

UNIVERSIDADE FEDERAL DE SANTA CATARINA
CENTRO TECNOLÓGICO
DEPARTAMENTO DE ENGENHARIA ELÉTRICA E ELETRÔNICA
CURSO DE GRADUAÇÃO EM ENGENHARIA ELÉTRICA

Henrique Pickler da Silva

**Predição de Incidência de Lesão por Pressão em UTI através de Registros
Eletrônicos e Aprendizado de Máquina**

Florianópolis
2021

Henrique Pickler da Silva

**Predição de Incidência de Lesão por Pressão em UTI através de Registros
Eletrônicos e Aprendizado de Máquina**

Trabalho de Conclusão de Curso de Graduação em Engenharia Elétrica do Centro Tecnológico da Universidade Federal de Santa Catarina para a obtenção do título de Bacharel em Engenharia Elétrica.
Orientador: Prof. Danilo Silva, Ph.D.

Florianópolis

2021

Ficha de identificação da obra elaborada pelo autor,
através do Programa de Geração Automática da Biblioteca Universitária da UFSC.

da Silva, Henrique Pickler

Predição de incidência de lesão por pressão em UTI
através de registros eletrônicos e aprendizado de máquina /
Henrique Pickler da Silva ; orientador, Danilo Silva, 2021.
66 p.

Trabalho de Conclusão de Curso (graduação) -
Universidade Federal de Santa Catarina, Centro Tecnológico,
Graduação em Engenharia Elétrica, Florianópolis, 2021.

Inclui referências.

1. Engenharia Elétrica. 2. Lesão por pressão. 3.
Aprendizado de máquina. 4. Registros eletrônicos. I. Silva,
Danilo. II. Universidade Federal de Santa Catarina.
Graduação em Engenharia Elétrica. III. Título.

Henrique Pickler da Silva

**Predição de Incidência de Lesão por Pressão em UTI através de Registros
Eletrônicos e Aprendizado de Máquina**

Este Trabalho Conclusão de Curso foi julgado adequado para obtenção do Título de Bacharel em Engenharia Elétrica e aprovado em sua forma final pelo Curso de Engenharia Elétrica .

Florianópolis, 23 de Setembro de 2021.

Prof. Jean Viane Leite, Dr.
Coordenador do Curso

Banca Examinadora:

Prof. Danilo Silva, Ph.D.
Orientador
Universidade Federal de Santa Catarina

Prof. Márcio Holsbach Costa, Dr.
Avaliador
Universidade Federal de Santa Catarina

Prof. Ricardo Faria Giglio, Ph.D.
Avaliador
Universidade Federal de Santa Catarina

AGRADECIMENTOS

Dou meus mais sinceros e profundos agradecimentos ao meu orientador Danilo Silva, por me orientar nesse trabalho e na minha vida acadêmica.

Aos meus colegas Arthur, Dominique e Saulo, por me auxiliarem e me acompanharem nesse projeto. Agradeço também à empresa 3778 em si por fornecer todos os recursos necessários para a execução desse projeto.

Aos meus queridos amigos que tanto me ajudaram na elaboração desse trabalho, Pedro, Vitor e Gabriel. Digo sem dúvida alguma que não teria chegado aqui se não fosse pelo companheirismo, dicas e bons momentos proporcionados por vocês.

À Lintsey por estar ao meu lado e sempre me apoiar.

À minha família e principalmente aos meus pais por, além do apoio e ensinamentos, me permitirem me dedicar exclusivamente aos meus estudos através do enorme esforço que fizeram por mim.

Por fim agradeço à UFSC por me proporcionar, além de muitas oportunidades, um ambiente de ensino maravilhoso com os melhores professores que já tive e aos quais também sou muito grato.

RESUMO

Lesões por pressão, também chamadas por escaras, são ocorrências que podem prolongar em meses a estada de um paciente. Em estágios iniciais é possível evitar o agravamento da ferida, assim como suas adversidades, com tratamentos especializados e atenção preventiva. Porém, é inviável oferecer esse nível de cuidado para todo paciente internado devido ao alto custo financeiro e de mão de obra. Atualmente existem ferramentas utilizadas por profissionais de enfermagem para identificar pacientes em risco auxiliando o direcionamento de medidas preventivas. Embora essas ferramentas sejam usadas largamente por hospitais do mundo todo, a aplicação de modelos de aprendizado de máquina para predição de incidência de lesão por pressão é uma área nova e com poucos resultados na literatura. Este trabalho busca desenvolver uma nova abordagem ao problema na tentativa de aproximar a avaliação de modelos ao uso prático das ferramentas usadas atualmente. Dessa forma, pode ser feita uma comparação justa entre os modelos desenvolvidos e as técnicas utilizadas na prática. Foram testados 6 modelos em três bases de dados diferentes, duas de momentos diferentes de um hospital de Boston e a terceira de um hospital brasileiro. Os resultados apontam que os modelos de aprendizado de máquina podem superar significativamente as técnicas atuais em todos os três conjuntos de dados. Além disso, devido à natureza dos dados esse modelo pode ser integrado em sistemas de monitoramento e utilizado de maneira automática, diminuindo inclusive a mão de obra associada à identificação de risco.

Palavras-chave: Lesão por pressão, aprendizado de máquina, registros eletrônicos.

ABSTRACT

Pressure ulcers are occurrences that may postpone up to months an hospital stay. They are easily preventable by means of specialized treatment. Nonetheless, due to high monetary and labor cost, it is impracticable to offer this level of care to every patient. Currently, there exists tools used by professional care workers to assess at-risk patients directing their labor towards those. Though these tools are largely used by hospitals around the world, the usage of machine learning models for predicting pressure ulcer incidence is a fairly new area with few results published. This work seeks to develop a new approach to this problem by trying to approximate the evaluation and training of models to the practical usage of tools currently in use. This way, a fair comparison can be made between the models developed and the techniques used in practice. Six different models were evaluated in three separate datasets, two from a hospital located in Boston and the third one from a Brazilian hospital. The results show that machine learning models can perform significantly better than what's used in practice. Furthermore, due to the nature of the data, this model can be integrated in critical care monitoring systems enabling patients to be automatically assessed reducing labor associated with manual assessment.

Keywords: Pressure ulcer, machine learning, electronic health records.

LISTA DE FIGURAS

Figura 1 – Exemplo de classificação binária	25
Figura 2 – Exemplo de curva ROC	26
Figura 3 – Exemplo de uma curva da precisão por sensibilidade	27
Figura 4 – Exemplo de árvore de decisão	29
Figura 5 – Exemplo de um modelo de floresta aleatória	30
Figura 6 – Abordagem usada por Cramer <i>et al.</i> (2019)	36
Figura 7 – Representação da abordagem de janelas	36
Figura 8 – Processo de extração dos dados da tabela <i>CHARTEVENTS.csv.gz</i>	38
Figura 9 – Diagrama representando o processo de criação da matriz de atributos	41
Figura 10 – Diagrama das filtragens realizadas e criação de janelas	42
Figura 11 – Sumarização dos pontos de operação para os melhores modelos no MIMIC-III.	47
Figura 12 – Sumarização dos pontos de operação para os melhores modelos no MIMIC-IV.	48
Figura 13 – Sumarização dos pontos de operação para os melhores modelos no HMD.	49
Figura 14 – Sumarização dos pontos de operação para todos os modelos no MIMIC-III.	64
Figura 15 – Sumarização dos pontos de operação para todos os modelos no MIMIC-IV.	65
Figura 16 – Sumarização dos pontos de operação para todos os modelos no HMD.	66

LISTA DE TABELAS

Tabela 1 – Classificações de lesões por pressão de acordo com o NPIAP	23
Tabela 2 – Resultados das métricas no MIMIC-III	47
Tabela 3 – Resultados das métricas no MIMIC-IV	48
Tabela 4 – Resultados das métricas no HMD	49
Tabela 5 – Tabela com estatísticas dos atributos disponíveis em cada conjunto de dados	59
Tabela 6 – Tabela com porcentagem de dados faltantes	61
Tabela 7 – Configurações dos hiperparâmetros usados na otimização	63

LISTA DE ABREVIATURAS E SIGLAS

AUC	Area Under the ROC Curve
CID-10	Classificação Internacional de Doenças, 10 ^a revisão
CID-9	Classificação Internacional de Doenças, 9 ^a revisão
HMD	Hospital Mater Dei
LP	Lesão por Pressão
LTP	Lesão Tecidual Profunda
MIMIC	Medical Information Mart for Intensive Care
MIMIC-III	Medical Information Mart for Intensive Care III
MIMIC-IV	Medical Information Mart for Intensive Care IV
NPIAP	National Pressure Injury Advisory Panel
ROC	Receiver Operating Characteristic
TFP	Taxa de Falso Positivo
TVP	Taxa de Verdadeiro Positivo
UTI	Unidade de Tratamento Intensivo
VPP	Valor Preditivo Positivo

LISTA DE SÍMBOLOS

Sen	Sensibilidade
vp	Número de verdadeiros positivos
fn	Número de falsos negativos
TFP	Taxa de falsos positivos
fp	Número de falsos positivos
vn	Número de verdadeiros negativos
N	Número total de amostras
P_p	Número de amostras classificadas positivas
P_n	Número de amostras classificadas negativas
ρ	Prevalência
N_p	Número de amostras positivas
Pre	Precisão

SUMÁRIO

1	INTRODUÇÃO	21
1.1	Objetivos	22
1.1.1	Objetivo geral	22
1.1.2	Objetivos específicos	22
2	FUNDAMENTAÇÃO TEÓRICA	23
2.1	Lesão por pressão e escala Braden	23
2.2	Classificação binária	24
2.3	Métricas	25
2.4	Modelos	27
2.4.1	Regressão logística	28
2.4.2	Árvores de decisão	28
2.4.3	Floresta aleatória	30
2.4.4	<i>Gradient boosting</i>	31
2.5	Trabalhos relacionados	31
3	DESENVOLVIMENTO	35
3.1	Formulações do problema	35
3.2	Bases de dados	35
3.2.1	MIMIC-III	36
3.2.2	MIMIC-IV	37
3.2.3	HMD	38
3.2.4	Extração de atributos	39
3.2.5	Criação da matriz de atributos e variável alvo	40
3.3	Métodos	41
3.3.1	Filtragem dos dados	41
3.3.2	Pré-processamento	41
3.3.3	Metodologia de treinamento e avaliação	42
4	RESULTADOS	45
4.1	Discussão	45
5	CONCLUSÃO	51
	REFERÊNCIAS	53
	APÊNDICE A – TABELAS DE ESTATÍSTICAS	59
	APÊNDICE B – TABELAS DE HIPERPARÂMETROS	63
	APÊNDICE C – CURVAS DE TODOS OS MODELOS	64

1 INTRODUÇÃO

A Lesão por Pressão (LP), também chamada de úlcera por pressão ou de decúbito, é um ferimento na pele, que pode se estender até o tecido subcutâneo, causado por pressão prolongada que geralmente ocorre em locais de protuberância óssea. Pacientes de Unidade de Tratamento Intensivo (UTI) possuem risco elevado de desenvolver LP com incidências de 3% a 33% devido à relação da LP com fatores comuns a pacientes de UTI como imobilidade, comorbidades, tempo de estada prolongado, baixa perfusão, uso de vasopressores, dentre outros fatores (ALDERDEN *et al.*, 2017).

O desenvolvimento de LP em pacientes representa um aumento no tempo de recuperação, podendo prolongar em até meses a estada no hospital e resultando em um período de afastamento prolongado. Isso ocorre pois em casos mais graves os danos podem estender-se até os músculos e outros tecidos, permanecendo mesmo após fechamento da ferida.

Em estágios iniciais a LP aparece como um avermelhamento e endurecimento da pele na região. Nesse estágio a recuperação é rápida e não representa risco ao paciente, porém para prevenir casos mais severos é necessário identificar pacientes em risco pois os tratamentos preventivos ou envolvem intensa mão de obra, ou estão associados a um custo elevado. Para esse propósito, utilizam-se na prática escalas que são atribuídas a cada paciente, como as escalas de Norton, de Gosnell, de Waterlow e de Braden (BERGSTROM *et al.*, 1987).

A escala utilizada com maior frequência é a escala Braden que envolve a atribuição pelo profissional de enfermagem de seis notas individuais relacionadas a fatores de risco, que ao serem somadas proporcionam um valor que identifica o risco do paciente de desenvolver LP. Quanto maior a nota atribuída, menor o risco. Por ser extensivamente utilizada, existem vários estudos que validam o poder preditivo da escala Braden. No entanto, por ser uma escala subjetiva, as atribuições das notas podem ser inconsistentes entre hospitais e entre profissionais de enfermagem.

Nos últimos anos, o uso de sistemas de monitoramento de pacientes em UTIs tornou-se cada vez mais frequente. Esses sistemas armazenam os dados obtidos em cada internação, permitindo o desenvolvimento de modelos de aprendizado de máquina. Neste trabalho serão desenvolvidos modelos de predição de incidência de LP a partir desses dados. Esses modelos podem auxiliar equipes de enfermagem a direcionar seus esforços a pacientes com risco, eliminando também a mão de obra associada à atribuição da escala Braden para cada paciente.

Essa abordagem já foi explorada por Cramer *et al.* (2019) utilizando o banco de dados Medical Information Mart for Intensive Care III (MIMIC-III) (JOHNSON *et al.*, 2016) que traz dados de registros eletrônicos de 61,532 internações em UTI de 46,476 pacientes distintos. O artigo analisa vários modelos diferentes e reporta o desempenho em

termos de precisão e sensibilidade, mostrando que o melhor modelo possui desempenho superior ao da escala Braden.

1.1 Objetivos

Tomou-se como base o trabalho de Cramer *et al.* (2019) tendo como propósito reproduzir os resultados apresentados no artigo. Visando aperfeiçoar os resultados, buscou-se melhorar vários processos realizados no trabalho, desde a extração dos dados e geração de atributos, até a avaliação dos modelos e a própria abordagem utilizada. Além disso, buscou-se validar os resultados em outras duas bases de dados.

1.1.1 Objetivo geral

O objetivo deste trabalho é criar um modelo que possa predizer o risco de pacientes em UTI de desenvolverem LP ao longo da sua estada que, após validação, possa substituir técnicas manuais utilizadas atualmente como a escala Braden.

1.1.2 Objetivos específicos

Os objetivos específicos identificados que guiam a elaboração deste trabalho são:

- a) Reproduzir os resultados apresentados no artigo de Cramer *et al.* (2019)
- b) Extrair mais dados para predição e melhorar o desempenho obtido pelo artigo
- c) Desenvolver uma abordagem que se aproxime à aplicação prática da escala Braden para uma avaliação mais justa e precisa
- d) Obter e comparar o desempenho de múltiplos modelos
- e) Validar o modelo na base Medical Information Mart for Intensive Care IV (MIMIC-IV)
- f) Validar o modelo na base do Hospital Mater Dei (HMD)

2 FUNDAMENTAÇÃO TEÓRICA

Nessa seção serão revisados alguns conceitos básicos relacionados às LPs, assim como uma revisão básica de problemas de classificação binária e de modelos de aprendizado de máquina. Por fim, será feita uma revisão dos trabalhos relacionados na literatura.

2.1 Lesão por pressão e escala Braden

Os principais locais de incidência de LP são nos calcânhares, tornozelos, quadril e cóccix, por serem os locais que ficam sob pressão em posições deitadas e por terem protuberâncias ósseas. As LPs são classificadas em 6 categorias diferentes pelo National Pressure Injury Advisory Panel (NPIAP) (EDSBERG *et al.*, 2016), indicadas na Tabela 1. Essa classificação é usada largamente por hospitais do mundo.

Tabela 1 – Classificações de lesões por pressão de acordo com o NPIAP

Nível	Descrição
1	Pele intacta, com rubor não branqueável, numa área localizada normalmente sobre uma proeminência óssea.
2	Perda parcial da espessura da pele que se apresenta como uma ferida com base avermelhada, podendo haver esfacelo na área, também podem estar presentes bolhas na pele.
3	Perda total da espessura tecidual - tecido adiposo subcutâneo pode ser visível, mas não estão expostos os ossos, tendões ou músculos
4	Danos em todas as camadas profundas da pele com exposição de ossos, tendões ou juntas.
LTP	Ferida roxa ou marrom sem abertura de ferida ou com bolhas de sangue devido a danos causados em tecidos profundos.
Inclassificáveis	LP nível 3 ou 4 não sendo possível determinar devido a detritos, descamação ou escarificação na ferida.

Fonte - adaptado do NPIAP

Em geral, o tempo de recuperação aumenta com o nível da LP, sendo de grande importância a identificação da lesão em níveis iniciais e de pacientes em risco (PROQUALIS, 2021). A prevenção de LP envolve medidas como frequente mudança de posição, hidratação da pele, intervenção nutricional, uso de colchões especiais, e uso de espumas multicamadas que distribuem a pressão em regiões de proeminência óssea. Além disso é importante evitar fontes de fricção como rugas no lençol do paciente (PROQUALIS, 2021). Essas medidas exigem esforço elevado da equipe de enfermagem, podendo haver também um custo elevado tanto no uso de equipamentos especiais quanto no treinamento para identificação de fatores ligados à LP.

Uma das ferramentas mais utilizadas para identificação de pacientes em risco de LP é a escala Braden (BERGSTROM *et al.*, 1987). A escala Braden avalia seis fatores de risco sendo eles percepção sensorial, mobilidade, nutrição, fricção, umidade e atividade física. São atribuídas notas de 1 a 4 para cada um dos fatores de risco, exceto para fricção que é atribuída uma nota entre 1 e 3. Para obtenção da nota Braden total é feita a soma dessas notas, onde uma nota menor indica maior risco de desenvolver LP.

Existem muitos trabalhos na literatura que avaliam o poder preditivo da escala Braden, como em Hyun *et al.* (2013), Serpa *et al.* (2011a) e em Paranhos e Santos (1999). Esses estudos avaliam o poder preditivo de atribuições subsequentes da escala Braden em internações de UTI, que na prática pode ser obtida diariamente ou até com maior frequência. Existem também estudos do poder preditivo das notas individuais da escala Braden, como em Serpa *et al.* (2011b). Além disso, algumas notas são de difícil avaliação sofrendo de falta de clareza na sua atribuição, como as notas individuais de nutrição e percepção sensorial, e portanto podem ser inconsistentes entre enfermeiros ou hospitais dificultando a validação preditiva da ferramenta.

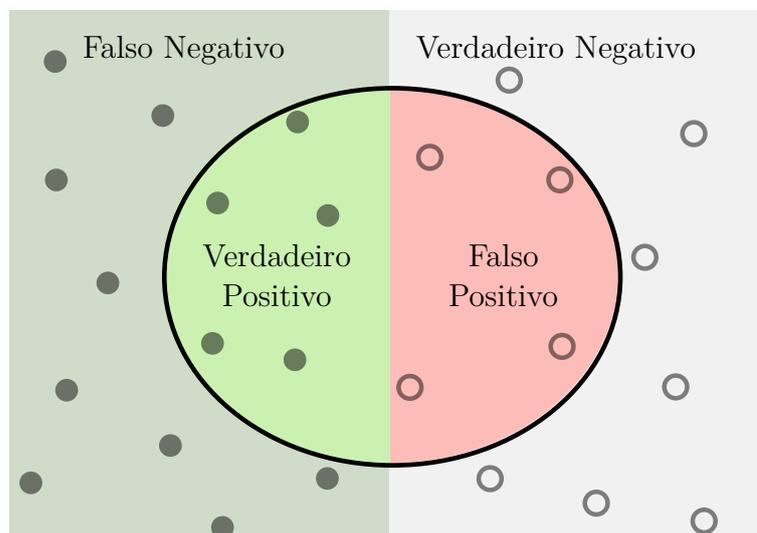
2.2 Classificação binária

A classificação binária consiste em decidir entre dois grupos a qual destes um dado elemento pertence. Um exemplo é um teste feito para identificar pacientes com uma doença. Pessoas com a doença são ditas da classe positiva e pessoas sem a doença são da classe negativa. A partir de alguma medição, possivelmente ruidosa, o teste irá dividir cada paciente entre as duas classes, geralmente apresentando erros de classificação devido a ruídos ou falta de informação. Como existem duas classes podem haver dois tipos de erros e dois tipos de acertos, como demonstrado na Figura 1.

No contexto de aprendizado de máquina geralmente existem muitas medições e busca-se uma função que, além de prever a classe do elemento, generalize para outros elementos e possivelmente outros conjuntos de dados. Os modelos de aprendizado de máquina implementam estruturas para obter essa função analisando amostras de treinamento e otimizando uma métrica associada ao problema. Existem algumas técnicas para verificar se há generalização, por exemplo separar um conjunto de teste com amostras não utilizadas no treinamento e verificar se o desempenho, em termos da métrica utilizada, é similar entre os dois conjuntos.

A função obtida em geral é uma função sobre o domínio dos reais, onde para obter o modelo de classificação binária é necessário definir um limiar ou ponto de operação. Ou seja, para um único modelo é possível obter inúmeras métricas ao variar seu limiar. Na escala Braden, por exemplo, um limiar utilizado para classificar pacientes em alto risco é Nota ≤ 12 . Para esse limiar pode-se determinar a quantidade de erros e acertos totais. E, por exemplo, ao aumentar o limiar para 16 certamente serão classificados mais pacientes

Figura 1 – Exemplo de classificação binária. Os círculos preenchidos são amostras positivas, enquanto que os círculos não preenchidos são amostras negativas. As amostras dentro do círculo são classificadas como classe positiva. As áreas indicam os tipos de erros e acertos.



Fonte - do autor

como positivo e portanto haverá um aumento tanto na quantidade de verdadeiros positivos quanto na quantidade de falsos positivos, sendo que a proporção do aumento entre as duas dependerá da qualidade do classificador.

2.3 Métricas

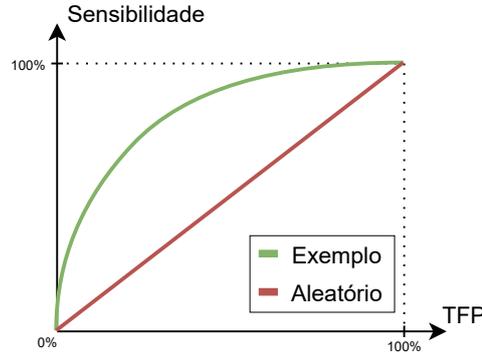
A comparação de modelos de classificação pode ser feita analisando um único ponto de operação, ou obtendo métricas que sumarizam o desempenho médio de todos os pontos de operação. Comumente usa-se a curva Receiver Operating Characteristic (ROC) que expressa a sensibilidade, conhecida também por *recall* (revocação) ou Taxa de Verdadeiro Positivo (TVP), em função da Taxa de Falso Positivo (TFP). A sensibilidade Sen e a TFP são dados por

$$Sen = \frac{vp}{vp + fn}, \quad (1)$$

$$TFP = \frac{fp}{fp + vn}. \quad (2)$$

A sensibilidade indica a proporção de amostras positivas que foram identificadas corretamente, enquanto a TFP indica a proporção de amostras classificadas como positivas sendo que eram negativas. Nota-se que a curva ROC inicia em $Sen = TFP = 0$ pois pode-se sempre escolher um limiar que classifique todas as amostras como negativas tal que $vp = 0$ e $fp = 0$, com fn e vn não nulos. A curva termina em $Sen = TFP = 1$ pois,

Figura 2 – Exemplo de curva ROC. Em verde um modelo de exemplo e em vermelho um modelo aleatório.



Fonte - do autor

do mesmo modo, sempre é possível escolher um limiar que classifique todas as amostras como positivas dando $vp = N_p$, $fn = 0$, $fp = N_n$ e $vn = 0$.

É comum tomar como base o classificador aleatório, portanto é importante entender como este é representado na curva ROC. Supondo um caso limite com a média de muitas realizações de um classificador ou um número muito grande de amostras N , um dado limiar irá classificar P_p amostras como positivas e P_n como negativas. Nesse caso, para uma prevalência $\rho = \frac{N_p}{N}$ tem-se que

$$vp = P_p \cdot \rho, \quad (3)$$

$$fp = P_p(1 - \rho), \quad (4)$$

$$fn = P_n \cdot \rho = (N - P_p)\rho, \quad (5)$$

$$vn = P_n(1 - \rho) = (N - P_p)(1 - \rho). \quad (6)$$

Portanto

$$\text{Sen} = \frac{vp}{vp + fn} = \frac{P_p \cdot N_p}{P_p \cdot \rho + (N - P_p)\rho} = \frac{P_p}{N}, \quad (7)$$

$$\text{TFP} = \frac{fp}{fp + vn} = \frac{P_p(1 - \rho)}{P_p(1 - \rho) + (N - P_p)(1 - \rho)} = \frac{P_p}{N}. \quad (8)$$

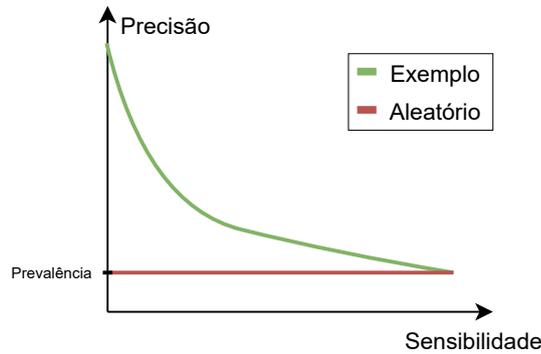
Ou seja, a tendência para um classificador aleatório é que $\text{Sen} = \text{TFP}$ indicando que a curva ROC será linha com inclinação unitária. Apresenta-se na Figura 2 um exemplo de curva ROC.

Quando não se sabe um ponto de operação definido para avaliar o modelo é comum utilizar uma métrica que sumariza a curva ROC em um número, como a Area Under the ROC Curve (AUC), sendo que um classificador aleatório teria $AUC = 0.5$.

Uma outra métrica utilizada nesse trabalho, é a precisão Pre, também conhecida por Valor Preditivo Positivo (VPP). A precisão é dada por

$$\text{Pre} = \frac{vp}{vp + fp}. \quad (9)$$

Figura 3 – Exemplo de uma curva da precisão por sensibilidade. Em verde um modelo de exemplo e em vermelho um classificador aleatório.



Fonte - do autor

A precisão indica a proporção de amostras positivas em relação a todas as predições positivas. Nota-se que a precisão e a sensibilidade estão associadas, comumente ao aumentar a sensibilidade diminui-se a precisão, porém essa relação nem sempre é respeitada dependendo da qualidade do classificador.

Um valor com interpretação importante para o problema de incidência de LP é o recíproco da precisão. Esse valor indica em média quantas predições positivas são feitas para cada verdadeiro positivo. No contexto de predição de incidência de LP esse valor estará relacionado ao custo necessário, em termos de tratamento preventivo, para evitar uma LP assumindo que a prevenção é 100% efetiva. Por outro lado, a sensibilidade irá indicar a quantidade de LP evitadas. Nota-se que há um *tradeoff* entre essas duas métricas, agora mais evidente para o problema a ser abordado nesse estudo. Devido a esse *tradeoff* entre precisão e sensibilidade, para comparar modelos de classificação de maneira justa, deve-se fixar uma das métricas que estará associada a um limiar.

Assim como na curva ROC, pode-se traçar uma curva da precisão em função da sensibilidade. Para entender o formato da curva, volta-se ao classificador aleatório, definido anteriormente nas Equações (3) até (6). Sua precisão será dada por

$$\text{Pre} = \frac{vp}{vp + fp} = \frac{P_p \cdot \rho}{P_p \cdot \rho + P_p(1 - \rho)} = \rho. \quad (10)$$

Ou seja, a precisão é constante e possui valor igual à prevalência. Na Figura 3 há um exemplo de curva precisão por sensibilidade.

2.4 Modelos

Existem vários modelos de aprendizado de máquina disponíveis para uso, neste estudo focou-se em modelos de árvores devido à facilidade de lidar com atributos categóricos, bom desempenho e robustez a pontos fora da curva (*outliers*). Foi incluída na análise a

regressão logística devido à simplicidade e à frequência de seu uso na literatura, inclusive sendo o melhor modelo reportado por Cramer *et al.* (2019).

2.4.1 Regressão logística

A regressão logística modela a probabilidade das amostras serem de uma dada classe a partir do conjunto de treinamento. Esse modelo ajusta uma função logística aos dados, geralmente buscando maximizar a verossimilhança (MURPHY, 2012). No caso de múltiplas variáveis de predição, a função logística é dada por

$$\tilde{y}(\mathbf{X}) = \frac{1}{1 + \exp(-c - \mathbf{X}^T \cdot \mathbf{w})}, \quad (11)$$

onde c está relacionado com o valor da ordenada ao ser interceptada e \mathbf{w} são os parâmetros do modelo que indicam o aumento do logaritmo das chances em relação a cada variável em \mathbf{X} . Uma perda comumente utilizada para o ajuste desse modelo é a entropia cruzada com regularização ℓ_2 dada por

$$\min_{\mathbf{w}, c} \left(\frac{1}{2} \mathbf{w}^T \mathbf{w} - C \sum_{i=0}^n [y_i \log \tilde{y}_i + (1 - y_i) \log(1 - \tilde{y}_i)] \right) \quad (12)$$

sendo C um hiper-parâmetro que controla o nível de regularização. A regularização ℓ_2 , representada pelo termo $\frac{1}{2} \mathbf{w}^T \mathbf{w}$, simplesmente pondera a norma dos parâmetros do modelo junto à perda associada ao ajuste. Dessa forma ao aumentar a regularização obtém-se modelos com parâmetros de menor magnitude, ou seja menor variância, e que possivelmente generalizam melhor para amostras fora do conjunto de treinamento. No entanto, uma regularização muito alta pode diminuir a importância do ajuste na perda, resultando em um modelo com baixo desempenho. Com o modelo ajustado, pode-se interpretar que $\tilde{y} = P(y = 1 | \mathbf{X})$, ou seja, o modelo aproxima a função de probabilidade a posteriori associada ao conjunto de dados.

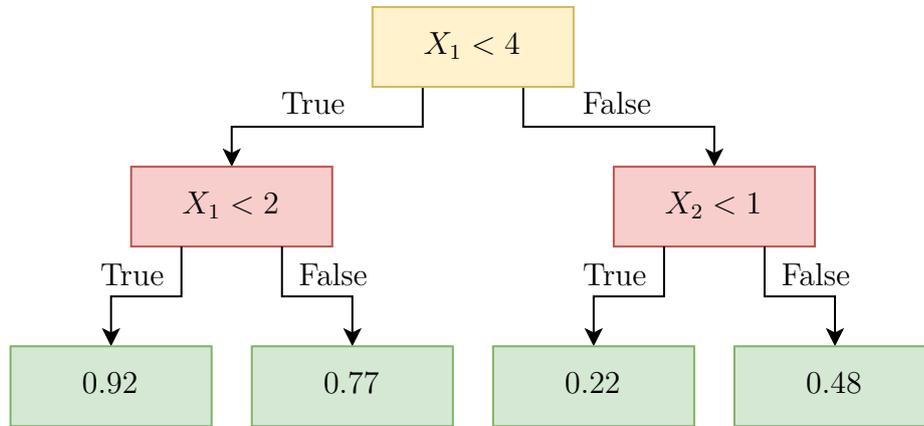
2.4.2 Árvores de decisão

Árvores de decisão (*Decision Trees*) são modelos que implementam uma árvore binária com regras treinadas. As regras binárias geralmente são limiarizações nas variáveis observadas que são seguidas até alcançar uma folha que dará o valor da predição. A Figura 4 mostra um exemplo de árvore de decisão. As regras são escolhidas a partir de uma perda medida nos dois nós resultantes da regra. Geralmente é utilizada a impureza de Gini que, para o problema multiclasse, é dada por

$$I_G(\mathbf{p}) = \sum_{i=1}^N p_i(1 - p_i) \quad (13)$$

onde N é a quantidade de classes com probabilidades p_i obtidas a partir da distribuição do conjunto de interesse. A impureza de Gini mede a probabilidade média de uma amostra ser

Figura 4 – Exemplo de árvore de decisão de profundidade 2 e com 4 folhas. Em vermelho estão os nós com as regras binárias, no topo está a raiz e em verde estão as folhas com o valor da predição.



Fonte - do autor

classificada erroneamente, ou seja, mede o nível de separação de classes em um conjunto de amostras.

Para a classificação binária, a impureza de Gini é dada por

$$I_G(p) = 2p(p - 1) \quad (14)$$

onde p é a probabilidade da classe positiva no conjunto de interesse. O treinamento é baseado em dividir cada nó a partir da regra que minimiza a perda nos dois nós gerados. A perda é ponderada pela quantidade de amostras nos dois nós e são somadas. A perda total de um nó com M amostras é expressa por

$$L = \frac{M_E}{M} I_G(p_E) + \frac{M_D}{M} I_G(p_D) \quad (15)$$

onde M_E refere-se à quantidade de amostras na folha esquerda com probabilidade (proporção) da classe positiva p_E , enquanto M_D e p_D referem-se às mesmas grandezas na folha à direita.

Geralmente as regras implementadas são simples e se restringem apenas a limia- rizações dos atributos, normalmente havendo uma limitação na quantidade de limiares testados escolhidos a partir dos quantis de cada atributo. O processo de divisão de nós é feito até não ser mais possível dividir o conjunto de amostras ou até atingir algum critério de parada. A partir disso a saída do modelo é dada pela porcentagem de representação da classe majoritária em cada folha. Existem várias implementações de algoritmos para trei- namento de árvores de decisão, a implementação utilizada neste trabalho é a do *SKLearn* e é conhecida como *CART* (*Classification and Regression Trees*) proposta por Breiman *et al.* (1984).

Árvores de decisão tendem a sofrer de sobreajuste (*overfitting*), ou seja, tendem a criar regras influenciadas pelo ruído do conjunto de treinamento. Para lidar com isso

existem técnicas que buscam limitar o tamanho da árvore, e portanto sua complexidade, evitando obter regras específicas demais que capturam o ruído no conjunto de treinamento. Geralmente é imposto um limite na profundidade da árvore ou a quantidade de folhas.

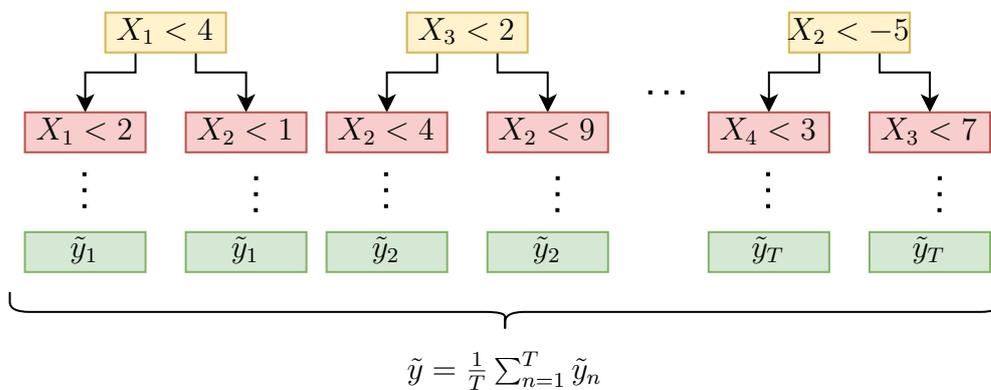
Embora tenham pontos fracos, árvores de decisão são modelos simples, com uma rotina de treinamento simples, e são fáceis de interpretar pois são auto-explicativas já que a predição estará associada às regras treinadas pelo modelo. Inclusive, são ferramentas utilizadas comumente fora do âmbito de aprendizado de máquina, definindo estratégias de operação ou de decisão a partir de regras heurísticas.

2.4.3 Floresta aleatória

O modelo de floresta aleatória (*Random Forest*) utiliza aprendizado em conjunto em que é feito o treinamento de múltiplas árvores de decisão independentes e a predição final é dada pelo voto majoritário entre elas. Florestas aleatórias são modelos caixa preta, ou seja, há uma grande complexidade em tentar explicar qual a regra de classificação implementada pelo modelo, e isso ocorre devido ao processo de aprendizado em conjunto.

O treinamento de cada árvore é feito individualmente utilizando um subconjunto dos atributos podendo ser feita uma subamostragem das amostras também. A partir disso, é possível obter árvores especializadas em sub-áreas do conjunto de treinamento, e em geral cada árvore individual sofrerá de sobreajuste. Na implementação utilizada¹ a saída do modelo será dada pela média das saídas das árvores, embora na abordagem original a saída é dada pelo voto majoritário (BREIMAN, 2001). Pelo processo de média há uma diminuição do sobreajuste no modelo como um todo, reduzindo sua variância, e geralmente provendo um desempenho superior ao de uma única árvore de decisão. Na Figura 5 há um exemplo de um modelo de floresta aleatória.

Figura 5 – Exemplo de um modelo de floresta aleatória com T árvores de decisão, representadas de um modo simplificado.



Fonte - do autor

¹ disponibilizada pelo *SKLearn*

2.4.4 Gradient boosting

Gradient boosting é uma técnica de aprendizado em conjunto onde é feita a associação em série de vários modelos com baixa complexidade, geralmente sendo árvores de decisão com pouca profundidade ou quantidade de folhas. A associação é feita de modo que um novo modelo h_m contribua com a predição F_m de modelos anteriores onde a partir de uma predição inicial F_0 tem-se

$$F_{m+1}(\mathbf{X}) = F_m(\mathbf{X}) + h_m(\mathbf{X}) = y \quad (16)$$

por outro lado

$$h_m(\mathbf{X}) = y - F_m(\mathbf{X}) \quad (17)$$

ou seja, cada modelo subsequente será treinado para reduzir o resíduo dos modelos anteriores. A função perda utilizada para problemas de classificação, assim como na regressão logística, geralmente está associada à maximização da verossimilhança. Quando utiliza-se árvores de decisão há um processo de encontrar o valor de saída que minimiza a perda junto com a predição do modelo anterior. É difícil encontrar o valor ótimo, porém pode-se utilizar uma aproximação equivalente a um passo do método de Newton-Raphson (FRIEDMAN, 1999) sendo, para uma folha t que possui conjunto de amostras L_t , dada por:

$$\gamma_t = \sum_{x \in L_t} \frac{y - F_m(x)}{F_m(x)(1 - F_m(x))}. \quad (18)$$

Por fim, é comum ponderar a predição de cada árvore subsequente por um parâmetro λ que pode ser interpretado como a taxa de aprendizado. Assim, a Equação 16 torna-se

$$F_{m+1}(\mathbf{X}) = F_m(\mathbf{X}) + \lambda h_m(\mathbf{X}) \quad (19)$$

Existem várias implementações de *gradient boosting* para árvores de decisões. Nesse estudo foram utilizadas três implementações diferentes, sendo elas a implementação disponibilizada na biblioteca em Python *SKLearn* (PEDREGOSA *et al.*, 2011), o *XGBoost* (CHEN; GUESTRIN, 2016) e o *CatBoost* (PROKHORENKOVA *et al.*, 2019). A implementação da *SKLearn* e do *XGBoost* são equivalentes na sua abordagem porém o segundo é consideravelmente mais rápido e eficiente em termos de uso de memória. Já o *CatBoost* propõe mudanças no algoritmo de *gradient boosting*. Nessa implementação foi resolvido um tipo especial de vazamento de informação que ocorre ao longo do treinamento das árvores em outras implementações, além de propor uma nova estratégia para gerar estatísticas de atributos categóricos de alta cardinalidade.

2.5 Trabalhos relacionados

A predição automática de LP através de registros eletrônicos já foi explorada em alguns artigos da literatura. Nesta seção serão citados alguns trabalhos publicados e será

feita uma discussão das suas metodologias e resultados. Em geral, as abordagens variam entre avaliar se um modelo consegue prever pacientes em risco de LP, e prever se um paciente irá ter LP na sua estadia. É importante notar a diferença entre as duas abordagens, em uma é feita a classificação de pacientes usando uma sumarização dos dados da estadia inteira, na outra a classificação é feita em uma estadia usando dados disponíveis até o momento da predição. É comum também excluir pacientes que não tiveram tempo de internação longo o suficiente (e.g. 24 horas) e pacientes que já tinham LP na hora de internação pois já foram expostos a riscos que não estarão inclusos nos dados disponíveis. Geralmente busca-se prever a ocorrência de LP nível 2 ou maior, já que a classificação de LP nível 1 pode ser facilmente atribuída erroneamente produzindo ruído significativo.

Avaliando o poder preditivo da escala Braden, obtida na admissão do paciente, de classificar pacientes em risco, Hyun *et al.* (2013), concluiu que ela não obteve desempenho satisfatório em pacientes de UTI. Em seguida, incluindo dados de diagnóstico e medicamento, Kaewprag *et al.* (2015) avaliou modelos clássicos de aprendizado de máquina obtendo um resultado superior à escala Braden. No entanto não fica claro se foram utilizados todos os dados registrados durante a estadia do paciente, incluindo após o desenvolvimento da LP, ou se houve uma filtragem temporal. Isso é importante pois os modelos são comparados com o desempenho da escala Braden de admissão, que intuitivamente terá um desempenho pior que a nota Braden mais recente ou um modelo que possui informações mais próximas à incidência de LP. Além disso, há uma comparação entre combinações dos três conjuntos de atributos, Medicação, Diagnósticos e Braden. Porém não é fixada uma das métricas, portanto não é possível compará-las a partir dessas métricas, sendo possível comparar apenas valores de AUC. Esses resultados foram melhorados em um estudo subsequente (KAEWPRAK *et al.*, 2017), utilizando redes Bayesianas.

Buscando trazer uma comparação mais justa entre a escala Braden e modelos de aprendizado de máquina, Cramer *et al.* (2019) usou uma abordagem na qual apenas as informações das primeiras 24 horas de internação são utilizadas para prever a incidência de LP futura. Foram extraídos 40 atributos a partir dos dados disponíveis no banco de dados MIMIC-III que os modelos de aprendizado de máquina utilizaram. Concluiu-se que o melhor modelo foi a regressão logística atingindo precisão de 9% com sensibilidade de 71% no conjunto de teste enquanto a escala Braden teve a mesma precisão com sensibilidade de 50%. O artigo também disponibiliza as importâncias de cada atributo para o melhor modelo, verificando-se que os atributos preditivos coincidiram com o conhecimento da área. Também mostrou-se que o modelo de regressão logística não foi melhorado ao incluir as notas Braden como atributos do modelo.

Uma abordagem de predição diária foi utilizada por Cho e Chung (2011), na qual o treinamento foi feito a partir de cada dia de internação utilizando apenas informações disponíveis no momento da predição. O modelo foi implementado em uma ferramenta,

chamada *PU manager*, e utilizado por 6 meses em um hospital na Coreia do sul (CHO *et al.*, 2013). No período de intervenção observou-se uma redução significativa na prevalência de LPs e no tempo de estadia médio dos pacientes. O autor ainda reflete em explicações alternativas para esse resultado além do desempenho do modelo de predição, pois a intervenção não se limitou apenas na implementação do sistema, houve mudanças no fluxo de trabalho. Além disso, a intervenção não foi testada junto a um grupo de controle, sendo comparada apenas ao período anterior onde não houve intervenção.

Em suma, nota-se que há um aparente consenso na área de que utilizando dados de registros eletrônicos é possível superar o desempenho de predição da escala Braden. Há ainda de se tomar cuidado pois em geral são feitos estudos retrospectivos em hospitais que já utilizam a escala Braden para iniciar cuidados em pacientes identificados em risco. Isso implica que a escala Braden está em desvantagem pois em um caso onde efetivamente preveniu-se uma LP, no estudo retrospectivo a prevenção desse caso será contabilizado como um erro. Além disso, havendo treinado algum modelo a partir desses dados, o modelo irá aprender a classificar tais amostras como negativas. Pode-se dizer no entanto, que a escala Braden, por ser simples, irá identificar apenas os casos mais fáceis que podem ser incluídos no modelo reajustando-o durante sua aplicação prática.

Embora não seja possível quantizar o nível desse enviesamento, o resultado obtido por Cho *et al.* (2013) é de grande importância, pois apesar desse viés ainda obteve-se uma diminuição significativa de incidências de LP motivando a busca por modelos de predição de incidência de LP a partir de registros eletrônicos.

3 DESENVOLVIMENTO

Nesse capítulo serão descritos a formulação do problema, estatísticas importantes dos três conjuntos de dados utilizados na nossa análise, os atributos utilizados e, por fim, o método de análise e treinamento dos modelos.

3.1 Formulações do problema

Nesse estudo, focou-se na abordagem de classificação para o problema de predição de incidência de LP de nível 2 ou maior. Porém, é possível ainda abordar do ponto de vista de regressão, usando como variável alvo o tempo até LP, por exemplo. No entanto, a regressão geralmente exige mais informações para ser possível realizar uma predição acurada. Existem outros problemas como a representação de internações sem nenhuma LP e a avaliação da qualidade dos modelos. Devido à existência desses problemas, focou-se apenas na abordagem de classificação.

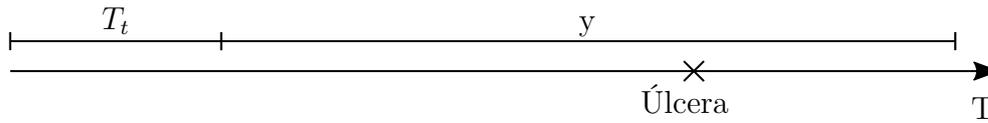
Uma das abordagens na qual foram realizados experimentos iniciais foi a de Cramer *et al.* (2019), representada na Figura 6, na qual a predição é feita apenas uma vez por paciente após 24 horas de internação usando informações disponíveis até esse momento. No entanto, a escala Braden pode ser calculada mais de uma vez por dia para cada paciente. Portanto para uma representação mais justa da aplicação usual da escala Braden, que também seria similar ao uso prático do modelo desenvolvido, propõe-se uma mudança na abordagem onde o modelo realizaria a predição diariamente. Para satisfazer essa modificação foram obtidas informações em uma janela de curto prazo, iniciando 24 horas antes da predição, assim como informações de longo prazo, agregando informações desde o início da internação. Essa modificação na abordagem pode ser vista como um método de *data augmentation*, embora nesse caso sirva para proporcionar uma avaliação do modelo mais próxima à realidade e portanto esse processo é feito no conjunto de teste também.

Uma última modificação na abordagem vem da observação de que, em geral, é mais importante saber quais pacientes estão prestes a desenvolver uma LP. Nesse sentido, propõe-se que em vez de predizer a incidência em qualquer momento, busca-se predizer a incidência apenas nos próximos dias, dessa maneira o modelo possibilitaria direcionar esforços preventivos para pacientes em risco iminente. A nova abordagem com as duas modificações está representada na Figura 7. Nesse trabalho, focou-se na nova abordagem, incluindo as duas modificações comentadas, que ao ser apresentada a profissionais da área foi vista com maior interesse.

3.2 Bases de dados

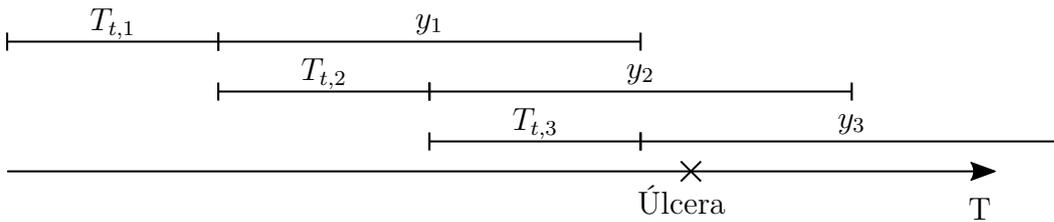
Para a validação dos experimentos foram utilizadas três bases de dados diferentes, a base de dados MIMIC-III, a base de dados MIMIC-IV e a base de dados do HMD. Em

Figura 6 – Abordagem usada por Cramer *et al.* (2019) para um paciente t .



Fonte - do autor

Figura 7 – Representação da abordagem de janelas para uma internação t , o segundo subíndice n de $T_{t,n}$ representa cada janela de curto prazo e cada y_n é em relação à internação t e janela n . Nota-se que, nesse exemplo, $y_1 = 0$ e $y_2 = y_3 = 1$.



Fonte - do autor

todas essas bases de dados estão disponíveis informações de exames, de sinais vitais, de internações, de transferências, demográficas, de remédios administrados, e outros dados. Inicialmente foram obtidos os atributos no conjunto de dados MIMIC-III, em seguida foi feita uma compatibilização do MIMIC-IV e do HMD. A partir dos dados compatibilizados, é feito um mapeamento de atributos e uma extração em tabelas separadas para que a construção da matriz de atributos seja feita.

No Apêndice A está a Tabela 5 na qual estão disponíveis estatísticas (média e desvio padrão) de cada atributo agregados por internações com e sem LP, e a Tabela 6 indicando porcentagem de dados faltantes em cada base. Nota-se em particular a diferença entre as médias da escala Braden nos conjuntos de dados MIMIC-III e MIMIC-IV em relação ao HMD, trazendo evidências de que de fato a aplicação da escala Braden varia significativamente entre hospitais.

3.2.1 MIMIC-III

O MIMIC-III¹ possui registros eletrônicos de 61,532 internações em UTI de 46,476 pacientes distintos que foram hospitalizados durante os anos 2001 e 2012 no Centro Médico Beth Israel em Boston. A base de dados é aberta para todos usuários credenciados na plataforma Physionet, o que inclui completar um curso sobre pesquisa utilizando dados obtidos de pacientes. A base de dados é composta da união de dois sistemas de monitoramento de pacientes, CareVue que foi utilizado entre 2001 e 2008 e MetaVision utilizado de 2008 a 2012.

¹ Link para acesso: <https://mimic.mit.edu/docs/iii/>

O MIMIC-III possui 26 tabelas das quais 5 possuem apenas definições de códigos utilizados em outras tabelas, enquanto as outras possuem dados demográficos, exames de laboratório, procedimentos, sinais vitais, remédios administrados, diagnósticos, transferências, entre outros.

Embora o MIMIC-III possua várias tabelas, a maior parte das informações de sinais vitais e exames está na tabela CHARTEVENTS.csv.gz que, com 330 milhões de linhas e 15 colunas, em formato descomprimido ocupa 46GB. Nela tabela estão os resultados de exames laboratoriais, sinais vitais, escala Braden, etc. Em cada linha da tabela está um registro de alguma dessas grandezas associadas a uma internação.

Cada informação é identificada por um ITEMID único e definido em outra tabela D_ITEMS.csv.gz. Foi feito um agrupamento de ITEMIDs equivalentes a partir de um dicionário, onde cada grupo corresponde a um atributo. Esse agrupamento é utilizado para separar as leituras de cada atributo na tabela CHARTEVENTS.csv.gz em arquivos individuais tornando mais fácil o processamento dos dados para a criação da matriz de atributos. Além disso, são extraídos também, de maneira similar, os dados da tabela LABEVENTS.csv.gz com um agrupamento separado. Existem atributos que estão presentes nas duas tabelas portanto são removidas leituras repetidas.

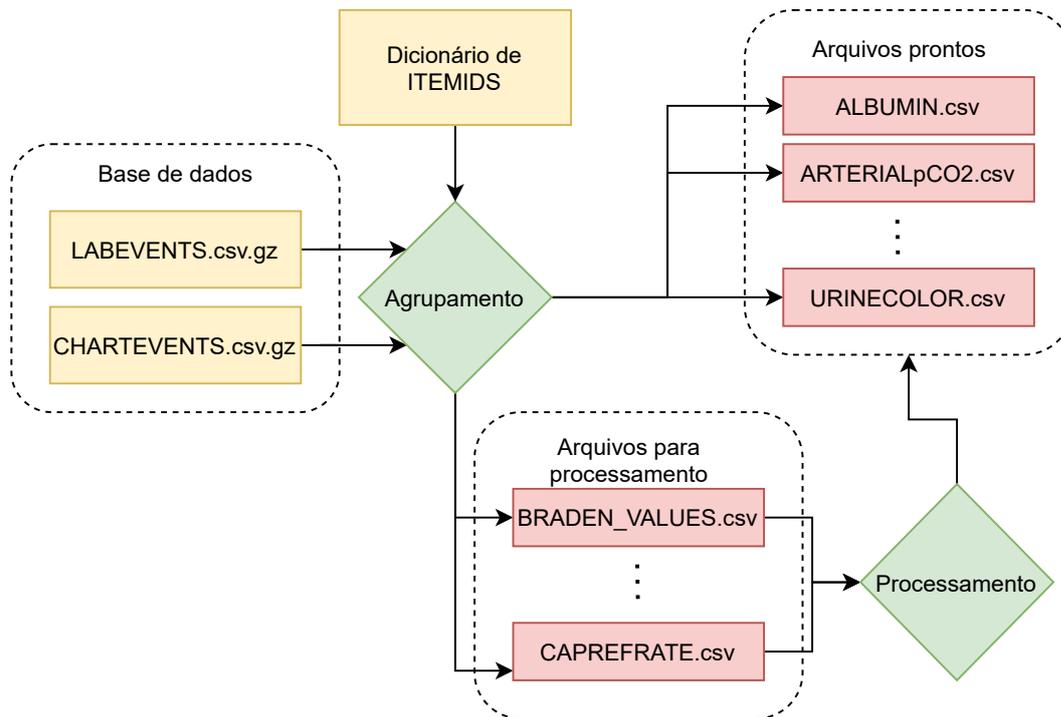
Em algumas das tabelas extraídas é necessário fazer um processamento extra para compatibilizar valores ou para agrupar e separar valores. Por exemplo, todas as notas Braden são extraídas para um único arquivo onde os valores em texto são mapeados e as notas são agrupadas pelo horário que foram obtidas e descartam-se leituras que não possuem uma ou mais notas. Em seguida cada nota é salva em um arquivo separado. No MIMIC-III, devido ao uso de dois sistemas de monitoramento, é necessário mapear para um valor comum os valores das tabelas de alguns atributos que possuem registros em texto, como as notas Braden e registros de nível de LP. Na Figura 8 há um diagrama do processo de extração de dados.

3.2.2 MIMIC-IV

A base de dados MIMIC-IV² é a sucessora da MIMIC-III. Esta base de dados possui um total de 76,540 internações de 53,150 pacientes distintos sendo estas internações ocorridas no período de 2008 a 2019 onde foi utilizado apenas o sistema de monitoramento MetaVision, tornando a base mais consistente e fácil de explorar. Deve-se notar que também há uma interseção temporal com o MIMIC-III nos anos de 2008 a 2012.

A disposição das tabelas mudou ligeiramente, e seguindo a filosofia de reaproveitar o código desenvolvido para o MIMIC-III buscou-se compatibilizar o MIMIC-IV com o formato do seu antecessor, um processo simples devido às numerosas similaridades entre os dois conjuntos de dados. Para isso, foi necessário reorganizar o local de algumas tabelas,

² Link para acesso: <https://mimic.mit.edu/docs/iv/>

Figura 8 – Processo de extração dos dados da tabela *CHARTEVENTS.csv.gz*

Fonte - do autor

recalcular a data de nascimento dos pacientes, fazer um mapeamento de Classificação Internacional de Doenças, 10^a revisão (CID-10) para Classificação Internacional de Doenças, 9^a revisão (CID-9) e reconstruir a tabela de transferências.

3.2.3 HMD

Em colaboração com esse projeto, o HMD forneceu uma base de dados com informações de internações em UTI de dezembro de 2015 a março de 2021, totalizando 10,177 internações de 6,875 pacientes distintos. Deve-se observar que nem todas as informações contidas nas bases MIMIC-III e MIMIC-IV estavam presentes na base do HMD, sendo assim no HMD foram utilizados apenas os atributos que possuíam equivalente aos existentes nas outras bases.

Os dados disponibilizados consistem de 8 tabelas com informações de atendimento, movimentações no hospital, cirurgias, dados vitais, prescrições, exames, escala Braden, e outros itens variados como ventilação mecânica e tabagismo. Novamente buscando reutilizar o código usado no MIMIC-III foi feita uma compatibilização, e dessa vez, devido às grandes diferenças tanto na disposição das tabelas quanto na formatação dos dados, foram necessárias numerosas mudanças.

Inicialmente, devido à ausência de uma tabela que definisse internações em UTI, foi necessário construir uma tabela equivalente à de ICUSTAYS.csv.gz do MIMIC-III a partir da tabela de transferências. Dessa maneira, foi possível extrair todas as internações em

UTI e associar códigos únicos as mesmas. Na tabela de transferências foi necessário criar os campos que indicavam a unidade original e à qual o paciente foi transferido.

Em seguida, buscando recriar a tabela CHARTEVENTS.csv.gz, juntou-se a tabela de internações com a tabela de sinais vitais, sendo removidos sinais vitais que não pertenciam a nenhuma internação em UTI e associando a tabela resultante diretamente às correspondentes internações. Foram selecionados os sinais vitais de interesses gerando uma nova tabela de sinais vitais. Esse processo se repetiu na tabela de exames, que, ao ser acrescentada à tabela de sinais vitais originou o equivalente da tabela CHARTEVENTS.csv.gz para esta base de dados.

Devido à disposição única da tabela com os registros de notas Braden e LP, foi necessário reorganizá-la e, também, associar as leituras com as internações correspondentes. As notas Braden e registros de LP foram extraídos diretamente, não sendo anexados à tabela de sinais vitais. Nota-se que no HMD não está presente a informação de nível de LP, sendo reportadas apenas as de nível 2 ou maior.

Por fim, a partir da tabela de atendimentos, foi possível recriar a tabela PATIENTS.csv.gz e ADMISSIONS.csv.gz. Com todas as tabelas necessárias criadas e compatibilizadas, foi possível utilizar o mesmo código de extração e criação da matriz de atributos.

3.2.4 Extração de atributos

Inicialmente, reproduziu-se o estudo feito por Cramer *et al.* (2019), no qual são extraídos 40 atributos relacionados a informações do paciente, local de internação, sinais vitais, exames, diagnósticos e tratamentos. Boa parte das informações, em particular os sinais vitais, possui múltiplas leituras por internação, sendo necessário sumarizá-las em atributos numéricos. O artigo lida com isso obtendo ou a média ou o último valor registrado.

Um dos objetivos deste trabalho é expandir a quantidade de informações extraídas do conjunto de dados, e para isso foi expandida a quantidade de atributos de 40 para 83. Também buscou-se extrair mais informações de atributos que possuem muitas leituras, por exemplo frequência cardíaca. Isso foi feito obtendo múltiplas estatísticas de cada atributo sendo geradas a partir de uma heurística que analisa a quantidade de leituras nas internações do MIMIC-III:

1. Atributos com 1 leitura ou menos em 50% das internações: última leitura.
2. Atributos que possuem entre 2 a 8 leituras em 50% das internações: última leitura, valor médio e tendência³.
3. Atributos que possuem mais de 8 leituras em 50% das internações: última leitura, valor médio, tendência, valor máximo, valor mínimo, desvio padrão e quantidade de

³ Calculada usando o coeficiente da regressão linear nos dados

leituras.

A partir disso gera-se uma lista de atributos de predição que foram utilizados nos outros conjuntos de dados. Cada atributo é colocado em uma de 3 categorias, compostas por valores numéricos, boolianos e categóricos. Essa diferenciação é importante pois cada categoria necessita de um pré-processamento específico antes de realizar o treinamento dos modelos.

Por fim, nota-se que na abordagem do artigo, os dados são obtidos nas primeiras 24 horas de internação. Na nova abordagem, com predição diária, esse processo pode ser interpretado de duas maneiras: obter as informações apenas nas 24 horas antes da predição, ou obter as informações desde o início da internação. Decidiu-se fornecer ambas as informações para o modelo, ou seja, para cada estatística de um atributo existem duas versões, uma com informações apenas nas últimas 24 horas e outra com informações desde o início da internação. Por exemplo, a frequência cardíaca terá, entre outras, uma estatística de valor médio nas últimas 24 horas e outra estatística de valor médio desde o início da internação. Desse modo, o modelo possui informações do passado assim como informações recentes. Para comparar com as notas Braden, decidiu-se utilizar o valor mais recente das notas Braden, estando dentro ou fora da janela de 24 horas.

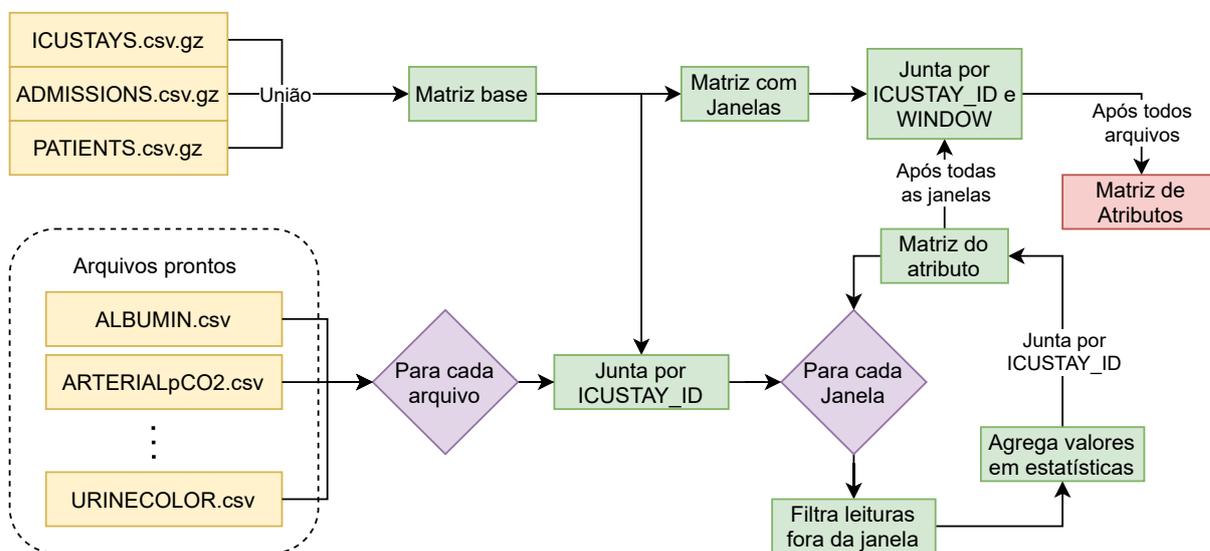
3.2.5 Criação da matriz de atributos e variável alvo

A partir da junção das tabelas com informações das internações em UTI, atendimentos e pacientes, cria-se uma matriz base com informações de cada internação. Usando a matriz base, pode-se gerar janelas diárias para cada internação repetindo-as pela quantidade de dias de internação. Assim obtém-se uma matriz com todas as janelas onde serão feitas as predições e que será usada para agregar os atributos presentes em cada janela.

Para cada atributo, deseja-se obter suas estatísticas para cada janela de cada internação. Esse processo é feito unindo os dados com a matriz base e filtrando as linhas de leituras que estão fora da janela. Para os atributos da estadia inteira, são removidos apenas os dados futuros. Em seguida, agregam-se os valores em estatísticas para cada internação resultando em uma tabela com uma linha por internação e uma coluna para cada estatística. Essa rotina se repete para cada janela, onde o resultado final é uma tabela com uma linha para cada janela de cada internação e uma coluna para cada estatística nas janelas. Essa tabela é anexada à tabela base com janelas e o processo se repete para cada atributo até ser obtida a matriz de atributos completa. Esse processo é representado pelo diagrama na Figura 9.

Para a construção da variável alvo (*target*) obtém-se o primeiro registro de uma LP de nível 2 ou mais para cada internação que possui LP. As duas classificações LTP e inclassificáveis são consideradas LP de nível 1. Para obter-se a variável alvo basta verificar se o momento de predição está 7 dias antes ou mais até o primeiro registro de LP da

Figura 9 – Diagrama representando o processo de criação da matriz de atributos, desde a criação da matriz base, a criação de matrizes das estatísticas dos atributos em janelas até obter a matriz de atributos completa.



Fonte - do autor

internação. Se o momento de predição é após o registro da primeira LP, remove-se a janela pois não há razão para realizar a predição de incidência de LP em um paciente que já foi diagnosticado com ela.

3.3 Métodos

Nessa seção serão descritos os processos de filtragem de dados, pré-processamento de atributos, treinamento e avaliação de modelos.

3.3.1 Filtragem dos dados

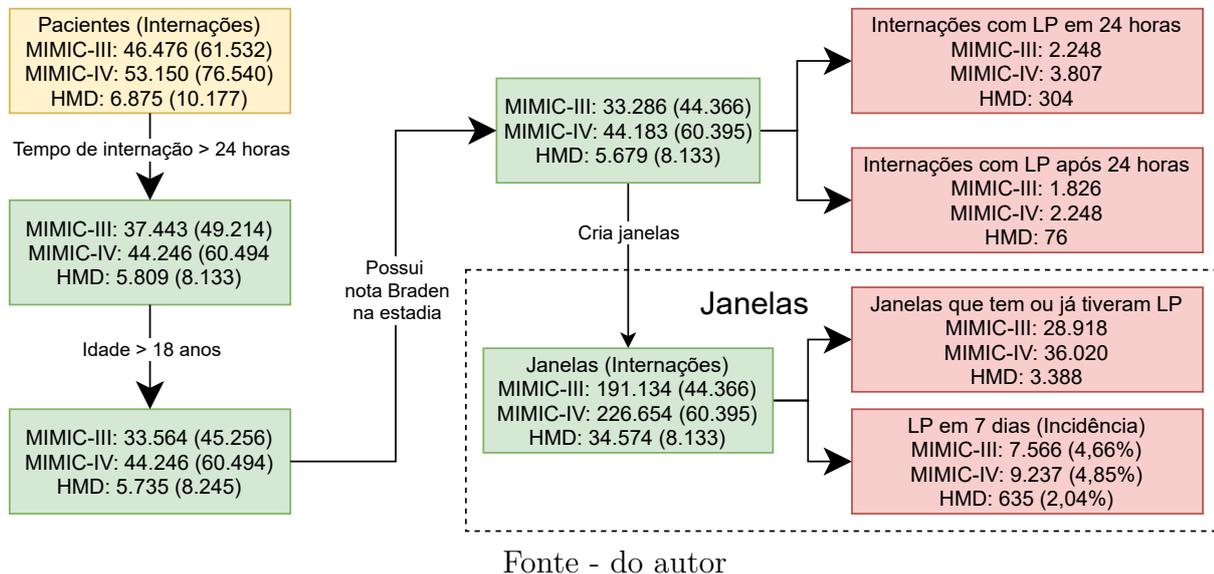
Antes de utilizar os conjuntos de dados, estes são processados de acordo com a Figura 10. Os pacientes menores de 18 anos são removidos pois possuem pequeno risco de desenvolver LP. Para comparar com a escala Braden, removem-se as janelas que ainda não possuem nenhum registro da escala Braden. E finalmente, removem-se também as janelas que já tiveram alguma notificação de LP nível 2 ou mais.

3.3.2 Pré-processamento

Devido às limitações do conjunto de dados e alguns modelos, é necessário realizar um pré-processamento consistindo de imputação, escalonamento e codificação *one-hot* de atributos categóricos.

O processo de imputação consiste em preencher dados faltantes nas janelas. Para atributos numéricos é feita a imputação a partir da mediana dos dados existentes em

Figura 10 – Diagrama das filtragens realizadas e procedimento de criação de janelas junto com figuras notáveis de cada caso. As janelas que possuem ou já tiveram incidência de lesão por pressão são removidas.



outras janelas. Os atributos booleanos são construídos para que ao serem preenchidos com *False* introduzam a menor quantidade de informação. Por exemplo, caso não haja nenhuma confirmação de que um paciente esteja usando ou não um aparelho de redução de pressão, assume-se que não. Para atributos categóricos, preenche-se com *None*, uma categoria representando a falta do dado.

Os atributos numéricos também serão escalonados para reduzir sua variância e remover a média. Isso é feito pois modelos como a regressão logística convergem melhor se todos os atributos possuem variância unitária e média nula. Foi utilizado um escalonamento robusto a partir do primeiro e terceiro quartil, para minimizar o impacto de *outliers*.

Esse pré-processamento faz, efetivamente, parte do modelo em si devido à existência de parâmetros treináveis como o cálculo dos quartis para o escalonamento robusto e a mediana para a imputação. Por essa razão, o pré-processamento é acoplado ao modelo de modo que, ao treinar o modelo treina-se o pré-processador junto, e durante a avaliação do modelo utilizam-se os parâmetros obtidos no treinamento. Dessa forma não há vazamento de informação do conjunto de teste para o de treinamento.

3.3.3 Metodologia de treinamento e avaliação

A avaliação do modelo é feita de duas maneiras, uma usando um conjunto de teste *holdout* e outra usando uma amostragem com repetição de conjuntos treino e teste. A divisão de todos os conjuntos treino e teste é feita de maneira estratificada nas internações, ou seja uma internação é considerada da classe 0 se em nenhum momento da internação foi reportada LP nível 2 ou maior, e da classe 1 se em algum momento foi reportada LP nível 2

ou maior. A partir disso, o conjunto de dados é dividido entre um conjunto de treinamento e teste, onde o conjunto de teste possui todas as janelas de 20% das interações de cada classe e o conjunto de treinamento possui os 80% restantes.

Para o treinamento dos modelos, faz-se uma separação de um conjunto de treino e teste *holdout* de maneira estratificada, onde no conjunto de treinamento é feita a busca de hiperparâmetros dos modelos, a partir da validação cruzada com *5-folds*, testando 30 realizações aleatórias de hiperparâmetros. São selecionados os hiperparâmetros do modelo que obteve o maior valor da métrica utilizada, precisão fixando a sensibilidade em 50%, e obtém-se os valores de outras métricas no conjunto de teste. No apêndice B há a Tabela 7 com os intervalos de hiperparâmetros definidos para cada modelo.

Em seguida, fixando os hiperparâmetros encontrados, são feitas outras 50 reamostragens de conjuntos treino e teste. Em cada uma dessas reamostragens é feito o treinamento do modelo e são registradas as variáveis alvo e as predições do modelo. Após obter todas as predições em cada conjunto, é feita uma varredura de todos os limiares calculando as métricas de cada conjunto e agregando-os pela média, permitindo a geração de curvas ROC e precisão por sensibilidade.

Como as amostras usadas na otimização dos hiperparâmetros são reutilizadas nos conjuntos de treinamento reamostrados, é esperado que haja um pequeno vazamento e portanto um aumento nos desempenhos obtidos. Por isso, informam-se também as métricas no primeiro conjunto de teste. No entanto, esse processo de reamostragem é útil para obtenção de curvas mais suaves.

Por fim, de modo a ter uma comparação justa, deve-se ressaltar que essas divisões de treinamento e teste são fixas para modelos diferentes.

4 RESULTADOS

Nesse capítulo serão apresentados os resultados obtidos em cada base de dados. Foram realizados o treinamento de 6 modelos diferentes, em 3 bases de dados e em dois conjuntos de atributos onde o primeiro conjunto se assemelha aos atributos do artigo de Cramer *et al.* (2019), referido por “Original”, e o segundo conjunto inclui todos os novos atributos e estatísticas extraídas referido por “Estendido”. Os principais resultados são mostrados nas Tabelas (2) até (4) nas quais são apresentados os valores da precisão com sensibilidade em 50% para cada modelo. Para fim de comparação são disponibilizados gráficos da curva precisão por sensibilidade e curva ROC para os dois melhores modelos em termos da métrica no conjunto de teste, para a escala Braden e escala Braden reponderada pela regressão logística.

Na Figura 11 e Tabela 2 estão os resultados no conjunto de dados do MIMIC-III. Nota-se que o melhor modelo no 1º conjunto foi o *Gradient Boosting* em ambos os atributos original e estendido. O modelo obteve o melhor desempenho com o conjunto de atributos original, o que indica que nesse conjunto de dados os atributos estendidos introduzem ruído significativo. Por fim, nota-se que mesmo realizando a reponderação da escala Braden seu desempenho é significativamente inferior a todos os modelos.

Os resultados no MIMIC-IV, presentes na Figura 12 e na Tabela 3, são similares ao MIMIC-III. No entanto, agora o melhor modelo foi o *Gradient Boosting* com os atributos novos. Além disso, nota-se que o desempenho da escala Braden está mais próximo do desempenho dos modelos de aprendizado de máquina, obtendo inclusive desempenho similar ao do modelo de árvore de decisão.

Por fim, no HMD (Figura 13, Tabela 4) o melhor modelo foi a floresta aleatória obtendo uma precisão quase duas vezes maior que a escala Braden. Nota-se que o valor numérico da precisão de todos os modelos no HMD é inferior aos valores vistos tanto no MIMIC-III quanto no MIMIC-IV. No entanto, não deve ser feita uma comparação direta devido ao fato de que a incidência de LP no HMD é mais de duas vezes menor. Uma observação a ser feita é que no HMD há uma quantidade reduzida de atributos, como pode ser verificado na Tabela 5 no Apêndice A, o que indica uma possibilidade de melhoria no desempenho dos modelos se houvessem mais dados.

4.1 Discussão

Em todos os conjuntos de dados, os modelos de aprendizado de máquina superam significativamente tanto o desempenho da escala Braden quanto a escala Braden reponderada a partir da regressão logística. Essa diferença aumenta ainda mais para valores menores de sensibilidade, indicando que, em geral, o modelo treinado em registros eletrônicos realiza um trabalho melhor de discriminar pacientes em risco.

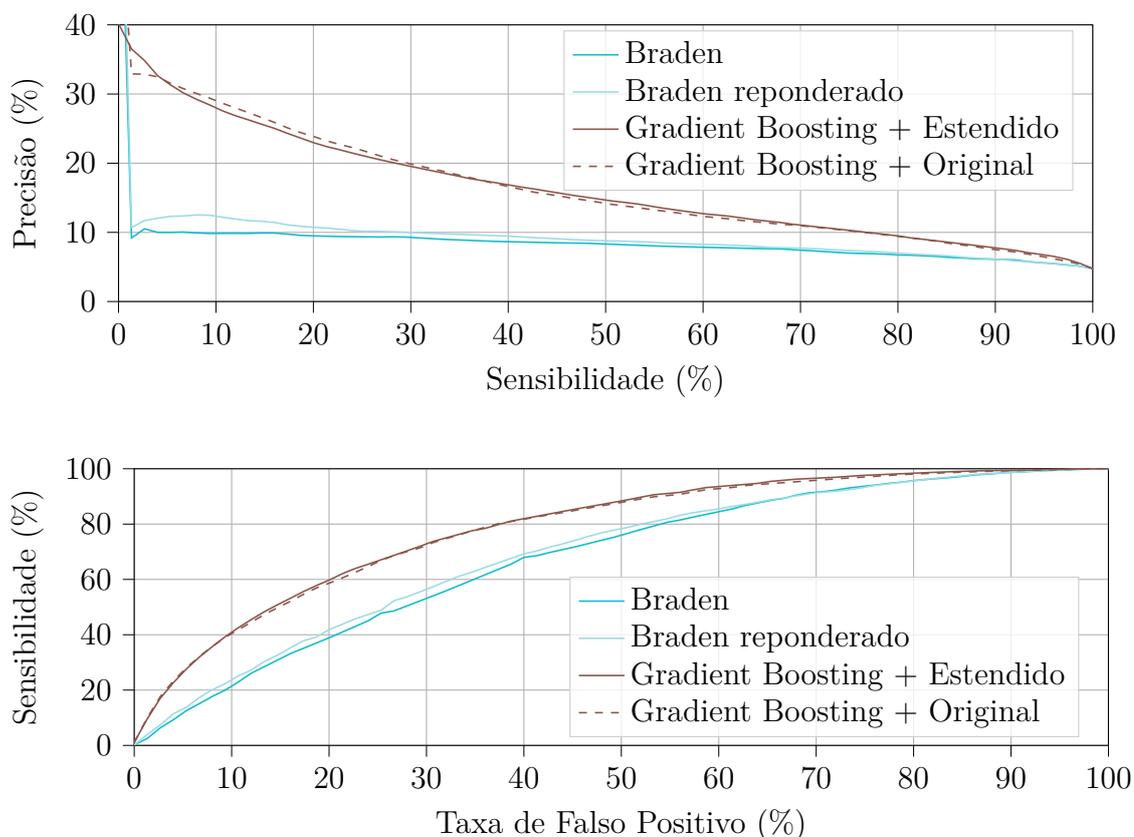
Esses resultados eram esperados, devido à grande quantidade de informações disponíveis para os modelos de aprendizado de máquina. Em particular, deve-se ressaltar que nem sempre a escala Braden utilizada na sua avaliação foi atribuída na janela das 24 horas anteriores à predição. Essa ocorrência é mais prevalente no HMD, em que a frequência de atribuição da nota Braden é significativamente menor que nas bases MIMIC-III e MIMIC-IV, que possuem frequências em média de 4 em 4 horas. Essa é uma das vantagens que os modelos que utilizam dados eletrônicos possuem em relação à escala Braden, já que os dados utilizados estão disponíveis e são coletados automaticamente.

Outra vantagem no uso desses modelos é que, como pode ser visto na Tabela 5 no Apêndice A, a diferença entre as notas médias no HMD e MIMIC é significante e, assumindo não haver uma diferença tão grande na população de pacientes de UTI entre as bases, indica uma possível inconsistência na atribuição da nota Braden entre os dois hospitais. Enquanto isso, muitos dos dados utilizados pelos modelos são similares entre as bases de dados, o que possibilitaria a criação de um modelo que poderia ser utilizado em múltiplos hospitais.

Deve-se notar também que as curvas apresentadas são as curvas obtidas nos conjuntos reamostrados. Era esperado que os resultados nos conjuntos reamostrados fossem levemente superiores ao primeiro conjunto de teste, no entanto essa variação foi maior que a esperada, variando de maneira diferente entre os modelos. Os modelos mais complexos, com mais atributos, apresentam uma diferença maior entre as métricas obtidas indicando haver um maior sobreajuste. Por exemplo, no MIMIC-III o melhor modelo no conjunto de teste não possui o melhor desempenho nos conjuntos reamostrados, possivelmente por ser um modelo com menos atributos. Portanto, o resultado mais confiável, como discutido no capítulo anterior, é o resultado no primeiro conjunto de teste. As curvas, embora não sirvam para comparar entre os modelos de aprendizado de máquina, servem para mostrar a diferença entre o desempenho da escala Braden e os modelos em pontos de operação diferentes.

Como discutido na Seção 2.1, é necessário tomar cuidado ao tirar conclusões desses resultados. Em particular, deve-se notar que a escala Braden é utilizada para identificação de pacientes em risco em todos os hospitais dos quais os dados originaram. Isso implica que, por exemplo, um paciente que iria desenvolver uma LP ao ser identificado como em risco corretamente pela escala Braden recebeu um tratamento especial e não desenvolveu a LP. Ou seja, no período de aquisição de dados a escala Braden trabalhou para a sua piora nessa análise. Ainda, os modelos de aprendizado de máquina irão reproduzir esse viés presente nos dados, necessitando haver cuidado na sua implementação prática ao substituir a escala Braden.

Figura 11 – Sumarização dos pontos de operação para os melhores modelos no MIMIC-III.



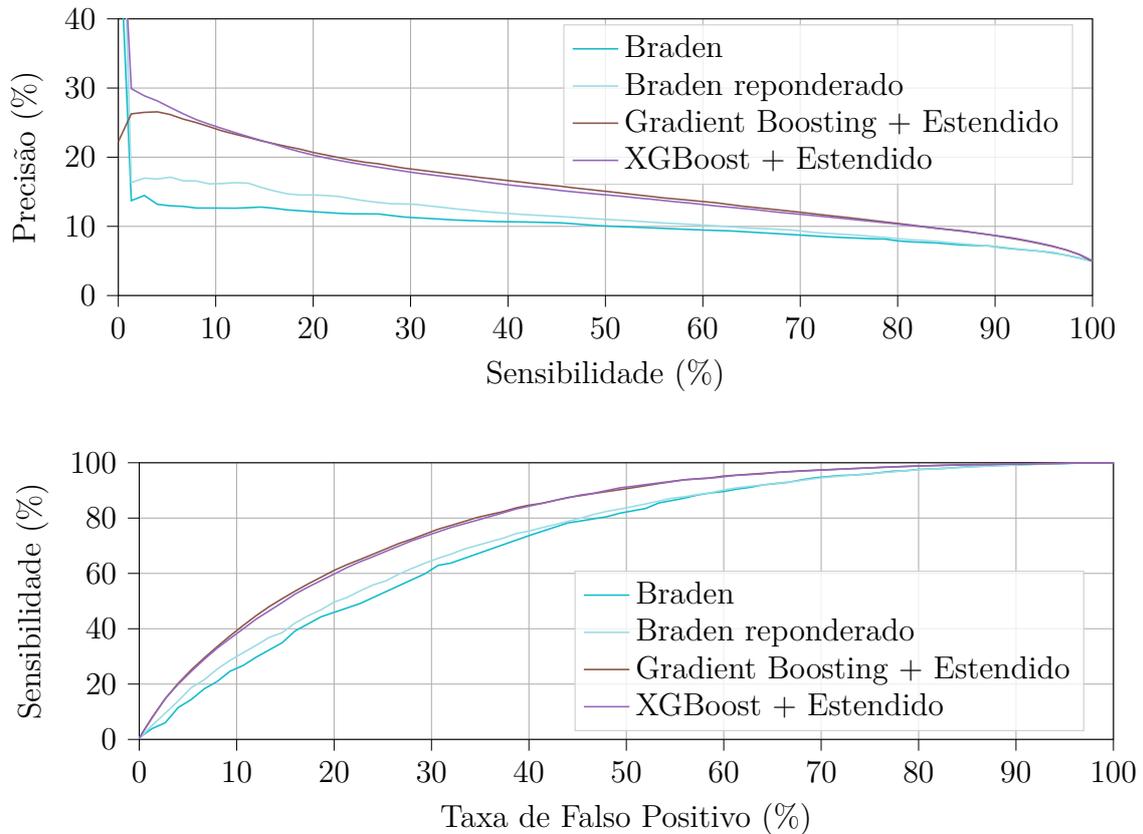
Fonte - do autor

Tabela 2 – Valores da precisão para sensibilidade em 50% obtidas no 1º conjunto e valor médio nas reamostragens de treino/teste no MIMIC-III

Modelos	Validação Cruzada	Teste	Reamostragens
Braden	8.31% (0.628%)	8.626%	8.353%
Braden reponderado	8.74% (0.547%)	9.025%	8.853%
Regressão Logística + Original	13.3% (0.924%)	13.44%	13.48%
Regressão Logística + Estendido	11.7% (1.37%)	10.74%	11.13%
Árvore de Decisão + Original	10.6% (0.421%)	10.27%	10.22%
Árvore de Decisão + Estendido	10.8% (0.486%)	10.39%	10.25%
Floresta Aleatória + Original	13.3% (1.24%)	12.21%	13.03%
Floresta Aleatória + Estendido	13.3% (0.862%)	11.89%	13.07%
<i>Gradient Boosting</i> + Original	14.0% (0.500%)	13.96%	14.29%
<i>Gradient Boosting</i> + Estendido	14.7% (0.355%)	13.77%	14.73%
<i>XGBoost</i> + Original	13.9% (0.661%)	13.51%	13.99%
<i>XGBoost</i> + Estendido	14.5% (0.794%)	13.08%	14.41%
<i>CatBoost</i> + Original	14.2% (0.782%)	13.22%	14.06%
<i>CatBoost</i> + Estendido	15.0% (1.27%)	13.53%	14.41%

Fonte - do autor

Figura 12 – Sumarização dos pontos de operação para os melhores modelos no MIMIC-IV.



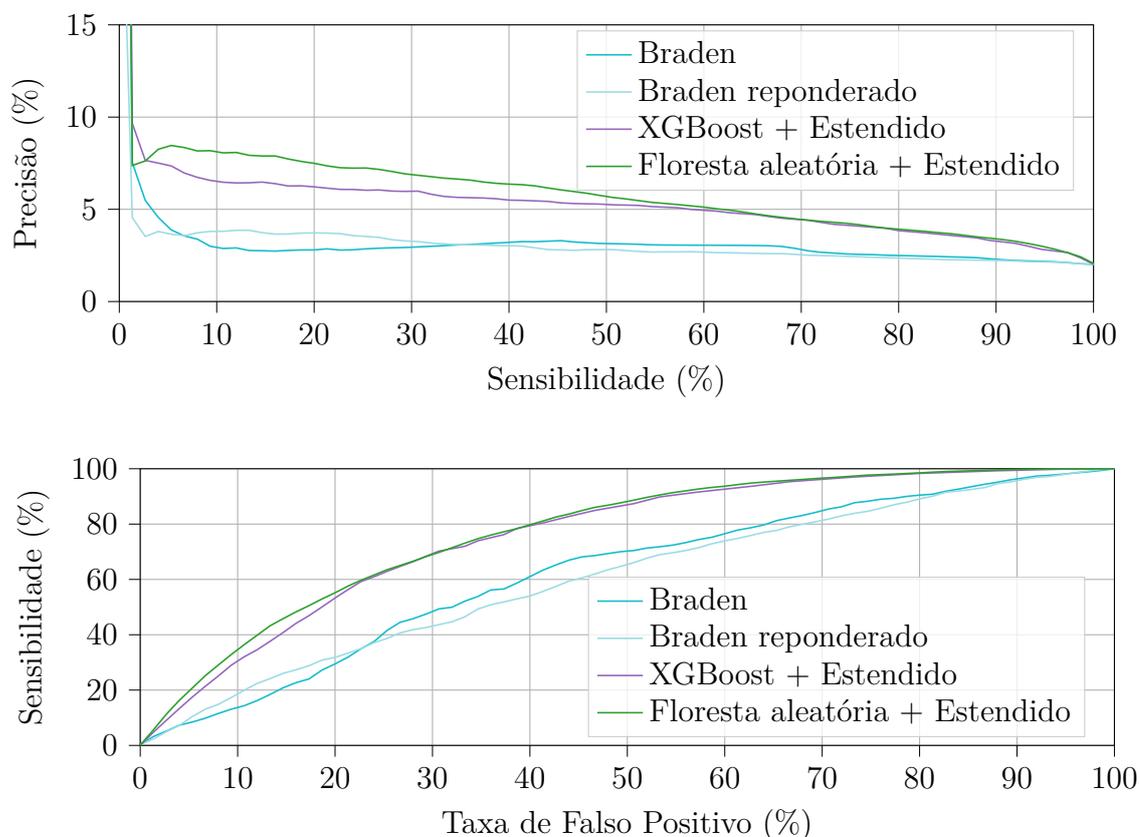
Fonte - do autor

Tabela 3 – Valores da precisão para sensibilidade em 50% obtidas no 1º conjunto e valor médio nas reamostragens de treino/teste no MIMIC-IV

Modelos	Validação Cruzada	Teste	Reamostragens
Braden	10.2% (0.780%)	10.2%	10.2%
Braden reponderado	11.1% (0.840%)	11.04%	11.03%
Regressão Logística + Original	13.8% (0.178%)	12.73%	13.71%
Regressão Logística + Estendido	7.47% (1.03%)	5.738%	6.814%
Árvore de Decisão + Original	11.5% (1.05%)	11.11%	11.64%
Árvore de Decisão + Estendido	11.6% (0.980%)	11.03%	11.47%
Floresta Aleatória + Original	14.3% (1.38%)	12.4%	13.89%
Floresta Aleatória + Estendido	14.3% (1.67%)	12.56%	13.82%
<i>Gradient Boosting</i> + Original	14.9% (1.12%)	13.04%	14.44%
<i>Gradient Boosting</i> + Estendido	15.3% (1.25%)	13.8%	15.08%
<i>XGBoost</i> + Original	14.8% (1.04%)	12.58%	14.16%
<i>XGBoost</i> + Estendido	15.1% (1.41%)	13.47%	14.6%
<i>CatBoost</i> + Original	14.7% (1.12%)	12.79%	14.44%
<i>CatBoost</i> + Estendido	15.3% (1.57%)	13.1%	14.78%

Fonte - do autor

Figura 13 – Sumarização dos pontos de operação para os melhores modelos no HMD.



Fonte - do autor

Tabela 4 – Valores da precisão, em porcentagem, para sensibilidade em 50% obtidas no 1º conjunto e valor médio nas reamostragens de treino/teste no HMD

Modelos	Validação Cruzada	Teste	Reamostragens
Braden	3.59% (1.41%)	2.961%	3.254%
Braden reponderado	4.10% (1.84%)	3.172%	3.129%
Regressão Logística + Original	4.68% (1.45%)	3.691%	4.229%
Regressão Logística + Estendido	5.14% (1.28%)	4.145%	4.782%
Árvore de Decisão + Original	4.31% (1.77%)	2.823%	3.284%
Árvore de Decisão + Estendido	5.12% (0.833%)	3.763%	3.889%
Floresta Aleatória + Original	4.42% (1.09%)	4.393%	5.082%
Floresta Aleatória + Estendido	5.80% (1.04%)	6.099%	6.31%
<i>Gradient Boosting</i> + Original	4.42% (1.52%)	4.716%	4.852%
<i>Gradient Boosting</i> + Estendido	5.77% (1.93%)	4.949%	5.355%
<i>XGBoost</i> + Original	4.24% (1.52%)	4.026%	4.444%
<i>XGBoost</i> + Estendido	5.75% (1.10%)	5.893%	5.543%
<i>CatBoost</i> + Original	4.60% (1.35%)	5.85%	5.131%
<i>CatBoost</i> + Estendido	6.50% (2.77%)	5.775%	5.632%

Fonte - do autor

5 CONCLUSÃO

Neste trabalho, foi realizado um estudo transversal analisando o uso de modelos de aprendizado de máquina para a predição de incidência de LP a partir de registros eletrônicos, que foram avaliados em três conjuntos de dados obtidos de dois hospitais diferentes. Ao todo foram treinados 36 modelos diferentes, referentes a 6 implementações de modelos de aprendizado de máquina, 3 conjuntos de dados e 2 conjuntos de atributos, um reproduzindo Cramer *et al.* (2019) com 40 atributos, e outro com a adição de mais 43 atributos. Todos os objetivos apresentados no início do trabalho foram alcançados com sucesso.

Os resultados indicam que os modelos superam significativamente a escala Braden, que é a ferramenta de uso comum na prática. Isso motiva o desenvolvimento de uma ferramenta automática para a identificação de pacientes em risco de desenvolverem LP. O uso dessa ferramenta poderia eliminar a mão de obra necessária para a atribuição manual da escala Braden, provendo inclusive uma redução de até 30% no custo de tratamento preventivo para um mesmo nível de qualidade de atendimento.

Por ser uma área com poucos trabalhos publicados, ainda existe muita pesquisa a ser realizada. Podem ser feitas melhorias no modelo, desde a extração de mais dados até a criação de novos atributos a partir dos já existentes. Trabalhos em conjunto com profissionais da área seriam de grande proveito na identificação de informações críticas para o modelo.

Também é de grande interesse a validação prática através de um estudo de intervenção como feito por Cho *et al.* (2013) ou ainda um estudo duplo cego randomizado comparando a escala Braden com o modelo. Esse trabalho permitiria confirmar se o uso do modelo traria benefícios de fato.

Outro estudo de interesse prático seria investigar a possibilidade de transferência de aprendizado entre conjuntos de dados. Assim seria possível, por exemplo, a partir de um modelo treinado em uma base de dados grande, reajustá-lo em uma base de dados menor permitindo o uso dessa ferramenta em hospitais menores.

Fora do objetivo desse trabalho, é possível a partir dos dados extraídos criar e testar novos modelos para predição de outras condições prevalentes em UTIs, como a predição de mortalidade, de problemas nutricionais, nível de consciência, etc. Esses possíveis trabalhos futuros são facilitados devido à compatibilização dos três conjuntos de dados utilizados nesse trabalho.

REFERÊNCIAS

- ALDERDEN, Jenny; RONDINELLI, June; PEPPER, Ginette; CUMMINS, Mollie; WHITNEY, JoAnne. RISK FACTORS FOR PRESSURE INJURIES AMONG CRITICAL-CARE PATIENTS: A SYSTEMATIC REVIEW. **International journal of nursing studies**, v. 71, p. 97–114, jun. 2017. ISSN 0020-7489. DOI: 10.1016/j.ijnurstu.2017.03.012. Disponível em: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5485873/>. Acesso em: 10 jul. 2021.
- BERGSTROM, Nancy *et al.* The Braden Scale for predicting pressure sore risk. **Nurs res**, v. 36, n. 4, p. 205–210, 1987.
- BREIMAN, Leo. Random forests. **Machine learning**, Springer, v. 45, n. 1, p. 5–32, 2001.
- BREIMAN, Leo; FRIEDMAN, Jerome H; OLSHEN, Richard A; STONE, Charles J. **Classification and regression trees**. [*S.l.*]: Routledge, 1984.
- CHEN, Tianqi; GUESTRIN, Carlos. XGBoost: A Scalable Tree Boosting System. *In*: PROCEEDINGS of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. San Francisco, California, USA: ACM, 2016. (KDD '16), p. 785–794. ISBN 978-1-4503-4232-2. DOI: 10.1145/2939672.2939785. Disponível em: <http://doi.acm.org/10.1145/2939672.2939785>.
- CHO, In Sook; CHUNG, Eunja. Predictive Bayesian Network Model Using Electronic Patient Records for Prevention of Hospital-Acquired Pressure Ulcers. **Journal of Korean Academy of Nursing**, v. 41, n. 3, p. 423–431, 13 jun. 2011. Publisher: Korean Society of Nursing Science. DOI: 10.4040/jkan.2011.41.3.423. Disponível em: <https://synapse.koreamed.org/articles/1002724>. Acesso em: 29 ago. 2021.
- CHO, In Sook; PARK, Ihnsook; KIM, Eunman; LEE, Eunjoon; BATES, David W. Using EHR data to predict hospital-acquired pressure ulcers: A prospective study of a Bayesian Network model. **International Journal of Medical Informatics**, v. 82, n. 11, p. 1059–1067, 1 nov. 2013. ISSN 1386-5056. DOI: 10.1016/j.ijmedinf.2013.06.012. Disponível em: <https://www.sciencedirect.com/science/article/pii/S1386505613001482>. Acesso em: 16 jun. 2021.

CRAMER, Eric M.; SENEVIRATNE, Martin G.; SHARIFI, Husham; OZTURK, Alp; HERNANDEZ-BOUSSARD, Tina. Predicting the Incidence of Pressure Ulcers in the Intensive Care Unit Using Machine Learning. **eGEMs**, v. 7, n. 1, p. 49, set. 2019. ISSN 2327-9214. DOI: 10.5334/egems.307. Disponível em:

<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6729106/>. Acesso em: 10 jul. 2021.

EDSBERG, Laura E.; BLACK, Joyce M.; GOLDBERG, Margaret; MCNICHOL, Laurie; MOORE, Lynn; SIEGGREEN, Mary. Revised National Pressure Ulcer Advisory Panel Pressure Injury Staging System: Revised Pressure Injury Staging System. **Journal of Wound, Ostomy & Continence Nursing**, v. 43, n. 6, p. 585–597, nov. 2016. ISSN 1071-5754. DOI: 10.1097/WON.0000000000000281. Disponível em:

<https://journals.lww.com/00152192-201611000-00003>. Acesso em: 10 jul. 2021.

FRIEDMAN, Jerome H. Greedy function approximation: a gradient boosting machine. **Annals of statistics**, JSTOR, p. 1189–1232, 1999.

HYUN, Sookyung; VERMILLION, Brenda; NEWTON, Cheryl; FALL, Monica; LI, Xiaobai; KAEWPRAG, Pacharmon; MOFFATT-BRUCE, Susan; LENZ, Elizabeth R. Predictive Validity of the Braden Scale for Patients in Intensive Care Units. **American Journal of Critical Care**, v. 22, n. 6, p. 514–520, 1 nov. 2013. ISSN 1062-3264. DOI: 10.4037/ajcc2013991. Disponível em: <https://doi.org/10.4037/ajcc2013991>. Acesso em: 30 jul. 2021.

JOHNSON, Alistair E W *et al.* MIMIC-III, a freely accessible critical care database. **Scientific data**, Nature Publishing Group, v. 3, n. 1, p. 1–9, 2016.

KAEWPRAG, Pacharmon; NEWTON, Cheryl; VERMILLION, Brenda; HYUN, Sookyung; HUANG, Kun; MACHIRAJU, Raghu. Predictive Modeling for Pressure Ulcers from Intensive Care Unit Electronic Health Records. **AMIA Summits on Translational Science Proceedings**, v. 2015, p. 82–86, 25 mar. 2015. ISSN 2153-4063. Disponível em: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4525237/>. Acesso em: 30 jul. 2021.

KAEWPRAG, Pacharmon; NEWTON, Cheryl; VERMILLION, Brenda; HYUN, Sookyung; HUANG, Kun; MACHIRAJU, Raghu. Predictive models for pressure ulcers from intensive care unit electronic health records using Bayesian networks. **BMC Medical Informatics and Decision Making**, v. 17, 5 jul. 2017. DOI: 10.1186/s12911-017-0471-z.

MURPHY, Kevin P. **Machine learning: a probabilistic perspective**. [S.l.]: MIT press, 2012.

PARANHOS, Wana Yeda; SANTOS, VLCG. Avaliação de risco para úlceras de pressão por meio da escala de Braden, na língua portuguesa, 1999.

PEDREGOSA, F. *et al.* Scikit-learn: Machine Learning in Python. **Journal of Machine Learning Research**, v. 12, p. 2825–2830, 2011.

PROKHORENKOVA, Liudmila; GUSEV, Gleb; VOROBEB, Aleksandr; DOROGUSH, Anna Veronika; GULIN, Andrey. CatBoost: unbiased boosting with categorical features. **arXiv:1706.09516** [cs], 20 jan. 2019. arXiv: 1706.09516. Disponível em: <http://arxiv.org/abs/1706.09516>. Acesso em: 8 ago. 2021.

PROQUALIS. **Protocolo para prevenção de úlcera por pressão | Proqualis**. Disponível em: <https://proqualis.net/protocolo/protocolo-para-preven%5C%C3%5C%A7%5C%C3%5C%A3o-de-%5C%C3%5C%BA1cera-por-press%5C%C3%5C%A3o>. Acesso em: 7 ago. 2021.

SERPA, Letícia Faria; SANTOS, Vera Lúcia Conceição de Gouveia; CAMPANILI, Ticiane Carolina Gonçalves Faustino; QUEIROZ, Moelisa. Validade preditiva da escala de Braden para o risco de desenvolvimento de úlcera por pressão, em pacientes críticos. **Revista Latino-Americana de Enfermagem**, v. 19, p. 50–57, fev. 2011a. Publisher: Escola de Enfermagem de Ribeirão Preto / Universidade de São Paulo. ISSN 1518-8345. DOI: 10.1590/S0104-11692011000100008. Disponível em: <http://www.scielo.br/j/rlae/a/pvfjgRw3q844YGt4LHMqNpQ/abstract/?lang=pt>. Acesso em: 4 set. 2021.

SERPA, Letícia Faria; SANTOS, Vera Lúcia Conceição de Gouveia; PERES, Giovana R. P.; CAVICCHIOLI, Maria G. S.; HERMIDA, Mirta M. Validity of the Braden and Waterlow subscales in predicting pressure ulcer risk in hospitalized patients. **Applied Nursing Research**, v. 24, n. 4, e23–e28, 1 nov. 2011b. ISSN 0897-1897. DOI: 10.1016/j.apnr.2010.05.002. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0897189710000388>. Acesso em: 4 set. 2021.

Apêndices

APÊNDICE A – TABELAS DE ESTATÍSTICAS

Tabela 5 – Tabela com estatísticas dos atributos disponíveis em cada conjunto de dados agregados por internações com e sem lesão por pressão. São apresentados os valores de média e desvio padrão em parênteses

Conjunto de dados Variável	MIMIC-III		MIMIC-IV		HMD	
	Sem LP	Com LP	Sem LP	Com LP	Sem LP	Com LP
Idade	64 (17)	66.3 (16)	65.2 (17)	67 (15)	64.9 (18)	69.1 (15)
Tempo até UTI	1.83 (5.7)	4.09 (11)	1.97 (6)	4.63 (12)	3.74 (15)	16.5 (36)
Gênero [% Feminino]	43.7 (50)	39.2 (49)	43.7 (50)	39.5 (49)	45.9 (50)	36 (48)
Lesão em Internação Prévia [%]	2.12 (14)	7.86 (27)	2.6 (16)	9.38 (29)	0.754 (8.7)	6.4 (25)
Redutor de Pressão [%]	2.26 (15)	8.42 (28)	22.1 (41)	30.8 (46)	-	-
Possui LP Nível 1 [%]	2.88 (17)	13.8 (34)	4.04 (20)	15 (36)	-	-
Albumina	2.89 (0.64)	2.62 (0.6)	2.99 (0.63)	2.73 (0.61)	2.33 (0.73)	1.72 (0.57)
Pressão Arterial de CO2	41.6 (9.6)	42.1 (9.9)	41.3 (9.6)	42 (10)	39.4 (10)	40.9 (10)
Pressão Arterial de O2	137 (80)	119 (55)	143 (87)	117 (58)	107 (38)	112 (36)
Saturação de O2	97.2 (2.7)	97.1 (3.1)	96.7 (2.8)	97.2 (2.8)	96.1 (3)	96.9 (2.9)
Bilirrubina	3.67 (6.5)	5.36 (8.6)	3.19 (5.9)	5.23 (8.3)	1.16 (2.2)	1.08 (2.3)
Nitrogênio Uréico no Sangue	30.4 (24)	41.6 (27)	29.4 (23)	39.3 (27)	-	-
Creatinina	1.47 (1.4)	1.72 (1.4)	1.47 (1.3)	1.77 (1.4)	1.35 (1.1)	1.68 (1.1)
Glicose	140 (51)	139 (49)	147 (56)	149 (54)	134 (48)	147 (52)
Frequência Cardíaca	86.7 (18)	89.1 (17)	85.7 (18)	88.7 (18)	85 (19)	90.6 (19)
Frequência Respirató- ria	18.9 (7.2)	19.7 (7.8)	19.1 (6.7)	20.3 (7.8)	19.9 (5.4)	20.1 (5.5)
Hematócrito	30.1 (4.7)	29.3 (4)	29.3 (5.8)	28.4 (5.4)	30.5 (7)	25.5 (5.1)
Hemoglobina	10.3 (1.6)	9.67 (1.4)	9.87 (1.9)	9.09 (1.6)	10.1 (2.4)	8.29 (1.7)
Protrombina INR	1.54 (0.74)	1.64 (0.75)	1.53 (0.72)	1.65 (0.75)	-	-
Pressão Arterial Mé- dia	79.7 (16)	76.9 (15)	79.2 (15)	76.6 (15)	84.8 (16)	83.6 (16)
Neutrófilos	77.5 (17)	77 (18)	76.3 (17)	76.9 (16)	41.7 (34)	43.9 (35)
Pressão Arterial Dias- tólica	60.3 (14)	57.4 (13)	63.2 (15)	59.9 (14)	63.8 (14)	61.9 (14)
Pressão Arterial Sis- tólica	122 (24)	119 (23)	120 (22)	118 (22)	126 (24)	127 (24)
Peso	83.4 (22)	86.7 (24)	86.7 (23)	91.5 (23)	-	-
Plaquetas	214 (131)	226 (156)	195 (117)	206 (143)	210 (108)	199 (123)
Potássio	4.12 (0.62)	4.09 (0.58)	4.26 (0.72)	4.11 (0.65)	4.01 (0.68)	4.01 (0.73)
Sódio	139 (4.9)	139 (5.4)	135 (6.9)	136 (6.5)	140 (6.2)	140 (8.4)
Temperatura	37.1 (0.82)	37.1 (0.86)	36.9 (0.92)	37 (0.94)	36.2 (0.74)	36.3 (0.84)
Troponina	1.25 (3.8)	1.2 (3.5)	0.667 (1.7)	0.706 (1.7)	298 (2413)	294 (1151)

Contagem de Leucócitos	11.8 (6.3)	13.1 (7.1)	12 (6.6)	14 (8.2)	11.7 (6.8)	13 (7.8)
Método de Avaliação de Confusão	0.368 (0.48)	0.524 (0.5)	0.489 (0.5)	0.627 (0.48)	-	-
Altura	66.5 (4.6)	66.9 (4.3)	66.6 (4.6)	66.8 (4.2)	-	-
Alanina	235 (682)	192 (550)	227 (660)	177 (511)	-	-
Fosfato	138 (135)	179 (172)	129 (117)	164 (141)	-	-
Intervalo Aniônico	13.4 (3.6)	14.1 (4.1)	14.2 (4)	15 (4.4)	-	-
Aspartato Amino-transferase	287 (902)	257 (802)	299 (948)	246 (777)	-	-
Cálcio	8.34 (0.77)	8.4 (0.84)	8.39 (0.75)	8.48 (0.84)	4.49 (0.4)	4.6 (0.49)
Ácido láctico	2.59 (2.5)	2.67 (2.5)	2.5 (2.2)	2.46 (2.1)	1.93 (1.3)	1.8 (0.99)
Magnésio	2.06 (0.34)	2.11 (0.33)	2.11 (0.35)	2.14 (0.34)	1.92 (0.38)	1.88 (0.37)
Tempo de Protrombina	43.4 (25)	48.1 (26)	43.6 (25)	49.6 (25)	-	-
Pressão Inspiratória Máxima	21.3 (8.8)	23.3 (9)	19.7 (7.4)	21.3 (7.6)	-	-
PEEP	8.21 (3.9)	9.28 (4.2)	8.19 (3.8)	9.27 (4.2)	-	-
Protrombina	11.7 (7.8)	12.4 (8.1)	16.6 (7.4)	17.9 (7.6)	-	-
Fornecimento de Oxigênio	49.2 (15)	48.5 (15)	49.7 (18)	47.4 (15)	-	-
PH Arterial	7.39 (0.075)	7.39 (0.076)	7.38 (0.083)	7.38 (0.085)	7.41 (0.092)	7.4 (0.091)
Tamanho da Pupila Esquerda	2.94 (0.92)	2.98 (0.91)	2.94 (0.81)	2.96 (0.82)	-	-
Tamanho da Pupila Direita	2.95 (0.94)	2.97 (0.92)	2.96 (0.88)	2.97 (0.86)	-	-
Glasgow Ocular	3.3 (1.1)	3.08 (1.1)	3.33 (1)	2.99 (1.2)	-	-
Glasgow Verbal	3.1 (1.9)	1.97 (1.6)	3.33 (1.8)	2.01 (1.6)	-	-
Glasgow Motor	5.32 (1.4)	4.86 (1.7)	5.38 (1.3)	4.71 (1.8)	-	-
Glasgow Total	11.7 (3.7)	9.9 (3.6)	12 (3.7)	9.71 (3.8)	15.1 (3.6)	13.3 (4.3)
Tamanho do Catéter	15.6 (1.5)	15.8 (1.6)	15.4 (1.2)	15.5 (1.3)	-	-
Braden Percepção	2.86 (0.9)	2.38 (0.83)	2.98 (0.9)	2.43 (0.85)	3.26 (0.95)	2.36 (1.1)
Braden Umidade	3.41 (0.66)	3.12 (0.71)	3.45 (0.63)	3.14 (0.66)	3.31 (0.61)	3.02 (0.7)
Braden Atividade	1.26 (0.6)	1.11 (0.35)	1.45 (0.75)	1.21 (0.45)	2.07 (1.2)	1.27 (0.71)
Braden Mobilidade	2.44 (0.74)	2.02 (0.65)	2.54 (0.75)	2.04 (0.67)	2.96 (0.85)	2.19 (0.83)
Braden Nutrição	2.29 (0.64)	2.24 (0.63)	2.34 (0.67)	2.25 (0.66)	2.93 (0.4)	2.82 (0.47)
Braden Fricção	2.12 (0.5)	1.76 (0.53)	2.16 (0.53)	1.74 (0.55)	2.4 (0.68)	1.84 (0.69)
Braden Total	14.4 (2.6)	12.7 (2.1)	14.9 (2.8)	12.8 (2.1)	17 (3.3)	13.6 (3)
Taxa de Reenchimento Capilar [% Lenta]	3.65 (19)	8.01 (27)	-	-	-	-

Tabela 6 – Tabela com a porcentagem de dados faltantes, em termos de proporção de internações com nenhuma leitura, para cada conjunto de dados.

Conjunto de dados Variável	MIMIC-III Faltantes (%)	MIMIC-IV Faltantes (%)	HMD Faltantes (%)
Albumina	62.8	61.3	37.5
Pressão Arterial de CO2	36.3	47.0	26.6
Pressão Arterial de O2	36.3	47.0	26.6
Saturação de O2	0.0	0.0	0.0
Bilirrubina	54.2	48.0	30.2
Nitrogênio Uréico no Sangue	0.5	0.4	-
Creatinina	0.5	0.4	0.6
Glicose	5.6	0.3	13.9
Frequência Cardíaca	0.0	0.0	0.0
Frequência Respirató- ria	0.0	0.0	0.0
Hematócrito	30.5	72.2	0.6
Hemoglobina	0.8	0.5	0.6
Protrombina INR	22.4	9.9	-
Pressão Arterial Mé- dia	0.0	0.0	0.0
Neutrófilos	79.5	53.9	0.6
Pressão Arterial Dias- tólica	0.0	0.0	0.0
Pressão Arterial Sis- tólica	0.0	0.0	0.0
Peso	16.8	2.3	99.9
Plaquetas	0.6	0.5	0.6
Potássio	26.7	63.2	2.9
Sódio	30.0	70.9	3.0
Temperatura	38.0	88.2	0.0
Troponina	76.5	65.1	52.8
Contagem de Leucóci- tos	13.5	0.5	0.6
Método de Avaliação de Confusão	85.7	48.1	-
Altura	76.1	48.1	-
Alanina	52.7	47.8	-
Fosfato	53.7	48.6	-
Intervalo Aniônico	52.2	0.5	-
Aspartato Amino- transferase	52.8	47.8	-
Cálcio	18.7	4.5	71.2

Ácido láctico	45.3	35.2	39.1
Magnésio	2.7	1.9	8.7
Tempo de Protrombina	11.6	10.4	-
Pressão Inspiratória Máxima	48.0	51.6	-
PEEP	86.6	62.7	-
Protrombina	11.3	9.9	-
Fornecimento de Oxigênio	74.5	42.8	-
PH Arterial	36.0	46.7	17.0
Tamanho da Pupila Esquerda	32.0	7.3	-
Tamanho da Pupila Direita	32.0	7.3	-
Glasgow Ocular	0.0	0.0	-
Glasgow Verbal	0.0	0.0	-
Glasgow Motor	0.0	0.0	-
Glasgow Total	0.0	0.0	84.0
Tamanho do Catéter	71.5	40.1	-
Braden Percepção	0.0	0.0	0.0
Braden Umidade	0.0	0.0	0.0
Braden Atividade	0.0	0.0	0.0
Braden Mobilidade	0.0	0.0	0.0
Braden Nutrição	0.0	0.0	0.0
Braden Fricção	0.0	0.0	0.0
Braden Total	0.0	0.0	0.0
Taxa de Reenchimento Capilar	66.7	-	-

APÊNDICE B – TABELAS DE HIPERPARÂMETROS

Tabela 7 – Configurações dos hiperparâmetros de cada modelo utilizados na otimização¹

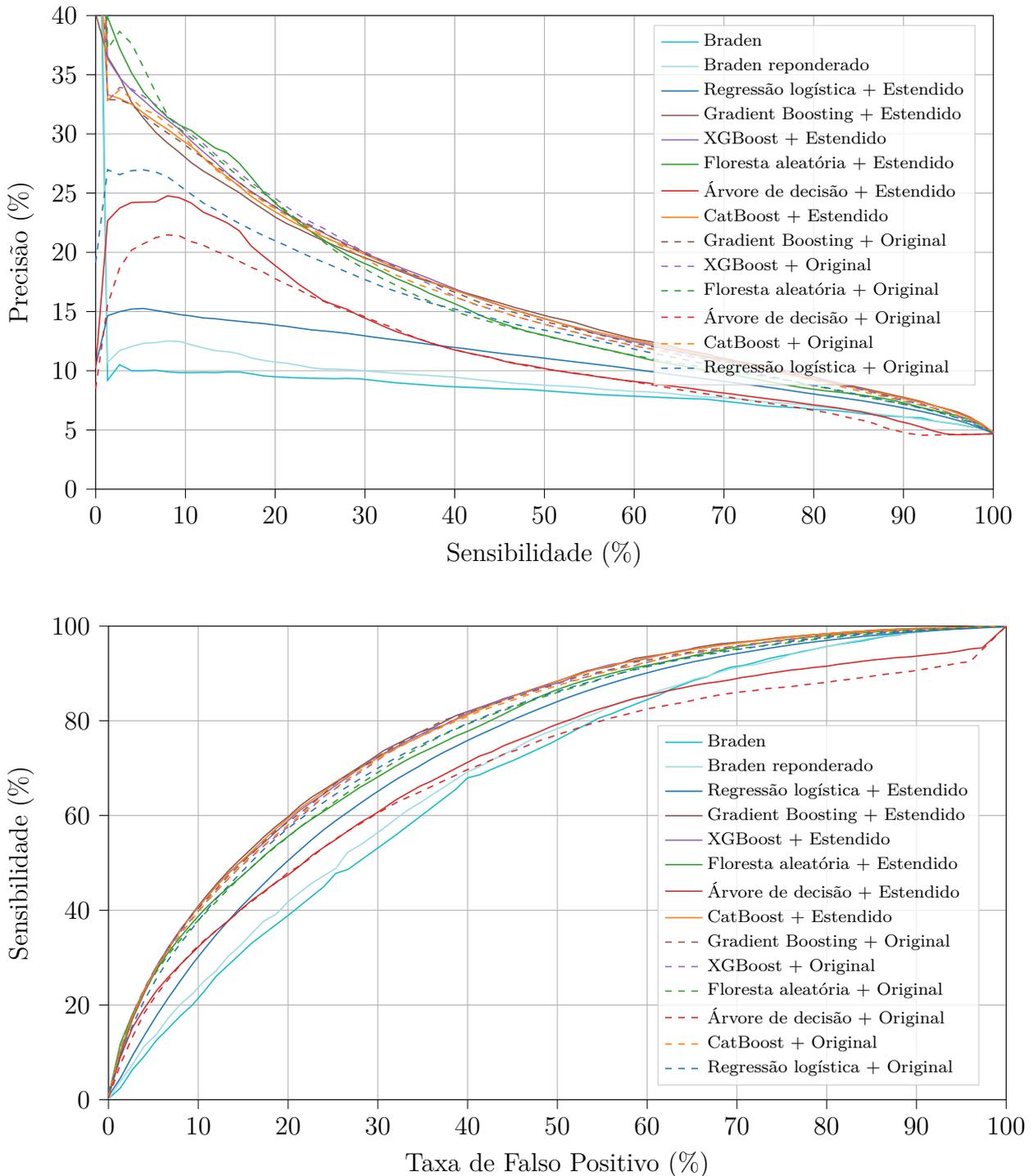
<i>Gradient Boosting</i>		
Parâmetro	Tipo	Valores
learning_rate	loguniform	[0.005, 0.2]
n_estimators	randint	[100, 800]
min_samples_split	randint	[4, 32]
max_depth	randint	[3, 6]
max_features	uniform	[0.04, 0.36]
subsample	uniform	[0.3, 0.7]
<i>XGBoost</i>		
Parâmetro	Tipo	Valores
learning_rate	loguniform	[0.005, 0.2]
max_depth	randint	[3, 8]
subsample	uniform	[0.3, 0.6]
colsample_bytree	uniform	[0.04, 0.36]
reg_lambda	uniform	[0, 4]
reg_alpha	uniform	[0, 1]
grow_policy	select	['depthwise', 'lossguide']
<i>CatBoost</i>		
Parâmetro	Tipo	Valores
grow_policy	select	['SymmetricTree', 'Depthwise', 'Lossguide']
learning_rate	loguniform	[0.005, 0.2]
iterations	randint	[100, 500]
depth	randint	[3, 6]
l2_leaf_reg	loguniform	[0.001, 4]
random_strength	uniform	[0.1, 2.9]
bagging_temperature	uniform	[0, 3]
Floresta Aleatória		
Parâmetro	Tipo	Valores
n_estimators	randint	[50, 500]
max_depth	randint	[3, 6]
min_samples_split	randint	[8, 128]
max_features	uniform	[0.04, 0.36]
max_samples	uniform	[0.3, 0.7]
Árvore de Decisão		
Parâmetro	Tipo	Valores
max_depth	randint	[8, 15]
min_samples_split	randint	[8, 128]
max_features	uniform	[0.1, 0.4]
Regressão Logística		
Parâmetro	Tipo	Valores
C	uniform	[0.01, 3.99]
max_iter	loguniform	[100, 600]

Fonte - do autor

¹ Observação: a distribuição *uniform*(a, b) é uma distribuição uniforme no intervalo [a, a+b], enquanto a *loguniform* é uniforme de maneira logarítmica no intervalo [a, b]. A distribuição *randint* fornece um número inteiro aleatório no intervalo [a, b].

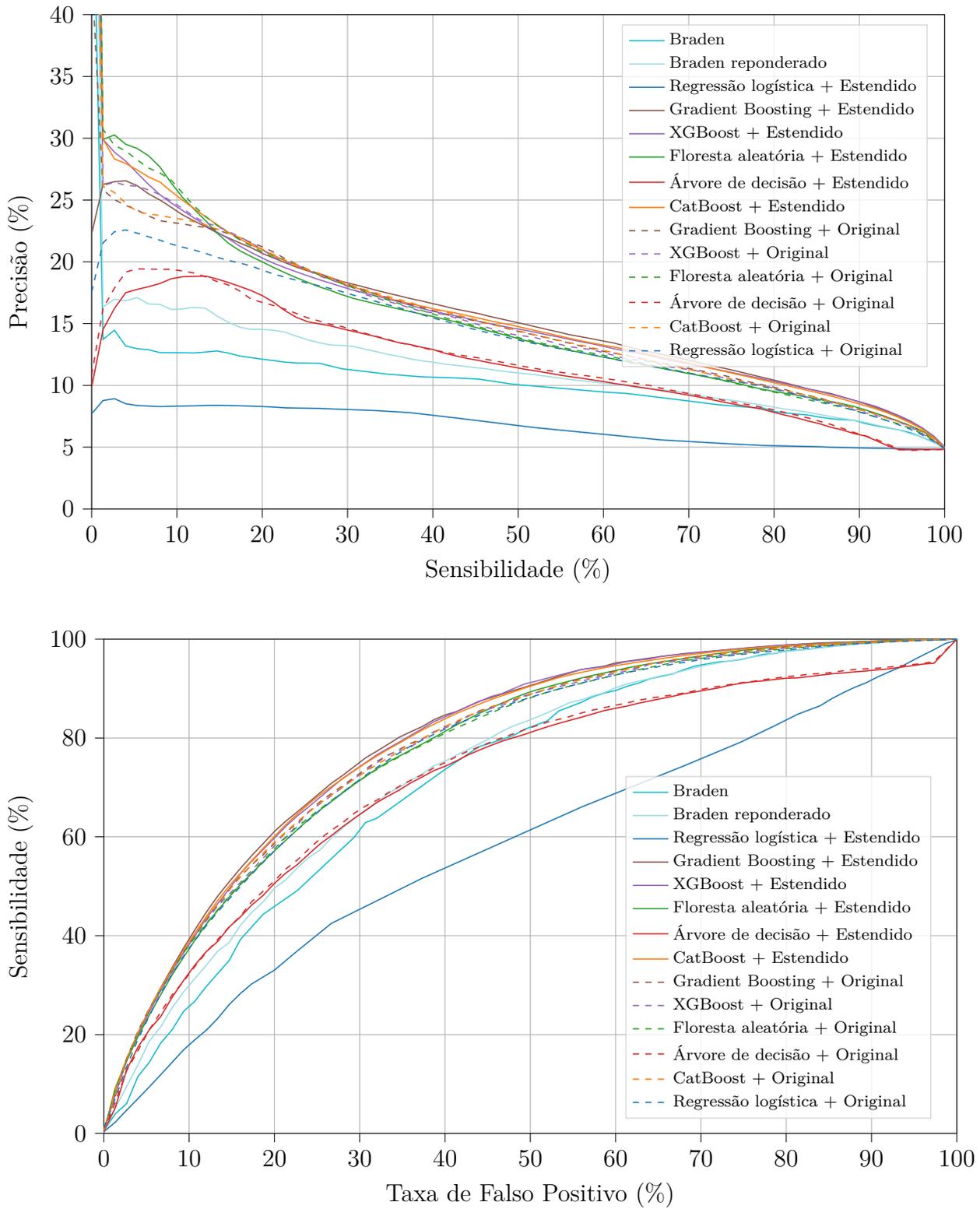
APÊNDICE C – CURVAS DE TODOS OS MODELOS

Figura 14 – Sumarização dos pontos de operação para todos os modelos no MIMIC-III.



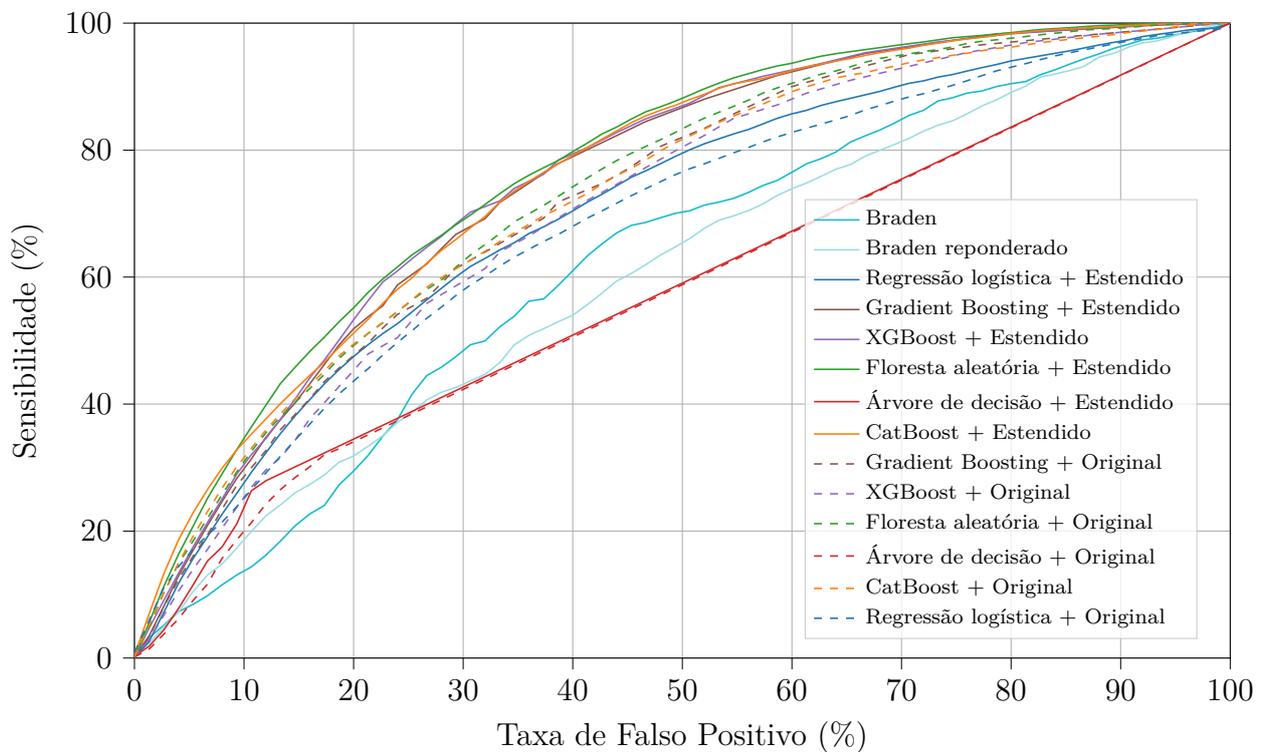
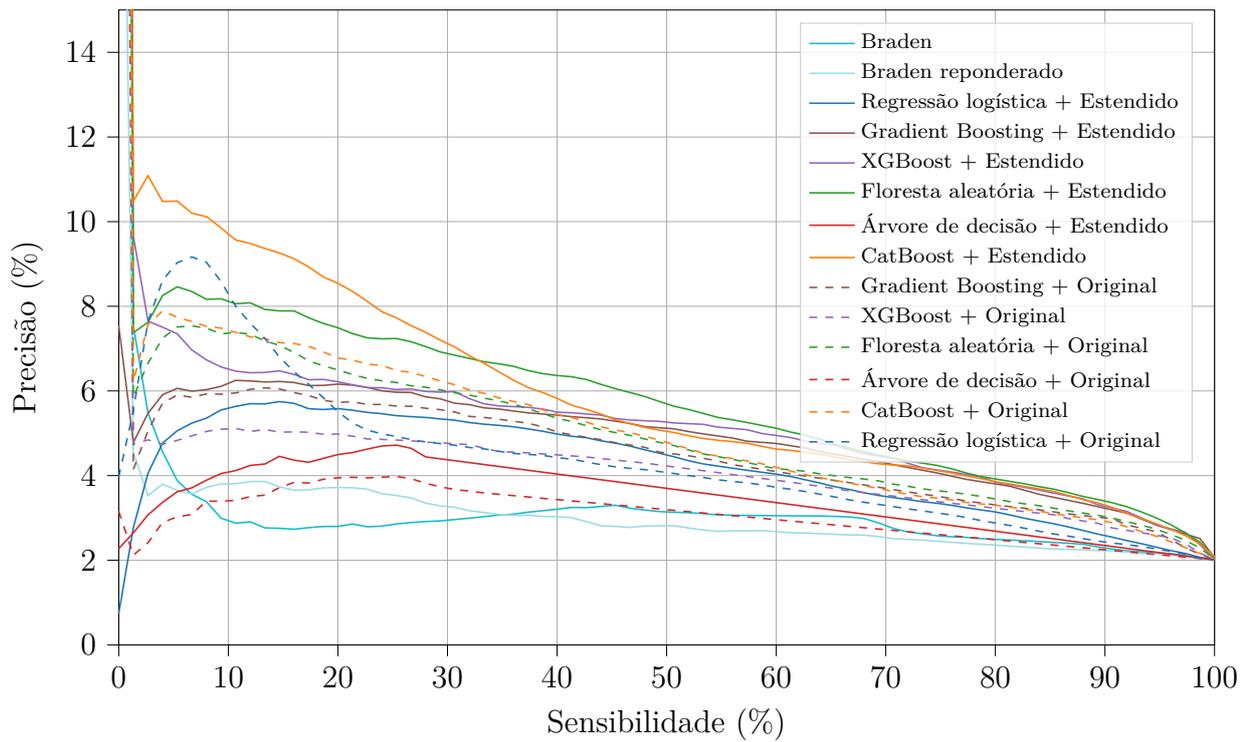
Fonte - do autor

Figura 15 – Sumarização dos pontos de operação para todos os modelos no MIMIC-IV.



Fonte - do autor

Figura 16 – Sumarização dos pontos de operação para todos os modelos no HMD.



Fonte - do autor