

# **GUILHERME FONSECA FERRARI**

Validação de um modelo de Machine Learning para a previsão de Length of Stay em Unidades de Tratamento Intensivo para Pneumonia Adquirida na Comunidade

PROJETO DE GRADUAÇÃO EM ENGENHARIA DE PRODUÇÃO  
APRESENTADO AO DEPARTAMENTO DE ENGENHARIA INDUSTRIAL  
DA PUC-RIO, COMO PARTE DOS REQUISITOS PARA OBTENÇÃO  
DO TÍTULO DE ENGENHEIRO DE PRODUÇÃO

Orientador: Prof. Igor Tona Peres  
Coorientador: Prof. Leonardo dos Santos Lourenço Bastos

Departamento de Engenharia Industrial  
Rio de Janeiro, 12 de junho de 2024.

## **Resumo**

A métrica do tempo de permanência (LoS) é uma das mais utilizadas para avaliar o uso de recursos em Unidades de Terapia Intensiva (UTI). Este trabalho propõe a atualização e validação de uma metodologia para análise de eficiência de UTIs em pacientes de Pneumonia Adquirida na Comunidade (CAP) utilizando uma abordagem baseada em aprendizado de máquina. Foram utilizados dados de três anos (2019, 2022 e 2023) para pacientes adultos consecutivos de UTIs médico-cirúrgicas não especializadas em hospitais de uma rede integrada de hospitais no Brasil, que inclui 220 UTIs de 57 hospitais. O modelo de previsão de LoS foi retreinado com um total de 163.503 admissões em 2022. A validação do indicador de eficiência SLOSR foi realizada com 16.985 admissões de teste (2023) no grupo de Pneumonia Adquirida na Comunidade (CAP) [ $R^2=0,89$ ; SLOSR = 1,13 (0,91-1,34)], mostrando uma performance calibrada e similar ao que havia sido notado no modelo para pacientes gerais. Portanto, o modelo proposto foi usado para atualizar o indicador SLOSR de eficiência de UTIs que foi validado para aplicação em um subgrupo específico de Pneumonia Comunitária no período pós-COVID-19.

## **Palavras-chave**

Ciência de Dados; Aprendizado de Máquina; Modelos Preditivos;

Tempo de Permanência; Unidades de Terapia Intensiva; Serviços de Saúde

## **Abstract**

The length of stay (LoS) metric is one of the most widely used to assess the use of resources in Intensive Care Units (ICU). This work proposes the updating and validation of a methodology for analyzing the efficiency of ICUs in patients with Community-Acquired Pneumonia (CAP) using a machine learning-based approach. Data from three years (2019, 2022, and 2023) were used for consecutive adult patients from non-specialized medical-surgical ICUs in hospitals of an integrated network of hospitals in Brazil, which includes 220 ICUs from 57 hospitals. The LoS prediction model was retrained with a total of 163,503 admissions in 2022. The validation of the SLOS efficiency indicator was performed with 16,985 test admissions (2023) in the Community-Acquired Pneumonia (CAP) group [ $R^2=0.89$ ; SLOS = 1.13 (0.91-1.34)], showing a calibrated performance similar to what had been observed in the model for general patients. Therefore, the proposed model was used to update the SLOS indicator of ICU efficiency that was validated for application in a specific subgroup of Community-acquired Pneumonia in the post-COVID-19 period.

## **Keywords**

Data Science, Machine Learning, Predictive Models, Length of Stay, Intensive Care Units, Health Services

## Sumário

1. Introdução.....	7
2. Referencial Teórico .....	9
2.1. Previsão do tempo de permanência em UTIs .....	9
2.2. Modelos de previsão para tempo de permanência na UTI .....	<b>Erro! Indicador não definido.</b>
2.3. Geração de um indicador de eficiência para UTIs .....	9
2.4. Pneumonia Adquirida na Comunidade (CAP) .....	11
3. Materiais e métodos.....	12
3.1. Design do Estudo.....	12
3.2. Participantes e Coleta de Dados .....	12
3.3. Desfecho de interesse .....	13
3.4. Análise estatística .....	13
4. Resultados.....	14
4.1. Atualização do modelo .....	18
4.3. Validação em pacientes com CAP.....	19
5. Conclusão .....	23
6. Referências Bibliográficas.....	24

## **Lista de Figuras**

Figura 1 - Exemplo de funnel plot para o SLOS

Figura 2 - Fluxograma da seleção dos dados para a base de 2022 de atualização do modelo

Figura 3 - Fluxograma da seleção dos dados para a base de 2023 para o teste da atualização do modelo

Figura 4 - Fluxograma da seleção dos dados para o conjunto de validação do modelo

Figura 5 - Calibração do modelo de Stacking atualizado, treinamento em 2022 e teste em 2023

Figura 6 - Calibração do modelo de Stacking proposto por Peres et al. [2022]

Figura 7 - Gráfico de calibração do modelo atualizado em 2022 e testado em pacientes de 2023 com CAP

Figura 8 - Calibração do modelo com os pacientes de 2023 com CAP

Figura 9 - Funnel plot da verificação em pacientes de CAP em 2023

## **Lista de Tabelas**

Tabela 1 - Análise descritiva das bases de dados após o tratamento para 2022 e 2023

Tabela 2 - Análise descritiva da base de dados após o tratamento para os pacientes com CAP em 2023

## 1. Introdução

Existe uma pressão contínua sobre os hospitais pela melhora de eficiência e redução de custos. As Unidades de Terapia Intensiva (UTIs) são ambientes complexos que fornecem cuidados caros [Halpern e Pastores, 2015; Halpern et al., 2004]. Diversos estudos apontam que os custos hospitalares estão fortemente relacionados ao tempo de permanência dos pacientes nas UTIs [Arabi et al., 2002; Halpern et al., 2004; Higgins et al., 2003; Kahn et al., 2008; Laupland et al., 2006; Schoffelen et al., 2010; Zampieri et al., 2014; Zimmerman et al., 2006]. Dessa forma, a previsão de tempo de permanência (LoS) dos pacientes nas UTIs permite um melhor planejamento de recursos necessários.

Para atingir os objetivos de eficiência em terapia intensiva, foi proposta uma metodologia baseada em dados para prever o LoS individual do paciente em UTI [Peres et al., 2022]. Atualmente, a análise de eficiência de UTIs é realizada por meio do SRU (*Standardized Resource Use*), um indicador proposto por Rothen et al. [2007], que realiza a previsão de permanência por meio de um score de severidade (SAPS3). Assim, em um artigo publicado no final do ano de 2023 com participação do autor [Peres et al., 2023], propusemos o indicador SLOS (*Standardized Length of Stay Ratio*) para a análise de eficiência de UTIs por meio da utilização do modelo de Peres et al. [2022], que mostrou ser mais calibrado do que o indicador SRU de Rothen et al. [2007] para a base de dados estudada. Dessa forma, parte deste trabalho de fim de curso já resultou na publicação de um artigo por parte do autor [Peres et al., 2023].

Este trabalho complementa estudos anteriores realizados. Inicialmente foi proposta uma metodologia baseada em dados para prever o LoS individual do paciente em UTI [Peres et al., 2022]. Em seguida foi proposta por Peres et al. [2023] uma medida de eficiência para UTIs.

O objetivo principal deste estudo foi a atualização e validação da metodologia para análise de eficiência de UTIs proposta para aplicação em um subgrupo de pacientes de Pneumonia Adquirida na Comunidade (*Community-acquired pneumonia - CAP*). O subgrupo de pacientes com CAP foi escolhido por representar o diagnóstico com o maior número de pacientes, sendo superior a 11% do total na base utilizada. Deseja-se responder as seguintes perguntas de pesquisa: “O modelo proposto por Peres et al. [2022] mantém sua capacidade preditiva ao longo do tempo?”; e “O indicador de eficiência se mantém calibrado para um subgrupo de pacientes com Pneumonia Adquirida na Comunidade?”.

Este trabalho foi subdividido da seguinte forma: o capítulo 2 apresenta referencial teórico sobre o modelo preditivo utilizado, o indicador de análise de eficiência SLOS<sub>R</sub> e o conceito de CAP; o capítulo 3 apresenta o desenho do estudo, a apresentação da base de dados utilizada e a metodologia aplicada no estudo; o capítulo 4 apresenta os resultados da aplicação da metodologia proposta em uma grande base de dados de pacientes de UTI com CAP; o capítulo 5 traz uma discussão sobre os resultados; e o capítulo 6 apresenta uma conclusão para o trabalho.

## 2. Referencial Teórico

### 2.1. Previsão do tempo de permanência em UTIs

Peres et al. [2021] realizaram uma revisão da literatura sobre modelos que poderiam ser aplicados para previsão do tempo de permanência em UTIs. Os autores notaram que a maioria dos estudos da literatura utilizava Regressão Linear, sem testar outras abordagens mais recentes como algoritmos de aprendizado de máquina. Os autores verificam que os estudos que incluíram modelos de aprendizado (*Support Vector Regression, Gradient Boosting Machine, and Random Forests*) reportaram melhores resultados e concluem sobre a importância de testar diferentes abordagens nos estudos de desenvolvimento de modelos preditivos.

Peres et al. [2022] propôs um modelo de aprendizado de máquina para o tempo de permanência em UTIs usando uma grande base de dados brasileira. A abordagem foi baseada em dados e incluiu o treinamento e validação de diversos modelos de machine learning. O modelo final proposto por Peres et al. [2022] resultou de um empilhamento (*stacking*) de modelos. Para tal, foi utilizado um algoritmo de meta-aprendizado para aprender a melhor maneira de combinar as previsões dos dois algoritmos de aprendizado de máquina. O funcionamento dos modelos utilizados está exposto no Apêndice I.

### 2.2. Geração de um indicador de eficiência para UTIs

A duração da estadia na UTI pode ser um substituto para medir custo e eficiência, refletindo vários aspectos do atendimento, desde políticas de admissão e alta até a aderência às melhores práticas e segurança do paciente. Uma das medidas mais comuns para avaliar a eficiência das UTIs é o *Standardized Resource Use (SRU)* proposto por Rothen et al. [2007]. Esta medida usa a duração da estadia na UTI para estimar a quantidade média de recursos usados por paciente sobrevivente em uma unidade específica. Para estimar a duração da estadia na UTI foram utilizados os decis do sistema de pontuação de gravidade SAPS3. Para chegar nesse resultado, os dados agregados dos pacientes foram estratificados de acordo com a pontuação de admissão SAPS3 e, para cada estrato, a soma de duração da estadia na UTI de todos os pacientes naquele estrato foi calculada e dividida pela quantidade de pacientes no extrato. Assim, a duração de estadia esperada dependia do SAPS3 de admissão do paciente, estando contido em um estrato que resultaria no LoS esperado.

Para medir a eficiência de UTIs, foi elaborada uma métrica alternativa baseada no *Standardized Length of Stay Ratio* (SLOS<sub>R</sub>) sugerido por Verburg et al. [2018]. A métrica proposta analisa o uso de recursos e a eficiência das unidades por meio do tempo de permanência (LoS) dos pacientes. O SLOS<sub>R</sub> tem uma formulação semelhante à SRU, entretanto, em vez de usar o SAPS3 para prever o LoS esperado, os autores utilizam um modelo de previsão usando regressão de mínimos quadrados ordinários com uma função de link logarítmico, resultando em uma métrica que tende a melhorar a correção do case-mix.

$$SLOS_{R_u} = \frac{\sum_{i=1}^n LoS_{Observado_{iu}}}{\sum_{i=1}^n LoS_{Esperado_{iu}}} \quad (1)$$

Uma das formas de avaliação do indicador de eficiência proposto é a partir de *Funnel Plots* (Figura 2), em que no eixo “y” temos o indicador SLOS<sub>R</sub> e no eixo “x” o número de admissões, sendo as UTIs representadas pelos pontos amarelos. Os limites de controle, ilustrados em linhas tracejadas, indicam uma faixa na qual se esperaria que os valores da medida de qualidade caíssem. Os limites de controle formam uma forma de “funil” ao redor do padrão de referência, apresentado como uma linha horizontal. Também é importante notar que o SLOS<sub>R</sub> demonstra a calibração geral da métrica proposta [Peres et al., 2022]. Dessa forma, UTIs eficientes têm SLOS<sub>R</sub> menor que um (total observado < total esperado de duração da estadia), enquanto UTIs ineficientes apresentam SLOS<sub>R</sub> maior que um (total observado > total esperado de duração da estadia).

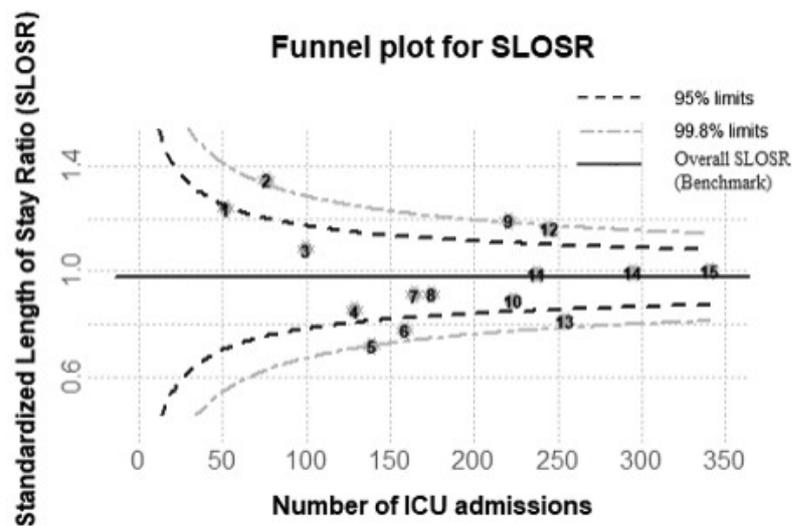


Figura. 1. Exemplo de funnel plot para o SLOS<sub>R</sub>

Peres et al. [2023] propõem um novo indicador SLOS<sub>R</sub> que substitui o uso da previsão usando a regressão linear de Verburg et al. [2018] pela abordagem *data-driven* desenvolvida por Peres et al. [2022] usando modelos de aprendizado. A métrica proposta de SLOS<sub>R</sub> para avaliar a eficiência da UTI foi comparada com o modelo tradicional SRU no nível da UTI por meio do coeficiente de determinação ( $R^2$ ), gráficos de calibração e *Funnel Plots*. Foi utilizado um subconjunto do conjunto de dados original relatado por Peres et al. [2022] que foi separado para validar o modelo.

### 2.3. Pneumonia Adquirida na Comunidade (CAP)

Pneumonia adquirida na comunidade (CAP) é qualquer pneumonia adquirida fora do hospital, podendo ser causada por agentes bacterianos ou agentes virais. Um paciente de CAP pode ser diagnosticado como severo ou não dependendo da presença dos seguintes critérios [Metlay et Al, 2019]:

Critérios maiores:

- Choque séptico com necessidade de vasopressores;
- Insuficiência respiratória que requeira ventilação mecânica;

Critérios menores:

- Taxa respiratória  $\geq 30$  respirações/min;
- Relação  $PaO_2/FiO_2 \leq 250$ ;
- Infiltrações multilobares;
- Confusão/desorientação;
- Uremia (nível de ureia no sangue  $\geq 20$  mg/dl);
- Leucopenia (contagem de leucócitos  $< 4.000$  células/ $\mu$ l);
- Trombocitopenia (contagem de plaquetas  $< 100.000$ / $\mu$ l);
- Hipotermia (temperatura central  $< 36^\circ\text{C}$ );
- Hipotensão que requeira ressuscitação agressiva com fluidos.

Pacientes com um dos critérios maiores ou pelo menos três dos critérios menores são classificados como pacientes com Pneumonia Adquirida na Comunidade Severa.

### **3. Materiais e métodos**

#### **3.1. Design do Estudo**

Para a atualização do modelo e validação em um subgrupo de pacientes com Pneumonia Adquirida na Comunidade (CAP), foi utilizada uma base atualizada dos dados originais coletados e utilizados pelos estudos de Peres et al [2022, 2023]. O banco de dados inclui dados de três anos (2019, 2022 e 2023) para pacientes adultos consecutivos de UTIs médico-cirúrgicas não especializadas em hospitais de uma rede integrada de hospitais no Brasil. Para evitar viés no modelo, não foram incluídos dados dos dois anos de pico da pandemia de COVID-19 (2020 e 2021). A base inclui dados de 109 UTIs em 38 hospitais (em 2019), 206 UTIs em 55 hospitais (em 2022) e 220 UTIs em 57 hospitais (em 2023). O Comitê de Ética Local e o Comitê Nacional de Ética do Brasil (CAAE: 17079119.7.0000.5249) aprovaram o estudo sem a necessidade de consentimento informado.

#### **3.2. Participantes e Coleta de Dados**

As informações anonimizadas foram extraídas de um sistema eletrônico utilizado pela rede hospitalar para melhoria da qualidade e fins de benchmarking para UTIs (Epimed Monitor®, Rio de Janeiro, Brasil) [Zampiere et al., 2017]. Este banco de dados contém dados estruturados coletados prospectivamente de todas as admissões nas UTIs participantes, com uma caracterização clínica abrangente dos pacientes, incluindo demografia, diagnóstico de admissão na UTI, comorbidades, pontuações de gravidade da doença, uso de suporte a órgãos e resultados da UTI e do hospital (mortalidade e tempo de permanência). Incluímos as admissões de pacientes adultos com qualquer diagnóstico. Para o ano de 2023 (escolhido como período de validação do modelo para CAP), aplicamos também o filtro de diagnóstico primário ou secundário de CAP nas primeiras 48h. Excluimos readmissões hospitalares em menos de 30 dias, admissões com tempo de permanência na UTI (LoS) inferior a 6 horas e que apresentaram tempo de permanência no hospital superior a 60 dias antes da admissão na UTI.

Os valores faltando para os dados fisiológicos (HighestHeartRate1h, HighestRespiratoryRate1h e LowestGlasgowComaScale1h) e para os dados laboratoriais (HighestCreatinine1h, HighestTemperature1h e HighestLeukocyteCount1h) foram imputados usando a mediana da amostra. Para o tratamento dos dados foram utilizados os pacotes caret [v 6.0.94; Kuhn, 2008] e tidyverse [v2.0.0; Wickham et al., 2021)] para a linguagem de

programação R [v4.1.3; R Core Team, 2021]. Para a geração dos *Funnel Plots* foi utilizado o pacote *ems* [1.3.11; Lunna, 2021].

### 3.3. Desfecho de interesse

Os desfechos de interesse são o tempo de permanência na UTI e a métrica de eficiência da UTI no uso de recursos. O tempo de permanência é calculado por meio da metodologia de dados proposta por Peres et Al. [2022] e a métrica de eficiência é o SLOS<sub>R</sub> proposta que utiliza o LoS na UTI como um substituto para medir o uso de recursos.

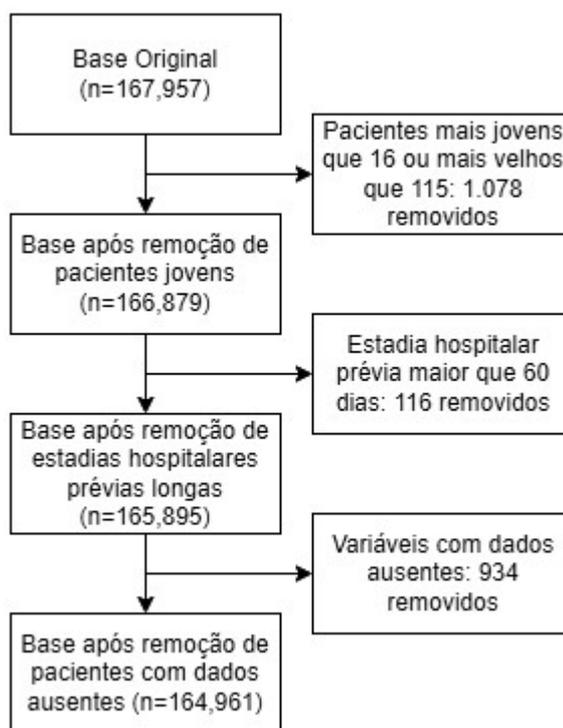
### 3.4. Metodologia

A metodologia proposta por Peres et al. [2022] para o desenvolvimento de um modelo de dados para a previsão do tempo de permanência consistiu em um pré-processamento do conjunto de dados usando os dados das primeiras 24 horas e a seleção do melhor conjunto de características. O tempo de permanência dos pacientes foi truncado em 21 dias e em seguida, foi aplicada a metodologia para estimar o tempo de permanência na UTI (no momento da admissão). Foi realizada uma atualização do modelo proposto utilizando as bases de 2019 e 2022 como treinamento e a base 2023 como teste. O modelo foi treinado em três cenários: somente os pacientes da base de 2019, somente os pacientes da base de 2022 e o conjunto dos pacientes das duas bases (2019 e 2022), todos os cenários foram testados na base de 2023 para verificação do melhor modelo. A verificação foi realizada pela comparação das métricas individuais (Raiz do Erro Quadrático Médio [RMSE]; Erro Absoluto Médio [MAE]; Coeficiente de Determinação [ $R^2$ ]) com ênfase no coeficiente de determinação e na calibração do modelo. O modelo treinado somente com pacientes em 2022 se mostrou o melhor modelo nos testes realizados e foi utilizado na atualização do modelo.

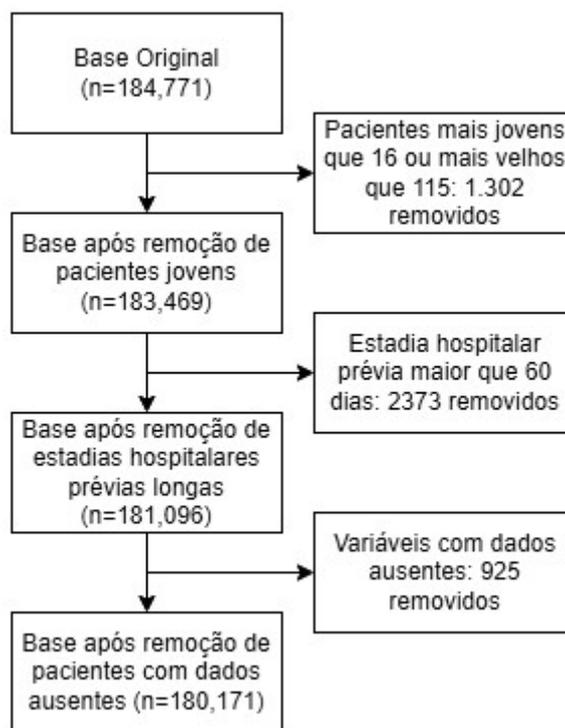
O diagnóstico de Pneumonia Adquirida na Comunidade (CAP) foi escolhido para validação do modelo de Peres et al. [2023] por subgrupo já que é o diagnóstico individual mais representativo na amostra utilizada. Para a validação do indicador no subgrupo de pacientes com CAP foi utilizado o gráfico de calibração por unidade, o *Funnel Plot* e o coeficiente de determinação no nível da unidade (UTIs).

## 4. Resultados

Foram extraídos dois conjuntos de dados para treinamento e teste do modelo. O conjunto de treinamento contém um total de 167.957 admissões independentes de primeiro de janeiro a 31 de dezembro de 2022 e o conjunto de teste contém um total de 184.771 admissões independentes de primeiro de janeiro a 31 de dezembro de 2023. Após a aplicação dos critérios, o conjunto de dados de 2022 usado para treino final ficou com 164.961 admissões (Figura 2). O conjunto de teste com dados de 2023 ficou com 180.171 admissões (para ser usado no escopo de atualização do modelo como grupo de teste). Em seguida, este conjunto de 2023 passou também pelo filtro de diagnóstico de CAP nas primeiras 48 e ficou com 16.985 admissões (para ser usado como validação do modelo SLOS por subgrupo). Foi utilizado também o conjunto de dados de Peres et al. [2022], que incluíram um total de 99.492 admissões de janeiro a dezembro de 2019.

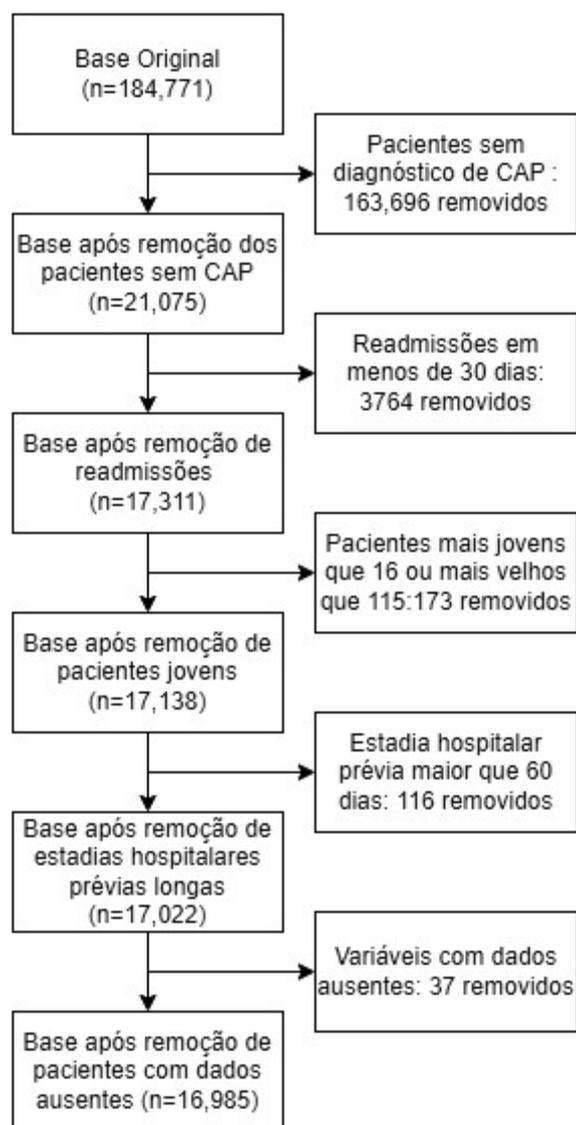


**Figura 2.** Fluxograma da seleção dos dados para a base de 2022 de atualização do modelo [Elaboração própria]



**Figura 3.** Fluxograma da seleção dos dados para a base de 2023 para o teste da atualização do modelo  
[Elaboração própria]

Os pacientes da base utilizada para treino possuíam em média 64 anos, eram majoritariamente mulheres (53%), 67% foram admitidos da emergência. A estadia mediana foi de três dias na UTI e seis dias no hospital. Os pacientes possuíam um SAPS3 médio de 42 e 53% eram hipertensos. 15,3% dos pacientes necessitaram do uso de ventilação (9% de ventilação não invasiva e 6,3% com ventilação invasiva) e 8% utilizaram vasopressores.



**Figura 4.** Fluxograma da seleção dos dados para o conjunto de validação do modelo [Elaboração própria]

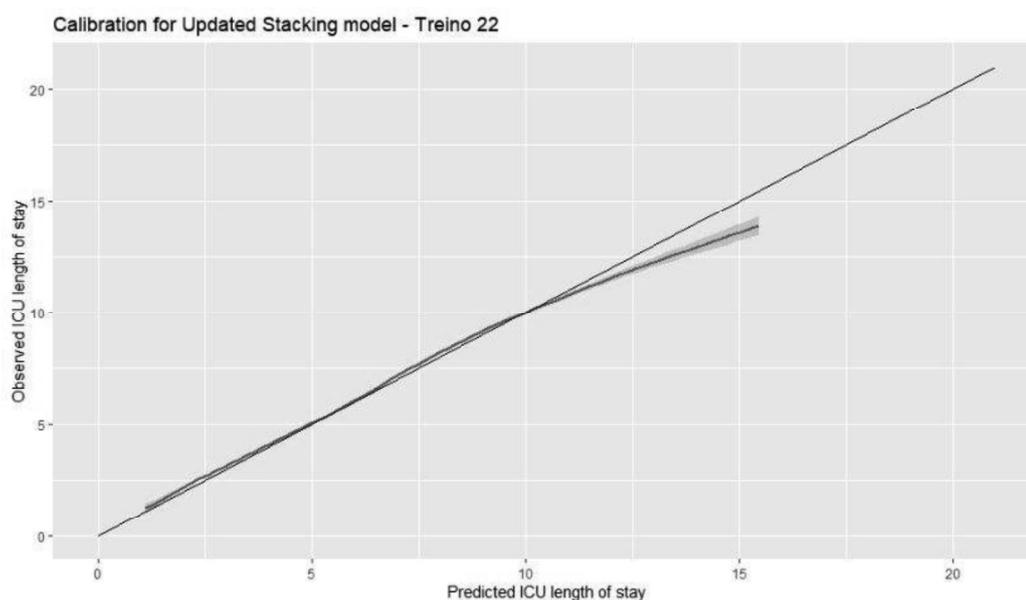
**Tabela 1.** Análise descritiva das bases de dados após o tratamento para 2022 e 2023

Fator	2022	2023
	N = 163,503	N = 180,171
Idade	64 (44, 79)	62 (43, 78)
Gênero		
F	86,628 (53%)	96,249 (53%)
M	76,875 (47%)	83,922 (47%)
Score MFI	0.09 (0.00, 0.18)	0.09 (0.00, 0.18)
Pacientes com MFI de fragilidade	17,765 (11%)	18,449 (10%)
Índice de comorbidade de Charlson	1 (0, 2)	1 (0, 2)
Origem da Admissão		
Sala de intervenção cardiovascular	6,433 (3.9%)	5,958 (3.3%)
Emergência	110,209 (67%)	126,785 (70%)
Sala de operação	17,946 (11%)	17,554 (9.7%)
Outro	5,033 (3.1%)	5,666 (3.1%)
Outra unidade no hospital	14,300 (8.7%)	14,820 (8.2%)
Ala/Andar	9,582 (5.9%)	9,388 (5.2%)
Pontuação SAPS 3	42 (34, 51)	41 (34, 50)
Pontuação Sofa	0 (0, 2)	0 (0, 2)
Hipertensão	86,414 (53%)	90,777 (50%)
Diabetes	45,967 (28%)	49,651 (28%)
Obesidade	6,128 (3.7%)	5,947 (3.3%)
Imunossupressão	30,093 (18%)	32,976 (18%)
Doença cardiovascular	40,390 (25%)	42,910 (24%)
Asma	12,894 (7.9%)	14,078 (7.8%)
Tumor maligno	23,443 (14%)	25,104 (14%)
Doença Cerebral	20,076 (12%)	21,516 (12%)
Insuficiência renal crônica	12,349 (7.6%)	12,707 (7.1%)
Fumante	7,912 (4.8%)	9,262 (5.1%)
Cirrose	2,030 (1.2%)	1,960 (1.1%)
Outras comorbidade	56,639 (35%)	63,243 (35%)
Escala de coma de Glasgow na primeira hora	15.00 (15.00, 15.00)	15.00 (15.00, 15.00)
Número de plaquetas na primeira hora	223 (175, 279)	223 (175, 277)
Ureia	37 (28, 50)	37 (28, 49)
BUN	17 (13, 25)	17 (13, 24)
Creatinina	0.90 (0.70, 1.10)	0.90 (0.70, 1.10)
Uso de ventilação não invasiva	14,590 (9.0%)	15,230 (8.5%)
Uso de ventilação mecânica	10,183 (6.3%)	8,324 (4.6%)
Oxigenação por Membrana Extra Corporal	67 (<0.1%)	44 (<0.1%)
Uso de vasopressores	12,874 (8.0%)	11,551 (6.4%)
Resultado hospitalar		
Alta	151,989 (93%)	167,901 (94%)
Morte	11,387 (7.0%)	10,105 (5.7%)

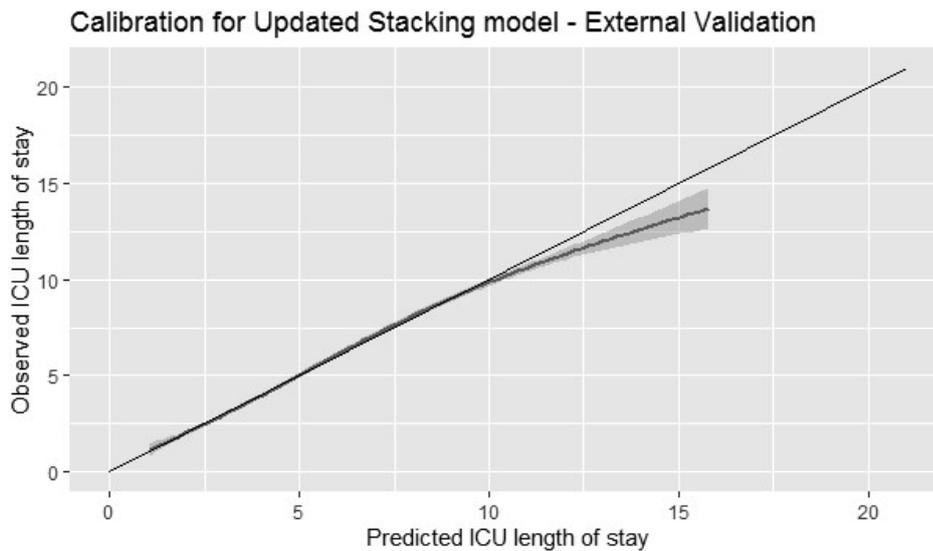
Resultado na UTI			
Alta		156,868 (96%)	174,161 (97%)
Morte		6,635 (4.1%)	6,010 (3.3%)
Tempo de permanência na UTI		2.9 (1.7, 5.3)	2.9 (1.7, 5.1)
Tempo de permanência no hospital		6 (3, 12)	5 (3, 11)

#### 4.1. Atualização do modelo

O modelo de *stacking* adotado foi retreinado na base tratada de 2022 com 163,503 pacientes e testado na base tratada de 2023 com 180,171 pacientes. A aplicação do modelo atualizado com dados de 2022 no conjunto de dados de validação de 2023 resultou em métricas (RMSE = 4.02; MAE = 2.66;  $R^2 = 0.17$ ; Figura 5). Comparado ao modelo original de Peres et al. [2022] (RMSE = 4.47; MAE = 3.05;  $R^2 = 0.25$ ; Figura 6), houve uma ligeira melhora nos erros com uma pequena piora no coeficiente de determinação. Os valores de erros e a calibração do modelo atualizado em 2022 e testado em 2023 está em linha com a validação original de Peres et al. [2022].



**Figura 5.** Calibração do modelo de Stacking atualizado, treinamento em 2022 e teste em 2023



**Figura 6.** Calibração do modelo de Stacking proposto por Peres et Al. [2022] nos dados de teste de 2023

### 4.3. Validação em pacientes com CAP

Os pacientes da base utilizada para o teste possuíam em média 70 anos, eram majoritariamente mulheres (54%), e 87% foram admitidos da emergência. A estadia mediana foi de quatro dias na UTI e sete dias no hospital. Os pacientes possuíam um SAPS3 médio de 50, e 51% eram hipertensos. 28,1% dos pacientes necessitaram do uso de ventilação (21% de ventilação não invasiva e 7,1% com ventilação invasiva) e 8,5% utilizaram vasopressores.

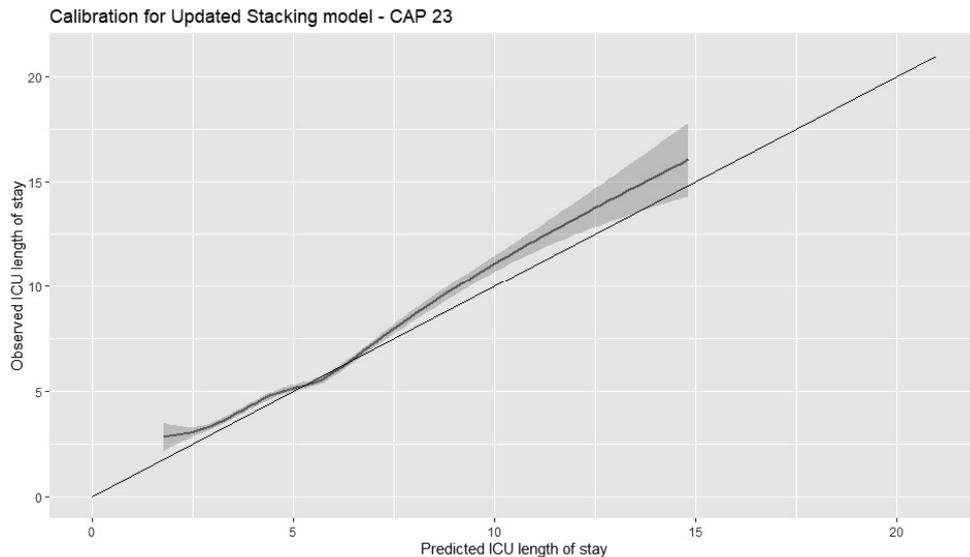
**Tabela 2.** Análise descritiva da base de dados após o tratamento para os pacientes com CAP em 2023

Fator	CAP - 2023
	N = 16,985
Idade	70 (46, 84)
Gênero	
F	9,158 (54%)
M	7,827 (46%)
Score MFI	0.18 (0.00, 0.27)
Pacientes com MFI de fragilidade	2,261 (13%)
Índice de comorbidade de Charlson	1 (0, 2)
Origem da Admissão	
Sala de intervenção cardiovascular	33 (0.2%)
Emergência	14,709 (87%)
Sala de operação	103 (0.6%)
Outro	683 (4.0%)

Outra unidade no hospital	896 (5.3%)
Ala/Andar	561 (3.3%)
Pontuação SAPS 3	50 (42, 57)
Pontuação Sofa	1 (0, 2)
Hipertensão	8,664 (51%)
Diabetes	4,852 (29%)
Obesidade	489 (2.9%)
Imunossupressão	3,723 (22%)
Doença cardiovascular	3,856 (23%)
Asma	2,902 (17%)
Tumor maligno	2,422 (14%)
Doença Cerebral	3,091 (18%)
Insuficiência renal crônica	1,118 (6.6%)
Fumante	1,077 (6.3%)
Cirrose	122 (0.7%)
Outras comorbidade	5,684 (33%)
Escala de coma de Glasgow na primeira hora	15.00 (15.00, 15.00)
número de plaquetas na primeira hora	223 (172, 282)
Ureia	37 (29, 54)
BUN	18 (13, 26)
Creatinina	0.90 (0.70, 1.14)
Uso de ventilação não invasiva	3,647 (21%)
Uso de ventilação mecânica	1,203 (7.1%)
Oxigenação por Membrana Extra Corporal	3 (<0.1%)
Uso de vasopressores	1,437 (8.5%)
Resultado hospitalar	
Alta	15,184 (91%)
Morte	1,565 (9.3%)
Resultado na UTI	
Alta	15,903 (94%)
Morte	1,082 (6.4%)
Tempo de permanência na UTI	3.9 (2.4, 6.9)
Tempo de permanência no hospital	7 (4, 12)

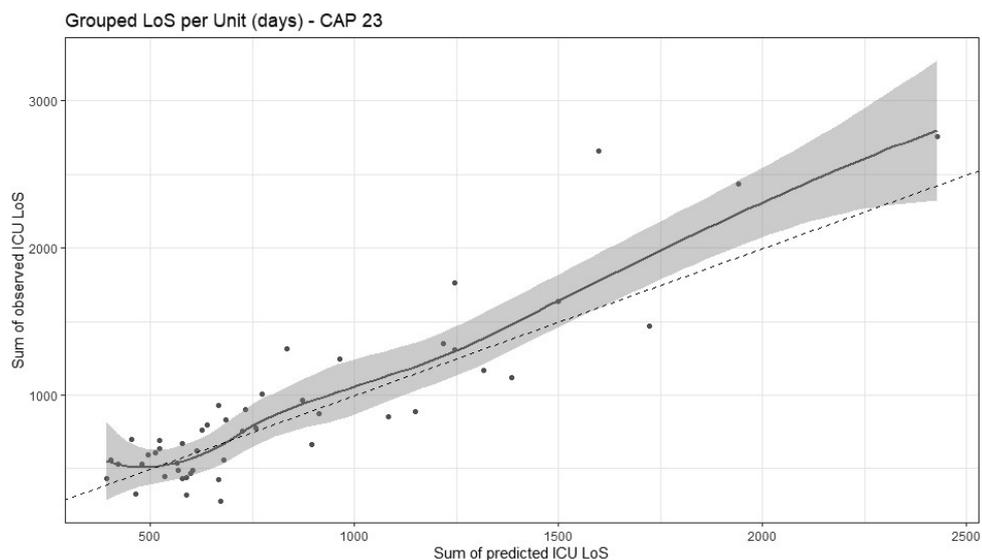
O modelo de *stacking* treinado na base tratada de 2022 com 163,503 pacientes foi testado na base tratada de 2023 com 16,985 pacientes com Pneumonia Adquirida na Comunidade. A aplicação do modelo atualizado no conjunto de dados de CAP (RMSE = 4.57;

MAE = 3,10;  $R^2 = 0.15$ ) resultou em métricas de erro comparáveis e com uma calibração similar à anteriormente vista nos dados completos de 2023 (Figura 7).

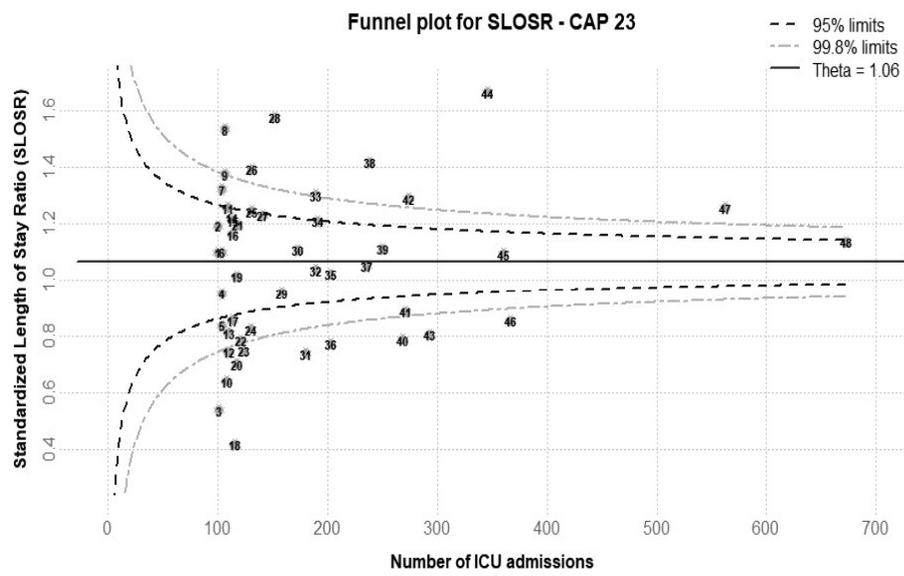


**Figura 7.** Gráfico de calibração do modelo atualizado em 2022 e testado em pacientes de 2023 com CAP

Ao aplicar o indicador de eficiência de UTIs, SLOSR, aos dados de CAP para o ano de 2023 (usando o modelo de LoS atualizado), é possível verificar que a calibração do modelo SLOSR [ $R^2=0,89$ ; SLORS = 1,13 (0,91-1,34); Figuras 8 e 9] foi ligeiramente pior que o resultado reportado por Peres et al. [2023] [ $R^2=0,90$ ; SLORS = 1,00 (0,87-1,10)]. Entretanto, os resultados mostram que, apesar da ligeira perda de calibração da análise de eficiência ao filtrar o subgrupo específico de CAP, o modelo SLOSR continua mantendo uma performance calibrada.



**Figura 8.** Calibração do modelo com os pacientes de 2023 com CAP



**Figura 9.** Funnel plot da verificação em pacientes de CAP em 2023

## 5. Conclusão

A previsão de LoS de pacientes de UTIs é uma forma importante de melhora do planejamento de recursos necessários em UTIs, por serem ambientes complexos e de cuidados caros, permitindo uma melhora de eficiência e redução de custos. O presente estudo descreveu a atualização de um modelo de predição da duração da estadia individual na UTI por meio de uma metodologia estruturada e orientada por dados usando algoritmos de aprendizado de máquina.

Foram realizados testes com uma base de 2019 a 2022 para o treinamento do modelo, com a utilização exclusiva de pacientes de 2022 para treinamento apresentando o melhor desempenho. O modelo atualizado com dados de 2022 demonstrou performance calibrada em uma base externa de 2023. Além disso, este modelo foi usado para gerar um indicador SLOSR de eficiência de UTIs que foi validado para aplicação em um subgrupo específico de Pneumonia Comunitária.

Existe espaço para futuros estudos no tema, principalmente na aplicação em outros subgrupos importantes de pacientes além de CAP, como em diagnósticos de sepse e AVC, e em pacientes de diferentes populações, pois os estudos desenvolvidos até o momento sobre o modelo proposto foram realizados em UTIs de hospitais brasileiros. Diante da importância da área de saúde e as crescentes necessidades com o envelhecimento da população, acompanhada de uma demanda de melhor eficiência e uso de recursos limitados, este estudo se soma como uma contribuição para uma melhor otimização dos recursos disponíveis ao setor de saúde, em busca da melhor utilização possível.

## 6. Referências Bibliográficas

- Arabi, Y., Venkatesh, S., Haddad, S., Al Shimemeri, A., and Al Malik, S. (2002). A prospective study of prolonged stay in the intensive care unit: predictors and impact on resource utilization. *International Journal for Quality in Health Care*, 14(5):403–410. Number: 5 Reporter: *International Journal for Quality in Health Care*.
- Breiman, L. (1996). Bagging predictors. *Machine learning*, 24(2):123–140.
- Breiman, L. (2001). Random forests. *Machine learning*, 45(1):5–32.
- Breiman, L., Friedman, J., Stone, C. J., and Olshen, R. A. (1984). *Classification and regression trees*. CRC press.
- Graybill, F. A. (1976). *Theory and application of the linear model*, volume 183. Duxbury press North Scituate, MA.
- Halpern, N. A. and Pastores, S. M. (2015). Critical care medicine beds, use, occupancy and costs in the United States: a methodological review. *Critical care medicine*, 43(11):2452. Number: 11 Reporter: *Critical care medicine*.
- Halpern, N. A., Pastores, S. M., and Greenstein, R. J. (2004). Critical care medicine in the United States 1985–2000: an analysis of bed numbers, use, and costs. *Critical care medicine*, 32(6):1254–1259. Number: 6 Reporter: *Critical care medicine*.
- Higgins, T. L., McGee, W. T., Steingrub, J. S., Rapoport, J., Lemeshow, S., and Teres, D. (2003). Early indicators of prolonged intensive care unit stay: impact of illness severity, physician staffing, and pre-intensive care unit length of stay. *Critical Care Medicine*, 31(1):45–51. Number: 1 Reporter: *Critical Care Medicine*.
- Kahn, J. M., Rubenfeld, G. D., Rohrbach, J., and Fuchs, B. D. (2008). Cost savings attributable to reductions in intensive care unit length of stay for mechanically ventilated patients. *Medical care*, pages 1226–1233. Reporter: *Medical care*.
- Kuhn, M. (2008). Building Predictive Models in R Using the caret Package. *Journal of Statistical Software*, 28(5), 1–26. <https://doi.org/10.18637/jss.v028.i05>

- Laupland, K. B., Kirkpatrick, A. W., Kortbeek, J. B., and Zuege, D. J. (2006). Long-term mortality outcome associated with prolonged admission to the ICU. *Chest*, 129(4):954–959. Number: 4 Reporter: *Chest*.
- Lunna Borges (2021). *ems: Epimed Solutions Collection for Data Editing, Analysis, and Benchmark of Health Units*. R package version 1.3.11. <https://CRAN.R-project.org/package=ems>
- Metlay, J. P., Waterer, G. W., Long, A. C., Anzueto, A., Brozek, J., Crothers, K., Cooley, L. A., Dean, N. C., Fine, M. J., Flanders, S. A., Griffin, M. R., Metersky, M. L., Musher, D. M., Restrepo, M. I., & Whitney, C. G. (2019). Diagnosis and Treatment of Adults with Community-acquired Pneumonia. An Official Clinical Practice Guideline of the American Thoracic Society and Infectious Diseases Society of America *American Journal of Respiratory and Critical Care Medicine*, 200(7), e45-e67
- Peres, I.T., Hamacher, S., Oliveira, F.L.C., Bozza, F.A., Salluh, J.I.F.: Prediction of intensive care units length of stay: a concise review. *Revista Brasileira de Terapia Intensiva* 33, 183–187 (2021)
- Peres, I.T., Hamacher, S., Oliveira, F.L.C., Bozza, F.A., Salluh, J.I.F.: Data-driven methodology to predict the icu length of stay: A multicentre study of 99,492 admissions in 109 brazilian units. *Anaesthesia Critical Care & Pain Medicine* 41(6), 101142 (2022)
- Peres, I.T., Ferrari, G.F., Quintairos, A., Bastos, L. dos S. L., & Salluh, J.I.F.: Validation of a new data-driven SLOSICU efficiency measure compared to the traditional SRU. *Intensive Care Medicine* (Vol. 49, Issue 12, pp. 1546–1548). (2023)
- R Core Team (2021). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- Rothen, H. U., Stricker, K., Einfalt, J., Bauer, P., Metnitz, P. G., Moreno, R. P., and Takala, J. (2007). Variability in outcome and resource use in intensive care units. *Intensive care medicine*, 33(8):1329–1336.
- Schoffelen, A. F., Hofhuis, J. G. M., Posthouwer, D., Rommes, J. H., and Spronk, P. E. (2010). Consumption of ICU resources by long-stay patients does not change over time: 9-year observation in a teaching hospital in the Netherlands. *Neth J Crit Care*, 14(5):313. Number: 5 Reporter: *Neth J Crit Care*.
- Verburg, I. W., de Jonge, E., Peek, N., and de Keizer, N. F. (2018). The association between outcome-based quality indicators for intensive care units. *PloS one*, 13(6):e0198522.

- Wickham H, Averick M, Bryan J, Chang W, McGowan LD, François R, Grolemund G, Hayes A, Henry L, Hester J, Kuhn M, Pedersen TL, Miller E, Bache SM, Müller K, Ooms J, Robinson D, Seidel DP, Spinu V, Takahashi K, Vaughan D, Wilke C, Woo K, Yutani H (2019). “Welcome to the tidyverse.” *Journal of Open Source Software*, \*4\*(43), 1686. doi: 10.21105/joss.01686 (URL: <https://doi.org/10.21105/joss.01686>).
- Zampieri, F. G., Ladeira, J. P., Park, M., Haib, D., Pastore, C. L., Santoro, C. M., and Colombari, F. (2014). Admission factors associated with prolonged (>14 days) intensive care unit stay. *Journal of Critical Care*, 29(1):60–65. Number: 1 Reporter: *Journal of Critical Care*.
- Zampieri FG, Soares M, Borges LP, Salluh JIF, Ranzani OT. The Epimed Monitor ICU Database®: a cloud-based national registry for adult intensive care unit patients in Brazil. *Rev Bras Ter Intensiva* 2017;29:418–26. <https://doi.org/10.5935/0103-507X.20170062>.
- Zimmerman, J. E., Kramer, A. A., McNair, D. S., Malila, F. M., and Shaffer, V. L. (2006). Intensive care unit length of stay: Benchmarking based on Acute Physiology and Chronic Health Evaluation (APACHE) IV. *Critical Care Medicine*, 34(10):2517–2529. Number: 10 Reporter: *Critical Care Medicine*.

## **APÊNDICE I – Modelos de predição de tempo de permanência**

Foram utilizados no trabalho os modelos de Regressão Linear e Random Forest, um modelo de árvores via aprendizado de máquina.

Os modelos de Regressão Linear (LR) são os mais comuns na literatura sobre tempo de permanência, principalmente para problemas de inferência (14). A vantagem desse tipo de modelo é a facilidade de compreensão e implementação. A Regressão Linear (ou Regressão Linear de Mínimos Quadrados Ordinários) visa encontrar o plano que minimiza a medida de erro entre a resposta observada e a resposta prevista. A LR escolhe os parâmetros de um conjunto de variáveis explicativas pelo princípio dos mínimos quadrados: minimizando a soma dos quadrados das diferenças entre a variável dependente observada nos dados de treinamento e aquela prevista pela função linear [Peres et al., 2022].

Modelos baseados em árvores são algoritmos de aprendizado que executam uma ou mais instruções condicionais (if-then) para particionar os dados. Esses modelos generalizam os dados de treinamento construindo uma estrutura de árvore. Dentro das partições da árvore, um modelo é usado para prever o resultado. Existem várias técnicas para construir árvores de regressão. Uma das mais antigas e amplamente utilizadas é a Árvore de Classificação e Regressão (CART). O CART é um algoritmo de árvore de decisão capaz de realizar regressão e classificação. Para a regressão, o modelo começa com o conjunto completo de dados,  $S$ , e busca cada valor distinto de cada preditor para encontrar o preditor e o valor de divisão que particionam os dados em dois grupos ( $S_1$  e  $S_2$ ) minimizando a medida geral de erro de previsão. Em seguida, dentro de  $S_1$  e  $S_2$ , esse método procura o preditor e o valor de divisão que melhor reduzem o erro. Portanto, devido à natureza recursiva da divisão em árvores de regressão, esse método também é conhecido como particionamento recursivo [Peres et al., 2022].

Bootstrap Aggregation (Bagging) foi uma das primeiras técnicas de conjunto desenvolvidas. O Bagging é uma abordagem geral que utiliza o bootstrapping em conjunto

com qualquer modelo de regressão (ou classificação) para construir um conjunto. Cada modelo no conjunto é então usado para gerar uma previsão para uma nova amostra, e essas previsões são médias para dar a previsão do modelo bagged. Modelos bagged oferecem vantagens sobre modelos que não são bagged. Enquanto a árvore de decisão “simples” CART costumava produzir previsões instáveis, o uso de árvores bagged tende a reduzir a variância de previsão, agregando muitas versões dos dados de treinamento gerados por  $k$  diferentes árvores de decisão CART [Peres et al., 2022].

Gerar amostras de bootstrap introduz um componente aleatório no processo de construção da árvore, o que induz uma distribuição de árvores e, portanto, uma distribuição de valores previstos para cada amostra. No entanto, as árvores no Bagging não são completamente independentes umas das outras, uma vez que todos os preditores originais são considerados em cada divisão de cada árvore. Considerando muitas amostras originais e uma relação entre preditores e resposta que uma árvore pode modelar, então árvores de diferentes amostras de bootstrap podem ter estruturas semelhantes entre si devido à relação subjacente. Essa característica (correlação entre árvores) impede o Bagging de otimizar, reduzindo a variância dos valores previstos. Dietterich desenvolveu a ideia de seleção de divisão aleatória, onde as árvores são construídas usando um subconjunto aleatório dos principais preditores  $k$  em cada divisão da árvore. Após avaliar esse método, Breiman construiu um algoritmo unificado chamado Random Forests. Cada modelo no conjunto é então usado para gerar uma previsão para uma nova amostra, e essas  $m$  previsões são médias para dar a previsão da floresta. Como o algoritmo seleciona aleatoriamente os preditores em cada divisão, a correlação entre árvores será necessariamente reduzida. Random forests (RF) é um método de aprendizado de máquina em conjunto baseado na construção de várias árvores de decisão CART para obter um desempenho melhor do que um modelo de “única” árvore (para regressão ou classificação). A principal técnica subjacente usada nas florestas aleatórias é a agregação de bootstrap (Bagging). O valor previsto pelo algoritmo Random Forests é o modo no caso de classificação ou o valor médio das  $k$  diferentes árvores de decisão no caso de regressão [Peres et al., 2022].