4899-1424-8_9).
- PEARSON, Karl. NOTES ON THE HISTORY OF CORRELATION. **Biometrika**, Oxford University Press (OUP), v. 13, n. 1, p. 25–45, 1920. DOI: [10.1093/biomet/13.1.25](https://doi.org/10.1093/biomet/13.1.25). Disponível em: [10.1093/biomet/13.1.25](https://doi.org/10.1093/biomet/13.1.25).
- PEARSON, Ronald K. **Exploratory Data Analysis Using R**. [S. l.]: Chapman e Hall/CRC, mai. 2018. DOI: [10.1201/9781315382111](https://doi.org/10.1201/9781315382111). Disponível em: [10.1201/9781315382111](https://doi.org/10.1201/9781315382111).

- PEDREGOSA, Fabian *et al.* Scikit-learn: Machine Learning in Python. arXiv, 2012. DOI: [10.48550/ARXIV.1201.0490](https://doi.org/10.48550/ARXIV.1201.0490). Disponível em: <https://arxiv.org/abs/1201.0490>.
- PELLATHY, Tiffany Purcell; PINSKY, Michael R.; HRAVNAK, Marilyn. Intensive Care Unit Scoring Systems. **Critical Care Nurse**, AACN Publishing, v. 41, n. 4, p. 54–64, ago. 2021. DOI: [10.4037/ccn2021613](https://doi.org/10.4037/ccn2021613). Disponível em: <https://doi.org/10.4037/ccn2021613>.
- PETERSEN, Kai *et al.* Systematic Mapping Studies in Software Engineering. In: ELECTRONIC Workshops in Computing. [S. l.]: BCS Learning & Development, jun. 2008. DOI: [10.14236/ewic/ease2008.8](https://doi.org/10.14236/ewic/ease2008.8).
- PINTO, L. Nelson Sanchez; LUO, Yuan; CHURPEK, Matthew M. Big Data and Data Science in Critical Care. **Chest**, Elsevier BV, v. 154, n. 5, p. 1239–1248, nov. 2018. DOI: [10.1016/j.chest.2018.04.037](https://doi.org/10.1016/j.chest.2018.04.037). Disponível em: <https://doi.org/10.1016/j.chest.2018.04.037>.
- PIRRACCHIO, Romain *et al.* Mortality prediction in intensive care units with the Super ICU Learner Algorithm (SICULA): a population-based study. **The Lancet Respiratory Medicine**, Elsevier BV, v. 3, n. 1, p. 42–52, jan. 2015. DOI: [10.1016/s2213-2600\(14\)70239-5](https://doi.org/10.1016/s2213-2600(14)70239-5). Disponível em: [https://doi.org/10.1016/s2213-2600\(14\)70239-5](https://doi.org/10.1016/s2213-2600(14)70239-5).
- PLUIM, J.P.W.; MAINTZ, J.B.A.; VIERGEVER, M.A. Mutual-information-based registration of medical images: a survey. **IEEE Transactions on Medical Imaging**, v. 22, n. 8, p. 986–1004, 2003. DOI: [10.1109/TMI.2003.815867](https://doi.org/10.1109/TMI.2003.815867).
- POLLARD, Tom J. *et al.* The eICU Collaborative Research Database, a freely available multi-center database for critical care research. **Scientific Data**, Springer Science e Business Media LLC, v. 5, n. 1, set. 2018. DOI: [10.1038/sdata.2018.178](https://doi.org/10.1038/sdata.2018.178). Disponível em: <https://doi.org/10.1038/sdata.2018.178>.
- PORTET, Stephanie. A primer on model selection using the Akaike Information Criterion. **Infectious Disease Modelling**, Elsevier BV, v. 5, p. 111–128, 2020. DOI: [10.1016/j.idm.2019.12.010](https://doi.org/10.1016/j.idm.2019.12.010). Disponível em: <https://doi.org/10.1016/j.idm.2019.12.010>.
- POTDAR, Kedar; PARDAWALA, Taher; PAI, Chinmay. A Comparative Study of Categorical Variable Encoding Techniques for Neural Network Classifiers. **International Journal of Computer Applications**, v. 175, p. 7–9, out. 2017. DOI: [10.5120/ijca2017915495](https://doi.org/10.5120/ijca2017915495).

- QUAN, Hude *et al.* Coding Algorithms for Defining Comorbidities in ICD-9-CM and ICD-10 Administrative Data. **Medical Care**, Ovid Technologies (Wolters Kluwer Health), v. 43, n. 11, p. 1130–1139, nov. 2005. DOI: [10.1097/01.mlr.0000182534.19832.83](https://doi.org/10.1097/01.mlr.0000182534.19832.83). Disponível em: [10.1097/01.mlr.0000182534.19832.83](https://doi.org/10.1097/01.mlr.0000182534.19832.83).
- QUINLAN, J.R. Simplifying decision trees. **International Journal of Man-Machine Studies**, Elsevier BV, v. 27, n. 3, p. 221–234, set. 1987. DOI: [10.1016/s0020-7373\(87\)80053-6](https://doi.org/10.1016/s0020-7373(87)80053-6). Disponível em: [10.1016/s0020-7373\(87\)80053-6](https://doi.org/10.1016/s0020-7373(87)80053-6).
- RASSENFOSSE, Gaétan de; SELIGER, Florian. Imputation of missing information in worldwide patent data. **Data in Brief**, Elsevier BV, v. 34, p. 106615, fev. 2021. DOI: [10.1016/j.dib.2020.106615](https://doi.org/10.1016/j.dib.2020.106615). Disponível em: [10.1016/j.dib.2020.106615](https://doi.org/10.1016/j.dib.2020.106615).
- ROJAS, Juan C *et al.* Accuracy of clinicians' ability to predict the need for intensive care unit readmission. en. **Ann. Am. Thorac. Soc.**, American Thoracic Society, v. 17, n. 7, p. 847–853, jul. 2020.
- ROMEIJN, Jan-Willem. Philosophy of Statistics. In: ZALTA, Edward N. (Ed.). **The Stanford Encyclopedia of Philosophy**. Spring 2022. [S. l.]: Metaphysics Research Lab, Stanford University, 2022.
- RONCO, Claudio; BELLASI, Antonio; LULLO, Luca Di. Cardiorenal Syndrome: An Overview. **Advances in Chronic Kidney Disease**, Elsevier BV, v. 25, n. 5, p. 382–390, set. 2018. DOI: [10.1053/j.ackd.2018.08.004](https://doi.org/10.1053/j.ackd.2018.08.004). Disponível em: [10.1053/j.ackd.2018.08.004](https://doi.org/10.1053/j.ackd.2018.08.004).
- ROSENBERG, Andrew L.; WATTS, Charles. Patients Readmitted to ICUs: A Systematic Review of Risk Factors and Outcomes. **Chest**, v. 118, n. 2, p. 492–502, 2000. ISSN 0012-3692. DOI: <https://doi.org/10.1378/chest.118.2.492>. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0012369215518138>.
- ROSS, Brian C. Mutual Information between Discrete and Continuous Data Sets. Edição: Daniele Marinazzo. **PLoS ONE**, Public Library of Science (PLoS), v. 9, n. 2, e87357, fev. 2014. DOI: [10.1371/journal.pone.0087357](https://doi.org/10.1371/journal.pone.0087357). Disponível em: [10.1371/journal.pone.0087357](https://doi.org/10.1371/journal.pone.0087357).

RUMELHART, David E.; HINTON, Geoffrey E.; WILLIAMS, Ronald J. Learning representations by back-propagating errors. **Nature**, Springer Science e Business Media LLC, v. 323, n. 6088, p. 533–536, out. 1986. DOI: [10.1038/323533a0](https://doi.org/10.1038/323533a0). Disponível em: <https://doi.org/10.1038/323533a0>.

RUSANOV, Alexander *et al.* Hidden in plain sight: bias towards sick patients when sampling patients with sufficient electronic health record data for research. **BMC Medical Informatics and Decision Making**, Springer Science e Business Media LLC, v. 14, n. 1, jun. 2014. DOI: [10.1186/1472-6947-14-51](https://doi.org/10.1186/1472-6947-14-51). Disponível em: <https://doi.org/10.1186/1472-6947-14-51>.

SAFRAN, C. *et al.* Toward a National Framework for the Secondary Use of Health Data: An American Medical Informatics Association White Paper. **Journal of the American Medical Informatics Association**, Oxford University Press (OUP), v. 14, n. 1, p. 1–9, jan. 2007. DOI: [10.1197/jamia.m2273](https://doi.org/10.1197/jamia.m2273). Disponível em: <https://doi.org/10.1197/jamia.m2273>.

SANHARAWI, M. El; NAUDET, F. Comprendre la regression logistique. **Journal Francais d Ophtalmologie**, Elsevier BV, v. 36, n. 8, p. 710–715, out. 2013. DOI: [10.1016/j.jfo.2013.05.008](https://doi.org/10.1016/j.jfo.2013.05.008). Disponível em: <https://doi.org/10.1016/j.jfo.2013.05.008>.

SARA, Jaskanwal D.S. *et al.* Clinical decision-making: Challenging traditional assumptions. **International Journal of Cardiology**, Elsevier BV, v. 326, p. 6–11, mar. 2021. DOI: [10.1016/j.ijcard.2020.10.077](https://doi.org/10.1016/j.ijcard.2020.10.077). Disponível em: <https://doi.org/10.1016/j.ijcard.2020.10.077>.

SARIPALLE, Rishi; RUNYAN, Christopher; RUSSELL, Mitchell. Using HL7 FHIR to achieve interoperability in patient health record. **Journal of Biomedical Informatics**, Elsevier BV, v. 94, p. 103188, jun. 2019. DOI: [10.1016/j.jbi.2019.103188](https://doi.org/10.1016/j.jbi.2019.103188). Disponível em: <https://doi.org/10.1016/j.jbi.2019.103188>.

SARWAR, Chaudhry M.S. *et al.* Hyperkalemia in Heart Failure. **Journal of the American College of Cardiology**, v. 68, n. 14, p. 1575–1589, 2016. ISSN 0735-1097. DOI: <https://doi.org/10.1016/j.jacc.2016.06.060>. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0735109716346563>.

SAURO, Jeff; LEWIS, James R. **Quantifying the user experience: Practical statistics for user research**. [S. l.]: Morgan Kaufmann, 2016.

- SAVARESE, Gianluigi *et al.* Global burden of heart failure: a comprehensive and updated review of epidemiology. **Cardiovascular Research**, Oxford University Press (OUP), fev. 2022. DOI: [10.1093/cvr/cvac013](https://doi.org/10.1093/cvr/cvac013). Disponível em: <https://doi.org/10.1093/cvr/cvac013>.
- SCHNOHR, P. Coronary heart disease risk factors ranked by importance for the individual and community. A 21 year follow-up of 12000 men and women from The Copenhagen City Heart Study. **European Heart Journal**, Oxford University Press (OUP), v. 23, n. 8, p. 620–626, abr. 2002. DOI: [10.1053/euhj.2001.2842](https://doi.org/10.1053/euhj.2001.2842). Disponível em: <https://doi.org/10.1053/euhj.2001.2842>.
- SCHOBER, Patrick; VETTER, Thomas R. Logistic Regression in Medical Research. **Anesthesia and Analgesia**, Ovid Technologies (Wolters Kluwer Health), v. 132, n. 2, p. 365–366, jan. 2021. DOI: [10.1213/ane.0000000000005247](https://doi.org/10.1213/ane.0000000000005247). Disponível em: <https://doi.org/10.1213/ane.0000000000005247>.
- _____. Survival Analysis and Interpretation of Time-to-Event Data. **Anesthesia & Analgesia**, Ovid Technologies (Wolters Kluwer Health), v. 127, n. 3, p. 792–798, set. 2018. DOI: [10.1213/ane.0000000000003653](https://doi.org/10.1213/ane.0000000000003653). Disponível em: <https://doi.org/10.1213/ane.0000000000003653>.
- SCHULZ, Wade L. *et al.* Temporal relationship of computed and structured diagnoses in electronic health record data. **BMC Medical Informatics and Decision Making**, Springer Science e Business Media LLC, v. 21, n. 1, fev. 2021. DOI: [10.1186/s12911-021-01416-x](https://doi.org/10.1186/s12911-021-01416-x). Disponível em: <https://doi.org/10.1186/s12911-021-01416-x>.
- SHAILAJA, K.; SEETHARAMULU, B.; JABBAR, M. A. Machine Learning in Healthcare: A Review. In: 2018 Second International Conference on Electronics, Communication and Aerospace Technology (ICECA). [S. l.: s. n.], 2018. p. 910–914. DOI: [10.1109/ICECA.2018.8474918](https://doi.org/10.1109/ICECA.2018.8474918).
- SHARMA, ZubinD *et al.* Sequential organ failure assessment score as prognostic marker in critically ill patients in a tertiary care intensive care unit. **International Journal of Medicine and Public Health**, EManuscript Technologies, v. 3, n. 3, p. 155, 2013. DOI: [10.4103/2230-8598.118956](https://doi.org/10.4103/2230-8598.118956). Disponível em: <https://doi.org/10.4103/2230-8598.118956>.
- SHIFFLER, Ronald E. Maximum Z Scores and Outliers. **The American Statistician**, Informa UK Limited, v. 42, n. 1, p. 79–80, fev. 1988. DOI:

[10.1080/00031305.1988.10475530](https://doi.org/10.1080/00031305.1988.10475530). Disponível em:

<<https://doi.org/10.1080/00031305.1988.10475530>>.

SIEST, Gerard *et al.* The theory of reference values: an unfinished symphony. **Clinical Chemistry and Laboratory Medicine (CCLM)**, Walter de Gruyter GmbH, v. 51, n. 1, p. 47–64, nov. 2012. DOI: [10.1515/cclm-2012-0682](https://doi.org/10.1515/cclm-2012-0682). Disponível em:

<<https://doi.org/10.1515/cclm-2012-0682>>.

SINGH, Harshinder *et al.* Nearest neighbor estimates of entropy. **Amer. J. Math. Management Sci.**, Informa UK Limited, v. 23, n. 3-4, p. 301–321, fev. 2003.

SINGH, Harvineet; MHASAWADE, Vishwali; CHUNARA, Rumi. Generalizability challenges of mortality risk prediction models: A retrospective analysis on a multi-center database. Edição: Tom J. Pollard. **PLOS Digital Health**, Public Library of Science (PLoS), v. 1, n. 4, e0000023, abr. 2022. DOI: [10.1371/journal.pdig.0000023](https://doi.org/10.1371/journal.pdig.0000023).

Disponível em: <<https://doi.org/10.1371/journal.pdig.0000023>>.

SINGH, Ritesh; MUKHOPADHYAY, Keshab. Survival analysis in clinical trials: Basics and must know areas. eng. **Perspectives in clinical research**, Medknow Publications & Media Pvt Ltd, v. 2, n. 4, p. 145–148, out. 2011. ISSN 2229-5488. DOI:

[10.4103/2229-3485.86872](https://doi.org/10.4103/2229-3485.86872). Disponível em:

<<https://doi.org/10.4103/2229-3485.86872>>.

SINGH, Vikash *et al.* Impact of train/test sample regimen on performance estimate stability of machine learning in cardiovascular imaging. **Scientific Reports**, Springer Science e Business Media LLC, v. 11, n. 1, jul. 2021. DOI:

[10.1038/s41598-021-93651-5](https://doi.org/10.1038/s41598-021-93651-5). Disponível em:

<<https://doi.org/10.1038/s41598-021-93651-5>>.

SMILDE, A. K. *et al.* Matrix correlations for high-dimensional data: the modified RV-coefficient. **Bioinformatics**, Oxford University Press (OUP), v. 25, n. 3, p. 401–405, dez. 2008. DOI: [10.1093/bioinformatics/btn634](https://doi.org/10.1093/bioinformatics/btn634). Disponível em:

<<https://doi.org/10.1093/bioinformatics/btn634>>.

SOMANI, Sulaiman S. *et al.* Characterization of Patients Who Return to Hospital Following Discharge from Hospitalization for COVID-19. **Journal of General Internal Medicine**, Springer Science e Business Media LLC, v. 35, n. 10, p. 2838–2844, ago. 2020. DOI: [10.1007/s11606-020-06120-6](https://doi.org/10.1007/s11606-020-06120-6). Disponível em:

<<https://doi.org/10.1007/s11606-020-06120-6>>.

- THOMAS, Mark E. *et al.* The definition of acute kidney injury and its use in practice. **Kidney International**, Elsevier BV, v. 87, n. 1, p. 62–73, jan. 2015. DOI: [10.1038/ki.2014.328](https://doi.org/10.1038/ki.2014.328). Disponível em: <<https://doi.org/10.1038/ki.2014.328>>.
- TOLLES, Juliana; MEURER, William J. Logistic Regression. **JAMA**, American Medical Association (AMA), v. 316, n. 5, p. 533, ago. 2016. DOI: [10.1001/jama.2016.7653](https://doi.org/10.1001/jama.2016.7653). Disponível em: <<https://doi.org/10.1001/jama.2016.7653>>.
- TORRALBA, Felipe *et al.* Os Escores HEART, TIMI e GRACE para Predicao de Eventos Cardiovasculares Adversos Maiores no Periodo de 30 Dias na Era de Troponina I de Alta Sensibilidade. **Arquivos Brasileiros de Cardiologia**, Sociedade Brasileira de Cardiologia, 2020. DOI: [10.36660/abc.20190206](https://doi.org/10.36660/abc.20190206). Disponível em: <<https://doi.org/10.36660/abc.20190206>>.
- TSALATSANIS, Athanasios; HOZO, Iztok; DJULBEGOVIC, Benjamin. Research synthesis of information theory measures of uncertainty: Meta-analysis of entropy and mutual information of diagnostic tests. **Journal of Evaluation in Clinical Practice**, Wiley, v. 27, n. 2, p. 246–255, set. 2020. DOI: [10.1111/jep.13475](https://doi.org/10.1111/jep.13475). Disponível em: <<https://doi.org/10.1111/jep.13475>>.
- VAN ROSSUM, Guido; DRAKE JR, Fred L. **Python tutorial**. [S. l.]: Centrum voor Wiskunde en Informatica Amsterdam, The Netherlands, 1995.
- VARGA, Csaba *et al.* ECG alterations suggestive of hyperkalemia in normokalemic versus hyperkalemic patients. **BMC Emergency Medicine**, Springer Science e Business Media LLC, v. 19, n. 1, mai. 2019. DOI: [10.1186/s12873-019-0247-0](https://doi.org/10.1186/s12873-019-0247-0). Disponível em: <<https://doi.org/10.1186/s12873-019-0247-0>>.
- VERBURG, Ilona Willempje Maria *et al.* Which models can I use to predict adult ICU length of stay? A systematic review. en. **Crit. Care Med.**, Ovid Technologies (Wolters Kluwer Health), v. 45, n. 2, e222–e231, fev. 2017.
- VINCENT, J L *et al.* The SOFA (Sepsis-related Organ Failure Assessment) score to describe organ dysfunction/failure. On behalf of the Working Group on Sepsis-Related Problems of the European Society of Intensive Care Medicine. **Intensive care medicine**, v. 22, n. 7, p. 707–10, jun. 1996.
- VIRANI, Salim S *et al.* Heart disease and stroke statistics-2020 update: A report from the American Heart Association. en. **Circulation**, Ovid Technologies (Wolters Kluwer Health), v. 141, n. 9, e139–e596, mar. 2020.

- VOHRA, Deepak. Using PostgreSQL Database. In: KUBERNETES Microservices with Docker. [S. l.]: Apress, 2016. p. 115–139. DOI: [10.1007/978-1-4842-1907-2_5](https://doi.org/10.1007/978-1-4842-1907-2_5). Disponível em: https://doi.org/10.1007/978-1-4842-1907-2_5.
- VUJOVIC, Zeljko. Classification Model Evaluation Metrics. **International Journal of Advanced Computer Science and Applications**, The Science e Information Organization, v. 12, n. 6, 2021. DOI: [10.14569/ijacsa.2021.0120670](https://doi.org/10.14569/ijacsa.2021.0120670). Disponível em: <https://doi.org/10.14569/ijacsa.2021.0120670>.
- WANG, Q Q *et al.* Overview of logistic regression model analysis and application. zh. **Zhonghua Yu Fang Yi Xue Za Zhi**, v. 53, n. 9, p. 955–960, set. 2019.
- WARNE, Russell T. *et al.* Statistical Methods Used in Gifted Education Journals, 2006-2010. **Gifted Child Quarterly**, SAGE Publications, v. 56, n. 3, p. 134–149, jun. 2012. DOI: [10.1177/0016986212444122](https://doi.org/10.1177/0016986212444122). Disponível em: <https://doi.org/10.1177/0016986212444122>.
- WHELEHAN, Dale F.; CONLON, Kevin C.; RIDGWAY, Paul F. Medicine and heuristics: cognitive biases and medical decision-making. **Irish Journal of Medical Science (1971 -)**, Springer Science e Business Media LLC, v. 189, n. 4, p. 1477–1484, mai. 2020. DOI: [10.1007/s11845-020-02235-1](https://doi.org/10.1007/s11845-020-02235-1). Disponível em: <https://doi.org/10.1007/s11845-020-02235-1>.
- WORLD MEDICAL ASSOCIATION. World Medical Association Declaration of Helsinki Ethical Principles for Medical Research Involving Human Subjects, 2013.
- XING, Eric P. *et al.* Strategies and Principles of Distributed Machine Learning on Big Data. **Engineering**, v. 2, n. 2, p. 179–195, 2016. ISSN 2095-8099. DOI: <https://doi.org/10.1016/J.ENG.2016.02.008>. Disponível em: <https://www.sciencedirect.com/science/article/pii/S2095809916309468>.
- YE, Siqin *et al.* Hospital Readmissions After Implementation of a Discharge Care Program for Patients with COVID-19 Illness. **Journal of General Internal Medicine**, Springer Science e Business Media LLC, v. 36, n. 3, p. 722–729, jan. 2021. DOI: [10.1007/s11606-020-06340-w](https://doi.org/10.1007/s11606-020-06340-w). Disponível em: <https://doi.org/10.1007/s11606-020-06340-w>.
- YUAN, Lin; KESAVAN, H.K. Minimum entropy and information measure. **IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)**, v. 28, n. 3, p. 488–491, 1998. DOI: [10.1109/5326.704595](https://doi.org/10.1109/5326.704595).

ZBILI, Mickael; RAMA, Sylvain. A Quick and Easy Way to Estimate Entropy and Mutual Information for Neuroscience. **Frontiers in Neuroinformatics**, Frontiers Media SA, v. 15, jun. 2021. DOI: [10.3389/fninf.2021.596443](https://doi.org/10.3389/fninf.2021.596443). Disponível em: [10.3389/fninf.2021.596443](https://doi.org/10.3389/fninf.2021.596443).

ZHANG, Xiaohua Douglas. Entropy for the Complexity of Physiological Signal Dynamics. In: HEALTHCARE and Big Data Management. [S. l.]: Springer Singapore, 2017. p. 39–53. DOI: [10.1007/978-981-10-6041-0_3](https://doi.org/10.1007/978-981-10-6041-0_3). Disponível em: [10.1007/978-981-10-6041-0_3](https://doi.org/10.1007/978-981-10-6041-0_3).

ZHANG, Zhongheng *et al.* Time-varying covariates and coefficients in Cox regression models. **Annals of Translational Medicine**, AME Publishing Company, v. 6, n. 7, p. 121–121, abr. 2018. DOI: [10.21037/atm.2018.02.12](https://doi.org/10.21037/atm.2018.02.12). Disponível em: [10.21037/atm.2018.02.12](https://doi.org/10.21037/atm.2018.02.12).

ZHENG, Chun-Hou *et al.* Nonnegative independent component analysis based on minimizing mutual information technique. en. **Neurocomputing**, Elsevier BV, v. 69, n. 7-9, p. 878–883, mar. 2006.

ZIAEIAN, Boback; FONAROW, Gregg C. Epidemiology and aetiology of heart failure. **Nature Reviews Cardiology**, Springer Science e Business Media LLC, v. 13, n. 6, p. 368–378, mar. 2016. DOI: [10.1038/nrcardio.2016.25](https://doi.org/10.1038/nrcardio.2016.25). Disponível em: [10.1038/nrcardio.2016.25](https://doi.org/10.1038/nrcardio.2016.25).

ZIMMERMAN, Jack E; KRAMER, Andrew A; KNAUS, William A. Changes in hospital mortality for United States intensive care unit admissions from 1988 to 2012. **Critical Care**, Springer Science e Business Media LLC, v. 17, n. 2, r81, 2013. DOI: [10.1186/cc12695](https://doi.org/10.1186/cc12695).

ZIMMERMAN, Jack E.; KRAMER, Andrew A.; MCNAIR, Douglas S. *et al.* Acute Physiology and Chronic Health Evaluation (APACHE) IV: Hospital mortality assessment for todays critically ill patients. **Critical Care Medicine**, Ovid Technologies (Wolters Kluwer Health), v. 34, n. 5, p. 1297–1310, mai. 2006. DOI: [10.1097/01.ccm.0000215112.84523.f0](https://doi.org/10.1097/01.ccm.0000215112.84523.f0). Disponível em: [10.1097/01.ccm.0000215112.84523.f0](https://doi.org/10.1097/01.ccm.0000215112.84523.f0).

APÊNDICE A - Tabelas de informações complementares

Tabela 5: Principais algoritmos de aprendizado de máquina supervisionado considerados para uso nos estudos com a base MIMIC-IV.

Algoritmo	Suporta relações não lineares?	Explicável	Aprende automaticamente	Hiperparâmetros
Regressão Logística	Não	Sim	Não	Solucionador, penalidade e força de regularização
Máquina de Vetor de Suporte	Sim	Não	Não	C (penalidade), kernel e hiperparâmetros específicos para o kernel
Árvore de Decisão	Sim	Parcialmente	Sim	α (custo de uma árvore)
<i>Random Forest</i>	Sim	Parcialmente	Sim	α , $mtry$ (número de variáveis aleatoriamente selecionadas como candidatos em cada árvore)
Rede Bayesiana	Não	Não	Não	Suavização
<i>Dependency Network</i>	Não	Não	Não	Regularização
Rede Neural	Sim	Não	Sim	Arquitetura de rede, função de erro, funções de ativação
<i>K-Nearest Neighbors</i>	Sim	Não	Não	Número de vizinhos

Tabela 6: Principais medidas de avaliação de desempenho para modelos de classificação. VP: Taxa de verdadeiro positivo; VN: Taxa de verdadeiro negativo; FP: Taxa de falso positivo; FN: Taxa de falso negativo; AUROC: área sob a curva *Receiver Operating Characteristics*.

Medida	Cálculo	Significado
Acurácia	$\frac{VP+VN}{VP+VN+FP+FN}$	Razão entre a quantidade de acertos e o total de entradas.
Precisão	$\frac{VP}{VP+FP}$	Razão entre a quantidade de exemplos classificados corretamente como positivos e o total de exemplos classificados como positivos.
Revocação (Sensibilidade)	$\frac{VP}{VP+FN}$	Razão entre a quantidade de exemplos classificados corretamente como positivos e a quantidade de exemplos que são de fato positivos.
F-score	$\frac{2*VP}{2*VP+FP+FN}$	Média harmônica calculada com base na precisão e na revocação.
Especificidade	$\frac{VN}{FP+VN}$	Razão entre a quantidade de exemplos classificados corretamente como negativos e o total de exemplos classificados como negativos.
AUROC	Área bidimensional abaixo da curva ROC	Indica visualmente a relação entre falsos positivos e verdadeiro positivos em diferentes limiares de classificação.

APÊNDICE B - Tabelas de análise

Tabela 7: Dados selecionados da base MIMIC-IV.

Dado	Tipo do dado	Ausentes (%)	Informação clínica
subject_id	int64	0,00%	Identificador do paciente
hadm_id	int64	0,00%	Identificador da admissão
age	int64	0,00%	Idade
dischtime	object	0,00%	Data e hora de alta da unidade
admittime	object	0,00%	Data e hora de admissão na unidade
ethnicity	object	0,00%	Etnia
stay_id	int64	0,00%	Identificador da internação
gender	object	0,00%	Gênero
heart_rate	float64	0,13%	Frequência cardíaca
resp_rate	float64	0,23%	Frequência respiratória
mbp	float64	0,24%	Pressão arterial média
spo2	float64	0,25%	Saturação periférica de oxigênio
sbp	float64	0,48%	Pressão arterial sistólica
dbp	float64	0,49%	Pressão arterial diastólica
creatinine	float64	0,85%	Creatinina
bun	float64	0,85%	Nitrogênio ureico no sangue
hematocrit	float64	0,92%	Hematócrito
wbc	float64	0,99%	Leucócitos
hemoglobin	float64	1,00%	Hemoglobina
platelets	float64	1,00%	Plaquetas
chloride	float64	1,13%	Cloro
sodium	float64	1,14%	Sódio
bicarbonate	float64	1,15%	Bicarbonato
potassium	float64	1,17%	Potássio
aniongap	float64	1,23%	<i>Anion Gap</i> sanguíneo
glucose	float64	2,70%	Glicemia
temperature	float64	3,23%	Temperatura
urineoutput_24hr	float64	3,86%	Débito urinário em 24 horas
weight	float64	6,80%	Peso corporal
calcium	float64	8,15%	Cálcio
pt	float64	11,13%	Tempo de atividade da protrombina
inr	float64	11,13%	Razão normalizada internacional
ptt	float64	11,79%	Tempo de tromboplastina parcial ativado
abs_lymphocytes	float64	40,52%	Linfócitos
abs_eosinophils	float64	40,56%	Eosinófilos
abs_neutrophils	float64	40,56%	Neutrófilos
abs_monocytes	float64	40,56%	Monócitos
abs_basophils	float64	40,56%	Basófilos
po2	float64	43,93%	Pressão de oxigênio no sangue
baseexcess	float64	43,93%	Excesso de base no sangue
ph	float64	43,93%	pH sanguíneo
ast	float64	48,69%	Aspartato aminotransferase
alt	float64	48,97%	Aminotransferase alanina
alp	float64	49,21%	Fosfatase alcalina
bilirubin_total	float64	49,30%	Bilirrubinas totais
lactate	float64	52,50%	Ácido láctico
height	float64	54,05%	Altura
albumin	float64	61,88%	Albumina
pao2fio2ratio	float64	62,09%	Relação PaO2/FiO2 no sangue
ck_mb	float64	65,60%	Creatina quinase fração mb
ck_cpk	float64	69,65%	Creatinofosfoquinase
ldh	float64	70,15%	Lactato desidrogenase
troponin_t	float64	73,88%	Troponina T
fibrinogen	float64	74,16%	Fibrinogênio
imm_granulocytes	float64	78,89%	Granulócitos
bands	float64	86,70%	Bastões
metas	float64	87,61%	Metamielócitos
atyps	float64	87,81%	Atipias linfocitárias
ntprobnp	float64	89,34%	NT-proBNP
crp	float64	92,73%	Proteína-C Reatina
amylase	float64	93,67%	Amilase
bilirubin_direct	float64	95,69%	Bilirrubina direta
bilirubin_indirect	float64	95,84%	Bilirrubina indireta
total_protein	float64	97,90%	Proteínas totais
globulin	float64	98,82%	Globulina
ggt	float64	99,20%	Gama glutamil transferase
d_dimer	float64	99,79%	D-Dímero
thrombin	float64	99,82%	Trombina

Tabela 8: Análise de dados *outliers*, com comparação entre os intervalos considerados normais para as variáveis do estudo, além da faixa de valores impossíveis.

Variável	Intervalo normal	Intervalo observado	Unidade
Idade	0-120+	18-102	anos
Frequência cardíaca	60-100	36-295	bpm
Pressão arterial sistólica	< 120	46-335	mmHg
Pressão arterial diastólica	< 80	20-228	mmHg
Pressão arterial média	< 60	26-299	mmHg
Frequência respiratória	12-20	6-69	irpm
Saturação periférica de oxigênio	95-100	56-100	%
Débito urinário em 24 horas	800-2000	0-10700	mL
Temperatura corporal	35,9-37,0	26,6-41,4	°
Hemoglobina	M: 13,5-17,5; F: 12-16	5,2-20,9	g/dL
Leucócitos	3,9-10,7	0,1-160,2	10 ⁹ cel/L
Plaquetas	150-400	14,0-1137,0	10 ⁹ cel/L
Sódio	136-145	97-175	mg/dL
Potássio	3,5-5	1,6-10,0	mg/dL
Cálcio	8,5-10,2	2,6-79,2	mg/dL
Cloro	96-106	73-137	mEq/L
BUN	6-24	2-263	mg/dL
Creatinina	0,1-1,2	0,1-23,2	mg/dL
Anion gap	4,0-12,0	5-54	mmol/L
Glicose	70-99	51-412	mg/dL
TAP	11-13,5	8,4-150	segundos
INR	0,8-1,1	0,8-21,8	-
PTT	25-25	17,7-150	segundos
Ck-mb	5-25	1-575	UI/L
Troponina T	0-0,01	0-39,7	ng/mL
NT-proBNP	< 125	20-68347,0	pg/mL

Tabela 9: Perfil epidemiológico, clínico e laboratorial da população do estudo. DP: Desvio-padrão; T-CICU: Tempo de internação na CICU; T-H: Tempo de internação hospitalar; FC: frequência cardíaca; FR: frequência respiratória; PAS: pressão arterial sistólica; PAD: pressão arterial diastólica; PAM: pressão arterial média; SO2: oximetria de pulso; TAP: tempo de protrombina; PTT: tempo de tromboplastina; NT-ProBNP: fragmento N-terminal do peptídeo natriurético tipo B.

	Ausentes	Todos	Sobreviventes	Óbitos	valor-p	Teste	DP
Número de pacientes		8035	6468	1567			
Idade, mediana [Q1,Q3]	0	72,1 [60,9,82,4]	71,0 [59,9,81,7]	76,6 [66,0,84,9]	0,001	Teste T	0,340
Sexo masculino, n (%)	0	4600 (57,2)	3722 (57,5)	878 (56,0)	0,290	Qui-quadrado	0,031
T-CICU, mediana [Q1,Q3]	0	1,9 [1,1,3,4]	1,9 [1,0,3,1]	2,2 [1,1,4,6]	0,001	Teste T	0,284
T-H, mediana [Q1,Q3]	0	6,0 [3,2,10,8]	5,9 [3,3,10,3]	6,6 [3,0,12,4]	0,290	Teste T	0,082
FC	8	99,4 (23,1)	98,1 (22,5)	104,7 (24,4)	0,001	Teste T	0,282
FR	14	28,0 (6,2)	27,6 (5,9)	29,7 (6,7)	0,001	Teste T	0,323
PAS	63	143,4 (24,3)	144,3 (23,5)	139,5 (27,1)	0,001	Teste T	-0,190
PAD	63	87,4 (18,7)	88,0 (18,2)	84,9 (20,5)	0,001	Teste T	-0,162
PAM	14	102,7 (24,1)	103,1 (22,7)	101,1 (29,0)	0,010	Teste T	-0,078
SO2	32	99,2 (1,6)	99,2 (1,2)	99,1 (2,7)	0,014	Teste T	-0,083
Temperatura	212	37,1 (0,7)	37,1 (0,6)	37,1 (1,0)	0,004	Teste T	-0,097
Glicemia	244	312,2 (11327,6)	335,2 (12601,9)	215,4 (118,5)	0,451	Teste T	-0,013
Creatinina	47	1,8 (1,6)	1,6 (1,5)	2,4 (1,9)	0,001	Teste T	0,447
Hematócrito	60	35,6 (6,3)	35,9 (6,3)	34,3 (6,5)	0,001	Teste T	-0,245
Leucometria	72	12,1 (7,0)	11,6 (6,4)	14,3 (9,0)	0,001	Teste T	0,344
Plaquetometria	68	236,3 (101,6)	236,6 (99,1)	234,9 (111,6)	0,567	Teste T	-0,017
Sódio	51	139,2 (4,6)	139,2 (4,2)	138,8 (5,9)	0,008	Teste T	-0,082
Potássio	50	4,6 (0,9)	4,6 (0,8)	4,9 (1,0)	0,001	Teste T	0,330
TAP	721	18,2 (12,2)	17,4 (10,9)	21,9 (15,8)	0,001	Teste T	0,336
PTT	763	59,3 (41,6)	57,3 (40,5)	67,3 (44,8)	0,001	Teste T	0,233
Troponina T	3880	1,5 (2,9)	1,4 (2,7)	1,7 (3,4)	0,031	Teste T	0,083
NT-proBNP	6541	8775,8 (11823,4)	7726,5 (10917,6)	13113,9 (14219,7)	0,001	Teste T	0,425

Tabela 10: Tempo médio de sobrevivência de diferentes grupos populacionais baseados em comorbidades. Um mesmo paciente pode pertencer a mais de um grupo. *: *valor – p* menor que 0,05.

Grupo	Tempo (dias)	IC 95%
Todos os pacientes (n = 8035)	28	26-30*
Infarto do Miocárdio		
Presente (n = 1741)	27	23-29*
Ausente	27	24-32*
Doença Arterial Periférica		
Presente (n = 710)	25	21-28
Ausente	28	25-31
Doença Cerebrovascular		
Presente (n = 340)	24	21-29*
Ausente	27	25-32*

Tabela 11: Riscos proporcionais pelo modelo de Regressão de Cox para o conjunto inicial de variáveis, que não incluiu exames laboratoriais. FC: frequência cardíaca; FR: frequência respiratória; PAS: pressão arterial sistólica; PAD: pressão arterial diastólica; PAM: pressão arterial média; SO2: oximetria de pulso.

Modelo	Cox			Weibull		
	HR	IC 95 %	valor – p	HR	IC 95 %	valor – p
Variável						
Idade	1,03	1,02 - 1,03	0,05	0,98	0,97 - 0,98	0,05
Sexo masculino	1,00	0,89 - 1,13	0,97	1,00	0,90 - 1,10	0,92
Doença cerebrovascular	1,17	0,95 - 1,43	0,14	0,86	0,73 - 1,02	0,09
Infarto do Miocárdio	1,22	1,04 - 1,43	0,01	0,85	0,74 - 0,96	0,05
Doença vascular periférica	1,04	0,87 - 1,25	0,64	0,96	0,83 - 1,11	0,60
PAD	0,99	0,99 - 1,00	0,01	1,00	1,00 - 1,01	0,05
FC	1,01	1,00 - 1,01	0,05	1,00	0,99 - 1,00	0,05
PAM	1,00	1,00 - 1,00	0,18	1,00	1,00 - 1,00	0,19
FR	1,02	1,02 - 1,03	0,05	0,98	0,97 - 0,99	0,05
PAS	1,00	1,00 - 1,00	0,16	1,00	1,00 - 1,00	0,20
SO2	0,92	0,88 - 0,96	0,05	1,07	1,04 - 1,11	0,05
Temperatura	0,78	0,72 - 0,85	0,05	1,23	1,14 - 1,32	0,05
Intercepto				1,20	1,15 - 1,25	0,05

Tabela 12: Riscos proporcionais, com variáveis de exames laboratoriais, pelo modelo de Cox ajustado para idade, gênero e comorbidades. TAP: Tempo de Protrombina; PTT: Tempo de Tromboplastina. *: *valor – p* < 0,05.

Variável	HR	IC 95%
Glicose	1,00	1,00-1,00
Creatinina	1,07	1,03-1,10*
Hematócrito	1,07	1,03-1,11*
Leucócitos	1,02	1,01-1,02*
Plaquetas	1,00	1,00-1,00
Sódio	1,00	0,99-1,01
Potássio	1,22	1,14-1,30*
TAP	1,00	0,99-1,02
PTT	1,00	1,00-1,00
Troponina T	1,04	0,96-1,13
NT-ProBNP	1,00	1,00-1,00

APÊNDICE C - Questionário de validação do sistema

Seção 1 de 4

System Usability Survey

This is a short survey about the [Mortality Prediction Tool](#) that has been shared with you.

This research is conducted by Pedro Gemal Lanzieri as part of a master degree from Universidade Federal Fluminense, oriented by Flávio Luis Seixas.

We kindly ask you to answer those questions in plain English if you do accept the terms below. It is expected not to take more than 3 minutes to answer this form.

I consent that I am a healthcare provider or researcher and that I understand that the results presented by the tool are experimental and hereby should not be used as clinical decision guidance. I also allow my answers to be shared anonymously.

Yes

No

Adicionar opção ou [adicionar "Outro"](#)

Caixas de seleção

Obrigatória

Seção 2 de 4

Demographic questions ✕ ⋮

Descrição (opcional)

How old are you? *

Texto de resposta curta

For how many years have you been practicing medicine? *

Texto de resposta curta

What is your area of expertise in medicine (medical specialty)? *

Texto de resposta curta

Seção 3 de 4

Graded questions



We request that you answer these questions after using the system. A suggestion of a use case would be the following: pick 3 options from the selection menu provided in the tool, fill the values with plausible information (regardless if they are in the normal range or not), accept the checkbox and submit the information to get a simple estimation of mortality prediction and some other analysis.

I think that I would like to use this system frequently *

	1	2	3	4	5	
Strongly disagree	<input type="radio"/>	Strongly agree				

I found the system unnecessarily complex *

	1	2	3	4	5	
Strongly disagree	<input type="radio"/>	Strongly agree				

⋮

I thought there was too much information in the system *

	1	2	3	4	5	
Strongly disagree	<input type="radio"/>	Strongly agree				

I would imagine that most people would learn to use the system very quickly *

	1	2	3	4	5	
Strongly disagree	<input type="radio"/>	Strongly agree				

I found the system very difficult to be used *

	1	2	3	4	5	
Strongly disagree	<input type="radio"/>	Strongly agree				

I felt very confident using the system *

	1	2	3	4	5	
Strongly disagree	<input type="radio"/>	Strongly agree				

I needed to learn a lot of things before I could get going with this system *

	1	2	3	4	5	
Strongly disagree	<input type="radio"/>	Strongly agree				

Seção 4 de 4

Open questions ✕ ⋮

Descrição (opcional)

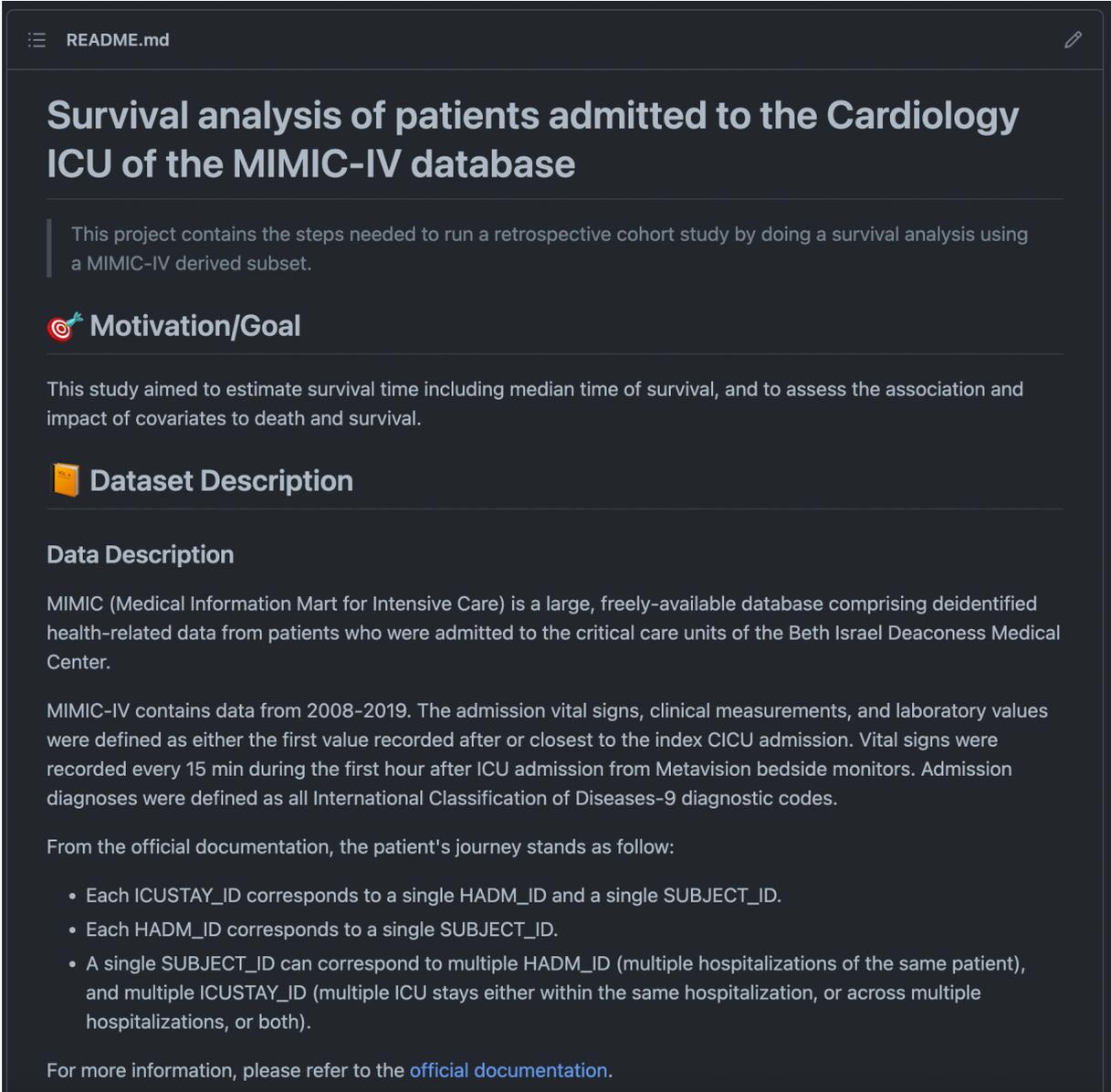
What feature or function did you like the most about the system? *

Texto de resposta longa

Do you have any suggestions or see any opportunities for improvement in the system? Please *
give many details as possible.

Texto de resposta longa

APÊNDICE D - Documentos auxiliares do repositório



☰ README.md ✎

Survival analysis of patients admitted to the Cardiology ICU of the MIMIC-IV database

This project contains the steps needed to run a retrospective cohort study by doing a survival analysis using a MIMIC-IV derived subset.

Motivation/Goal

This study aimed to estimate survival time including median time of survival, and to assess the association and impact of covariates to death and survival.

Dataset Description

Data Description

MIMIC (Medical Information Mart for Intensive Care) is a large, freely-available database comprising deidentified health-related data from patients who were admitted to the critical care units of the Beth Israel Deaconess Medical Center.

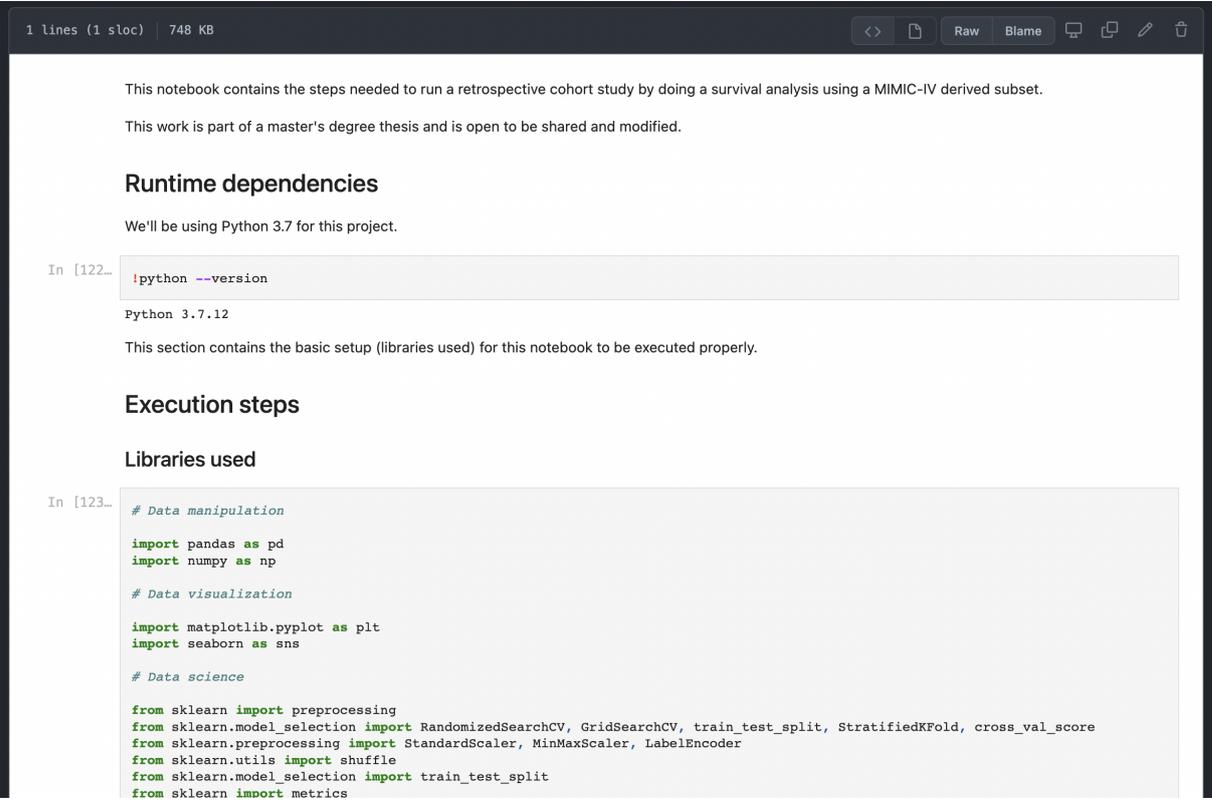
MIMIC-IV contains data from 2008-2019. The admission vital signs, clinical measurements, and laboratory values were defined as either the first value recorded after or closest to the index CICU admission. Vital signs were recorded every 15 min during the first hour after ICU admission from Metavision bedside monitors. Admission diagnoses were defined as all International Classification of Diseases-9 diagnostic codes.

From the official documentation, the patient's journey stands as follow:

- Each ICUSTAY_ID corresponds to a single HADM_ID and a single SUBJECT_ID.
- Each HADM_ID corresponds to a single SUBJECT_ID.
- A single SUBJECT_ID can correspond to multiple HADM_ID (multiple hospitalizations of the same patient), and multiple ICUSTAY_ID (multiple ICU stays either within the same hospitalization, or across multiple hospitalizations, or both).

For more information, please refer to the [official documentation](#).

Figura 13: Representação do arquivo de ajuda disponível no repositório.



The screenshot shows a Jupyter Notebook interface with a dark theme. At the top, it displays '1 lines (1 sloc) | 748 KB' and navigation icons. The main content area contains the following text:

This notebook contains the steps needed to run a retrospective cohort study by doing a survival analysis using a MIMIC-IV derived subset.

This work is part of a master's degree thesis and is open to be shared and modified.

Runtime dependencies

We'll be using Python 3.7 for this project.

In [122..] `!python --version`

Python 3.7.12

This section contains the basic setup (libraries used) for this notebook to be executed properly.

Execution steps

Libraries used

In [123..]

```
# Data manipulation

import pandas as pd
import numpy as np

# Data visualization

import matplotlib.pyplot as plt
import seaborn as sns

# Data science

from sklearn import preprocessing
from sklearn.model_selection import RandomizedSearchCV, GridSearchCV, train_test_split, StratifiedKFold, cross_val_score
from sklearn.preprocessing import StandardScaler, MinMaxScaler, LabelEncoder
from sklearn.utils import shuffle
from sklearn.model_selection import train_test_split
from sklearn import metrics
```

Figura 14: Reprodutibilidade da pesquisa usando o notebook.

prediction_full.py

This is the core code of the MIMIC-IV based prediction tool. Currently we support a single view application, divided in the following sections:

1. **Model** - The data handling (jump to section in [prediction_full.py](#))
2. **View** - The user interface (jump to section in [prediction_full.py](#))
3. **Controller** - The services that are runned by the application (jump to section in [prediction_full.py](#))

importing libraries

streamlit configuration

model layer

```
import streamlit as st
import os
import pickle
import pandas as pd
import time
import math
import numpy as np
import sksurv

st.set_page_config(
    page_title="MIMIC-IV based prediction tool",
    page_icon="🏠",
    layout="wide"
)

def clear_multi():
    st.session_state.multiselect = []
    return

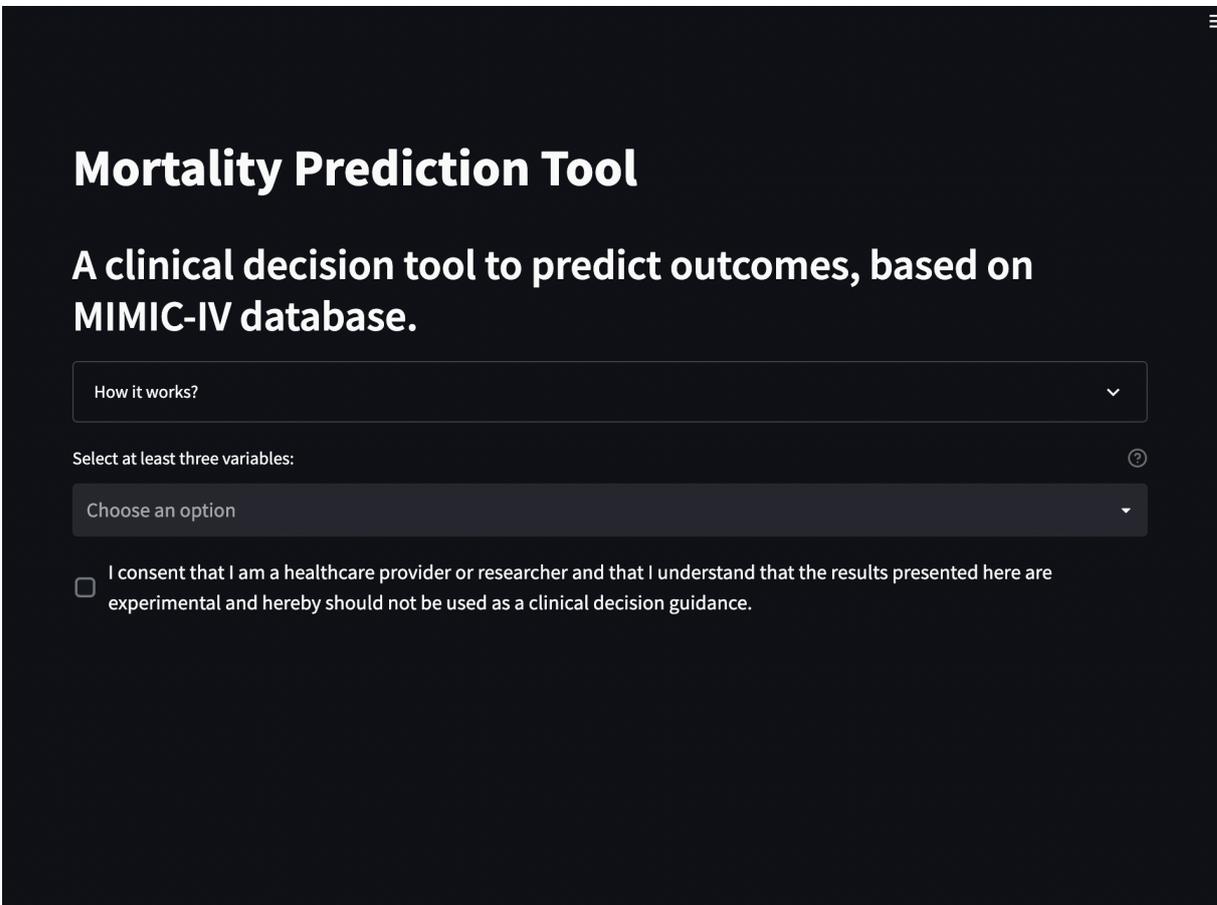
@st.cache(allow_output_mutation=True, show_spinner=False)
def get_data():
    path = "../data/df_cicu_imputed_data.csv"
    return pd.read_csv(path)

def train_model(features_provided, input_data):
    df = get_data()

    from sklearn.model_selection import train_test_split
```

Figura 15: Documentação técnica do código principal do sistema.

APÊNDICE E - Documentos representativos do sistema



The screenshot shows the initial screen of the 'Mortality Prediction Tool'. The title 'Mortality Prediction Tool' is prominently displayed in white text on a dark background. Below the title, a subtitle reads 'A clinical decision tool to predict outcomes, based on MIMIC-IV database.' The interface includes a dropdown menu with the text 'How it works?' and a downward arrow. Below this, there is a prompt 'Select at least three variables:' followed by a question mark icon. Underneath is another dropdown menu with the text 'Choose an option' and a downward arrow. At the bottom, there is a consent checkbox and the text: 'I consent that I am a healthcare provider or researcher and that I understand that the results presented here are experimental and hereby should not be used as a clinical decision guidance.'

Figura 16: Tela inicial do sistema.

Mortality Prediction Tool

A clinical decision tool to predict outcomes, based on MIMIC-IV database.

How it works? ^

This tool is developed for theoretical evaluation of a prediction model, based on a subset of cardiac patients from MIMIC-IV database.

A cohort of patients with known diagnosis of Heart Failure was used to train and test a prediction model.

This prediction was developed based on MIMIC-IV clinical data and shall not be used as a single guidance for patient decision making.

Keep in mind the restrictions of generability of a machine learning model.

Please input at least three variables from the patient in order to get estimates about intra-hospitalar mortality.

Select at least three variables: ?

HR × DBP × Creatinine × ⓧ

Heart rate (per minute): ?

80 - +

Diastolic blood pressure (mmHg): ?

150 - +

Creatinine (mg/dL): ?

2.12 - +

I consent that I am a healthcare provider or researcher and that I understand that the results presented here are experimental and hereby should not be used as a clinical decision guidance.

Figura 17: Exemplo de preenchido do sistema com os requisitos mínimos.

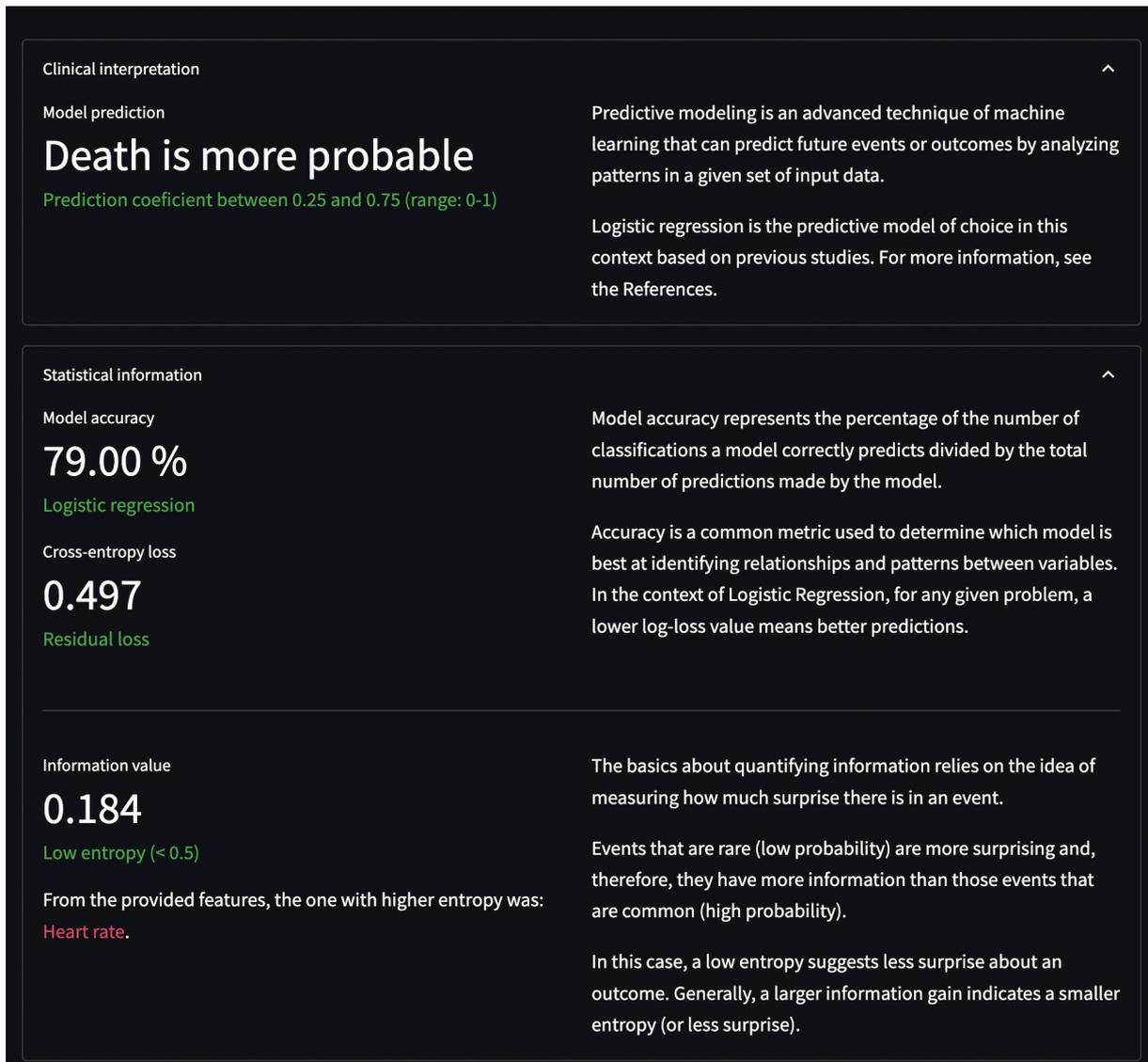
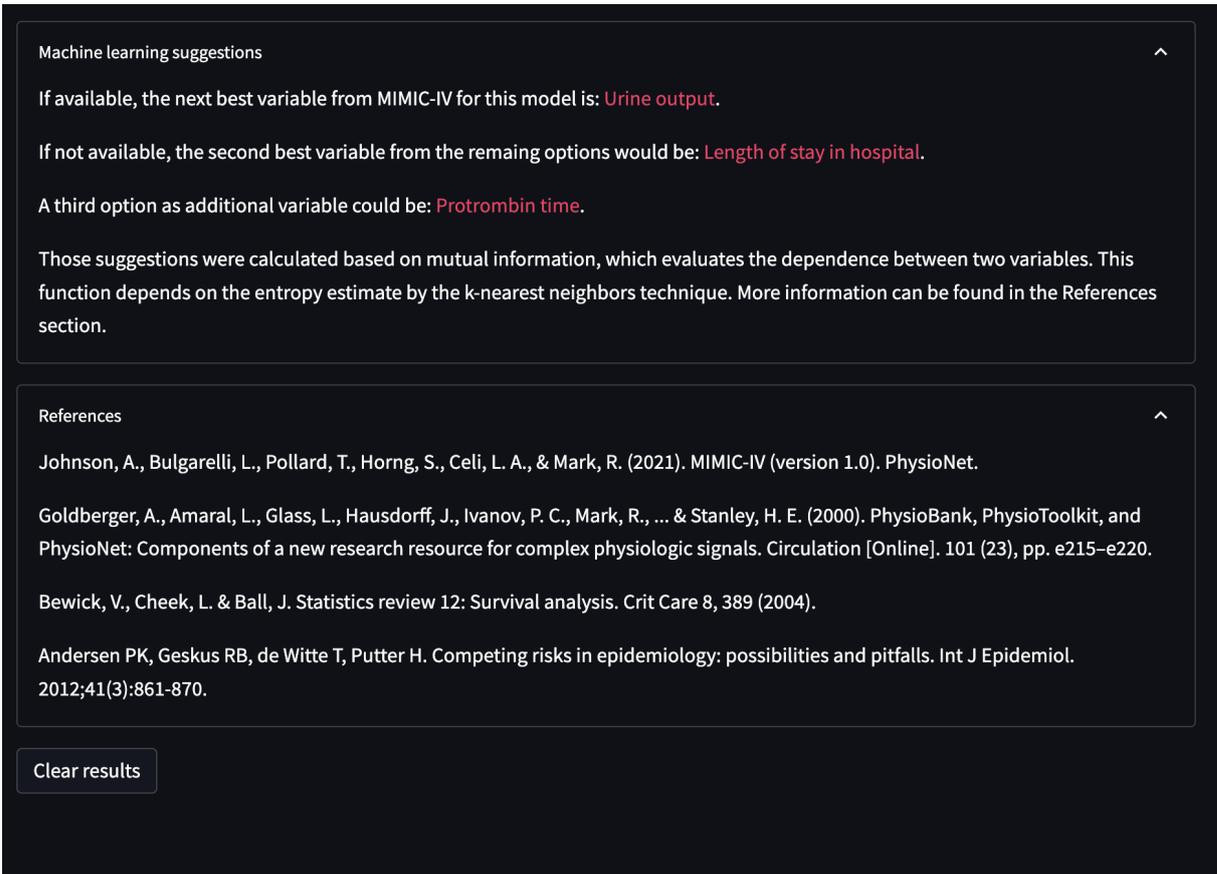


Figura 18: Primeiro e segundo bloco de informações.



Machine learning suggestions ^

If available, the next best variable from MIMIC-IV for this model is: **Urine output**.

If not available, the second best variable from the remaining options would be: **Length of stay in hospital**.

A third option as additional variable could be: **Protrombin time**.

Those suggestions were calculated based on mutual information, which evaluates the dependence between two variables. This function depends on the entropy estimate by the k-nearest neighbors technique. More information can be found in the References section.

References ^

Johnson, A., Bulgarelli, L., Pollard, T., Horng, S., Celi, L. A., & Mark, R. (2021). MIMIC-IV (version 1.0). PhysioNet.

Goldberger, A., Amaral, L., Glass, L., Hausdorff, J., Ivanov, P. C., Mark, R., ... & Stanley, H. E. (2000). PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals. *Circulation [Online]*. 101 (23), pp. e215–e220.

Bewick, V., Cheek, L. & Ball, J. *Statistics review 12: Survival analysis*. Crit Care 8, 389 (2004).

Andersen PK, Geskus RB, de Witte T, Putter H. Competing risks in epidemiology: possibilities and pitfalls. *Int J Epidemiol*. 2012;41(3):861-870.

Clear results

Figura 19: Terceiro e quarto bloco de informações.

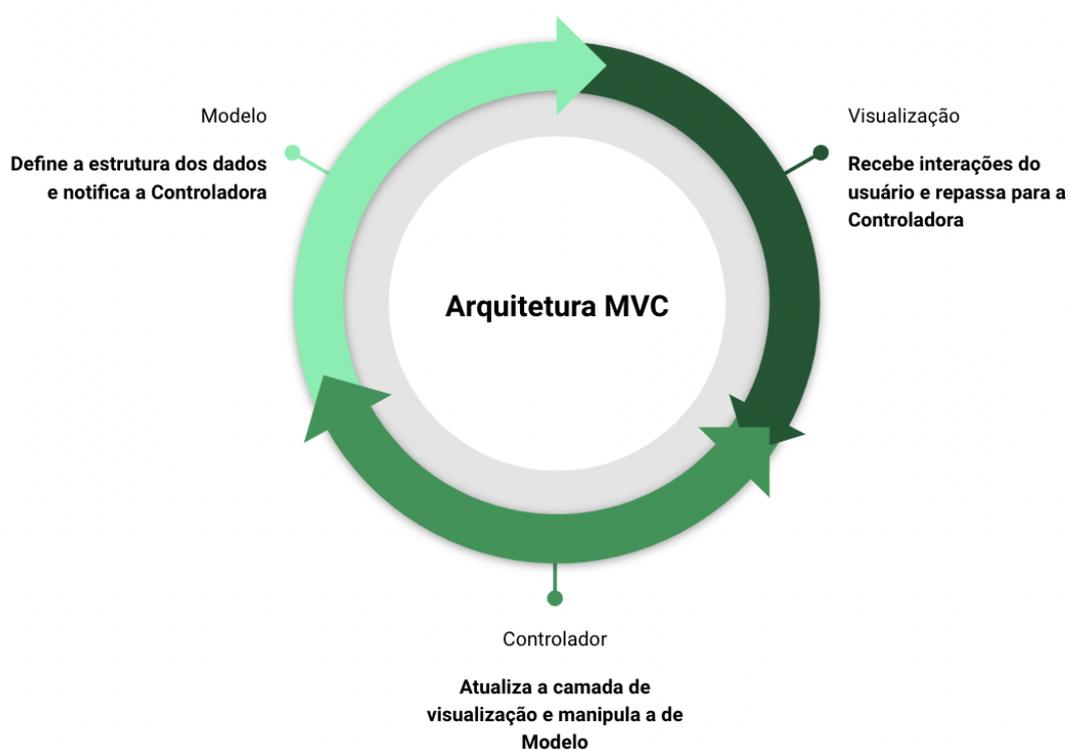


Figura 20: Arquitetura do código desenvolvido conforme o padrão MVC

APÊNDICE F - Figuras complementares aos estudos

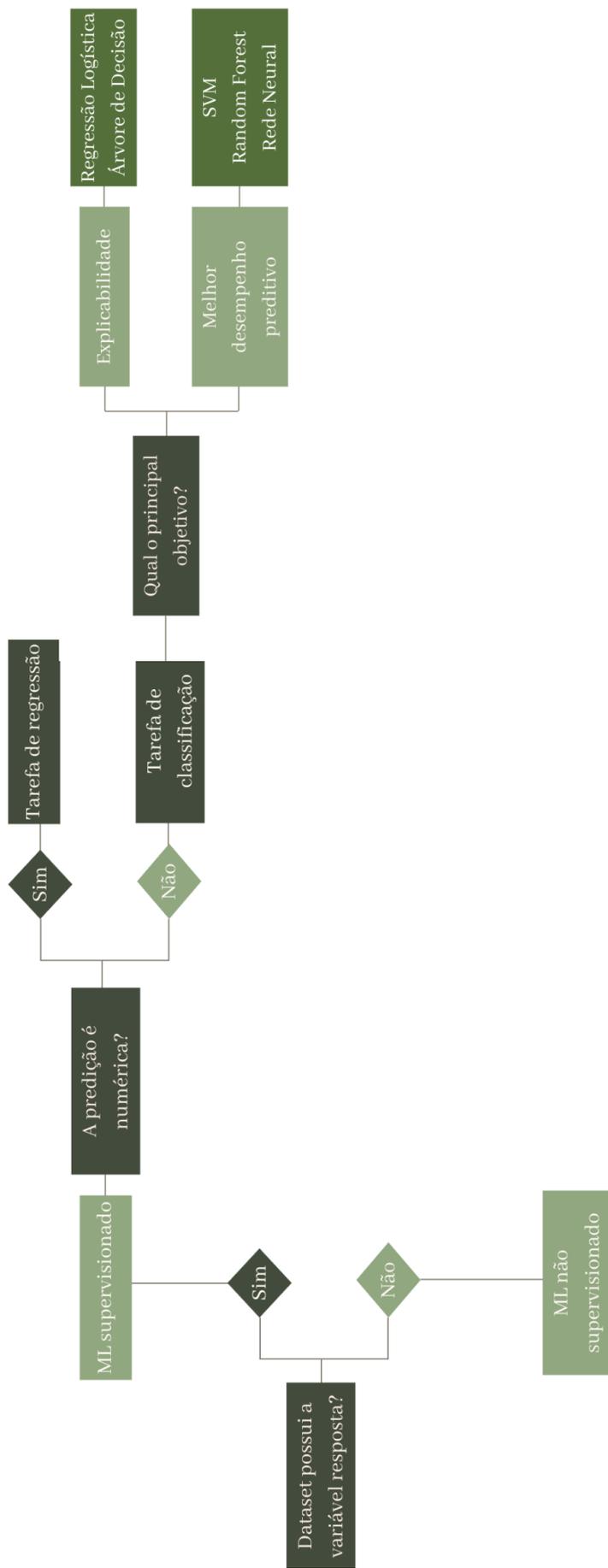


Figura 21: Estudos realizados com a base MIMIC-IV.