

ARTIGO DE REVISÃO

Técnicas de aprendizagem de máquina aplicadas à identificação de pacientes com sepse

Machine learning techniques applied to the identification of patients with sepsis

Técnicas de aprendizaje automático aplicadas a la identificación de pacientes con sepsis

Fabiana Guerra Pimenta,¹ André Luiz Alvim,² Paulo Meirelles.³

¹Graduação em Enfermagem. Especialista em Controle de Infecção Hospitalar. Mestre em Ciências Cirúrgicas e Oftalmológicas. Fundação Hospitalar do Estado de Minas Gerais – FHEMIG, Belo Horizonte, MG, Brasil.

²Doutorando e mestre em Enfermagem pela Universidade Federal de Minas Gerais. Especialista em Controle de Infecções. MBA em Auditoria e Gestão da Qualidade. Belo Horizonte, MG, Brasil.

³Professor Doutor, Departamento de Informática em Saúde, Universidade Federal de São Paulo – UNIFESP. Pesquisador-colaborador do Instituto de Matemática e Estatística da Universidade de São Paulo. São Paulo, SP, Brasil.

Recebido em: 29/01/2020

Aceito em: 13/xx04/2020

Disponível online: 13/04/2020

Autor correspondente:

Fabiana Guerra Pimenta

faguepi@gmail.com

RESUMO

Objetivo: Descrever com base na literatura, as técnicas de aprendizagem de máquina na identificação de casos de sepse. **Método:** Revisão integrativa da literatura que foi realizada nas bases de dados LILACS e PUBMED, no período de 2013 a 2019. **Resultados:** Foram incluídos seis estudos (100%) que atenderam aos critérios estabelecidos. Verificou-se que a metade (50,0%) foi publicada em revistas científicas que tiveram fator de impacto avaliado acima de 2,0, destacando os periódicos: *Journal of Biomedical Informatics* (2.95), *BMJ Open* (2.376) e *Computers in Biology and Medicine* (2.286). Todos os artigos mostraram predomínio da utilização do modelo de aprendizado de máquina supervisionado, pelo fato de ter inserido os padrões de identificação de sepse nos algoritmos que identificam os casos. **Conclusão:** Os algoritmos para a identificação e predição de sepse vêm sendo cada vez mais utilizados na prática clínica, mostrando uma tendência favorável para o crescimento e disseminação desses modelos entre os serviços de saúde.

Descritores: Inteligência artificial; Aprendizagem de máquina; Infecção; Sepsis.

ABSTRACT

Objective: We describe based on the literature about machine learning techniques in the identification of sepsis cases. **Method:** Method: Integrative literature review that was conducted in the LILACS and PUBMED databases, from 2013 to 2019. **Results:** Six studies (100%) that met the established criteria were included. It was found that half (50.0%) were published in scientific journals that had an impact factor rated above 2.0, highlighting the journals: *Journal of Biomedical Informatics* (2.95), *BMJ Open* (2,376) and *Computers in Biology and Medicine* (2,286). All articles showed a predominance of the use of the supervised machine learning model, due to the fact that it inserted the sepsis identification patterns in the algorithms that identify the cases. **Conclusion:** Algorithms for the identification and prediction of sepsis have been increasingly used in clinical practice, showing a favorable trend for the growth and dissemination of these models.

Descriptors: Artificial intelligence; Machine learning; Infection; Sepsis.

RESUMEN

Objetivo: Describir, con base en la literatura, técnicas de aprendizaje automático en la identificación de casos de sepsis. **Método:** Revisión bibliográfica integral que se realizó en las bases de datos LILACS y PUBMED, de 2013 a 2019. **Resultados:** Se incluyeron seis estudios (100%) que cumplieron con los criterios establecidos. Se encontró que la mitad (50.0%) se publicaron en revistas científicas que tenían un factor de impacto calificado por encima de 2.0, destacando las revistas: *Journal of Biomedical Informatics* (2.95), *BMJ Open* (2.376) y *Computers in Biology and medicina* (2,286). Todos los artículos mostraron un predominio del uso del modelo supervisado de aprendizaje automático, debido a que insertó los patrones de identificación de sepsis en los algoritmos que identifican los casos. **Conclusión:** Los algoritmos para la identificación y predicción de sepsis se han utilizado cada vez más en la práctica clínica, mostrando una tendencia favorable para el crecimiento y la difusión de estos modelos entre los servicios de salud.

Descriptores: *Inteligencia artificial; Aprendizaje automático; Infección; Septicemia.*

INTRODUÇÃO

A sepsé é definida como a disfunção orgânica causada por uma resposta desordenada do organismo a um processo infeccioso. Já o choque séptico, diz respeito a uma anormalidade circulatória e celular/metabólica secundária a sepsé, com a presença de hipotensão e utilização de vasopressores para manutenção da pressão arterial média do paciente acima de 65 mmHg e o lactato maior que 2 mmol/L, mesmo após adequada ressuscitação volêmica.¹

Sepsé e choque séptico são considerados problemas de saúde pública que afetam milhões de pessoas em todo o mundo, e matam um a cada quatro pacientes diagnosticados. A identificação precoce e o manejo apropriado, nas horas iniciais dos quadros de sepsé e choque séptico, proporcionam melhores resultados clínicos e reduzem consideravelmente as taxas de morbimortalidade.²

Com o objetivo de auxiliar a identificação dos casos, os profissionais de saúde realizam a avaliação dos sinais e sintomas utilizando prontuários e também, a beira leito. Além disso, destaca-se o método que pode ser utilizado na identificação precoce da sepsé de maneira rápida e objetiva, sendo descrito como “técnica de aprendizagem de máquina”.

O aprendizado de máquina desenvolve técnicas computacionais e a construção de sistemas que podem adquirir conhecimento de forma programada ou automática, como por exemplo, a identificação dos sinais e sintomas da sepsé.³ A aprendizagem de máquina aplicada na área da saúde consiste na elaboração de uma sequência finita de ações executáveis, conhecida como *algoritmo* que auxilia a tomada de decisão baseada em experiências acumuladas por meio da solução de problemas e/ou identificação de casos de maneira correta.³

Esse algoritmo pode ser do tipo supervisionado ou não supervisionado. O método supervisionado fornece ao programa de aprendizado um conjunto de exemplos para o treinamento do *software*, com padrões esperados ou pré estabelecidos.³ Normalmente, determina-se previamente um conjunto de parâmetros com a finalidade de construir um classificador que possa identificar corretamente a situação em que estes dados estão definidos. No aprendizado não supervisionado, o algoritmo não necessita de exemplos pelo fato de procurar e determinar os padrões, agrupando-os de forma aprendida e de maneira automática.³

Essas técnicas de aprendizagem de máquina vêm sendo empregadas nos serviços de saúde visando à melhoria da qualidade na assistência ao paciente e a identificação mais precisa de casos associados ao agravo. No entanto, a temática ainda é incipiente na literatura e desconhecida por alguns profissionais, justificando a elaboração deste estudo.

Objetivou-se descrever, com base na literatura, as técnicas de aprendizagem de máquina na identificação dos casos de sepsé.

MÉTODOS

Trata-se de uma revisão integrativa da literatura. Para a elaboração do estudo, as seguintes etapas foram percorridas:⁴

- Estabelecimento da hipótese ou questão da pesquisa
- Amostragem ou busca na literatura
- Categorização dos estudos
- Avaliação dos estudos incluídos na revisão
- Interpretação dos resultados
- Síntese do conhecimento ou apresentação da revisão.

Na primeira etapa, foi definida a questão norteadora. Formulou-se a seguinte pergunta de pesquisa: <Como a literatura aborda as técnicas de aprendizagem de máquina na identificação dos casos de sepsé?>

Para a seleção dos estudos, foi realizado levantamento bibliográfico nas bases de dados PubMed e Literatura Latino-americana em Ciências da Saúde (LILACS). Para estratégia de recuperação da informação científica na primeira base de dados, respectivamente, utilizou-se os seguintes descritores: “*Artificial Intelligence*”, “*Machine Learning*”, “*Un-supervised Machine Learning*”, “*Supervised Machine Learning*”, “*Infection*”, “*Sepsis*”, “*Neonatal Sepsis*”. Todas as palavras-chave fazem parte dos descritores MeSH. Já a busca na LILACS contemplou os termos: “Inteligência artificial”, “Infeção”, “Sepsé” e “Sepsé neonatal”.

Utilizaram-se os seguintes critérios de inclusão: artigos publicados em espanhol, inglês e português, no período de 2013 a 2019. Foram excluídos da pesquisa os textos em forma de apostilas, manuais e/ou livros que não abordavam sobre as técnicas de aprendizagem de máquina aplicadas na identificação da sepsé.

Para coleta de dados foi elaborado um formulário pelos próprios autores, contendo as seguintes variáveis: título do artigo/referência; objetivo e resultados. Para aprofundamento da temática, foi realizada uma síntese da produção científica, por meio de um quadro sinóptico, descrevendo os principais achados dos pesquisadores em relação ao objeto de estudo.

A análise descritiva dos dados ocorreu por meio da técnica de estatística descritiva simples para apresentação de valores absolutos e relativos. Posteriormente, os estudos foram discutidos por meio da síntese dos diálogos dos autores, e analisados com base na literatura.

RESULTADOS E DISCUSSÃO

Foram incluídos seis estudos (100%) que atenderam aos critérios estabelecidos. O Quadro 1 apresenta a síntese dos artigos que compuseram a amostra.

Do total de artigos sobre o assunto em questão, verificou-se que a metade (50,0%) foi publicada em revistas científicas que tiveram fator de impacto avaliado acima de 2,0, destacando os periódicos: *Journal of Biomedical Informatics* (2.95), *BMJ Open* (2.376) e *Computers in Biology and Medicine* (2.286) (Tabela 1).

Quadro 1. Síntese dos artigos incluídos no estudo, no período de 2013 a 2019.

Título do Artigo/Referência	Objetivo	Resultados
Development and validation of a diagnostic model for early differentiation of sepsis and non-infectious SIRS in critically ill children - a data-driven approach using machinelearning algorithms. ⁵	Desenvolver e validar um modelo de diagnóstico para a identificação/diferenciação do diagnóstico de Sepse e síndrome de resposta inflamatória sistêmica (SIRS) não infecciosa durante o curso clínico da doença, baseado em parâmetros que são utilizados rotineiramente e que podem ser implementados facilmente na prática clínica em pacientes pediátricos.	Modelo que inclui quatro parâmetros clínicos e quatro laboratoriais foi identificado como o melhor modelo após avaliação estatística. Apresentando resultados de 0,98 de acurácia na diferenciação do diagnóstico de Sepse / SIRS não infecciosa. Elaborado um segundo modelo de de predição de que obteve uma acurácia de 0,78 (IC 95%: 0,70-0,87). Quando solicitando que todos os casos de sepsse fossem classificados, 100% dos casos (IC 95%: 87–100%) e, 28% (IC 95%: 20–38%) da SIRS não infecciosa foram classificados corretamente.
Multicentre validation of a sepsis prediction algorithm using only vital sign data in the emergency department, general ward and ICU. ⁶	Validar uma aprendizagem de máquina baseada em algoritmo de predição de sepsse (InSight) para a detecção e predição baseada em três padrões de ouro relacionados à sepsse, usando seis sinais vitais. Avaliar a robustez do modelo de aprendizagem de máquina na presença de falta de dados e, realizar a personalização para que dados específicos, usando transferência de aprendizagem e generalização para novas configurações.	Ao avaliar o InSight, o mesmo apresenta uma acurácia para a detecção de sepsse de 0,92 (IC 95% 0,90 a 0,93), para início da sepsse grave de 0,87 (IC 95% 0,86 a 0,88) e para choque séptico de 0,99 (IC 95% 0,9991 a 0,9994), que é superior a acurácia identificadas com as escalas utilizadas atualmente. O InSight também foi capaz de realizar a detecção e predição de casos de sepsse grave 4 horas antes do início da mesma, apresentando uma acurácia de 0,85 (IC 95% 0,79 a 0,91). Significativamente maior que o tempo de detecção / predição das escalas de SIRS, MEWS (Modified Early Warning Scoring), SOFA (Sepsis-related Organ Failure Assessment). Estas previsões possuem uma sensibilidade e especificidade superiores a destas mesmas escalas.
Early Sepsis Detection in Critical Care Patients Using Multiscale Blood Pressure and Heart rate Dynamics. ⁷	Predizer o início da sepsse com quatro horas de antecedência, utilizando sinais vitais comumente mensurados e, características derivadas das formas de onda do ECG e da pressão sanguínea.	Os recursos do modelo 1 (Entropy) sozinhos alcançaram uma acurácia na curva ROC 0,67, com precisão geral de 47% no total testes (p menores que $\alpha = 0,05$). Em seguida, o modelo baseado na combinação dos recursos de EMR e características sociodemográficas do paciente (Modelo 2) acurácia na curva ROC de 0,70, com uma precisão geral de 50%. O último modelo, este combinando todas as características dos anteriores (Modelo 3), o classificador alcançou um acurácia na curva ROC 0,78, com uma precisão geral de 61%.
Learning representations for the early detection of sepsis with deep neural networks. ⁸	Desenvolver modelos de detecção para o estágio inicial de sepsse através de metodologia de deep learning (aprendizagem profunda) e, comparar a viabilidade e desempenho desta nova metodologia de aprendizagem com as do método de regressão com métodos temporais extração de recursos.	Este novo modelo foi denominado DNF e, apresentou como resultado uma acurácia um pouco maior do a apresentada pelo InSight (de 5,7 a 8,5 maior). Outro modelo foi testado, (denominado LSTM), apresentou performance de acurácia de 0,929, também maior do que a da apresentada na última versão do InSight. Conclui-se que sensibilidade do modelo LSTM foi de 91% e ligeiramente menor quando comparada ao critério SIRS (93%). LSTM apresentou uma especificidade maior, de 94%.
Septic shock prediction for ICU patients via coupled HMM walking on sequential contrast patterns. ⁹	Explorar dados vitais (pressão arterial média, frequência cardíaca e pressão arterial) para desenvolver a previsão antecipada de choque séptico. Transformar os dados de treinamento originais (dados vitais básicos) em uma série temporal de padrões capaz de prever o choque séptico.	Evidenciado que a integração dos padrões pode ajudar a fornecer uma avaliação do risco de choque séptico. Também evidenciado que os modelos desenvolvidos podem prever eventos de choque séptico usando séries temporais de padrões de contraste, que têm desempenhos comparativos como modelos anteriores. No entanto, deve-se notar que a aplicação de modelos complexos de choque séptico e as taxas de detecção aceitáveis na prática atual foram limitadas ao uso de medidas clínicas tradicionais, como o APACHE-III.
High-performance detection and early prediction of septic shock for alcohol-use disorder patients. ¹⁰	Desenvolver um algoritmo de predição de sepsse de alto desempenho (InSight), que supera os métodos existentes para populações de pacientes com transtornos do uso de álcool.	Como resultado para pacientes que apresentam transtornos do uso de álcool, o InSight apresentou 93% de sensibilidade, possibilitando reduzir alarmes falsos em mais de 80% em relação a outras ferramentas de detecção. O odds ratio de diagnóstico do InSight é 30 vezes maior do que as ferramentas mais utilizadas atualmente para detecção de casos de sepsse. O algoritmo também supera os métodos comparáveis em relação às horas de previsão do choque séptico antes do início.

Nota: SIRS = Síndrome da Resposta Inflamatória Sistêmica.

Tabela 1. Fator de impacto das revistas científicas avaliadas para o estudo.

Periódico	Fator de impacto
Journal of Biomedical Informatics	2.95
BMJ Open	2.376
Computers in Biology and Medicine	2.286
BMC Pediatr	1.983
Journal of Electrocardiology	1.363
Annals of Medicine and Surgery	Não consta

Em relação à delimitação temporal, observa-se nesta revisão integrativa que três artigos selecionados são do ano de 2017 (50,0%), dois (33,3%) de 2018 e apenas um (16,7%) de 2016. Ainda no que tange ao recorte temporal, foi notória a ausência de estudos publicados entre 2013 e 2015, predominando estudos recentes em relação ao tema (Tabela 2).

Tabela 2. Delimitação temporal dos artigos incluídos na revisão integrativa.

Ano	Número	Porcentagem (%)
2019	0	0,0
2018	2	33,3
2017	3	50,0
2016	1	16,7
2015	0	0,0
2014	0	0,0
2013	0	0,0

No primeiro artigo selecionado para esta revisão,⁵ os autores relatam que na construção inicial do modelo de avaliação de aprendizagem de máquina, 46 variáveis foram consideradas para o desenvolvimento, o que caracterizou um aprendizado supervisionado de máquina. Além disso, os diagnósticos de SIRS e sepse foram realizados prospectivamente e revisados por dois médicos (duplo cego) para confirmação. Os resultados da aplicação deste modelo constam no quadro 1.

O segundo trabalho mostrou que o modelo é caracterizado como um aprendizado supervisionado de máquina, uma vez que na sua construção, foram assinalados os critérios das escalas *SOFA* e *MEWS* (ambas escalas, utilizadas em terapia intensiva para a avaliação do paciente). Além de critérios de SIRS, com o objetivo de comparar o desempenho do modelo utilizando cada escala, ou uma combinação delas, na identificação de quadros de sepse através de uma transferência de aprendizado.⁶

O terceiro estudo observou que o início da sepse está associado a uma queda na pressão arterial média (PAM), assim com como a pressão sistólica e diastólica, além do aumento da frequência cardíaca (FC). Assim, foram desenvolvidos três modelos de aprendizagem por máquina utilizando o modelo de aprendizagem supervisionado, uma vez que os dados a serem avaliados foram pré-determinados pelos programadores.⁷

Os pesquisadores do quarto artigo realizaram um estudo retrospectivo utilizando um banco de dados com aproximadamente 400 mil pacientes. Neste novo modelo desenvolvido,

foram realizadas melhorias na acurácia de detecção da sepse por meio de modelos de aprendizagem computacional comparados com o modelo desenvolvido no *InSight*. Os recursos de referência incluíram 156 variáveis, porém, o novo modelo de predição utilizou 109 variáveis, denominando o novo modelo como DNF.⁸

O quinto artigo mostrou a aplicação do modelo de aprendizagem de máquina nos conjuntos de dados de sepse e choque séptico. As variáveis extraídas para uso foram a PAM, a FC e a frequência respiratória (FR). Para desenvolvimento deste modelo foi utilizada o aprendizado supervisionado de máquina. Este estudo explorou o uso de uma nova técnica de transformação de dados, unindo padrões de contraste sequenciais discretos para representação de uma instância de dados. A análise dos resultados da pressão arterial mostrou-se útil para previsão de eventos relacionados ao choque séptico. O estudo demonstrou, também, a importância do treinamento de modelos classificadores de UTI utilizando padrões sequenciais informativos, além de medidas clínicas convencionais.⁹

No último estudo incluído nesta revisão,¹⁰ os autores relataram que pacientes que apresentam o transtorno por uso de álcool possuem condições de saúde que aumentam a dificuldade de detectar e prever o início do choque séptico. Por esse motivo, os pesquisadores desenvolveram um algoritmo de predição de sepse de alto desempenho (*InSight*), que superou os métodos existentes para esta população específica. O *InSight* analisa uma combinação de dados e medições clínicas ao longo do tempo para gerar um escore de risco.

Os artigos mostraram predomínio da utilização do modelo de aprendizado de máquina supervisionado, pelo fato de ter inserido os padrões de identificação de sepse nos algoritmos que identificam os casos.^{3,5-10} Estudo desenvolvido no Brasil com o objetivo de prever as mortes de pacientes em Unidades de Terapia Intensiva (UTI) encontrou dados similares aos descritos no Quadro 1, e destacou o aprendizado de máquina supervisionado como estratégia para a detecção de novos casos.¹¹

Percebe-se, também, que todos os estudos utilizaram bancos de dados para avaliar a capacidade de predição dos modelos desenvolvidos e, estes algoritmos foram capazes de identificar a sepse com maior precisão, quando comparados às escalas utilizadas pelos profissionais de saúde para o diagnóstico. Além disso, destaca-se um estudo que utilizou o Robô Laura* para o gerenciamento de riscos através da aprendizagem autônoma e agnóstica em tempo real. Esta tecnologia pode ser adaptada ao prontuário eletrônico do próprio serviço de saúde, utilizando as informações contidas nos bancos de dados dos sistemas e equipamentos gravadores de informação.

CONCLUSÃO

Este estudo descreveu, com base na literatura, as técnicas de aprendizagem de máquina na identificação dos casos de sepse. A detecção precoce dos sinais e sintomas poderá promover o tratamento imediato, o que possibilitará a redução das taxas de morbimortalidade, uma vez que a agilidade do diagnóstico está relacionada a maiores chances de sobrevida do paciente.

Os algoritmos para a identificação e predição de sepse vêm sendo cada vez mais utilizados na prática clínica, mostrando uma tendência favorável para o crescimento e disseminação desses modelos, visto que a maioria dos serviços de saúde ainda identifica a sepse de forma observacional, sem o auxílio desta tecnologia. Mesmo com modelos bastante precisos, o aperfeiçoamento para identificação de sepse pode ser realizado de maneira multidisciplinar através de parcerias entre os profissionais de saúde e da informática, sugerindo ser um possível campo de atuação desses profissionais.

Por se tratar de uma revisão integrativa de literatura, destaca-se que os trabalhos realizados para a identificação de sepse com o auxílio de técnicas de aprendizagem de máquina podem não estar indexados nas bases avaliadas, podendo ser uma limitação deste estudo.

REFERÊNCIAS

1. Singer M et al. *The Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3)*. JAMA. 2016; 315(8):801-10. Disponível em: <https://www.ncbi.nlm.nih.gov/pubmed/26903338>
2. Rhodes A et al. *Surviving Sepsis Campaign: International Guidelines for Management of Sepsis and Septic Shock: 2016*. Intensive Care Med 2017;43(3):304-77. Disponível em: <https://www.ncbi.nlm.nih.gov/pubmed/28101605>
3. MC Monard, JA Baranauskas. *Conceitos sobre aprendizado de máquina Sistemas Inteligentes-Fundamentos e Aplicações*. Barueri-SP: Manole Ltda; 2003.
4. Mendes KS, Silveira RCCP, Galvão CM. *Revisão integrativa: método de pesquisa para a incorporação de evidências na saúde e na enfermagem*. Texto context enferm. 2008; 17(4).
5. Lamping F et al. *Development and validation of a diagnostic model for early differentiation of sepsis and non-infectious SIRS in critically ill children - a data-driven approach using machine-learning algorithms*. BMC Pediatr 2018;18(1):112.
6. Mao Q et al. *Multicentre validation of a sepsis prediction algorithm using only vital sign data in the emergency department, general ward and ICU*. BMJ Open 2018;8(1):e017833.
7. Shashikumar SP et al. *Early sepsis detection in critical care patients using multiscale blood pressure and heart rate dynamics*. J Electrocardiol 2017;50(6):739-43.
8. Kam HJ, Kim HY. *Learning representations for the early detection of sepsis with deep neural networks*. Comput Biol Med 2017;89:248-55.
9. Ghosh S et al. *Septic shock prediction for ICU patients via coupled HMM walking on sequential contrast patterns*. J Biomed Inform 2017;66:19-31.
10. Calvert J. et al. *High-performance detection and early prediction of septic shock for alcohol-use disorder patients*. Ann Med Surg (Lond) 2016;8:50-5. Disponível em: <https://www.ncbi.nlm.nih.gov/pubmed/27489621>
11. Schmidt D et al. *Um Modelo de Predição de Mortalidade em Unidades de Terapia Intensiva Baseado em Deep Learning*. Simpósio Brasileiro de Computação Aplicada à Saúde (SBCAS_CSBC) 2018;18(1).
12. Kalil AJ. *Avaliação do impacto na identificação de pacientes com risco de sepse após implantação de um robô cognitivo gerenciador de risco (Robô Laura)*. 2017. 55 f. Dissertação (Mestrado em Engenharia Biomédica) - Universidade Tecnológica Federal do Paraná, Curitiba; 2017.