

EVERSON COSER

**APLICAÇÃO TÉCNICA DE APRENDIZADO DE
MÁQUINA NA PREDIÇÃO DE MORTALIDADE
HOSPITALAR EM PACIENTES ACOMETIDOS POR
AVC HEMORRÁGICO - UM ESTUDO DE CASO
MIMIC-IV**

Trabalho de Conclusão de Curso apresentado ao Curso de Ciência da Computação da Universidade Federal da Fronteira Sul (UFFS), como requisito para obtenção do título de Bacharel em Ciência da Computação.

Orientador: Prof. Dr. Denio Duarte

**CHAPECÓ
2025**

EVERSON COSER


APLICAÇÃO TÉCNICA DE APRENDIZADO DE MÁQUINA NA PREDIÇÃO DE MORTALIDADE HOSPITALAR EM PACIENTES ACOMETIDOS POR AVC HEMORRÁGICO - UM ESTUDO DE CASO MIMIC-IV

Trabalho de Conclusão de Curso apresentado ao Curso de Ciência da Computação da Universidade Federal da Fronteira Sul (UFFS), como requisito para obtenção do título de Bacharel em Ciência da Computação.


Orientador: Prof. Dr. Denio Duarte

Este Trabalho de Conclusão de Curso foi avaliado e aprovado pela banca avaliadora em: 27/11/2025.


BANCA AVALIADORA

Documento assinado digitalmente
 **DENIO DUARTE**
Data: 03/12/2025 09:21:24-0300
Verifique em <https://validar.iti.gov.br>

Dr. Denio Duarte - UFFS

Documento assinado digitalmente
 **JULYANE FELIPETTE LIMA**
Data: 03/12/2025 14:16:00-0300
Verifique em <https://validar.iti.gov.br>

Profa. Dra. Julyane Felipette Lima - UFFS


Documento assinado digitalmente
 **GIANCARLO DONDONI SALTON**
Data: 03/12/2025 18:07:26-0300
Verifique em <https://validar.iti.gov.br>


Prof. Dr. Giancarlo Dondoni Salton - UFFS

Aplicação Técnica de Aprendizado de Máquina na Predição de Mortalidade Hospitalar em Pacientes Acometidos por AVC Hemorrágico - um estudo de caso MIMIC-IV

Predicting Hospital Mortality in Hemorrhagic Stroke Patients using Machine Learning: A MIMIC-IV Study

Everson Coser   [Universidade Federal da Fronteira Sul | everson.coser@estudante.uffs.edu.br]

Denio Duarte  [Universidade Federal da Fronteira Sul | duarte@uffs.edu.br]

 Universidade Federal da Fronteira Sul, Rodovia SC 484, km 02, Bairro Fronteira Sul, Chapecó, SC CEP 89815-899
Tel. (49) 2049-2600 CNPJ 11.234.780/0007-46, Brasil.

Resumo. Este estudo tem como objetivo desenvolver modelos preditivos baseados em aprendizado de máquina para a predição de óbitos de pacientes em ambiente hospitalar, acometidos por Acidente Vascular Cerebral (AVC) hemorrágico, em razão da elevada taxa de mortalidade associada à doença, o que permite realizar o tratamento do enfermo a fim de evitar a fatalidade. Para tanto, houve a utilização de resultados de exames coletados nas primeiras 24 horas, após admissão na unidade de saúde. O banco de dados MIMIC-IV foi utilizado para este trabalho. Nele, 2.976 pacientes maiores de 18 foram selecionados, tendo sido acometidos por AVC hemorrágico. Deste total, 1.367 faleceram, sendo que 513 deles morreram no ambiente hospitalar. Testes empíricos foram realizados com algoritmos de aprendizado de máquina supervisionado, no qual XGBoost e LightGBM foram selecionados para aprofundamento em razão de apresentarem uma performance superior aos demais com base na área sob a curva (AUC). Da mesma forma, as curvas de calibração dos modelos resultantes foram analisadas além da importância de cada atributo nas decisões dos modelos por meio da técnica SHAP, a qual demonstrou que o indicador final da Escala de Coma de Glasgow é o mais impactante na predição do óbito hospitalar. Tanto o XGBoost quanto o LightGBM apresentaram uma grande capacidade de discriminação das classes, $AUC = 0.915$ e $AUC = 0.911$, respectivamente. A análise das curvas de calibração mostrou que os modelos não tendem a superestimar ou subestimar os riscos significativamente, o que, conseqüentemente, garante sua utilidade para a tomada de decisões no ambiente hospitalar. Igualmente, as curvas de decisão clínicas evidenciaram que ambos os modelos apresentam um benefício clínico associado, principalmente nos baixos e médios limiares probabilísticos. Por sua vez, as curvas Precision-Recall demonstraram a relação entre a precisão e a revocação para o conjunto de testes que se encontra desbalanceado. Tanto o XGBoost quanto o LightGBM apresentaram resultados significativos para a identificação precoce de óbitos em pacientes com acidente vascular cerebral hemorrágico, isso contribui na tomada de decisão clínica ao empregar dados de exames laboratoriais coletados dos pacientes, o que garante a estratificação de risco e a potencialidade de reduzir complicações futuras.

Abstract. This study aims to develop machine learning-based predictive models for predicting deaths in hospitalized patients suffering from hemorrhagic stroke, given the high mortality rate associated with the disease, thus enabling treatment to prevent fatalities. To this end, results from tests collected within the first 24 hours after admission to the healthcare unit were used. The MIMIC-IV database was used for this work. In this database, 2,976 patients over 18 years of age who had suffered hemorrhagic stroke were selected. Of this total, 1,367 died, with 513 of them dying in the hospital setting. Empirical tests were performed with supervised machine learning algorithms, in which XGBoost and LightGBM were selected for further investigation due to their superior performance based on the area under the curve (AUC). Similarly, the calibration curves of the resulting models were analyzed, in addition to the importance of each attribute in the models' decisions, using the SHAP technique, which demonstrated that the final indicator of the Glasgow Coma Scale is the most impactful in predicting hospital mortality. Both XGBoost and LightGBM showed a high capacity for class discrimination, $AUC = 0.915$ and $AUC = 0.911$, respectively. The analysis of the calibration curves showed that the models do not tend to overestimate or underestimate risks significantly, which consequently guarantees their usefulness for decision-making in the hospital environment. Likewise, the clinical decision curves showed that both models present an associated clinical benefit, mainly in the low and medium probabilistic thresholds. In turn, the Precision-Recall curves demonstrated the relationship between precision and recall for the unbalanced set of tests. Both XGBoost and LightGBM have shown significant results for the early identification of deaths in patients with hemorrhagic stroke. This contributes to clinical decision-making by using laboratory data collected from patients, ensuring risk stratification and the potential to reduce future complications.

Palavras-chave: Aprendizado de máquina, AVC hemorrágico, MIMIC - IV, Aprendizado supervisionado, Métricas

Keywords: Machine learning, hemorrhagic stroke, MIMIC - IV, Supervised learning, Metrics

1 Introdução

O acidente vascular cerebral (AVC) é uma síndrome caracterizada por eventos rápidos da perda das funções cerebrais, que pode variar em termos de gravidade, com o indivíduo

podendo se recuperar em horas, ter uma recuperação incompleta e até mesmo levar a óbito [Warlow, 1998]. Essa síndrome é uma das principais causas de morte em todo o mundo, sendo responsável por um elevado número de in-

tenções anuais, em virtude de dois tipos: isquêmico ¹ e hemorrágico [Feigin *et al.*, 2024]. Segundo Rajashekar and Liang [2023], a hemorragia intracerebral (HIC) é um sub-tipo de acidente vascular cerebral, no qual um hematoma se forma dentro do parênquima cerebral. Ainda, a lesão pode ser caracterizada como primária ou secundária: a HIC primária corresponde a 85% dos casos e está associada à hipertensão crônica ou à angiopatia amilóide. Por outro lado, a hemorragia secundária está associada a distúrbios hemorrágicos decorrentes do sangramento na região intracraniana.

Dentre suas formas clínicas, o AVC hemorrágico constitui aproximadamente 20% de todos os casos associados aos eventos cerebrovasculares [Qureshi *et al.*, 2009], afetando milhões de pessoas todos os anos [Virani *et al.*, 2020]. Estima-se que, somente em 2021, 3.444 milhões de novos casos de HIC ocorreram ao redor do mundo, com um total de 3.308 milhões de mortes associadas à doença [Xu *et al.*, 2024]. Estudos indicam que a taxa de mortalidade em pacientes nos primeiros dias após o evento são superiores a 40%, enquanto, apenas 12% a 39% são considerados funcionalmente independentes em até 6 meses posteriores à ocorrência do episódio [Faghih-Jouybari *et al.*, 2021]. Com base nisso, indivíduos acometidos por hemorragia intracerebral tendem a apresentar piores problemas neurológicos e uma maior possibilidade de desenvolver incapacidades em comparação com àqueles que sofreram um AVC isquêmico [Van Asch *et al.*, 2010].

Tendo em vista este panorama, os modelos de predição baseados em dados clínicos se mostram cada vez mais importantes para a tomada de decisões no âmbito hospitalar [Van Smeden *et al.*, 2021]. Dessa forma, dois modelos de predição são empregados: a previsão diagnóstica, que está relacionada com o fator de calcular o risco do paciente ter uma determinada doença; e a previsão prognóstica, que busca calcular o risco do paciente desenvolver uma determinada doença no futuro [Hendriksen *et al.*, 2013]. Tradicionalmente, escores clínicos, como a Escala de Coma de Glasgow e SOFA, são empregados para avaliar a gravidade de pacientes em casos médicos [Tian *et al.*, 2025]. Embora esses indicadores auxiliem no tratamento de enfermos, limitações são encontradas, seja em razão da dificuldade de estabelecer relações com outros fatores que influenciam a evolução clínica ou em virtude da falta de informações para calcular os escores [Lieh-Lai *et al.*, 1992; Molinnus *et al.*, 2024].

Modelos preditivos baseados em aprendizado de máquina também apresentam um significativo potencial na prevenção de agravos ao identificar pacientes com maiores riscos de deterioração clínica. Isso é possível por meio da análise e identificação de padrões complexos presentes em conjuntos de dados, uma vez que o modelo é capaz de captar mudanças sutis que podem anteceder eventos adversos [Ribeiro *et al.*, 2020].

À vista disso, a predição precisa e antecipada do acidente vascular cerebral hemorrágico é relevante para estimar o risco de óbito hospitalar e otimizar a alocação de recursos. Nos últimos anos, técnicas de aprendizado de máquina passaram a ser utilizadas nos centros de saúde, pois fornecem um suporte para a tomada de decisões clínicas em razão da pos-

sibilidade de criar modelos preditivos capazes de identificar padrões em dados médicos [Shehab *et al.*, 2022]. Isso se deve principalmente em virtude do avanço tecnológico, uma vez que contribuiu para que os bancos de dados clínicos crescessem substancialmente por meio da digitalização de prontuários e exames [Paixão *et al.*, 2022]. Logo, bases de registros médicos coletados de pacientes no ambiente hospitalar constituem um importante recurso para pesquisas, pois fornecem informações determinantes para criar estratégias diagnósticas eficazes que visam reverter fatalidades.

Considerando isso, o Beth Israel Deaconess Medical Center, em parceria com o Instituto de Tecnologia de Massachusetts, disponibilizou o banco de dados público Medical Information Mart for Intensive Care IV (MIMIC-IV), com a finalidade de proporcionar uma fonte de dados estruturada, com informações coletadas de pacientes reais entre 2008 e 2022 [Johnson *et al.*, 2023]. Consequentemente, modelos de aprendizado de máquina podem ser utilizados em pesquisas e treinados com uma heterogeneidade apta a promover a generalização de modelos preditivos para serem usufruídos em diferentes contextos hospitalares.

Portanto, este trabalho está centrado em aplicar técnicas de aprendizado de máquina para prever mortalidade hospitalar em pacientes com AVC hemorrágico utilizando a base de dados MIMIC-IV com informações coletadas nas primeiras 24 horas após a admissão na unidade de saúde. Os algoritmos escolhidos para os modelos preditores que obtiveram um desempenho superior são XGBoost e LightGBM, com uma AUC de 0.915 e 0.911, respectivamente. Ambos os modelos preditivos estão baseados no aprendizado supervisionado de classificação.

A escolha por esse tema se justifica pela pequena quantidade de trabalhos produzidos, aproximadamente sete estudos, e a importância dele em fornecer informações adicionais aos profissionais de saúde para a rápida identificação de possíveis desfechos clínicos com o intuito de auxiliar no tratamento de doentes.

O artigo está estruturado da seguinte maneira. Na Seção 2, os trabalhos relacionados com o presente estudo são apresentados. Na Seção 3, a metodologia empregada no trabalho é descrita. Na Seção 4, os resultados obtidos por meio dos modelos de aprendizado de máquina são apresentados. Na Seção 5, uma análise comparativa entre o presente estudo e outros trabalhos é realizada e, por fim, na Seção 6, a conclusão é exposta.

2 Trabalhos relacionados

Um estudo produzido por Miao *et al.* [2024], com o intuito de prever um possível risco de readmissão de pacientes com hemorragia intracerebral na UTI empregou quatro modelos de aprendizado de máquina: AdaBoost, RandomForest, Light-GBM e XGBoost utilizando dados do MIMIC-III e MIMIC-IV. Um total de 2.242 pacientes com hemorragia intracerebral (HIC) foram selecionados, sendo 747 advindos do MIMIC-III e 1.495 do MIMIC-IV.

Para realizar a filtragem dos pacientes com HIC, os códigos da classificação CID-9 (Classificação Internacional de Doenças) foram utilizados: 431, I610-I616 e I618-I619. Esse código foi criado pela Organização Mundial da Saúde

¹ ocorre em razão da interrupção do fluxo sanguíneo para o cérebro.

(OMS) para a padronização na identificação de enfermidades. Pacientes com menos de 18 anos, com internação inicial em UTI inferior a 24 horas e aqueles que faleceram após a primeira transferência para a UTI foram excluídos do conjunto de dados de 2.242 pacientes. No caso de indivíduos possuírem mais de uma internação associada, apenas a primeira foi considerada.

Para lidar com as informações faltantes a respeito dos indivíduos selecionados, a técnica de imputação múltipla foi utilizada. Outrossim, o conjunto de dados foi dividido na proporção 70:30, ou seja, 70% para treinamento e 30% para testes. Por fim, a técnica Recursive Feature Elimination with Cross-Validation (RFECV) foi empregada para identificar as possíveis variáveis que estavam associadas à readmissão na unidade de terapia intensiva e a técnica SMOTE-ENN (Synthetic Minority Oversampling Technique-Edited Nearest Neighbours), utilizada para balancear o conjunto de dados que estava inicialmente desbalanceado com a criação de informações sintéticas e a eliminação de ruídos da classe dominante, resultando em uma fronteira de decisão mais definida.

Os resultados apresentados indicaram uma taxa de readmissão de 9,6% para MIMIC-III e 10,6% para MIMIC-IV. O modelo Light-GBM apresentou o melhor desempenho, com uma área sob a curva (AUC) de 0,736. Além disso, a técnica SHAP (SHapley Additive exPlanations) foi utilizada com o intuito de verificar quais foram os atributos com maior relevância na predição, e, como resposta, hidrocefalia, sexo, neutrófilos, Escala de Coma de Glasgow (GCS), níveis específicos de saturação de oxigênio (SpO₂) e creatinina foram indicados.

Por sua vez, Chen *et al.* [2025] apresentaram um estudo cujo objetivo é avaliar o risco de morte de pacientes com hemorragia intracerebral não traumática, utilizando, para tal fim, um modelo de aprendizado de máquina. O trabalho contou com um conjunto de dados de 1.274 pacientes com HIC do banco de dados MIMIC-IV, versão 3.0, e 129 pacientes obtidos de informações do Segundo Hospital da Universidade de Lanzhou para validação externa.

Visando a realização do estudo, um conjunto de 34 variáveis relacionadas com 11.178 dados atribuídos à primeira internação de pacientes com HIC foram coletados com base na CID - versão 10. Na sequência, o tratamento dos dados foi efetuado da seguinte maneira: variáveis com uma taxa de dados faltantes superiores a 30% foram excluídas, resultando em um total de 22 atributos restantes para análise, e a técnica Multivariate Imputation via Chained Equations (MICE) foi empregada para atribuir valores nos campos nulos. Com tais informações, o conjunto resultante foi dividido em três grupos: treinamento, validação e teste, na proporção de 6:2:2.

Para identificar quais as variáveis mais importantes, a regressão LASSO - (Least Absolute Shrinkage and Selection Operator) foi aplicada, a qual resultou em um total de sete atributos potenciais para a estrutura preditiva, além de uma análise de regressão logística multivariada. Essa análise concluiu que idade, albumina, lipoproteína de baixa densidade - (LDL), fibrilação atrial, resposta motora na Escala de Coma de Glasgow, resposta ocular na Escala de Coma de Glasgow e gênero foram apontados como os atributos mais informativos para a predição.

O modelo de aprendizado de máquina escolhido para a previsão foi o XGBoost, que apresentou uma AUC de 0,814 para o conjunto de treinamento, uma AUC de 0,715 para a validação, uma AUC de 0,797 para o conjunto de testes e uma AUC de 0,847 para a validação externa. Ademais, a técnica SHAP foi utilizada para identificar o grau de importância de cada atributo e apresentou como resultado, em ordem decrescente: resposta motora na Escala de Coma de Glasgow, idade, resposta ocular na Escala de Coma de Glasgow, LDL, albumina, fibrilação atrial e gênero.

De forma semelhante, um estudo com o intuito de realizar a predição precoce do risco de mortalidade de pacientes com hemorragia intracerebral admitidos na UTI foi conduzido por Mao *et al.* [2024]. Neste sentido, um conjunto 1.486 pacientes presentes no banco de dados MIMIC-IV e 110 pacientes do Hospital de Zhejiang internados entre dezembro de 2018 e fevereiro de 2023 foram utilizados. A seleção ocorreu seguindo cinco regras: os pacientes devem ser admitidos na unidade de terapia intensiva pela primeira vez; devem possuir o diagnóstico confirmado para HIC; devem possuir idade entre 18 e 90 anos; devem possuir um período de internação na UTI superior a um dia e devem possuir dados clínicos completos.

A extração dos dados ocorreu da seguinte maneira: variáveis clínicas coletadas nas primeiras 24 horas após a admissão na unidade de terapia intensiva foram utilizadas, em caso de múltiplas medições relacionadas aos exames e informações fisiológicas o valor médio foi empregado para a análise. Com este método, 46 variáveis foram selecionadas e identificadas com base nos códigos da Classificação Internacional de Doenças (CID-9 e CID-10). O trabalho considerou também a relação existente entre as comorbidades e o impacto no quadro de saúde do paciente. Assim, as patologias consideradas para o estudo foram: hipertensão arterial, diabetes mellitus, doença pulmonar obstrutiva crônica (DPOC), insuficiência cardíaca congestiva, doença renal, doença hepática e neoplasia maligna.

Os modelos de aprendizado de máquina escolhidos para o desenvolvimento do trabalho foram: Logistic regression (LR), K-nearest neighbors (KNN), Adaptive boosting (AdaBoost), Random forest (RF) e eXtreme Gradient Boosting (XGBoost). Os resultados obtidos no conjunto de treinamento indicaram para o XGBoost uma AUC de 0,907, para o KNN uma AUC de 0,808, para a LR uma AUC de 0,851, para o RF uma AUC de 0,897, e para o AdaBoost uma AUC de 0,900. No conjunto de validação externa, o XGBoost apresentou uma AUC de 0,788, seguido pela LR com uma AUC de 0,790. Como o modelo XGBoost apresentou alta acurácia em ambas as predições foi escolhido como o melhor para prever o risco de mortalidade de pacientes com hemorragia intracerebral admitidos na UTI.

Outro estudo com o intuito de prever as readmissões na UTI de pacientes com hemorragia intracerebral foi conduzido por Li *et al.* [2025]. No total, 2.144 pacientes foram selecionados dos bancos de dados MIMIC-III e MIMIC-IV, 730 do MIMIC-III e 1.414 do MIMIC-IV, por meio do código '431' para CID-9 e 'I610-I616', 'I618-I619' CID-10. Além disso, quatro critérios de seleção foram seguidos: pacientes com idade inferior a 18 anos foram excluídos, assim como os indivíduos associados a uma internação inicial infe-

rior a 24 horas na UTI. Da mesma forma, casos clínicos que indicaram óbito, após a primeira transferência para a UTI, e, em última instância, em casos de múltiplas internações, apenas a primeira foi considerada.

Para a obtenção dos dados necessários para previsão, uma técnica de eliminação recursiva de características com validação cruzada (RFECV) foi empregada, a qual diminuiu o número de potenciais atributos de 50 para 18, entre eles estão: idade, gênero, frequência cardíaca, pressão sanguínea, taxa de saturação de oxigênio e Escala de Coma de Glasgow. Variáveis com mais de 90% de valores ausentes foram eliminadas e a imputação pela média foi realizada para as informações nulas dos atributos restantes. A fim de balancear os dados, a estratégia aplicada foi o método `RandomUnderSampler`, que reduziu a quantidade de amostras da classe majoritária, ou seja, pacientes que não foram readmitidos na UTI.

Da mesma forma que os trabalhos anteriores, o conjunto de dados foi dividido em treinamento e teste. Neste caso, 70% para treino e 30% para testes. Para realizar a predição, quatro algoritmos de aprendizado de máquina foram utilizados: `AdaBoost`, `Random Forest`, `XGBoost` e `LightGBM`. A otimização dos modelos preditivos foi realizada definindo hiperparâmetros comuns, como o número de estimadores, taxa de aprendizagem e profundidade máxima da árvore, e por meio da biblioteca `Optuna`, que busca maximizar o desempenho mediante a verificação de diferentes parâmetros definidos com base em um intervalo determinado.

Considerando o processo de seleção e manipulação, o modelo `AdaBoost` apresentou os melhores resultados, com uma acurácia de 0,810 e uma área sob a curva roc - `AUROC` de 0,877, seguido do modelo `XGBoost` com acurácia de 0,771 e `AUROC` de 0,865. Ademais, o `Random Forest` obteve uma acurácia de 0,782 e `AUROC` de 0,841, e o `LightGBM` uma acurácia de 0,714 e `AUROC` de 0,828. Por fim, a utilização do método `SHAP` identificou que as características mais preditivas foram contagem de leucócitos, hemoglobina, plaquetas e doença cardiovascular.

Os estudos listados se assemelham à proposta deste trabalho em razão de abordarem pacientes com acidente vascular cerebral hemorrágico. No entanto, diferem em determinados aspectos, como por exemplo, o objetivo principal dos artigos, tendo em vista que alguns tratam a predição de readmissões e internações na UTI, enquanto outros abordam a previsão de óbitos. De modo semelhante, esta proposta está centrada na predição de mortalidade de pacientes com AVC hemorrágico. Para isso, foram considerados pacientes que obtiveram uma estadia inferior a 24 horas no hospital, para indivíduos com tempo de permanência superior a um dia, apenas dados das primeiras 24 horas foram utilizados e, múltiplas admissões associadas também foram incluídas nas análises, algo que não ocorre nos demais artigos.

3 Métodos

Para o desenvolvimento do projeto, visando predizer o óbito de pacientes admitidos com AVC hemorrágico, foram estipuladas quatro etapas. A primeira diz respeito a seleção dos pacientes na base de dados, por meio da utilização do Código Internacional de Doenças. A segunda se refere à extração e manipulação dos atributos com o auxílio de um *script* em

Python. A terceira é designada para a construção e otimização do modelo preditor. Por fim, a quarta etapa está atrelada a análise dos modelos e verificação de atributos com maiores impactos preditivos. A sequência cronológica da metodologia pode ser visualizada na Figura 1.

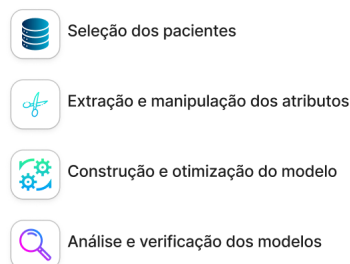


Figura 1. Etapas de desenvolvimento

3.1 Seleção dos pacientes

Essa etapa é dividida em duas subetapas: a primeira diz respeito ao entendimento da organização do banco de dados MIMIC-IV e a segunda na identificação dos pacientes admitidos por AVC hemorrágico.

3.1.1 Base de Dados

O banco de dados MIMIC - IV, versão 3.1, é amplamente utilizado em pesquisas médicas devido à grande disponibilidade de dados mapeados ao longo dos anos, como, por exemplo, resultados de exames laboratoriais, sinais vitais coletados no leito hospitalar, procedimentos e medicamentos. Também estão disponíveis inúmeras informações de pacientes reais internados em unidades de terapia intensiva (UTI) do BIDMC entre 2008 e 2022 que foram coletadas e relacionadas a várias doenças [Johnson *et al.*, 2023].

Para o desenvolvimento deste trabalho, foram coletados dados relacionadas aos pacientes internados com AVC hemorrágico, no intervalo de até 24 horas após a admissão hospitalar, em dois módulos do MIMIC-IV: *hosp* e *icu*.

O módulo *hosp* é o mais abrangente e armazena informações referentes às ações que ocorreram com os pacientes na hospitalização, como transferência de doentes, administração de medicamentos e resultados de exames laboratoriais. Neste módulo, seis tabelas foram utilizadas para a extração dos dados: *patients*, *admissions*, *labevents*, *d_labitems*, *diagnoses_icd* e *d_icd_diagnoses*. A Figura 2 indica as tabelas utilizadas, seus relacionamentos e principais atributos.

A tabela *patients* é utilizada para aplicar o processo de seleção dos pacientes que possuem idade superior a 18 anos. Ela contém as informações referentes aos dados do paciente: o atributo **subject_id** é o identificador do indivíduo e chave primária do objeto, **gender** indica o sexo, **anchor_age** a idade e **dod** a data do óbito. Os pacientes são identificados de forma numérica em razão da necessidade de anonimato.

Na sequência, a tabela *admissions*, a qual explora as informações de admissão, é utilizada com o propósito de verificar a data e hora que o paciente deu entrada no hospital. Dentre os principais atributos estão: **subject_id**, chave estrangeira, **hadm_id** que é o identificador da admissão e chave primária, além de **admittime** que indica o dia e hora da admissão. Convém salientar que o mesmo paciente pode ser admitido no hospital mais de uma vez, o que implica em um

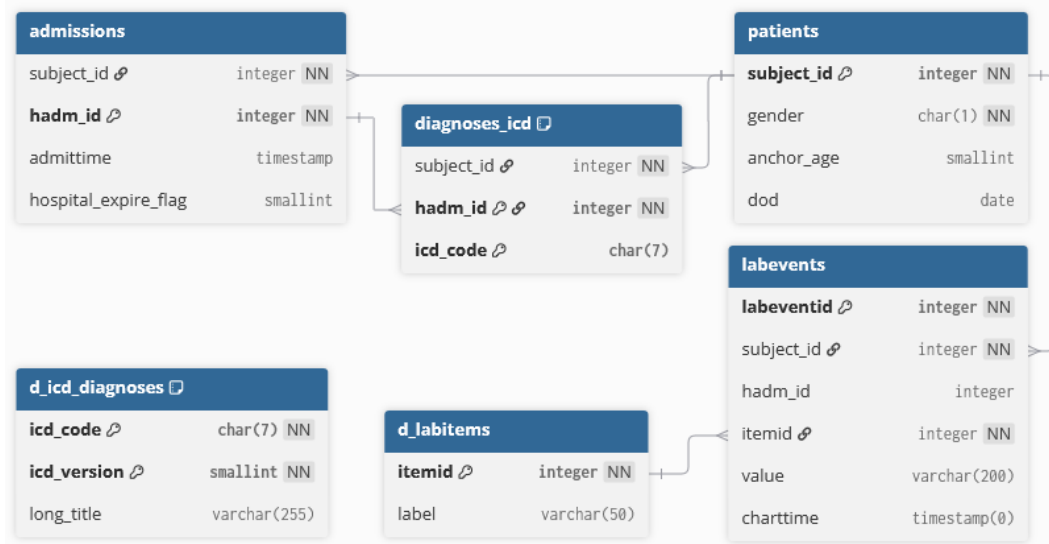


Figura 2. Modelo Lógico das principais tabelas do módulo hosp.

hadm_id diferente, ou seja, um indivíduo pode estar associado a mais de uma admissão. Na sequência, o atributo **hospital_expire_flag** indica se o paciente faleceu no hospital ou fora dele.

Além disso, duas tabelas atuam em conjunto para disponibilizar as informações relacionadas aos códigos de identificação das doenças. A primeira é *diagnoses_icd*, a qual contém os atributos **subject_id**, chave estrangeira de *patients*, **hadm_id**, chave estrangeira de *admissions*, e **icd_code** que indica o código da doença e a capacidade de relacionamentos por meio da chave primária composta (hadm_id, icd_code) e **icd_version** que apresenta o versionamento do código. A segunda tabela é *d_icd_diagnoses*, que possui como atributos o **icd_code**, **icd_version** e **long_title** que apresenta o nome da doença, possui uma chave primária composta para relacionamentos (icd_code, icd_version).

Por fim, outras duas tabelas atuam em conjunto. A primeira é *labevents*, a qual contém todas as informações referentes aos exames laboratoriais realizados pelo paciente. Dentre os principais atributos, estão **labeventid** que é a chave primária da tabela, **subject_id**, chave estrangeira de *patients*, **hadm_id**, **itemid** que é o identificador do exame realizado, **value** que contém o resultado do procedimento e **charttime** que indica o momento em que a coleta foi feita. A segunda tabela é *d_labitems* que possui os atributos **itemid**, chave primária, e **label**, que indica o nome do exame.

Por sua vez, o módulo *icu* é inteiramente composto por dados coletados pelo sistema MetaVision (iMDsoft, Israel) nos leitos da UTI. Nesta unidade, três tabelas foram utilizadas para a coleta de dados: *icustays*, *chartevents* e *d_items*. A Figura 3 apresenta essas tabelas, seus relacionamentos e principais atributos.

Inicialmente, a tabela *icustays* é responsável por armazenar as informações referentes às internações dos pacientes na Unidade de Terapia Intensiva (UTI). Dentre os principais atributos estão: **subject_id**, **hadm_id**, e **stay_id**, que é o identificador da internação e **intime** que representa o momento do internamento.

Na sequência, as tabelas *chartevents* e *d_items* atuam

em conjunto no armazenamento e identificação dos sinais vitais e exames coletados. Os principais atributos na tabela *chartevents* são: **subject_id**, **hadm_id**, **stay_id**, **charttime**, esse último apresenta o momento em que a coleta do parâmetro fisiológico foi realizado; **value** contém o dado coletado a beira do leito e **itemid** é um valor utilizado para verificar o nome do sinal. Por sua vez, a tabela *d_items* apresenta os atributos **itemid**, chave primária, e **label** apresenta o nome do item, como por exemplo, a resposta motora da Escala de Coma de Glasgow.

3.1.2 Identificação dos pacientes

Com o intuito de descobrir quais pacientes foram admitidos no hospital por AVC hemorrágico, uma consulta em SQL (Structured Query Language) foi realizada na tabela *d_icd_diagnoses*. Assim, é possível descobrir os códigos da Classificação Internacional de Doenças - CID para AVC hemorrágico. Neste caso, as informações utilizadas para a identificação desses pacientes são 'I61', 'I610', 'I611', 'I612', 'I619' para CID na versão 10 e '431' para CID na versão 9.

Em seguida, com a utilização do relacionamento entre as tabelas *diagnoses_icd* e *patients*, foi possível identificar com exatidão os pacientes acometidos pela doença, assim como, garantir que apenas maiores de 18 anos fossem selecionados. Desta forma, 2.976 indivíduos foram encontrados, dos quais 1.367 faleceram, ou seja, aproximadamente 46% do conjunto.

Como o MIMIC-IV mantém as datas de óbito de pacientes que faleceram em até um ano após a última admissão [Johnson et al., 2023], optou-se, neste trabalho, por prever as mortes que ocorreram no ambiente hospitalar, como mostra a seleção na Figura 4. Na base de dados, não existe qualquer informação que relaciona o óbito de um paciente a uma determinada doença catalogada no momento da admissão. Em virtude disso, preferiu-se pela escolha das mortes hospitalares, pois é possível identificar que no momento da admissão e ao decorrer de sua hospitalização o paciente estava com AVC hemorrágico, o que não é viável afirmar para as mortes que ocorreram fora do hospital.

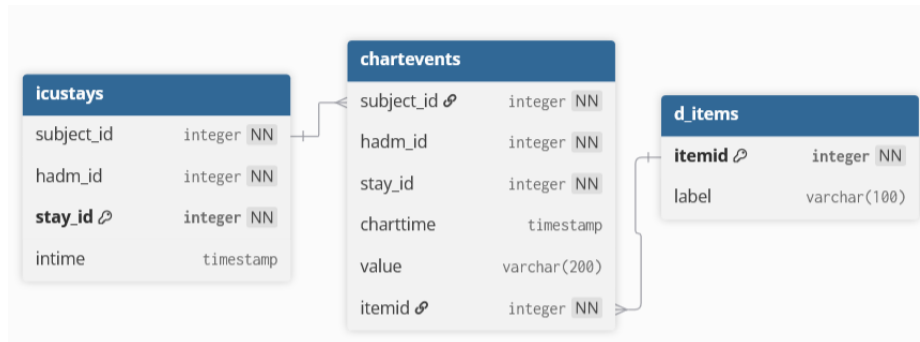


Figura 3. Modelo Lógico das principais tabelas do módulo icu.

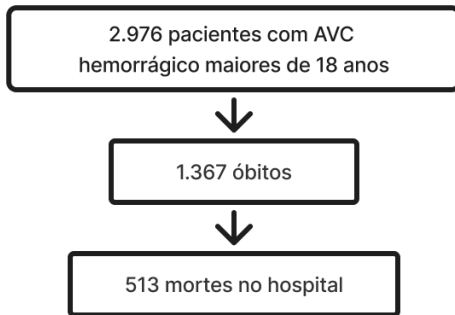


Figura 4. Seleção dos pacientes

subject_id	hadm_id	stay_id	label	value	charttime
1001	2001	1	glucose	90	2133-01-10 08:00:00
1001	2001	1	glucose	135	2133-01-10 12:00:00
1001	2002	2	glucose	95	2133-01-11 09:00:00
1001	2002	2	glucose	70	2133-01-11 12:00:00
1001	2002	2	glucose	86	2133-01-11 15:00:00
1001	2002	2	glucose	86	2133-01-11 18:00:00
1001	2002	2	glucose	75	2133-01-11 21:00:00

subject_id	hadm_id	stay_id	glucose_MAX	glucose_MIN	glucose_MODE	glucose_LAST
1001	2001	1	135	90	112.5	135
1002	2002	2	95	70	86	75

Figura 5. Exemplo fictício da transformação dos exames laboratoriais.

3.2 Extração e manipulação dos atributos

Para realizar a extração das informações pertinentes, adotou-se um *script* em Python que acessa o MIMIC-IV e retorna os dados no módulo *hosp* e do módulo *icu*. Esse *script* utiliza como restrição o atributo **admittime**, em razão da necessidade de utilizar o momento da admissão hospitalar para encontrar os exames e demais informações coletadas nas primeiras 24 horas. É possível que um mesmo paciente tenha tido várias admissões hospitalares; desta forma, no conjunto de dados final, cada admissão representará uma nova linha na tabela.

Nesse sentido, ao relacionar as tabelas *labevents* e *d_labitems* no módulo *hosp* e as tabelas *chartevents* e *d_items* no módulo *icu* foi possível identificar quais exames os pacientes realizaram durante o período em que estiveram no hospital.

Como existem várias medições para um mesmo exame dentro de um mesmo período, optou-se por fazer um pré-processamento. As informações relacionadas com avaliações laboratoriais sofreram uma análise estatística, que resultou na criação de quatro colunas para cada exame. Uma representando o valor máximo encontrado; a outra o valor mínimo; na sequência a moda, quando não existem medições iguais, ou seja, um mesmo valor, a mediana é escolhida; e, por fim, o último valor encontrado para a avaliação em questão. Essa reconfiguração dos atributos é feita respeitando o período em que o paciente esteve hospitalizado e, assim, garante-se que, a cada nova admissão, os resultados laboratoriais estarão de acordo com a respectiva admissão. Um exemplo fictício dessa transformação pode ser visualizado na Figura 5.

Ademais, no módulo *icu*, os dados passaram pela mesma transformação do módulo *hosp* em relação à criação de atributos. Outro ponto importante a ser destacado, diz res-

peito à não utilização de atributos relacionados com os sinais vitais dos pacientes presentes neste módulo. A coleta de sinais vitais é de grande relevância para iniciar qualquer tipo de atendimento [Sapra *et al.*, 2023]. No entanto, o MIMIC-IV não possui informações relacionadas com tais indicadores fisiológicos fundamentais no módulo *hosp*, o que não permite que ocorra uma comparação utilizando essas informações para verificar o aumento ou diminuição do grau de gravidade em que o paciente se encontrava quando foi internado na UTI. Assim, os atributos relacionados com os sinais vitais não foram considerados no modelo por não existirem no momento da admissão hospitalar. Essa limitação, portanto, pode impactar a capacidade de predição em relação à captura de informações iniciais do estado de saúde do paciente, as quais seriam identificadas pelos indicadores fisiológicos.

No módulo *hosp*, utilizou-se a estratégia de criar um atributo para verificar se o paciente foi readmitido no hospital após a primeira alta. Na primeira admissão, o valor zero é atribuído. Já para casos de reincidência, o valor um é designado. Além disso, mais dois atributos de readmissão foram criados, um para verificar se o paciente foi reincidente em até 30 dias após a última alta hospitalar e outro para identificar se o indivíduo ingressou novamente no hospital em até 60 dias após a última alta hospitalar.

Outros três atributos relacionados com as comorbidades foram criados. O primeiro contém a quantidade de patologias associada a cada paciente no momento da admissão. Para isso, utilizou-se o Índice de Comorbidade Elixhauser (ICE) [Elixhauser *et al.*, 1998], o qual descreve 30 comorbidades iniciais, que são condições que indicam um maior grau de riscos de problemas de saúde.

O segundo atributo se refere a utilização do S-Elixhauser (Van Walraven Score) [Van Walraven *et al.*, 2009] que depende do Elixhauser original. Esse índice correlaciona cada comorbidade a uma determinada pontuação ponderada para refletir o impacto dessas doenças na questão da mor-

talidade hospitalar. O terceiro atributo tem a utilização do Índice de Comorbidade de Charlson (ICC) [Charlson et al., 1987], com 19 comorbidades catalogadas. Esse parâmetro visa quantificar a carga de comorbidades em pacientes adultos e atribui peso de 1 a 6 para cada patologia. Por fim, o escore final utilizado é a soma dos pesos das comorbidades presentes.

Finalmente, mais dois atributos relacionados com a Escala de Coma de Glasgow foram criados, GCSInitial e GCSFinal. Eles estão relacionados com a soma total das três vertentes do indicador, resposta motora, resposta verbal e abertura dos olhos. O atributo GCSInitial corresponde a realização da primeira medição do exame e GCSFinal à última medição. Logo, se o paciente foi avaliado apenas uma vez dentro do período de 24 horas a soma total do exame é atribuída para ambos os atributos criados. Caso exista mais de uma medição, os elementos recebem seus respectivos resultados. Dessa forma, a Tabela 1 indica os atributos utilizados para a criação do modelo.

Label	Module	Feature
Anion Gap	Hosp	max,min,mode,last
Bicarbonate	Hosp	max,min,mode,last
Chloride	Hosp	max,min,mode,last
Creatinine	Hosp	max,min,mode,last
Glucose	Hosp	max,min,mode,last
Hematocrit	Hosp	max,min,mode,last
Hemoglobin	Hosp	max,min,mode,last
MCH	Hosp	max,min,mode,last
MCHC	Hosp	max,min,mode,last
MCV	Hosp	max,min,mode,last
Platelet Count	Hosp	max,min,mode,last
Potassium	Hosp	max,min,mode,last
RDW	Hosp	max,min,mode,last
Red Blood Cells	Hosp	max,min,mode,last
Sodium	Hosp	max,min,mode,last
Urea Nitrogen	Hosp	max,min,mode,last
White Blood Cells	Hosp	max,min,mode,last
Gender	Hosp	original
Anchor_age	Hosp	original
Death	Hosp	original
Readmitted	Novo	calculado
Readmitted_30	Novo	calculado
Readmitted_60	Novo	calculado
GCSInitial	Novo	calculado
GCSFinal	Novo	calculado
S_Elixhauser	Novo	calculado
Num_Comorbidades	Novo	calculado
Charlson	Novo	calculado
GCS - Eye Opening	ICU	max,min,mode,last
GCS - Motor Response	ICU	max,min,mode,last
GCS - Verbal Response	ICU	max,min,mode,last
Admission Weight (Kg)	ICU	original

Tabela 1. Atributos utilizados para predição

Na sequência, o tratamento dos dados faltantes foi realizado. Segundo Emmanuel et al. [2021], a frequência de valores ausentes em atributos utilizados para o aprendizado de máquina pode gerar análises imprecisas ou tendenciosas. Existem várias formas de realizar a gestão de valores ausentes, entre elas, a eliminação da coluna ou linha em que existe a falta de informação e a imputação do valor, por exemplo, com a média da respectiva coluna [Saar-Tsechansky and Provost, 2007]. Em vista da necessidade de realizar o preenchimento de tais lacunas no conjunto de dados, optou-se pela técnica de imputação avançada MICE (Multiple Imputation by Chained Equations), que é implementada pelo *IterativeImputer* no scikit-learn. Esse método atua com uma abordagem multivariada, ao tentar prever os valores ausentes com base nas informações fornecidas pelos outros atributos. Inicialmente um valor comum é atribuído em todos os campos nulos, no caso deste trabalho, o valor mais frequente foi empregado.

O segundo passo é estabelecer as relações entre os atributos de forma a detectar como eles se conectam com a informação que está faltando, por meio de relações estatísticas entre as variáveis do conjunto de dados. Nesse sentido, o algoritmo entra em processo iterativo, no qual cada variável com campos nulos é tratada de forma dependente dentro de um modelo de regressão, ao usar os demais atributos como recursos independentes. Esse processo se repete várias vezes, até que os valores praticamente não sofram mais alterações.

Esta abordagem de imputação se mostra mais eficiente, pois possibilita realizar análises mais complexas em razão da identificação de padrões mais complexos. Diante disso, o MICE pode ser implementado por meio de uma regressão linear ou árvores de decisão [Burgette and Reiter, 2010].

Esta abordagem de imputação se mostra mais eficiente, pois possibilita realizar análises mais complexas em razão da identificação de padrões mais complexos. Diante disso, o MICE pode ser implementado por meio de uma regressão linear ou árvores de decisão [Burgette and Reiter, 2010].

3.3 Construção do modelo

O aprendizado de máquina é uma subcategoria das áreas de inteligência artificial e estatística e foi definida por Mitchell [1997]. Para ele, os computadores são capazes de aprender a partir de experiências, ou seja, um conjunto de dados. Logo, quanto mais experiências ele recebe, maior é sua capacidade de desenvolver determinada tarefa e, conseqüentemente, sua capacidade de resolução é aprimorada empregando os padrões e regularidades encontrados nos conjuntos de dados de entrada.

O método escolhido para a realização do trabalho está centrado no aprendizado supervisionado de classificação com a utilização de modelos preditivos baseados em árvores de decisão, as quais são compostas por nodos que contêm regras responsáveis por prever resultados [Quinlan, 1996]. Nesse sentido, os algoritmos escolhidos para criar as ferramentas de previsões prévias são Random Forest, XGBoost, LightGBM e CatBoost. Após testes empíricos, apenas o XGBoost e o LightGBM foram utilizados em razão de um melhor desempenho, quando comparado com os demais.

A aprendizagem de máquina supervisionada é a mais utilizada em razão da facilidade na análise dos dados, uma vez que as instâncias são rotuladas, i.e., cada conjunto de dados é acompanhado pela resposta (rótulo) [Cunningham et al., 2008]. Esse método utiliza dois conjuntos de dados, um para treinamento e outro para testes. Assim, com os dados destinados ao treinamento, o algoritmo constrói um modelo capaz de analisar novos dados de entrada, ou seja, o conjunto de testes [Duarte and Ståhl, 2018].

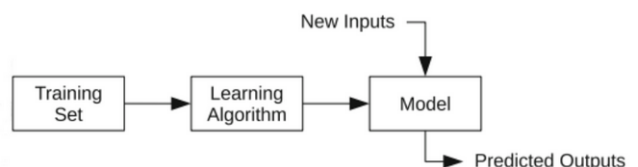


Figura 6. Representação do modelo preditor [Duarte and Ståhl, 2018]

A Figura 6 representa de forma gráfica como o modelo de aprendizado é organizado. Inicialmente, um conjunto de dados de treinamento é passado para o algoritmo, o qual é responsável por aprender como as informações se comportam, além do resultado esperado da manipulação desses dados, que são recebidos como rótulos (*labels*). Após o treinamento, um novo conjunto de dados de testes é enviado ao modelo preditivo. Este por sua vez deverá analisar os parâmetros e retornar uma saída que representa a predição para determinada tarefa. Assim, os rótulos que foram previstos são comparados com os rótulos verdadeiros, por meio de métricas, como acurácia, precisão e revocação, a fim de verificar o desempenho do modelo na tarefa de predição.

Portanto, para a construção do modelo, estabeleceu-se a divisão de 70% do conjunto de dados para treinamento e 30% para testes. Além disso, a técnica SMOTEENN, que consiste na junção do Synthetic Minority Over-sampling Technique e do Edited Nearest Neighbors, foi empregada a fim de balancear o conjunto de dados de treinamento e remover os exemplos ruidosos ou sobrepostos das classes. Além disso, em razão das múltiplas admissões hospitalares, as quais podem estar associadas ao mesmo paciente, optou-se por utilizar a técnica GroupShuffleSplit, para que as informações de um mesmo indivíduo aparecessem sempre no treino ou no teste, de forma a evitar a contaminação dos dados (*data leakage*).

Por sua vez, a técnica baseada na calibração isotônica foi utilizada em 20% dos dados do conjunto de treinamento para ajustar as probabilidades previstas. Essa estratégia corrige desvios que o modelo preditivo subestima ou superestima em relação à probabilidade de ocorrência da classe positiva, neste caso, o óbito hospitalar.

Por fim, outro aspecto relevante está associado à validação cruzada (*cross-validation*) realizada tanto para a escolha dos hiperparâmetros quanto para a calibração do modelo. Neste trabalho, optou-se por realizar uma validação cruzada de 3 dobras (*folder*). Por conseguinte, os dados do conjunto de treinamento foram subdivididos em três subconjuntos, destes, dois são usados para treinar o modelo e um para validar.

3.3.1 Extreme Gradient Boosting (XGBoost)

O XGBoost é uma técnica de aprendizado de máquina supervisionado que trabalha com o conceito de árvores de decisão. Inicialmente, uma árvore simples é criada para a predição. Após a análise do modelo construído, o erro de previsão é calculado e uma nova árvore é criada com o intuito de ajustar a predição e os erros. Nesse sentido, o algoritmo utiliza a técnica de *boosting*, pois esse processo ocorre inúmeras vezes e, ao final dele, todas as árvores criadas, que caracterizam-se por serem executadas sequencialmente, são consideradas para a obtenção do resultado final. Esse modelo de aprendizado de máquina é amplamente utilizado em razão da alta velocidade de processamento para um grande conjunto de dados [Chen and Guestrin, 2016].

3.3.2 LightGBM

É um algoritmos de aprendizado de máquina baseado em árvores de decisão e utilizado para problemas de classificação e regressão. Assim como o XGBoost [Chen and Guestrin, 2016], o LightGBM atua com a criação de várias árvores pe-

quenas de forma sequencial e tenta corrigir os erros das anteriores. Além disso, utiliza histogramas para agrupar as variáveis em faixas, com o intuito de encontrar o melhor ponto de divisão possível, GOSS (Gradient-based One-Side Sampling), que reduz a quantidade de dados utilizada em cada iteração, mantendo os mais importantes, e o EFB (Exclusive Feature Bundling), que junta variáveis que dificilmente assumem valores diferentes ao mesmo tempo. Outra diferença está centrada em como a árvore de decisão é criada. No LightGBM, a técnica *leaf-wise* (folha a folha) é aplicada em razão do algoritmo escolher sempre a divisão que apresenta a maior redução possível de erros [Ke et al., 2017].

3.4 Otimização do modelo

A seleção da melhor configuração possível de hiperparâmetros em modelos de aprendizagem de máquina desempenham um papel fundamental no impacto do desempenho dos modelos preditivos [Yang and Shami, 2020]. Em vista disso, para garantir um melhor desempenho, a técnica baseada em Grid-SearchCV foi empregada em todos os modelos, uma vez que a ferramenta garante que todas as potenciais combinações pré-definidas sejam testadas. A Tabela 2 mostra os melhores hiperparâmetros encontrados para cada modelo preditivo, hiperparâmetros padrões não foram inclusos na tabela:

Modelo	Hiperparâmetros
XGBoost	colsample_bytree = 0.9 learning_rate = 0.02 max_depth = 4 min_child_weight = 3 n_estimators = 600 scale_pos_weight = 1.0 subsample = 0.9
LightGBM	colsample_bytree = 0.7 learning_rate = 0.1 max_depth = -1 min_child_samples = 40 n_estimators = 200 num_leaves = 31 subsample = 0.8

Tabela 2. Melhores hiperparâmetros dos modelos

Para o modelo construído com o XGBoost, o hiperparâmetro *colsample_bytree* igual a 0.9 indica que 90% dos atributos escolhidos aleatoriamente serão utilizados para a construção da árvore. Já *learning_rate* indica o tamanho do passo dado em cada iteração, *max_depth* a profundidade máxima de cada árvore de decisão, *min_child_weight* utilizado para evitar que o modelo crie divisões com poucos dados. Por sua vez, *n_estimators* se refere a quantidade de árvores construídas, *scale_pos_weight* para ajustar o peso da classe em razão do desbalanceamento e *subsample* para definir a quantidade de linhas utilizada para treinar cada árvore.

A configuração do LightGBM tem um conjunto de hiperparâmetros similares ao XGBoost, mas com valores diferentes, são eles: *colsample_bytree*, *learning_rate*, *max_depth*, *n_estimators* e *subsample*. Além disso, existem hiperparâmetros diferentes, como *min_child_samples* utilizado para controlar o tamanho mínimo dos nós e *num_leaves* para indicar a quantidade de folhas que cada árvore pode ter.

Esses modelos são analisados por meio de métricas de desempenho, que são essenciais na avaliação de modelos de aprendizado de máquina em virtude de apresentar informações relacionadas com a performance de um modelo [Opitz, 2024]. Existem várias métricas para avaliar modelos de classificação, as mais comuns e apresentadas neste trabalho são: acurácia, precisão, recall, F1-score e curva ROC. Todas elas são baseadas na matriz de confusão.

A matriz de confusão é uma ferramenta fundamental na avaliação de modelos preditivos baseados em classificação. Com ela, é possível visualizar o desempenho dos modelos com base em acertos e erros, ou seja, é plausível identificar onde o modelo está errando, verificar se existe tendência em confundir as classes, além de auxiliar no ajuste de limiares [Sathyanarayanan and Tantri, 2024].

Classe Real \ Classe Prevista	Positiva	Negativa
Positiva	VP	FN
Negativa	FP	VN

VP (Verdadeiros Positivos): o modelo previu positivo e era realmente positivo.

FP (Falsos Positivos): o modelo previu positivo, mas era negativo.

FN (Falsos Negativos): o modelo previu negativo, mas era positivo.

VN (Verdadeiros Negativos): o modelo previu negativo e era realmente negativo.

A acurácia diz respeito à proporção de previsões corretas em relação ao total de previsões. Essa métrica é adequada para conjuntos de dados balanceados, mas pode ser enganosa em bancos de dados desbalanceados que apresentam poucas classes-alvo para a detecção.

$$A = \frac{VP + VN}{VP + VN + FP + FN} \tag{1}$$

A precisão (P) é a proporção entre a quantidade de respostas corretas entre todas as respostas positivas que foram preditas.

$$P = \frac{VP}{VP + FP} \tag{2}$$

A revocação (R), também conhecida como *recall*, é a proporção de casos positivos corretamente identificados pelo modelo, e pode ser calculada pela seguinte fórmula:

$$R = \frac{VP}{VP + FN} \tag{3}$$

O F1-score é uma métrica que diz respeito à média harmônica entre a precisão e a revocação, assim, ela busca encontrar um equilíbrio entre essas duas medidas. Pode ser calculada com base na seguinte fórmula:

$$F1 = 2 \times \frac{P}{P + R} \tag{4}$$

A curva ROC é uma métrica é amplamente utilizada para identificar a capacidade de separação do modelo entre a classe alvo e as demais. Ela é composta por uma linha central $x = y$, que indica a proporção igualitária de falsos positivos e verdadeiros positivos, como mostra a Figura 7. Além disso,

a área abaixo da linha azul é denominada de AUC (Area Under the Curve) e é esse indicador que define a capacidade de distinção entre as classes, quando a AUC assume valor igual a 0.5 significa que o modelo não possui capacidade discriminatória. Por outro lado, quando o valor 1 é encontrado, a ferramenta de previsão possui uma discriminação perfeita [Hoo et al., 2017]. Desta forma, uma AUC maior que 0.5 e menor ou igual a 0.6, indica uma discriminação pobre; uma AUC maior que 0.6 e menor ou igual a 0.7, indica uma discriminação aceitável; uma AUC maior que 0.7 e menor ou igual a 0.8, indica uma boa discriminação; uma AUC maior que 0.8 e menor ou igual a 0.9, indica uma discriminação excelente; e valores acima de 0.9, indicam uma discriminação notável ou quase perfeita [Yang and Berdine, 2017].

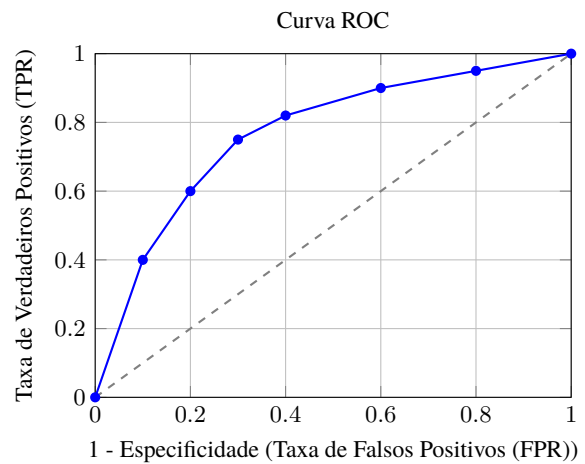


Figura 7. Curva ROC ilustrativa

4 Resultados

Esta seção apresenta os resultados obtidos para os modelos de predição XGBoost e LightGBM, os quais apresentaram um melhor desempenho. A matriz de confusão resultante do XGBoost apresentou 620 casos verdadeiros negativos, 77 falsos positivos, 32 falsos negativos e 125 verdadeiros positivos. Já para o LightGBM, 636 verdadeiros negativos, 61 falsos positivos, 44 falsos negativos e 113 verdadeiros positivos. A Tabela 3 apresenta os valores obtidos para cada métrica avaliada com base nas informações listadas.

Métrica	XGBoost	LightGBM
Acurácia	0.88	0.88
Precisão	0.62	0.65
Recall	0.80	0.72
F1-Score	0.70	0.68
AUC	0.915	0.911

Tabela 3. Desempenho dos modelos nas métricas avaliadas

Percebe-se que os modelos XGBoost e LightGBM obtiveram um desempenho parecido em todas as métricas avaliadas. Uma acurácia de 0.88 indica que os modelos preditivos erraram apenas 12% das previsões totais. A precisão do modelo construído com o LightGBM demonstra que dentre todos os pacientes que faleceram, 65% foram classificados

corretamente, enquanto esse valor ficou em 62% no XGBoost. Além disso, o modelo baseado no XGBoost apresentou uma pequena vantagem em recall e F1-Score, o que indica uma melhora na identificação de casos positivos, quando comparado com a outra ferramenta preditiva. Por fim, a AUC de 0.915 para o XGBoost e uma AUC de 0.911 para o LightGBM indicam que os modelos possuem uma grande capacidade de distinção entre as classes, como mostra a Figura 8.

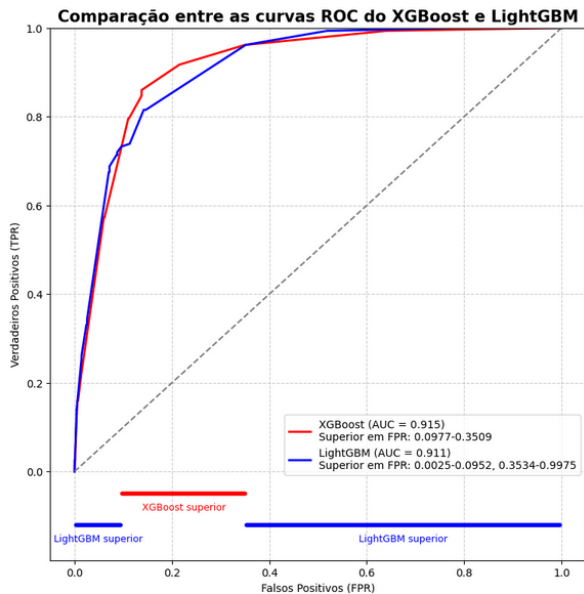


Figura 8. Representação comparativa da curva ROC entre XGBoost e LightGBM

A representação gráfica das curvas ROC de ambos os modelos demonstram que resultados semelhantes foram obtidos. Isso ocorre porque ao tentar capturar uma quantidade maior de exemplos positivos, o modelo passa a incluir mais falsos positivos na previsão. Durante o intervalo da taxa de falsos positivos de 0.003 a 0.095 e 0.353 a 0.997 o modelo preditivo baseado no LightGBM (Linha azul e indicador na cor azul abaixo do Eixo X) foi ligeiramente superior em relação à discriminação das classes. Por outro lado, nos intervalos de 0.098 a 0.351 o XGBoost (Linha vermelha e indicador na cor vermelha abaixo do Eixo X) se mostrou um pouco mais eficiente na questão da sensibilidade para a taxa de falsos positivos. Isso indica que os modelos baseados tanto no XGBoost, quanto no LightGBM atuam de forma complementar no intervalo de 0.003 a 0.997. Portanto, ocorrem momentos em que um modelo preditivo é mais eficaz na identificação de casos positivos, quando comparados com os negativos.

Por fim, nos intervalos de 0.000 a 0.002, 0.096 a 0.097, 0.998 a 1 e no ponto 0.352, ambos os modelos apresentaram resultados iguais na distinção das classes. Dessa forma, segundo a classificação apresentada por Yang and Berdine [2017], este trabalho alcançou uma discriminação notável com base nos resultados da curva ROC ao obter um desfecho superior a 0.9 para ambos os modelos preditivos.

Para além, cabe mencionar a análise das curvas Precision-Recall de ambos os modelos. Essas curvas descrevem a relação entre a precisão e a recall, ao aumentar o recall, normalmente, a precisão diminui.

Essa informação é normalmente utilizada em conjun-

tos de dados desbalanceados nos quais a classe positiva tende a apresentar uma ocorrência reduzida. Nesses casos, a curva ROC pode apresentar resultados otimistas, enquanto a curva Precision-Recall fornece uma avaliação mais condizente com o real desempenho obtido pelo modelo na identificação da classe minoritária, neste caso, a classe alvo [Ozenne et al., 2015]. De forma igual à curva ROC, a curva Precision-Recall tem como valor máximo 1, esse resultado é indicado pela *Average Precision* (AP), que é a área sob a curva. A Figura 9 mostra uma análise comparativa entre os modelos, na qual o XGBoost apresentou uma AP = 0.6596 e o LightGBM uma AP = 0.6602.

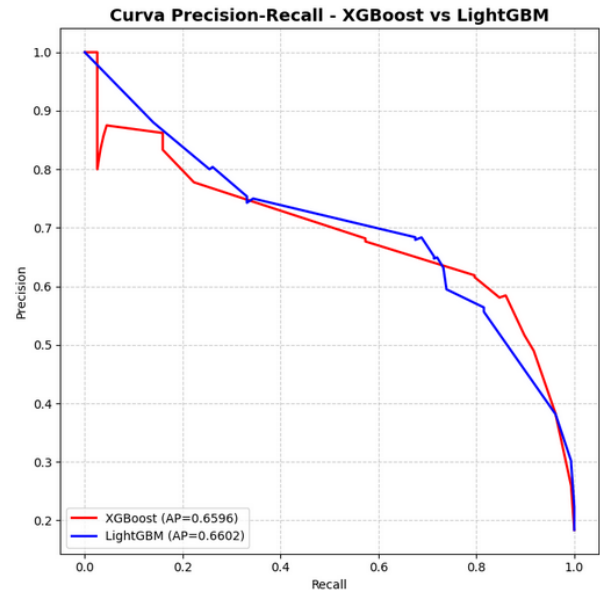


Figura 9. Representação comparativa da curva Precision-Recall entre XGBoost e LightGBM

Outro ponto importante a ser analisado se refere às curvas de calibração dos modelos. Esse indicador compara as probabilidades previstas (Eixo X) pelo modelo, isto é, a chance do modelo preditivo classificar um paciente como falecido, com as probabilidades observadas (Eixo Y), ou seja, a real frequência dos eventos, neste caso, os óbitos.

A Figura 10 apresenta as curvas de calibração para os modelos baseados no XGBoost e LightGBM. É possível observar que a linha pontilhada de 45° indica uma calibração perfeita, ou seja, as probabilidades previstas coincidem exatamente com a frequência observada. Para casos em que a marcação ocorre abaixo da reta central, o modelo preditivo tende a superestimar o risco de morte. De forma oposta, marcações acima da reta central indicam que o modelo subestima o risco.

Nesse sentido, percebe-se que o modelo baseado no XGBoost tende a subestimar o risco de óbitos em baixas e médias probabilidades, ou seja, prediz valores mais baixos do que os observados. No entanto, a partir de 0.6 o cenário se inverte e começa a superestimar. De maneira contrária, o modelo preditivo baseado no LightGBM tende a superestimar em praticamente todo o intervalo, apenas entre 0.2-0.4 e próximo a 0.6 subestima.

Logo, com base nas informações apresentadas na Figura 10, percebe-se que ambos os modelos apresentaram suas curvas próximo da reta de calibração perfeita. Esse

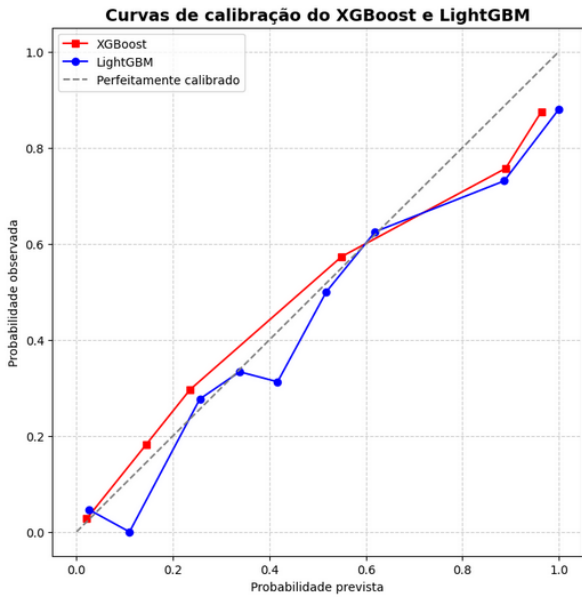


Figura 10. Representação comparativa da curva de calibração entre XGBoost e LightGBM

resultado, conseqüentemente, indica que as probabilidades previstas pelos modelos preditivos possuem uma boa correspondência com as reais probabilidades do paciente ir a óbito.

Outra análise realizada se refere ao gráfico de decisão clínica, apresentado na Figura 11. Esse diagrama é utilizado para avaliar a utilidade prática do modelo preditivo. O eixo X apresenta o limiar de decisão, ou seja, a probabilidade mínima para realizar o tratamento. Valores baixos para esse indicador apontam alta sensibilidade, o que significa que é possível realizar o tratamento mesmo com uma baixa chance de morte. Já valores altos evidenciam uma alta especificidade, o que denota que o tratamento só deve ser feito quando existe uma alta probabilidade de morte do paciente. Por outro lado, o eixo Y apresenta o benefício líquido do modelo. Esse indicador aponta quando o modelo preditivo manifesta ganhos em relação a tratar todos os pacientes ou não tratar ninguém.

Além disso, para a construção do gráfico, a fórmula do benefício líquido apresentada por Vickers and Elkin [2006] foi utilizada para descrever o comportamento dos modelos e a estratégia de "Tratar todos":

$$NB(p_t) = \frac{TP}{N} - \frac{FP}{N} \cdot \frac{p_t}{1 - p_t} \quad (5)$$

- TP : Número de verdadeiros positivos.
- FP : Número de falsos positivos.
- N : Tamanho total da amostra.
- p_t : Probabilidade-limiar, ou seja, é a menor probabilidade de risco a partir da qual uma intervenção é considerada justificável.
- $\frac{TP}{N}$: Proporção de indivíduos corretamente identificados como positivos.
- $\frac{FP}{N}$: Proporção de indivíduos incorretamente identificados como positivos.
- $\frac{p_t}{1 - p_t}$: Representa o custo relativo de um falso positivo em relação a um verdadeiro positivo.
- $NB(p_t)$: Benefício líquido (Net Benefit), que combina ganhos (TP) e perdas ponderadas (FP) para avaliar o valor clínico do modelo em cada limiar p_t .

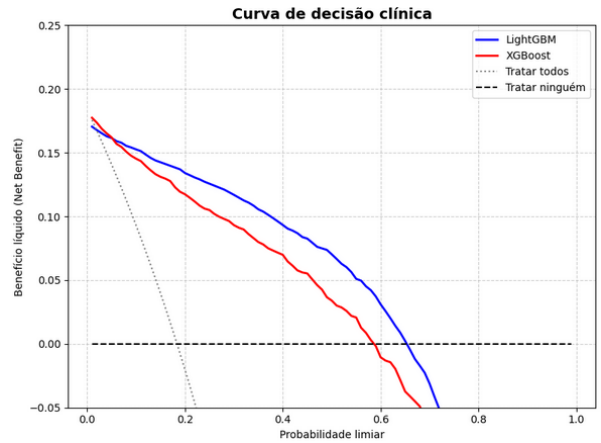


Figura 11. Representação comparativa da curva de benefício entre XGBoost e LightGBM

Sob essa perspectiva, quando a curva do modelo estiver acima de "Tratar todos" (linha pontilhada mais fina) e "Tratar ninguém" (linha pontilhada mais grossa) existe um ganho clínico associado. Em contrapartida, se a curva do modelo preditivo estiver abaixo desses indicadores, não existe vantagem correlacionada. Deste modo, o modelo só é útil clinicamente nas probabilidades limiares na qual a curva se encontra acima das estratégias simples. Diante disso, ao avaliar o desempenho do XGBoost e do LightGBM, verifica-se que ambos os modelos preditivos apresentam ganhos em baixos e médios limiares. No entanto, a partir do limiar 0.6, uma queda é evidenciada para o XGBoost, e entre 0.6-0.8 para o LightGBM, o que garante perda de ganho associado.

Além disso, uma análise para a verificação dos atributos com maiores impactos preditivos foi realizada. Nesse sentido, para tal tarefa, utilizou-se a técnica SHAP (SHapley Additive exPlanations), que é um método de interpretação de modelos de aprendizado de máquina baseado na teoria dos valores de Shapley no contexto da teoria dos jogos cooperativos. Cada variável é considerada como um jogador que contribui positivamente ou negativamente para o resultado final, assim, o SHAP calcula qual o grau de influência de cada atributo na predição [Lundberg and Lee, 2017].

Para o modelo baseado no XGBoost, a Figura 12 mostra as 20 variáveis no eixo Y com maiores impactos na previsão do óbito de pacientes. Em ordem decrescente, tem-se: GCSFinal, GCS - Motor Response_MAX, GCS - Eye Opening_MODE, Glucose_MAX, GCS - Motor Response_MODE, GCS - Motor Response_MIN, GCS - Eye Opening_MIN, S_Elixhauser, Potassium_MAX, Admission Weight (Kg), GCS - Eye Opening_MAX Hemoglobin_MIN, anchor_age, Glucose_MODE, Platelet Count_MIN, White Blood Cells_MAX, Hemoglobin_MODE, Creatinine_LAST, gender e Platelet Count_MODE.

De forma adicional, no eixo X, é apresentado o valor do SHAP atribuído a cada variável na predição. Valores positivos indicam que o atributo aumenta a probabilidade de ocorrência do evento; já valores negativos diminuem essa probabilidade. Ademais, a cor vermelha informa que o elemento possui um valor alto e azul um valor baixo. Quando existem muitos pontos sobrepostos, é possível verificar a densidade das informações: faixas largas demonstram que o atributo possui um comportamento variável; já faixas estreitas mostram um efeito mais estável da informação.

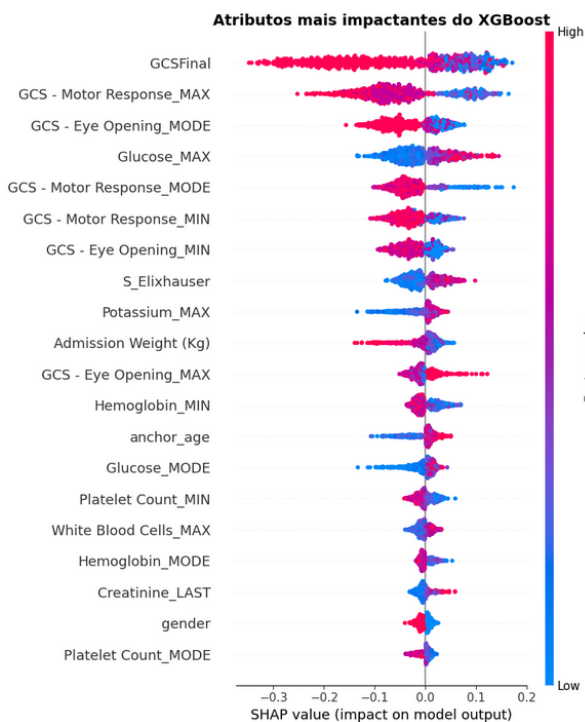


Figura 12. Representação dos atributos com maiores impactos preditivos do XGBoost

Em conformidade ao exposto, se os pontos vermelhos estão mais à direita, valores altos para o atributo aumentam a chance do evento acontecer. Caso contrário, se estão no lado esquerdo, valores altos diminuem essa chance. Logo, é possível identificar que valores altos para **GCSFinal** tendem a diminuir a probabilidade do óbito; já valores baixos indicam que o paciente tem maiores chances de morte. Um ponto importante a ser destacado está atrelado ao atributo **Glucose**, valores altos obtidos na última medição, nas primeiras 24 horas, após a admissão hospitalar, mostraram que as chances de óbito aumentam. **S_Elixhauser** e **anchor_age**; também atuam em conformidade: pacientes mais velhos tendem a possuir maior probabilidade de óbito, assim como aqueles que possuem comorbidades mais graves. Outra característica visível é que a grande maioria dos atributos se encontram em faixas estreitas, indicando que existe uma relação mais consistente entre os valores.

Em razão da idade influenciar as chances de óbito, é possível realizar uma análise com esse atributo, como mostra a Figura 13. O gráfico A apresenta que as probabilidades de um paciente falecer em admissões por AVC hemorrágico tendem a crescer com o passar dos anos. Assim, percebe-se que indivíduos com idade superior aos 60 anos possuem um nível maior de risco. Em contrapartida, existem idades em que a probabilidade de óbito é praticamente zero. A suavização entre as idades não ocorre pois a amostra empregada possui poucos dados por idade inteira, ou seja, ao agrupar todos os pacientes com a mesma idade e calcular a média de probabilidade, cada idade pode estar associada a um pequeno grupo de indivíduos e à uma grande variabilidade nos dados.

Já no gráfico B, é realizada uma comparação entre as chances de óbito em relação à idade para os gêneros feminino e masculino. É possível observar, em linhas gerais, que as mulheres são mais suscetíveis a um maior grau de risco, principalmente até os 35 anos, entre 40 e 50 anos e entre 60 e

70 anos. Por outro lado, próximo aos 40 e 76 anos os homens apresentam uma maior probabilidade de óbito.

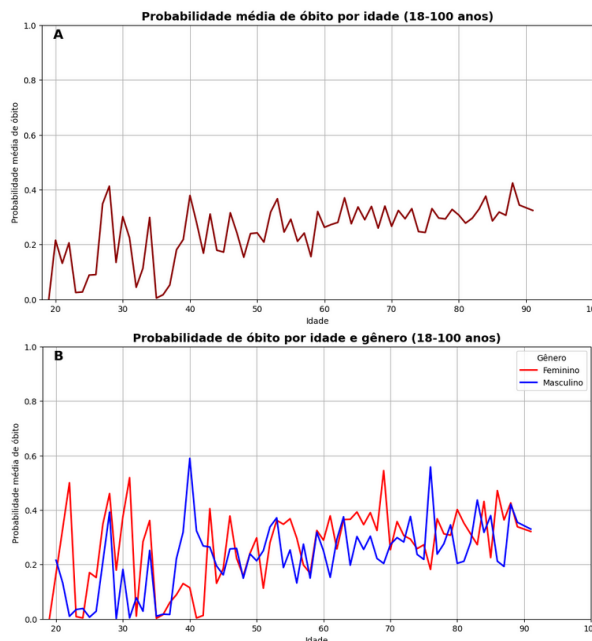


Figura 13. Análise da idade nas probabilidades de óbito por AVC hemorrágico, XGBoost

Para o modelo baseado no LightGBM, a Figura 14 mostra, em ordem decrescente, os 20 atributos mais impactantes no resultado final: **GCSFinal**, **GCS - Motor Response_MAX**, **Glucose_MAX**, **GCS - Motor Response_MIN**, **S_Elixhauser**, **Admission Weight (Kg)**, **GCS - Eye Opening_MODE**, **White Blood Cells_MAX**, **Glucose_MODE**, **GCS - Motor Response_LAST**, **GCS - Motor Response_MODE**, **Platelet Count_MAX**, **GCS - Eye Opening_MAX**, **Platelet Count_MIN**, **GCSInitial**, **Hemoglobin_MIN**, **anchor_age**, **GCS - Eye Opening_LAST**, **MCV_MODE** e **Bicarbonate_LAST**.

De forma semelhante aos atributos do XGBoost, o modelo preditivo com base no LightGBM detectou o atributo **GCSFinal** como o mais importante. Neste caso, valores altos para esse indicador demonstram que a chance de ocorrer o óbito é menor; o mesmo é aplicado para os outros recursos relacionados com a Escala de Coma de Glasgow a exceção de **GCS - Eye Opening_LAST**, em razão de valores altos também influenciarem o óbito. Para os atributos **anchor_age**, **S_Elixhauser**, **Glucose_MAX**, **White Blood Cells_MAX**, **Glucose_MODE**, **MCV_MODE** e **Bicarbonate_LAST**, valores baixos indicam uma probabilidade menor de ocorrer o óbito.

De modo semelhante ao modelo baseado no XGBoost, a Figura 15 mostra a relação entre as probabilidades de óbito por idade e por gênero do LightGBM. O gráfico A indica que, com o passar dos anos, a probabilidade de óbito tende a aumentar, além de apresentar informações de que os indivíduos entre 40 anos e 60 anos possuem chances de óbito que oscilam entre 15% a praticamente 40%. Já o gráfico B, mostra que as mulheres possuem uma maior probabilidade de falecer em decorrência de um AVC hemorrágico, quando comparadas com os homens.

5 Discussão

Este trabalho faz parte de um pequeno grupo de estudos produzidos com o intuito de verificar as probabilidades de óbitos de pacientes admitidos em unidades hospitalares com AVC hemorrágico. Algumas limitações foram encontradas durante a realização do projeto, sendo as principais delas a falta de dados relacionados com os sinais vitais no momento da admissão e a falta de informações relacionadas às admissões dos pacientes que podem ou não ter sido encaminhados à UTI. Desta forma, o trabalho contribui para a rápida identificação de desfechos clínicos ao utilizar dados coletados nas primeiras 24 horas.

Assim, para a realização deste trabalho, é válido ressaltar que o maior empecilho encontrado diz respeito à falta de informações relacionadas com os sinais vitais e exames laboratoriais dos pacientes. Com base nisso, o presente estudo não fez uso de sinais vitais para a criação dos modelos preditivos em razão de tais indicadores serem coletados apenas na UTI e não estarem presentes no módulo com as informações iniciais da hospitalização. Observações faltantes em conjuntos de dados médicos, como o MIMIC - IV, são normais em virtude de não existir um monitoramento eletrônico nas primeiras etapas dos atendimentos, problema que seria minimizado com a utilização de totens para a medição de indicadores fisiológicos distribuídos em unidades de saúde. Além disso, valores de exames podem ser omitidos nas fichas técnicas dos pacientes. Em vista disso, o tratamento de tais lacunas no conjunto de dados é necessário, para garantir que os modelos consigam criar uma generalização adequada para a correta classificação.

Outro ponto importante a ser destacado se refere ao balanceamento do conjunto de treinamento com a utilização do SMOTEENN, o que ajuda o modelo ao criar novas instâncias sintéticas da classe minoritária e eliminar os exemplos ruidosos dos dados para garantir que a classe majoritária não cause inconsistências na predição.

Com base nos resultados obtidos, relacionados aos modelos baseados no XGBoost e LightGBM, foi possível identificar um grande poder de discriminação entre as classes, $AUC = 0.915$ e $AUC = 0.911$, respectivamente, para prever a mortalidade de pacientes admitidos na unidade hospitalar com AVC hemorrágico. O presente trabalho superou as AUCs obtidas por todos os trabalhos relacionados ao estudo, isso aponta que a utilização de múltiplas admissões associadas ao paciente influenciam de maneira positiva no treinamento do modelo. A Tabela 4 demonstra essa análise comparativa.

Autor	Versão MIMIC	Modelo	AUC
Miao et al. [2024]	III e IV	LightGBM	0.736
Chen et al. [2025]	IV	XGBoost	0.797
Mao et al. [2024]	IV	XGBoost	0.907
Li et al. [2025]	III e IV	AdaBoost	0.877
Este trabalho	IV	XGBoost	0.915
Este trabalho	IV	LightGBM	0.911

Tabela 4. Desempenho comparativo dos modelos preditivos

Por sua vez, a utilização da técnica SHAP foi imprescindível para a identificação dos atributos que mais impactam na ocorrência de óbitos relacionados com acidente vas-

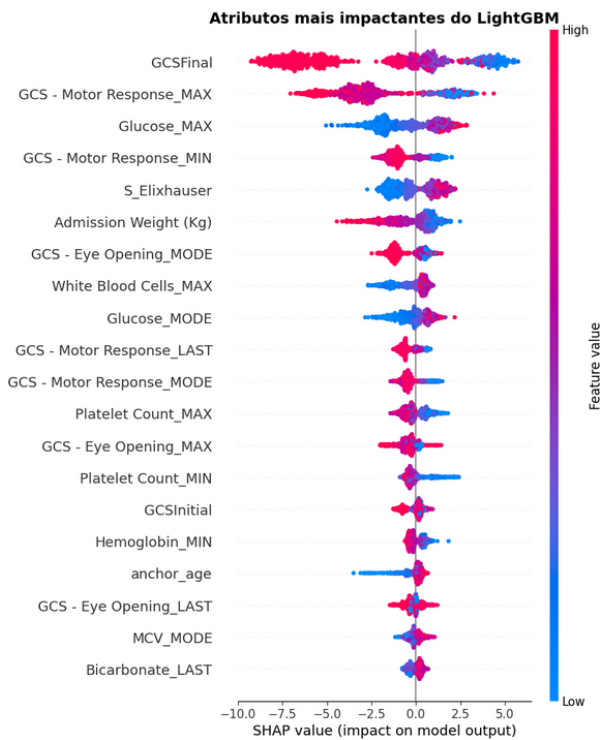


Figura 14. Representação dos atributos com maiores impactos preditivos do LightGBM

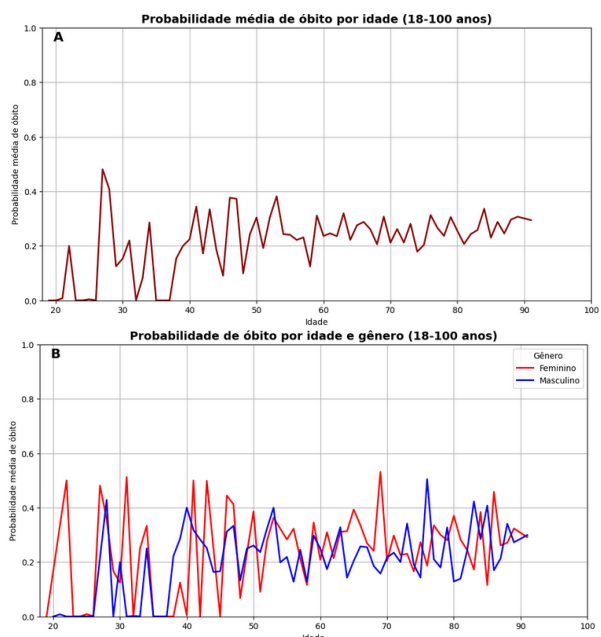


Figura 15. Análise da idade nas probabilidades de óbito por AVC hemorrágico, LightGBM

cular cerebral hemorrágico. Essa análise apontou que, para o XGBoost, os cinco indicadores mais preditivos em ordem decrescente são: GCSFinal, GCS - Motor Response_MAX, GCS - Eye Opening_MODE, Glucose_MAX e GCS - Motor Response_MODE. Já para o LightGBM tem-se: GCSFinal, GCS - Motor Response_MAX, Glucose_MAX, GCS - Motor Response_MIN e S_Elixhauser.

A Escala de Coma de Glasgow foi criada com o objetivo de padronizar a avaliação em pacientes com comprometimento neurológico, auxiliando na identificação da gravidade da lesão e orientando os profissionais da saúde no tratamento desses indivíduos [Maserati *et al.*, 2016]. Neste trabalho, assim como nos demais estudos relacionados, a Escala de Coma de Glasgow apresentou uma grande relevância nos modelos preditivos, principalmente no que se refere ao indicador de resposta motora do paciente que, ao apresentar valores altos, indica que tal atributo impacta negativamente na ocorrência do óbito por AVC hemorrágico.

Outro aspecto está atrelado às relações da idade e as probabilidades de morte de pacientes. Um trabalho produzido por Wang *et al.* [2022] indicou que as chances de óbito aumentam com a idade, em especial, a partir dos 40 anos. Em consonância com tal estudo, o presente projeto, por meio da análise SHAP dos atributos do XGBoost e do LightGBM, também identificou essa relação, ao apresentar que idades mais elevadas aumentam a probabilidade de falecer em decorrência de um AVC hemorrágico. Além disso, foi possível identificar que as mulheres apresentaram uma taxa de mortalidade maior do que os homens em praticamente todas as idades anteriores aos 50 anos. Por outro lado, em idade mais avançadas, os homens possuíram uma maior taxa de mortalidade.

Tendo como base os modelos de aprendizado de máquina supervisionados desenvolvidos neste trabalho, percebe-se que a utilização da inteligência artificial no campo médico é capaz de gerar inúmeros benefícios no tratamento de doentes. Nessa perspectiva, isso é verificável ao encontrar possíveis desfechos no quadro clínico dos pacientes, o que, consequentemente, auxilia na alocação de recursos e no tratamento dos enfermos. No entanto, urge a necessidade de proteger esses dados sensíveis para que não ocorra vazamentos e prejuízos à sociedade. Desse modo, estudos futuros ainda são necessários, principalmente, no que se refere ao uso de indicadores fisiológicos e na aplicação real em ambientes hospitalares dos modelos de aprendizado de máquina.

6 Conclusão

Este trabalho teve como objetivo propor uma abordagem para implementar um modelo de previsão para pacientes acometidos de AVC hemorrágico que vão a óbito com o resultados dos exames coletados nas primeiras 24 horas da admissão hospitalar.

A incompletude de dados clínicos é um dos principais fatores que afeta o desempenho de generalização dos modelos de aprendizado de máquina. Para mitigar esse problema é necessário padronizar os protocolos de registro nas unidades de saúde, além de usar sistemas eletrônicos que preenchem campos críticos.

Os modelos XGBoost e LightGBM obtiveram os me-

lhores desempenhos nessa tarefa de predição, com valores de curva ROC de 0.915 e 0.911, respectivamente. Além disso, foi possível identificar que fatores relacionados com a Escala de Coma de Glasgow, comorbidades, idade e exames iniciais se mostraram com um forte impacto na predição do risco de mortalidade para ambos os modelos. Por fim, espera-se que as estruturas preditivas desenvolvidas sejam postas em prática nas unidades de saúde.

Trabalhos futuros devem incorporar mais variáveis clínicas, como sinais vitais e exames de imagem. Além disso, outros algoritmos de aprendizado de máquina, como redes neurais profundas e modelos híbridos podem ser avaliados. A necessidade de otimizar o balanceamento das classes com técnicas mais robustas como SMOTE-Tomek, bem como o tratamento de dados faltantes com outras ferramentas é viável para a obtenção e validação de novos resultados. Outro aspecto está relacionado com a utilização de outras bases de dados para a validação externa do modelo. Finalmente, verificar a aplicabilidade do modelo em diferentes contextos clínicos.

Referências

- Burgette, L. F. and Reiter, J. P. (2010). Multiple imputation for missing data via sequential regression trees. *American journal of epidemiology*, 172(9):1070–1076. DOI: <https://doi.org/10.1093/aje/kwq260>.
- Charlson, M. E., Pompei, P., Ales, K. L., and MacKenzie, C. R. (1987). A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *Journal of chronic diseases*, 40(5):373–383. DOI: [doi.org/10.1016/0021-9681\(87\)90171-8](https://doi.org/10.1016/0021-9681(87)90171-8).
- Chen, T. and Guestrin, C. (2016). Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, pages 785–794. DOI: doi.org/10.1145/2939672.2939785.
- Chen, Y., Liu, X., Yan, M., and Wan, Y. (2025). Death risk prediction model for patients with non-traumatic intracerebral hemorrhage. *BMC Medical Informatics and Decision Making*, 25(1):35. DOI: [10.1186/s12911-025-02865-4](https://doi.org/10.1186/s12911-025-02865-4).
- Cunningham, P., Cord, M., and Delany, S. J. (2008). Supervised learning. In *Machine learning techniques for multimedia: case studies on organization and retrieval*, pages 21–49. Springer. DOI: doi.org/10.1007/978-3-540-75171-7.
- Duarte, D. and Ståhl, N. (2018). Machine learning: a concise overview. *Data Science in Practice*, pages 27–58. DOI: doi.org/10.1007/978-3-319-97556-6_3.
- Elixhauser, A., Steiner, C., Harris, D. R., and Coffey, R. M. (1998). Comorbidity measures for use with administrative data. *Medical care*, 36(1):8–27. DOI: doi.org/10.1097/00005650-199801000-00004.
- Emmanuel, T., Maupong, T., Mpoeleng, D., Semong, T., Mphago, B., and Tabona, O. (2021). A survey on missing data in machine learning. *Journal of Big data*, 8(1):140. DOI: doi.org/10.1186/s40537-021-00516-9.
- Faghih-Jouybari, M., Raof, M. T., Abdollahzade, S., Jamshidi, S., Padegane, T., Ehteshami, S., and Fateh, S.

- (2021). Mortality and morbidity in patients with spontaneous intracerebral hemorrhage: a single-center experience. *Current Journal of Neurology*, 20(1):32. DOI: doi.org/10.18502/cjn.v20i1.6377.
- Feigin, V. L., Abate, M. D., Abate, Y. H., Abd ElHafeez, S., Abd-Allah, F., Abdelalim, A., Abdelkader, A., Abdelmasseh, M., Abd-Elsalam, S., Abdi, P., et al. (2024). Global, regional, and national burden of stroke and its risk factors, 1990–2021: a systematic analysis for the global burden of disease study 2021. *The Lancet Neurology*, 23(10):973–1003. DOI: doi.org/10.1016/S1474-4422(24)00369-7.
- Hendriksen, J. M., Geersing, G.-J., Moons, K. G., and de Groot, J. A. (2013). Diagnostic and prognostic prediction models. *Journal of Thrombosis and Haemostasis*, 11:129–141. DOI: doi.org/10.1111/jth.12262.
- Hoo, Z. H., Candlish, J., and Teare, D. (2017). What is an roc curve?
- Johnson, A. E., Bulgarelli, L., Shen, L., Gayles, A., Shammout, A., Horng, S., Pollard, T. J., Hao, S., Moody, B., Gow, B., et al. (2023). MIMIC-IV, a freely accessible electronic health record dataset. *Scientific data*, 10(1):1. DOI: doi.org/10.1038/s41597-022-01899-x.
- Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., and Liu, T.-Y. (2017). LightGBM: a highly efficient gradient boosting decision tree. In *Proceedings of the 31st International Conference on Neural Information Processing Systems, NIPS'17*, page 3149–3157, Red Hook, NY, USA. Curran Associates Inc.. DOI: dl.acm.org/doi/10.5555/3294996.3295074.
- Li, H., Monger, R., Pishgar, E., and Pishgar, M. (2025). ICU readmission prediction for intracerebral hemorrhage patients using MIMIC III and MIMIC IV databases. *medRxiv*, pages 2025–01. DOI: doi.org/10.1101/2025.01.01.25319859.
- Lieh-Lai, M. W., Theodorou, A. A., Sarnaik, A. P., Meert, K. L., Moylan, P. M., and Canady, A. I. (1992). Limitations of the Glasgow coma scale in predicting outcome in children with traumatic brain injury. *The Journal of Pediatrics*, 120(2):195–199. DOI: doi.org/10.1016/S0022-3476(05)80426-3.
- Lundberg, S. M. and Lee, S.-I. (2017). A unified approach to interpreting model predictions. In *Proceedings of the 31st International Conference on Neural Information Processing Systems, NIPS'17*, page 4768–4777, Red Hook, NY, USA. Curran Associates Inc.. DOI: 10.5555/3295222.3295230.
- Mao, B., Ling, L., Pan, Y., Zhang, R., Zheng, W., Shen, Y., Lu, W., Lu, Y., Xu, S., Wu, J., et al. (2024). Machine learning for the prediction of in-hospital mortality in patients with spontaneous intracerebral hemorrhage in intensive care unit. *Scientific Reports*, 14(1):14195. DOI: doi.org/10.1038/s41598-024-65128-8.
- Maserati, M., Fetzick, A., and Puccio, A. (2016). The Glasgow coma scale (GCS): deciphering the motor component of the GCS. *Journal of Neuroscience Nursing*, 8(6):311–314. DOI: doi.org/10.1097/JNN.0000000000000242.
- Miao, J., Zuo, C., Cao, H., Gu, Z., Huang, Y., Song, Y., and Wang, F. (2024). Predicting ICU readmission risks in intracerebral hemorrhage patients: Insights from machine learning models using MIMIC databases. *Journal of the Neurological Sciences*, 456:122849. DOI: doi.org/10.1016/j.jns.2023.122849.
- Mitchell, T. M. (1997). *Machine Learning*. McGraw-Hill Science/Engineering/Math, New York. Hardcover, 432 pages.
- Molinnus, D., Beulertz, M., Bickenbach, J., Marx, G., and Benstoen, C. (2024). Observational study of missing sofa score data frequency in RCTs relative to ICU length of stay. *Scientific Reports*, 14(1):16160. DOI: doi.org/10.1038/s41598-024-67089-4.
- Opitz, J. (2024). A closer look at classification evaluation metrics and a critical reflection of common evaluation practice. *Transactions of the Association for Computational Linguistics*, 12:820–836. DOI: doi.org/10.1162/tacl_a_00675.
- Ozenne, B., Subtil, F., and Maucort-Boulch, D. (2015). The precision–recall curve overcame the optimism of the receiver operating characteristic curve in rare diseases. *Journal of Clinical Epidemiology*, 68(8):855–859. DOI: doi.org/10.1016/j.jclinepi.2015.02.010.
- Paixão, G. M. d. M., Santos, B. C., Araujo, R. M. d., Ribeiro, M. H., Moraes, J. L. d., and Ribeiro, A. L. (2022). Machine learning na medicina: revisão e aplicabilidade. *Arquivos Brasileiros de Cardiologia*, 118(1):95–102. DOI: doi.org/10.36660/abc.20200596.
- Quinlan, J. R. (1996). Learning decision tree classifiers. *ACM Comput. Surv.*, 28(1):71–72. DOI: 10.1145/234313.234346.
- Qureshi, A. I., Mendelow, A. D., and Hanley, D. F. (2009). Intracerebral haemorrhage. *The Lancet*, 373(9675):1632–1644. DOI: doi.org/10.1016/s0140-6736(09)60371-8.
- Rajashakar, D. and Liang, J. W. (2023). Intracerebral hemorrhage. In *StatPearls [Internet]*. StatPearls Publishing, Treasure Island (FL). [Atualizado em 6 de fevereiro de 2023; citado em 11 de novembro de 2025.].
- Ribeiro, J. K., dos Santos, H. D., Barletta, F., da Silva, M. C., Vieira, R., Morales, H. M., and da Costa Rocha, C. (2020). A machine learning early warning system: multicenter validation in Brazilian hospitals. In *2020 IEEE 33rd International Symposium on Computer-Based Medical Systems (CBMS)*, pages 321–326. IEEE. DOI: doi.org/10.1109/CBMS49503.2020.00067.
- Saar-Tschekansky, M. and Provost, F. (2007). Handling missing values when applying classification models. *J. Mach. Learn. Res.*, 8:1623–1657. DOI: 10.5555/1314498.1314553.
- Sapra, A., Malik, A., and Bhandari, P. (2023). Vital sign assessment. In *StatPearls [internet]*. StatPearls Publishing. [Atualizado em 1 de maio de 2023; consultado em 11 de novembro de 2025.].
- Sathyanarayanan, S. and Tantri, B. R. (2024). Confusion matrix-based performance evaluation metrics. *African Journal of Biomedical Research*, 27(4S):4023–4031. DOI: doi.org/10.53555/AJBR.v27i4S.4345.
- Shehab, M., Abualigah, L., Shambour, Q., Abu-Hashem, M. A., Shambour, M. K. Y., Alsalihi, A. I., and Gandomi, A. H. (2022). Machine learning in medical applications: A review of state-of-the-art methods.

- Computers in Biology and Medicine*, 145:105458. DOI: doi.org/10.1016/j.compbiomed.2022.105458.
- Tian, Y., Zhuang, Z., Su, M., Yao, X., Wang, X., and Li, G. (2025). Predictive value of seven scoring systems for 28-day mortality in icu patients with sleep apnea–hypopnea syndrome (sahs) and clinical indicators. *European Journal of Medical Research*, 30(1):856. DOI: doi.org/10.1186/s40001-025-03120-2.
- Van Asch, C. J., Luitse, M. J., Rinkel, G. J., van der Tweel, I., Algra, A., and Klijn, C. J. (2010). Incidence, case fatality, and functional outcome of intracerebral haemorrhage over time, according to age, sex, and ethnic origin: a systematic review and meta-analysis. *The Lancet Neurology*, 9(2):167–176. DOI: doi.org/10.1016/s1474-4422(09)70340-0.
- Van Smeden, M., Reitsma, J. B., Riley, R. D., Collins, G. S., and Moons, K. G. (2021). Clinical prediction models: diagnosis versus prognosis. *Journal of clinical epidemiology*, 132:142–145. DOI: doi.org/10.1016/j.jclinepi.2021.01.009.
- Van Walraven, C., Austin, P. C., Jennings, A., Quan, H., and Forster, A. J. (2009). A modification of the elixhauser comorbidity measures into a point system for hospital death using administrative data. *Medical care*, 47(6):626–633. DOI: doi.org/10.1097/mlr.0b013e31819432e5.
- Vickers, A. J. and Elkin, E. B. (2006). Decision curve analysis: a novel method for evaluating prediction models. *Medical Decision Making*, 26(6):565–574. DOI: doi.org/10.1177/0272989X06295361.
- Virani, S. S., Alonso, A., Benjamin, E. J., Bittencourt, M. S., Callaway, C. W., Carson, A. P., Chamberlain, A. M., Chang, A. R., Cheng, S., Delling, F. N., et al. (2020). Heart disease and stroke statistics—2020 update: a report from the american heart association. *Circulation*, 141(9):e139–e596. DOI: doi.org/10.1161/CIR.0000000000000757.
- Wang, S., Zou, X.-L., Wu, L.-X., Zhou, H.-F., Xiao, L., Yao, T., Zhang, Y., Ma, J., Zeng, Y., and Zhang, L. (2022). Epidemiology of intracerebral hemorrhage: a systematic review and meta-analysis. *Frontiers in neurology*, 13:915813. DOI: doi.org/10.3389/fneur.2022.915813.
- Warlow, C. (1998). Epidemiology of stroke. *The Lancet*, 352:S1–S4. DOI: doi.org/10.1016/s0140-6736(98)90086-1.
- Xu, L., Wang, Z., Wu, W., Li, M., and Li, Q. (2024). Global, regional, and national burden of intracerebral hemorrhage and its attributable risk factors from 1990 to 2021: results from the 2021 global burden of disease study. *BMC Public Health*, 24(1):2426. DOI: doi.org/10.1186/s12889-024-19923-7.
- Yang, L. and Shami, A. (2020). On hyperparameter optimization of machine learning algorithms: Theory and practice. *Neurocomputing*, 415:295–316. DOI: doi.org/10.1016/j.neucom.2020.07.061.
- Yang, S. and Berdine, G. (2017). The receiver operating characteristic (roc) curve. *The Southwest Respiratory and Critical Care Chronicles*, 5(19):34–36. DOI: doi.org/10.12746/swrccc.v5i19.391.