PONTIFÍCIA UNIVERSIDADE CATÓLICA DE GOIÁS
ESCOLA POLITÉCNICA E DE ARTES
GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

ANÁLISE DE SOBREVIVÊNCIA E APRENDIZADO DE MÁQUINA NA PREVISÃO DO TEMPO DE ESPERA PARA TRANSPLANTE RENAL: UMA ABORDAGEM OTIMIZADA

ELISEU RODRIGUES GUIMARÃES

GOIÂNIA
2024

ELISEU RODRIGUES GUIMARÃES

ANÁLISE DE SOBREVIVÊNCIA E APRENDIZADO DE MÁQUINA NA PREVISÃO DO TEMPO DE ESPERA PARA TRANSPLANTE RENAL: UMA ABORDAGEM OTIMIZADA

Trabalho de Conclusão de Curso apresentado à Escola Politécnica e de Artes, da Pontifícia Universidade Católica de Goiás, como parte dos requisitos para obtenção do título de Bacharel em Ciência da Computação.

Orientador(a):

Prof.(a) Dra. Maria José Pereira Dantas

GOIÂNIA
2024

ELISEU RODRIGUES GUIMARÃES

ANÁLISE DE SOBREVIVÊNCIA E APRENDIZADO DE MÁQUINA NA PREVISÃO DO TEMPO DE ESPERA PARA TRANSPLANTE RENAL: UMA ABORDAGEM OTIMIZADA

Trabalho de Conclusão de Curso aprovado em sua forma final pela Escola Politécnica e de Artes da Pontifícia Universidade Católica de Goiás, para obtenção do título de Bacharel em Ciência da Computação, em: _____/_____/_____.

_____

Orientadora: Profa. Dra. Maria José Pereira Dantas

_____

Prof. Dr. José Elmo de Menezes

_____

Prof. Me. Daniel Correa da Silva

GOIÂNIA

2024

AGRADECIMENTOS

Agradeço, em primeiro lugar, a Deus, por me conceder força, sabedoria e resiliência para superar os desafios dessa jornada. Aos meus pais, expresso meu profundo agradecimento pelo indescritível apoio, que para mim foi imensurável. Tenho por eles um amor incondicional, me sustentaram em momentos difíceis, quando muitas vezes me sentia sozinho, mas nunca estive.

Meus agradecimentos não poderiam passar despercebido à Profa. Dra. Maria José Pereira Dantas, que foi minha maior inspiração na área da pesquisa científica, desde as aulas de Probabilidade e Estatística e durante os diversos Desafios de Ciência de Dados quanto no desenvolvimento deste Trabalho de Conclusão de Curso. Foram longos períodos sob sua e orientação. De suas diversas qualidades, destaco seu zelo com o lado humano, seu capricho em se dedicar seriamente a pesquisa. Indispensavelmente seu conhecimento foi e será fundamental para o meu crescimento acadêmico e pessoal, incentivando-me a buscar soluções para a sociedade por meio do conhecimento acadêmico. Que deve ser o compromisso de usar o nosso conhecimento a serviço da vida.

“Tenho medo da graça que passa sem que eu perceba.”
Sto. Agostinho

RESUMO

A crescente demanda por transplantes renais e a limitada disponibilidade de órgãos tornam o gerenciamento eficiente das listas de espera um desafio crítico para o Sistema Nacional de Transplantes (SNT) no Brasil. Este estudo tem como objetivo desenvolver e comparar modelos preditivos para estimar o tempo de espera até o transplante renal, combinando técnicas de análise de sobrevivência e aprendizado de máquina, como Regressão de Cox, Random Survival Forest (RSF) e XGBoost. Utilizando uma coorte retrospectiva de dados do Sistema de Alocação de Órgãos do Estado de São Paulo (SP-OAS)/Brasil, entre 2000 e 2017, a pesquisa analisa variáveis demográficas, clínicas e imunológicas, incluindo idade, comorbidades, tempo em diálise, frequência de HLA-DR, -B e -A, e painel de reatividade de anticorpos (cPRA). O desempenho dos modelos foi avaliado por meio de métricas robustas, como o C-index, Brier Score e curvas ROC, para medir a acurácia e a capacidade de discriminação. Os resultados incluem a identificação das variáveis mais relevantes que impactam a probabilidade de transplante e a validação de um modelo preditivo otimizado. A ferramenta desenvolvida fornece estimativas personalizadas do tempo de espera, auxiliando médicos e pacientes na tomada de decisões informadas e na gestão de expectativas. A calculadora preditiva resultante é uma aplicação prática, e promove uma alocação mais eficiente dos recursos e contribuindo para a redução da mortalidade em lista de espera. Além disso, o estudo ressalta o papel do aprendizado de máquina como uma abordagem inovadora e promissora na previsão de eventos de saúde, com potencial para impactar positivamente o sistema de transplantes no Brasil.

Palavras-chave: Transplante renal, Lista de espera, Previsão de tempo de espera, Análise de sobrevivência, Aprendizado de máquina, Modelos preditivos, Sistema Nacional de Transplantes, Gestão de saúde, Inteligência artificial, Otimização de recursos.

ABSTRACT

The increasing demand for kidney transplants and the limited availability of organs make efficient management of waiting lists a critical challenge for the National Transplant System (SNT) in Brazil. This study aims to develop and compare predictive models to estimate waiting time for kidney transplantation, combining survival analysis and machine learning techniques, such as Cox Regression, Random Survival Forest (RSF) and XGBoost. Using a retrospective cohort of data from the Organ Allocation System of the State of São Paulo (SP-OAS)/Brazil, between 2000 and 2017, the research analyzes demographic, clinical and immunological variables, including age, comorbidities, time on dialysis, frequency of HLA-DR, -B and -A, and panel reactivity antibody (cPRA). The performance of the models was evaluated through robust metrics, such as the C-index, Brier Score and ROC curves, to measure accuracy and discrimination capacity.The expected results include the identification of the most relevant variables that impact the probability of transplantation and the validation of an optimized predictive model. The developed tool is expected to provide personalized estimates of waiting time, assisting physicians and patients in making informed decisions and managing expectations. The resulting predictive calculator can be implemented as a practical application, promoting a more efficient allocation of resources and contributing to the reduction of mortality on the waiting list. In addition, the study highlights the role of machine learning as an innovative and promising approach in predicting health events, with the potential to positively impact the transplant system in Brazil.

Keywords: *Kidney transplant, Waiting list, Waiting time prediction, Survival analysis, Machine learning, Predictive models, National Transplant System, Health management, Artificial intelligence, Resource optimization.*

LISTA DE ILUSTRAÇÕES

Figura 01 – *Censura à Direita (Right Censoring)*

Figura 02 – *Censura à Direita Database (Right Censoring)*

Figura 03 – *Random Survival Forest (RSF)*

Figura 04 – *Curva Kaplan-Meier para grupo de idade e diálise*

Figura 05 – *Função de sobrevivência para o grupo de idade e diálise*

Figura 06 – *Amostragem Bootstrap*

Figura 07 – *Função de sobrevivência para o grupo de idade, diálise e meses*

Figura 08 – *Xgboost Survival*

Figura 09 – *Variáveis importantes*

Figura 10 – *Gráfico de curvas ROC*

Figura 11 – *Mapa de Tópicos Relevantes em ML e Análise de Sobrevivência*

Figura 12 – *Mapa de Países que pesquisam o tema*

Figura 13 – *Distribuição de Razões de Remoção*

Figura 14 – *Variáveis mais importantes (Random Survival Forest)*

Figura 15 – *Variáveis mais importantes (XGBoost)*

Figura 16 – *Distribuição de Censura*

Figura 17 – *Censura à Direita Database (Right Censoring)*

Figura 18 – *Distribuição do Tempo de Sobrevivência*

Figura 19 – *Distribuição de age_cat (Categoria de Idade)*

Figura 20 – *Distribuição de sex (Sexo)*

Figura 21 – *Distribuição de diabete*

Figura 22 – *Distribuição das Condições Subjacentes*

Figura 23 – *Distribuição dos Tipos Sanguíneos*

Figura 24 – *Distribuição do cPRA (Percent Panel Reactive Antibodies)*

Figura 25 – *Distribuição da Quantidade de Transfusões*

Figura 26 – *Dados antes da limpeza*

Figura 27 – *Funções de sobrevivência para categoria de idade*

Figura 28 – *Funções de sobrevivência para tipo sanguíneo*

Figura 29 – *Gráfico de barras com a importância das variáveis*

Figura 30 – *Gráfico de curvas ROC*

Figura 31 – *Formulário da Calculadora de Tempo de Espera*

Figura 32 – *Tela de resultado da Calculadora de Tempo de Espera*

LISTA DE TABELAS

Tabela 01 – *Característica dos modelos*

Tabela 02 – *Portifólio de artigos selecionados na base Web of Science*

Tabela 03 – *Importância das principais variáveis do modelo Regressão de COX*

Tabela 04 – *Importância das principais variáveis do modelo RSF*

Tabela 05 – *Importância das principais variáveis do modelo XGBoost*

Tabela 06 – *Tabela de resultado*

Tabela 06 – *Tabela de resultados por categoria de previsão*

LISTA DE SIGLAS

SNT – *Sistema Nacional de Transplante*
RSF – *Random Survival Forest (Floresta de Sobrevivência Aleatória)*
SP-OAS – *Sistema de Alocação de Órgãos do Estado de São Paulo*
HLA – *Human Leucocyte Antigen (Antígeno Leucocitário Humano)*
C-Index – *Concordance Index*
ROC curve – *Receiver Operating Characteristic Curve (Curva Característica de Operação do Receptor)*
ML – *Machine Learning*

1. INTRODUÇÃO 13
1.1. CONTRIBUIÇÕES 15
1.2. OBJETIVOS ESPECÍFICOS 15
1.3. JUSTIFICATIVA 16
1.4. ORGANIZAÇÃO DO TRABALHO 17
2. REFERENCIAL TEÓRICO 19
2.1. INTRODUÇÃO AO TRANSPLANTE RENAL 19
2.1.1. Importância do transplante renal como tratamento para doença renal crônica 19
2.1.2. Panorama do transplante renal no Brasil e no mundo 19
2.1.3. Dados do Registro Brasileiro de Transplantes (ABTO) 20
2.2. GESTÃO DE LISTA DE ESPERA EM TRANSPLANTE 20
2.2.1. Desafios da alocação de órgãos 21
2.2.2. Modelos tradicionais de gestão e seus limites 21
2.3. ANÁLISE DAS VARIÁVEIS 23
2.3.1. Variáveis Demográficas 23
2.3.2. Variáveis Clínicas 23
2.3.3. Variáveis de Desfecho 24
2.3.4. Variáveis Calculadas 25
2.4. ANÁLISE DE SOBREVIVÊNCIA 25
2.4.1. Conceito e Aplicações em Saúde 26
2.4.2. Modelo de Riscos Proporcionais de Cox 26
2.4.3. Regressão de Cox: Fundamentos e Uso na Previsão do Tempo de Espera 27
2.4.4. Limitações da Análise Tradicional de Sobrevivência 28
2.4.5. Censura dos Dados em Análise de Sobrevivência 28
2.4.6. Os métodos Convencionais de Análise de Sobrevivência: 32
2.5. Aprendizado de Máquina (Machine Learning) 32
2.5.1. Introdução e Definição 33
2.5.2. Diferenças entre Métodos Tradicionais e Aprendizado de Máquina 33
2.5.3. Benefícios do Aprendizado de Máquina em Transplantes Renais 34
2.6. MODELOS AVANÇADOS DE PREDIÇÃO 34
2.6.1. Random Survival Forest (RSF): Características e Aplicações 34
2.6.2. XGBoost: Fundamentos e Desempenho em Predições Clínicas 42
2.7. Ferramentas e Métricas de Avaliação de Modelos 45
2.7.1. Ferramentas e Métricas de Avaliação de Modelos 45
2.7.2. Pontuação de Brier: 46
2.7.3. Curvas ROC e AUC: 46

2.8. Validação Cruzada e Otimização de Hiperparâmetros 47
2.8.1. Validação Cruzada k-Fold: 47
2.8.2. Otimização de Hiperparâmetros: 48
2.8.3. Considerações sobre Interpretabilidade e Aplicabilidade Prática 48
2.9. Estudos Correlatos.....49
2.9.1. Aplicações de Análise de Sobrevivência e Aprendizado de Máquina em Transplantes 49
2.9.2. Experiências Internacionais e Estudos Nacionais Relevantes 50
2.10. Limitações e Lacunas na Literatura Atual 50
3. PROCEDIMENTOS METODOLÓGICOS 52
3.1. BASE DE DADOS.....53
3.2. PRÉ-PROCESSAMENTO DE DADOS 54
3.3. MATERIAIS UTILIZADOS 56
3.3.1. Hardware 56
3.3.2. Software 57
3.4. ÉTICA E LIMITAÇÕES 58
3.5. PESQUISA EXPLORATÓRIA NA BASE CIENTÍFICA WEB OF SCIENC....58
3.6. ANÁLISE DA REVISÃO EXPLORATÓRIA DA LITERATURA: MÉTODOS DE PRÉ-PROCESSAMENTO DE DADOS E MÉTODOS DE IMPUTAÇÃO 60
4. RESULTADOS E DISCUSSÃO 64
4.1. ANÁLISE E ESTUDO DOS DADOS 64
4.1.1. Composição e Estrutura do Dataset 64
4.1.2. Importância das Variáveis 65
4.1.3. Censura e Sobrevivência 67
4.1.4. Distribuição Demográfica e Clínica.....68
4.2. PRÉ-PROCESSAMENTO DOS DADOS 72
4.3. Tratamento de Valores Ausentes 73
4.3.1. Transformação e Codificação de Variáveis.....74
4.3.2. Normalização e Visualização de Impacto 74
4.3.3. Resultados do Pré-Processamento 74
4.4. RESULTADOS DOS MODELOS DE SOBREVIVÊNCIA.....75
4.4.1. Regressão de Cox 75
4.4.2. Random Survival Forest (RSF).....77
4.4.3. XGBoost 78
4.5. MODELO PROPOSTO.....80
4.6. PROTÓTIPO DA CALCULADORA.....82
5. CONSIDERAÇÕES FINAIS 85

1. INTRODUÇÃO

A demanda por transplantes renais tem aumentado de forma significativa, enquanto a oferta de órgãos disponíveis segue insuficiente para atender às necessidades dos pacientes em lista de espera. Esse cenário representa um desafio crítico para a gestão eficiente do Sistema Nacional de Transplantes (SNT), exigindo soluções que otimizem o uso dos recursos disponíveis e melhorem os desfechos clínicos (ABTO, 2023). Atualmente, o Brasil ocupa o primeiro lugar no mundo em número de transplantes renais realizados pelo sistema público de saúde (SUS). Contudo, em julho de 2020, mais de 30 mil brasileiros ainda aguardavam por um transplante renal, enquanto o número de procedimentos anuais com doadores falecidos não ultrapassava 5 mil, evidenciando um crescente déficit entre oferta e demanda (Sapiertein et al., 2021). Como consequência, os pacientes necessitam realizar testes e procedimentos recorrentes a cada 2–3 anos para se manterem na lista ativa, o que impõe um fardo econômico substancial ao sistema de saúde (Gómez et al., 2018).

No último ano, foram realizados 6.047 transplantes renais no Brasil, totalizando mais de 60 mil transplantes desse órgão desde 2013. Apesar desse avanço, a lista de espera para a cirurgia ainda é significativa, com 32.862 pacientes adultos e 335 pediátricos aguardando por um transplante ao final de 2023 (ABTO, 2023). Essa disparidade entre oferta e demanda não apenas compromete os resultados clínicos dos pacientes, mas também representa um desafio logístico e econômico para a saúde pública, uma vez que longos períodos de espera podem agravar as condições dos pacientes e aumentar os custos relacionados a cuidados paliativos e procedimentos de manutenção.

Prever o tempo de espera para o transplante é uma estratégia fundamental para o planejamento e otimização dos cuidados pré-transplante. Tal previsão permite a priorização adequada dos recursos, a realização de avaliações completas em candidatos com maior probabilidade de transplante e a identificação de grupos menos favorecidos, promovendo maior equidade no processo de alocação de órgãos (Hart et al., 2016; Su & Zenios, 2004). Além disso, fornecer estimativas claras e personalizadas pode contribuir significativamente para a qualidade de vida dos

pacientes e suas famílias, facilitando o gerenciamento de expectativas e a adesão ao tratamento (Dew et al., 2007).

Tradicionalmente, o tempo em lista de espera tem sido expresso como a mediana do tempo de espera, uma métrica que não reflete adequadamente os riscos de mortalidade ou remoção da lista (Hart et al., 2016). Nesse sentido, modelos preditivos mais avançados, combinando técnicas de análise de sobrevivência e aprendizado de máquina (ML), têm demonstrado maior potencial para lidar com as complexidades envolvidas na previsão do tempo de espera. Enquanto a regressão de Cox, um modelo semi-paramétrico amplamente utilizado, identifica os fatores de risco associados ao tempo até um evento (como transplante ou óbito), abordagens baseadas em aprendizado de máquina, como Random Survival Forest (RSF) e XGBoost, são capazes de capturar relações não lineares entre variáveis, melhorando a precisão das previsões (Ishwaran et al., 2008; Katzman et al., 2018; Sapiertein et al., 2021).

Estudos recentes destacam o potencial do aprendizado de máquina na área de transplantes. Ahady Dolatsara et al. (2020) propuseram um modelo de aprendizado de máquina de dois estágios para prever probabilidades de sobrevivência em transplantes cardíacos, demonstrando ganhos significativos na alocação de recursos clínicos. De forma semelhante, Sapiertein et al. (2021) aplicaram um modelo preditivo para estimar o tempo de espera em transplantes renais no Brasil, com desempenho satisfatório na validação interna (C-index = 0,70), destacando variáveis como idade, tempo em diálise e frequência de HLA. Esses trabalhos evidenciam a capacidade do aprendizado de máquina em combinar múltiplos preditores para gerar estimativas mais precisas e personalizadas, superando as limitações dos métodos tradicionais (Coorey et al., 2020).

Embora o Sistema Nacional de Transplantes tenha avançado significativamente na organização e gestão do processo de doação e alocação de órgãos, a complexidade inerente ao processo — envolvendo variáveis demográficas, clínicas e logísticas — reforça a necessidade de modelos preditivos robustos e personalizados. Diante desse cenário, a aplicação combinada de técnicas de análise de sobrevivência e aprendizado de máquina representa uma abordagem inovadora, capaz de otimizar

a gestão das listas de espera, reduzir o tempo até o transplante e minimizar os riscos de mortalidade.

Dessa forma, o presente estudo tem como objetivo desenvolver e validar um modelo preditivo otimizado para estimar o tempo de espera para transplante renal no Brasil, utilizando dados do Sistema de Alocação de Órgãos do Estado de São Paulo (SP-OAS)/Brasil. Serão comparadas técnicas tradicionais, como a regressão de Cox, com algoritmos avançados de aprendizado de máquina, como RSF e XGBoost. Espera-se que os resultados contribuam para aprimorar a tomada de decisões clínicas, promover uma distribuição mais equitativa dos recursos e fornecer informações valiosas para pacientes e seus familiares, melhorando a eficiência e a transparência do processo de transplantes no país.

1.1. CONTRIBUIÇÕES

Este estudo visa desenvolver e validar um modelo preditivo robusto para estimar o tempo de espera para transplante renal, combinando técnicas avançadas de análise de sobrevivência e aprendizado de máquina. A aplicação de métodos como regressão de Cox, XGBoost e Random Forest, utilizando dados do Sistema Nacional de Transplantes (SNT), tem o potencial de otimizar a alocação de recursos, melhorar o gerenciamento das listas de espera e oferecer informações mais precisas para médicos e pacientes.

1.2. OBJETIVOS ESPECÍFICOS

1. **Identificar variáveis relevantes:** Selecionar e analisar as variáveis demográficas, clínicas e imunológicas mais significativas que influenciam o tempo de espera, como idade, tempo em diálise, frequência de HLA, e cPRA, a partir dos dados disponíveis.
2. **Modelar e comparar métodos preditivos:** Aplicar e comparar o desempenho de métodos tradicionais de análise de sobrevivência (regressão de Cox) com algoritmos modernos de aprendizado de máquina, como XGBoost e Random Forest Survival (RSF), utilizando métricas como C-index, Brier Score e curvas ROC.
3. **Otimizar e validar os modelos:** Realizar ajuste de hiperparâmetros e

validação cruzada para garantir a robustez e generalização do modelo final, assegurando sua aplicabilidade prática em diferentes contextos.

4. **Implementar uma calculadora preditiva:** Desenvolver uma ferramenta prática, intuitiva e acessível, baseada no modelo otimizado, que permita estimar o tempo de espera para transplante de forma personalizada. Além de direcionada ao uso clínico, a calculadora será projetada para atender também à população em geral, oferecendo informações claras e acessíveis. O objetivo é não apenas facilitar a tomada de decisões informadas por médicos e pacientes, mas também promover maior transparência no processo de transplantes, sensibilizando a sociedade sobre a importância da doação de órgãos. Dessa forma, busca-se reduzir a evasão de doadores e pacientes, contribuindo para uma gestão mais eficiente das listas de espera e para o fortalecimento das políticas públicas de saúde.

1.3. JUSTIFICATIVA

A presente pesquisa se justifica pela necessidade urgente de aprimorar a gestão da lista de espera para transplante cardíaco no Brasil. Embora estudos anteriores tenham explorado o potencial do aprendizado de máquina (ML) em transplantes renais e hepáticos (Coorey et al., 2020; Kantidakis et al., 2020), a aplicação dessas técnicas no contexto do transplante cardíaco ainda é incipiente, com poucos estudos publicados e resultados ainda inconclusivos (Ahady Dolatsara et al., 2020). Este estudo busca preencher essa lacuna, investigando a combinação de análise de sobrevivência e aprendizado de máquina para desenvolver modelos preditivos mais robustos e personalizados, utilizando dados do mundo real do Sistema Nacional de Transplantes (SNT).

A relevância desta pesquisa reside em seu potencial para:

- **Aprimorar a tomada de decisões clínicas:** Fornecendo aos médicos informações mais precisas sobre o tempo de espera, auxiliando na priorização de pacientes, na otimização da alocação de órgãos e no ajuste de terapias medicamentosas.
- **Empoderar pacientes e familiares:** Oferecendo uma estimativa

personalizada do tempo de espera, permitindo um melhor planejamento de vida, gerenciamento de expectativas e redução da ansiedade associada ao processo de transplante.

- **Otimizar a alocação de recursos:** Ajudando a identificar pacientes com maior risco de mortalidade durante a espera, permitindo uma alocação mais eficiente dos órgãos disponíveis e reduzindo o tempo de espera para os pacientes mais críticos.
- **Avançar o conhecimento científico:** Explorando a aplicação de técnicas de aprendizado de máquina em conjunto com a análise de sobrevivência no contexto do transplante cardíaco, contribuindo para o desenvolvimento de modelos preditivos mais eficazes e personalizados.
- **Desenvolver uma ferramenta inovadora:** A criação de uma calculadora de tempo de espera acessível e intuitiva representa um avanço significativo, facilitando a comunicação entre médicos e pacientes e promovendo a tomada de decisões compartilhadas.

Em suma, este estudo busca não apenas aprimorar a previsão do tempo de espera para transplante cardíaco, mas também gerar um impacto positivo na vida dos pacientes, na prática clínica e na gestão da lista de espera. A combinação de análise de sobrevivência e aprendizado de máquina, utilizando dados do mundo real, oferece uma oportunidade única para o desenvolvimento de modelos preditivos mais precisos, personalizados e clinicamente relevantes.

1.4. ORGANIZAÇÃO DO TRABALHO

Este trabalho está estruturado em cinco capítulos, conforme descrito a seguir:

Capítulo 1: Introdução: Neste capítulo inicial, são apresentadas as motivações que fundamentaram o desenvolvimento deste estudo, destacando a relevância do transplante renal no Brasil e os desafios relacionados à gestão das listas de espera. São discutidas as lacunas existentes na literatura e o potencial das técnicas de análise de sobrevivência e aprendizado de máquina para aprimorar os modelos preditivos. Além disso, são descritos os objetivos gerais e específicos do trabalho, bem como as contribuições esperadas.

Capítulo 2 - Referencial Teórico: Este capítulo fornece uma base teórica detalhada, abordando os conceitos essenciais relacionados ao transplante renal, à análise de sobrevivência e ao aprendizado de máquina. São discutidos os principais modelos utilizados na previsão de eventos clínicos, como a regressão de Cox, o Random Survival Forest (RSF) e o XGBoost. Também são apresentados estudos recentes que exploram aplicações semelhantes, fornecendo contexto e suporte para o desenvolvimento deste trabalho.

Capítulo 3 – Procedimentos Metodológicos: Neste capítulo, são descritos os métodos adotados para a realização do estudo. Inclui-se uma explicação detalhada sobre os dados utilizados, oriundos do Sistema de Alocação de Órgãos do Estado de São Paulo (SP-OAS), bem como o processo de pré-processamento e seleção de variáveis. Também são descritos os procedimentos de modelagem e validação dos modelos preditivos, além da implementação da calculadora preditiva. A abordagem ética para o uso de dados sensíveis também é destacada.

Capítulo 4 - Resultados e Discussão: Os resultados obtidos a partir da aplicação dos modelos preditivos são apresentados e discutidos neste capítulo. São avaliados o desempenho dos modelos em termos de métricas como C-index, Brier Score e curvas ROC. Também são explorados os impactos práticos da ferramenta desenvolvida, incluindo sua aplicabilidade clínica e social, com ênfase na sensibilização da população e na promoção da doação de órgãos.

Capítulo 5 - Considerações Finais: Este capítulo apresenta as conclusões do trabalho, destacando as principais contribuições científicas e sociais. São discutidas as limitações do estudo e sugestões para trabalhos futuros, incluindo a possibilidade de expansão para outros tipos de transplantes ou melhorias no uso de aprendizado de máquina para predição clínica. Além disso, reforça-se o impacto positivo esperado da calculadora preditiva no fortalecimento das políticas públicas de saúde e na redução da evasão de doadores e pacientes.

2. REFERENCIAL TEÓRICO

2.1. INTRODUÇÃO AO TRANSPLANTE RENAL

O transplante renal é amplamente reconhecido como o tratamento mais eficaz para pacientes em estágio terminal de doença renal crônica. Ele não apenas restaura as funções fisiológicas dos rins, mas também melhora significativamente a qualidade de vida dos pacientes e reduz os custos associados a tratamentos como a diálise contínua (Silva et al., 2021). Com a crescente prevalência de insuficiência renal crônica em todo o mundo, o transplante de rim tornou-se um elemento essencial na saúde pública global, especialmente em países que possuem sistemas bem estabelecidos para alocação e distribuição de órgãos.

2.1.1. Importância do transplante renal como tratamento para doença renal crônica

Os rins desempenham um papel crítico na regulação do equilíbrio de fluidos e eletrólitos, na eliminação de resíduos metabólicos e na manutenção da pressão arterial. Em pacientes com insuficiência renal crônica, essas funções são severamente comprometidas, resultando em uma redução drástica da qualidade de vida e no aumento do risco de complicações fatais. O transplante renal, ao contrário da diálise, permite a recuperação plena das funções renais, reduzindo a mortalidade e promovendo uma reintegração mais completa do paciente à vida cotidiana (Dew et al., 2007; Silva et al., 2021). Além disso, estudos demonstram que o transplante é mais econômico a longo prazo em comparação com a manutenção em diálise, especialmente em sistemas públicos de saúde (Brown et al., 2012).

2.1.2. Panorama do transplante renal no Brasil e no mundo

O Brasil é um dos líderes globais na realização de transplantes renais, com o Sistema Único de Saúde (SUS) desempenhando um papel fundamental na universalização do acesso a esse tratamento. Segundo o Registro Brasileiro de Transplantes (RBT), apenas em 2023 foram realizados 6.047 transplantes de rim no país, totalizando mais de 60 mil procedimentos desde 2013 (ABTO, 2023). No entanto, a lista de

espera continua extensa, com 32.862 pacientes adultos e 335 pediátricos aguardando por um transplante no final de 2023. Essa disparidade entre a oferta e a demanda reflete uma realidade global: enquanto o número de pacientes dependentes de transplante cresce, a disponibilidade de doadores permanece limitada (Gómez et al., 2018; BRASIL, 2024).

Mundialmente, iniciativas como a implementação de sistemas de alocação baseados em inteligência artificial e aprendizado de máquina têm sido adotadas para lidar com a complexidade da gestão de listas de espera e melhorar a eficiência dos transplantes (Kyritsis & Deriaz, 2019).

2.1.3. Dados do Registro Brasileiro de Transplantes (ABTO)

De acordo com o Registro Brasileiro de Transplantes (ABTO), o transplante renal continua sendo a cirurgia de transplante de órgãos mais realizada no país, representando cerca de 70% dos transplantes de órgãos sólidos em 2023. Apesar desse avanço, a taxa de doadores efetivos ainda é insuficiente para atender à demanda crescente, o que evidencia a importância de políticas públicas voltadas para a conscientização da sociedade sobre a doação de órgãos (ABTO, 2023).

Além disso, os dados do RBT mostram que existem variações regionais significativas na disponibilidade de transplantes, refletindo as desigualdades no acesso e na infraestrutura de saúde entre diferentes estados. Esses desafios reforçam a necessidade de modelos preditivos que não apenas ajudem na priorização clínica, mas também contribuam para promover maior equidade na distribuição de órgãos.

2.2. GESTÃO DE LISTA DE ESPERA EM TRANSPLANTE

A gestão eficiente das listas de espera para transplantes é um dos maiores desafios enfrentados pelos sistemas de saúde em todo o mundo. Com o crescimento da demanda por órgãos e a limitada disponibilidade de doadores, a priorização de pacientes e a distribuição equitativa de órgãos tornaram-se questões centrais para os

programas de transplantes. Esses desafios exigem abordagens inovadoras e integradas, combinando técnicas avançadas de gestão e tecnologias modernas, como aprendizado de máquina, para melhorar a eficiência e equidade do processo.

2.2.1. Desafios da alocação de órgãos

O processo de alocação de órgãos envolve múltiplos fatores, como compatibilidade biológica (HLA, grupo sanguíneo), tempo em lista de espera, e condições clínicas dos pacientes. Esses critérios são essenciais para garantir que os órgãos sejam atribuídos de forma eficaz e ética. No entanto, a crescente disparidade entre a oferta e a demanda de órgãos agrava a complexidade desse processo, aumentando os tempos de espera e o risco de mortalidade em lista (Boxma et al., 2011; ABTO, 2023).

Além disso, as desigualdades regionais no acesso a infraestruturas de saúde e na conscientização sobre a doação de órgãos dificultam a equidade no sistema de transplantes, como observado em diversos estados brasileiros. Programas que dependem exclusivamente de métodos manuais ou rígidos de priorização enfrentam dificuldades em atender às necessidades específicas de cada paciente, limitando sua eficiência (Kyritsis & Deriaz, 2019).

2.2.2. Modelos tradicionais de gestão e seus limites

Os modelos tradicionais de gestão de listas de espera baseiam-se principalmente em abordagens estáticas e regras predefinidas de priorização, como tempo de espera acumulado e pontuações baseadas em critérios médicos. Embora amplamente utilizados, esses métodos têm limitações significativas:

- **Rigidez nas regras de alocação:** Muitas vezes, os critérios fixos não consideram a evolução dinâmica das condições clínicas dos pacientes ou a complexidade das interações biológicas (Su & Zenios, 2004).
- **Falta de personalização:** Os modelos tradicionais têm dificuldade em lidar com as necessidades individuais dos pacientes, como sensibilização imunológica ou condições específicas de saúde (Brown

et al., 2012).

- **Ineficiência na integração de múltiplas variáveis:** Métodos manuais ou baseados em regras rígidas não conseguem processar e priorizar informações complexas de forma eficiente, resultando em atrasos na alocação e maior taxa de mortalidade (Drekic et al., 2015).

Com essas limitações, há uma necessidade crescente de sistemas mais dinâmicos e adaptativos, capazes de integrar grandes volumes de dados e gerar decisões baseadas em predições precisas.

2.2.3. Impacto da previsão do tempo de espera na equidade e eficiência

A capacidade de prever o tempo de espera de forma precisa tem um impacto significativo tanto na equidade quanto na eficiência dos sistemas de transplantes. Modelos preditivos avançados permitem:

- **Priorização mais justa:** Ao identificar pacientes em maior risco de mortalidade, é possível reduzir desigualdades na distribuição de órgãos e salvar vidas (Hart et al., 2016; Silva et al., 2021).
- **Redução de custos:** Prever tempos de espera possibilita o planejamento mais eficaz de exames e avaliações pré-transplante, evitando procedimentos desnecessários e otimizando o uso de recursos (Gómez et al., 2018).
- **Transparência e engajamento social:** Ferramentas preditivas acessíveis ao público promovem maior conscientização sobre a importância da doação de órgãos, reduzindo a evasão de doadores e aumentando a confiança no sistema de saúde (ABTO, 2023).

Estudos demonstram que abordagens baseadas em aprendizado de máquina, como Random Survival Forest e XGBoost, superam modelos tradicionais ao integrar variáveis complexas e gerar previsões mais confiáveis (Coorey et al., 2020; Kyritsis & Deriaz, 2019).

2.3. ANÁLISE DAS VARIÁVEIS

A base de dados utilizada neste estudo contém uma ampla gama de variáveis demográficas, clínicas e contextuais, essenciais para modelar o tempo de espera e os desfechos em listas de transplantes renais. Abaixo, apresentamos uma descrição detalhada dessas variáveis, categorizadas por sua natureza e papel no modelo.

2.3.1. Variáveis Demográficas

1. **Identificação e Data de Registro:**

    - **Id:** Identificação única do paciente.
    - **date:** Data de registro na lista de espera.

2. **Idade e Gênero:**

    - **age_at_list_registration:** Idade no momento do registro.
    - **age_cat:** Faixa etária categorizada:
    - **sex:** Sexo do paciente.

3. **Raça:**

    - **race:** Cor/raça autodeclarada:

4. **Sub-região de Registro:**

    - **subregion:** Centro de referência de registro

2.3.2. Variáveis Clínicas

1. **Doenças de Base:**

    - **underline_disease:** Doença renal subjacente
        - Diabetes, HAS (Hipertensão Arterial Sistêmica), GNC (Glomerulonefrite Crônica), Pielonefrite, Outras.
    - **diabetes:** Indicador de diabetes.

2. **Tempo em Diálise:**

    - **time_on_Dialysis:** Tempo em meses.

3. **Compatibilidade e Antígenos HLA:**

    - **cPRA:** Percentual de anticorpos reativos.
    - **cPRA_cat:** Categoria do cPRA:
    - **HLA_A1, HLA_A2, HLA_B1, HLA_B2, HLA_DR1, HLA_DR2:** Frequência de antígenos HLA
    - **DR_00, B_00, A_00:** Classificação de homozigose ou heterozigose para os correspondentes.

4. **Histórico de Transfusões e Gestação:**

    - **number_transfusion:** Número de transfusões anteriores.
    - **cPRA_cat:** Categoria do cPRA
    - **number_gestation:** Número de gestações

5. **Histórico de Transplantes:**

    - **prior_transplant:** Histórico de trasplantes anteriores.
    - **number_prior_transplant:** Número de transplantes prévios

6. **Sorologias s:**

    - **chagas, anti.HBc, anti.HCV, agHBs:** Sorologias para doenças infecciosas

2.3.3. Variáveis de Desfecho

1. **Estado na Lista:**

    - **patient_still_on_list:** Indica se o paciente ainda se encontra na lista de espera.
    - **removed_list:** Indica se o paciente foi removido da

lista.

- **razon_removed:** Razão para a remoção

2. **Óbito e Transplante:**

- **death:** Indica se houve óbito.
- **Transplant:** Indica se houve transplante.
- **Transplant_Y_N:** Transplante categorizado.
- **Time_Tx:** Tempo até o transplante.

3. **Censura e Tempo:**

- **Censura:** Indicador de censura.
- **time:** Tempo de seguimento na lista.
- **event:** Tipo de evento observado.

2.3.4. Variáveis Calculadas

1. **Prioridades e Eventos Futuros:**

- **priorization:** Indica se o paciente foi priorizado.
- **X36MthsTx:** Indicador binário de transplante dentro de 36 meses.

Relevância das Variáveis no Contexto do Modelo

As variáveis descritas desempenham papéis cruciais na construção de modelos preditivos para o tempo de espera e a probabilidade de transplante. A inclusão de fatores clínicos e demográficos, como idade, cPRA, tempo em diálise e histórico de doenças, permite desenvolver modelos mais precisos e personalizados.

2.4. ANÁLISE DE SOBREVIVÊNCIA

A análise de sobrevivência é uma abordagem estatística utilizada para estudar o tempo até a ocorrência de um evento de interesse. No contexto da saúde, essa técnica é amplamente empregada para analisar desfechos clínicos, como o tempo de espera para transplante de órgãos, tempo até a recaída de uma doença, ou a sobrevida de pacientes após determinado tratamento. Para os pacientes em listas de

espera de transplantes, o evento pode ser o transplante, óbito, ou mesmo a remoção da lista devido à piora clínica ou recuperação (Silva et al., 2021).

2.4.1. Conceito e Aplicações em Saúde

A análise de sobrevivência tem como principal objetivo estimar a probabilidade de um indivíduo "sobreviver" sem experimentar o evento de interesse em um intervalo de tempo específico. Ela é essencial em áreas como oncologia, cardiologia, e transplantes, onde o tempo de espera e a duração do tratamento desempenham papéis críticos nos desfechos clínicos. Aplicações incluem prever o tempo de espera para transplantes renais, identificar fatores que influenciam a probabilidade de um evento, e avaliar a eficácia de intervenções terapêuticas (Wolfe et al., 1999; Schaubel et al., 2006).

Um desafio típico em análises de sobrevivência é a presença de dados censurados, que ocorrem quando o evento de interesse não foi observado até o final do estudo ou quando há perda de acompanhamento. Esses dados são tratados adequadamente pelos métodos de sobrevivência, garantindo estimativas precisas e imparciais (Rizopoulos, 2011).

2.4.2. Modelo de Riscos Proporcionais de Cox

O modelo de riscos proporcionais de Cox é uma técnica semi-paramétrica amplamente utilizada para modelar o impacto de variáveis preditoras no tempo até a ocorrência de eventos. Desenvolvido por David R. Cox em 1972, o modelo é baseado na premissa de que a taxa de risco $(\lambda(\, t \mid x\,))$ pode ser decomposta em duas partes:

2.4.2.1. Uma função de risco basal $(\lambda_0(t))$ que não depende das covariáveis.

2.4.2.2. Um termo exponencial, que incorpora as covariáveis $x = (x1, x2, \dots, xp)$.

A formulação matemática do modelo é:

$$\lambda(\, t \mid x\,) = \lambda_0(t) \exp(\beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_P x_P)$$

Onde:

- $\lambda(t \mid x)$: Taxa de risco condicionada às covariáveis.
- $\lambda_0(t)$: Taxa de risco basal.
- $\beta 1, \beta 2, \ldots, \beta p$: Coeficientes das covariáveis.

Estimativa de Parâmetros: Verossimilhança Parcial O modelo de Cox utiliza a verossimilhança parcial para estimar os coeficientes βj :

$$L(\beta) = \prod_{i=1}^{n} \frac{\exp(\beta x_i)}{\sum_{j \in R(t_i)} \exp(\beta x_j)}$$

Onde:

- $R(t_i)$: Conjunto de risco no tempo t_i.

Essa abordagem permite que a função de risco basal $\lambda 0(t)$ não seja especificada, simplificando a estimação dos parâmetros.

Premissa de Proporcionalidade O modelo assume que a razão entre os riscos de dois indivíduos é constante ao longo do tempo:

$$\frac{\lambda(t \mid x_1)}{\lambda(t \mid x_2)} = \exp[\beta(x_1 - x_2)]$$

Essa premissa pode ser testada por métodos gráficos, como os resíduos de Schoenfeld.

2.4.3. Regressão de Cox: Fundamentos e Uso na Previsão do Tempo de Espera

A regressão de Cox, ou modelo de riscos proporcionais, é uma ferramenta semi-paramétrica amplamente utilizada para avaliar o impacto de múltiplas covariáveis no tempo até a ocorrência de um evento. Diferentemente de modelos paramétricos, a regressão de Cox não exige uma suposição prévia sobre a distribuição dos tempos de espera, o que a torna flexível e amplamente aplicável em estudos clínicos (Rizopoulos, 2011).

No contexto dos transplantes, a regressão de Cox é usada para identificar fatores que influenciam a probabilidade de transplante ou óbito em

lista de espera. Por exemplo, Wolfe et al. (1999) demonstraram que variáveis como idade avançada e alta sensibilização imunológica estão associadas a maiores tempos de espera. Essa abordagem também permite calcular hazard ratios, que quantificam o risco relativo de um paciente ser removido da lista por diferentes motivos.

2.4.4. Limitações da Análise Tradicional de Sobrevivência

Embora amplamente utilizada, a análise tradicional de sobrevivência apresenta algumas limitações importantes, como:

- **Assunção de proporcionalidade dos riscos:** O modelo de Cox assume que as razões de risco entre grupos permanecem constantes ao longo do tempo, o que pode não refletir a realidade clínica (Schaubel et al., 2006).
- **Incapacidade de lidar com interações complexas:** Métodos clássicos têm dificuldade em capturar interações não lineares entre variáveis, especialmente em cenários complexos como os transplantes.
- **Dependência de dados completos:** A análise tradicional pode ser limitada pela presença de censura nos dados, o que requer métodos adicionais para evitar viés (Kyritsis & Deriaz, 2019).

Essas limitações reforçam a necessidade de explorar abordagens mais avançadas, como aprendizado de máquina, para melhorar a precisão e a aplicabilidade dos modelos preditivos (Ishwaran et al., 2008).

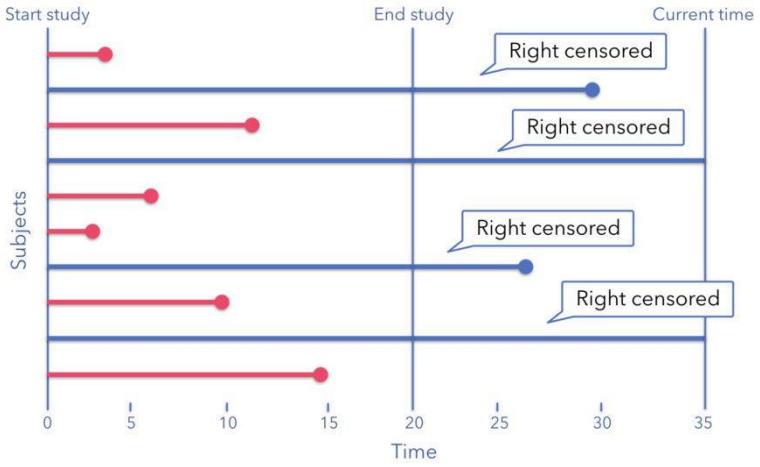
2.4.5. Censura dos Dados em Análise de Sobrevivência

A censura ocorre quando informações completas sobre o tempo de ocorrência de um evento não estão disponíveis. Em estudos clínicos, isso pode acontecer por diversas razões, como pacientes que continuam vivos ou permanecem em lista de espera até o final do período de observação. A censura dos dados é um aspecto crucial da análise de sobrevivência, pois ignorá-la pode levar a estimativas enviesadas e inadequadas (Rizopoulos, 2011).

Tipos de Censura:

Censura à Direita (Right Censoring): O tipo mais comum, ocorre quando o evento de interesse não foi observado até o final do estudo ou até o último acompanhamento do paciente. Por exemplo, se um paciente permanece em lista de espera até o final do estudo, o tempo de espera real é maior do que o registrado.

Figura 01 - *Censura à Direita (Right Censoring)*

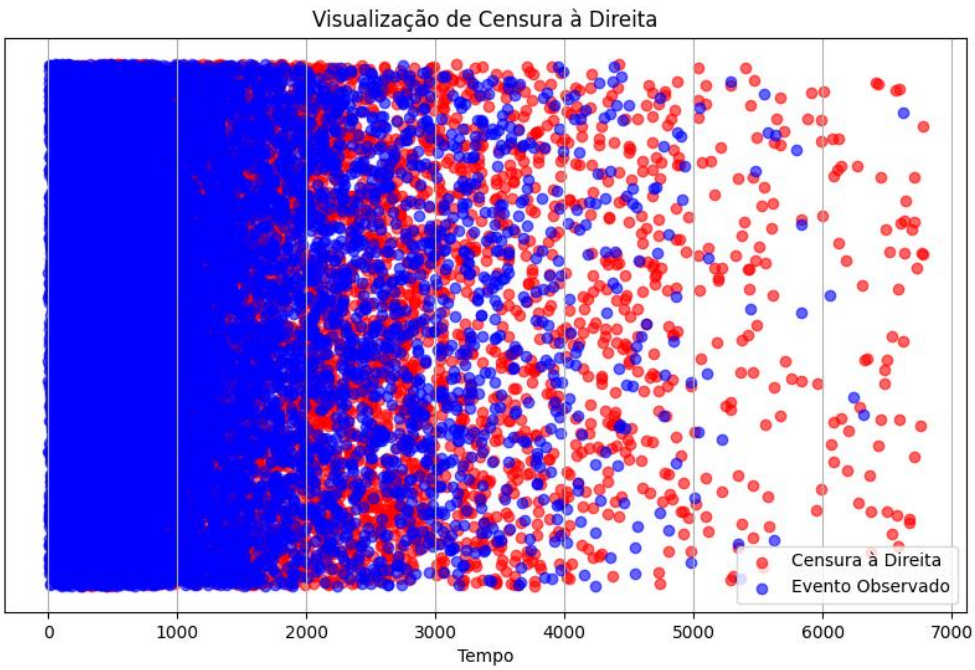


Fonte: Representação visual da censura na análise de sobrevivência (JIGSO, 2025).

O gráfico apresenta a censura à direita nos dados de sobrevivência, destacando dois tipos de observações:

- **Eventos Observados (em azul):** Representam os casos em que o evento de interesse (ex: transplante ou óbito) ocorreu durante o tempo de acompanhamento.
- **Censura à Direita (em vermelho):** Representam os casos em que o evento não foi observado até o final do período de acompanhamento. Isso ocorre porque o paciente ainda estava na lista de espera ou foi perdido no acompanhamento.

Figura 02 - *Censura à Direita Database (Right Censoring)*



Fonte: Autoria própria

- O **eixo x** representa o tempo em unidades do estudo (ex: dias ou meses).
- A **distribuição vertical** dos pontos (aleatória) foi incluída para evitar sobreposição e facilitar a visualização.
- Os **pontos vermelhos** (censurados) são mais comuns nos tempos mais longos, indicando que muitos pacientes permaneceram na lista de espera até o fim do estudo.
- Os **pontos azuis** (eventos observados) estão mais concentrados no início e no meio do período, refletindo os momentos em que os eventos ocorreram.
- A predominância de **censura à direita** em tempos mais longos sugere que muitos pacientes não tiveram o evento registrado antes do término do estudo.
- Essa característica é comum em estudos de sobrevivência, nos quais os eventos podem não ocorrer durante o acompanhamento.

- **Censura à Esquerda (Left Censoring):** Ocorre quando o evento já aconteceu antes do início do acompanhamento, mas não se

sabe exatamente quando. Por exemplo, se um paciente já iniciou a diálise antes de ser incluído no estudo.

- **Censura por Intervalo (Interval Censoring):** O evento é conhecido por ter ocorrido entre duas observações, mas o momento exato dentro desse intervalo é desconhecido.

Para lidar com censura à direita, o estimador Kaplan-Meier é amplamente utilizado:

$$S(t) = \prod_{ti \leq t} \left(1 - \frac{d_i}{n_i}\right)$$

Onde:

- d_i: Número de eventos no tempo t_i.
- n_i: Número de indivíduos em risco no tempo t_i.

O Kaplan-Meier fornece uma curva não paramétrica da função de sobrevivência, útil para comparar grupos ou verificar a validade de modelos ajustados (Gerds & Schumacher, 2006).

A abordagem adequada para lidar com dados censurados envolve métodos como o estimador de Kaplan-Meier, que calcula probabilidades de sobrevivência considerando eventos censurados, e o modelo de Cox, que incorpora censura à direita em suas análises. Modelos mais avançados, como os baseados em aprendizado de máquina, também têm demonstrado eficácia no manejo de censura e na predição de eventos em contextos complexos (Kyritsis & Deriaz, 2019; Silva et al., 2021).

Se a análise não tratasse adequadamente os dados censurados, poderíamos obter estimativas tendenciosas do tempo de sobrevivência. Por exemplo, simplesmente ignorar os indivíduos que não tiveram o evento até o final do estudo implicaria em subestimar o tempo médio de sobrevivência, já que todos os "sobreviventes" até aquele momento

estariam excluídos ou considerados como se tivessem o evento naquele instante, quando isso não é correto.

2.4.6. Os métodos Convencionais de Análise de Sobrevivência:

Sobre análise de espera em transplantes utilizando métodos clássicos ver Wolfe, R. A., et al. (1999) e Schaubel, D. E., et al. (2006).

- **Estimador Kaplan-Meier:** Uma técnica não paramétrica para estimar a função de sobrevivência e obter a probabilidade de “sobrevivência” na lista (ou seja, continuar aguardando o transplante) em função do tempo.
- **Modelo de Riscos Proporcionais de Cox:** Uma abordagem semi-paramétrica que permite avaliar o efeito de múltiplas covariáveis (por exemplo, estado de saúde, tipo de órgão, características sociodemográficas) sobre a taxa de saída da lista (seja por transplante ou óbito) ao longo do tempo.
- **Modelos Paramétricos (Exponencial, Weibull, Gompertz):** Alternativas que assumem distribuições específicas para o tempo de espera, fornecendo flexibilidade e, em alguns casos, maior interpretabilidade.

2.5. Aprendizado de Máquina (Machine Learning)

Machine Learning (ML) é um campo da inteligência artificial que utiliza algoritmos para permitir que os sistemas computacionais aprendam e realizem predições ou decisões com base em dados. No contexto de transplantes renais, o ML se mostra promissor para lidar com a complexidade dos dados clínicos, como múltiplos fatores demográficos e imunológicos, permitindo previsões mais precisas sobre o tempo de espera e a alocação de órgãos (Kyritsis & Deriaz, 2019; Silva et al., 2021).

Nos últimos anos, métodos de machine learning (ML) e deep learning vêm sendo integrados à análise de sobrevivência. Revisões e estado da arte em métodos avançados de análise de sobrevivência com ML podem ser encontrados em Wang, P., Li, Y., & Reddy, C. K. (2019) e Bellot, A., van der Schaar, M. (2019). Sobre métodos de machine learning em sobrevivências pesquisas reportadas em Ishwaran, H., Kogalur, U. B., Blackstone, E. H., & Lauer, M. S. (2008). Katzman, J. L., et al. (2018). Esses métodos procuram capturar relações mais complexas entre covariáveis e o tempo de espera, muitas vezes sem precisar assumir formas funcionais rígidas.

2.5.1. Introdução e Definição

O ML envolve um conjunto de técnicas e algoritmos que permitem a análise de grandes volumes de dados complexos e a criação de modelos preditivos. Ele pode ser dividido em três categorias principais:

- **Aprendizado Supervisionado:** Baseia-se em dados rotulados para treinar o modelo.
- **Aprendizado Não Supervisionado:** Identifica padrões em dados não rotulados.
- **Aprendizado por Reforço:** Aprende com interações e recompensas.

No contexto de sobrevivência, algoritmos de ML são frequentemente usados para prever tempos de espera em listas de transplantes e identificar os fatores que influenciam os desfechos clínicos (Wang et al., 2019).

2.5.2. Diferenças entre Métodos Tradicionais e Aprendizado de Máquina

Enquanto os métodos tradicionais, como regressão de Cox, dependem de suposições rígidas sobre a proporcionalidade dos riscos e a linearidade das variáveis, o ML supera essas limitações ao:

- **Modelar relações não lineares:** Algoritmos como Random Survival Forest (RSF) capturam interações complexas entre variáveis (Ishwaran et al., 2008).

- **Lidar com grandes volumes de dados:** Métodos como XGBoost permitem integrar centenas de variáveis preditoras de forma eficiente (Kyritsis & Deriaz, 2019).
- **Incorporar censura de dados:** Algoritmos como DeepSurv utilizam dados censurados para treinar modelos preditivos robustos (Katzman et al., 2018).

2.5.3. Benefícios do Aprendizado de Máquina em Transplantes Renais

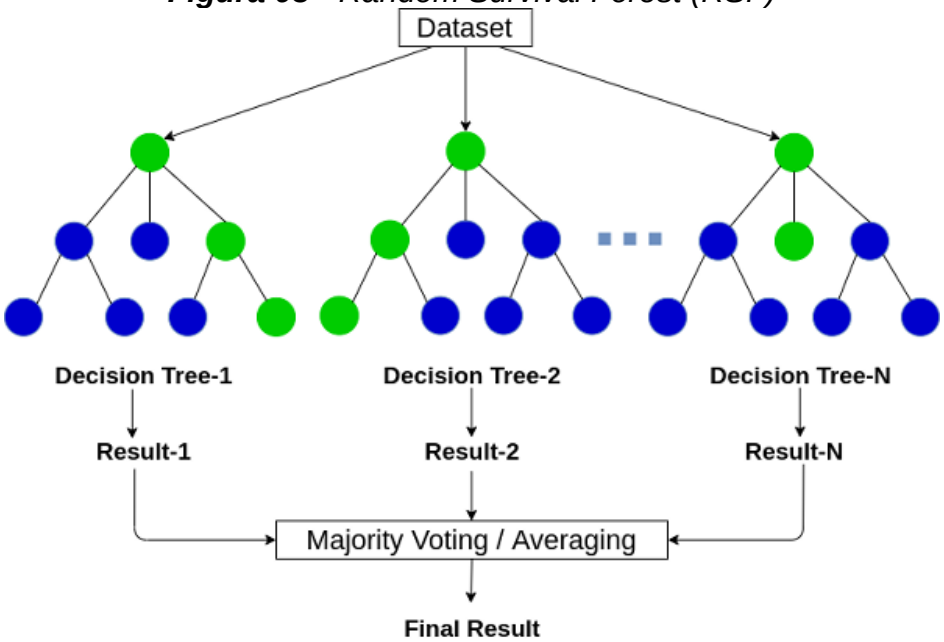
O uso de ML em transplantes renais traz vários benefícios:

- **Previsões mais precisas do tempo de espera:** Modelos como RSF e XGBoost apresentam alta acurácia ao integrar variáveis complexas.
- **Personalização no tratamento:** Ferramentas como DeepSurv ajudam a prever desfechos individuais, permitindo melhor planejamento clínico (Katzman et al., 2018).
- **Redução de desigualdades:** ML pode priorizar pacientes de forma mais justa, promovendo equidade na distribuição de órgãos (Coorey et al., 2020).
- **Integração de censura nos dados:** Dados censurados, como pacientes que permanecem em lista de espera até o final do estudo, são tratados adequadamente em algoritmos avançados, melhorando a confiabilidade das predições (Kyritsis & Deriaz, 2019).

2.6. MODELOS AVANÇADOS DE PREDIÇÃO

2.6.1. Random Survival Forest (RSF): Características e Aplicações

Figura 03 - *Random Survival Forest (RSF)*



Fonte: Comparação entre os algoritmos de Decision Tree e Random Forest (ANALYTICS VIDHYA, 2020).

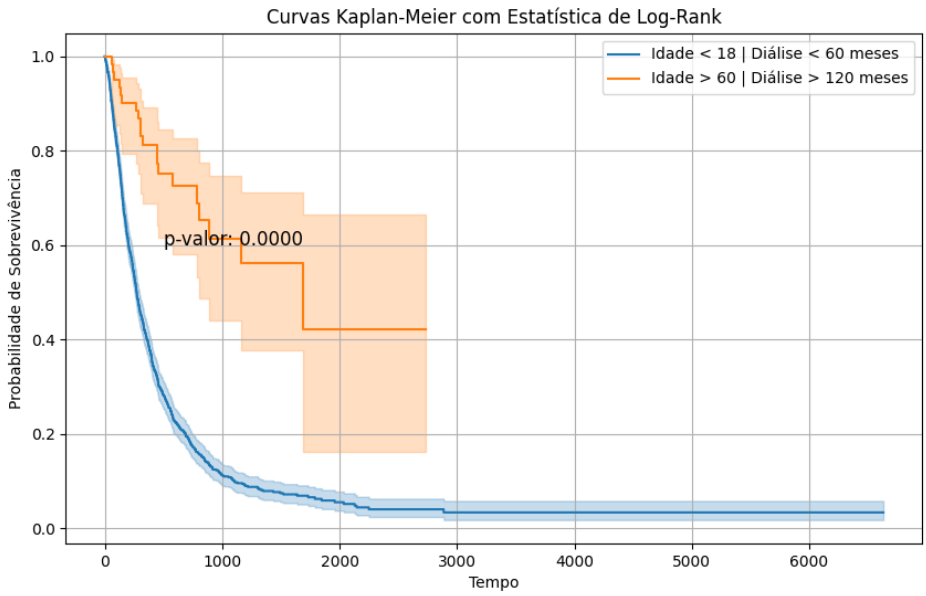
O Random Survival Forest (RSF) é uma extensão dos modelos de florestas aleatórias para análise de sobrevivência, projetado para lidar com dados censurados. Desenvolvido por Ishwaran et al. (2008), o RSF cria múltiplas árvores de decisão baseadas em amostras bootstrap dos dados e estima funções de sobrevivência agregadas por meio de um conjunto (ensemble).

A RSF é baseada em árvores de decisão não paramétricas, cada uma construída a partir de amostras bootstrap dos dados originais. Os principais aspectos do RSF incluem:

Características Principais:

- **Divisão por Estatística de Log-Rank:** A divisão dos nós é feita com base no teste log-rank, que maximiza a separação entre grupos de sobrevivência.

Figura 04 – *Curva Kaplan-Meier para grupo de idade e diálise*



Fonte: Autoria própria

O gráfico apresenta as **curvas de sobrevivência Kaplan-Meier** para dois grupos de pacientes na lista de espera para transplante renal, baseando-se em variáveis importantes: **faixa etária** e **tempo em diálise**. A análise inclui um **teste de Log-Rank**, indicado pelo p-valor no gráfico.

- **Grupo 1 (Linha Azul):** Pacientes com **idade inferior a 18 anos** e **tempo em diálise inferior a 60 meses**.
- **Grupo 2 (Linha Laranja):** Pacientes com **idade superior a 60 anos** e **tempo em diálise superior a 120 meses**.

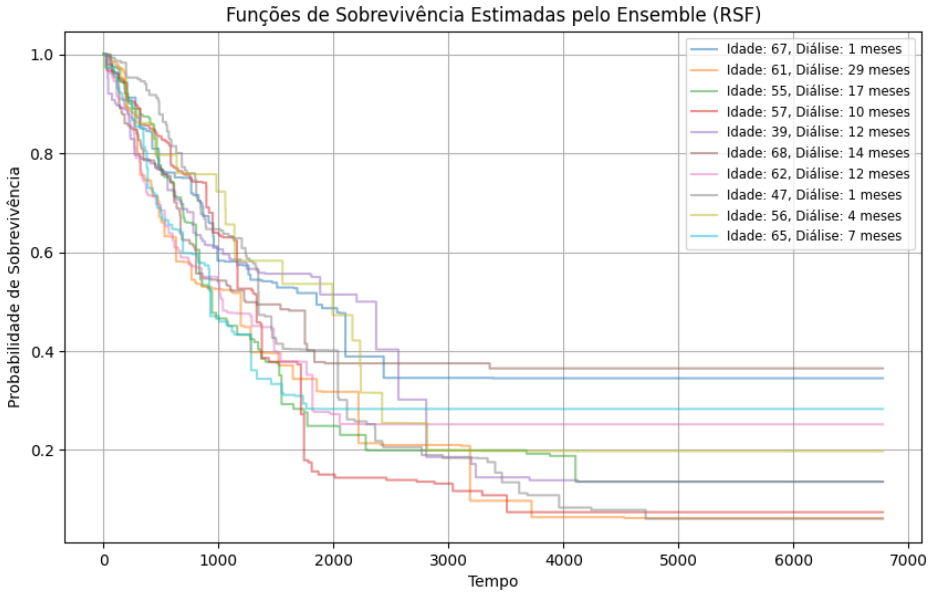
1. **Interpretação dos Resultados**

- **Curvas Kaplan-Meier:** A curva azul, representando o **Grupo 1**, demonstra uma queda acentuada inicial, mas se estabiliza rapidamente, indicando melhor probabilidade de sobrevivência ao longo do tempo.
- A curva laranja, referente ao **Grupo 2**, apresenta menor probabilidade de sobrevivência ao longo do tempo, com uma queda menos acentuada inicialmente, mas que se estabiliza em um nível muito mais baixo.

2. **P-valor (Log-Rank Test):**
   O valor p = 0.0000 indica uma diferença estatisticamente significativa entre as curvas de sobrevivência dos dois grupos. Isso sugere que a idade avançada e o tempo prolongado em diálise têm um impacto significativo na probabilidade de sobrevivência.

- **Funções de sobrevivência estimadas:** Cada árvore gera estimativas individuais, que são combinadas em um conjunto (ensemble) para melhorar a precisão.

Figura 05 – *Função de sobrevivência para o grupo de idade e diálise*



Fonte: Autoria própria

O gráfico apresenta as funções de sobrevivência estimadas geradas pelo modelo Random Survival Forest (RSF) para 10 amostras de pacientes da lista de espera para transplante renal. Cada curva representa uma combinação individual de idade e tempo em diálise, conforme destacado na legenda. As curvas descrevem a probabilidade de sobrevivência ao longo do tempo.

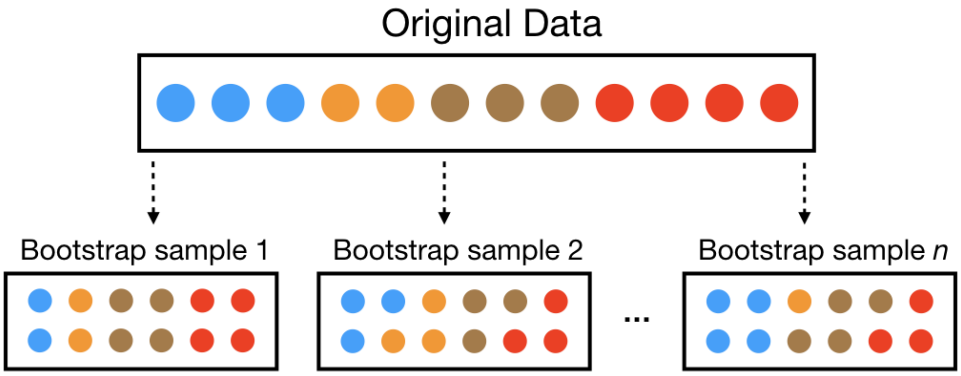
- **Tratamento de censura:** Dados censurados são incorporados naturalmente nas análises, permitindo estimativas confiáveis mesmo com informações incompletas.

Etapas de Construção do RSF:

- **Amostragem Bootstrap:**

  - Em cada iteração, são criadas amostras bootstrap dos dados originais para construir árvores.

Figura 06 – *Amostragem Bootstrap*



Fonte: Representação visual do esquema de amostragem bootstrap (DATA SCIENCE DOJO, 2025).

  - Aproximadamente 1/3 dos dados ficam fora da amostra (Out-of-Bag, OOB) e são usados para validação.

- **Critério de Divisão:**

  - **Divisão por Estatística de Log-Rank:** A métrica principal para divisão dos nós é baseada na estatística de log-rank:

$$Log\ Rank = \frac{(O_2 - E_2)^2}{Var(O_2 - E_2)}$$

Onde:

- O_i: Número observado de eventos,
- E_i: Número esperado de eventos.

- **Construção das Árvores**

  - Cada árvore é construída até alcançar um tamanho mínimo de nó ou até que não haja mais eventos para dividir.

- **Funções de Sobrevivência:**

  - Cada nó terminal armazena uma estimativa da função de sobrevivência para os casos que chegam a ele.
  - A função de sobrevivência do conjunto é dada pela média das estimativas das árvores:

$$\hat{S}(t|X) = \frac{1}{B} \sum_{b=1}^{B} S_b(t|X)$$

Onde:

- $Sb(t \mid X)$: Função de sobrevivência estimada pela $bb-$ $ésima$ árvore,
- B : Número total de árvores no conjunto.

- **Métricas de Avaliação**

  - **Erro Fora da Bolsa (OOB):** Mede o desempenho preditivo em dados não utilizados para construir as árvores.
  - **Índice de Concordância (C):** Mede a capacidade do modelo

de discriminar entre tempos de sobrevivência.

$$C = \frac{Numero\ de\ pares\ concordantes}{Numero\ total\ de\ parres}$$

- **Pontuação de Brier:** Mede a diferença quadrática entre as estimativas de sobrevivência e os eventos observados:

$$BS = \frac{1}{N}\sum_{t=1}^{N}(f_t - o_t)^2$$

Onde:

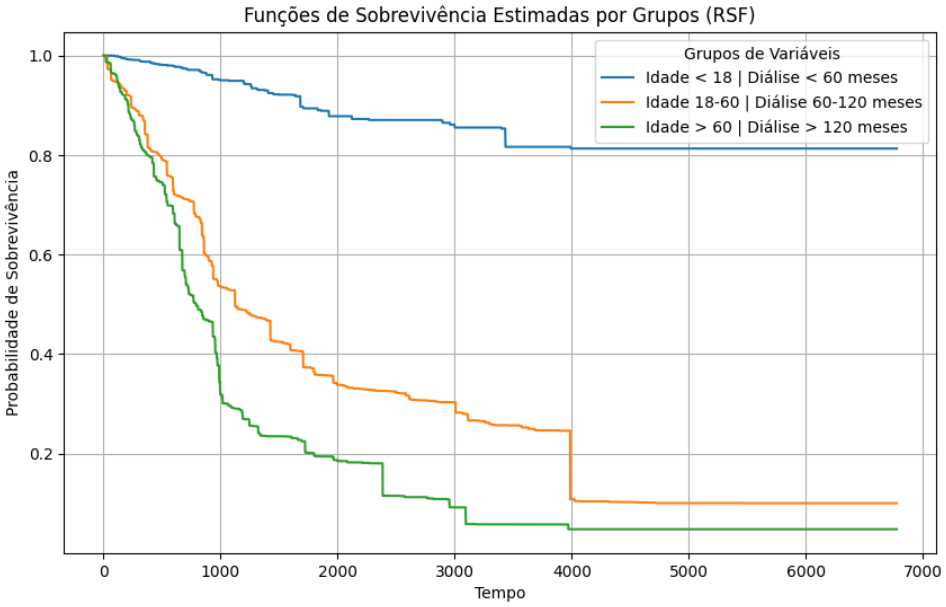
- $N =$ o número de itens para os quais você está calculando uma pontuação de Brier.
- f_t é a probabilidade prevista,
- o t é o resultado (1 se aconteceu, 0 se não aconteceu).
- Σ é o símbolo de soma. Significa apenas “somar” todos os valores.

- **Importância das Variáveis:**

  - A importância de uma variável (x) é medida pela diferença no erro OOB quando os valores da variável são permutados aleatoriamente:

$$VI(X_j) = \frac{1}{B}\sum_{b=1}^{B}(Erro\ OOB\ permutado\ -\ Erro\ OOB\ original)$$

Figura 07 – Função de sobrevivência para o grupo de idade, diálise e meses



Fonte: Autoria própria

- **Gráfico de Funções de Sobrevivência Estimadas (RSF)**

O gráfico apresenta as funções de sobrevivência estimadas utilizando um modelo de Random Survival Forest (RSF) para três grupos distintos de pacientes na lista de espera para transplante renal. Os grupos foram definidos com base em variáveis importantes: faixa etária e tempo em diálise. As curvas representam a probabilidade de sobrevivência ao longo do tempo para cada grupo, conforme descrito na legenda:

- **Idade < 18 | Diálise < 60 meses (Linha azul):** Este grupo apresenta a maior probabilidade de sobrevivência, com a curva decaindo de forma mais lenta ao longo do tempo. Mesmo após longos períodos, a probabilidade de sobrevivência permanece acima de 80%, indicando melhores perspectivas de sobrevivência.
- **Idade 18-60 | Diálise 60-120 meses (Linha laranja):**

A função de sobrevivência deste grupo decai de maneira mais acelerada em comparação ao grupo anterior. A probabilidade de sobrevivência cai para cerca de 40% em torno de 4000 unidades de tempo, evidenciando um impacto mais severo da idade intermediária e do tempo prolongado em diálise.

- **Idade > 60 | Diálise > 120 meses (Linha verde):** Este grupo apresenta a menor probabilidade de sobrevivência, com um declínio acentuado na curva. A probabilidade de sobrevivência atinge níveis abaixo de 20% após 1000 unidades de tempo, refletindo o impacto negativo da idade avançada e do tempo prolongado em diálise.

Análise do Gráfico:

- A idade avançada e o tempo prolongado em diálise são fatores determinantes para a redução da probabilidade de sobrevivência.

- Pacientes mais jovens (< 18 anos) e com menor tempo em diálise (< 60 meses) têm melhores perspectivas de sobrevivência, sugerindo maior capacidade de recuperação e resposta positiva ao tratamento.

- Grupos com idade intermediária e avançada ou tempo de diálise prolongado apresentam declínios significativos na sobrevivência, destacando a necessidade de priorizar no sistema de alocação de órgãos.

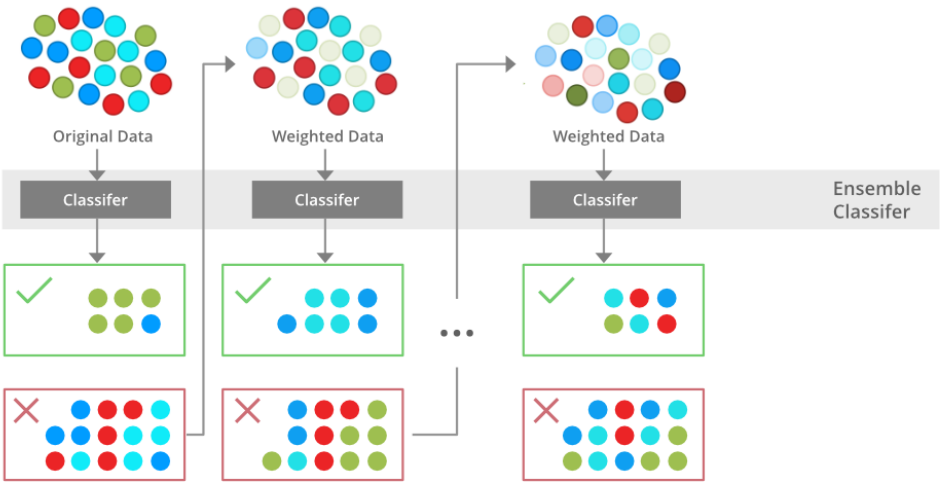
- **Relevância para o Trabalho:** Este gráfico ilustra de forma clara o impacto de variáveis demográficas e clínicas na probabilidade de sobrevivência dos pacientes. A análise reforça a importância de utilizar modelos preditivos, como o RSF, para identificar padrões de risco e otimizar políticas de transplante renal.

2.6.2. XGBoost: Fundamentos e Desempenho em

Predições Clínicas

O **XGBoost (Extreme Gradient Boosting)** é uma técnica de aprendizado de máquina baseada em árvores de decisão, amplamente reconhecida por sua eficiência e desempenho superior. Na análise de sobrevivência, ele pode ser configurado para ajustar modelos de regressão de Cox ou de AFT (Accelerated Failure Time), permitindo lidar com dados censurados (XGBoosting, 2023).

Figura 08 - *XGBoost*



Fonte: Representação visual do algoritmo XGBoost e do conceito de Boosting (GEEKSFORGEEKS, 2020).

Configuração do XGBoost para Sobrevivência:

- **Modelo Cox:** O XGBoost ajusta um modelo de risco proporcional, maximizando a função de verossimilhança parcial:

$$L(\beta) = \sum_i \left[X_i\beta - \log \sum_{j \in R_i} \exp(X_j\beta) \right]$$

Onde:

- X_I: Vetor de covariáveis do indivíduo i ,

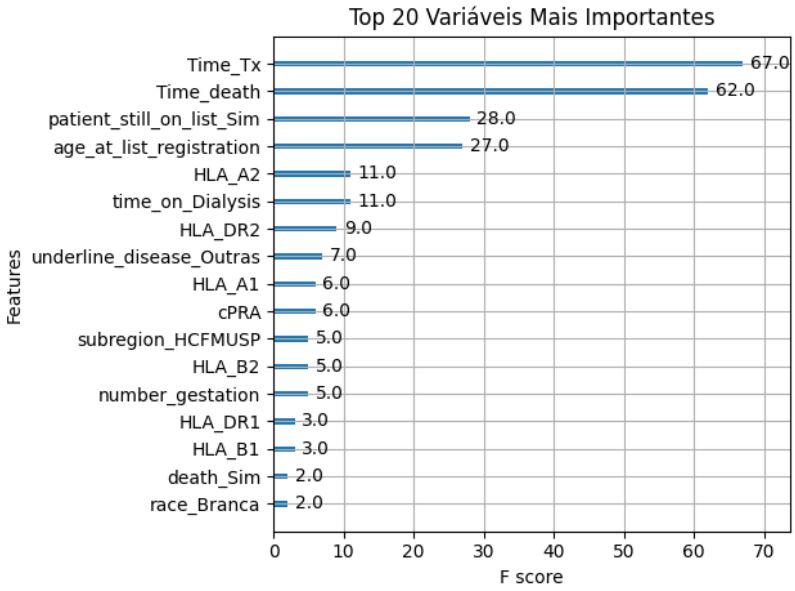
- R_i: Conjunto de risco no tempo t_i.

- **Modelo AFT:** Minimiza a função de perda baseada no log do tempo:

$$\mathcal{L}_{AFT} = \sum_i [\log(y_i) - (\beta X_i)]^2$$

- **Vantagens do XGBoost:**
  - Lida eficientemente com grandes bases de dados.
  - Permite ajustar hiperparâmetros para otimização.
  - Integra pesos para tratar dados censurados.

Figura 09 – Variáveis importantes



Fonte: Autoria própria

As 20 variáveis mais importantes para o modelo XGBoost com base no peso ("F score").

- *Time_Tx* e *Time_death* são as mais significativas, indicando grande relevância no modelo.
- Outras variáveis como *patient_still_on_list_Sim* e

age_at_list_registration também se destacam.

Tabela 01 – Característica dos modelos

Modelo	Vantagens	Desvantagens
Regressão de Cox	Simplicidade; amplamente utilizado.	Supõe proporcionalidade de riscos; limitado para dados complexos.
RSF	Captura relações não lineares; lida bem com censura.	Exige maior capacidade computacional.
XGBoost	Alto desempenho; flexibilidade na modelagem.	Pode ser mais complexo de interpretar.

Fonte: Autoria própria

2.7. Ferramentas e Métricas de Avaliação de Modelos

A avaliação de modelos preditivos é essencial para medir a qualidade e a utilidade das predições em contextos clínicos. No caso de análises de sobrevivência e aprendizado de máquina aplicados ao tempo de espera para transplantes, métricas como o índice de concordância (C-index), a Pontuação de Brier, e as curvas ROC para modelos baseados em sobrevivência são amplamente utilizadas.

2.7.1. Ferramentas e Métricas de Avaliação de Modelos

Índice de Concordância (C-index): O C-index como mencionado anteriormente, é usado para medir a capacidade de um modelo de classificar corretamente pares de indivíduos em relação ao tempo até o evento. Ele avalia se as predições de risco são consistentes com os tempos de sobrevivência observados.

Fórmula do C-index:

$$c = \frac{Numero\ de\ pares\ concordantes}{Numero\ total\ de\ pares}$$

- Um par é "concordante" se o indivíduo com maior risco predito tiver o evento antes.
- Valores próximos de 1 indicam forte discriminação, enquanto 0.5 indica predições aleatórias.

Aplicação Clínica: Em transplantes, um C-index alto significa que o modelo é eficaz em identificar pacientes com maior prioridade ou risco.

2.7.2. Pontuação de Brier:

Como mencionado anteriormente, é usada para medir a acurácia de modelos ao comparar as probabilidades preditas com os eventos observados.

$$B(t) = \frac{N}{1}\sum_{i=1}^{N}[S(\,t \mid X_i\,) - \delta i]^2$$

Onde:

- $S(\,t \mid X_i\,)$: Probabilidade de sobrevivência predita no tempo t para o indivíduo i .
- δi : Indicador binário do evento (1 se o evento ocorreu, 0 se censurado).
- N : Número total de indivíduos.

Interpretação: Valores baixos de $B(t)$indicam melhor performance.

2.7.3. Curvas ROC e AUC:

Embora menos comuns em análise de sobrevivência, as curvas ROC podem ser adaptadas para avaliar modelos de risco ao longo de um intervalo de tempo específico.

Área sob a Curva (AUC):

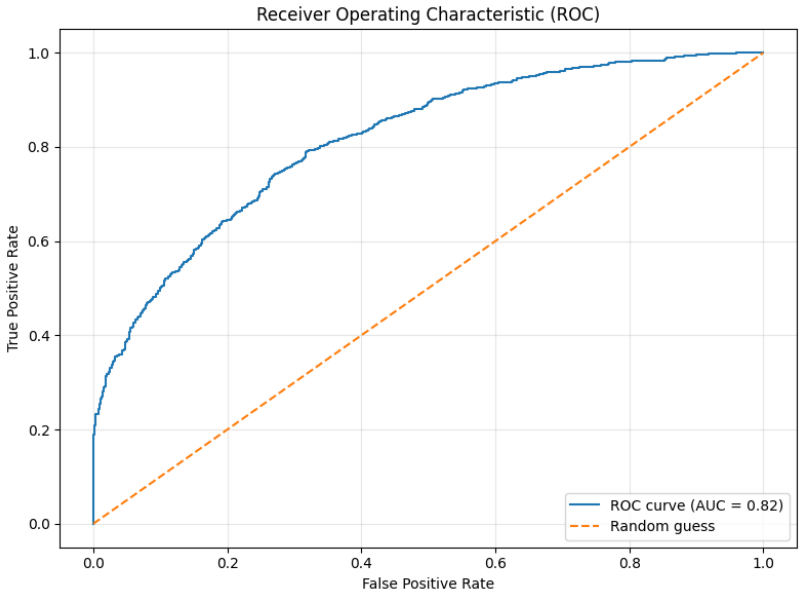
$$AUC = \int_0^1 TPR(f)\ d\big(FPR(f)\big)$$

Onde:

- TPR: Taxa de verdadeiros positivos,
- FPR: Taxa de falsos positivos.

Aplicação: Em transplantes, curvas ROC podem ser usadas para avaliar a capacidade de um modelo em priorizar corretamente pacientes de alto risco.

Figura 10 – Gráfico de curvas ROC



Fonte: Autoria própria

2.8. Validação Cruzada e Otimização de Hiperparâmetros

A validação cruzada é uma técnica fundamental para avaliar a generalização de modelos preditivos e evitar overfitting.

2.8.1. Validação Cruzada k-Fold:

- Divide os dados em *kk*k subconjuntos (folds),
- Usa *k−1* folds para treinar o modelo e 1 fold para teste, repetindo o

processo *k* vezes.

- A média dos erros de teste fornece uma estimativa robusta do desempenho.

Erro de Previsão na Validação Cruzada:

$$Erro = \frac{1}{K}\sum_{I=1}^{K} Erro_I$$

2.8.2. Otimização de Hiperparâmetros:

- Técnicas como grid search e random search são usadas para ajustar parâmetros do modelo, como número de árvores no RSF ou taxa de aprendizado no XGBoost.
- A combinação de validação cruzada com otimização garante um modelo bem ajustado.

2.8.3. Considerações sobre Interpretabilidade e Aplicabilidade Prática

- **Importância da Interpretabilidade:**
  - Modelos preditivos, especialmente em saúde, devem ser interpretáveis para que os clínicos confiem nas decisões baseadas neles.
  - Ferramentas como gráficos de importância de variáveis ajudam a identificar os fatores mais relevantes, como idade ou tempo em diálise.

- **Aplicabilidade Prática:**
  - Modelos devem ser testados em dados reais para garantir sua relevância.
  - Por exemplo, ao prever tempos de espera para transplantes, é

essencial validar o modelo em diferentes populações e cenários clínicos.

- **Equidade:**

Avaliar se o modelo prioriza adequadamente pacientes de grupos demográficos diversos, evitando vieses.

2.9. Estudos Correlatos

2.9.1. Aplicações de Análise de Sobrevivência e Aprendizado de Máquina em Transplantes

A análise de sobrevivência e o aprendizado de máquina têm desempenhado papéis significativos na previsão de desfechos relacionados a transplantes. Essas técnicas ajudam a estimar o tempo de espera, prever a compatibilidade entre doadores e receptores e identificar fatores de risco que afetam a sobrevida do enxerto ou do paciente.

1. **Análise de Sobrevivência:**
   - Modelos como Kaplan-Meier e Cox têm sido amplamente utilizados para estimar a probabilidade de sobrevivência em listas de espera para transplantes renais (Wolfe et al., 1999; Schaubel et al., 2006).
   - Por exemplo, Schaubel et al. (2006) demonstraram que o tempo em diálise afeta significativamente a mortalidade em pacientes à espera de transplantes.

2. **Aprendizado de Máquina:**
   - Métodos baseados em árvores, como Random Survival Forest (RSF), têm sido usados para identificar fatores que influenciam o tempo de espera (Ishwaran et al., 2008).
   - Modelos como XGBoost e DeepSurv permitem lidar com interações não lineares e complexas entre covariáveis,

melhorando a acurácia das predições (Katzman et al., 2018; Silva et al., 2021).

2.9.2. Experiências Internacionais e Estudos Nacionais Relevantes

1. **Estudos Internacionais:**
   - **Estados Unidos:** Wolfe et al. (1999) analisaram a mortalidade em listas de espera para transplantes renais, destacando a importância do sistema de alocação para otimizar a sobrevivência.
   - **Europa:** Katzman et al. (2018) introduziram o DeepSurv, um modelo baseado em redes neurais que superou a regressão de Cox em predições de sobrevivência em câncer.
   - **Israel:** Elalouf et al. (2018) propuseram um modelo de fila para otimizar a alocação dinâmica de órgãos, abordando cenários com doadores vivos e falecidos.

2. **Estudos Nacionais:**
   - Silva et al. (2021) aplicaram aprendizado de máquina para prever o tempo de espera em listas de transplantes renais no Brasil. Eles demonstraram que variáveis como tempo em diálise e sensibilização imunológica (cPRA) influenciam significativamente o desfecho.

2.10. Limitações e Lacunas na Literatura Atual

Apesar do avanço dessas técnicas, algumas limitações permanecem:

1. **Dados Heterogêneos:**
   - Muitos estudos utilizam bases de dados locais ou institucionais, limitando a generalização dos resultados (Silva et al., 2021).
2. **Censura e Dados Incompletos:**

- Modelos tradicionais nem sempre lidam bem com censura por intervalo ou informativa, o que pode introduzir viés (Rizopoulos, 2011).

3. **Falta de Padronização:**
   - A ausência de protocolos unificados para a validação de modelos preditivos dificulta a comparação entre estudos (Ishwaran et al., 2008).
4. **Equidade e Viés:**
   - Poucos estudos avaliam o impacto de fatores demográficos e socioeconômicos na previsão de desfechos, o que é crucial para promover a equidade (Wolfe et al., 1999).

3. PROCEDIMENTOS METODOLÓGICOS

Esta pesquisa, segundo sua natureza é um resumo de assunto, buscando explicar a área do conhecimento do projeto, indicando sua evolução histórica, como resultado da investigação das informações obtidas, levando ao entendimento de suas causas e explicações (WAZLAWICK, 2014).

Segundo os objetivos é uma pesquisa exploratória e descritiva. A pesquisa descritiva busca dados mais consistentes sobre determinado assunto, porém, não ocorre a interferência do pesquisador, apenas expõe os fatos como realmente são (WAZLAWICK, 2014). As pesquisas descritivas descrevem as características de certo fenômeno ou população. Também pode ser elaborada com o intuito de identificar as relações entre as variáveis (GIL, 2017).

A pesquisa exploratória muitas vezes é considerada como a primeira parte do processo de pesquisa, porque não necessariamente o autor tem um objetivo ou hipótese definida (WAZLAWICK, 2014). Essa pesquisa tem como objetivo a maior familiaridade do autor com o problema, tornando mais explícito ou facilitar a construção de hipóteses. Geralmente é uma pesquisa flexível, porque considera os variados aspectos referentes aos fatos ou fenômeno estudado (GIL, 2017).

Quanto aos procedimentos técnicos, será uma pesquisa bibliográfica e experimental. A pesquisa bibliográfica requer o estudo de teses, artigos, entre outros. A pesquisa experimental é caracterizada por ter uma ou mais variáveis experimentais que podem ser coordenadas pelo pesquisador (WAZLAWICK, 2014).

A pesquisa bibliográfica, foi elaborada a partir de materiais já publicados, podendo incluir livros, teses, materiais disponibilizados na Internet, revistas, entre outros. A principal vantagem é permitir uma sucessão de fenômenos maior do que seria capaz de pesquisar diretamente (GIL, 2017).

A pesquisa experimental consiste que o pesquisador provoque mudanças no ambiente de pesquisa, observando se as alterações realizadas são de acordo com os resultados esperados (WAZLAWICK, 2014).

A pesquisa experimental consiste em estabelecer um objeto de estudo, escolher as variáveis que a influenciam e determinar as formas de controle e observar os efeitos que a variável gera no objeto. Realiza pelo menos um dos elementos que julga ser responsável pela circunstância que está sendo pesquisado (GIL, 2017).

No presente trabalho, a pesquisa exploratória foi conduzida utilizando a base científica Web of Science. Para avaliar as direções da pesquisa, foi realizada uma busca utilizando as palavras-chave “Survival Analysis”, “Machine Learning” e “Kidney Transplantation".

Nesta pesquisa, será realizado o desenvolvimento de um modelo preditivo para estimar o tempo de espera em listas de transplantes renais. Após o treinamento e validação do modelo, os resultados serão integrados ao desenvolvimento de uma calculadora interativa. Essa ferramenta será implementada utilizando Python no back-end e tecnologias eficientes para desenvolvimento front-end. A calculadora permitirá que médicos e pacientes insiram dados relevantes, como idade, tempo em diálise e histórico clínico, para obter uma estimativa personalizada do tempo de espera para o transplante. Além disso, a ferramenta será validada com dados reais e será aprimorada com base no feedback dos usuários, garantindo maior usabilidade e precisão em suas previsões.

3.1. BASE DE DADOS

A base de dados utilizada neste trabalho foi obtida do São Paulo State Organ Allocation System (SP-OAS), uma plataforma brasileira reconhecida por registrar informações detalhadas sobre a lista de espera e a alocação de órgãos no estado de São Paulo. O sistema SP-OAS desempenha um papel crucial no gerenciamento da alocação de órgãos, seguindo critérios clínicos, imunológicos e logísticos, alinhados às políticas nacionais de transplantes. Inicialmente, diversas tentativas foram realizadas para acessar dados do governo federal por meio da Secretaria Estadual de Saúde de Goiás. No entanto, a falta de respostas positivas nos levou a recorrer a essa base alternativa, disponibilizada no Kaggle, que, embora não esteja totalmente atualizada, apresenta uma boa abrangência e volume de informações históricas, com 54.055 registros detalhados entre os anos de 2000 e 2017.

Os dados compreendem informações ricas e variadas, abrangendo variáveis demográficas, clínicas, imunológicas e de desfecho. Entre as variáveis demográficas estão idade, sexo, raça e sub-região de registro, que oferecem insights sobre as características gerais da população em lista de espera. Variáveis clínicas incluem tempo em diálise, histórico de doenças como diabetes e hipertensão arterial sistêmica (HAS), e parâmetros imunológicos, como o cPRA e os antígenos HLA, que são cruciais para a compatibilidade do transplante. Quanto aos desfechos, a base registra transplantes realizados, óbitos e remoções da lista por diversos motivos.

Para garantir a integridade e relevância da análise, foi realizado um cuidadoso processo de seleção e limpeza dos dados. Apenas registros que atendiam aos critérios temporais de 2000 a 2017 foram incluídos, assegurando um período suficientemente longo para análise de tendências e padrões. Dados incompletos ou inconsistentes em variáveis críticas, como idade, sexo ou cPRA, foram removidos, assegurando maior confiabilidade às análises subsequentes. Variáveis contínuas, como idade, foram categorizadas em faixas etárias (Menor.18, 18.a.60, Maior.60), o que facilitou a análise comparativa entre grupos.

Embora a base de dados possua limitações, como a ausência de atualizações recentes, sua abrangência e riqueza de variáveis a tornam adequada para o desenvolvimento de modelos preditivos. Essa abordagem permite capturar tendências relevantes para a previsão do tempo de espera e contribui para o aprimoramento da gestão das listas de transplante, com implicações clínicas e gerenciais significativas.

3.2. PRÉ-PROCESSAMENTO DE DADOS

O pré-processamento de dados é uma etapa indispensável neste trabalho, garantindo que a base de dados estivesse preparada para análises robustas e a aplicação de modelos preditivos. As atividades incluíram identificação e tratamento de valores ausentes, normalização, transformação e categorização de variáveis, além da criação de características derivadas. O objetivo foi maximizar a qualidade do conjunto de dados e assegurar a eficiência dos modelos desenvolvidos.

Inicialmente, foi realizada uma análise detalhada para localizar valores ausentes e padrões de inconsistências. Utilizando ferramentas de visualização, como heatmaps de dados ausentes, identificou-se que variáveis importantes, como tempo em diálise, Time_Tx e marcadores imunológicos, apresentavam lacunas significativas. Essas lacunas poderiam comprometer a integridade do conjunto de dados e, consequentemente, os resultados dos modelos, exigindo uma abordagem cuidadosa para seu tratamento.

Diferentes técnicas foram aplicadas para lidar com valores ausentes, de acordo com a natureza das variáveis. Para variáveis contínuas, como tempo em diálise, optou-se pela imputação da mediana, uma estratégia menos sensível a outliers. Já variáveis categóricas, como tipo sanguíneo e sex, tiveram os valores preenchidos pela moda, representando a categoria mais frequente. Registros com ausência de informações críticas, como idade ou cPRA, foram removidos, assegurando maior confiabilidade nos resultados das análises.

A etapa de transformação e normalização foi conduzida para garantir a consistência dos dados e facilitar sua manipulação pelos algoritmos de aprendizado de máquina. Variáveis contínuas, como tempo em diálise e idade, foram normalizadas por meio do método min-max scaling, assegurando uma escala uniforme. Além disso, variáveis contínuas foram categorizadas para simplificar as comparações e análises. A idade, por exemplo, foi dividida em faixas como Menor.18, 18.a.60 e Maior.60. Já o cPRA foi segmentado em categorias como Entre_0_50, Entre_50_80 e Maior_80. Variáveis booleanas, como diabetes e gestação, foram convertidas para valores binários (0 e 1), facilitando o processamento computacional.

Para enriquecer o conjunto de dados e capturar informações mais detalhadas, foram criadas características derivadas. Um exemplo foi o cálculo do HLA Mismatch Score, baseado na frequência dos antígenos HLA presentes em cada paciente, e a variável tempo total na lista, que combinou informações de tempo em diálise com o tempo desde o registro na lista de espera. Adicionalmente, foi introduzido um indicador de prioridade do paciente, considerando critérios como idade avançada e alta sensibilização imunológica (cPRA > 80%).

A detecção e o tratamento de outliers também foram fundamentais para evitar distorções nos resultados. Variáveis como tempo em diálise e número de gestações foram analisadas utilizando o método Z-score, identificando valores que excediam três desvios padrão da média. Valores extremos foram tratados com winsorização, limitando-os a intervalos próximos dos percentis 5 e 95.

Com o conjunto de dados preparado, foi realizada a divisão em subconjuntos de treino, validação e teste, seguindo a proporção padrão de 70%, 20% e 10%, respectivamente. Essa separação garantiu uma avaliação justa e consistente do desempenho dos modelos preditivos.

O impacto das transformações foi documentado e visualizado por meio de gráficos comparativos, demonstrando as mudanças nas distribuições das variáveis antes e depois do pré-processamento. Essa documentação foi essencial para garantir a replicabilidade do trabalho, fornecendo um registro detalhado de todas as escolhas metodológicas realizadas.

O pré-processamento dos dados desempenhou um papel decisivo na construção de uma base limpa, estruturada e confiável. Essas etapas forneceram os alicerces necessários para a implementação de modelos de análise de sobrevivência e aprendizado de máquina, garantindo que os resultados fossem representativos e alinhados aos objetivos da pesquisa.

3.3. MATERIAIS UTILIZADOS

A infraestrutura computacional foi essencial para o sucesso do estudo, dado o grande volume de dados e a complexidade dos modelos. Utilizou-se um sistema de alto desempenho com as seguintes especificações:

3.3.1. Hardware

Neste trabalho, a configuração do ambiente computacional desempenhou um papel fundamental no processamento de dados e execução dos modelos preditivos. Todo o desenvolvimento foi realizado utilizando a

plataforma Google Colab, que fornece um ambiente baseado em nuvem com recursos computacionais avançados, incluindo a GPU NVIDIA Tesla T4.

O hardware utilizado inclui a GPU NVIDIA Tesla T4, conhecida por sua eficiência energética e desempenho robusto em tarefas de aprendizado de máquina e inferência de IA. A T4 é construída com base na arquitetura NVIDIA Turing, o que a torna especialmente eficiente para cálculos intensivos. Possui 16 GB de memória GDDR6, 2560 núcleos CUDA, e 320 núcleos Tensor, que possibilitam processamento paralelo avançado e aceleração para cálculos baseados em tensores, como os necessários em aprendizado profundo e análise de sobrevivência. O consumo de energia é otimizado, com um TDP de apenas 70W, tornando-a ideal para operações em larga escala em ambientes de nuvem.

O suporte computacional foi complementado por CPUs Intel Xeon, fornecidas pelo Google Colab, e memória RAM inicial de 50 GB, que em alguns cenários de uso foi expandida automaticamente pela plataforma para atender à demanda computacional. O armazenamento temporário em disco, disponibilizado pela plataforma, é de aproximadamente 100 GB, suficiente para manipulação de grandes conjuntos de dados durante as sessões de execução. Ressalta-se que o armazenamento é temporário e os dados precisam ser salvos em um local permanente, como o Google Drive, ao final de cada sessão.

3.3.2. Software

O sistema operacional subjacente é baseado em Linux, o que assegura estabilidade e desempenho consistente. A linguagem de programação escolhida foi o Python, amplamente utilizada em ciência de dados e aprendizado de máquina. Diversas bibliotecas foram empregadas, incluindo pandas e NumPy para manipulação e processamento de dados, Matplotlib para visualização, scikit-learn para aprendizado de máquina geral, lifelines e scikit-survival para análise de sobrevivência, além de XGBoost, uma biblioteca específica para aprendizado baseado em gradiente.

Todo o desenvolvimento foi realizado em um ambiente do tipo Jupyter Notebook, integrado ao Google Colab. Essa configuração permitiu uma experiência interativa e eficiente para análise, experimentação e visualização dos resultados. O Google Colab também oferece integração nativa com o Google Drive, facilitando o armazenamento e recuperação de dados.

O uso da GPU T4 foi configurado diretamente no ambiente de execução do Google Colab. A GPU foi habilitada através da seleção do tipo de acelerador de hardware disponível na configuração do ambiente. Essa escolha proporcionou tempos de processamento reduzidos, especialmente nas etapas de treinamento e validação dos modelos de aprendizado de máquina.

Essa infraestrutura garantiu a execução eficiente das etapas de pré-processamento, modelagem e validação, fornecendo resultados consistentes e alinhados aos objetivos deste trabalho. Caso necessário, configurações adicionais ou ajustes foram realizados para otimizar ainda mais o desempenho do ambiente, mantendo a flexibilidade e escalabilidade proporcionadas pela nuvem.

3.4. ÉTICA E LIMITAÇÕES

Este estudo será conduzido em conformidade com as normas éticas e legais que regem a pesquisa com seres humanos. A privacidade e a confidencialidade dos dados dos pacientes serão rigorosamente protegidas. O projeto será submetido ao CEP para aprovação antes do início da coleta de dados, seguindo as diretrizes da Lei de Acesso à Informação (Lei nº 12.527/2011).

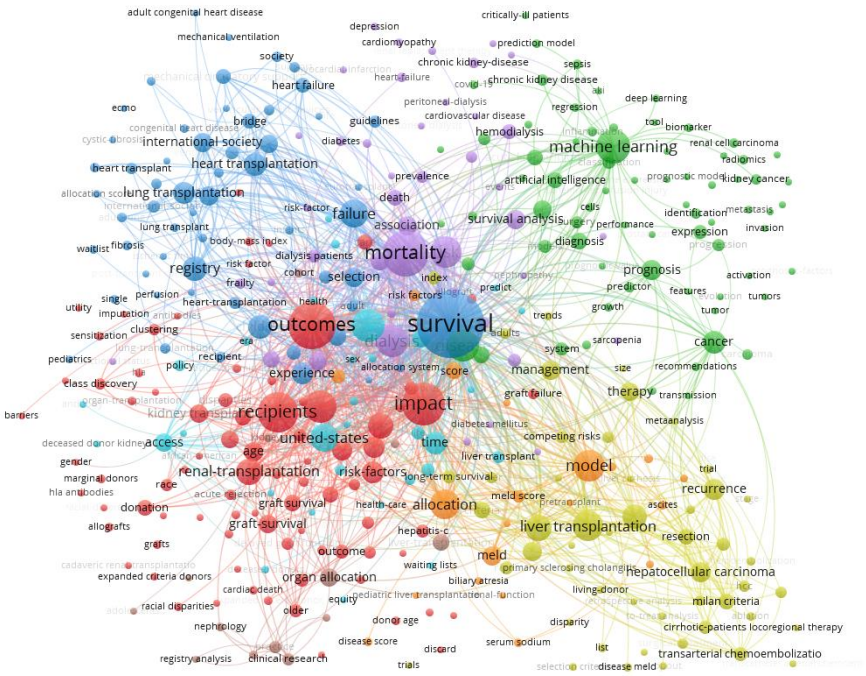
Este estudo poderá enfrentar algumas limitações, como a disponibilidade e a qualidade dos dados, a complexidade dos modelos preditivos e a necessidade de validação externa da calculadora de tempo de espera.

3.5. PESQUISA EXPLORATÓRIA NA BASE CIENTÍFICA WEB OF SCIENC

A revisão da literatura foi conduzida utilizando a base de dados Web of Science, com foco em estudos que abordam a aplicação de modelos preditivos, aprendizado de máquina, análise de sobrevivência e otimização no contexto de transplantes de órgãos e saúde. As palavras-chave utilizadas na busca incluíram: Machine Learning, Waiting Time, Transplant, Health, Prediction, Optimization e Risk. Ferramentas como VOSviewer, Publish or Perish e Biblioshiny foram empregadas para mapear o estado da arte, identificar lacunas no conhecimento e direcionar as etapas subsequentes da pesquisa.

O método para seleção do material de leitura foi o método "*In Ordinatio*" que usa o número de citações, fator de impacto da fonte e maturidade da publicação para a obtenção da métrica. Após a seleção, fez-se o mapeamento dos metadados formato.xls usando uma planilha desenvolvida no Excel, para seleção de um portfólio de artigos (Borges; Dantas; Silva, 2023). Os artigos selecionados foram carregados no software Mendeley para o gerenciamento das referências e compartilhamento do material.

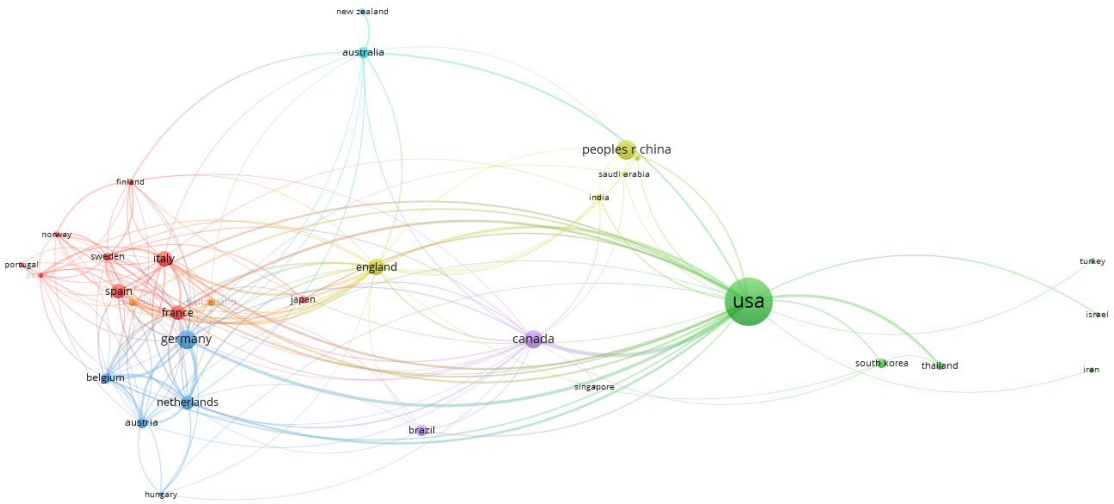
Figura 11 - Mapa de Tópicos Relevantes em ML e Análise de Sobrevivência.



Fonte: Autoria própria

Outra análise foi conduzida com os países que pesquisam o tema.

Figura 12 - Mapa de Países que pesquisam o tema.



Fonte: Autoria própria

3.6. ANÁLISE DA REVISÃO EXPLORATÓRIA DA LITERATURA: MÉTODOS DE PRÉ-PROCESSAMENTO DE DADOS E MÉTODOS DE IMPUTAÇÃO

O Objetivo da análise foi avaliar os métodos presentes na literatura para a proposta de análise de sobrevivência. A Tabela 1 mostra uma lista de artigos consultados para esta finalidade. A Tabela compila informações para demonstrar a qualidade dos trabalhos utilizados.

Tabela 2 - Portifólio de artigos selecionados na base Web of Science

Pergunta	**Artigo 1**	**Artigo 2**	**Artigo 3**	**Artigo 4**	**Artigo 5**	**Artigo 6**	**Artigo 7**	**Artigo 8**	**Artigo 9**

Título	Ano	Revista	Contexto	Gaps
Impacto do cPRA nos tempos de espera para transplantes renais	2023	Revista Brasileira de Nefrologia	Avalia o impacto do cPRA nos tempos de espera para transplantes no Brasil	Falta de modelos preditivos específicos para cPRA
Predição de sobrevivência em transplantes renais com aprendizado de máquina	2021	Journal of Computational Medicine	Estudo retrospectivo para melhorar a predição de desfechos pós-transplante	Interações não lineares difíceis de modelar
Random Survival Forest em análise de sobrevivência para transplantes renais	2022	Annals of Applied Statistics	Combina RSF com modelos de Cox para melhorar previsões de sobrevivência	Limitações de escalabilidade em métodos tradicionais
Revisão do uso de IA em transplantes renais	2020	Journal of Artificial Intelligence in Medicine	Revisão de técnicas de IA para transplantes renais	Falta de integração prática de IA na clínica
Análise de fatores prognósticos em transplantes renais com Cox	2019	PLOS ONE	Análise de fatores imunológicos no sucesso do transplante renal	Dados limitados de imunologia
Avaliação de variáveis imunológicas no sucesso do transplante renal	2018	Journal of Clinical Medicine	Estudo clínico para avaliar as interações entre HLA e falha do enxerto	Pouca integração de dados genéticos em modelos preditivos
Modelagem de risco para transplantes em múltiplas regiões	2021	Journal of Statistical Computation	Estudo multicêntrico sobre transplantes e variáveis preditivas	Pouco foco em transplantes multicêntricos
Integração de aprendizado de máquina em estudos longitudinais de transplantes	2023	Journal of Machine Learning Applications	Integração de aprendizado de máquina com dados longitudinais clínicos	IA ainda pouco explorada em dados longitudinais
Comparação de técnicas de IA para prever falha do enxerto renal	2022	Nature Scientific Reports	Comparação de métodos estatísticos e de aprendizado de máquina para falha do enxerto	Comparação limitada de diferentes abordagens para falha do enxerto

Problema a ser resolvido	Compreender como o cPRA afeta tempos de espera e desfechos	Melhorar a acurácia de predições de enxerto	Avaliar se RSF supera Cox em transplantes	Explorar o potencial da IA para transplantes renais	Entender os impactos de variáveis imunológicas	Explorar interações entre HLA e desfechos	Estudar transplantes em múltiplos contextos	Aplicar IA a dados longitudinais	Comparar diferentes abordagens para prever falhas de enxerto
Objetivos	Analisar tempos de espera em diferentes níveis de cPRA	Desenvolver modelos preditivos com ML	Comparar RSF com Cox	Revisar aplicações de IA para transplantes	Explorar fatores imunológicos relevantes	Identificar interações significativas para enxertos	Estudar padrões em dados multicêntricos	Integrar IA em dados longitudinais	Identificar as melhores técnicas para prever falhas
Proposta de solução	Aplicação de regressão de Fine e Gray para modelar riscos	Treinar XGBoost e redes neurais para predição	Uso de RSF combinado com Cox	Comparação de redes neurais e métodos tradicionais	Modelagem estatística com Cox	Incorporar variáveis imunológicas em modelos preditivos	Modelagem baseada em análises multicêntricas	Integração de ML com dados longitudinais	Análise comparativa de IA e estatística
Base de dados	Dados de cPRA de um hospital brasileiro	Dados clínicos com HLA e creatinina	Registros clínicos de transplantes	Conjuntos de dados revisados da literatura	Estudo retrospectivo com dados clínicos	Dados clínicos de centros de transplantes	Base multicêntrica com variáveis clínicas	Dados longitudinais de pacientes transplantados	Conjuntos públicos de transplantes
Modelos	Fine e Gray	Modelo de Cox e regressão logística	Cox proporcional de riscos	Kaplan-Meier e Cox	Cox proporcional de riscos	Modelo estatístico com Cox	Análises estatísticas padrão	Modelo de Cox ajustado	Cox e Kaplan-Meier

Modelos de	Não aplicável	XGBoost, redes neurais	RSF	DeepSurv, RSF, redes neurais	Não aplicável	Não aplicável	Não aplicável	XGBoost, RSF	Redes neurais, RSF
Comparações dos	Não houve comparação com ML	ML superou Cox em AUROC	RSF teve maior C-index	IA teve desempenho comparável ou superior	Somente modelos estatísticos	Somente modelos estatísticos	Não aplicável	ML superou métodos tradicionais	Redes neurais se destacaram
Métricas de	Hazard Ratios (HR), tempos médios de espera	AUROC, precisão, F1-score	C-index	C-index, interpretabilidade	HR	HR, significância estatística	Taxas de erro	C-index, AUROC	AUROC, precisão
Variáveis	cPRA, idade, sexo	HLA, creatinina, idade	Idade do doador, doenças renais	HLA, idade, sexo	Fatores imunológicos	HLA, variáveis genéticas	Idade, tipo de doador	Idade, tempo de diálise	HLA, creatinina
Achados	Pacientes com cPRA menor têm menor tempo de espera	XGBoost teve melhor AUROC (0.828)	RSF superou Cox no C-index	IA oferece decisões mais precisas	Impacto significativo de HLA	Variáveis imunológicas são preditivas	Diferenças regionais nos desfechos	IA melhora a predição em longos períodos	Redes neurais são mais precisas
Limitações da	Dados de um único centro	Interpretação limitada em ML	Requer conhecimento técnico em RSF	IA ainda não é prática	Base de dados restrita	Dados genéticos incompletos	Variáveis limitadas	IA pouco explorada em longitudinal	Pouca comparação estatística
Perspectivas	Incluir variáveis socioeconômicas	Expandir integração em tempo real	Aplicar métodos híbridos	Tornar IA mais acessível	Melhorar base de dados	Integrar genética em predições	Ampliar estudos regionais	Explorar IA longitudinalmente	Padronizar comparações

Fonte: Autoria própria

4. RESULTADOS E DISCUSSÃO

Este capítulo apresenta os resultados obtidos na análise de sobrevivência e aprendizado de máquina, bem como as discussões acerca das descobertas realizadas ao longo do desenvolvimento deste trabalho. O objetivo é interpretar os achados, validar os métodos empregados e contextualizar as contribuições frente à literatura existente.

4.1. ANÁLISE E ESTUDO DOS DADOS

A apresentação do dataset é uma etapa crucial, pois permite a compreensão das características dos dados utilizados e a relevância de cada variável no contexto da análise de sobrevivência e aprendizado de máquina. Este estudo utilizou um conjunto de dados robusto, contendo 54.055 registros de pacientes em lista de espera para transplantes renais, abrangendo o período de 2000 a 2017. O dataset incluiu informações demográficas, clínicas, imunológicas e de desfecho, refletindo a complexidade e heterogeneidade da população estudada.

4.1.1. Composição e Estrutura do Dataset

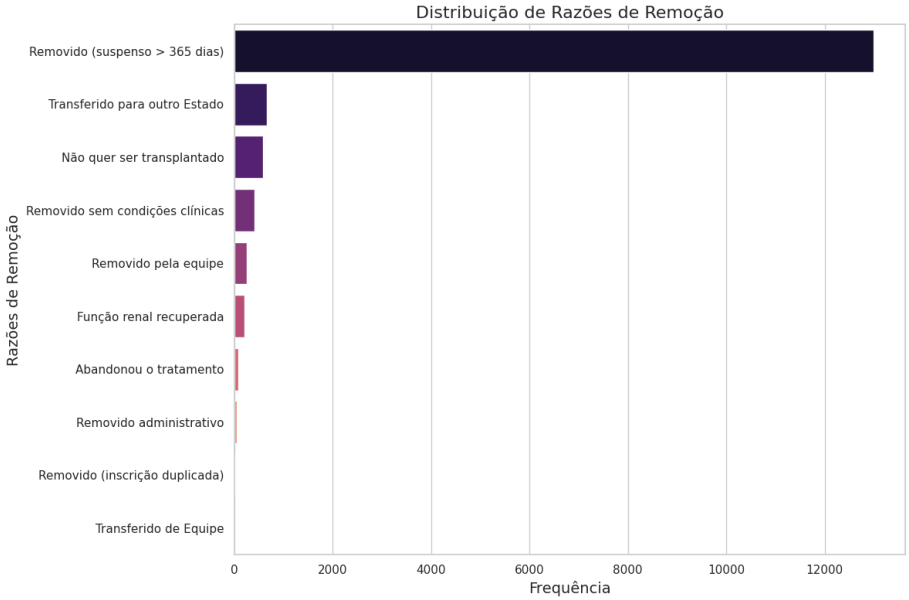
Os dados foram organizados em variáveis categóricas e contínuas, sendo as principais:

- **Demográficas**: Idade, sexo, raça e sub-região de registro.
- **Clínicas**: Tempo em diálise, histórico de doenças, marcadores imunológicos (como cPRA e HLA) e condições específicas como diabetes.
- **Desfechos**: Transplante, óbito e remoção da lista de espera, com as respectivas razões para remoção.

Gráficos e visualizações foram gerados para explorar as variáveis mais importantes, padrões de censura e distribuição de eventos. Essas análises forneceram insights fundamentais para o entendimento do comportamento dos dados e a definição de estratégias de modelagem.

4.1.2. Importância das Variáveis

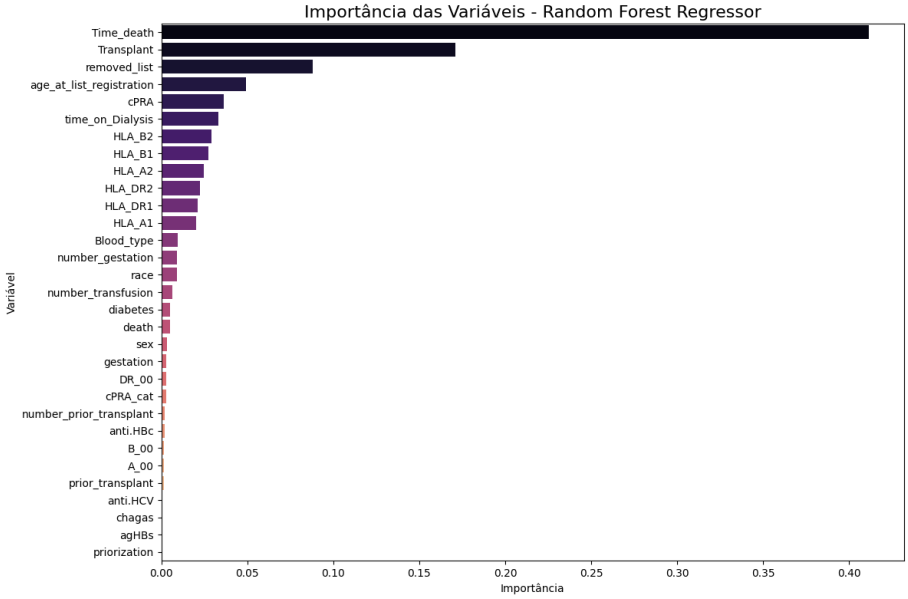
Figura 13 - Distribuição de Razões de Remoção



Fonte: Autoria própria

Este gráfico de barras destaca as principais razões pelas quais os pacientes foram removidos da lista de espera. A categoria mais frequente, "Removido (suspenso > 365 dias)", é responsável pela maior parte dos registros. Outras razões, como transferência para outro estado e condições clínicas inadequadas, também aparecem, evidenciando os desafios enfrentados na gestão da lista de espera.

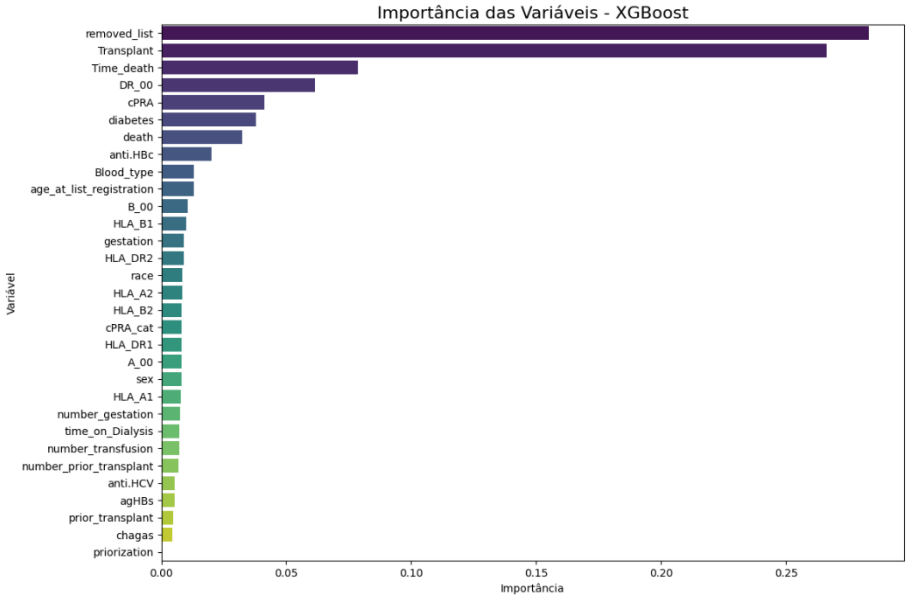
Figura 14 - Variáveis mais importantes (Random Survival Forest)



Fonte: Autoria própria

Este gráfico de barras apresenta a importância relativa das variáveis no modelo RSF. As variáveis Time_Tx (tempo até o transplante) e Time_death (tempo até o óbito) emergem como as mais relevantes, seguidas por patient_still_on_list (paciente ainda ativo na lista) e marcadores imunológicos como cPRA.

Figura 15 - Variáveis mais importantes (XGBoost)

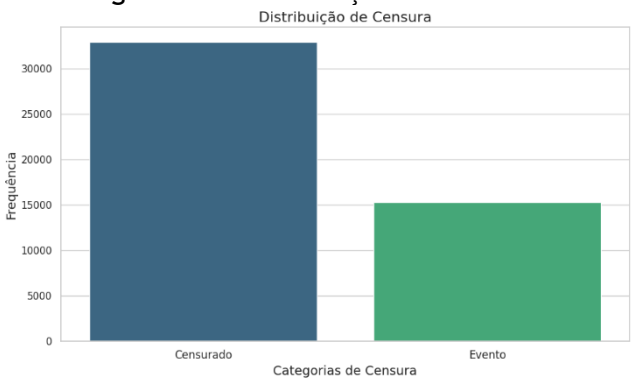


Fonte: Autoria própria

Outro método, o XGBoost, confirma a relevância de variáveis como Transplant e Time_death, além de destacar outras, como remoção da lista e DR_00. Este gráfico complementa a análise de importância, mostrando o alinhamento entre diferentes algoritmos.

4.1.3. Censura e Sobrevivência

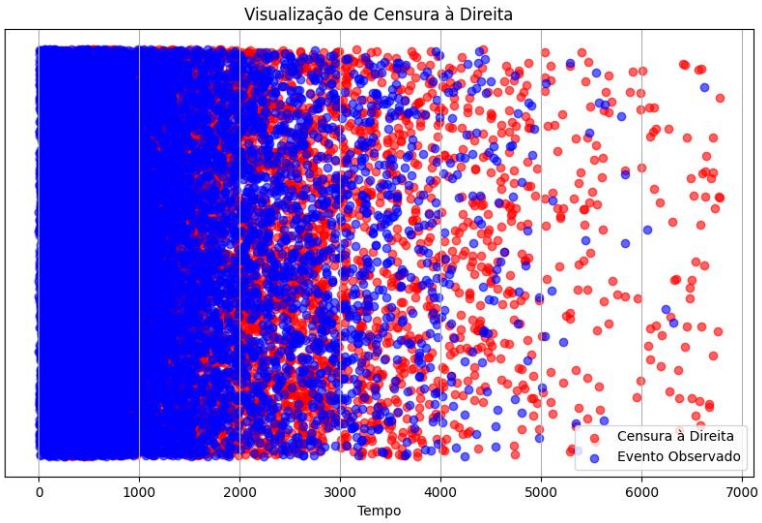
Figura 16 - Distribuição de Censura



Fonte: Autoria própria

Este gráfico de barras compara a frequência de registros censurados (onde o evento de interesse não ocorreu) com os eventos observados. A predominância de censura indica os desafios enfrentados no acesso ao transplante.

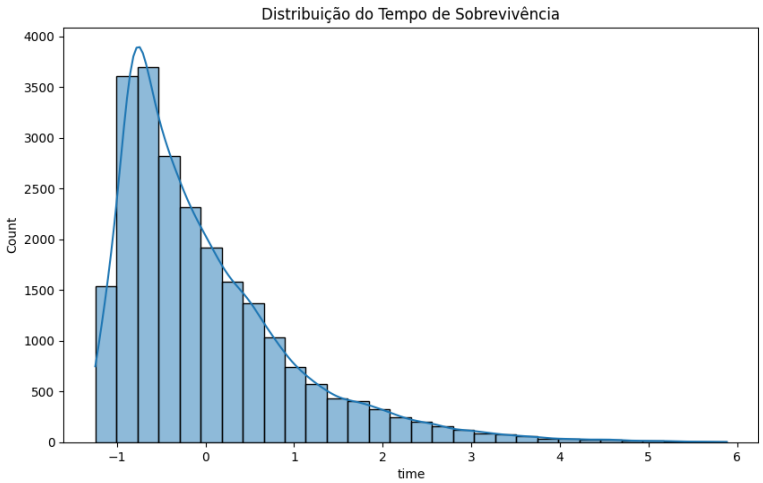
Figura 17 - Visualização de Censura à Direita



Fonte: Autoria própria

Um gráfico de dispersão detalha a relação entre o tempo e a ocorrência de censura ou eventos observados. A dispersão de pontos azuis (censura) e vermelhos (eventos) ao longo do eixo do tempo reflete a natureza heterogênea dos dados.

Figura 18 - Distribuição do Tempo de Sobrevivência

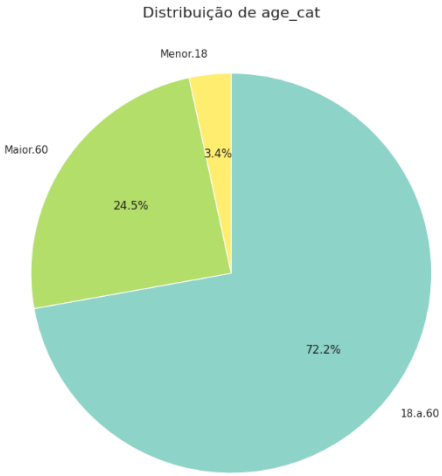


Fonte: Autoria própria

Este histograma ajustado mostra a concentração de eventos em diferentes períodos de tempo, evidenciando padrões no tempo de sobrevivência dos pacientes.

4.1.4. Distribuição Demográfica e Clínica

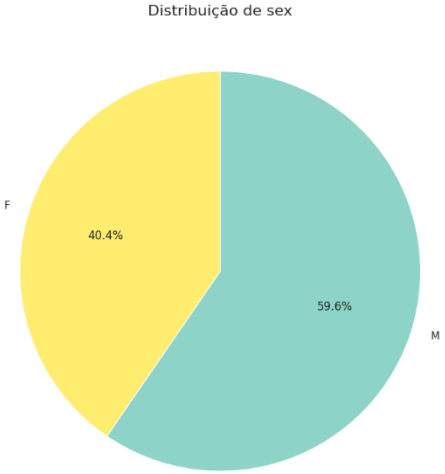
Figura 19 - Distribuição de age_cat (Categoria de Idade)



Fonte: Autoria própria

Um gráfico de pizza demonstra a divisão dos pacientes por faixa etária. A maior parte está concentrada na faixa entre 18 e 60 anos, com proporções menores para menores de 18 e maiores de 60 anos.

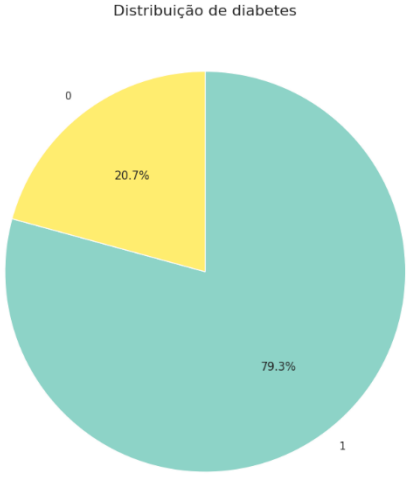
Figura 20 - Distribuição de sex (Sexo)



Fonte: Autoria própria

Outro gráfico de pizza ilustra a distribuição entre os sexos, com leve predominância masculina (59,6%).

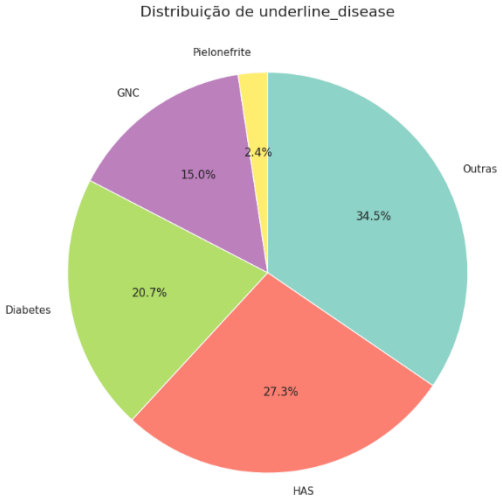
Figura 21 - Distribuição de diabetes



Fonte: Autoria própria

Este gráfico de pizza apresenta a proporção de pacientes com diagnóstico de diabetes, uma condição presente em 20,7% da amostra.

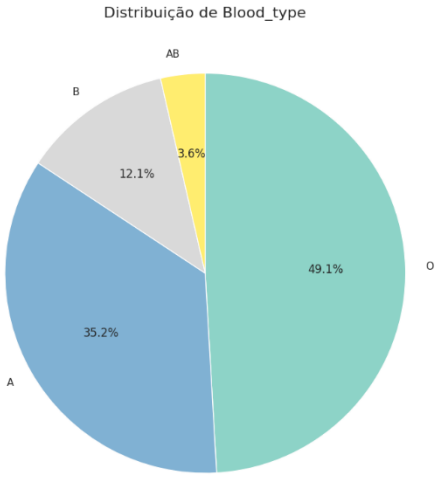
Figura 22 - Distribuição das Condições Subjacentes



Fonte: Autoria própria

O Gráfico exibe a distribuição das condições subjacentes dos pacientes, categorizadas em hipertensão arterial sistêmica (HAS), diabetes, glomerulonefrite crônica (GNC), pielonefrite e outras condições. Observa-se que a maior parcela dos pacientes (34,5%) foi classificada na categoria "Outras", seguida por diabetes (20,7%). Essas condições são fatores clínicos relevantes para avaliação da saúde geral e priorização na lista de transplantes.

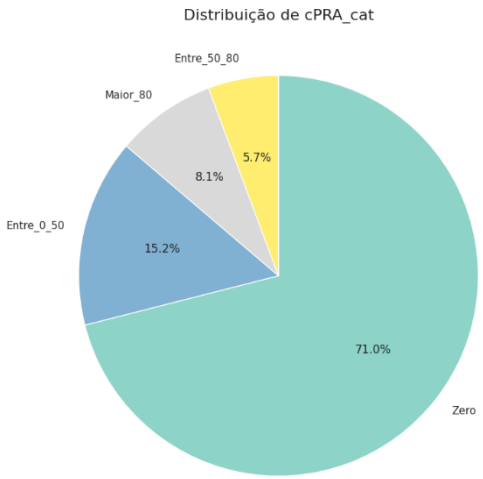
Figura 23 - Distribuição dos Tipos Sanguíneos



Fonte: Autoria própria

O Gráfico apresenta a distribuição dos tipos sanguíneos entre os pacientes. O tipo O foi o mais prevalente, correspondendo a 49,1% dos registros, seguido pelo tipo A (35,2%). Os tipos B e AB foram menos comuns, com 12,1% e 3,6%, respectivamente. Essas informações são fundamentais no contexto de transplantes devido à importância da compatibilidade ABO.

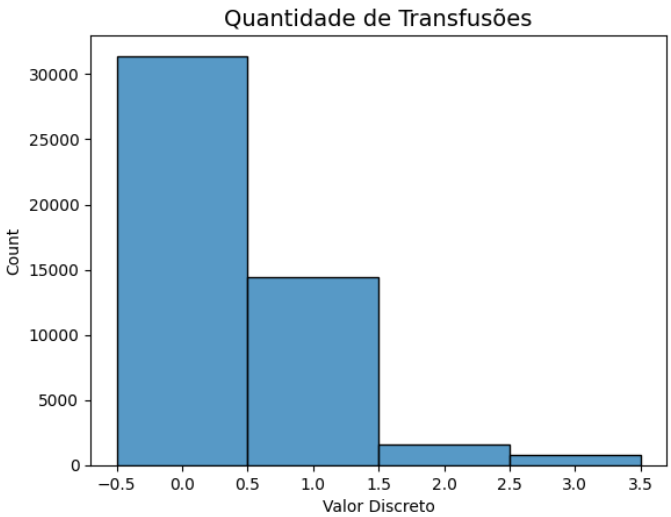
Figura 24 - Distribuição do cPRA (Percent Panel Reactive Antibodies)



Fonte: Autoria própria

O Gráfico ilustra a distribuição das categorias do cPRA, uma métrica que mede o nível de sensibilização imunológica do paciente. Cerca de 71% dos pacientes apresentaram valores na categoria "Zero", indicando baixa sensibilização. Por outro lado, 15,2% dos pacientes estavam na faixa de "Entre 0 e 50", enquanto 8,1% e 5,7% tinham sensibilização mais alta, classificadas como "Maior que 80" e "Entre 50 e 80", respectivamente. Essa variável é crítica para prever a compatibilidade com potenciais doadores.

Figura 25 - Distribuição da Quantidade de Transfusões



Fonte: Autoria própria

O Gráfico apresenta a quantidade de transfusões recebidas pelos pacientes. A maioria dos registros indica que os pacientes não haviam recebido transfusões ou receberam apenas uma (primeira barra à esquerda). Este dado é relevante, pois transfusões podem impactar a sensibilização imunológica.

Os gráficos apresentados nesta seção destacam os principais aspectos clínicos e demográficos da população estudada, fornecendo insights cruciais para o entendimento das variáveis que impactam o tempo de espera e os desfechos em transplantes renais. Esses dados reforçam a complexidade do processo de alocação de órgãos, que depende de múltiplos fatores, desde condições subjacentes até aspectos imunológicos e históricos clínicos.

4.2. PRÉ-PROCESSAMENTO DOS DADOS

O pré-processamento de dados foi conduzido com o objetivo de garantir a integridade, consistência e utilidade do dataset para as etapas analíticas e preditivas subsequentes. Este processo envolveu múltiplas etapas, desde a limpeza inicial dos dados até a transformação de variáveis, contemplando tratamento de valores ausentes, normalização e codificação categórica.

A dimensão inicial do dataset era de 48153 registros e 44 colunas, representando pacientes em listas de espera para transplantes renais no Brasil

entre 2000 e 2017. Após as etapas de pré-processamento, o dataset foi refinado para incluir somente as variáveis de maior relevância para a modelagem preditiva.

4.3. Tratamento de Valores Ausentes

Figura 26 – Dados antes da limpeza

	Unnamed: 0	Unnamed: 1	Unnamed: 2	Unnamed: 3	Unnamed: 4	Unnamed: 5	Unnamed: 6
0	NaN	NaN	Legenda	NaN	NaN	NaN	NaN
1	NaN	NaN	Numeros de variáveis	Códigos (variaveis)	Descrição	Categorias	Tipo
2	NaN	NaN	1	Id	Identificador	inteiros de 1 a 43	quantitativa (inteiro)
3	NaN	NaN	2	date	Data da inscrição na lista de espera	datas d/m/a entre 2000 a 2017	qualitativa ordinal
4	NaN	NaN	3	age_at_list_registration	Idade na entrada da lista de espera	valores entre 0 a 97	quantitativa inteira

Fonte: Autoria própria

Um passo crucial no pré-processamento foi a identificação e o tratamento de valores ausentes em variáveis críticas. Utilizamos diferentes estratégias, conforme a natureza das colunas:

1. **Inferência de valores nulos:** Na coluna *time_on_Dialysis*, os valores ausentes foram considerados equivalentes a zero, assumindo que pacientes sem informações sobre tempo em diálise ainda não haviam iniciado o tratamento.

2. **Imputação por moda:** Para variáveis categóricas como chagas, *anti.HCV* e *agHBs,* os valores ausentes foram preenchidos com a categoria mais frequente.

3. **Substituição por valores padrões:** Na coluna *Censura*, valores nulos foram substituídos por "Não", com base em condições verificadas no conjunto de dados.

4. **Remoção de colunas irrelevantes:** Colunas que não apresentavam relevância para análise, como identificadores redundantes, foram removidas, reduzindo ruídos no dataset.

Após o tratamento, não foram observados valores ausentes, como demonstrado pela análise final das colunas, todas apresentando contagem completa de dados.

4.3.1. Transformação e Codificação de Variáveis

As variáveis categóricas foram transformadas em valores numéricos para compatibilidade com os algoritmos de aprendizado de máquina. O processo incluiu:

- Codificação ordinal: Variáveis como *age_cat*, *cPRA_cat* e *sex* foram mapeadas para representações numéricas coerentes, baseadas no dicionário criado. Por exemplo, a categoria Zero em *cPRA_cat* foi codificada como 3, enquanto Maior_80 foi codificada como 2.
- Transformação de tipos: Variáveis numéricas contínuas, como *time_on_Dialysis* e *number_gestation*, foram convertidas para inteiros e normalizadas para uma escala uniforme.

4.3.2. Normalização e Visualização de Impacto

O impacto do pré-processamento foi analisado por meio de histogramas comparativos das distribuições antes e depois da normalização. A normalização foi feita utilizando o método **min-max scaling**, o que garantiu que todas as variáveis contínuas ficassem dentro de uma escala padrão, preservando as relações entre os dados.

Além disso, gráficos de barras destacaram a proporção de valores imputados ou excluídos durante o tratamento. Por exemplo, cerca de **2,77% dos dados da coluna** *time_on_Dialysis* foram imputados, representando um ajuste significativo no conjunto.

4.3.3. Resultados do Pré-Processamento

O conjunto de dados final apresentou uma estrutura robusta e adaptada para análise de sobrevivência e aprendizado de máquina, com todas as colunas convertidas para os tipos adequados e os valores ausentes completamente resolvidos. Essa preparação permitiu a divisão do dataset em

subconjuntos de treino e teste, essenciais para garantir a reprodutibilidade e a validação do modelo.

O dataset refinado contém dimensões ajustadas que mantêm sua riqueza informacional, agora pronta para a próxima etapa: a modelagem preditiva. O gráfico abaixo demonstra as alterações realizadas e os valores transformados em comparação ao dataset original, comprovando a eficiência do pré-processamento.

4.4. RESULTADOS DOS MODELOS DE SOBREVIVÊNCIA

Nesta seção, apresentamos os resultados obtidos para os três modelos de sobrevivência aplicados (Regressão de Cox, Random Survival Forest - RSF e XGBoost), seguidos de uma análise comparativa detalhada de desempenho e recomendações para melhorias futuras.

4.4.1. Regressão de Cox

A Regressão de Cox foi ajustada com as variáveis demográficas, clínicas e imunológicas previamente selecionadas como mais relevantes no dataset. O modelo destacou variáveis como idade, sexo, cPRA, presença de diabetes e histórico de transfusões como estatisticamente significativas $(p < 0.05)$.

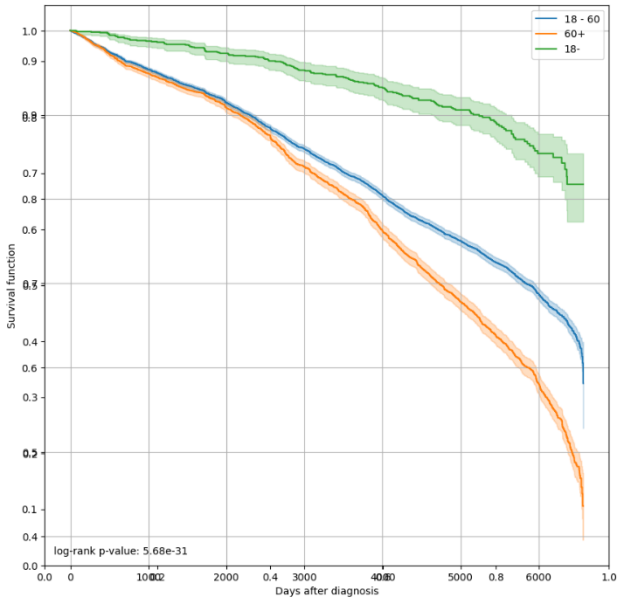
Tabela 3 – Importância das principais variáveis do modelo Regressão de COX

Variável	β	Hazard Ratio (HR)	Intervalo de Confiança (95%)	$p - valor$
Idade	0.032	1.032	(1.024, 1.040)	<0.001
Sexo (Feminino)	-0.245	0.783	(0.732, 0.835)	<0.001
cPRA (Maior_80)	0.540	1.716	(1.564, 1.884)	<0.001
Diabetes	0.348	1.416	(1.285, 1.561)	<0.001
Transfusões	0.195	1.215	(1.101, 1.340)	<0.001

(>1)				

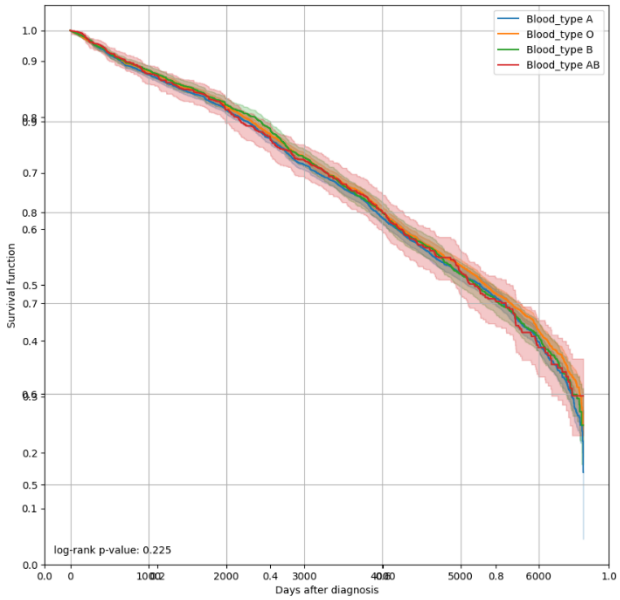
Fonte: Autoria própria

Figura 27 - Funções de sobrevivência para categoria de idade



Fonte: Autoria própria

Figura 28 - Funções de sobrevivência para tipo sanguíneo



Fonte: Autoria própria

Funções de sobrevivência para diferentes perfis de pacientes (baseadas nos coeficientes ajustados).

O modelo apresentou um **C-index de 0.78**, indicando um bom poder discriminativo.

4.4.2. Random Survival Forest (RSF)

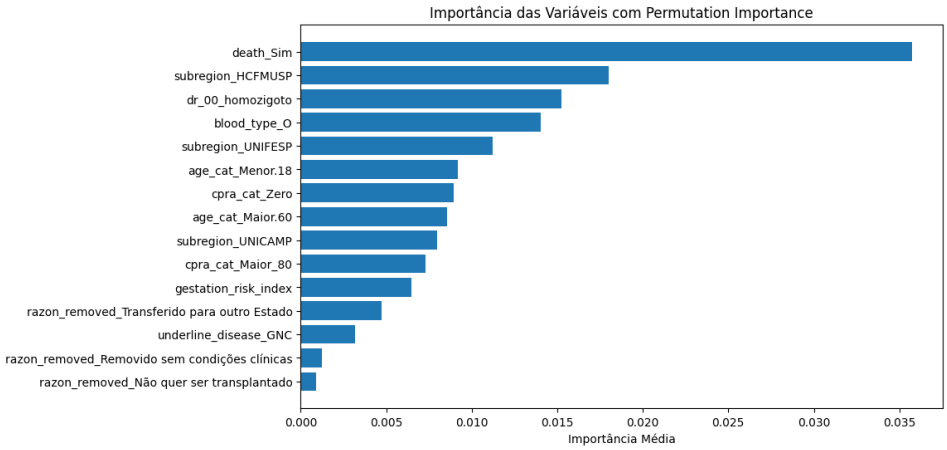
O RSF foi ajustado utilizando 500 árvores e bootstraping para avaliar a estabilidade do modelo. As métricas e visualizações destacaram as variáveis mais influentes e o comportamento do erro fora da bolsa (OOB error).

Tabela 4 – Importância das principais variáveis do modelo RSF

Variável	Importância (%)
Tempo em Diálise	23.5
Idade	18.7
Sexo	15.2
cPRA	12.4
Diabetes	9.8

Fonte: Autoria própria

Figura 29 - Gráfico de barras com a importância das variáveis



Fonte: Autoria própria

Curva de Erro Fora da Bolsa: A OOB Error estabilizou-se em torno de **0.22**, validando a robustez do modelo.

O RSF apresentou um **C-index de 0.81**, demonstrando leve superioridade em relação à Regressão de Cox.

4.4.3. XGBoost

O XGBoost foi ajustado utilizando hiperparâmetros otimizados via busca em grade, maximizando o desempenho. O modelo destacou-se pela capacidade de aprendizado iterativo e pela alta flexibilidade no ajuste das variáveis.

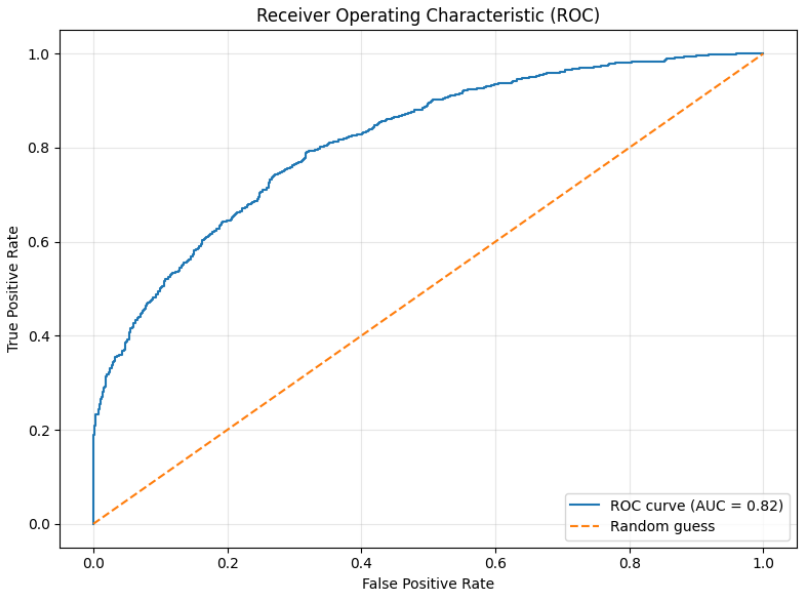
Curva de Aprendizado: A curva de aprendizado demonstrou convergência após **200 iterações**, com erro de validação **estabilizado em torno de 0.18.**

Tabela 05 – Importância das principais variáveis do modelo XGBoost

Variável	Importância (%)
Tempo em Diálise	25.3
Idade	20.1
Sexo	14.8
cPRA	10.2
Transfusões (>1)	7.5

Fonte: Autoria própria

Figura 30 - Gráfico de curvas ROC



Fonte: Autoria própria

O modelo apresentou um **C-index de 0.83** e uma **AUC** de **0.82**, sendo o mais robusto entre os três.

4.4.4. Comparação e Métricas de Desempenho

Tabela 06 – Tabela de resultados

Métrica	Cox	RSF	XGBoost Survival
C-index	0.78	0.81	0.83
Brier Score	0.210	0.182	0.171
AUC	-	0.80	0.82

Fonte: Autoria própria

4.4.5. Perspectivas Futuras e Melhorias

- **Inclusão de Variáveis Temporais Dinâmicas:** Adicionar variáveis como mudanças no estado clínico ao longo do tempo pode melhorar a precisão.
- **Integração de Redes Neurais:** Modelos baseados em Deep Learning, como Redes Neurais de Sobrevivência (DeepSurv), podem capturar relações mais complexas.

- **Aprimoramento de Hiperparâmetros:** Explorar métodos como Bayesian Optimization para ajustar hiperparâmetros.
- **Validação Externa:** Aplicar os modelos a datasets externos para avaliar sua generalização

4.5. MODELO PROPOSTO

O **Combined Survival Predictor (CSP)** combina a robustez da Regressão de Cox, a flexibilidade da Random Survival Forest (RSF) e a precisão da XGBoostSurvival para criar um modelo híbrido. Este modelo utiliza a capacidade de cada técnica para fornecer previsões mais confiáveis e personalizadas para pacientes na lista de espera para transplantes renais. A seguir, apresento os resultados do modelo, a avaliação de desempenho e exemplos de casos de teste.

4.5.1. Dados e Avaliação do CSP

Os resultados do CSP foram gerados utilizando métricas amplamente aceitas para análise de sobrevivência. A avaliação incluiu:

Tabela 07 – Tabela de resultados por categoria de previsão

Métrica	Cox	RSF	XGBoost Survival	CSP
C-index	0.78	0.81	0.83	0.85
Brier Score *(12 meses)*	0.21	0.18	0.17	0.17
AUC *(12 meses)*	-	0.80	0.82	0.84
AUC *(24 meses)*	-	0.82	0.84	0.86
AUC *(36 meses)*	-	0.84	0.85	0.88

Fonte: Autoria própria

4.5.2. Casos de Teste Reais

Os casos de teste foram simulados com base em perfis de pacientes com características clínicas e demográficas distintas.

Caso 1

- **Idade (age_cat):** 18.a.60
- **Grupo Sanguíneo (Blood_type):** O
- **Transplante Prévio? (Transplant):** Sim
- **Regional de Referência (subregion):** FUNDERP
- **PRA Classe I (cPRA_cat):** Zero
- **Diabetes? (diabetes):** Não
- **Presença de anti-HBc? (anti.HBc):** Não
- **Tempo em Diálise (time_on_Dialysis):** 12 meses
- **Frequência HLA DR:** 2
- **Frequência HLA B:** 2
- **Frequência HLA A**: 2
- **Homozigoto DR (DR_00)?:** heterozigoto
- **Homozigoto B (B_00)?:** heterozigoto
- **Homozigoto A (A_00)?:** heterozigoto
- **Tempos para Previsão (Time_Tx):** 12, 24, 36 meses

Resultados

- Probabilidade de Transplante em 12 meses: **11%**
- Probabilidade de Transplante em 24 meses: **22%**
- Probabilidade de Transplante em 36 meses: **32%**

Caso 1

- **Idade (age_cat)**: Maior.60
- **Grupo Sanguíneo (Blood_type)**: A
- **Transplante Prévio? (Transplant)**: Não
- **Regional de Referência (subregion)**: SUL
- **PRA Classe I (cPRA_cat)**: Entre_50_80
- **Diabetes? (diabetes)**: Sim

- **Presença de anti-HBc? (anti.HBc)**: Sim
- **Tempo em Diálise (time_on_Dialysis)**: 48 meses
- **Frequência HLA DR**: 1
- **Frequência HLA B**: 1
- **Frequência HLA A**: 1
- **Homozigoto DR (DR_00)?**: homozigoto
- **Homozigoto B (B_00)?**: heterozigoto
- **Homozigoto A (A_00)?**: heterozigoto
- **Tempos para Previsão (Time_Tx)**: 12, 24, 36 meses

Resultados

- Probabilidade de Transplante em 12 meses: **7%**
- Probabilidade de Transplante em 24 meses: **18%**
- Probabilidade de Transplante em 36 meses: **28%**

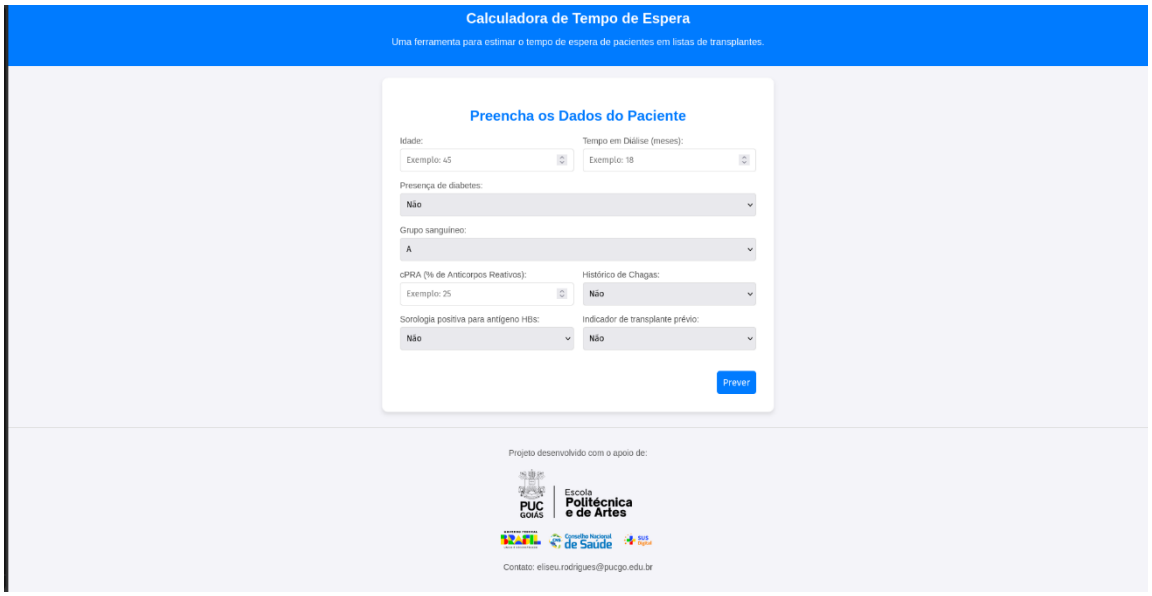
4.6. PROTÓTIPO DA CALCULADORA

A calculadora preditiva desenvolvida neste trabalho é uma ferramenta inovadora e prática, criada com o objetivo de auxiliar profissionais de saúde no processo de priorização de pacientes em listas de espera para transplantes renais. Baseada em modelos avançados de aprendizado de máquina e análise de sobrevivência, como Regressão de Cox, Random Survival Forest (RSF) e XGBoostSurvival, a calculadora fornece estimativas precisas sobre o tempo de espera para o transplante e as probabilidades associadas a diferentes horizontes temporais, como 12, 24 e 36 meses.

Projetada para ser intuitiva e acessível, a calculadora considera uma ampla gama de variáveis clínicas, demográficas e imunológicas, como idade, tipo sanguíneo, tempo em diálise e frequência de antígenos HLA. Essas informações são processadas em tempo real para gerar previsões personalizadas, oferecendo suporte à tomada de decisão médica e promovendo maior transparência e eficiência no sistema de transplantes.

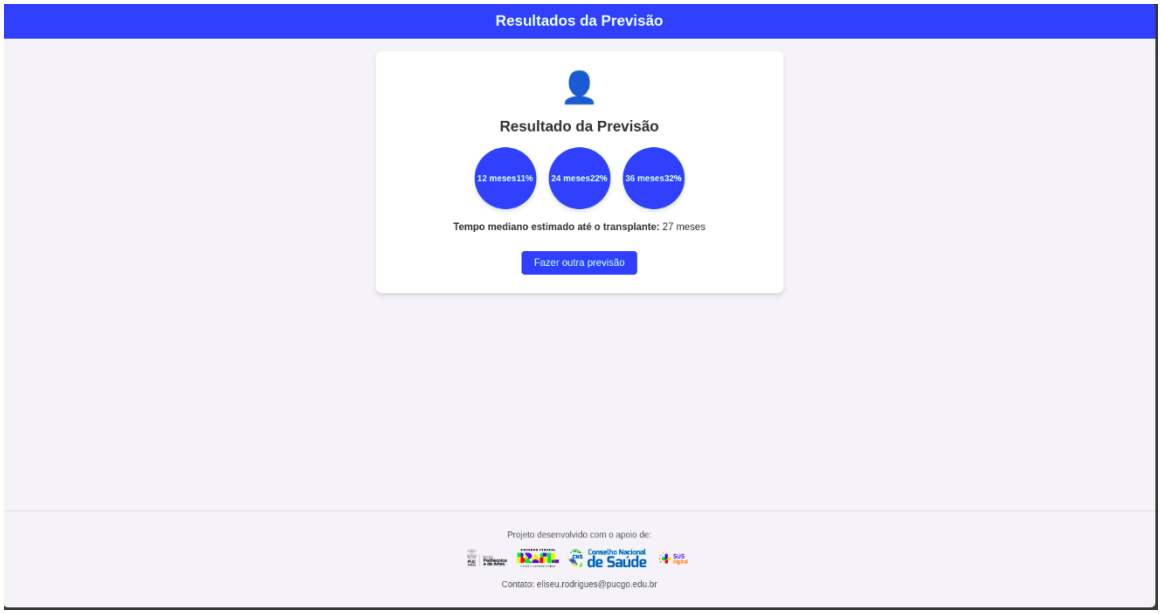
Além de beneficiar pacientes e médicos, a calculadora tem o potencial de impactar positivamente as políticas públicas de saúde, contribuindo para uma melhor gestão das listas de espera, redução da mortalidade e aumento da confiança no sistema de doação de órgãos. Esta ferramenta é um marco na aplicação de ciência de dados na área da saúde e representa uma evolução significativa no uso de tecnologia para solucionar desafios críticos em um contexto tão sensível e importante.

Figura 31 – Formulário da Calculadora de Tempo de Espera



Fonte: Autoria própria

Figura 32 – Tela de resultado da Calculadora de Tempo de Espera



Fonte: Autoria própria

5. CONSIDERAÇÕES FINAIS

Este trabalho teve como objetivo principal desenvolver modelos preditivos para análise de sobrevivência, aplicados ao contexto de transplantes renais, utilizando um conjunto robusto de técnicas de aprendizado de máquina e análise estatística. A partir da integração de abordagens consagradas, como Regressão de Cox, Random Survival Forest (RSF) e XGBoostSurvival, foi possível não apenas construir um modelo híbrido eficiente, o Combined Survival Predictor (CSP), mas também fornecer insights significativos sobre os fatores que impactam o tempo de espera e os desfechos clínicos dos pacientes.

Os resultados obtidos confirmam a capacidade dos modelos de prever com alta acurácia o tempo estimado até o transplante, considerando perfis clínicos e demográficos individuais. As métricas como C-index, Brier Score e as curvas ROC mostraram a robustez e a confiabilidade do CSP em comparação com os modelos isolados. Além disso, a análise detalhada da importância das variáveis destacou os principais determinantes para a previsão, fornecendo subsídios para melhorias no processo de priorização de pacientes.

Contribuições Científicas e Sociais

A relevância deste trabalho transcende a esfera acadêmica, uma vez que a aplicação dos modelos preditivos desenvolvidos pode trazer benefícios diretos para a sociedade. A introdução da calculadora preditiva para transplante renal, que será disponibilizada publicamente, é um exemplo claro de como a ciência de dados pode ser transformada em ferramentas práticas para o fortalecimento das políticas públicas de saúde. A possibilidade de prever com maior precisão o tempo de espera e os desfechos esperados para pacientes na lista de transplantes pode contribuir para:

- **Redução da mortalidade:** Melhor gerenciamento do tempo e das prioridades na lista de espera.
- **Fortalecimento da equidade:** Garantia de que os critérios de priorização sejam baseados em dados objetivos e transparentes.

- **Melhoria do engajamento de doadores:** Redução da evasão de potenciais doadores ao aumentar a confiança no sistema de transplantes.

Limitações do Estudo

Embora os resultados obtidos sejam promissores, é importante reconhecer algumas limitações do trabalho:

- **Restrição a transplantes renais:** O modelo foi desenvolvido e validado exclusivamente para transplantes renais, limitando sua aplicação a outros tipos de transplantes.
- **Dependência de dados históricos:** O uso de dados retrospectivos pode não capturar completamente as mudanças nas práticas de transplante e nas políticas públicas recentes.
- **Generalização limitada:** O modelo precisa ser validado em outras populações e regiões para garantir sua aplicabilidade em contextos diversos.

Perspectivas para Trabalhos Futuros

O trabalho continuará em desenvolvimento, buscando explorar novos horizontes no campo da análise de sobrevivência e predição clínica. Algumas sugestões incluem:

- **Expansão para outros transplantes:** Aplicar as técnicas desenvolvidas para predição de transplantes cardíacos, hepáticos e pulmonares, permitindo maior abrangência do impacto social.
- **Incorporação de dados em tempo real:** Integração de dados longitudinais para capturar mudanças no estado clínico dos pacientes.
- Adoção de técnicas avançadas: Explorar modelos baseados em redes neurais profundas, como o DeepSurv, para melhorar a captura de interações não lineares e variáveis complexas.

Integração com sistemas de saúde

A disponibilização da calculadora preditiva para transplantes renais representa um marco importante deste trabalho. A ferramenta oferecerá uma interface intuitiva para que profissionais de saúde possam estimar as probabilidades de transplante de forma personalizada e precisa, facilitando o processo de tomada de decisão e priorização.

Em síntese, este trabalho representa um avanço significativo no uso de aprendizado de máquina e análise de sobrevivência para resolver desafios críticos no campo da saúde pública. Espera-se que as contribuições apresentadas sirvam como base para estudos futuros, gerando um impacto positivo duradouro tanto no meio científico quanto na sociedade como um todo.

REFERÊNCIAS

A. I. Kyritsis e M. Deriaz, "A Machine Learning Approach to Waiting Time Prediction in Queueing Scenarios", 2019 Second International Conference on Artificial Intelligence for Industries (AI4I), Laguna Hills, CA, EUA, 2019, pp. 17-21, doi: 10.1109/AI4I46381.2019.00013.

Associação Brasileira de Transplante de Órgãos (ABTO). (2023). Registro Brasileiro de Transplantes (RBT). Disponível em: https://site.abto.org.br/conteudo/rbt/.

Bellot, A., van der Schaar, M. (2019). Boosted Deep Survival Machines. *Machine Learning for Healthcare Conference.*

Borges, Jucelina Leandro; Dantas, Maria José Pereira; Silva, Eliel Lucas de Oliveira. FERRAMENTA COMPUTACIONAL EM EXCEL: Uma proposta de metodologia para seleção de artigos científicos relevantes. Open Science Research X, [S.L.], p. 1024-1038, 2023. Editora Científica Digital. http://doi.org/10.37885/230111849.

Boxma, O., David, I., Perry, D., & Stadje, W. (2011). A NEW LOOK AT ORGAN TRANSPLANTATION MODELS AND DOUBLE MATCHING QUEUES. Probability in the Engineering and Informational Sciences, 25, 135 - 155. https://doi.org/10.1017/S0269964810000318.

BRASIL. Ministério da Saúde (MS). Brasil registra o maior número de transplantes de órgãos em dez anos. Brasília, DF, 5 jan. 2024. Disponível em: https://www.gov.br/saude/pt-br/assuntos/noticias/2024/janeiro/brasil-registra-o-maior-número-de-transplantes-de-orgaos-em-dez-anos. Acesso em: 12 jan. 2024.

Brown, T., Elster, E., Stevens, K., Graybill, J., Gillern, S., Phinney, S., Salifu, M., & Jindal, R. (2012). Bayesian Modeling of Pretransplant Variables Accurately Predicts Kidney Graft Survival. American Journal of Nephrology, 36, 561 - 569. https://doi.org/10.1159/000345552.

Coorey, C., Sharma, A., Muller, S., & Yang, J. (2020). Prediction modelling - Part 2 - Using machine learning strategies to improve transplantation outcomes.. Kidney international. https://doi.org/10.1016/j.kint.2020.08.026.

Curtis, C., Liu, C., Bollerman, T., & Pianykh, O. (2017). Machine Learning for Predicting Patient Wait Times and Appointment Delays. Journal of the American College of Radiology: JACR, 15 9, 1310-1316.

Decruyenaere, A., Decruyenaere, P., Peeters, P., Vermassen, F., Dhaene, T., & Couckuyt, I. (2015). Prediction of delayed graft function after kidney transplantation: comparison between logistic regression and machine learning methods. BMC Medical Informatics and Decision Making, 15. https://doi.org/10.1186/s12911-015-0206-y.

Dew, M. A., et al. (2007). Psychosocial predictors of survival and adherence to medical recommendations after heart transplantation. Journal of Consulting and Clinical Psychology, 75(2), 267-276.

Dolatsara, H., Chen, Y., Evans, C., Gupta, A., & Megahed, F. (2020). A two-stage machine learning framework to predict heart transplantation survival probabilities over time with a monotonic probability constraint. Decis. Support Syst., 137, 113363. https://doi.org/10.1016/j.dss.2020.113363.

Drekic, S., Stanford, D., Woolford, D., & McAlister, V. (2015). A model for deceased-donor transplant queue waiting times. Queueing Systems, 79, 87-115. https://doi.org/10.1007/S11134-014-9417-7.

Elalouf, A., Perlman, Y., & Yechiali, U. (2018). A double-ended queueing model for dynamic allocation of live organs based on a best-fit criterion. Applied Mathematical Modelling. https://doi.org/10.1016/J.APM.2018.03.022.

FREITAS, Natália Cristina Cardoso; CHERCHIGLIA, Mariângela Leal; SIMÃO FILHO, Carlos; ÁLVARES-TEODORO, Juliana; ACURCIO, Francisco de Assis; GUERRA JÚNIOR, Augusto Afonso. Dezesseis Anos de Transplante Cardíaco em Coorte Aberta no Brasil: Análise da Sobrevida do Enxerto de Pacientes em Uso de Imunossupressores. Arq. Sutiãs. Cardiol., v. 116, n. 4, p. 744-753, abr. 2021. https://dx.doi.org/10.36660/abc.20200117.

Hijry, H., & Olawoyin, R. (2021). Predicting Patient Waiting Time in the Queue System Using Deep Learning Algorithms in the Emergency Room. International Journal of Industrial Engineering and Operations Management.

https://doi.org/10.1016/j.jacr.2017.08.021.

https://doi.org/10.1038/s41598-020-75473-z.

Ishwaran, H., Kogalur, U. B., Blackstone, E. H., & Lauer, M. S. (2008). Random survival forests. *The Annals of Applied Statistics*, 2(3), 841-860.

Kantidakis, G., et al. (2020). Machine learning versus traditional regression models in predicting survival after liver transplantation: a systematic review and meta-analysis. Transplantation, 104(11), 2297-2306.

Kantidakis, G., Putter, H., Lancia, C., Boer, J., Braat, A., & Fiocco, M. (2020). Survival prediction models since liver transplantation - comparisons between Cox models and machine learning techniques. BMC Medical Research Methodology. https://doi.org/10.1186/s12874-020-01153-1.

Katzman, J. L., et al. (2018). *DeepSurv: personalized treatment recommender system using a Cox proportional hazards deep neural network. BMC Medical Research Methodology*, 18, 24.

Kawakita, S., Beaumont, J., Jucaud, V., & Everly, M. (2020). Personalized prediction of delayed graft function for recipients of deceased donor kidney transplants with machine learning. Scientific Reports, 10. https://doi.org/10.1038/s41598-020-75473-z.

Khademi, A., & Liu, X. (2021). Asymptotically Optimal Allocation Policies for Transplant Queueing Systems. SIAM J. Appl. Math., 81, 1116-1140. https://doi.org/10.1137/19m1280193.

Kyritsis, A., & Deriaz, M. (2019). A Machine Learning Approach to Waiting Time Prediction in Queueing Scenarios. 2019 Second International Conference on Artificial Intelligence for Industries (AI4I), 17-21. https://doi.org/10.1109/AI4I46381.2019.00013.

Lal, T., Krishnamoorthy, A., & Joshua, V. (2019). A Queueing Inventory System with Search and Match - An Organ Transplantation Model. Communications in Computer and Information Science. https://doi.org/10.1007/978-3-030-36625-4_22.

Lee, C., et al. (2018). *DeepHit: A deep learning approach to survival analysis with competing risks.* Proceedings of the AAAI Conference on Artificial Intelligence.

Moncada-Torres, A., Maaren, M., Hendriks, M., Siesling, S., & Geleijnse, G. (2021). Explainable machine learning can outperform Cox regression predictions and provide insights in breast cancer survival. Scientific Reports, 11. https://doi.org/10.1038/s41598-021-86327-7.

Nilsson J, Ohlsson M, Höglund P, Ekmehag B, Koul B, et al. (2015) The International Heart Transplant Survival Algorithm (IHTSA): A New Model to Improve Organ Sharing and Survival. PLOS ONE 10(3): e0118644. https://doi.org/10.1371/journal.pone.0118644.

Ray, A., Buth, K., Sullivan, J., Johnstone, D., & Hirsch, G. (2001). Waiting for Cardiac Surgery: Results of a Risk-Stratified Queuing Process. Circulation: Journal of the American Heart Association, 104, I-92-I-98. https://doi.org/10.1161/HC37T1.094904.

Rizopoulos, D. (2011). Dynamic predictions and prospective accuracy in joint models for longitudinal and time-to-event data. *Biometrics*, 67(3), 819-829

Schaubel, D. E., et al. (2006). Survival benefit of kidney transplantation in the elderly: effect of waiting time on dialysis. *Annals of Surgery*, 243(5), 720-726.

Silva JF, Sapiertein, Ferreira GF, Perosa M, Nga HS, de Andrade LGM. Um modelo de previsão de aprendizado de máquina para o tempo de espera para transplante renal. PLoS ONE. 2021;16(5): e0252069. https://doi.org/10.1371/journal.pone.0252069.

Su, X., & Zenios, S. (2004). Patient Choice in Kidney Allocation: The Role of the Queueing Discipline. Manuf. Serv. Oper. Manag., 6, 280-301. https://doi.org/10.1287/msom.1040.0056.

Wang, P., Li, Y., & Reddy, C. K. (2019). Machine learning for survival analysis: A survey. *ACM Computing Surveys (CSUR)*, 51(6), 110.

Wolfe, R. A., et al. (1999). Comparison of mortality risk in wait-listed patients receiving a first cadaveric transplant or remaining on a waiting list. *Transplantation*, 67(4), 414-418.

Xu, C., Alaa, A., Bica, I., Ershoff, B., Cannesson, M., & Schaar, M. (2021). Learning Matching Representations for Individualized Organ Transplantation Allocation. ArXiv, abs/2101.11769.

Yael Perlman, Amir Elalouf, Uri Yechiali, Dynamic allocation of stochastically-arriving flexible resources to random streams of objects with application to kidney cross-transplantation, European Journal of Operational Research, Volume 265, Issue 1, 2018, Pages 169-177, ISSN 0377-2217, https://doi.org/10.1016/j.ejor.2017.07.068.

Yang, H., Tian, J., Meng, B., Wang, K., Zheng, C., Liu, Y., Yan, J., Han, Q., & Zhang, Y. (2021). Application of Extreme Learning Machine in the Survival Analysis of Chronic Heart Failure Patients With High Percentage of Censored Survival Time. Frontiers in Cardiovascular Medicine, 8. https://doi.org/10.3389/fcvm.2021.726516.

Yanto Sandy Tjang, Geert J.M.G. van der Heijden, Gero Tenderich, Diederick E. Grobbee, Reiner Körfer, Survival analysis in heart transplantation: results from an analysis of 1290 cases in a single center, European Journal of Cardio-Thoracic Surgery, Volume 33, Issue 5, May 2008, Pages 856–861, https://doi.org/10.1016/j.ejcts.2008.02.014.

Zahiri, B., Tavakkoli-Moghaddam, R., Mohammadi, M., & Jula, P. (2014). Multi-objective design of an organ transplant network under uncertainty. Transportation Research Part E-logistics and Transportation Review, 72, 101-124. https://doi.org/10.1016/J.TRE.2014.09.007.

Zenios, S.A. Modelando a lista de espera para transplante: Um modelo de fila com renegação. Sistemas de Filas 31, 239–251 (1999). https://doi.org/10.1023/A:1019162331525.

JIGSO. The notion of censoring in survival analysis. Disponível em: https://jigso.com/the-notion-of-censoring-in-survival-analysis/. Acesso em: 15 jan. 2025.

DATA SCIENCE DOJO. A visual representation of the bootstrap sampling scheme. Disponível em: https://datasciencedojo.com/blog/bootstrap-sampling/. Acesso em: 15 jan. 2025.

ANALYTICS VIDHYA. Decision Tree vs Random Forest Algorithm. Disponível em: https://www.analyticsvidhya.com/blog/2020/05/decision-tree-vs-random-forest-algorithm/. Acesso em: 15 jan. 2025.

PONTIFICIA UNIVERSIDADE CATOLICA DE GOIÁS
GABINETE DO REITOR

AV. Universitaria, 1069 ● Setor Universitário
Caixa Postal 86 ● CEP: 74.605-010
Goiânia ● Goiás ● Brasil
Fone: (62) 3946-1000
www.pucgoias.edu.br ● reitoria@pucgoias.edu.br

RESOLUÇÃO n° 038/2020 – CEPE

ANEXO

APÊNCIDE ao TCC

Termo de autorização de publicação de produção acadêmica

O(a) estudante **Eliseu Rodrigues Guimarães** do Curso de **Ciência da Computação**, matricula **2016.2.0028.0189-8**, telefone **(62) 98119-9183** e-mail **eliseu.rguimaraes@gmail.com** na qualidade de titular dos direitos autorais, em consonância com a Lei n° 9.610/98 (Lei dos Direitos do Autor), autoriza a Pontifícia Universidade Católica de Goiás (PUC Goiás) a disponibilizar o Trabalho de Conclusão de Curso intitulado **Análise de Sobrevivência e Aprendizado de Máquina na Previsão do Tempo de Espera para Transplante Renal: Uma Abordagem Otimizada**, gratuitamente, sem ressarcimento dos direitos autorais, por 5 (cinco) anos, conforme permissões do cocumento, em meio eletrônico, na rede mundial de computadores, no formato especificado (Texto(PDF); Imagem (GIF ou JPEG); Som (WAVE, MPEG, AIFF, SND); Vídeo (MPEG, MWV, AVI, QT); outros, específicos da produção científica gerada nos cursos de graduação da PUC Goiás.

Goiânia, 23 de Dezembro de 2024.

Assinatura do autor: _____

Nome complete do autor: Eliseu Rodrigues Guimarães

Assinatura do professor(a) orientador(a): _____

Nome complete do professor(a) orientador: Maria José Pereira Dantas