

Análise Preditiva de Sepsis em Unidades de Terapia Intensiva utilizando Machine Learning

Ana F. Santos¹, Gustavo A. Alves¹, Mariana F. Tavares¹, Michelle Andrade¹

¹Pontifícia Universidade Católica de Minas Gerais (PUC Minas)
Belo Horizonte – MG – Brasil

{ana.santos.1380776,mariana.tavares}@sga.pucminas.br,
gustavoalves02@gmail.com, michelleandrade@pucminas.br

Abstract. Sepsis is the leading cause of preventable death worldwide, with mortality rates in Brazil reaching 60%. This paper presents a study based on real-world ICU data from the Santa Casa de Misericórdia of Belo Horizonte for the early identification of sepsis. Our approach integrates vital sign data and the NEWS protocol with Machine Learning models to predict clinical deterioration. Preliminary results using Ensemble models demonstrate promising performance, with Gradient Boosting achieving an F1-Score of 89% for sepsis prediction 6 hours in advance, 88% at 12 hours, and 85% at 24 hours, effectively balancing the trade-off between sensitivity and diagnostic precision.

Resumo. A sepsis é a principal causa de morte evitável no mundo, com taxas de mortalidade no Brasil chegando a 60%. Este artigo apresenta um trabalho baseado em dados reais da UTI da Santa Casa de Misericórdia de Belo Horizonte na identificação precoce da sepsis. Nossa abordagem integra dados de sinais vitais e o protocolo NEWS com modelos de Machine Learning para prever a deterioração clínica. Resultados preliminares utilizando modelos de Ensemble mostram ser promissores, como o Gradient Boosting com F1-Score de 89% para predição da sepsis com 6h de antecedência, 88% para 12h e 85% para 24h, equilibrando o trade-off entre detecção e precisão diagnóstica.

1. Introdução

A sepsis é uma condição crítica caracterizada por uma resposta inflamatória sistêmica desregulada, sendo a principal causa de mortalidade em Unidades de Terapia Intensiva (UTIs) no Brasil, com taxas que chegam a 60% dos casos diagnosticados [ILAS 2024]. Do ponto de vista da Ciência de Dados, a detecção precoce da sepsis apresenta-se como um problema complexo de classificação, cujo desafio é prever com antecedência para uma rápida intervenção médica. O desafio reside na natureza dos Registros Eletrônicos de Saúde (RES), que são inerentemente ruidosos, apresentam alta taxa de dados faltantes (“missing values”) e possuem uma janela temporal crítica para intervenção, onde cada hora de atraso no diagnóstico aumenta a mortalidade em aproximadamente 8% [Singer et al. 2016].

Embora protocolos tradicionais, como o *National Early Warning Score* (NEWS), ofereçam uma base linear de pontuação para deterioração clínica, eles falham em capturar correlações não lineares entre sinais vitais e biomarcadores laboratoriais. Neste contexto, este estudo propõe a avaliação de possíveis soluções algorítmicas de Machine

Learning (ML) para previsão de sepse no intervalo de 6h, 12h e 24h. A pesquisa investiga como modelos de aprendizado supervisionado podem ser treinados sobre dados reais de prontuários para antecipar o diagnóstico precoce da sepse.

O objetivo principal é avaliar o desempenho de modelos de Machine Learning na previsão de sepse, tendo como referência o *score* NEWS. Pretende-se demonstrar que a aplicação de algoritmos como Random Forest, Gradient Boosting, AdaBoost e RF Boruta oferece uma base mais precisa para o apoio à intervenção médica no caso de sepse. Dentre os objetivos específicos destacam-se: (i) criar a ABT (*Analytical Base Table*) com os dados reais, cedidos para este estudo. (ii) organizar safras, agrupamento de dados com base na data de coleta de dados clínicos, como FC (Frequência Cardíaca), FR (Frequência Respiratória), Temperatura Corporal, sexo e outros. (iii) realizar análise comparada com algoritmos de Machine Learning por meio de métricas como precisão, revocação (*recall*), curva ROC e F1-score.

2. Trabalhos Relacionados

As bases conceituais para a detecção precoce de sepse transitam de *scores* estáticos, como o NEWS, para modelos preditivos avançados. Embora amplamente adotados, esses *scores* apresentam sensibilidade limitada frente à deterioração clínica não linear [Covino et al. 2023]. Modelos de *Machine Learning* (ML) superam esses métodos ao identificar padrões complexos em dados longitudinais horas antes do diagnóstico convencional [Bomrah et al. 2024, Tan et al. 2025].

No âmbito da Ciência de Dados, algoritmos de Gradient Boosting, como o XGBoost, consolidaram-se como estado da arte para dados clínicos tabulares, correlacionando sinais vitais e biomarcadores [Li et al. 2023]. Contudo, sua aplicação enfrenta desafios como heterogeneidade e alta taxa de dados faltantes nos Registros Eletrônicos de Saúde. Para mitigar isso, técnicas de *Uncertainty Quantification* e *Active Sensing* garantem previsões estatisticamente válidas mesmo em cenários esparsos e de alta incerteza [Yin et al. 2024, Ren et al. 2024].

A natureza multimodal dos dados hospitalares impõe desafios adicionais de integração. O framework Diffmv aborda esse problema com imputação baseada em difusão para modalidades ausentes, mitigando *view laziness* por reponderação e preservando a consistência intra-paciente [Zhao et al. 2025].

3. Metodologia

A pesquisa caracteriza-se como experimental, conforme os critérios metodológicos propostos por [Gil 2002] e [Lakatos and Marconi 2003]. Neste projeto, o arranjo experimental foi estruturado em cinco etapas sequenciais: (i) construção da ABT com dados reais da UTI (Unidade de Terapia Intensiva) da Santa Casa de Misericórdia de Belo Horizonte, em conformidade com a Lei Geral de Proteção de Dados (LGPD, Lei nº 13.709/2018); (ii) cálculo do *score* NEWS; (iii) definição das safras com base em dados clínicos e sinais vitais; (iv) implementação dos algoritmos de Machine Learning, com execução de treinamento e testes; e (v) apuração e análise dos resultados.

3.1. Construção da ABT e Safras

A construção da ABT (*Analytical Base Table*) envolveu 74 arquivos CSV, extraídos por meio de um Data Warehouse, entre o período de janeiro de 2024 a maio de 2025. Tota-

lizando 52.719 atendimentos distintos e 1.845.668 registros. Os dados disponíveis nesta pesquisa são:

- **Sinais Vitais:** Fonte principal para extração de identificadores como número do atendimento, dados de admissão como data e CID (Classificação Internacional de Doenças) da internação e desfecho (data e CID da alta/óbito), além das medições aferidas por equipamentos médicos (monitor cardíaco, termômetro e oxímetro de pulso).
- **Documentos Eletrônicos (Clínicos):** Registros de evolução e prescrição médica.
- **Dados Laboratoriais:** Resultados de exames laboratoriais de pacientes.

Este estudo ateve-se somente aos dados de sinais vitais para a construção da ABT inicial. Os atributos que constituem essa base são descritos na Tabela 1.

Tabela 1. Descrição das variáveis do conjunto de dados clínicos

Variável	Descrição
id	Código do atendimento
data_coleta	Data da coleta dos sinais vitais
hora_coleta	Hora da coleta dos sinais vitais
tempo_internacao	Tempo de internação decorrido em horas
genero	Gênero do paciente
TEMP	Temperatura corporal (°C)
FC	Frequência cardíaca (bpm)
FR	Frequência respiratória (rpm)
PAS	Pressão arterial sistólica (mmHg)
CID_internacao	Código CID de internação
CID_saida	Código CID de saída ou de óbito, se houver
SEPSE	Marcação de sepse caso o CID de saída esteja na lista de CIDs relacionados à septicemia (ex.: A400 – Septicemia por <i>Streptococcus</i> do grupo A)
OBITO	Marcação de óbito (0 ou 1)
NEWS_SCORE	Cálculo do escore NEWS

O cálculo do *score* NEWS baseou-se na estratificação de risco fisiológico conforme preconizado pelo *Royal College of Physicians*. O algoritmo de pontuação atribui valores inteiros variando de 0 (dentro da normalidade) a 3 (desvio extremo) para cada parâmetro vital, dependendo do intervalo em que a medição se encontra.

Devido à ausência de dados referentes à saturação de oxigênio (SpO_2), uso de oxigênio suplementar e nível de consciência na base de dados original, o cálculo foi adaptado para considerar exclusivamente a soma ponderada das quatro variáveis hemodinâmicas e fisiológicas disponíveis. O *score* total para cada aferição no tempo t foi calculado conforme a Equação 1:

$$NEWS_{total} = S_{FR} + S_{Temp} + S_{PAS} + S_{FC} \quad (1)$$

Onde S_x representa a pontuação individual atribuída a cada variável x de acordo com os intervalos de gravidade. A Tabela 2 detalha os critérios de pontuação implementados no algoritmo para a atribuição dos pesos.

Tabela 2. Critérios de pontuação do NEWS para as variáveis utilizadas.

Parâmetro Fisiológico	3	2	1	0	1	2	3
Frequência Respiratória (rpm)	≤ 8		9–11	12–20		21–24	≥ 25
Temperatura (°C)	≤ 35.0		35.1–36.0	36.1–38.0	38.1–39.0	≥ 39.1	
Pressão Arterial Sistólica (mmHg)	≤ 90	91–100	101–110	111–219			≥ 220
Frequência Cardíaca (bpm)	≤ 40		41–50	51–90	91–110	111–130	≥ 131

O score NEWS foi calculado para cada registro de medição individual, considerando os dados de sinais vitais. Esses dados apresentam variação de medição paciente/dia. Em média, são 8 coletas/dia, porém, dependendo da gravidade, esse número pode ser maior. Posteriormente, foram criadas três colunas de médias (móvel de 6h, 12h e 24h) para cada variável de sinais vitais usadas no cálculo da sepse, totalizando 12 novas medidas, agrupadas pelo atendimento. Exemplo: para a TEMP (temperatura) foram adicionadas as colunas de média móvel (MEDIA_MOVEL_6H_TEMP, MEDIA_MOVEL_12H_TEMP e MEDIA_MOVEL_24H_TEMP).

Após o cálculo do SCORE News e das colunas de médias móveis, foram definidas três momentos de predição da sepse, período de 6h, 12h e 24h de antecedência baseado na medição atual do paciente. Desse modo, foram definidas três safras: Sepse6h, Sepse12h e Sepse24h, atribuído um valor binário que representa a presença ou ausência de sepse no tempo futuro. Para determinar a presença ou ausência da sepse no futuro, adotou-se o *score* NEWS ≥ 7 pontos, conforme protocolo médico adotado na Santa Casa de Misericórdia de Belo Horizonte.

3.2. Algoritmos de Machine Learning

Neste trabalho são utilizados algoritmos supervisionados de classificação disponíveis na biblioteca scikit-learn ¹, também foi utilizado o algoritmo Boruta ².

Entre os algoritmos de *Ensemble*, destacam-se o **AdaBoost**, o **Gradient Boosting** e o **XGBoost**, variantes de *Boosting* que minimizam a função de perda sequencialmente, com ênfase crescente em eficiência e regularização. O **Random Forest**, por sua vez, emprega *Bagging* com múltiplas árvores independentes. Complementarmente, o algoritmo **Boruta** realiza seleção de variáveis comparando a importância dos atributos reais frente a variáveis aleatórias geradas internamente.

3.3. Métricas de Avaliação

A avaliação de modelos de classificação em Machine Learning fundamenta-se na análise da **Matriz de Confusão**, a partir da qual derivam-se métricas essenciais como a **Acurácia**, que mede a proporção global de acertos, a **Precisão** (capacidade de não classificar um exemplo negativo como positivo) e o **Recall** ou Sensibilidade (capacidade de identificar todos os exemplos positivos). Também foi utilizado o **F1-Score**, definido pela média harmônica entre precisão e recall, além da **AUC-ROC** (*Area Under the Receiver Operating Characteristic Curve*), que sumariza a performance do classificador em diferentes limiares de decisão.

¹<https://scikit-learn.org/stable/>

²Biblioteca BorutaPy em Python

4. Análise de Resultados

Os dados deste estudo revelam uma associação entre a deterioração clínica monitorada pelo score NEWS, o diagnóstico de sepse e o desfecho de óbito. A análise do desfecho hospitalar evidenciou que pacientes diagnosticados com sepse apresentam uma taxa de mortalidade de 88,2%, em comparação aos 6,4% de óbitos observados no grupo sem a patologia. Pacientes nos grupos de sepse e que tiveram óbito registraram medianas de 2.0 pontos e Q3 de 4.0 pontos, enquanto os grupos sem sepse mantiveram medianas de 1.0 e Q3 de 3.0 pontos.

Previamente ao treinamento dos classificadores, empregou-se o algoritmo Boruta para seleção de variáveis, o qual identificou, de forma consistente nos três horizontes temporais avaliados (6h, 12h e 24h), 19 variáveis (*features*) como estatisticamente relevantes para a predição de sepse. As variáveis selecionadas abrangeram sinais vitais, suas respectivas médias móveis nos intervalos de 6h, 12h e 24h, além de informações clínicas como tempo decorrido de internação, gênero e CID de internação.

O treinamento foi estruturado mediante validação cruzada ($k = 5$). O processo de *5-fold cross-validation* foi adotado para assegurar a consistência dos resultados, mitigando vieses de seleção na divisão entre os dados de treino e teste.

Os quatro algoritmos avaliados: Random Forest, XGBoost, AdaBoost e Gradient Boosting apresentaram desempenho elevado nos três horizontes temporais (6h, 12h e 24h), com ROC AUC $> 0,93$ em todos os cenários, com degradação progressiva das métricas conforme o horizonte se amplia. O XGBoost destacou-se pelo maior *recall* (0,91 em 6h; 0,88 em 12h; 0,85 em 24h) para prever a classe sepse=1, porém com precisão inferior ($\approx 0,53-0,54$), indicando maior taxa de falsos positivos. Random Forest e Gradient Boosting demonstraram melhor equilíbrio entre precisão e *recall*, com F1-Score de 0,89 para 6h e acurácia de 0,97. No horizonte de 24h, o Gradient Boosting sustentou o melhor F1-Score (0,85), sugerindo maior robustez para predições antecipadas. A escolha do modelo deve considerar o *trade-off* clínico: XGBoost é preferível quando a prioridade é maximizar a detecção de casos; Gradient Boosting e Random Forest são mais adequados quando se busca equilíbrio entre sensibilidade e precisão.

A Tabela 3 exibe a comparação dos algoritmos de Machine Learning para a classe positiva (sepse=1).

No que diz respeito ao desempenho dos algoritmos na identificação da ausência de sepse (sepse=0), todos os modelos apresentaram comportamento notavelmente uniforme e consistente ao longo dos três horizontes temporais avaliados. Em termos de precisão, os quatro classificadores mantiveram valores variando entre 0,95 a 0,99 em todos os cenários, com o **XGBoost** atingindo o maior valor pontual (0,99) nos horizontes de 6h e 12h e 0,98 em 24h.

5. Considerações Finais e Trabalhos Futuros

Os resultados demonstram que algoritmos *Ensemble* predizem a ocorrência de sepse com até 24 horas de antecedência, mantendo desempenho discriminativo elevado (ROC AUC $> 0,93$) mesmo nos horizontes temporais mais desafiadores. Em cenários nos quais a sensibilidade é prioritária, o Gradient Boosting apresenta-se como a alternativa mais indicada; em ambientes que demandam equilíbrio entre precisão e *recall*, Gradient Boos-

Tabela 3. Comparativo dos algoritmos de ML para predição de sepse

Alvo	Algoritmo	Prec.	Rec.	F1	Acur.	ROC AUC
SEPSE_6H	Random Forest	0.88	0.71	0.79	0.97	0.9706
	XGBoost	0.54	0.91	0.68	0.93	0.9757
	AdaBoost	0.84	0.75	0.79	0.97	0.9717
	Gradient Boosting	0.85	0.75	0.80	0.97	0.9747
SEPSE_12H	Random Forest	0.91	0.67	0.77	0.96	0.9579
	XGBoost	0.54	0.88	0.67	0.91	0.9637
	AdaBoost	0.87	0.70	0.77	0.96	0.9577
	Gradient Boosting	0.88	0.70	0.78	0.96	0.9624
SEPSE_24H	Random Forest	0.91	0.60	0.72	0.94	0.9395
	XGBoost	0.53	0.85	0.66	0.89	0.9468
	AdaBoost	0.85	0.64	0.73	0.94	0.9364
	Gradient Boosting	0.88	0.64	0.74	0.94	0.9442

ting e Random Forest mostram-se superiores. Os achados reforçam o potencial da Inteligência Artificial como ferramenta de suporte à decisão clínica na identificação precoce de pacientes em risco. Estudos futuros devem explorar a validação externa em diferentes populações e ambientes hospitalares, a integração de variáveis laboratoriais e a criação de alertas em tempo real integrados ao sistema de gestão hospitalar.

Referências

- Bomrah, S. et al. (2024). A scoping review of machine learning for prediction of sepsis and septic shock. *Critical Care*, 28:180.
- Covino, M. et al. (2023). Predicting icu admission and death in the emergency department: A comparison of six early warning scores. *Resuscitation*, 190:109876.
- ILAS (2024). Instituto latino-americano de sepse - dados e estatísticas. Acesso em: 21 dez. 2025.
- Li, X. et al. (2023). Machine learning algorithm to predict mortality in critically ill patients with sepsis-associated acute kidney injury. *Scientific Reports*, 13:5223.
- Ren, J. et al. (2024). Inverse set estimation and inversion of simultaneous confidence intervals. *Journal of the Royal Statistical Society Series C*.
- Singer, M., Deutschman, C. S., Seymour, C. W., et al. (2016). The third international consensus definitions for sepsis and septic shock (sepsis-3). *JAMA*, 315(8):801–810.
- Tan, K. M. et al. (2025). Advancing multi-organ disease care: A hierarchical multi-agent reinforcement learning framework. *Preprint*.
- Yin, C. et al. (2024). Sepsislab: Early sepsis prediction with uncertainty quantification and active sensing. In *Proceedings of the 30th ACM SIGKDD Conference*, pages 459–461.
- Zhao, C. et al. (2025). Diffmv: A unified diffusion framework for healthcare predictions with random missing views and view laziness. In *Proceedings of KDD*.