

Predição de sepse em pacientes de unidade de terapia intensiva: desenvolvimento e avaliação de algoritmos de inteligência artificial

Prediction of sepsis in intensive care unit patients: Development and evaluation of artificial intelligence algorithms

Antônio Diego Costa Bezerra^{1*}, Lucilane Maria Sales da Silva², Francisco José Maia Pinto³ e Robert Vinícius Oliveira Gonçalves⁴

¹Mestre em Saúde Coletiva, Universidade Estadual do Ceará (UECE); ^{2,3}Doutor em Saúde Coletiva, UECE; ⁴Engenheiro de Computação, Universidade Federal do Ceará (UFC)

Resumo

Introdução: a sepse, síndrome clínica com alto potencial de infecção e impacto nas unidades de terapia intensiva, necessita de rápida identificação para melhorar as chances de sobrevivência dos pacientes que apresentam essa complexa resposta desregulada do sistema à infecção. No entanto, devido à sobrecarga de informações clínicas do paciente, a identificação precoce é um desafio. **Objetivo:** avaliar algoritmos de inteligência artificial para predição da sepse em pacientes de unidade de terapia intensiva. **Metodologia:** foram utilizados dados de 659 pacientes de uma unidade intensiva de hospital público de referência, no Brasil, por meio de estudo de desenvolvimento com as tarefas de regressão e classificação. **Resultados:** foram testados cinco algoritmos, com resultados que apontam para *Random Forest* e *Gradient Boosting* com o melhor desempenho, com altos valores de F1-score (0,98) e acurácia (98,48%). **Conclusão:** o estudo apontou oportunidades por meio da avaliação de sistemas de apoio à decisão baseados em algoritmos de aprendizado de máquina.

Palavras-chave: Inteligência artificial; sepse; predição; Unidade de Terapia Intensiva. algoritmo.

Abstract

Introduction: Sepsis, a clinical syndrome with a high potential for infection and impact on Intensive Care Units, requires rapid identification to improve the chances of survival of patients who present this complex dysregulated response of the system to infection. However, early identification is a challenge due to the overload of clinical information from the patient. **Objective:** to evaluate artificial intelligence algorithms for predicting sepsis in Intensive Care Unit patients. **Methodology:** data from 659 patients from an intensive care unit of a reference public hospital in Brazil were used through a development study with regression and classification tasks. **Results:** five algorithms were tested, with results indicating that Random Forest and Gradient Boosting had the best performance, with high F1-score values (0.98) and accuracy (98.48%). **Conclusion:** the study identified opportunities by evaluating decision support systems based on machine learning algorithms.

Keywords: Artificial intelligence; Sepsis; Prediction; Intensive Care Unit; Algorithm.

INTRODUÇÃO

A sepse é uma síndrome clínica complexa, caracterizada por uma resposta desregulada do hospedeiro à infecção, representando um desafio significativo na saúde contemporânea, devido às alterações fisiopatológicas que os pacientes apresentam no internamento hospitalar. A sepse geralmente ocorre em resposta a uma infecção bacteriana, viral, fúngica ou parasitária, podendo levar à disfunção ou à falência de órgãos como coração, pulmões, rins e fígado. Seus sinais e sintomas incluem febre alta ou hipotermia, taquicardia, taquipneia, letargia e hipotensão, o que pode levar ao choque séptico, uma forma grave da síndrome que resulta em

pressão arterial baixa persistente, mesmo após tratamento adequado¹.

Com alta taxa de mortalidade e impacto significativo na qualidade de vida dos sobreviventes, a infecção é reconhecida como uma das principais causas de morte em todo o mundo. A incidência de sepse, no Brasil, é de aproximadamente 200 mil casos por ano, com mortalidade de 35% a 45% em casos de sepse grave². A terapia direcionada precoce pode estar associada à redução da mortalidade em pacientes com sepse e choque séptico, com importância para intervenções precoces no tratamento dessas condições³.

Diante desse quadro, se a rápida identificação e intervenção são cruciais para melhorar as perspectivas de sobrevivência dos pacientes, por outro lado, a identificação precoce da infecção continua sendo um problema clínico desafiador⁴, a começar pela unidade de terapia intensiva (UTI), que desempenha papel fundamental

Correspondente/Corresponding: *Antônio Diego Costa Bezerra – End: Av. Dr. Silas Munguba, 1700 – Itaperi, Fortaleza. Fortaleza (CE) (NUPEINSC). – E-mail: diegocostamjc@gmail.com

no tratamento de pacientes graves, principalmente, daqueles com sepse. Nesse contexto, a identificação precoce da patologia é uma tarefa desafiadora para os profissionais de saúde, devido à complexidade dos sinais e sintomas e à sobrecarga de informações clínicas. Frequentemente, a sepse só é diagnosticada quando já está em um estágio avançado, o que compromete as chances de recuperação do paciente⁵.

A identificação precoce desses pacientes com sepse em UTI é de grande importância clínica, pois permite a intervenção rápida e adequada, com a consequente melhoria das chances de sobrevivência e redução dos custos do tratamento. Essa condição é a principal geradora de custos nos setores público e privado, devido à necessidade de uso de equipamentos sofisticados, medicamentos caros, além de exigir a assistência qualificada e adequada da equipe de saúde⁶. Assim, surgem as tecnologias como potenciais inovações que podem contribuir para as práticas profissionais, sem substituir o olhar clínico, mas que aprimoram e agilizam os atendimentos.

Atualmente, o reconhecimento da sepse depende principalmente da experiência clínica e da interpretação subjetiva dos profissionais de saúde, o que pode levar a diagnósticos tardios ou imprecisos. Nos últimos anos, avanços significativos em algoritmos de inteligência artificial (IA) e aprendizado de máquina têm gerado um interesse crescente na aplicação dessas técnicas na saúde, substancialmente na identificação precoce de doenças complexas^{7,8}.

Os recursos para utilização de IA na predição de doenças indicam uma área em constante desenvolvimento. Embora ainda relativamente recente, a Organização Mundial da Saúde (OMS) tem destacado o potencial da IA para aprimorar a velocidade e a precisão do diagnóstico e da triagem de patologias. Além disso, a tecnologia tem sido reconhecida como uma ferramenta capaz de apoiar diversas ações de saúde coletiva, como a vigilância de agravos, incluindo a sepse, e a gestão de sistemas de saúde. O Brasil, por exemplo, tem avançado na adoção de tecnologias de IA na área da saúde, com iniciativas que buscam integrar esses recursos em diferentes aspectos do cuidado ao paciente hospitalizado. Nesse sentido, embora ainda haja desafios a serem enfrentados, como questões éticas e de segurança dos dados, a utilização da IA na predição de doenças apresenta um potencial significativo para aprimorar a prática clínica e melhorar os resultados de saúde, em diferentes contextos.

A OMS e organizações como o Ministério da Saúde (MS), no Brasil, incentivaram, recentemente, pesquisas na área de IA e de *Machine Learning* (ML), com foco nas UTIs⁹. No entanto, pesquisas literárias sistemáticas, realizadas em bases como *National Library of Medicine* (MEDLINE/PubMed), *Cochrane Library* e *Web of Science* apontaram fragilidades nos estudos já publicados, pois a maioria foi conduzida em hospitais privados, com

pouca comparação entre algoritmos, além de serem encontrados apenas dois estudos brasileiros, apenas um deles em hospital público.

No presente estudo, possibilitou-se a avaliação de melhores técnicas, com a coleta de dados em contextos clínicos variados. Ademais, avanços em tecnologias de sensores e coleta de dados via prontuários eletrônicos de saúde permitiram que elas fossem mais detalhadas e em tempo real, melhorando a qualidade dos dados de entrada para os modelos de ML^{10,11}.

Há ainda um potencial impacto no sistema público de saúde e na qualidade de vida dos pacientes, além de orientações para gestores em saúde. Por fim, este trabalho pode servir como um modelo para a aplicação de IA em outros contextos complexos, demonstrando o potencial da tecnologia na melhoria da assistência de saúde^{12,13}. De forma científica, esta pesquisa pode ainda ser considerada para implementação em larga escala, no Brasil, considerando o perfil epidemiológico dos pacientes de hospitais públicos gerais e possibilitando que os resultados sejam compreensíveis para os profissionais de saúde.

Diante disso, a presente pesquisa visa avaliar algoritmos para predição de sepse em pacientes adultos de UTI.

METODOLOGIA

Tipo de estudo e contexto

Este é um estudo de desenvolvimento por meio das tarefas de regressão e classificação. Os estudos de desenvolvimento tratam da criação, validação e avaliação de ferramentas e métodos¹⁴. A regressão e a classificação são utilizadas para prever classes com base em variáveis independentes (ou características), que são fornecidas como entrada ao modelo¹⁵, por meio de modelagem preditiva, e são realizadas com dados de pacientes admitidos em UTI neurológica de um hospital público localizado na Região Nordeste do Brasil. A unidade possui 15 leitos e atende pacientes com variadas condições clínicas e cirúrgicas associadas ao quadro neurológico. Foram considerados, para a coleta, dados dos últimos cinco anos, obtidos a partir dos prontuários dos pacientes.

As etapas foram seguidas com base no processo *Knowledge Discovery in Databases* (KDD), formalizado em 1989, de referência à procura de conhecimento a partir de bases de dados. Esse processo orienta o fluxo de mineração de dados, com etapas já estabelecidas: seleção dos dados, pré-processamento, mineração de dados, avaliação e análise, e pós-processamento¹⁶.

A primeira etapa foi a seleção dos dados, que compreendeu a coleta na UTI, entre os meses de junho a agosto de 2023. Na segunda etapa de pré-processamento, concebeu-se a limpeza dos dados, especialmente por se tratar de prontuários preenchidos por diferentes profissionais, entre os quais se verificou discrepância

semântica. No terceiro passo, os dados foram numericamente codificados e, no quarto, houve a normalização dos dados em escalas de valores com intervalos similares. Por último, procedeu-se à construção de variáveis.

Foram incluídos 659 pacientes na amostra, com 18 anos ou mais, e que foram internados na UTI entre julho de 2018 e julho de 2023. Pacientes que apresentavam dados incompletos para qualquer uma das variáveis utilizadas no estudo foram excluídos das análises. A coleta de dados nos prontuários foi realizada considerando a última evolução clínica, realizada por profissionais médicos ou de enfermagem. Diariamente, realizava-se uma auditoria para verificar o correto preenchimento dos dados coletados, evitando perdas ou erros de registro.

Variável-alvo

A ocorrência de sepse e choque séptico foi avaliada utilizando-se a evolução médica, a partir da definição da classificação de internação de doenças (CID-10) para o termo “septicemia” (A419). A variável-alvo foi composta de uma variável dicotômica, com valores sim ou não para sepse ou choque séptico, durante o tempo de internação.

Variáveis preditoras

A etapa de mineração das variáveis preditoras foi restrita aos dados disponíveis na data da última evolução do paciente na UTI, a qual resultou na identificação de 22 variáveis: “raça ou cor”, “sexo”, “idade”, “intervalo de tempo para diagnóstico”, “diagnosticado para algum critério de infecção”, “valores de creatinina sérica”, “valores de glicemia”, “valores de plaquetas”, “contagem de glóbulos brancos”, “valores de bilirrubina total”, “valores de Ph do sangue”, “valores de oximetria de pulso de saturação de oxigênio”, “escala de coma de Glasgow”, “histórias sugestivas de infecção”, “uso de vasopressor”, “critérios de disfunção orgânica”, “quantidade total de tempo permanecendo na UTI”, “frequência cardíaca”, “frequência respiratória”, “temperatura”, “pressão arterial sistólica”, “pressão arterial diastólica” e “valores de lactato”.

Como a maioria dos casos de sepse ocorre nas unidades intensivas, as diretrizes de 2016 do Sepsis-3¹⁷ incluíram novo modelo, denominado avaliação rápida de falha de órgãos relacionada à sepse (qSOFA), que considera distúrbios na pressão arterial sistólica, no estado mental e na frequência respiratória, justificando, assim, a necessidade de se trabalhar com variáveis clínicas e laboratoriais para melhor precisão¹⁸. Além disso, o estudo demonstrou que o desempenho é melhorado pela adição de resultados laboratoriais¹⁹.

Os pacientes com dados faltantes foram excluídos das análises. Optou-se por não realizar imputação de dados, uma vez que somente um número pequeno de pacientes apresentou dados faltantes, sem impacto sobre a capacidade preditiva do modelo.

A mineração de dados, terceira etapa do processo KDD, representou a fase de descoberta dos padrões por meio da aplicação de algoritmos. Foram testados os algoritmos de regressão e classificação do tipo Gradient Boosting (GB), K-Nearest Neighbors (KNN), Random Forest (RF), Naive Bayes (NB), Decision Tree e Logistic Regression (LR), na busca pelo melhor resultado na predição da sepse.

Treinamento do modelo

Para a construção do modelo, os dados foram importados para o *software Colaboratory Google*, e foram utilizadas as bibliotecas *Pandas*, *NumPy* e *Scikit-Learn* e da linguagem Python para a replicação de modelos, ou seja, conjuntos de códigos de computador já disponibilizados²⁰. Outros algoritmos como Redes Neurais Profundas (*Deep Learning*) e *Support Vector Machines (SVM)* com *Kernels Complexos* foram testados e apresentaram resultados abaixo de 70% nas métricas avaliadas.

Para a validação do modelo, foram realizados os seguintes passos: inicialmente, foi feita a importação das bibliotecas; em seguida, foi realizada a importação da função *train_test_split* do *scikit-learn*, para dividir os dados em conjuntos de treinamento e teste, sendo 80% e 20% dos dados, respectivamente. Na sequência, a classe *GridSearchCV* foi importada do módulo *model_selection*, técnica que permitiu encontrar os melhores parâmetros para o modelo por meio da busca na grade de valores especificados. Por serem dados consideravelmente não lineares, foram realizados cruzamentos multivariados entre as variáveis.

Foi utilizada, posteriormente, a função *cross_val_score* para realizar validação cruzada, visando avaliar o desempenho do modelo, especialmente por obtermos uma amostra considerada limitada.

Métricas de avaliação de desempenho

Na quarta etapa do processo KDD, realizou-se a importação das métricas de avaliação, utilizando o método *k-fold cross validation*. Nesse método, o banco foi subdividido em cinco conjuntos de dados. Em cada uma das cinco validações, uma parte diferente do modelo foi escolhida aleatoriamente para representar o grupo de teste, e o restante dos dados fez parte do conjunto de treinamento. As métricas de avaliação final são médias aritméticas dos cinco resultados obtidos ao fim de cada validação.

Utilizou-se o F1-score, métrica que combinou precisão (*precision*) e *recall*.

Ele foi calculado como:

$$\frac{2 \times \text{precision} \times \text{recall}}{\text{precision} + \text{recall}}$$

O F1-score varia de 0 a 1, em que 1 é a melhor pontuação possível. Utilizou-se ainda a acurácia, que mediu

a proporção de previsões corretas feitas pelo modelo. E, por último, a precisão, que é a proporção de verdadeiros positivos (instâncias corretamente classificadas como positivas) em relação a todas as instâncias classificadas como positivas pelo modelo²¹.

A área sob a curva (AUC) Receiver Operating Characteristic (ROC) também foi utilizada. Ela representa a taxa de verdadeiros positivos (sensibilidade) no eixo vertical e a taxa de falsos positivos (1 – especificidade) no eixo horizontal²².

Essas métricas foram escolhidas com o intuito de diminuir o número de falsos positivos, tendo em vista a gravidade da condição a ser detectada, assim como de falsos negativos, para reduzir a possibilidade de aplicação inadequada dos modelos na UTI.

Declaração ética

Esta pesquisa seguiu todos os preceitos éticos preconizados pela Resolução 466/12 do Conselho Nacional de Saúde, sendo aprovada pelo Comitê de Ética em Pesquisa com Seres Humanos do Hospital Geral de Fortaleza, sob o número 6.168.884. O trabalho seguiu ainda as normas gerais estabelecidas pela Lei nº 13.853, redação dada em 2019, denominada Lei Geral de Proteção de Dados Pessoais (LGPD)²³. A aplicação de Termo de Consentimento Livre e Esclarecido (TCLE) foi dispensada, já que todas as informações foram coletadas dos prontuários e com o mínimo de riscos para os pacientes.

Os autores declaram que a pesquisa foi realizada na ausência de quaisquer relações comerciais ou financeiras que pudessem ser interpretadas como potencial conflito de interesses.

RESULTADOS

Preencheram os critérios de inclusão do estudo 681 pacientes, dos quais 22 foram excluídos por apresentarem dados faltantes em uma ou mais variáveis preditoras, o que resultou na inclusão de 659 pacientes para as análises.

Os dados demográficos e as características gerais da população do estudo estão apresentados na Tabela 1.

Tabela 1 – Número e percentual de pacientes com sepse por variáveis sociodemográficas, Fortaleza, Ceará, 2024.

Variável	N	%
Sexo		
Feminino	341	51,7
Masculino	318	48,3
Idade (anos)		
18 a 24	55	8,3
25 a 31	22	3,3

32 a 38	80	12,1
39 a 45	12	1,8
46 a 52	57	8,6
53 a 59	108	16,4
60 a 67	127	19,4
68 a 74	89	13,5
≥ 75	109	16,6
Raça ou cor		
Branca	266	42,8
Preta	177	28,4
Parda	141	22,7
Amarela	38	6,1

Fonte: dados da pesquisa.

Entre os modelos testados, o GB construiu uma série de modelos fracos de árvore de decisão e os combinou para formar um modelo forte. De modo semelhante, o DT dividiu os dados em subconjuntos puros, permitindo que o modelo tome decisões com base nas características dos dados. Já o KNN classificou as instâncias com base na maioria das classes de seus k vizinhos mais próximos. A “proximidade” foi medida usando-se uma métrica de distância, como a distância euclidiana.

O RF funcionou como um *ensemble* de árvores de decisão. Ele criou várias árvores de decisão durante o treinamento e as combinou visando obter uma decisão mais precisa e estável. Já NB utilizou-se da probabilidade com base no teorema de Bayes. Ele assumiu que as *features* são independentes. O último analisado foi a RL que relacionou as variáveis independentes e a probabilidade de pertencer a uma classe aproximadamente linear.

As métricas de avaliação dos modelos estão apresentadas na Tabela 2.

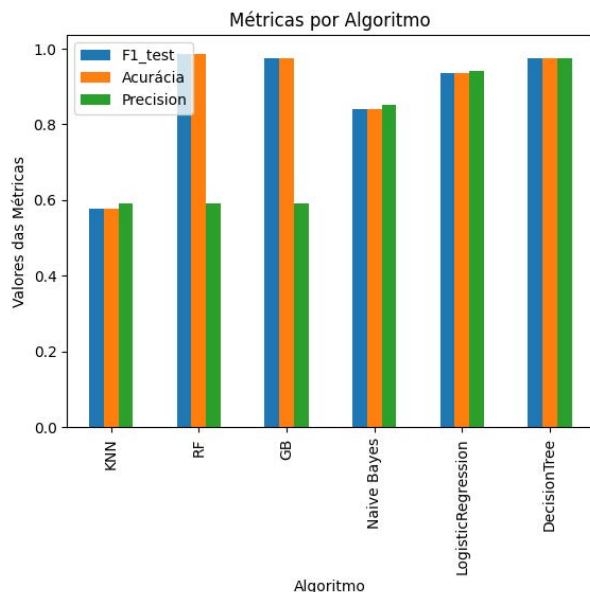
Quadro 1 – Métricas de avaliação de desempenho dos modelos na predição de sepse. Fortaleza, Ceará, 2024.

Algoritmo	F1_test	Acurácia	Precisão
Random Forest	0.98	0.98	0.59
Gradient Boosting	0.97	0.97	0.59
DecisionTree	0.96	0.96	0.96
Logistic Regression	0.91	0.91	0.91
Naive Bayes	0.83	0.83	0.85
K-Nearest Neighbors	0.57	0.57	0.59

Fonte: dados da pesquisa.

Entre os modelos comparados, estavam o melhor F1_test bem como a acurácia e, desse modo, o menor número de falsos negativos foi observado com os algoritmos RF e GB (Gráfico 1).

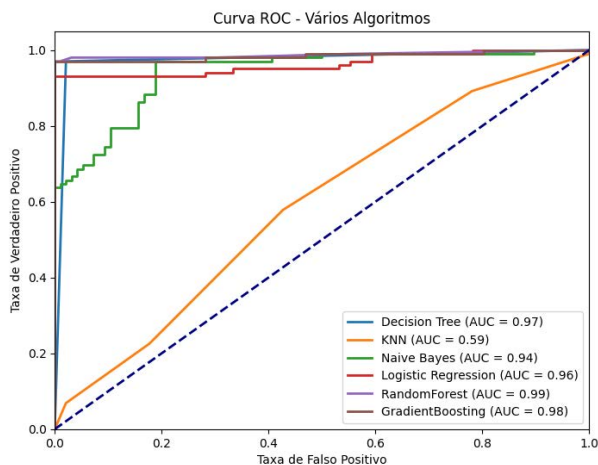
Gráfico 1 – Visualização gráfica da avaliação de desempenho dos modelos na predição de sepse. Fortaleza, Ceará, 2024.



Fonte: autoria própria.

O modelo com o algoritmo de RF apresentou o melhor desempenho quando consideradas as três métricas de avaliação, com uma AUC-ROC de 0,99 (Figura 1).

Figura 1 – Curva ROC dos modelos na predição da sepse. Fortaleza, Ceará, 2024.



Fonte: autoria própria.

DISCUSSÃO

Diversas são as formas de diagnóstico e manejo da sepse, especialmente por apresentar-se como uma res-

posta imune desregulada à infecção, podendo revelar-se de maneiras distintas no organismo humano. Logo, seu diagnóstico se torna, por vezes, impreciso e abstruso, devido aos sinais e sintomas não específicos, podendo ser desafiador, pois, mesmo na hodiernidade, não há um teste padrão-ouro que confirme o diagnóstico²⁴. Assim, o desenvolvimento de novas tecnologias e avaliação de preditores, como os que aqui serão apresentados, são importantes ferramentas para as unidades intensivas.

A predição precoce do início da sepse é fundamental para fornecer qualidade na assistência de saúde. Modelos de previsão de aprendizado de máquina têm demonstrado um bom desempenho conjunto com os sistemas tradicionais ou escalas das unidades hospitalares. Ao utilizar esses modelos, pode ser possível prever a sepse em pacientes e orientar as condutas dos profissionais de saúde, permitindo que eles ajam com maior urgência, monitorem os pacientes de forma mais eficaz e adotem medidas preventivas para melhorar o quadro de saúde dos indivíduos²⁴. Este estudo demonstrou ainda que o uso de técnicas de aprendizado profundo permitiu uma previsão precoce e precisa da sepse, com a utilização de dados de prontuários eletrônicos, o que pode melhorar os resultados clínicos e reduzir a morbidade e mortalidade associadas à sepse²⁵.

A partir dos resultados obtidos na análise dos 6 algoritmos desta pesquisa, identificou-se que o RF e o GB podem ser considerados bons em termos de F1-score, ambos alcançando valores acima de 0.97, o que indica uma excelente capacidade de balancear precisão e *recall*. Individualmente, o algoritmo RF obteve o melhor desempenho nas métricas avaliadas, com um F1-score de 0.98, uma acurácia de 98.48% e uma precisão de 59%. O GB também apresentou resultados positivos, com um F1-score de 0.97, uma acurácia de 97.47% e uma precisão também de 59%. Esses resultados sugerem que ambos os algoritmos são promissores para aplicação em previsão de dados de sepse.

A DT também apresentou um desempenho considerável, com um F1-score e precisão de 0.96. A LR, embora tenha uma pontuação inferior em comparação com os algoritmos anteriores, ainda demonstrou um desempenho de 0.91 na F1-score e precisão. O NB, apesar de ter uma precisão alta, apresentou um F1-score um pouco inferior, indicando um desafio na obtenção de um equilíbrio ideal entre precisão e *recall*. Por fim, o KNN obteve um F1-score de 0.57, uma acurácia de 57.58%, e uma precisão de 59%, tendo avaliações inferiores. Esses resultados dos diferentes algoritmos na previsão da sepse destacam as nuances entre eles e suas respectivas capacidades.

A partir das semelhanças entre RF e GB nos valores de F1_test e acurácia – sendo elas muito próximas (0.98), mas com uma precisão relativamente baixa (0.59) –, indica-se que esses modelos têm um bom equilíbrio entre precisão e *recall*, mas podem estar prevendo muitos falsos positivos²⁶. Ambos os modelos são baseados

no modelo de árvore, herdando suas vantagens, sendo robustos a *outliers*, e têm pouca interferência de ruído, possibilitando bons resultados. Essas observações sugerem ainda que, embora os modelos RF e GB tenham um desempenho promissor em termos de acurácia e F1-score, é importante considerar os *trade-offs* entre precisão, *recall* e a ocorrência de falsos positivos ao interpretar seus resultados.

Pesquisa com 2.436 pacientes, na China, demonstrou, através dos resultados gerais, que o modelo RF foi o melhor modelo de classificação para todos os três problemas de predição (gravidade, mortalidade e tempo de tratamento para sepse), com um F1 de 0,72, ainda menores do que nesta pesquisa²⁷. Os bons resultados apresentados aqui podem ser decorrentes da variabilidade e completude dos dados, bem como das características das variáveis consideradas. No entanto, é importante ressaltar a limitação em relação ao tamanho da amostra, que é consideravelmente menor, o que constitui uma restrição significativa na generalização dos resultados. Essa diferença no tamanho da amostra pode afetar a robustez e a validade externa dos resultados obtidos.

Um estudo que utilizou dados do Medical Information Mart for Intensive Care III (MIMIC-III), com dados de pacientes adultos internados em unidades de cuidados intensivos no Beth Israel Deaconess Medical Center, em Boston, obteve semelhanças em relação ao GB. Os resultados mostraram que ele apresentou maior precisão (0,948), o que pode apontar para uma limitação da pesquisa realizada aqui em relação à amostra, visto que a pesquisa anteriormente relatada²⁶ foi conduzida com 3.937 pacientes. No entanto, a pesquisa obteve uma acurácia de 0,954 e um escore F1 de 0,933, aproximando-se dos resultados obtidos neste estudo. Essa comparação destaca a importância do tamanho da amostra na generalização e na confiabilidade dos resultados em estudos de predição de sepse.

O DT, apesar de ser um modelo único, também apresentou um desempenho notável, destacando-se pela sua capacidade de apreender relações não lineares nos dados. Isso é confirmado por sua alta precisão. São resultados semelhantes aos encontrados em um único estudo brasileiro, que utilizou dados de um hospital do Nordeste, com 720 pacientes analisados. Nesse estudo, observou-se uma precisão de 0,97 para o algoritmo supracitado²⁸. A capacidade de a DT combinar e analisar várias variáveis dos pacientes com sepse permite que ela identifique quais são as mais importantes para fazer previsões precisas, como no caso da sepse. Ao examinar cada variável e como elas se relacionam umas com as outras, a árvore consegue determinar quais são os fatores mais relevantes para tomar decisões confiáveis. Essa capacidade de capturar relações complexas nos dados é uma das principais vantagens como modelo de aprendizado de máquina. Isso significa que ela pode lidar com dados complicados, como dos prontuários dos

pacientes e fornecer alertas que ajudam na tomada de decisões frente ao diagnóstico de infecção.

A LR, apesar de ser um modelo linear, mostrou-se eficaz, sugerindo que, em certos cenários, modelos mais simples podem ser suficientes. Isso pode significar que, em algumas situações, é melhor usar algo menos complexo. Por exemplo, um estudo²⁹ aponta que outros tipos de modelos podem ser melhores para entender como diferentes fatores afetam um resultado para predição de sepse. Além disso, a LR só funciona quando o resultado que queremos prever é uma categoria, o que nem sempre é o que precisamos nos pacientes com a infecção. Embora seja mais simples de entender e usar nos cenários de saúde, é importante lembrar que a escolha do modelo certo depende do problema que se está tentando resolver e dos dados disponíveis. Em alguns casos, um modelo mais complexo pode ser necessário para obter resultados mais precisos.

O NB apresentou um desempenho ligeiramente inferior em termos de F1_test e acurácia (0,83), mas demonstrou ter a maior precisão (0,85) entre os modelos testados. Isso sugere que o modelo pode ser mais confiável ao prever resultados positivos. Essa observação é consistente com um estudo recente³⁰, que analisou dados com características variáveis semelhantes às do presente estudo. A pesquisa destacou que as técnicas de ML possuem alta flexibilidade e escalabilidade em comparação com os métodos estatísticos tradicionais nas ciências da saúde, mesmo quando a amostra de dados é reduzida. Essas descobertas ressaltam a robustez e a utilidade dos modelos de ML, como o NB, na predição de resultados em contextos clínicos da saúde pública.

A alta precisão alcançada pelo NB indica que, embora outros algoritmos tenham uma precisão relativamente baixa, eles ainda podem ser úteis em contextos específicos, especialmente quando a precisão é de extrema importância, como nas UTIs. Os classificadores NB são simples e assumem independência condicional entre variáveis, dada a variável de resultado³¹. Essa observação é consistente com a ideia de que, mesmo em situações em que a precisão é crucial para o diagnóstico de pacientes, modelos mais simples podem fornecer resultados confiáveis e úteis. O fato de que o NB, apesar de sua simplicidade, alcançou alta precisão, destaca sua relevância e eficácia em determinados cenários hospitalares.

O modelo KNN teve um desempenho inferior, em comparação com os demais testados, apresentando valores significativamente mais baixos para F1_test e acurácia (0,57), apesar de ter a mesma precisão que RF e GB. Esses resultados destacam a importância da escolha adequada do algoritmo na predição da sepse em pacientes de unidades intensivas. É evidente que diferentes algoritmos podem produzir resultados substancialmente diferentes, mesmo quando a precisão é semelhante. Portanto, ao selecionar um modelo para tarefas críticas, como a predição da sepse, é crucial

considerar não apenas a precisão, mas também outras métricas de desempenho, como *F1-score* e acurácia, para garantir resultados confiáveis e precisos no manejo dos pacientes hospitalizados.

Estudos com pontuações semelhantes na avaliação apresentam comparações diretas difíceis, devido aos diferentes desenhos de estudo subjacentes, bem como definições incompletas de coortes e resultados, ou o uso de dados indisponíveis em tempo real^{29,32-34}. O possível impacto clínico da melhoria deste estudo é sugerido pela qualidade dos dados e modelos avaliados, resultando em menos falsos positivos³⁵. Esses resultados destacam a importância da contínua investigação e do refinamento de modelos de previsão de sepse para melhorar os cuidados de saúde e reduzir a morbidade e mortalidade associadas a essa condição grave.

Como já foi mencionado, o estudo apresentou limitações quanto à quantidade de dados disponíveis. Tal obstáculo se desdobra nos sistemas de saúde mundiais. Apesar dos avanços, a interoperabilidade dos sistemas digitais e o fornecimento de dados continuam a constituir uma questão em aberto, conduzindo a desafios na integração. Como resultado, o potencial que as informações hospitalares oferecem, em termos de compreensão e melhoria dos cuidados, ainda não foi plenamente concretizado.

Paralelamente, destaca-se a necessidade de aplicações de pesquisas como esta nos serviços de saúde, além de novos desenvolvimentos de algoritmos e avaliações, com o uso de dados semelhantes e com maior disponibilidade³⁴. Por fim, destacam-se as premissas futuras do aprendizado de máquina na previsão da sepse, mas ressalta-se a importância de abordagens interpretáveis e transparentes para garantir a aceitação e a implementação desses modelos nos sistemas de saúde, especialmente na saúde pública³⁶.

CONCLUSÃO

Esta pesquisa oferece resultados importantes sobre a acurácia e a precisão, após o desenvolvimento e a avaliação de diferentes algoritmos de ML na predição da sepse em pacientes de UTI. Com base nos resultados obtidos, é recomendável o uso dos modelos RF, GB ou DT para tarefas de classificação semelhantes às deste estudo, pois esses modelos demonstraram um equilíbrio entre precisão e *recall*. No entanto, é fundamental considerar o contexto clínico específico ao selecionar o algoritmo apropriado para aplicação prática.

Além disso, considerando os benefícios da utilização de modelos de aprendizagem automática na detecção precoce da sepse em pacientes de risco, há potencial para a implementação desses modelos como sistemas de apoio à decisão clínica. Integrando-os com recomendações de diretrizes para o tratamento e gestão da doença, esses sistemas podem oferecer suporte aos profissionais de saúde, facilitando a identificação rápida

e precisa da sepse e melhorando os resultados clínicos dos pacientes.

Portanto, este estudo além de destacar a eficácia dos modelos de ML na predição da sepse, também aponta para oportunidades de aprimorar a prática assistencial de saúde e melhorar o cuidado ao paciente por meio da implementação de sistemas de apoio à decisão baseados em algoritmos de aprendizado de máquina.

REFERÊNCIAS

1. Singer M, Deutschman CS, Seymour SW, Shankar-Hari M, Annane D, Bauer M, et al. The third international consensus definitions for sepsis and septic shock (Sepsis-3). *Jama*. 2016;315(8):801-10. doi:10.1001/jama.2016.0287
2. Almeida NRC de, Pontes GF, Jacob FL, Deprá JVS, Porto JPP, Lima FR de, et al. Análise de tendência de mortalidade por sepse no Brasil e por regiões de 2010 a 2019. *Rev Saúde Pública*. 2022;56:25. doi:https://doi.org/10.11606/s1518-8787.2022056003789
3. Herrán-Monge R, Muriel-Bombín A, García-García MM, Merino-García PA, Martínez-Barrios M, Andaluz D, et al. Epidemiology and changes in sepsis mortality after the implementation of the Surviving Sepsis Campaign guidelines. *J Intensive Care Med*. 2019;34(9):740-750. doi:10.1177/0885066617711882.
4. Rababa M, Bani Hamad D, Hayajneh AA. Sepsis assessment and management in critically ill adults: a systematic review. *PLoS One*. 2022;17(7):e0270711. doi: 10.1371/journal.pone.0270711
5. Husabø G, Nilsen RM, Flaatten H, Solligård E, Frich JC, Bondevik GT. Early diagnosis of sepsis in emergency departments, time to treatment, and association with mortality: an observational study. *PLoS One*. 2020;15(1):e0227652. doi: 10.1371/journal.pone.0227652
6. Gonçalves LS, Amaro ML de M, Romero Ade LM, Schamne FK, Fresatto JL, Bezerra CW. Implantação de algoritmo de inteligência artificial para detecção da sepse. *Rev Bras Enferm*. 2020;73(3):e20180421. doi:https://doi.org/10.1590/0034-7167-2018-0421
7. Pepic I, Feldt R, Ljungström L, Torkar R, Dalev D, Söderholm HM, et al. Early detection of sepsis using artificial intelligence: a scoping review protocol. *Syst Rev*. 2021;10(1):1-7. doi: 10.1186/s13643-020-01561-w
8. Moor M, Rieck B, Horn M, Jutzeler CR, Borgwardt K. Early prediction of sepsis in the ICU using machine learning: a systematic review. *Front Med*. 2021;8:607952. doi: 10.3389/fmed.2021.607952
9. Ministério da Saúde (BR). Revolução da inteligência artificial: uso na saúde traz novas possibilidades. Biblioteca Virtual em Saúde. 2023.
10. Teng AK, Wilcox AB. A review of predictive analytics solutions for sepsis patients. *Appl Clin Inform*. 2020;11(3):387-98. doi: 10.1055/s-0040-1710525
11. Vaz TA. Modelo de dados para treinamento de inteligência artificial na pesquisa em saúde: um estudo prático sobre infecções hospitalares [dissertação]. Porto Alegre: Hospital de Clínicas de Porto Alegre; 2017. 104p.
12. Schinkel M, van der Poll T, Wiersinga WJ. Artificial Intelligence for Early Sepsis Detection: A Word of Caution. *Am J Respir Crit Care Med*. 2023;207(7):853-4. doi: 10.1164/rccm.202212-2284VP
13. Wu M, Du X, Gu R, Wei J. Artificial intelligence for clinical decision support in sepsis. *Front Med*. 2021;8:665464. doi: 10.3389/fmed.2021.665464

14. Polit DF, Beck CT, Hungler BP. Fundamentos de pesquisa em enfermagem: métodos, avaliação e utilização. São Paulo: Artmed Editora; 2004.
15. Hair JF, Black WC, Babin BJ, Anderson RE, Tatham RL. Análise multivariada de dados. Porto Alegre: Bookman editora; 2009. 139p.
16. Clark P, Niblett T. The CN2 induction algorithm. Mach Learn. 1989;3:261-83.
17. Seymour CW. Assessment of clinical criteria for sepsis: for the Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3). Jama. 2016;315(8):762-74. doi: 10.1001/jama.2016.0288
18. Churpek MM, Snyder A, Han X, Sokol S, Petti N, Howell MD, et al. Quick sepsis-related organ failure assessment, systemic inflammatory response syndrome, and early warning scores for detecting clinical deterioration in infected patients outside the intensive care unit. Am J Respir Crit Care Med. 2017;195(7):906-11. doi: 10.1164/rccm.201604-0854OC
19. Benitz WE, Achten NB. Technical assessment of the neonatal early-onset sepsis risk calculator. Lancet Infect Dis. 2021;21(5):e134-40. doi: 10.1016/S1473-3099(20)30490-4
20. Spinak E. Reprodução e replicação na pesquisa científica – parte 1 [online]. SciELO em Perspectiva, 2023 [viewed 2025 Feb 12]. Available from: <https://blog.scielo.org/blog/2023/05/19/reproducao-e-replicacao-na-pesquisa-cientifica-parte-1/>
21. Strauss E, Villas Bôas Júnior M, Ferreira WLL. A importância de utilizar métricas adequadas de avaliação de performance em modelos preditivos de machine learning. Projectus. 2022;7(2):52-62.
22. Santos MK, Ferreira Junior JR, Wada DT, Tenório APM, Nogueira-Barbosa MH, Azevedo-Marques PM de. Inteligência artificial, aprendizado de máquina, diagnóstico auxiliado por computador e radiômica: avanços da imagem rumo à medicina de precisão. Radiol Bras. 2019;52(06):387-96. doi: <https://doi.org/10.1590/0100-3984.2019.0049>
23. Brasil. Lei Geral de Proteção de Dados Pessoais (LGPD). Lei nº 13.853, de 8 de julho de 2019. Altera a Lei nº 13.709, de 14 de agosto de 2018, para dispor sobre a proteção de dados pessoais e para criar a Autoridade Nacional de Proteção de Dados; e dá outras providências. Diário Oficial da União. 2019.
24. Bezerra ADC, Maciel N de S, Silva Filho LS da, Mendes AS, Gois FNB, Silva LMS da. Efetividade de algoritmos de inteligência artificial para predição de sepse em adultos de unidades de terapia intensiva: revisão de escopo. Rev Interfaces: Saúde, Humanas e Tecnologia. 2023;11(4):3180–90. doi: <https://doi.org/10.16891/2317-434X.v11.e3.a2023.pp3180-3190>
25. Islam KR, Prithula J, Kumar J, Tan TL, Reaz MBI, Sumon MSI, et al. Machine learning-based early prediction of sepsis using electronic health records: a systematic review. J Clin Med. 2023;12(17):5658. doi:10.3390/jcm12175658.
26. Li K, Shi Q, Liu S, Xie Y, Liu J. Predicting in-hospital mortality in ICU patients with sepsis using gradient boosting decision tree. Medicine (Baltimore). 2021;100(19): e25813. doi: 10.1097/MD.00000000000025813
27. Su L, Xu Z, Chang F, Ma Y, Liu S, Jiang H, et al. Early prediction of mortality, severity, and length of stay in the intensive care unit of sepsis patients based on sepsis 3.0 by machine learning models. Front Med. 2021;8:664966. doi: 10.3389/fmed.2021.664966
28. Pessoa SMB, Oliveira BS de S, Santos WG dos, Oliveira ANM, Camargo MS, Matos DLAB de, et al. Predição de choque séptico e hipovolêmico em pacientes de unidade de terapia intensiva com o uso de machine learning. Rev Bras Ter Intensiva. 2023;34(4):477-83. doi: <https://doi.org/10.5935/0103-507X.20220280-pt>
29. Bekhit O El SM, Algameel AA, Eldash HH. Application of pediatric index of mortality version 2: score in pediatric intensive care unit in an African developing country. Pan Afr Med J. 2014;17:185. doi: 10.11604/pamj.2014.17.185.2818
30. Gholamzadeh M, Abtahi H, Safdari R. Comparison of different machine learning algorithms to classify patients suspected of having sepsis infection in the intensive care unit. Inform Med Unlocked. 2023;38:101236. doi: <https://doi.org/10.1016/j.imu.2023.101236>
31. Gupta A, Liu T, Shepherd S. Clinical decision support system to assess the risk of sepsis using tree augmented Bayesian networks and electronic medical record data. Health Informatics J. 2020;26(2):841-61. doi: 10.1177/1460458219852872
32. Bloch E, Rotem T, Cohen J, Singer P, Apterstein Y. Machine learning models for analysis of vital signs dynamics: a case for sepsis onset prediction. J Healthc Eng. 2019;5930379. doi: 10.1155/2019/5930379
33. Churpek MM, Yuen TC, Winslow C, Meltzer DO, Kattan MW, Edelson DP. Multicenter comparison of machine learning methods and conventional regression for predicting clinical deterioration on the wards. Crit Care Med. 2016;44(2):368. doi: 10.1097/CCM.0000000000001571
34. Johnson AEW, Poll TJ, Shen L, Lehman L-W H, Feng M, Ghassemi M, et al. MIMIC-III, a freely accessible critical care database. Sci Data. 2016;3(1):1-9. doi: 10.1038/sdata.2016.35.
35. Rajkomar A, Oren E, Chen K, Dai AM, Hajaj NM, Hardt M, et al. Scalable and accurate deep learning with electronic health records. NPJ Digit Med. 2018;1(1):18. doi: 10.1038/s41746-018-0029-1
36. Fleuren LM, Klausch TLT, Zwager CL, Schoonmade LJ, Guo T, Roggeveen LF, et al. Machine learning for the prediction of sepsis: a systematic review and meta-analysis of diagnostic test accuracy. Intensive Care Med. 2020;46(3):383-400. doi:10.1007/s00134-019-05872-y.

Submetido em 11/11/2024

Aceito em 29/01/2025