



## TÉCNICAS DE INTELIGÊNCIA ARTIFICIAL PARA RECONHECIMENTO DE SEPSE EM AMBIENTES HOSPITALARES: REVISÃO INTEGRATIVA

*ARTIFICIAL INTELLIGENCE TECHNIQUES FOR SEPSIS RECOGNITION IN HOSPITAL  
ENVIRONMENTS: INTEGRATIVE REVIEW*



**Everton O. Cesario**

Mestrando pela Pontifícia Universidade Católica do Paraná  
(PUCPR) – Curitiba (PR) – Brasil.  
[evertoncesario@hotmail.com](mailto:evertoncesario@hotmail.com)



**Cristiane Y. Nakamura**

Mestranda pela Pontifícia Universidade Católica do Paraná  
(PUCPR) – Curitiba (PR) – Brasil  
[cristy\\_nakamura@hotmail.com](mailto:cristy_nakamura@hotmail.com)



**Yohan B. Gumiel**

Mestre e doutorando pela Pontifícia Universidade Católica do Paraná  
(PUCPR) – Curitiba (PR) – Brasil.  
[yohan.gumiel@hotmail.com](mailto:yohan.gumiel@hotmail.com)



**Deborah R. Carvalho**

Doutora em Computação pela Universidade Federal do Rio de Janeiro (UFRJ),  
docente dos cursos de Ciência e  
Engenharia da Computação da Pontifícia Universidade Católica do Paraná  
(PUCPR), Curitiba, PR – Brasil.  
[ribeiro.carvalho@pucpr.br](mailto:ribeiro.carvalho@pucpr.br)

**Resumo:** a sepse é uma inflamação generalizada com elevada morbidade e mortalidade, cujo reconhecimento e tratamento precoce são fatores essenciais para uma melhor qualidade de vida para o paciente; caso não seja identificada e tratada rapidamente, poderá levar a óbito. Este artigo de revisão integrativa objetiva identificar as técnicas baseadas em inteligência artificial adotadas, sua respectiva acurácia, sensibilidade e especificidade para a identificação precoce nos casos de sepse em ambiente hospitalar. A pesquisa, adaptada do método PRISMA, foi realizada em cinco bases de dados indexadas a partir dos seguintes descritores: sepse, *septic, sepsis, forecasting, predict, prediction, detection, predicting, diagnosis, assessment, machine learning, artificial intelligence, data mining e deep learning*. Foram identificados 333 artigos, sendo 21 com referência ao reconhecimento precoce da sepse por meio de 16 técnicas. Os resultados demonstram que as redes neurais tiveram melhor desempenho, variando a acurácia entre 76% e 93%, as árvores de decisão entre 69,0% e 91,5% e os métodos estatísticos entre 56% e 89%. Conclui-se que o fator mais influente na identificação precoce do diagnóstico são a variedade e a qualidade dos dados. Também se evidencia o desafio em relação ao pré-processamento, visto que os dados em geral são oriundos de fontes heterogêneas, coletados com critérios, métodos e objetivos distintos.

**Palavras-chave:** Sepse. Inteligência artificial. Modelo preditivo. *Machine learning*.

**Abstract:** Sepsis is a generalized inflammation with high morbidity and mortality, whose recognition and early treatment are essential factors for a better patient's quality of life; if not identified and treated promptly, could lead to death. This integrative review article aims to identify the techniques based on artificial intelligence adopted, their respective accuracy, sensitivity and specificity for early identification in cases of sepsis in a hospital environment. The research, adapted from the PRISMA method, was performed in five databases indexed from the following descriptors: sepsis, septic, forecasting, predict, prediction, detection, predicting, diagnosis, assessment, machine learning, artificial intelligence, data mining and deep learning. A total of 333 articles were identified, 21 with reference to the early recognition of sepsis by 16 techniques. The results showed that the neural networks performed better, varying the accuracy between 76% and 93%, the decision trees between 69.0% and 91.5% and the statistical methods between 56% and 89%. It is concluded that the most influential factor in the early identification of the diagnosis is the variety and quality of the data. The challenge in relation to pre-processing is also evident, since the data are generally from heterogeneous sources, collected with different criteria, methods and objectives.

**Keywords:** Sepsis. Artificial Intelligence. Predictive Model. Machine learning.

### Cite como

American Psychological Association (APA)

Cesario, E. O., Nakamura, C. Y., Gumiel, Y. B., & Carvalho D. R. (2020). Técnicas de inteligência artificial para reconhecimento de sepse em ambientes hospitalares: revisão integrativa. *Rev. gest. sist. saúde.*, São Paulo, 9(1), 15-31. <https://doi.org/10.5585/rgss.v9i1.13932>.





## 1 Introdução

No Brasil, estima-se a taxa de mortalidade em 67,4% para os pacientes diagnosticados com sepse, comparável apenas com a Malásia (66,1%) e bem distante de países como Alemanha (43,4%) e Estados Unidos (42,9%) (Beale et al., 2009). Além de ser uma condição médica fatal, a sepse está sempre associada a altos custos, tanto no processo de identificação quanto no processo de tratamento, tendo sido seu tratamento e diagnóstico considerados a condição clínica que mais gerou custos em 2013 nos Estados Unidos (Calvert et al., 2016). Esses custos têm íntima relação com a gravidade e o tempo de internação (Viana, Machado, & Souza, 2017) – pacientes que apresentam septicemia permanecem em média 75% mais tempo internados se comparados a pacientes hospitalizados por outras condições; a tendência é que essa média de permanência dobre para pacientes com idade próxima aos 65 anos (Hall, Willians, DeFrances, & Golosinskiy, 2011).

A dificuldade na identificação precoce dos pacientes em risco de sepse deve-se ao fato de os sinais e sintomas apresentados não serem exclusivos daqueles que estão em estado séptico; indicativos importantes, que poderiam levar ao reconhecimento precoce, podem passar despercebidos, provocando o reconhecimento tardio. Para auxiliar a identificação, os profissionais da saúde podem fazer uso de sistemas de score, que levam em conta exames físicos e laboratoriais para avaliar e graduar a gravidade dos pacientes com sepse, porém esses critérios ainda falham na identificação de um a cada oito pacientes com sepse severa (Kaukonen, Bailey, Pilcher, Cooper, & Bellomo, 2015); portanto, as intervenções iniciais ainda dependem diretamente da habilidade do profissional de reconhecer o quadro clínico séptico (Viana et al., 2017). Essa dependência é um aspecto que pode causar tanto o não reconhecimento de um quadro de sepse quanto o reconhecimento tardio, ambos sendo fatores que interferem diretamente no tratamento e recuperação do paciente.

Como consequência do reconhecimento tardio, foi demonstrado que cada hora de atraso no tratamento ocasiona uma redução de até 8% na expectativa de sobrevivência do paciente (Kumar et al., 2006), comprovando também que um rápido diagnóstico, aliado ao início precoce da terapia, tem impacto positivo na redução da mortalidade e morbidade em longo prazo e previne a redução da qualidade de vida dos pacientes sobreviventes.

Estudos revelam o potencial de métodos baseados em *Machine Learning* (ML) para apoiar o reconhecimento precoce de casos de sepse. Gultepe, Nguyen, Albertson e Tagkopoulos (2012), Faisal et al. (2018) e Mao et al. (2018) apresentam um comparativo entre técnicas de ML e três escores utilizados para identificação de sepse em ambientes hospitalares: *Modified*





*Early Warning Score* (MEWS), *Systemic Inflammatory Response Syndrome* (SIRS) e *Sequential Organ Failure Assessment* (SOFA). Os resultados apresentados por Mao et al. (2018) indicam que a acurácia, sensibilidade e especificidade dos algoritmos de ML obtiveram melhores resultados, além de demonstrar sua resiliência, pois, mesmo o algoritmo sendo treinado com a ausência de 60% dos dados dos pacientes, obteve melhores resultados que o sistema de escore SIRS. Além do reconhecimento precoce, a ML tem sido utilizada para diversas finalidades, como na criação de modelos para o mapeamento de diversas progressões relacionadas à sepse, como mortalidade (Purushotham, Meng, Che, & Lui, 2018), falha sistêmica de órgãos (Gultepe et al., 2014) e possíveis terapias para o tratamento de pacientes (Ghalwash, Radosavljevic, & Obradovic, 2013).

Dado o potencial apresentado e por não ter sido encontrada nenhuma revisão de literatura sobre esse tema, o objetivo deste artigo é realizar uma revisão integrativa para identificar quais técnicas baseadas em ML estão sendo adotadas e sua respectiva acurácia, sensibilidade e especificidade para identificação de paciente em risco de sepse no ambiente hospitalar.

## 2 Métodos

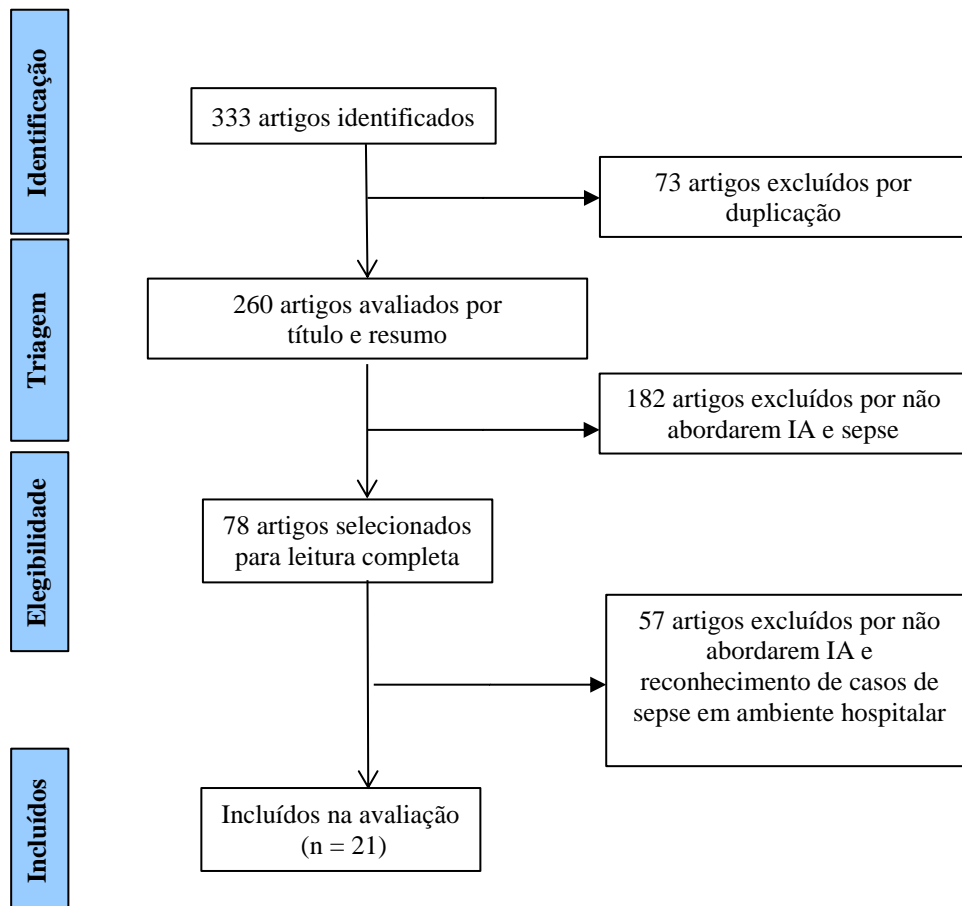
As etapas da revisão foram adaptadas do método *Preferred Reporting Items for Systematic Reviews and Meta-Analyse* (PRISMA) (Altman, Moher, Liberati, & Tetzlaff, 2009). A busca pelos documentos ocorreu em julho de 2018, tendo sido adotados descritores oriundos do *Medical Subject Headings* (MeSH) e Descritores em Ciências da Saúde (DeCS), quais sejam: sepse, detecção, *forecasting*, *data mining* e *artificial intelligence*. O termo de pesquisa foi construído com o objetivo de atender aos critérios das bases de publicações que utilizam descritores em saúde (PubMed, Biblioteca Virtual em Saúde – BVS, Literatura Latino-Americana e do Caribe em Ciências da Saúde – Lilacs), como também às demais bases que não os empregam (Institute of Electrical and Electronics Engineers – IEEE, Scopus, Web of Science). Sendo assim, foi necessário desenvolver um termo de pesquisa único para todas as bases pesquisadas, sendo o final: *(sepsis or septic or sepse) And (forecasting or predict or prediction or detection or predicting or diagnosis or assessment) And (“machine learning” or “artificial intelligence” or “data mining” or “deep learning”)*.

A Figura 1 apresenta os processos da revisão e o número de artigos identificados em cada etapa da pesquisa. Na etapa 1 (identificação), a procura foi realizada utilizando o termo de pesquisa desenvolvido, excluindo os editoriais, prefácios, resumos, entrevistas e cartas ao



editor. Na etapa 2 (triagem), os pesquisadores avaliaram a relevância dos artigos com o tema, por meio do título e *abstract*, levando em conta os critérios de elegibilidade, que foram: o artigo deve tratar de identificação de sepse em ambientes hospitalares e apresentar um ou mais métodos de ML descrevendo ou comparando os resultados. Entretanto, alguns artigos apresentaram falta ou dificuldade na apresentação do tema no *abstract*, dificultando a identificação do objetivo do artigo; nesses casos, o artigo foi selecionado para leitura integral. Na etapa 3 (elegibilidade), os 78 artigos foram lidos na íntegra pelos três pesquisadores e excluídos se o tema ou foco principal não abordasse os critérios de elegibilidade.

**Figura 1** – Processo de seleção dos artigos



**Fonte:** Adaptado de Altman et al. (2009).

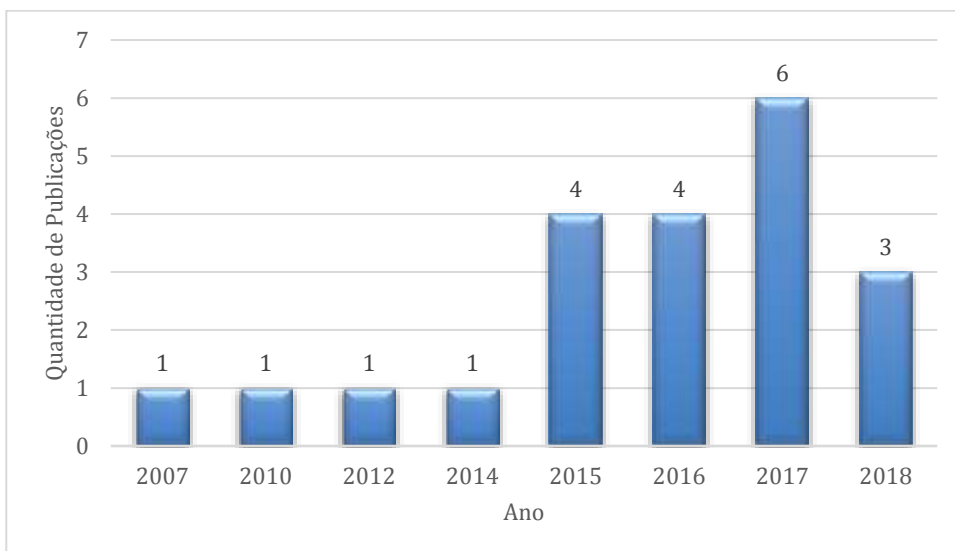
O coeficiente Kappa (Kosh & Landis, 1977) foi aplicado para medir o grau de concordância entre as seleções realizadas pelos três pesquisadores e obteve uma média de concordância de 0,79%, o que, segundo a tabela de concordância de Landis (Kosh & Landis, 1977), indica uma concordância “substancial”, confirmando que a seleção foi realizada mediante critérios. Todas as discordâncias foram resolvidas por meio de discussão entre os três





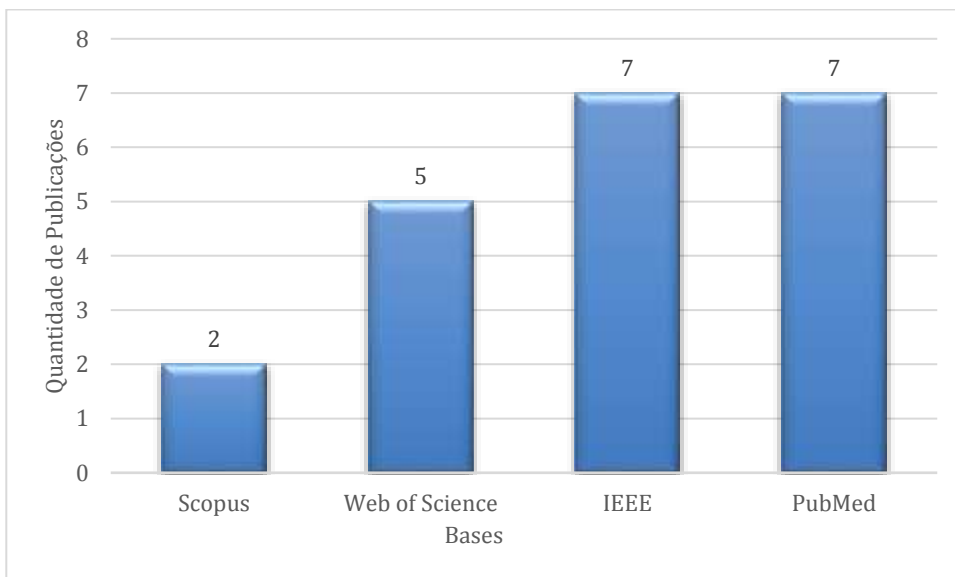
pesquisadores antes de passar para a próxima etapa. Os 21 artigos finais estão distribuídos da seguinte forma:

**Figura 2** – Quantidade de publicações selecionadas por ano



Fonte: Os autores (2019).

**Figura 3** – Quantidade de publicações selecionadas por base



Fonte: Os autores (2019).

O objetivo deste artigo é identificar as técnicas de ML e seus respectivos resultados de acurácia, sensibilidade e especificidade. Para essa finalidade, foram extraídas algumas características dos artigos, como modelo preditivo utilizado, critério de avaliação da predição,



variáveis disponíveis e utilizadas, fonte e volume de dados, valores de acurácia, sensibilidade, especificidade e desbalanceamento das classes.

Os resultados serão apresentados por meio de dois critérios: (i) em função dos valores de acurácia, sensibilidade e especificidade (Ashal, Baratloo, Hosseini, & Negida, 2015; Parikh, Mathai, Parikh, Sekhar, & Thomas, 2008); (ii) o tempo de reconhecimento da sepse em relação à frequência da coleta dos dados. A acurácia é a proporção de resultados classificados corretamente (Simundic, 2009), enquanto a sensibilidade é a capacidade de classificar corretamente um indivíduo como Verdadeiro Positivo (VP) (Ashal et al., 2015), definida por:

$$\text{Sensibilidade} = \frac{VP}{VP + FN}$$

Em que VP corresponde à quantidade de pacientes classificados corretamente em estado séptico e FN, ao número de casos Falsos Negativos, situação em que o paciente foi diagnosticado como não estando em estado séptico, porém estava.

Já a especificidade é a capacidade de classificar corretamente um indivíduo como fora do estado séptico (Ashal et al., 2015); em outros termos, os Verdadeiros Negativos (VN) são definidos por:

$$\text{Especificidade} = \frac{VN}{VN + FP}$$

Em que VN são os verdadeiros negativos, pacientes que foram corretamente classificados como não estando em estado séptico, e FP corresponde aos Falsos Positivos, ou seja, aqueles que foram erroneamente classificados como pertencendo ao estado séptico, porém não estavam.

Para ter uma ideia melhor do desempenho dos modelos, extraiu-se também, para esta análise, a quantidade de registros, variáveis utilizadas, fonte dos dados e desbalanceamento do conjunto de dados.

### 3 Resultados

A Tabela 1 apresenta os resultados por ordem decrescente de sensibilidade, pois o objetivo dos algoritmos é identificar os pacientes em risco de sepse, idealmente em fase precoce. A coluna “Fonte de Dados” foi preenchida somente com os estudos que utilizaram a fonte *Medical Information Mart for Intensive Care* (MIMIC), pois foi o conjunto de dados que mais se repetiu. A coluna “Desbalanceamento” representa o percentual de dados referente aos



casos de sepse. Em alguns estudos, o algoritmo foi treinado levando em consideração apenas os casos de sepse; nesses estudos foi utilizado o sinal “-” para identificar que não houve desbalanceamento. Alguns estudos, como Taylor et al. (2016) e Horng et al. (2017), apresentaram os resultados utilizando diferentes modelos; sendo assim, a tabela de resultados contém 23 itens, em vez de 21, quantidade final de artigos selecionados no processo de revisão.

**Tabela 1** – Desempenho dos algoritmos ordenados pelo valor de sensibilidade

Referência	Modelo utilizado	Sensibilidade	Especificidade	Acurácia	Variáveis utilizadas	Fonte de dados	Número de registros	Desbalanceamento
Mao et al. (2018)	InSight	98%	67%	76%	20	MIMIC - III	90.353	-
Lamping et al. (2018)	Random Forest	94,4%	94,4%	91,5%	44		807	39,86%
Kam e Kim (2017)	LSTM	91,4%	94,4%	93%	9	MIMIC - II	5.789	6,21%
Calvert et al. (2016)	InSight	90%	81%	80%	10	MIMIC - II	1.394	-
Wang et al. (2018)	Extreme Machine Learning	89,57%	65,77%	81,6%	160		42	54,54%
Hornig et al. (2017)	Random Forest	87%	83%	70%	12+PLN		230.936	14%
Hornig et al. (2017)	Regressão Logística	86%	84%	67%	12+PLN		230.936	14%
Nemati et al. (2018)	AI Sepsis Expert	85%	72%	72%	65	MIMIC - III	27.527	8,6%
Henry, Hager, Provonost e Saria (2015)	TREWScore	85%	67%	83%	54	MIMIC - II	16.234	14,1%
Ghosh, Li, Cao e Ramamohanarao (2017)	MCP+CHMM	85%		87%	3	MIMIC - II	1.519	4,36%
Rothman et al. (2017)	Logistic Regression	75%	88%	21%	36		258.836	4,79%
Schuh (2007)	Multlayer Perceptron	70%	73%	72%	11		1.674	-
Parente, Lee, Lin, Chase e Shaw (2010)	Kernel Classifiers	69%	75%	78%	8		9.286	-
Zhang, Lin, Chi, Ivy e Huddleston (2017)	LSTM	61,9%	97,5%	72,3%	31		1.446.225	-
Guillén et al. (2015)	Regressão Logística	60,3%	93,2%	87,7%	15	MIMIC - II	3.446	17,87







Shashikumar, Li, Clifford e Nemati (2017)	MSNR+EMR+MSE	57%	85%	80%	15		250	22%
Mitchell et al. (2016)	Regressão Logística	57%	77,3%	74%	7	MIMIC - II	235.000	3,64%
Taylor et al. (2016)	Random Forest			86%	+500		5.278	-
Faisal et al. (2018)	Regressão Logística			80%	20		73.651	18,5%
Taylor et al. (2016)	Regressão Logística			76%	+500		5.278	-
Gultepe et al. (2012)	Bayesian Networks				7	Mapa	1.492	15,61
Jiang et al. (2016)	AdapDBN						Dados simulados	

Fonte: Os autores (2019).





Os 21 estudos foram classificados dentro de categorias de aprendizado, a saber: métodos estatísticos, redes neurais e árvores de decisão. A Tabela 2 indica a quantidade de estudos em cada categoria, percentual de estudos e maior e menor valor de acurácia. Alguns utilizaram mais de um método para efeito de comparação, portanto a quantidade total de métodos utilizados é maior que a de estudos.

**Tabela 2** – Categoria de aprendizado

Categoria	Quantidade	Percentual	Acurácia
Redes neurais	6	25,00%	76,0% a 93,0%
Árvores de decisão	8	33,33%	69,0% a 91,5%
Métodos estatísticos	10	41,66%	56,0% a 89,0%

Fonte: Os autores (2019).

Em termos gerais, as redes neurais atingiram melhores valores de acurácia, talvez devido às possibilidades de customização do modelo, tornando-o mais aderente ao dado e, conseqüentemente, alcançando melhores resultados. Métodos estatísticos e árvores de decisão não possuem tantas possibilidades de configuração, diminuindo a possibilidade de aderência, porém já é possível atingir níveis satisfatórios de acurácia com modelos mais simples e que não necessitem de uma quantidade tão grande de registros e atributos. Um detalhe importante com relação aos resultados é que os estudos que apresentaram melhores resultados de desempenho possuem menores quantidades de registros.

O tempo para reconhecimento da sepse é de fundamental importância, pois o reconhecimento precoce pode dar ao paciente importantes resultados no seu tratamento e recuperação (Wyk, Khojandi, & Kamaleswaran, 2017). A Tabela 3 apresenta os resultados publicados por Nemati et al. (2018), em que foi implementado um algoritmo de inteligência artificial chamado *Artificial Intelligence Sepsis Expert*, com o propósito de prever com precisão o início da sepse em um paciente na unidade de terapia intensiva de quatro a 12 horas antes do reconhecimento clínico.

**Tabela 3** – Acurácia ao longo do tempo em horas

Critério utilizado	Técnica utilizada	4h	6h	8h	12h
Sepsis	<i>Artificial Intelligence Sepsis Expert</i>	85%	85%	84%	83%
SOFA	<i>Artificial Intelligence Sepsis Expert</i>	87%	85%	85%	82%

Fonte: Nemati et al. (2018).



Wyk et al. (2017) realizaram um estudo cujo objetivo foi comparar a acurácia de duas implementações de *deep learning* diferentes – redes neurais convolucionais (*Convolutional Neural Network* – CNN) e modelo perceptron multicamadas (*Multilayer Perceptron Model* – MLP), com o objetivo de classificar pacientes em grupos, com e sem sepse, utilizando dados coletados nas 12 primeiras horas após a admissão no hospital, com distintos intervalos de tempo entre as coletas de dados. Os resultados mostraram que a CNN superou o MLP na classificação dos pacientes. A Tabela 4 apresenta a acurácia da CNN obtida por meio dos dados coletados com intervalos (em minutos), ou seja, o classificador teve um resultado melhor para as coletas de dados mais frequentes, com o intervalo de até 10 min.

**Tabela 4** – Acurácia ao longo do tempo em minutos

Técnica utilizada	0-10 min	11-20 min	21-30 min	31-60 min	>60 min
CNN	86%	78,2%	75%	74%	72%

Autor: Wyk et al. (2017).

Wyk et al. (2017) observaram que a acurácia da detecção da sepse diminui na ordem de quase 10% quando os intervalos entre as coletas de dados variam de um a dez minutos e mantém uma queda gradual e menos acentuada nos intervalos de 15 a 60 minutos.

Ainda utilizando redes neurais, Mao et al. (2018) compararam um algoritmo de *machine-learning* proprietário chamado InSight, com três sistemas de escore “padrão ouro” (Seymour et al., 2016) empregados para identificação de sepse em ambientes hospitalares, sendo: MEWS, SIRS e SOFA. A Tabela 5 apresenta os valores para avaliação de *performance* da detecção de casos de sepse cujos sintomas iniciaram dentro da hora zero, sob os valores de *Area Under the Receiver Operating Characteristic* (Auroc), sensibilidade e especificidade, ficando evidenciada a melhor *performance* do algoritmo InSight frente às técnicas tradicionais utilizadas. O diferencial desse estudo está em ter sido utilizado apenas seis dados de sinais vitais para realização das análises, a saber: pressão sistólica, pressão diastólica, frequência cardíaca, frequência respiratória, saturação de oxigênio e temperatura, o que pode resultar em uma identificação mais rápida dos casos, pois independe de exames laboratoriais.

**Tabela 5** – Comparativo de desempenho do algoritmo de *machine-learning* InSight, com escores MEWS, SOFA e SIRS, na identificação de pacientes com sepse dentro da hora zero

	Gravidade	InSight	MEWS	SOFA	SIRS
<b>Auroc</b>	Sepse	0,92	0,76	0,63	0,75
	Sepse severa	0,87	0,77	0,65	0,72
	Choque séptico	0,99	0,94	0,86	0,82
<b>Sensibilidade</b>	Sepse	0,98	0,98	0,82	0,82
	Sepse severa	0,996	0,98	0,90	0,81
	Choque séptico	1,0	1,0	0,99	0,91
<b>Especificidade</b>	Sepse	0,95	0,72	0,32	0,51
	Sepse severa	0,85	0,72	0,37	0,50
	Choque séptico	0,99	0,91	0,58	0,49

Fonte: Mao et al. (2018).

Mao et al. (2018) ainda demonstraram a resiliência do algoritmo – mesmo com 60% de dados faltantes, o algoritmo alcançou melhores resultados que o sistema de escore SIRS, conforme Tabela 6.

**Tabela 6** – Desempenho do algoritmo de *machine-learning* InSight com falta de dados em comparação com o escore SIRS com o conjunto de dados completo

Percentual de dados faltantes	InSight					SIRS
	0%	10%	20%	40%	60%	0
<b>Auroc</b>	0,90	0,82	0,79	0,76	0,75	0,72
<b>Sensibilidade</b>	0,80	0,80	0,80	0,80	0,80	0,80
<b>Especificidade</b>	0,84	0,66	0,57	0,50	0,49	0,51

Fonte: Mao et al. (2018).

## 4 Discussão

De modo geral, não foi possível fazer um comparativo entre eficiência dos métodos utilizados, pois, apesar de o objetivo ser o reconhecimento precoce de sepse, as fontes de dados, conjuntos de atributos e métodos de pré-processamento variaram muito entre si. Na Tabela 1, identificam-se essas diferenças; por exemplo, Lamping et al. (2018) utilizaram o método Random Forest, com 807 registros e 44 variáveis, e obtiveram sensibilidade de 94,4%, enquanto Horneq et al. (2017), que também aplicaram Random Forest, com 230 mil registros e 12 variáveis, encontraram sensibilidade de 87%. Os dados utilizados foram de fontes distintas e com características diferentes, como, por exemplo, o desbalanceamento da primeira, que foi de 39,86%, e da segunda, 14,00%.



Na avaliação individual dos trabalhos, ficaram evidentes algumas restrições com relação aos dados. O grande desafio foi tornar o conjunto de dados adequado ao escopo sugerido. Para tanto, Kam e Kim (2017) reduziram o conjunto de dados de aproximadamente 32 mil casos para aproximadamente cinco mil casos de controle, tendo apenas 360 as características necessárias para o escopo definido. Na etapa de pré-processamento, cada estudo definiu seus critérios e objetivos de acordo com a técnica de inteligência artificial utilizada. Horng et al. (2017) precisaram pré-processar o texto aplicando processos de tokenização, que consistem em uma operação para condução da análise das frases; esse processo foi necessário para poder utilizar técnicas de Processamento de Linguagem Natural (PLN) e mostrou que o desempenho do algoritmo de ML, utilizando PLN, superou em mais de 10% o uso de apenas sinais vitais dos pacientes. Alguns estudos empregaram mais de uma fonte de dados, de modo que a fonte dos dados de treinamento por vezes foi diferente da fonte de dados de validação (Mitchell et al., 2016). No estudo de Rothman et al. (2017), foram utilizados dados de quatro hospitais, separados em um grupo de pacientes que desenvolveram sepse no hospital e outro de pacientes que já apresentavam sepse na admissão. Esse tipo de situação ocorreu devido à quantidade e qualidade dos dados, conjunto de atributos disponíveis, mas, principalmente, baixa frequência das coletas, o que impacta diretamente na *performance* de alguns métodos.

A fonte de dados mais utilizada foi a MIMIC, presente em oito estudos (38%), a qual está atualmente na sua terceira versão e é uma fonte de acesso público e livre, composta por dados de pacientes relacionados à saúde e informações demográficas, com mais de 40 mil pacientes que permaneceram em unidades de cuidados intensivos do Centro Médico Beth Israel Deaconess, entre 2001 e 2012. A versão MIMIC – II compreende os dados dos pacientes de 2001 a 2008 e a versão MIMIC – III, dos pacientes de 2008 a 2012.

No entanto, estudos que utilizaram a mesma fonte de dados não apresentaram resultados semelhantes – por exemplo, a sensibilidade variou de 98% (Mao et al., 2018) a 57% (Mitchell et al., 2016), a acurácia dos modelos, de 93% (Kam & Kim, 2017) a 72% (Nemati et al., 2018) –, mesmo porque, nesses estudos, a utilização dos dados teve diferentes critérios de seleção devido ao algoritmo de ML escolhido. Essa diferença de critério de seleção fica clara ao observar o desbalanceamento dos conjuntos de dados finais utilizados – variando de 17,87% a 6,21%. Por outro lado, não é possível deixar de levar em consideração que, entre os dez primeiros lugares da tabela de resultados (Tabela 1), seis usaram a base MIMIC.

Uma limitação presente em alguns estudos foi o uso da *International Classification of Disease Ninth Revision (ICD9)*, que é similar à Classificação Internacional de Doenças (CID) brasileira, como seleção dos pacientes com sepse. Nesse sentido, Wyk et al. (2017) apontaram



em seu estudo que essa codificação se mostrou errada várias vezes e, portanto, pode ter comprometido os resultados dos algoritmos, tanto na fase de treinamento quanto na fase de testes. Faisal et al. (2018) também indicaram a utilização do código ICD-10 como limitação, ao ponto de não ficar claro se seu uso, no lugar do consenso para identificação de sepse, afetou a avaliação do algoritmo ou não.

Foi possível perceber que, utilizando modelos mais simples e dados mais fáceis de serem obtidos, se podem obter alguns resultados expressivos. Os modelos mais complexos necessitam de um conjunto maior de dados e, normalmente, apenas são passíveis de serem desenvolvidos em hospitais que possuem uma infraestrutura maior e mais complexa.

Wyk et al. (2017) abordam em seu estudo a relação complexa entre o resultado da acurácia do método de classificação e o custo econômico dos sistemas de aquisição de dados, indicando que a alta frequência da coleta de dados revela melhor *performance* na classificação dos pacientes. Já Taylor et al. (2016) apontam como limitação a necessidade de uma infraestrutura crescente para permitir análises preditivas em tempo real.

## 5 Considerações finais

O objetivo do artigo foi fazer uma revisão integrativa, no intuito de identificar as técnicas mais eficientes de inteligência artificial utilizadas para reconhecimento de sepse em ambiente hospitalar. Por meio das etapas da pesquisa, foram encontrados estudos envolvendo ML e sepse em vários contextos, como, por exemplo, avaliação de mortalidade, avaliação da gravidade dos casos, simulação de tratamentos e, principalmente, avaliação dos escores de sepse utilizados. Sendo assim, trata-se de um trabalho de aplicabilidade social e gerencial, pois os modelos de inteligência artificial apresentados tiveram o objetivo de realizar a identificação precoce dos casos, provendo melhores resultados na recuperação do paciente e reduzindo o custo do tratamento da sepse nos ambientes hospitalares.

## Referências

- Altman, D. G., Moher, D., Liberati, A., & Tetzlaff, J. (2009 Jul.). Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med*, 6(7), e100097.
- Ashal, G. E., Baratloo, A., Hosseini, M., & Negida, A. (2015). Simple definition and calculation of accuracy, sensitivity and specificity. *Emerg (Tehran)*, 3(2), 48-49.
- Beale, R., Reinhart, K., Brunkhorst, F. M., Dobb, G., Levy, M., Martin, G., Ramsey, G., Silva, E., Vallet, B., Vincent, J. L., Janes, J. M., Sarwar, S., Williams, M. D., &



- PROGRESS Advisory Board (2009 Jun.). Promoting Global Research Excellence in Severe Sepsis (PROGRESS): lessons from an international sepsis registry. *Infection*, 37(3), 222-232.
- Calvert, J. S., Price, D. A., Chettipaly, U. K., Barton, C. W., Feldman, M. D., Hoffman, J. L., & Jay, M. (2016 Jul). A computational approach to early sepsis detection. *Comput Biol Med*, 74, 69-73.
- Faisal, M., Scally, A., Richardson, D., Beatson, K., Howes, R., Speed, K., & Mohammed, A. M. (2018 Apr.). Development and external validation of an automated computer-aided risk score for predicting sepsis in emergency medical admissions using the patient's first electronically recorded vital signs and blood test results. *Crit Care Med*, 46(4), 612-618.
- Ghalwash, M., Radosavljevic, V., & Obradovic, Z. (2013). Early diagnosis and its benefits in sepsis blood purification treatment. *Proceedings of IEEE International Conference on Healthcare Informatics*, Philadelphia, PA, USA.
- Ghosh, S., Li, J., Cao, L., & Ramamohanarao, K. (2017 Feb.). Septic shock prediction for ICU patients via coupled HMM walking on sequential contrast patterns. *J Biomed Inform*, 66, 19-31.
- Guillén, J., Liu, J., Furr, M., Wang, T., Strong, S., Moore, C. C., Flower, A., & Barnes, L. E. (2015). Predictive models for severe sepsis in adult ICU patients. *Proceedings of IEEE Systems and Information Engineering Design Symposium*, Charlottesville, VA, USA.
- Gultepe, E., Green, J. P., Nguyen, H., Adams, J., Albertson, T., & Tagkopoulos, I. (2014 Mar./Apr.). From vital signs to clinical outcomes for patients with sepsis: a machine learning -basis for a clinical decision support system. *J Am Med Inform Assoc*, 21(2), 315-325.
- Gultepe, E., Nguyen, H., Albertson, T., & Tagkopoulos, I. (2012). A Bayesian network for early diagnosis of sepsis patients: a basis for a clinical decision support system. *Proceedings of IEEE International Conference on Computational Advances in Bio and medical Sciences*, Las Vegas, NV, USA, 2.
- Hall, M. J., Willians, S. N., DeFrances, C. J., & Golosinskiy, A. (2011 Jun.). Inpatient care for septicemia or sepsis: a challenge for patients and hospitals. *NCHS Data Brief*, 62, 1-8.
- Henry, K. E., Hager, D. N., Pronovost, P. J., & Saria, S. (2015 Aug.). A targeted real-time early warning score (TREWScore) for septic shock. *Sci Transl Med*, 7(299), 299ra122.
- Hornig, S., Sontag, D. A., Halpern, Y., Jernite, Y., Shapiro, N., & Nathanson, L. (2017 Apr.). Creating an automated trigger for sepsis clinical decision support at emergency department triage using machine learning. *PLoS One*, 12(4), e0174708.
- Jiang, Y., Tan, P., Song, H., Wan, B., Hosseini, M., & Sha, L. (2016). A self-adaptively evolutionary screening approach for sepsis patient. *Proceedings of IEEE International Symposium on Computer-Based Medical Systems*, Dublin, 29.
- Kam, H. J., & Kim, H. Y. (2017 Oct.). Learning representations for the early detection of sepsis with deep neural networks. *Comput Biol Med*, 89, 248-255.



- Kaukonen, K. M., Bailey, M., Pilcher, D., Cooper, D., & Bellomo, R. (2015 Aug.). Systemic inflammatory response syndrome criteria in defining severe sepsis. *N Engl J Med*, 373(9), 881.
- Kosh, G., & Landis, R. J. (1977 Mar.). The measurement of observer agreement for categorical data. *Biometrics*, 33(1), 159-174.
- Kumar, A., Roberts, D., Wood, K. E., Light, B., Parrillo, J. E., Sharma, S., Suppes, R., Feinstein, D., Zanotti, S., Taiberg, L., Gurka, D., Kumar, A., & Cheang, M. (2006 Jun.). Duration of hypotension before initiation of effective antimicrobial therapy is the critical determinant of survival in human septic shock. *Crit Care Med*, 34(6), 1589-1596.
- Lamping, F., Jack, T., Rubsam, N., Sasse, M., Beerbaum, P., Mikolajczyk, R. T., Boehne, M., & Karch, A. (2018 Mar.). 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. *BMC Pediatr*, 18(1), 112.
- Mao, Q., Jay, M., Hoffman, J. L., Calvert, J., Barton, C., Shimabukuro, D., Shieh, L., Chettipally, U., Fletcher, G., Kerem, Y., Zhou, Y., Das, R. (2018 Jan.). Multicentre validation of a sepsis prediction algorithm using only vital sign data in the emergency department, general ward and ICU. *BMJ Open*, 8(1), e017833.
- Mitchell, S., Schinkel, K., Song, Y., Wang, Y., Ainsworth, J., Halbert, T., Strong, S., Zhang, J., Moore, C. C., & Barnes, L. E. (2016). Optimization of sepsis risk assessment for ward patients. *Proceedings of IEEE Systems and Information Engineering Design Conference*, Charlottesville, VA, USA.
- Nemati, S., Holder, A., Razmi, F., Stanley, M., Clifford, G. D., & Buchman, T. G. (2018 Apr.). An interpretable machine learning model for accurate prediction of sepsis in the ICU. *Crit Care Med*, 46(4), 547-553.
- Parente, J. D., Lee, D., Lin, J., Chase, G. J., & Shaw, G. M. (2010). A fast and accurate diagnostic test for severe sepsis using Kernel classifiers. *Proceedings of UKACC International Conference on Control*, Coventry.
- Parikh, R., Mathai, A., Parikh, S., Sekhar, C., & Thomas, R. (2008 Jan./Feb.). Understanding and using sensitivity, specificity and predictive values. *Indian J Ophthalmol*, 56(1), 45-50.
- Purushotham, S., Meng, C., Che, Z., & Lui, Y. (2018 Jul.). Benchmarking deep learning models on large healthcare datasets. *J Biomed Inform*, 83, 112-134.
- Rothman, M., Levy, M., Dellinger, P. R., Jones, S. L., Robert, F. L., Voelker, K. G., Gross, B., Marchetti, A., & Beals, J. (2017 Apr.). Sepsis as 2 problems: identifying sepsis at admission and predicting onset in the hospital using an electronic medical record-based acuity score. *J Crit Care*, 38, 237-244.
- Schuh, C. J. (2007). Sepsis and septic shock analysis using neural networks. *Proceedings of Annual Meeting of the North American Fuzzy Information Processing Society*.
- Seymour, W. C., Liu, X. V., Iwashyna, J. T., Brunkhorst, M. F., Rea, D. T., Scherag, A., Rubenfeld, G., Kahn, J. M., Shankar-Hari, M., Singer, M., Deutschman, C. S., Escobar, G. J., & Angus, C. D. (2016 Feb.). Assessment of clinical criteria for sepsis







- for the third international consensus definitions for sepsis and septic shock (Sepsis-3). *JAMA*, 315(8), 762-774.
- Shashikumar, P. S., Li, Q., Clifford, G. D., & Nemati, S. (2017 Nov.). Multiscale network representation of physiological time series for early prediction of sepsis. *Physiol Meas*, 38(12), 2235-2248.
- Simundic, A. (2009 Jan.). Measures of diagnostic accuracy: basic definitions. *EJIFCC*, 19(4), 203-211.
- Taylor, A. R., Pare, J. R., Venkatesh, A. K., Mowafi, H., Melnick, E. R., Fleischman, W., & Hall, K. M. (2016 Mar.). Prediction of in-hospital mortality in emergency department patients with sepsis: a local big data-driven, machine learning approach. *Acad Emerg Med*, 23(3), 269-278.
- Viana, R. A., Machado, F. R., & Souza, J. L. (2017). *Sepse: um problema de saúde pública: a atuação e colaboração da Enfermagem na rápida identificação e tratamento da doença*. São Paulo: Coren-SP.
- Wang, X., Wang, Z., Weng, J., Wen, C., Chen, H., & Wang, X. (2018). *A new effective machine learning framework for sepsis diagnosis*. New Jersey: IEEE.
- Wyk, F. V., Khojandi, A., & Kamaleswaran, R. (2017 Nov.). How much data should we collect? A case study in sepsis detection using deep learning. *Proceedings of IEEE Healthcare Innovations and Point of Care Technologies*, Bethesda, MD, USA.
- Zhang, Y., Lin, C., Chi, M., Ivy, C., & Huddleston, M. (2017). LSTM for septic shock: adding unreliable labels to reliable predictions. *Proceedings of IEEE International Conference on Big Data*, Boston, MA, USA.