

Inteligência Artificial na Análise de Dados Clínicos da Hemorragia Puerperal: Estratégias de Risco e Predição

Júlia Bomfá¹, Flávia Oliveira^{2,3}, Zilma Reis², Cristiane Nobre^{1*}

¹Instituto de Ciências Exatas e Informática – Pontifícia Universidade Católica de Minas Gerais (PUCMG) – Belo Horizonte MG Brasil

²Faculdade de Medicina - Universidade Federal de Minas Gerais (UFMG) Belo Horizonte MG Brasil

³Maternidade Odete Valadares - Fundação Hospitalar do Estado de Minas Gerais (FHEMIG)

jbomfa@sga.pucminas.br, flaviagob@ufmg.br
zilma.medicina@gmail.com, nobre@pucminas.br

Abstract. *Postpartum hemorrhage (PPH) accounts for 14.2% of maternal deaths in Brazil (1997-2009). This study develops machine learning models to predict PPH using data from a Brazilian public hospital (2024), where clinical information exists predominantly as unstructured Portuguese narratives. We analyzed 3,162 deliveries implementing a hybrid pipeline combining regex-based extraction with TF-IDF text mining to engineer 52 features from five variables. Evaluating 20 configurations (5 algorithms × 4 balancing strategies), Random Forest without balancing achieved 60.8% recall, 31.7% precision, and 0.746 AUC-ROC despite severe class imbalance (11.8% prevalence). Feature importance validated uterine atony (20%) and anemia (15%) as primary drivers, with text-derived features comprising 85% of predictive power.*

Resumo. *A hemorragia pós-parto (HPP) é responsável por 14,2% das mortes maternas no Brasil (1997–2009). Este estudo desenvolve modelos de aprendizado de máquina para predição de HPP utilizando dados de um hospital público brasileiro (2024), no qual as informações clínicas estão predominantemente registradas em narrativas textuais não estruturadas em português. Foram analisados 3.162 partos, com a implementação de um pipeline híbrido combinando extração baseada em expressões regulares e mineração de texto via TF-IDF, resultando na engenharia de 52 atributos a partir de cinco variáveis. Na avaliação de 20 configurações (5 algoritmos × 4 estratégias de balanceamento), o Random Forest sem balanceamento obteve 60,8% de recall, 31,7% de precisão e AUC-ROC de 0,746, apesar do forte desbalanceamento de classes (11,8% de prevalência). A análise de importância das variáveis confirmou atonia uterina (20%) e anemia (15%) como principais fatores associados, com atributos derivados de texto representando 85% do poder preditivo do modelo.*

*Os autores gostariam de agradecer o financiamento recebido pelo Centro de Inovação e Inteligência Artificial para Saúde (CI-IA Saúde), em parte com recursos da Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP, Processo nº 2020/09866-4), da Fundação de Amparo à Pesquisa de Minas Gerais (FAPEMIG, PPE-00030-21, APQ-03104-24 e APQ-05058-23) e da UNIMED Belo Horizonte. Agradecemos ainda ao Conselho Nacional de Desenvolvimento Científico e Tecnológico do Brasil (CNPq: 311573/2022-3), à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES - Auxílio PROAP 88887.842889/2023-00 - PUC/MG, Informática e Código de Financiamento 001).

1. Introdução

A hemorragia puerperal (HPP), caracterizada pela perda sanguínea após o parto, é a principal causa de morte materna evitável no mundo, respondendo por mais de 20% dos óbitos maternos globais [World Health Organization 2024]. Essa problemática é ainda mais acentuada em países em desenvolvimento, onde a HPP responde por 27,1% das mortes maternas, em comparação a 16,3% em países desenvolvidos [Say et al. 2014]. No contexto brasileiro, a HPP figura entre as principais causas de mortalidade materna [OPAS 2018], cenário que motivou a iniciativa “Estratégia Zero Morte Materna por Hemorragia Pós-Parto”, coordenada pelo Ministério da Saúde em parceria com a Organização Pan-Americana da Saúde (OPAS) e a Organização Mundial da Saúde (OMS).

Apesar dos avanços em estratégias de prevenção, aprimorar a predição de HPP antes de sua ocorrência permanece um desafio crítico. Em um estudo de coorte retrospectiva com 1.936 partos no Hospital das Clínicas da UFMG [Martins et al. 2024], observou-se que quase metade dos casos de HPP ocorreu em mulheres inicialmente classificadas como baixo ou médio risco, evidenciando limitações importantes nos métodos atuais de estratificação. No contexto brasileiro, esse desafio é agravado pelo fato de que grande parte das informações clínicas relevantes em hospitais públicos é registrada como texto livre em português, em vez de dados estruturados. Esse fator limita a aplicação direta de técnicas tradicionais de aprendizado de máquina (ML) e reforça a necessidade de métodos capazes de extrair conhecimento clínico a partir de narrativas não estruturadas.

Este trabalho tem como objetivo desenvolver e avaliar um pipeline de aprendizado de máquina para predição de HPP que integre dados tabulares e textuais de narrativas clínicas em português. Analisamos 3.162 partos realizados na Fundação Hospitalar do Estado de Minas Gerais (FHEMIG) no ano de 2024, desenvolvendo um pipeline híbrido que combina extração de entidades clínicas via expressões regulares com mineração estatística de texto (TF-IDF). A partir de apenas cinco variáveis retidas (idade, raça/cor, escolaridade, tipo de parto e texto clínico), geramos 52 *features* e avaliamos 20 configurações de modelos. O algoritmo Random Forest, sem aplicação de técnicas de balanceamento, obteve 60,8% de recall e 31,7% de precisão (AUC-ROC de 0,746), tendo *atonía uterina* e *anemia* como principais variáveis preditoras. Destaca-se que informações derivadas de texto responderam por 85% do poder preditivo total do modelo.

O restante deste texto está organizado da seguinte forma: a Seção 2 apresenta o referencial teórico sobre HPP; a Seção 3 discute trabalhos relacionados; a Seção 4 descreve os materiais e métodos; a Seção 5 apresenta os resultados; a Seção 6 traz a conclusão e trabalhos futuros.

2. Referencial Teórico

A hemorragia pós-parto é definida como perda sanguínea superior a 500 mL após parto vaginal ou 1000 mL após cesariana, ou qualquer perda associada à instabilidade hemodinâmica. Pode ser classificada em primária (até 24 horas) ou secundária (até seis semanas), sendo suas principais causas atonia uterina, retenção placentária, lacerações e distúrbios de coagulação [OPAS 2018].

Diretrizes internacionais convergem na adoção de estratégias baseadas na estratificação de risco e na padronização de protocolos de manejo. A Organização Pan-Americana da Saúde propõe a classificação dos pacientes em três níveis de risco (Tabela 1). De forma complementar, a Organização Mundial da Saúde recomenda a mensuração quantitativa da perda sanguínea e a implementação de pacotes padronizados de intervenção [World

Health Organization 2023]. Já o American College of Obstetricians and Gynecologists (ACOG) adota uma abordagem tripartite com condutas específicas [American College of Obstetricians and Gynecologists 2017], enquanto o Royal College of Obstetricians and Gynaecologists (RCOG) enfatiza o uso de protocolos estruturados para prevenção e manejo [Mavrides et al. 2017].

Tabela 1. Estratificação de risco para hemorragia pós-parto, segundo OPAS (2018)

BAIXO RISCO	MÉDIO RISCO	ALTO RISCO
<ul style="list-style-type: none"> • Ausência de cicatriz uterina • Gravidez única • ≤ 3 partos vaginais prévios • Ausência de distúrbio de coagulação • Sem história de HPP 	<ul style="list-style-type: none"> • Cesariana ou cirurgia uterina prévia • Pré-eclâmpsia leve • Hipertensão gestacional leve • Superdistensão uterina (gestação múltipla, polidrâmnio, macrossomia fetal) • ≥ 4 partos vaginais • Corioamnionite • História prévia de atonia uterina¹ ou hemorragia obstétrica • Obesidade materna (IMC² $> 35kg/m$) 	<ul style="list-style-type: none"> • Placenta prévia ou de inserção baixa • Pré-eclâmpsia grave • Hematócrito $< 30\%$ + fatores de risco • Plaquetas $< 100.000/mm$ • Sangramento ativo à admissão • Coagulopatias • Uso de anticoagulantes • Descolamento prematuro de placenta • Placentação anômala (acretismo) • Presença de ≥ 2 fatores de médio risco

Apesar da ampla adoção dessas diretrizes, análise de 1.936 partos no HC-UFGM revelou que 48,1% dos casos de HPP ocorreram em mulheres classificadas como baixo ou médio risco [Martins et al. 2024], evidenciando que os critérios atuais deixam de identificar parcela significativa das pacientes. Essa limitação motivou pesquisas sobre métodos preditivos mais acurados, incluindo aprendizado de máquina.

3. Trabalhos relacionados

Diversos estudos têm aplicado técnicas de aprendizado de máquina para predição de HPP, predominantemente em bases de dados estruturadas. Modelos como Generalized Additive Models (GAM) e Gradient Boosting apresentam AUC-ROC entre 0,67 e 0,83 [Lengerich et al. 2024; Ahmadzia et al. 2024], enquanto redes neurais também demonstram desempenho competitivo em contextos específicos [Jun-Yu et al. 2024]. Em cenários com menor volume de dados, Random Forest tem se mostrado uma alternativa eficiente, mantendo desempenho clínico relevante [Holcroft et al. 2024].

Mais recentemente, abordagens baseadas em processamento de linguagem natural têm sido exploradas para incorporar informações textuais. Woo et al. [2025] utilizaram Large Language Models (LLMs) aplicados a notas clínicas pré-natais, demonstrando ganho preditivo em relação a modelos baseados exclusivamente em dados estruturados. Apesar dos avanços, esses estudos são majoritariamente conduzidos em língua inglesa e em ambientes com alta disponibilidade de dados e recursos computacionais.

No contexto brasileiro, Baêta [2025] explorou aprendizado não supervisionado para identificar agrupamentos de risco em dados hospitalares estruturados, evidenciando

¹ Atonia uterina: condição caracterizada pela falha do útero em contrair adequadamente após o parto.

² IMC: Índice de Massa Corporal

padrões distintos associados à HPP. Ainda assim, observa-se a ausência de abordagens que integrem dados estruturados e narrativas clínicas em português em modelos supervisionados.

Nesse cenário, este trabalho avança ao propor um pipeline híbrido que combina dados tabulares e textuais em português, explorando mineração de texto clínico para extração de conhecimento preditivo em um contexto de dados imperfeitos e não estruturados, característico do sistema público brasileiro. Diferentemente da maioria dos estudos, que se baseiam exclusivamente em dados estruturados ou em ambientes internacionais, nossa abordagem demonstra que informações textuais podem representar parcela significativa do poder preditivo do modelo, mesmo em bases de menor escala.

A Tabela 2 sintetiza as principais abordagens metodológicas, contextos de aplicação e desempenho dos estudos selecionados, permitindo uma visão comparativa das estratégias atualmente utilizadas.

Tabela 2. Comparação dos trabalhos relacionados

Referência	Abordagem	Contexto	Tipo de Base	Região
[Lengerich et al. 2024]	ML Supervisionado (GAM)	19 hospitais, 85K partos	Tabular-EN	EUA
[Ahmadzia et al. 2024]	ML Supervisionado (comparativo)	12 hospitais, 228K partos	Tabular-EN	EUA
[Jun-Yu et al. 2024]	ML Supervisionado (MLP)	23K pacientes, minoria étnica	Tabular-EN	China
[Holcroft et al. 2024]	ML Supervisionado (comparativo)	5 unidades, 430 casos	Tabular-EN	Ruanda
[Woo et al. 2025]	ML Supervisionado (LLMs)	Rede hospitalar, 20K partos	Textual-EN	EUA
[Baêta 2025]	ML Não Supervisionado (clustering)	HC-UFGM, 1.9K partos	Tabular-PT	Brasil
Este trabalho	ML Supervisionado (comparativo)	3.2K partos, mineração de texto-PT	Mista-PT	Brasil

ML: Machine Learning; EN: Inglês; PT: Português; HC-UFGM: Hospital das Clínicas da UFGM.

4. Materiais e métodos

4.1. Descrição da base de dados

A base de dados utilizada neste trabalho foi extraída do sistema TASY da Maternidade Odete Valadares, unidade da Fundação Hospitalar do Estado de Minas Gerais (FHEMIG), referente ao ano de 2024. Cada registro corresponde a um parto, totalizando 3.352 casos (3.162 após remoção de duplicatas, conforme Seção 4.2.1), originalmente distribuídos em 18 variáveis. Destas, 7 foram mantidas (3 textuais e 4 tabulares, conforme Tabela 3), enquanto as demais foram descartadas por apresentarem elevada proporção de valores ausentes, inconsistências de preenchimento ou caráter exclusivamente administrativo, sem relevância preditiva. A base resultante configura-se, portanto, como mista, combinando

variáveis estruturadas e registros clínicos não estruturados, incluindo notas de evolução, resumos de caso e descrições cirúrgicas, que fornecem informações potencialmente relevantes para a identificação de fatores associados à hemorragia pós-parto.

Tabela 3. Variáveis retidas da base de dados

Variável	Descrição	Tipo
IDADE	Idade materna (em anos)	Numérica
DS_ESCOLARIDADE	Grau de escolaridade	Catégorica
DS_COR	Cor/raça autodeclarada	Catégorica
PROCEDIMENTO	Tipo de parto realizado	Catégorica
DS_CIRURGIA	Descrição da cirurgia (texto livre)	Texto
DS_EVOLUCAO	Evolução clínica (texto livre)	Texto
DS_RESUMO	Resumo do caso clínico	Texto

Nota: Base original continha 18 variáveis; foram retidas 7 após pré-processamento (4 tabulares + 3 textuais).

4.2. Metodologia

A metodologia adotada neste estudo segue um pipeline estruturado composto por quatro etapas principais: (i) pré-processamento e limpeza dos dados, (ii) identificação e rotulagem de casos de HPP a partir de narrativas clínicas, (iii) engenharia de *features* com integração de dados tabulares e textuais, e (iv) desenvolvimento e avaliação de modelos preditivos. As decisões metodológicas foram orientadas pela natureza não estruturada dos dados e pelo objetivo de avaliar a viabilidade de predição em um cenário clínico real com limitações de qualidade e padronização.

4.2.1. Pré-processamento dos dados

O conjunto de 3.352 registros descrito na Seção 4.1 passou por um processo de limpeza e validação para garantir qualidade e consistência dos dados. Primeiramente, foram removidos 12 registros com valores nulos. Em seguida, 178 casos duplicados foram identificados e excluídos por correspondência exata em todas as variáveis (incluindo textuais), resultando em um conjunto final de *3.162 registros únicos* de partos para desenvolvimento dos modelos preditivos.

Variáveis temporais (datas de admissão, alta e parto) foram excluídas após identificar que alguns registros continham valores preenchidos aleatoriamente que introduziriam ruído. Também foram removidas variáveis administrativas sem valor preditivo: identificador hospitalar, números de caso, códigos de procedimento e município de residência.

As variáveis estruturadas correspondiam majoritariamente a informações do pré-parto ou da admissão, enquanto os campos textuais incluíam registros do intraparto e pós-parto, refletindo um cenário de predição em tempo real durante a internação. Variáveis posteriores à hemorragia foram excluídas, por não serem compatíveis com a predição de HPP antes de sua ocorrência.

Os três campos textuais em português (DS_CIRURGIA, DS_EVOLUCAO, DS_RESUMO) foram concatenados em uma variável unificada TEXTO_COMPLETO para cada registro. A limpeza textual inicial normalizou todo o texto para minúsculas e removeu sequências de escape Unicode, artefatos do Excel, tags HTML, múltiplos espaços consecutivos e caracteres especiais.

As variáveis categóricas remanescentes (idade materna, raça/etnia, escolaridade e tipo de parto) foram codificadas numericamente: codificação binária para variáveis onde uma categoria correspondia a uma prevalência de 50% ou mais (como por exemplo etnia que foi classificado entre “pardo” e “não pardo”), e codificação categórica nos demais casos.

4.2.2. Identificação de Casos de HPP

Como a base de dados não possuía rótulos explícitos de HPP, desenvolveu-se, em colaboração com uma especialista em obstetrícia, um classificador baseado em regras textuais com validação clínica. Definiram-se três níveis de confiança baseados na força dos indicadores clínicos presentes nas narrativas. Casos de *alta confiança* incluem diagnóstico explícito de HPP, intervenções maiores (histerectomia, ligadura arterial, balão intrauterino), complicações críticas (choque hipovolêmico, parada cardíaca) ou transfusão massiva. Casos de *média confiança* compreendem atonia uterina com tratamento documentado, sangramento significativo requerendo manejo ativo (ocitocina, misoprostol, massagem uterina) ou complicações placentárias requerendo extração manual. Por fim, casos de *baixa confiança* caracterizam-se por hipotonia uterina transitória ou episódios menores de sangramento.

Para distinguir eventos reais de menções irrelevantes, implementou-se detecção de negação através de análise de contexto. Textos clínicos frequentemente contêm menções a HPP em contextos não indicativos de ocorrência real, como orientações preventivas (“orientada sobre sinais de hemorragia”), negações explícitas (“nega sangramento”, “sem sinais de hemorragia”) e cenários hipotéticos em condutas (“se houver sangramento excessivo”).

Para filtrar essas ocorrências, analisou-se uma janela de caracteres precedentes às palavras-chave identificadas, buscando termos negadores como “não”, “sem”, “nega” ou “ausência de”. Uma janela de 40 caracteres foi definida empiricamente apresentando o melhor equilíbrio entre captura de negações simples típicas de registros clínicos e minimização de ruído contextual. Termos negados não foram contabilizados como indicadores de HPP.

O algoritmo foi validado através de revisão manual de uma amostra estratificada de 90 casos (30 de cada nível de confiança), atingindo 96,1% de concordância com avaliação clínica especializada e zero falsos positivos de alta confiança. Embora a amostra tenha priorizado casos positivos para validação dos critérios de identificação, reconhece-se como limitação a ausência de uma avaliação sistemática da classe negativa, o que pode impactar a estimativa de desempenho do modelo. Para a análise final, casos classificados como média confiança (3,10%) e alta confiança (8,67%) foram agrupados como HPP positivo resultando em 372 casos positivos (11,8% de prevalência).

4.2.3. Engenharia de Features e mineração de texto

A partir de cinco variáveis retidas após o pré-processamento (*idade materna, cor/raça, escolaridade, tipo de parto e texto completo*), foram engenhadas 52 *features* por quatro estratégias complementares. A primeira consistiu em indicadores binários de idade para faixas de risco conhecidas ($idade < 18$ anos, $idade \geq 35$ anos). A segunda utilizou expressões regulares para extrair idade gestacional e histórico obstétrico (ex.: “G3PN2A1” = 3 gestações, 2 partos normais, 1 aborto), gerando variáveis numéricas. A terceira

aplicou 39 padrões regex para identificar fatores de risco clínicos, incluindo comorbidades, complicações obstétricas, condições fetais e intervenções. Por fim, empregou-se a vetorização textual por TF-IDF (`max_features=500`, `min_df=5`), seguida de seleção qui-quadrado (*chi-score* > 0.10), resultando em 9 variáveis textuais finais.

A matriz final compreende 52 variáveis: 4 tabulares originais, 2 derivadas de idade, 37 baseadas em regex (e derivadas) e 9 TF-IDF estatísticas. O conjunto de dados foi dividido em treino e teste usando estratificação (80-20), resultando em 2.529 casos de treino (298 HPP, 11,8%) e 633 casos de teste (74 HPP, 11,7%).

Embora as variáveis textuais utilizadas tenham sido cuidadosamente selecionadas com base em conhecimento clínico prévio, priorizando fatores de risco, comorbidades e eventos intraparto anteriores à hemorragia, reconhece-se que algumas expressões presentes nas narrativas podem estar temporalmente próximas ao desfecho.

4.2.4. Desenvolvimento e Avaliação de Modelos

Avaliaram-se 20 configurações diferentes, combinando quatro estratégias de balanceamento de classes com cinco algoritmos de classificação. As estratégias de balanceamento testadas foram: sem balanceamento (dados originais), SMOTE (*Synthetic Minority Oversampling Technique*), DSTO-GAN (*Data Synthesis with Tabular Generative Adversarial Networks*) e *Bootstrap oversampling*. Todas as técnicas de *oversampling* foram aplicadas apenas no conjunto de treino, gerando amostras sintéticas da classe minoritária até atingir proporção 50-50 com a classe majoritária.

Para cada estratégia de balanceamento, implementaram-se cinco algoritmos: Random Forest (RF), XGBoost (*Extreme Gradient Boosting*), MLP (*Multi-Layer Perceptron*), SVM (*Support Vector Machine*) e KNN (*K-Nearest Neighbors*). Os hiperparâmetros foram mantidos em valores padrão, com ajuste apenas do limiar de classificação (*threshold*). Os limiares foram otimizados no intervalo [0,1; 0,9] com passo de 0,05 para maximizar *recall* sujeito às restrições de acurácia $\geq 0,70$ e precisão $\geq 0,30$, refletindo o contexto clínico onde falsos negativos (casos de HPP não identificados) têm custo significativamente maior que falsos positivos (preparação preventiva desnecessária).

A avaliação dos modelos baseou-se em cinco métricas: *recall*¹ (métrica primária, representando a capacidade de identificar casos de HPP), precisão² (proporção de predições positivas corretas), *F1-score*³ (média harmônica entre precisão e *recall*), acurácia (taxa de acertos independente da classe) e AUC-ROC (área sob a curva ROC, medindo capacidade discriminativa geral do modelo).

4.2.5. Ambiente Computacional

Todas as análises foram realizadas em Python 3.13, utilizando Pandas 2.3 e NumPy 2.3 para manipulação de dados; NLTK 3.9 e spaCy 3.7 (`pt_core_news_sm`) para pré-processamento textual; Scikit-learn 1.7 para TF-IDF, seleção de *features* (qui-quadrado) e implementação de Random Forest, SVM, KNN e MLP; XGBoost 3.0 para *gradient boosting*; Imbalanced-learn 0.14 para técnicas de balanceamento (SMOTE e *bootstrap*

¹Recall = $\frac{TP}{TP+FN}$

²Precision = $\frac{TP}{TP+FP}$

³F1-Score = $2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$

oversampling); e Matplotlib 3.10 com Seaborn 0.13 para visualizações. As execuções foram realizadas em um computador Intel Core i7-10750H com 16GB de RAM, executando Windows 11 Home Single Language (build 26100).

5. Resultados

Foram avaliadas 20 configurações diferentes combinando cinco algoritmos de classificação com quatro estratégias de balanceamento de classes. A Tabela 4 apresenta as cinco configurações com melhor desempenho, ordenadas por *recall*, métrica primária deste estudo devido ao contexto clínico em que a não identificação de casos de HPP pode resultar em desfechos graves.

Tabela 4. Métricas de desempenho das cinco melhores configurações, por classe

Modelo	Classe	Precisão (%)	Recall (%)	F1-Score (%)
RF (sem balanceamento)	Sem HPP	94,1	82,6	88,0
	HPP	31,7	60,8	41,7
RF + Bootstrap	Sem HPP	93,7	88,4	91,0
	HPP	38,7	55,4	45,6
XGB + Bootstrap	Sem HPP	93,3	85,2	89,1
	HPP	32,5	54,1	40,6
XGB (sem balanceamento)	Sem HPP	93,2	85,9	89,4
	HPP	33,1	52,7	40,6
MLP + Bootstrap	Sem HPP	93,1	84,3	88,5
	HPP	30,7	52,7	38,8

O modelo Random Forest sem balanceamento externo apresentou o maior *recall* (60,8%), identificando 45 dos 74 casos de HPP no conjunto de teste, com precisão de 31,7% e AUC-ROC de 0,746. Esse desempenho pode ser atribuído ao uso de pesos de classe balanceados (*class_weight='balanced'*) e ao ajuste do limiar de decisão priorizando sensibilidade. O modelo classificou corretamente 462 dos 559 casos sem HPP (82,6%), resultando em 97 falsos positivos e 29 falsos negativos. Em comparação, o Random Forest com *bootstrap oversampling* externo apresentou maior AUC-ROC (0,762), porém com menor *recall* (55,4%) e maior precisão (38,7%), evidenciando o *trade-off* entre sensibilidade e especificidade.

Esses resultados são comparáveis aos reportados na literatura internacional, em que modelos supervisionados baseados em dados estruturados têm alcançado AUC-ROC moderadas a elevadas, variando entre aproximadamente 0,67 e 0,83 [Lengerich et al. 2024; Ahmadzia et al. 2024]. Considerando que o presente estudo utiliza uma base de menor escala e predominantemente não estruturada, o desempenho obtido reforça a viabilidade da aplicação de aprendizado de máquina em contextos com limitações de dados, como o sistema público brasileiro.

A análise manual dos falsos negativos mostrou que a maioria dos casos correspondia a HPP leve ou rapidamente revertida após aplicação do protocolo institucional e uso de medicações. Nesses prontuários, a documentação enfatizava a resolução clínica, com menor detalhamento dos sinais iniciais de sangramento, dificultando sua identificação automática.

A análise dos 97 falsos positivos revelou dois padrões principais: pacientes com múltiplos fatores de risco que não evoluíram para HPP, possivelmente devido a intervenções preventivas eficazes, e limitações na rotulagem e extração textual, associadas a variações na documentação clínica.

Do ponto de vista operacional, a presença de falsos positivos implica aumento na demanda por monitoramento e possíveis intervenções preventivas. Em ambientes com recursos limitados, esse fator pode impactar o fluxo assistencial, exigindo protocolos bem definidos para manejo de pacientes classificados como alto risco. Nesse sentido, o modelo deve ser interpretado como uma ferramenta de triagem inicial, complementando a avaliação clínica e não substituindo o julgamento médico.

A Figura 1 compara as curvas ROC dos cinco melhores modelos. O Random Forest com *bootstrap* apresentou o maior AUC-ROC (0,762), enquanto o RF sem balanceamento, com AUC ligeiramente inferior (0,746), obteve o melhor *recall* (60,8%), evidenciando o *trade-off* entre discriminação global e sensibilidade. O limiar de decisão foi ajustado para priorizar *recall*, estratégia adequada ao contexto clínico. Os modelos baseados em árvores apresentaram desempenho consistente (AUCs entre 0,739–0,762), superando o MLP (0,740) e o classificador aleatório (AUC=0,500).

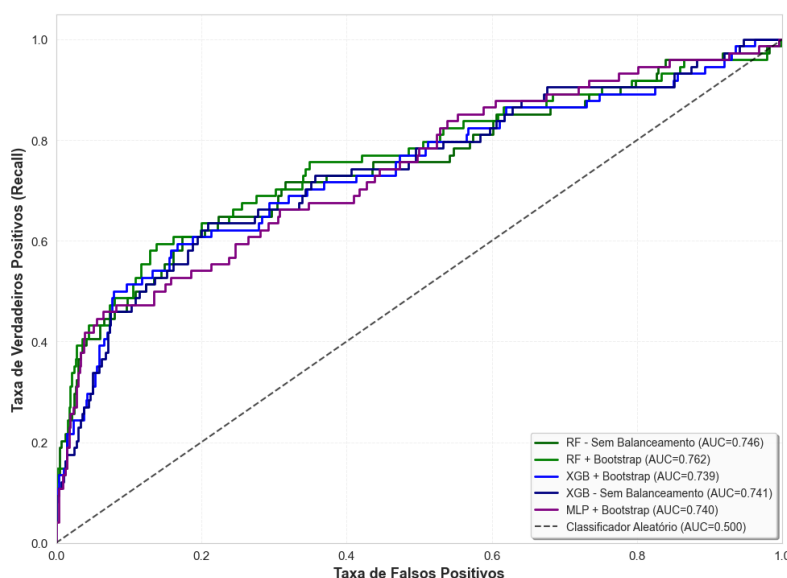


Figura 1. Curva ROC comparando os cinco modelos com melhor desempenho no conjunto de teste (n=633)

A análise de importância das *features* (Figura 2) revelou uma composição equilibrada entre termos TF-IDF, padrões regex e variáveis estruturadas. A atonia uterina foi a *feature* mais relevante (20%), em consonância com a literatura [OPAS 2018; Mavrides et al. 2017]. Considerando as 20 *features* mais relevantes, aproximadamente 50% correspondem a termos TF-IDF, 35% a padrões regex e 15% a variáveis estruturadas. Esses resultados indicam que a maior parte do poder preditivo do modelo deriva de informações contidas nas narrativas clínicas, reforçando a importância da exploração de dados não estruturados no contexto analisado.

A comparação entre as *features* do modelo e as variáveis descritas na literatura demonstra alinhamento clínico consistente. A anemia (15%) é amplamente reconhecida como fator de risco relevante [OPAS 2018; Jun-Yu et al. 2024], enquanto lacerações correspondem ao componente “Trauma” dos “4 Ts”⁴. Variáveis como idade materna,

⁴Os “4 Ts” correspondem aos principais mecanismos da hemorragia pós-parto: Tônus, Trauma, Tecido e Trombina.

tipo de parto e idade gestacional também estão entre os principais preditores descritos na literatura. Ressalta-se que variáveis laboratoriais importantes (hemoglobina, plaquetas, hematócrito) não puderam ser incluídas devido à ausência de registros estruturados.

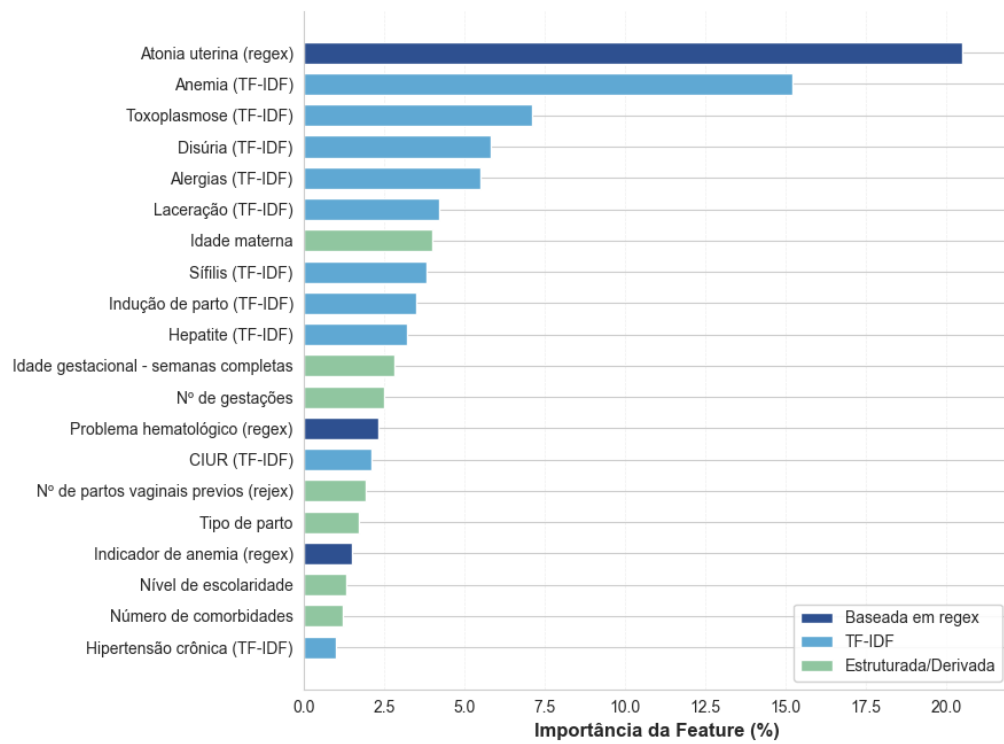


Figura 2. Importância das 20 features mais relevantes do modelo Random Forest sem balanceamento

Em termos clínicos, o melhor modelo gerou 142 predições de HPP (45 verdadeiros positivos e 97 falsos positivos), resultando em uma proporção aproximada de 2:1 entre falsos alarmes e casos reais. Embora essa relação represente aumento na carga operacional, deve ser interpretada à luz do custo assimétrico dos erros em saúde materna. A não identificação de casos de HPP pode resultar em atraso no manejo de uma condição potencialmente fatal, enquanto falsos positivos tendem a implicar apenas em monitoramento adicional e medidas preventivas de baixo risco.

Ainda assim, é importante reconhecer que taxas elevadas de falsos positivos podem gerar sobrecarga assistencial e potencial fadiga da equipe clínica, especialmente em ambientes com recursos limitados. Dessa forma, a utilização do modelo deve estar associada a protocolos clínicos bem definidos, permitindo estratificar níveis de intervenção conforme o risco estimado e mitigando impactos operacionais negativos.

6. Conclusão e trabalhos futuros

Este estudo demonstra a viabilidade de predição de HPP utilizando aprendizado de máquina aplicado a narrativas clínicas em português. O modelo Random Forest sem balanceamento alcançou *recall* de 60,8% com precisão de 31,7% e AUC-ROC de 0,746, resultados relevantes considerando prevalência baixa de HPP (11,8%), informações registradas predominantemente em texto não estruturado e variabilidade na documentação clínica.

A análise de importância revelou atonia uterina (20%) e anemia (15%) como fatores mais relevantes, alinhado com a literatura. Atributos derivados de texto representaram 85% do poder preditivo total, validando a estratégia híbrida regex-TF-IDF. A superioridade do Random Forest sem balanceamento representa vantagem prática ao eliminar geração de dados sintéticos.

Em termos clínicos, o melhor modelo gerou 142 predições de HPP (45 verdadeiros positivos e 97 falsos positivos), resultando em uma proporção aproximada de 2:1 entre alertas incorretos e casos corretamente identificados. Embora essa relação aumente a demanda por monitoramento e preparo preventivo, a HPP é uma emergência obstétrica potencialmente grave, na qual atrasos no reconhecimento e manejo elevam o risco de complicações maternas. Nesse contexto, os falsos negativos tendem a representar maior impacto clínico que os falsos positivos, tornando a priorização de *recall* uma escolha adequada para triagem inicial.

As limitações decorrem da natureza não estruturada e da qualidade variável dos dados. A dependência de narrativas textuais (85% do poder preditivo) evidencia insuficiência dos dados tabulares. Observou-se documentação clínica inconsistente devido à falta de treinamento padronizado no preenchimento de registros, resultando em variabilidade no estilo de escrita, terminologia e presença de informações essenciais. A ausência de campos estruturados para dados laboratoriais (hemoglobina, plaquetas, hematócrito) e amostra de instituição única (372 casos positivos) limitam a generalização.

Como trabalhos futuros, destacam-se a implementação prospectiva do modelo em ambiente clínico real, a validação em múltiplas instituições e bases públicas, e a expansão para cenários com maior diversidade populacional. Além disso, técnicas mais avançadas de processamento de linguagem natural, como modelos baseados em arquiteturas transformer, podem melhorar o desempenho preditivo ao explorar representações semânticas mais profundas das narrativas clínicas. Por fim, recomenda-se a padronização da coleta de dados em sistemas hospitalares brasileiros, com campos estruturados para variáveis essenciais, como exames laboratoriais e histórico obstétrico, favorecendo modelos mais robustos e sua aplicação em larga escala.

Este trabalho indica que predição significativa de saúde materna é alcançável mesmo com dados imperfeitos, fornecendo fundação para ferramentas de avaliação de risco assistidas por inteligência artificial. Contudo, os resultados também evidenciam que o verdadeiro avanço na predição de complicações obstétricas no Brasil virá não apenas de algoritmos mais sofisticados, mas fundamentalmente da melhoria na infraestrutura de dados de saúde, com priorização de captura estruturada, padronizada e validada de informações clínicas essenciais.

Referências

- H. K. Ahmadzia, A. C. Dzienny, M. Bopf, J. M. Phillips, J. J. Federspiel, R. Amdur, M. M. Rice, e L. Rodriguez. Machine learning models for prediction of maternal hemorrhage and transfusion: Model development study. *JMIR Bioinformatics and Biotechnology*, 5(1):e52059, 2024.
- American College of Obstetricians and Gynecologists. Practice bulletin no. 183: Postpartum hemorrhage. *Obstetrics & Gynecology*, 130(4):e168–e186, 2017. URL <https://www.acog.org/clinical/clinical-guidance/practice-bulletin/articles/2017/10/postpartum-hemorrhage>.
- T. Baêta. Estudo da hemorragia pós-parto com modelos de aprendizagem de máquina em busca de aprimorar os perfis de risco. Dissertação de mestrado em andamento,

- universidade federal de minas gerais, belo horizonte, brasil, Universidade Federal de Minas Gerais, Belo Horizonte, Brasil, 2025.
- S. Holcroft, I. Karangwa, F. Little, J. Behoor, e O. Bazirete. Predictive modelling of postpartum haemorrhage using early risk factors: A comparative analysis of statistical and machine learning models. *International Journal of Environmental Research and Public Health*, 21(5):600, 2024.
- P. Jun-Yu, W. Ling, L. Kai-Feng, e X. Sheng-Yun. Risk of postpartum hemorrhage in an ethnic minority region: Prediction results based on iterative machine learning. *J Clin Images Med Case Rep*, 5(12):3379, 2024.
- B. J. Lengerich, R. Caruana, I. Painter, W. B. Weeks, K. Sitcov, e V. Souter. Interpretable machine learning predicts postpartum hemorrhage with severe maternal morbidity in a lower-risk laboring obstetric population. *American Journal of Obstetrics & Gynecology MFM*, 6(8), Aug 2024. ISSN 2589-9333. doi: 10.1016/j.ajogmf.2024.101391. URL <https://doi.org/10.1016/j.ajogmf.2024.101391>.
- R. I. L. Martins, J. d. S. M. Novais, e Z. S. N. Reis. Postpartum hemorrhage in electronic health records: risk factors at admission and in-hospital occurrence. *Revista Brasileira de Ginecologia e Obstetrícia*, 46:e-rbgo14, 2024.
- E. Mavrides, S. Allard, E. Chandrarahan, P. Collins, L. Green, B. J. Hunt, S. Riris, e A. J. Thomson. Prevention and management of postpartum haemorrhage. *BJOG: An International Journal of Obstetrics & Gynaecology*, 124(5):e106–e149, 2017. doi: <https://doi.org/10.1111/1471-0528.14178>. URL <https://obgyn.onlinelibrary.wiley.com/doi/abs/10.1111/1471-0528.14178>.
- OPAS. Recomendações Assistenciais Para Prevenção, Diagnóstico E Tratamento Da Hemorragia, 2018. URL <https://iris.paho.org/handle/10665.2/34879>.
- L. Say, D. Chou, A. Gemmill, Ö. Tunçalp, A.-B. Moller, J. Daniels, A. M. Gülmezoglu, M. Temmerman, e L. Alkema. Global causes of maternal death: a who systematic analysis. *The Lancet global health*, 2(6):e323–e333, 2014.
- E. G. Woo, I. Zigelboim, T. Gifford, J. G. Bell, H. Milthorpe, E. Alsentzer, R. E. Longman, J. E. Tolosa, e B. K. Beaulieu-Jones. Leveraging large language models to develop an interpretable prediction model for postpartum hemorrhage prior to the onset of labor. *medRxiv*, pages 2025–03, 2025.
- World Health Organization. *WHO recommendations on the assessment of postpartum blood loss and use of a treatment bundle for postpartum haemorrhage*. World Health Organization, Geneva, 2023. ISBN 978-92-4-008539-8. URL <https://www.who.int/publications/i/item/9789240085398>. Licence: CC BY-NC-SA 3.0 IGO.
- World Health Organization. Second global call for data on postpartum haemorrhage. WHO News Room, 2024. URL <https://www.who.int/news-room/articles-detail/second-global-call-for-data-on-postpartum-haemorrhage>.