

**PONTIFÍCIA UNIVERSIDADE CATÓLICA DO PARANÁ
ESCOLA POLITÉCNICA
PROGRAMA DE PÓS-GRADUAÇÃO EM TECNOLOGIA EM SAÚDE**

EVERTON OSNEI CESARIO

**IDENTIFICAÇÃO PRECOCE DE PACIENTES COM RISCO DE SEPSE EM
AMBIENTE HOSPITALAR**

CURITIBA

2020

EVERTON OSNEI CESARIO

**IDENTIFICAÇÃO PRECOCE DE PACIENTES COM RISCO DE SEPSE EM
AMBIENTE HOSPITALAR**

Dissertação de mestrado apresentada ao Programa de Pós-Graduação em Tecnologia em Saúde da Pontifícia Universidade Católica do Paraná, como requisito parcial à obtenção do título de Mestre em Tecnologia em Saúde.

Área de concentração: Informática em Saúde.

Orientadora: Profa. Dra. Deborah Ribeiro Carvalho

CURITIBA

2020

Dados da Catalogação na Publicação
 Pontifícia Universidade Católica do Paraná
 Sistema Integrado de Bibliotecas – SIBI/PUCPR
 Biblioteca Central
 Edilene de Oliveira dos Santos CRB-9/1636

	Cesario, Everton Osnei
C421i	Identificação precoce de pacientes com risco de sepse em ambiente hospitalar /
2020	Everton Osnei Cesario ; orientadora, Deborah Ribeiro Carvalho. -- 2020
	117 f. : il. ; 30 cm
	Dissertação (mestrado) – Pontifícia Universidade Católica do Paraná,
	Curitiba, 2020
	Bibliografia: f. 80-87
	1. Inteligência artificial. 2. Tecnologia médica. 3. Informática na medicina. 4.
	Septicemia. 5. Valor preditivo dos testes. I. Carvalho, Deborah Ribeiro.
	II. Pontifícia Universidade Católica do Paraná. Programa de Pós-Graduação em
	CDD. 20.ed. – 006.3



Pontifícia Universidade Católica do Paraná

Escola Politécnica

Programa de Pós-Graduação em Tecnologia em Saúde

TERMO DE APROVAÇÃO DE DISSERTAÇÃO Nº 273

A Dissertação de Mestrado intitulada **IDENTIFICAÇÃO PRECOCE DE PACIENTES COM RISCO DE SEPSE EM AMBIENTE HOSPITALAR**, defendida em sessão pública pelo candidato **Everton Osnei Cesário**, no dia **17 de fevereiro de 2020**, foi julgada para a obtenção do grau de Mestre em Tecnologia em Saúde, e aprovada em sua forma final pelo Programa de Pós-Graduação em Tecnologia em Saúde.

BANCA EXAMINADORA:

Prof. Dr. Deborah Ribeiro Carvalho - Presidente – (PUCPR)

Prof. Dr. Sergiu Ossamu Ioshii – (PUCPR)

Prof. Dr. Priscyla Waleska Targino de Azevedo Simões – (UFACB)

A via original deste documento encontra-se arquivada na Secretaria do Programa, contendo a assinatura da Coordenação após a entrega da versão corrigida do trabalho.

Curitiba, 04 de maio de 2020.

Prof. Dr. Percy Nohama
Coordenador do Programa de
Pós-Graduação em Tecnologia em Saúde
PUCPR

AGRADECIMENTOS

Agradeço a Deus por propiciar esse momento e me abençoar com pessoas incríveis que me apoiaram, incentivaram e conduziram durante todo esse processo.

Agradeço minha família, na figura da minha mãe, Solange, pelo apoio incondicional, pelas infindáveis garrafas de café, por todas as conversas, ou apenas pela presença nas madrugadas, cuidando para que eu não dormisse fazendo um trabalho ainda na graduação. Te amo mãe, muito obrigado. Minha sobrinha Michelli, pelo seu apoio, sua preocupação, seu abraço forte e as incontáveis horas de conversa e aos meus irmãos que propiciaram todo suporte necessário nessa fase, muito obrigado.

À minha orientadora, Professora Dra. Deborah Ribeiro Carvalho, como exemplo de professora, com muito carinho, paciência e dedicação me concedeu toda atenção, compreensão e ensinamento durante todo o processo. Suas palavras sempre gentis, aliados ao olhar rigoroso e experiente me encaminharam em todos os momentos. Muito obrigado.

À Dra. Viviane Carvalho, por todas as conversas e orientações, por todo seu exíguo tempo livre, e parte deste, dedicado a este estudo. Seu conhecimento, preocupação e carinho no cuidado à vida inspiram todos a sua volta. Muito obrigado.

Aos professores que participaram da minha banca de qualificação, Professor Dr. Sérgio Ioshii e Professora Dra. Priscyla Simões, pelo tempo dedicado e pelas sugestões que enriqueceram e transformaram esse estudo. Muito obrigado.

Ao meu grupo de pesquisa, Cristiane, João, Yohan, Elziane, Lorena, Ana Sartorelli, Cris Meurer, Marcelo Dallagassa, Franciele, Georgia, Ana Luiza, Marcia e Adans, por todas as conversas, risadas e contribuições que forneceram diversos olhares que corrigiram e enriqueceram o trabalho. Esse grupo me ensinou o real significado do termo “apoio”, pois a preocupação e disponibilidade de todos em ajudar me deram segurança e confiança para encarar esse desafio e superar cada etapa. Deixo aqui meu agradecimento e apoio para sempre que necessário, muito obrigado.

À Celepar, por incentivar e custear parte do investimento. Meus colegas do setor, a equipe do GSUS e a Secretaria de Saúde do Estado, muito obrigado pelo apoio.

Ao meu amigo Danilo Souto, por todas as discussões que tivemos sobre este estudo, pela sua ajuda nas revisões e correções, pela participação ativa em todas as etapas e por segurar as pontas nas minhas ausências, muito obrigado.

À Professora Andreia Malucelli, que sempre com um sorriso no rosto e um olhar atento e carinhoso, já na graduação, nos apresentou a paixão e dedicação pela ciência, muito obrigado.

Para todos envolvidos que me ajudaram direta ou indiretamente, meu mais sincero agradecimento. Muito obrigado.

RESUMO

Introdução: Sepsis é uma inflamação generalizada com elevada morbidade e mortalidade, sendo a principal causa de morte nas unidades de terapia intensiva. Seu reconhecimento e tratamento precoce são fatores essenciais para uma melhor qualidade de vida para o paciente. A cada hora de atraso no início do tratamento, reduz-se a taxa de sobrevivência em até 8%. **Objetivo:** Desenvolver uma ferramenta que identifique precocemente os casos de sepsis, por meio de *deep learning*, para os hospitais gerenciados pelo Sistema de Gestão Hospitalar e Ambulatorial do Sistema Único de Saúde do governo do Estado do Paraná. **Método:** Trata-se de uma pesquisa descritiva, retrospectiva e de abordagem quantitativa, composta por quatro etapas. A primeira envolveu uma revisão de literatura para identificar as técnicas baseadas em inteligência artificial utilizadas na identificação precoce dos casos de sepsis em ambiente hospitalar, tendo sido levantados os valores de acurácia, sensibilidade e especificidade. A segunda etapa foi composta por duas atividades, com o objetivo de selecionar e pré-processar os dados extraídos do Sistema de Gestão Hospitalar e Ambulatorial do Sistema Único de Saúde do governo do estado do Paraná. O resultado desta etapa gerou um conjunto de dados inicial composto por 55.590 registros. Esse conjunto originou outros três conjuntos de séries temporais: o primeiro foi criado utilizando o código internacional de doenças, cadastrado no prontuário eletrônico do paciente; o segundo foi criado utilizando os critérios do escore *Sequential Organ Failure Assessment*; e o terceiro foi criado utilizando os critérios do escore *Quick Sequential Organ Failure Assessment*. A terceira etapa foi a implementação dos algoritmos *Long Short Term Memory* e *random forest*, identificados na revisão de literatura e selecionados por obterem os melhores valores de acurácia, sensibilidade e especificidade. A quarta etapa foi a avaliação dos resultados, tendo como métricas adotadas: acurácia, sensibilidade, especificidade e curva *Receiver Operator Characteristic* dos experimentos. **Resultados:** Utilizando o código internacional de doenças como critério de seleção de pacientes em risco de sepsis, o algoritmo *Long Short Term Memory* obteve acurácia de 31,6%, sensibilidade de 8,4% e especificidade de 75,0%; com o escore *Sequential Organ Failure Assessment*, alcançou 68,7% de acurácia, 56,3% de sensibilidade e 76,8% de especificidade; e, empregando o escore *Quick Sequential Organ Failure Assessment*, obteve 87,2% de acurácia, 87,7% de sensibilidade e 86,9% de especificidade. Para o algoritmo *random forest*, os resultados obtidos foram: 97,1% de acurácia, 61,1% de sensibilidade e 99,8% de especificidade. **Conclusão:** A partir dos resultados obtidos, é possível observar que o algoritmo proposto (*Long Short Term Memory*) apresentou bons resultados no reconhecimento precoce de pacientes, indicando resultados ainda mais promissores na identificação baseada em séries temporais. O produto final é uma ferramenta que pode ser utilizado pelos hospitais por meio de uma aplicação funcional, auxiliando o corpo clínico hospitalar a identificar os pacientes em risco.

Palavras-chave: Sepsis. Inteligência artificial. Modelo preditivo. *Machine learning*. *Deep learning*. Informática em saúde. Informática biomédica.

ABSTRACT

Introduction: Sepsis is a generalized inflammation with high morbidity and mortality, being the leading cause of death in intensive care units. Early treatment and recognition are essential factors for a better quality of life for the patient. Each hour in delay in starting treatment reduces the survival rate by up to 8%. **Objective:** Develop a tool that identifies early sepsis cases through Deep Learning for hospitals managed by the Unique System of Health Hospital and Ambulatory Management System of the Paraná State Government. **Method:** This is a descriptive, retrospective and quantitative research composed by four steps. The first involved a literature review to identify the techniques based on artificial intelligence used in early recognition of sepsis cases in hospital environment. Accuracy, sensitivity and specificity values were raised. The second step consisted by two activities aiming to select and pre-process the data extracted from Hospital and Ambulatory Management System of the Paraná State Government. The result of this step generated an initial data set with 55,590 records. This initial set generated three others time series data sets, the first set was created using the registered the international classification of disease in the patient's electronic medical record, the second set was created using the *Sequential Organ Failure Assessment* score criteria and the third set was created using the Quick *Sequential Organ Failure Assessment* score criteria. The third step is the implementation of the *Long Short Term Memory* and Random Forest algorithms, identified in the literature review and selected for obtaining the best accuracy, sensitivity and specificity values. The fourth step was the evaluation of the results, using the following metrics: accuracy, sensitivity and specificity and the ROC curve of each experiment. **Results:** Using the international disease code as criteria for selecting patients at risk for sepsis, the *Long Short Term Memory* algorithm obtained accuracy of 31.6%, sensitivity 8.4% and specificity of 75%, using the *Sequential Organ Failure Assessment* score as criteria for selecting patients at sepsis risk the algorithm obtained 68.7% accuracy, 56.3% sensitivity and 76.8% specificity and using the Quick *Sequential Organ Failure Assessment* as criteria for selecting patients at sepsis risk, the algorithm obtained 87.2% accuracy, 87.7% sensitivity and 86.9% specificity. For Random Forest algorithm, the results obtained were 97.1% accuracy, 61.1% sensitivity and 99,8% specificity. **Conclusion:** From the obtained results it can be observed that the proposed algorithm, *Long Short Term Memory*, presented good results in the early recognition of patients, indicating even more promising results in the identification based on time series. The final product is a tool that can be used by hospitals through a functional application to help hospital staff identify patients at risk.

Keywords: Sepsis. Artificial intelligence. Predictive model. Machine learning. Deep learning. Health informatics. Biomedical Informatics

LISTA DE FIGURAS

Figura 1 – Locais e módulos do GSUS implantados.	26
Figura 2 – Relação entre IA, ML e <i>deep learning</i>	27
Figura 3 – Hierarquia adotada: algoritmos de ML.	28
Figura 4 - Representação das situações de sub-ajuste e sobre-ajuste em modelos treinados.	32
Figura 5 – Árvore de decisão – representação do modelo de classificação.	38
Figura 6 – Gráfico de representação da função sigmoide.	40
Figura 7 – Fluxo das etapas da pesquisa.	44
Figura 8 - Exemplo de curva ROC.	52
Figura 9 – Modelo de dados desenvolvido para tratamento dos dados.	56
Figura 10 – Percentual de dados faltantes para os atributos de sinais vitais.	58
Figura 11 – Percentual de dados faltantes para os atributos de sinais vitais para os anos de 2017 e 2018.	59
Figura 12 – Curvas ROC obtidas do algoritmo LSTM para as bases criadas por meio dos critérios qSOFA, SOFA e GSUS.	65
Figura 13 – Curva ROC obtida por meio do treinamento do algoritmo <i>random forest</i>	67
Figura 14 – Gráfico de importância <i>mean decrease accuracy</i> , referente aos atributos preditores para o conjunto de dados selecionados a partir do critério qSOFA.	68
Figura 15 – Gráfico de importância <i>mean decrease GINI</i> , referente aos atributos preditores para o conjunto de dados selecionados a partir do critério qSOFA.	69
Figura 16 – Processo de tratamento dos dados, treinamento e predição do modelo proposto.	70
Figura 17 – Tela de identificação dos pacientes que satisfazem os critérios do escore qSOFA e seu valor preditivo respectivo.	71
Figura 18 – Valores dos registros de sinais vitais e valores preditivos diários de um único paciente.	72
Figura 19 – Gráfico de visualização dos padrões de falta de dados e combinações entre os atributos.	73

LISTA DE QUADROS

Quadro 1 – Critérios avaliados por meio do escore SOFA.....	22
Quadro 2 – Critérios qSOFA, utilizados para investigação de disfunção orgânica....	22
Quadro 3 – Módulos do GSUS desenvolvidos, em desenvolvimento e previstos para 2020.	24
Quadro 4 – Mapeamento entre o referencial teórico e os encaminhamentos metodológicos.	41
Quadro 5 – Atributos de pacientes e internamentos selecionados a partir de sua relevância preditiva.	45
Quadro 6 – Descrição dos códigos da família A41 – Outras septicemias, conforme CID10.	45
Quadro 7 – Termos utilizados na definição da sensibilidade e especificidade em um teste de diagnóstico.	51

LISTA DE TABELAS

Tabela 1 – Auroc, sensibilidade e especificidade, por gravidade da sepse, obtidos a partir do algoritmo de ML InSight; e dos escores MEWS, SOFA e SIRS.	18
Tabela 2 – Auroc, sensibilidade e especificidade, por percentual de dados faltantes, obtidos a partir do algoritmo de ML InSight em comparação com o escore SIRS.....	18
Tabela 3 – Número total de pacientes internados por ano, segundo hospitais.	56
Tabela 4 – Número total de pacientes maiores de 18 anos internados por ano, segundo hospitais.	57
Tabela 5 – Número total de internamentos por ano, segundo hospitais.	57
Tabela 6 – Número total de pacientes maiores de 18 anos internados por ano, segundo hospitais.	58
Tabela 7 – Perfil dos internamentos nos grupos com e sem CID de sepse do hospital H3.....	60
Tabela 8 – Quantidade de internamentos com suspeita de sepse identificados por meio dos critérios GSUS, qSOFA e SOFA e suas respectivas quantidades de óbitos.	61
Tabela 9 - Coeficientes Kappa para avaliação entre os escores.....	62
Tabela 10 – Parâmetros propostos e identificados para implementação do LSTM a partir do <i>grid-search</i>	63
Tabela 11 – Acurácia, sensibilidade e especificidade obtidos a partir das bases criadas por meio dos critérios GSUS, SOFA e qSOFA, intervalo de confiança de 95%.....	64
Tabela 12 – Matriz de confusão obtida para o modelo M-qSOFA, treinado com a série temporal qSOFA.....	64
Tabela 13 – Matriz de confusão obtida para o modelo M-SOFA, treinado com a série temporal SOFA.....	65
Tabela 14 – Matriz de confusão obtida para o modelo M-GSUS, treinado com a série temporal GSUS.	65
Tabela 15 – Valores de acurácia, sensibilidade e especificidade obtidos por meio do treinamento do LSTM utilizando <i>timesteps</i> de 8, 5, 3 e 1 dia de internamento, com intervalo de confiança de 95%.	66
Tabela 16 – Comparativo com o teste do modelo M-qSOFA utilizando os dados do hospital H2, com , intervalo de confiança de 95%.....	66

Tabela 17 – Resultados obtidos de acurácia, sensibilidade, especificidade, valor preditivo positivo e valore preditivo negativo, utilizando o algoritmo <i>random forest</i> , com intervalo de confiança de 95%.	67
Tabela 18 – Matriz de confusão para o algoritmo <i>random forest</i>	67

LISTA DE ABREVIATURAS E SIGLAS

Auroc	<i>Area Under the Receiver Operating Characteristic</i>
AUC	<i>Area Under the ROC Curve</i>
CEC	<i>Constant Error Carousel</i>
Celepar	Companhia de Tecnologia, Informação e Comunicação do Estado do Paraná
Cemepar	Centro de Medicamentos do Paraná
CEPDH	Centro de Estudos, Pesquisa e Desenvolvimento Humano
CID	Classificação Internacional de Doenças
CNN	<i>Convolutional Neural Network</i>
CRAID	Centro Regional de Atendimento Integrado ao Deficiente
CRE	Centro Regional de Especialidades
DBN	<i>Deep Belief Network</i>
DNN	<i>Deep Neural Network</i>
ESICM	European Society of Intensive Medicine
FE	<i>Feature engineering</i>
FP	<i>False Positive</i>
FN	<i>False Negative</i>
GSUS	Sistema de Gestão da Assistência de Saúde do SUS
Hemepar	Centro de Hematologia e Hemoterapia do Paraná
HMA	Histórico de Moléstia Atual
HMM	<i>Hidden Markov Model</i>
HT	Hospital do Trabalhador
HU	Hospital Universitário
IA	Inteligência Artificial
ILAS	Instituto Latino Americano de Sepsis
IRAS	Infecções Relacionadas à Assistência à Saúde
LSTM	<i>Long Short-Term Memory</i>
MEWS	<i>Modified Early Warning Score</i>
ML	<i>Machine Learning</i>
MLP	<i>Multilayer Perceptron</i>
MSE	<i>Mean Square Error</i>
PUCPR	Pontifícia Universidade Católica do Paraná

qSOFA	<i>Quick Sequential Organ Failure Assessment</i>
RNA	Rede Neural Artificial
RNN	<i>Recurrent Neural Network</i>
ROC	<i>Receiver Operator Characteristic</i>
SAME	Serviço de Arquivo Médico e Estatística
SCCM	Society of Critical Care Medicine
SESA	Secretaria de Estado da Saúde do Paraná
SETI	Secretaria da Ciência, Tecnologia e Ensino Superior do Paraná
SGD	<i>Stochastic Gradient Descendent</i>
SIRS	Síndrome da Resposta Inflamatória Sistêmica
SOFA	<i>Sequential Organ Failure Assessment</i>
SUS	Sistema Único de Saúde
SVM	<i>Support Vector Machine</i>
TN	<i>True Negative</i>
TP	<i>True Positive</i>
UTI	Unidade de Terapia Intensiva

SUMÁRIO

1	INTRODUÇÃO	16
1.1	OBJETIVO GERAL.....	19
1.2	OBJETIVOS ESPECÍFICOS	19
1.3	CONTRIBUIÇÃO CIENTÍFICA E SOCIAL	19
2	REFERENCIAL TEÓRICO	20
2.1	SEPSE.....	21
2.2	GSUS	23
2.3	<i>MACHINE LEARNING</i>	27
2.3.1	Engenharia de atributos	30
2.3.2	Regularização, sobreajuste e sub-ajuste	31
2.3.3	<i>Deep learning</i> e LSTM	34
2.3.4	Otimização de hiperparâmetros	35
2.3.5	<i>Random forest</i>	37
2.3.6	Regressão logística (<i>logistic regression</i>)	39
2.4	TRABALHOS RELACIONADOS.....	40
3	MAPEAMENTO ENTRE O REFERENCIAL TEÓRICO E OS ENCAMINHAMENTOS METODOLÓGICOS	41
4	ENCAMINHAMENTOS METODOLÓGICOS	43
4.1	CENÁRIO DA PESQUISA.....	43
4.2	ETAPAS DO PROJETO	43
4.2.1	Etapa 1 – revisão de literatura e Comitê de Ética	44
4.2.2	Etapa 2 – seleção dos dados e pré-processamento	44
4.2.3	Etapa 3 – implementação do algoritmo LSTM	49
4.2.4	Etapa 4 – avaliação dos resultados	50
4.2.5	Protocolo de testes	53
5	RESULTADOS	55
5.1	RESULTADO DA ETAPA 4 – DESENVOLVIMENTO DO MODELO PREDITIVO	63
6	DISCUSSÃO	74
6.1	LIMITAÇÕES.....	77
6.2	TRABALHOS FUTUROS.....	78
7	CONSIDERAÇÕES FINAIS	80

REFERÊNCIAS.....	81
APÊNDICE A – ARTIGO DE REVISÃO	89
ANEXO A – APROVAÇÃO DO COMITÊ DE ÉTICA DA PUCPR	105
ANEXO B – APROVAÇÃO DO COMITÊ DE ÉTICA DA SESA.....	109
ANEXO C – TERMO DE COMPROMISSO DE UTILIZAÇÃO DE DADOS	111
ANEXO D – TERMO DE DISPENSA DO TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO	112
ANEXO E – PROCESSO ADMINISTRATIVO DE SOLICITAÇÃO DE PERMISSÃO PARA UTILIZAÇÃO DOS DADOS DO GSUS PARA PESQUISA ACADÊMICA ..	113

1 INTRODUÇÃO

Sepse pode ser definida como a resposta sistêmica a uma doença infecciosa, seja ela causada por bactérias, fungos, vírus ou protozoários, manifestando-se em diferentes estágios clínicos de um mesmo processo fisiopatológico (ILAS, 2017). Os estágios clínicos são definidos, em ordem crescente de gravidade, em: sepse, sepse grave ou severa e choque séptico (DESAUTELS et al., 2016). Essas definições representam a evolução temporal da mesma síndrome, com diferentes espectros de gravidade, associados à crescente taxa de mortalidade. Inicia-se como um processo inflamatório agudo frente à infecção do paciente, podendo evoluir para uma inflamação exacerbada e crônica, chamada Síndrome da Resposta Inflamatória Sistêmica (SIRS) (HENKIN et al., 2009).

Nos Estados Unidos, pelo menos 750 mil pessoas desenvolvem sepse grave ou choque séptico todo ano; a estimativa é de três casos para cada mil habitantes e 2,26 casos por cem altas hospitalares (ANGUS et al., 2001). Mais da metade desses casos foi admitido em uma Unidade de Terapia Intensiva (UTI) e corresponde de 20% a 30% do total de mortes do hospital, com um custo anual de 15,4 milhões de dólares (ANGUS et al., 2001; HENRY et al., 2015). No Reino Unido, a sepse afeta mais de 260 mil pessoas por ano, gerando um custo direto de dois milhões de libras (FAISAL et al., 2018).

No Brasil, Machado et al. (2017) publicaram um estudo de prevalência em ambiente nacional, selecionando 230 UTIs em um só dia, e identificaram que 30% dos leitos estavam ocupados por pacientes com sepse ou choque séptico. Lobo et al. (2019) divulgaram, em seu artigo, as tendências temporais da prevalência e da mortalidade da sepse, contando com a participação de cerca de 30% dos leitos de UTIs adulto do país. Os dados demonstraram aumento progressivo de casos de sepse nas UTIs, de 19,4% em 2010 para 25,2% em 2016, além da queda nas taxas de mortalidade, de 39,0% em 2010 para 30,0% em 2016.

Não existem dados oficiais da Secretaria de Estado da Saúde do Paraná (SESA) com relação ao número de casos de sepse, porém há alguns estudos, como o de Souza et al. (2015), que mapearam as Infecções Relacionadas à Assistência à Saúde (IRAS) no Hospital Universitário (HU) de Londrina. Este foi realizado de dezembro de 2009 a janeiro de 2011, período durante o qual se identificaram 889 pacientes com algum tipo de infecção; 413 deles (46,45%) apresentaram sepse,

sendo que 69,00% evoluíram a óbito. Sales Júnior et al. (2006) realizaram um estudo prospectivo em 65 hospitais de todas as regiões do Brasil, analisando os pacientes admitidos no mês de setembro de 2003. As regiões Sul e Sudeste tiveram a maior participação – 73,8% do total dos dados. Com a participação de oito hospitais, a região Sul relatou 103 pacientes e apresentou mortalidade global de 52,4%, ficando acima da média de 46,6% constatada pelos autores.

Uma das principais razões para um número tão alto de casos de morte reside na compreensão e conhecimento limitados sobre o complexo mecanismo de resposta inflamatória (GHALWASH; RADOSAVLJEVIC; OBRADOVIC, 2013); essa situação resulta no reconhecimento tardio da sepse, pois suas manifestações podem ser confundidas com as de outros processos não infecciosos ou mesmo podem passar despercebidas (ILAS, 2017). As intervenções iniciais dependem do reconhecimento precoce, que é um processo contínuo e manual do profissional de saúde, sendo esta uma das maiores dificuldades na prática clínica, por depender diretamente da habilidade do profissional de identificar os pacientes em risco (WESTPHAL; LINO, 2015). Essas identificações são realizadas por meio de escores específicos que mensuram a gravidade e são utilizados para fazer o reconhecimento dos casos, porém falham na identificação de um a cada oito pacientes com sepse grave (KAUKONEN et al., 2015).

Kumar et al. (2006) demonstraram que, a cada hora de atraso no tratamento, a taxa de sobrevivência reduz em até 8%, e um diagnóstico rápido, aliado ao início precoce da terapia, administrando antimicrobianos, preferencialmente em até 60 minutos após o reconhecimento da hipotensão, causa impacto positivo na redução da mortalidade. O reconhecimento precoce provê benefícios para o paciente no tratamento e nos resultados, chegando a reduzir de 46,5% a 30,5% a mortalidade (RIVERS et al., 2001). Kumar et al. (2006) também observaram que a introdução de antibióticos eficientes, durante a primeira hora de hipotensão documentada, aumenta a sobrevida dos pacientes com choque séptico.

Estudos como de Gultepe et al. (2012), Lipton et al. (2016), Kam e Kim (2017), Zhang et al. (2017) e Mao et al. (2018) apresentaram o potencial de métodos baseados em *Machine Learning* (ML) para apoiar a predição de sepse. Além de modelos preditores, outros estudos realizaram o mapeamento de diversas progressões relacionadas à sepse, como mortalidade em Garcia-Gallo et al. (2018),

falha sistêmica de órgãos em Gultepe et al. (2014) e possíveis terapias para o tratamento de pacientes (GHALWASH; RADOSAVLJEVIC ; OBRADOVIC, 2013).

Mao et al. (2018) compararam um algoritmo de ML de detecção e predição de sepse chamado InSight com três sistemas de escore utilizados para identificação de sepse em ambientes hospitalares: *Modified Early Warning Score* (MEWS), SIRS e *Sequential Organ Failure Assessment* (SOFA). Ainda no estudo de Mao et al. (2018), foram utilizados 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. Os resultados mostraram que a acurácia, demonstrada pelos valores *Area Under de ROC Curve* (AUC), sensibilidade e especificidade do algoritmo de ML InSight, obteve melhores resultados em relação aos outros três escores, conforme Tabela 1.

Tabela 1 – Auroc, sensibilidade e especificidade, por gravidade da sepse, obtidos a partir do algoritmo de ML InSight; e dos escores MEWS, SOFA e SIRS.

	Gravidade	InSight	MEWS	SOFA	SIRS
Auroc	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
Sensibilidade	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
Especificidade	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, p. 6).

Mao et al. (2018) ainda demonstraram que, mesmo com 60% de dados faltantes, o algoritmo de ML InSight alcançou melhores resultados que o sistema de escore SIRS, conforme Tabela 2.

Tabela 2 – Auroc, sensibilidade e especificidade, por percentual de dados faltantes, obtidos a partir do algoritmo de ML InSight em comparação com o escore SIRS.

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

Fonte: Mao et al. (2018, p. 7).

Portanto, considerando a complexidade envolvida na detecção precoce de sepse, este estudo tem como objetivo desenvolver uma ferramenta para identificação precoce de risco de sepse nos hospitais gerenciados pelo Sistema de Gestão da Assistência de Saúde do SUS (GSUS), baseado em estratégias de Inteligência Artificial (IA).

1.1 OBJETIVO GERAL

Desenvolver uma ferramenta para identificação dos casos de risco de sepse nos hospitais que são gerenciados pelo GSUS do governo do estado do Paraná.

1.2 OBJETIVOS ESPECÍFICOS

- a) Realizar uma revisão integrativa para identificação da eficiência das técnicas de IA adotadas para reconhecimento de sepse em ambientes hospitalares.
- b) Construir um modelo preditivo que identifique os casos de sepse nos hospitais gerenciados pelo GSUS do governo do estado do Paraná.
- c) Avaliar o modelo preditivo construído.

1.3 CONTRIBUIÇÃO CIENTÍFICA E SOCIAL

Este estudo visa a contribuir para a disseminação do uso da IA para suporte aos processos de saúde (CARVALHO; ESCOBAR; TSUNODA, 2014). A identificação precoce dos casos de sepse promove melhoria à qualidade de vida dos pacientes, bem como segurança às atividades desenvolvidas pelos profissionais de saúde no cuidado ao paciente (RIVERS et al., 2001; GULTEPE et al., 2012; KAUKONEN et al., 2015).

Também propicia uma visão das ocorrências dos casos de sepse em um hospital do estado, permitindo comparar os dados obtidos com os achados na literatura, oportunizando a identificação de pontos a ser melhorados tanto do ponto de vista de registros no sistema hospitalar quanto de procedimentos de segurança no cuidado ao paciente.

2 REFERENCIAL TEÓRICO

Nesta seção, são apresentados conceitos e trabalhos relacionados que sustentam o desenvolvimento desta dissertação. Na subseção 2.1, traz-se o conceito de sepse, bem como a importância do reconhecimento precoce e sua respectiva dificuldade de identificação. Na subseção 2.2, é apresentado o GSUS, fonte dos dados utilizados para o desenvolvimento do modelo de reconhecimento de sepse.

Na subseção 2.3, são abordados a ML e suas respectivas características, para fundamentar a utilização do método de treinamento supervisionado, justificando a escolha do *Long Short-Term Memory* (LSTM) como algoritmo de classificação. Na subseção 2.3.2, são apresentados os problemas recorrentes na regularização dos modelos: o sobreajuste (*overfitting*) e o sub-ajuste (*underfitting*). Na subseção 2.3.3, exploram-se o conceito de *deep learning* e o algoritmo LSTM, bem como sua característica e aplicações no contexto clínico. Em 2.3.5 e 2.3.6, são abordados, respectivamente, os métodos de classificação *random forest* e regressão logística, identificados na revisão integrativa como segundo e terceiro lugares na avaliação de desempenho na identificação de pacientes com sepse.

A subseção 2.4 é a de trabalhos relacionados, que, por meio de uma revisão integrativa, teve o objetivo de levantar na literatura as técnicas de IA adotadas no processo de identificação precoce de casos de sepse em ambientes hospitalares, permitindo comparar os resultados das técnicas identificadas. Os resultados foram ordenados conforme teste de diagnóstico, por meio dos valores de sensibilidade, especificidade e acurácia. Nesta análise, também foi possível identificar as bases utilizadas, possibilitando uma comparação dos resultados dos modelos frente às informações sobre os conjuntos de dados, como origem da fonte de dados, quantidade de registros e desbalanceamento. O resultado da revisão foi um artigo, intitulado: *Técnicas de inteligência artificial para reconhecimento de sepse em ambientes hospitalares: revisão integrativa*, submetido e aceito em 27 de agosto de 2019 pela Revista de Gestão em Sistemas de Saúde (*Health Systems Management Journal*). O artigo na sua íntegra está no Apêndice A.

2.1 SEPSE

Sepse é uma resposta inflamatória descontrolada disparada por uma infecção, pode ser desencadeada por uma infecção, trauma, queimadura ou outro motivo clínico e atacar qualquer um, em qualquer idade (ANGUS et al., 2001). O critério SIRS foi proposto na conferência de 1992, por meio da American College of Chest Physicians/Society of Critical Care Medicine Consensus (BONE et al., 1992).

Quando o tratamento é tardio, a sepse pode rapidamente avançar para um choque séptico ou para sepse severa e acabar em morte. A taxa de mortalidade estimada para pacientes com sepse severa é de 60%, sendo que os sobreviventes apresentam baixa qualidade de vida (AMLAND; HAHN-COVER, 2016). Devido à sua alta taxa de mortalidade, é imprescindível uma rápida identificação, porém essa não é uma atividade trivial, pois suas manifestações podem ser confundidas com as de outros processos não infecciosos, visto que os sinais e sintomas atualmente utilizados para o diagnóstico, como alteração na contagem de leucócitos, febre, taquicardia e taquipneia, não são específicos da sepse (SINGER et al., 2016).

Em 2016, a Society of Critical Care Medicine (SCCM) e a European Society of Intensive Care Medicine (ESICM) atualizaram as definições de sepse. O novo conceito, conhecido como Sepse-3, define-a como uma disfunção orgânica ameaçadora à vida secundária à resposta desregulada do hospedeiro a uma infecção (SINGER et al., 2016), que pode ser avaliada por meio do escore SOFA. As definições anteriores são conhecidas por Sepsis-1, de 1991, e Sepsis-2, de 2001.

O escore SOFA foi criado pela ESICM em outubro de 1994, com o objetivo de descrever quantitativamente a disfunção ou a falha de órgãos decorrentes do quadro clínico séptico. Ele é utilizado para melhorar o entendimento da história natural da disfunção/falha do órgão e a inter-relação da falha em vários órgãos, além de permitir a avaliação dos efeitos de novas terapias no tratamento da disfunção.

É importante perceber que a pontuação do SOFA é projetada não para prever o resultado, mas para descrever uma sequência de complicações no doente crítico (VINCENT; MORENO; TAKALA, 1996). A pontuação é baseada em dados clínicos disponíveis e parâmetros laboratoriais, porém o escore não leva em conta outros fatores associados, como idade ou outras comorbidades (MACDONALD et al., 2017). O escore SOFA é apresentado no Quadro 1.

Quadro 1 – Critérios avaliados por meio do escore SOFA.

Descrição	Pontuação			
	1	2	3	4
Respiração PaO ₂ /FiO ₂ mmHg	< 400	< 300	< 200	< 100
Coagulação Plaquetas 10 ³ /mm ³	< 150	< 100	< 50	< 20
Fígado (bilirrubina mg/dl)	1,2 – 1,9 (20 – 32)	2,0 – 5,9 (33 – 101)	6,0 – 11,9 (102 – 204)	> 12,0 (< 204)
Hipotensão cardiovascular	PAM < 70 mmHg	Dopamina ≤ 5a ou dobutamina qualquer dose	Dopamina > 5 ou epinefrina/ norepinefrina ≤ 0,1	Dopamina > 15 ou epinefrina/ norepinefrina > 0,1
Sistema nervoso central Escala de coma de Glasgow	13 – 14	10 – 12	6 – 9	<6
Renal creatinina (mg/dL) ou débito urinário (ml/dia)	1,2 – 1,9 (110 – 170)	2,0 – 3,4 (171 – 299)	3,5 – 4,9 (300 – 440) ou < 500 ml/dia	> 5,0 (> 440) ou < 200 ml/dia

Fonte: Adaptado de Vincent, Moreno e Takala (1996).

Notas: Agentes adrenérgicos administrados por pelo menos 1h (as doses administradas são em gg/kg-min). A pontuação é somada, com o escore total se associando com o maior risco de óbito. PaO₂ = pressão de oxigênio arterial. FiO₂ = fração inspirada de oxigênio. mmHg = milímetros de mercúrio. PAM = pressão arterial média.

De acordo com os critérios estabelecidos no último consenso publicado (Sepse-3), a sepse passou a ser definida com dois ou mais pontos do escore SOFA basal, em consequência da infecção. Esse critério baseou-se na identificação de uma curva *Receiver Operator Characteristic* (ROC) mais acurada para a variação SOFA do que a da presença da SIRS (ILAS, 2017).

Singer et al. (2016) observaram que alguns dos critérios do SOFA eram mais frequentemente gravados nos registros eletrônicos dos pacientes com sepse e poderiam servir como alerta para facilmente sinalizar pacientes com infecção. Esse escore resumido foi chamado *Quick SOFA* (qSOFA). Entretanto, ele é menos robusto que o SOFA e não requer exames laboratoriais, o que facilita a análise do paciente a beira do leito. Os critérios desse escore são apresentados no Quadro 2.

Quadro 2 – Critérios qSOFA, utilizados para investigação de disfunção orgânica.

Critério	Valor
Frequência respiratória	≥ 22 respirações/min
Pressão arterial: sistólica	≤ 100 mmHg
Escala de coma de Glasgow	<15

Fonte: Singer et al. (2016, p. 8).

Inicialmente, o qSOFA foi desenvolvido para ser utilizado na identificação de pacientes de enfermaria que precisavam de maior atenção ou para determinar a necessidade de início imediato de tratamento apropriado. Alguns trabalhos, como o de Gupta et al. (2018), compararam os resultados do escore qSOFA com outros escores e concluíram que ele foi mais preciso para avaliação do risco de mortalidade entre os pacientes do departamento de emergência do hospital, obtendo um melhor balanço entre a sensibilidade e a especificidade. Quando comparados os escores SOFA e qSOFA, este se mostrou mais robusto na validação preditiva quando as condições de medidas foram variadas; entretanto, foi estatisticamente inferior se comparado ao SOFA para pacientes de UTI. Sendo assim, a recomendação é usar o escore SOFA de dois pontos ou mais para pacientes em UTIs e qSOFA para pacientes fora da UTI (KUSHIMOTO et al., 2018; SERAFIM et al., 2018).

2.2 GSUS

Motivada pela necessidade de gerir os serviços de assistência à saúde de seus hospitais próprios, a SESA, em conjunto com os HUs do estado, fomentou a busca de um sistema multi-hospitalar para gestão integrada dos hospitais do estado. Nessa oportunidade, foi designada à Companhia de Tecnologia, Informação e Comunicação do Estado do Paraná (Celepar) a elaboração do projeto de informatização, conhecido como GSUS.

O projeto foi elaborado com a característica de ser um sistema multi-institucional que deveria servir aos hospitais próprios da SESA, Hospital de Clínicas do Paraná, HUs da Secretaria da Ciência, Tecnologia e Ensino Superior do Paraná (SETI) e unidades administrativas da SESA relacionadas diretamente com a assistência de saúde do Sistema Único de Saúde (SUS). As instituições tinham algumas dificuldades em comum, tais como:

- a) As soluções informatizadas adotadas, próprias ou terceiras, não atendiam às necessidades dos processos atuais.
- b) Dificuldades de evoluir as soluções atuais ou mesmo desenvolver uma nova solução, devido à falta de recursos humanos e tecnológicos nas instituições.

- c) Caracterização da necessidade de uniformizar os processos e funcionalidades entre as instituições, visando à integração dos serviços de assistência à saúde.
- d) Interesse da SESA na gestão integrada das informações de assistência.

Sendo assim, o GSUS foi construído com o objetivo de ser uma solução de gestão hospitalar e ambulatorial atendendo a hospitais, prontos-socorros, clínicas ambulatoriais, laboratórios, farmácias, centros de diagnóstico, centrais de regulação e demais unidades de saúde que operam segundo as regras do SUS. Os usuários são pessoas do corpo administrativo, médicos, enfermeiros, terapeutas, profissionais de saúde, auditores/reguladores e gestores da assistência.

O sistema é composto por 25 módulos, dos quais 16 já estão desenvolvidos e implantados, quatro novos estão em desenvolvimento e outros seis ainda fazem parte do projeto, com seu desenvolvimento previsto para 2020. Os módulos estão descritos no Quadro 3:

Quadro 3 – Módulos do GSUS desenvolvidos, em desenvolvimento e previstos para 2020.

Desenvolvidos	Em desenvolvimento	Previstos para 2020
Serviço de Arquivo Médico e Estatística (SAME)	Regulação Ambulatorial de Consultas	Centro Cirúrgico
Infraestrutura	Exames, Serviço de Transporte e Remoção	Unidade de Terapia Intensiva
Serviço Ambulatorial	Controle, Avaliação, Faturamento e Auditoria	Serviço de Controle de Infecção Hospitalar
Pronto Atendimento		Controle de Assistência por Especialidade
Internação		Serviço de Engenharia e Manutenção Hospitalar
Atendimento Corpo Clínico		Serviço de Hotelaria
Serviço de Enfermagem		
Serviço de Nutrição e Lactário		
Serviço de Farmácia		
Serviço de Laboratório		
Interfaceamento de Equipamentos		
Centro de Diagnóstico		
Centro de Material Esterilizável		
Agência Transfusional de Sangue		
Módulo Gerencial		
Regulação de Leitos		

Fonte: Grein (2018, p. 33).

O GSUS foi desenvolvido com o objetivo de integrar todos os dados e processos de gestão da assistência à saúde em um único sistema. Sua estrutura foi planejada para suportar o acoplamento de módulos que visam à implantação integral de modo gradativo. Esses módulos compreendem um conjunto de serviços internos referentes à organização dos estabelecimentos e estão integrados a serviços externos por meio de *web services* de sistemas de bases nacionais e municipais.

Os seguintes objetivos orientaram seu desenvolvimento:

- a) Prover um sistema multi-hospitalar *on-line*.
- b) Organizar os dados e processos da assistência à saúde integrada em um mesmo sistema.
- c) Servir como instrumento de trabalho para gestão de estabelecimentos de assistência do SUS.
- d) Automatizar os processos operacionais de assistência nos serviços internos das unidades.
- e) Estruturar a prescrição eletrônica de medicamentos, laudos, exames, dietas, cuidados, consultas e terapias.
- f) Controlar o prontuário único do paciente integrado e compartilhado entre os estabelecimentos.
- g) Estabelecer documentos e procedimentos operacionais padronizados, conforme protocolos do SUS.
- h) Gerenciar a regulação da assistência à saúde entre os gestores e os prestadores de serviços do SUS.
- i) Promover o uso do prontuário eletrônico do paciente nos estabelecimentos do estado do Paraná.
- j) Permitir o monitoramento de indicadores de produtividade e qualidade da assistência.

O sistema fechou o ano de 2019 implantado em 11 hospitais, 15 ambulatórios e 372 farmácias, com diferentes conjuntos de módulos implantados. Gerando aproximadamente mais um milhão de prontuários de pacientes, gerenciando 12 mil usuários, 318 mil internações, um milhão de atendimentos em pronto atendimento, três milhões de consultas agendadas, 4,8 milhões de evoluções e prescrições de corpo clínico e 3,9 milhões de laudos de exames liberados (referência de setembro de 2019).

A Figura 1 apresenta os municípios que possuem mais módulos implantados e aqueles que possuem apenas o módulo de farmácia (informações de setembro de 2019).

Figura 1 – Locais e módulos do GSUS implantados.



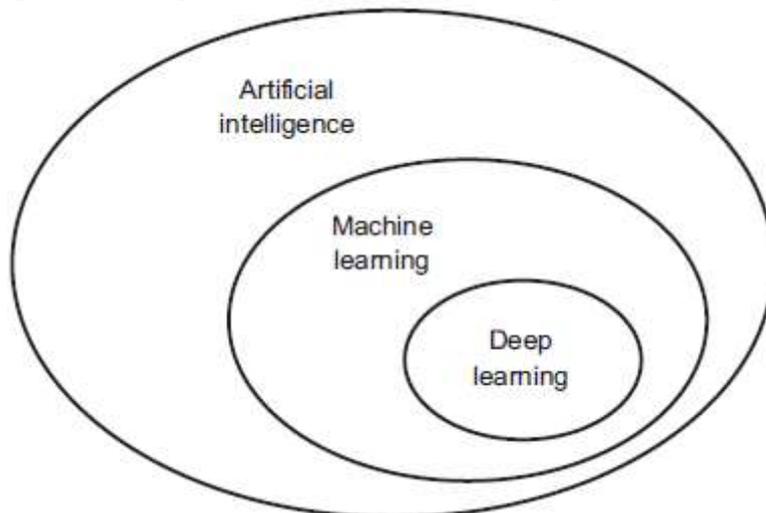
Fonte: GSUS (2019).

As implantações previstas para 2019 que contemplam os módulos de Centro Cirúrgico, Unidade de Terapia Intensiva, Serviço de Controle de Infecção Hospitalar, Controle de Assistência por Especialidade, Serviço de Engenharia e Manutenção Hospitalar e Serviço de Hotelaria, ficaram para 2020, pois o módulo de Regulação de Leitos foi priorizado pelo Ministério da Saúde. Segundo a legislação vigente – Portaria GM/MS nº 1.559, de 1º de agosto de 2008 –, é de competência das secretarias estaduais e municipais organizar o atendimento dos pacientes na rede assistencial, definindo os estabelecimentos para os quais os pacientes que precisam do cuidado deverão ser encaminhados.

2.3 MACHINE LEARNING

ML é um subcampo da ciência da computação que evoluiu a partir do estudo do reconhecimento de padrões e da teoria da aprendizagem computacional em IA (ALPAYDIN, 2014). É definido originalmente como “campo de estudo que dá ao computador a habilidade de aprender sem ser explicitamente programado” (SAMUEL, 1959, p.17). Goodfellow, Bengio e Courville (2015, p. 99) definem o termo “aprendizagem de máquina” como a capacidade computacional de “aprender a partir dos dados”. Chollet (2018) apresenta a ML como uma subárea específica da IA, assim como a *deep learning* é uma subárea específica da ML, conforme Figura 2.

Figura 2 – Relação entre IA, ML e *deep learning*.



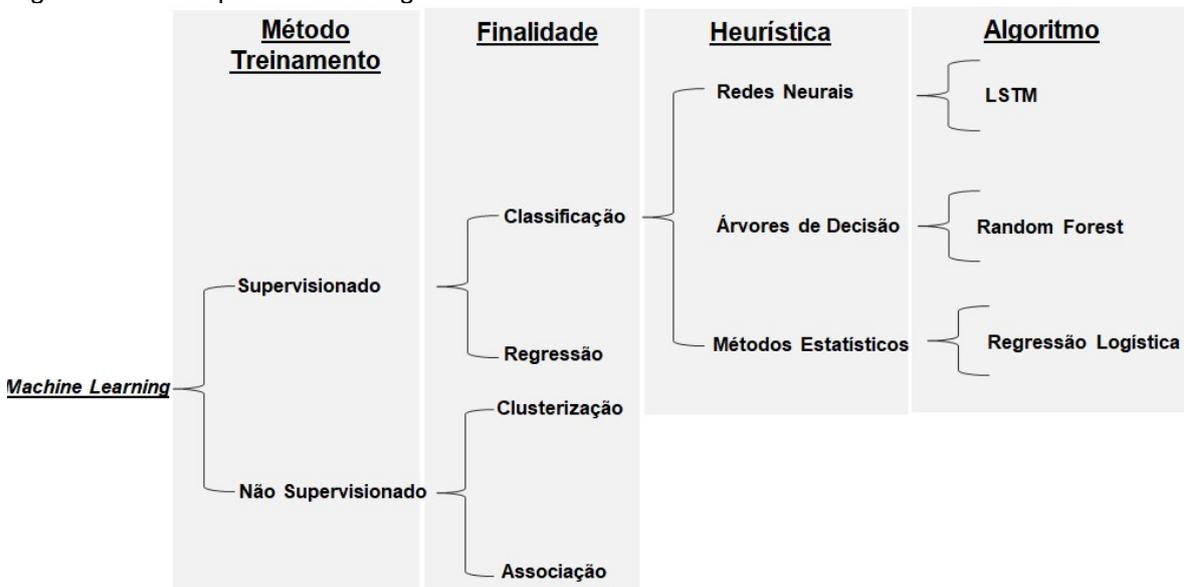
Fonte: Chollet (2018, p. 4).

O campo de estudo da ML trata das formas de aprendizado que orientam o algoritmo; entre as mais comuns, destacam-se o supervisionado e o não supervisionado. No aprendizado supervisionado, o processo ocorre por meio de dados que precisam ser rotulados e bem definidos ou, como descrito por Fellow e Rosenfeld (1976), é o aprendizado em que se aprende com um “professor”. O processo de aprendizado ocorre por um conjunto de dados que contém as respostas corretas, sendo o objetivo que o algoritmo aprenda a partir dos exemplos e desenvolva a habilidade de responder corretamente a todas as entradas possíveis (MARSLAND, 2015). Já no treinamento não supervisionado, o algoritmo precisa reconhecer padrões a partir de um conjunto de dados não rotulados, dando-se o reconhecimento dos

padrões pela identificação de características semelhantes entre os atributos contidos no conjunto de dados (GOODFELLOW; BENGIO; COURVILLE, 2015).

Cada método de treinamento tem uma finalidade, no caso desta dissertação, classificar um paciente em duas classes possíveis: com ou sem risco de sepse. Para executar a classificação, pelo menos um tipo de heurística deve ser utilizado; essa escolha é baseada no problema a ser resolvido, sendo que a característica dos dados também influencia a escolha do algoritmo. A Figura 3 apresenta a hierarquia seguida.

Figura 3 – Hierarquia adotada: algoritmos de ML.



Fonte: Adaptado de Eke et al. (2019).

A qualidade e a quantidade dos dados são cruciais para o sucesso no desenvolvimento do algoritmo de ML (MOHRI; ROSTAMIZADEH; TALWALKAR, 2012); além de a quantidade precisar ser significativa, os dados devem ser preferencialmente sem muito ruído, porém, com o aumento de dados, o custo computacional tende a aumentar (MARSLAND, 2015).

Outra técnica que possibilita a melhoria de desempenho dos algoritmos de ML é a engenharia de atributos (*Feature Engineering* - FE), que é o ato de extrair atributos de dados brutos e transformá-los em formatos adequados para o modelo de ML. Esse é um passo fundamental no processo de aprendizado, pois os atributos corretos podem facilitar a modelagem e, portanto, permitir que o processo produza resultados de maior qualidade (ZHENG; CASARI, 2018). Devido à sua importância no processo de desenvolvimento, as definições, técnicas e utilizações serão aprofundadas na subseção 2.3.1.

A habilidade de aprender automaticamente a partir de um conjunto de dados torna a utilização das Redes Neurais Artificiais (RNAs) na ML uma ferramenta atrativa, pois, em vez de seguir um conjunto de regras especificado por humanos, as redes aprendem por meio de associações subjacentes, como a relação entre os pesos de entrada e saída de uma dada coleção de exemplos representativos. Essa é uma das maiores vantagens das redes neurais sobre os sistemas especialistas tradicionais (JAIN; MAO, 1996).

Muitos trabalhos têm sido desenvolvidos com o aprendizado não supervisionado utilizando RNA, uma vez que seu potencial tem se mostrado eficiente nas mais diversas áreas, como aprendizado regular, linguagens livres de contexto e predição de séries temporais contínuas (RYBICKA; SCHRAUDOLPH; SCHMIDHUBER, 2001).

Como exemplos de aplicação na área médica, Gultepe et al. (2012) identificaram atributos que eram indicativos de sepse, para serem utilizados em uma rede bayesiana com o objetivo de desenvolver um sistema de apoio à decisão. Ghalwash, Radosavljevic e Obradovic (2013) utilizaram ML para mapear as progressões, mortalidade e possíveis tratamentos para as ocorrências de sepse, enquanto Lipton et al. (2016), Kam e Kim (2017) e Mao et al. (2018) empregaram para identificação precoce dos casos de sepse.

As RNAs foram classificadas por Jain e Mao (1996), por padrões de conexão, em duas classes diferentes de redes: *feed-forward* e *recurrent/feedback*. As redes do tipo *feed-forward* possuem apenas conexões unidirecionais entre as camadas, não havendo ligação entre os neurônios da mesma camada. Possuem a característica de serem estáticas, ou seja, produzem apenas um conjunto de valores de saída, em vez de uma sequência de valores; assim, são conhecidas por não possuírem memória, pois o valor de saída depende única e exclusivamente dos valores de entrada, sendo independente do estado prévio da rede (JAIN; MAO, 1996).

Já as redes recorrentes (*recurrent/feedback*) têm ligações entre os neurônios da mesma e de outras camadas anteriores, permitindo o *feedback* das informações (JAIN; MAO, 1996). São conhecidas por ser um sistema dinâmico: quando um novo padrão é apresentado, a camada de saída é recalculada. Devido às conexões de *feedback*, as entradas de cada neurônio são modificadas, o que permite à rede entrar em um novo estado (JAIN; MAO, 1996). Esse tipo de arquitetura permite a resolução de problemas, em que a solução não depende apenas do estado atual da rede, mas

de todo o seu estado prévio (COPPIN, 2004), permitindo sua aplicação a problemas resolvidos utilizando longas séries temporais.

Como o objetivo do algoritmo de ML é a classificação de pacientes, nas subseções seguintes, serão detalhados os problemas mais comuns encontrados nos treinamentos dos algoritmos: sobreajuste (*overfitting*) e *sub-ajuste* (*underfitting*), assim como os algoritmos de classificação que obtiveram os melhores resultados na revisão integrativa apresentada na subseção 2.4, a saber: LSTM, *random forest* e regressão logística.

O LSTM é apresentado em mais detalhes na subseção 2.3.3; *random forest*, na subseção 2.3.5; e regressão logística, na subseção 2.3.6. Tendo em vista a importância dos dados no treinamento dos modelos de ML, a subseção 2.3.1 abordar a técnica de engenharia de atributos, que tem o objetivo de selecionar atributos na intenção de facilitar a modelagem para produzir resultados de melhor qualidade.

2.3.1 Engenharia de atributos

Engenharia de atributos (feature engineering - FE) é o processo de extrair ou selecionar características, a partir de dados brutos, que possam ser utilizadas para reduzir a complexidade e melhorar a acurácia dos modelos de ML. Para Chollet (2018), ela faz uso do conhecimento do próprio profissional sobre o negócio, para aplicar transformações nos dados com o objetivo de fazer o algoritmo de ML melhorar o desempenho. Zheng e Casari (2018) definem-na como o ato de selecionar atributos e transformá-los em formatos adequados para o algoritmo de ML, podendo os atributos corretos somente ser definidos no contexto do problema e do dado; como os modelos e os dados são diversos, é difícil de generalizar a prática entre projetos.

De modo geral, as seleções de atributos enquadram-se em três classes: *filtering*, *wrapper methods* e *embedded methods* (ZHENG; CASARI, 2018). A técnica de *filtering* ou filtragem pré-processa os atributos para remover aqueles atributos que provavelmente não serão úteis para o modelo. Essa técnica tem um custo computacional muito menor que a técnica de *wrapper*, mas não leva em conta o modelo a ser utilizado. Portanto, pode-se não conseguir selecionar os atributos corretos para o modelo. Como recomendação, é melhor fazer uma pré-filtragem de maneira conservadora, para não eliminar algum atributo que inicialmente possa ser

útil antes mesmo de chegar à etapa de treinamento do modelo (ZHENG; CASARI, 208).

A técnica de *wrapper*, apesar de ter um custo computacional alto, permite que o experimento analise um subconjunto de atributos criados por meio da combinação entre eles. Esse processo evita a remoção de atributos que não são úteis à finalidade do algoritmo de ML por si só, porém podem tornar-se úteis quando combinados com outros. Ele é executado em um método separado, onde a seleção dos atributos para a criação do subconjunto, é realizado iterativamente (ZHENG; CASARI, 2018).

A técnica de *embedded* realiza a seleção de atributos como parte do processo de treinamento do modelo. Ela não é tão eficiente quanto o método de *wrapper*, porém o custo computacional é baixíssimo. Em comparação com o *filtering*, a técnica seleciona recursos específicos para o modelo, atingindo um equilíbrio entre custo computacional e qualidade dos resultados (ZHENG; CASARI, 2018).

No entanto, não foi necessária a aplicação de nenhum desses métodos, pois os atributos disponíveis no sistema hospitalar são os mesmos identificados nos estudos que tiveram os melhores resultados de desempenho na revisão integrativa; a maioria dos atributos utilizados, como, por exemplo, sinais vitais, pertence aos escores SOFA e qSOFA, amplamente utilizados como parâmetros no reconhecimento e avaliação da sepse, conforme os estudos de Mitchell et al. (2016), Shashikumar et al. (2017), Lamping et al. (2018), Mao et al. (2018) e Nematy et al. (2018), Garcia-Gallo et al., (2018).

2.3.2 Regularização, sobreajuste e sub-ajuste

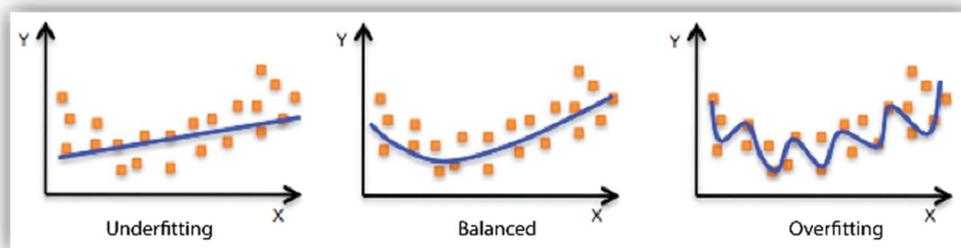
Determinar o modelo apropriado de arquitetura de um modelo de ML é um desafio, pois as configurações dos parâmetros diferem para cada conjunto de dados e, portanto, requerem ajustes personalizados para cada situação (ANDRADE, 2013). Por esse motivo, no processo de desenvolvimento do modelo, sempre será necessário acertar a combinação dos melhores ajustes de configuração, como por exemplo, a escolha do número ou do tamanho das camadas. Esses parâmetros são chamados hiperparâmetros (*hyperparameters*) do modelo, e a configuração dos parâmetros (*parameters*) são os pesos da rede (CHOLLET, 2018).

O objetivo fundamental de determinar o melhor ajuste das configurações dos modelos é o equilíbrio entre a otimização (*optimization*) e a generalização

(*generalization*), representada na figura 4 como *balanced*; enquanto a otimização foca no desempenho de treinamento do modelo, a generalização refere-se ao desempenho do modelo testado com dados nunca vistos antes (GOODFELLOW; BENGIO; COURVILLE, 2015; CHOLLET, 2018).

No início do treinamento, a otimização e a generalização estão correlacionadas; quanto menor for o valor da função da perda (*loss function*) na fase de treinamento, menor será o valor da função de perda na fase de testes. Enquanto isso está ocorrendo, o modelo é dito *underfit* ou sub-ajustado, ou seja, a rede ainda não modelou todos os padrões relevantes nos dados de treinamento, figura 4 - *underfitting*. Depois de certo número de iterações nesses dados, a generalização para de melhorar e as métricas de validação começam a se degradar; nesse ponto, o modelo está começando a realizar um sobreajuste, isto é, começa a apresentar *overfitting*; em outras palavras, aprendeu padrões que são específicos para os dados de treinamento, mas enganosos ou irrelevantes quando se trata de novos dados (MARSLAND, 2015; CHOLLET, 2018). A figura 4 apresenta as representações gráficas das situações de *underfitting*, ocasionada pela incapacidade do modelo treinado em capturar a relação entre os exemplos de treinamento e os valores a serem classificados.

Figura 4 - Representação das situações de sub-ajuste e sobre-ajuste em modelos treinados.



Fonte: Amazon Machine Learning, 2020

As causas do sobreajuste variam, podendo ser o conjunto de dados, os parâmetros configurados no modelo (quantidade de camadas ou unidades em cada camada) ou treinamento do modelo por tempo demais (MARSLAND, 2015). O processo de controlar a ocorrência de sobreajuste é chamado regularização (*regularization*) (CHOLLET, 2018). Goodfellow, Bengio e Courville (2015) definem a regularização como qualquer modificação feita em um algoritmo de aprendizado, com o objetivo de reduzir seu erro de generalização, mas não seu erro de treinamento.

A melhor maneira de prevenir que o modelo aprenda padrões enganosos ou irrelevantes é obter mais dados de treinamento; um modelo treinado com mais dados naturalmente irá generalizar melhor. Quando a obtenção de mais dados não é possível, outra solução é controlar a quantidade de informação permitida para o modelo armazenar ou adicionar validações para quais informações armazenar. Esse processo fará com que o modelo se concentre nos padrões proeminentes, ou seja, aqueles que têm a melhor chance de generalizar (BUDUMA; LOCASCIO, 2017; CHOLLET, 2018).

A redução do tamanho da rede neural também ajuda a prevenir o sobreajuste, sendo o número de parâmetros de aprendizagem no modelo determinado pelo número de camadas e pelo número de unidades em cada camada, conhecido como capacidade do modelo (*capacity*). Conseqüentemente, um modelo com mais parâmetros tem uma capacidade de memorização maior, permitindo mapear os padrões de um conjunto de treinamento, perdendo qualquer poder de generalização. Por outro lado, se o modelo tiver menos recursos de memorização, ele não será capaz de criar um mapeamento com facilidade e terá que recorrer a aprender as representações que têm poder preditivo em relação aos dados, que, resumidamente, é a representação que aqui interessa (CHOLLET, 2018).

O *dropout* é uma das técnicas mais efetivas na regularização das redes, consistindo em randomicamente “desligar” (atribuir com zeros) uma camada de neurônios da rede durante o treinamento. A *dropout rate* é a proporção de valores que serão zerados, normalmente variando entre 0,2 a 0,5. Na fase de testes, nenhuma camada é desligada, apenas as saídas são reduzidas por um fator igual à *dropout rate*, para balancear o fato de que existem mais camadas ativas na fase de teste do que na fase de treinamento (CHOLLET, 2018). Tecnicamente, o que está se fazendo é adicionar ruídos à saída das camadas para evitar que a rede decore casos não significativos.

Infelizmente, não há uma fórmula para determinar o número ideal de camadas ou o tamanho das camadas, sendo necessário avaliar diferentes arquiteturas de validação e testes para determinar o tamanho correto da rede; a melhor maneira de fazer isso é começar com uma rede pequena e avaliar a perda de validação em relação ao aumento da rede (CHOLLET, 2018).

2.3.3 Deep learning e LSTM

Dentre as técnicas identificadas na revisão de literatura apresentada na subseção 2.4, os estudos que obtiveram os melhores resultados utilizaram algoritmos LSTM que pertencem à arquitetura *Recurrent Neural Network* (RNN), utilizando o conceito de *deep learning*. Nos estudos clínicos apresentados na revisão, Lipton et al. (2016), Kam e Kim (2017) e Zhang (2017) utilizaram dados com séries temporais, o que justifica a utilização de RNN, tendo sido o LSTM empregado por resolver um dos maiores problemas das RNNs convencionais, o *vanishing gradient*.

A *deep learning* surgiu como uma nova área de pesquisa em ML, sendo desenvolvida para agir como um cérebro humano, com capacidade de aprender e resolver situações a partir de dados complexos. É uma nova visão sobre o processo de aprendizagem por meio de dados que enfatiza o aprendizado com sucessivas camadas de significativas representações (CHOLLET, 2018). Arquiteturas de *deep learning* como *Deep Neural Networks* (DNNs), *Convolutional Deep Neural Networks* (CNNs), *Deep Belief Networks* (DBNs) e RNNs estão sendo aplicadas em várias áreas, entre elas, médicas e de bioinformática (ALPAYDIN, 2014).

O processo de aprendizado normalmente envolve dezenas ou até mesmo centenas de sucessivas camadas de representação, todas elas aprendendo automaticamente por exposição aos dados de treinamento (CHOLLET, 2018). Uma das premissas da *deep learning* é substituir as configurações manuais por algoritmos mais eficientes para o aprendizado dos diferentes tipos de modelo (ALPAYDIN, 2014).

O LSTM, proposto por Hochreiter e Schmidhuber (1997), é uma RNN que possui a característica de gerar seus resultados a partir de todo o contexto histórico dos dados e não somente dos últimos valores de entrada. Sua arquitetura é composta por algumas células de estado que atuam como memórias longas e curtas, se autogerenciando por meio do aprendizado, de quando “lembrar” e quando “esquecer” a informação (HOCHREITER; SCHMIDHUBER, 1997). Sua unidade básica é um bloco de memória contendo uma ou mais células de memórias (*cells*) e três *adaptive gates* (*input*, *output* e *forget gates*), que são os responsáveis por controlar o fluxo dos dados, permitindo a entrada e a saída do dado no bloco de memória e, eventualmente, sua exclusão, prevenindo que o resto da rede altere o conteúdo da memória por longos períodos (GRAVES, 2012).

Ele foi desenvolvido para modelar sequências temporais de longo alcance com maior precisão do que as RNNs convencionais e ainda resolve sua principal fraqueza, conhecida na literatura como *vanishing gradient problem* (GRAVES et al., 2008), além de ter mostrado melhor desempenho no aprendizado de contextos livres e linguagem sensível ao contexto (SAK; SENIOR; BEAUFAYS, 2014). Resumidamente, o *vanishing gradient problem* representa a dificuldade de saber a direção em que os parâmetros precisam se mover para melhorar a função de custo (*loss function*) (GOODFELLOW; BENGIO; COURVILLE, 2015). Essa propriedade permite que a rede neural processe dados com uma interdependência complexa e separada, ou seja, esse comportamento possibilita vincular causas e efeitos que estão separados por vários passos de tempo, proporcionando à rede a habilidade de utilizar todo o contexto histórico na tarefa atual (KALCHBRENNER; DANIHELKA; GRAVES, 2016).

A limitação do LSTM dá-se por conta dos *gates*, *input* e *output* não terem nenhuma ligação com a unidade *Constant Error Carousel* (CEC), que supostamente é a unidade de controle. O resultado dessa falta de informação pode prejudicar o desempenho da rede, pois é função dos *gates* aprender quando mandar informação para dentro da rede (*forward pass*) ou não (*backward pass*). Essa limitação já foi discutida, e a solução proposta foi a utilização de *peephole* pelo mesmo autor (GERS; SCHMIDHUBER, 2000).

Além das RNNs, os *Hidden Markov Models* (HMM) são capazes de representar dependências temporais. Inicialmente descrito por Andrey Markov, em 1906, foi proposto um modelo probabilisticamente dependente de estados não observados. Entretanto, os modelos Markov tradicionais são limitados, pois, em algumas situações, seu tamanho cresce exponencialmente de acordo com o tamanho da janela temporal, tornando-os computacionalmente impraticáveis. Sendo assim, a principal diferença entre as RNNs e o HMM é que aquelas conseguem trabalhar com janelas temporais maiores (LIPTON; BERKOWITZ; ELKAN, 2015).

2.3.4 Otimização de hiperparâmetros

O processo de treinamento de um algoritmo tem o objetivo de gerar um modelo que tenha um bom desempenho na tarefa de prever ou classificar em dados nunca antes vistos (ZHENG, 2015). Para melhorar os resultados, é possível ajustar alguns

parâmetros (ou hiperparâmetros), em que o processo de ajuste chama-se otimização de hiperparâmetros (*hyperparameters optimization*) (CHOLLET, 2018).

As decisões iniciais, mesmo com uma boa intuição, são quase sempre sub-otimizadas; o método de refinamento ocorre por meio do ajuste manual e treinamento do modelo, sendo necessário realizar esse processo repetidamente. A necessidade de ajuste vai depender dos dados e do problema. Nesta subseção, são abordados os hiperparâmetros adotados como referência, baseados no estudo de Lipton et al. (2016), a saber: época, *loss function*, *weight decay* e *dropout*, utilizados no desenvolvimento do algoritmo LSTM.

A época (*epoch*) é a passagem de todo o conjunto de dados pela rede neural, sendo assim, o número de épocas determina quantas vezes o conjunto de dados deve passar pela rede. Se forem utilizadas poucas épocas, pode haver problemas de *underfitting* ou sub-ajuste, ou seja, o algoritmo não tem a possibilidade de aprender tudo que deveria com os dados de treinamento, e, se forem utilizadas muitas épocas, ocorre o oposto, *overfitting* ou sobreajuste, em que o modelo “decora” os dados de treinamento, tornando a generalização ineficiente (BUDUMA; LOCASCIO, 2017).

A avaliação do modelo é feita por meio da *loss function*, que é a diferença do valor predito para o valor real; quanto menor for o valor, menor será o erro. Sua função é mensurar o quão bem um modelo se comportou em determinada situação dado seu valor; esse valor é conhecido como *loss score* (CHOLLET, 2018). Existem diversos tipos de *loss function*, onde cada função tem uma aplicação específica. A Equação 1 representa o *mean squared error*, utilizado em problemas de regressão e classificação binária (CHOLLET, 2018).

$$Mse_{test} = \frac{1}{m} \sum_i (y'^{(test)} - y^{(test)})^2 \quad (1)$$

O ajuste por meio da *loss function* é fundamental, pois é possível usar *loss score*, resultado da função, como um sinal de resposta para acertar o valor dos pesos na rede, indicando sua direção para que a perda seja cada vez menor para o atual exemplo. Esse ajuste é trabalho do otimizador (*optimizer*), que é quem implementa o chamado algoritmo de *backpropagation* (CHOLLET, 2018).

Inicialmente, os pesos são atribuídos de forma aleatória. Naturalmente, os valores de saída produzem um *loss score* alto. Mas, a cada exemplo que a rede processa, os pesos são ajustados um pouco na direção correta, diminuindo o escore.

Essa repetição de treinamento é realizada quantas vezes forem necessárias (normalmente, dezenas de iterações sobre centenas de exemplos), para que os valores dos pesos diminuam o *loss score*. O resultado é uma rede treinada que gera os resultados mais próximos possíveis do objetivo previsto (CHOLLET, 2018).

Uma vez que o objetivo é diminuir a *loss function* para aumentar o acerto, Lipton et al. (2016) utilizaram um algoritmo de otimização chamado *Stochastic Gradient Descent* (SGD) para minimizar a perda. Quando um algoritmo de otimização é empregado, o custo computacional é proporcional ao tamanho do conjunto de dados. O SGD reduz o custo computacional a cada iteração, calculando o gradiente por meio de amostras. Isso significa que, na média, ele é um bom estimador de gradientes (ZHANG et al., 2018). Sua utilização é indicada para grandes conjuntos de dados para treinamento.

O *weight decay* é responsável por reduzir o tamanho dos pesos da rede, ao passo que o número de iterações aumenta (MARSLAND, 2015). O resultado prático é que a escolha dos pesos tende a ser menor para valores menores, proporcionando um controle maior das situações de sobreajuste e sub-ajuste por meio dos pesos da rede neural (GOODFELLOW; BENGIO; COURVILLE, 2015).

Por fim, Lipton et al. (2016) aplicaram o método de *dropout*, apresentado na subseção 2.3.2, cuja finalidade é “desligar” (atribuir com zeros) randomicamente determinado número de neurônios da rede durante o treinamento, com o objetivo de adicionar ruídos à saída das camadas para evitar que o algoritmo tenha um sobreajuste ao conjunto de treinamento, perdendo o poder de generalização do modelo (COPPIN, 2004).

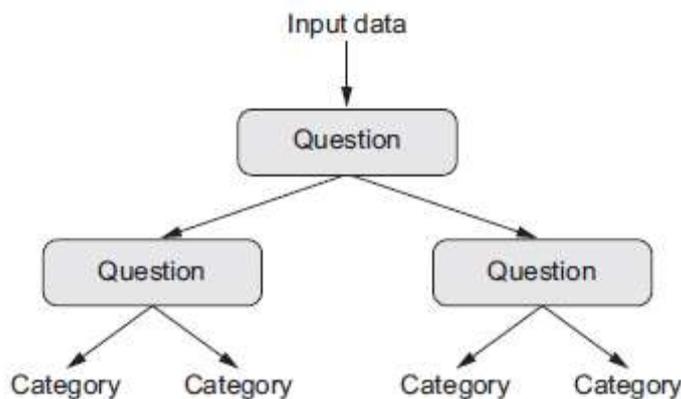
2.3.5 Random forest

A técnica de *random forest*, proposta por Breiman (2001), é aqui apresentada por ter sido o método que obteve o segundo melhor resultado de desempenho na revisão integrativa, com acurácia de 91,5%, sensibilidade de 94,4% e especificidade de 94,4% (LAMPING et al., 2018). Guillén et al. (2015) e Horng et al. (2017) compararam a técnica com outros algoritmos classificadores, como *Support Vector Machine* (SVM), regressão logística, entre outros; em ambos os estudos, a *random forest* apresentou melhor desempenho, se mostrando uma boa opção de segundo plano para um algoritmo de classificação não linear. Como exemplo de estudos que

utilizaram o *random forest* na saúde, destaca-se Taylor et al. (2016), Horng et al. (2017) e Lamping et al. (2018), aqui já citados.

As árvores de decisão são estruturas no formato de fluxogramas, conforme figura 5, que permitem a classificação dos dados fornecidos para prever o resultado de futuras entradas de dados (CHOLLET, 2018). Os parâmetros aprendidos são os nós, que são perguntas (*questions*) sobre os dados, como, por exemplo, se determinado valor é menor ou maior que cinco; a resposta é a categoria (*category*) em que o exemplo se enquadra.

Figura 5 – Árvore de decisão – representação do modelo de classificação.



Fonte: Chollet (2018, p.16).

A *random forest* é um tipo de árvore de decisão que foi desenvolvido a partir da combinação de *bagging*, proposto por Breiman (2001), com a seleção aleatória de atributos para a construção de uma coleção de árvores com grau de variação controlada (ALPAYDIN, 2014). A ideia básica é: se uma árvore é boa, então uma *forest* (floresta) é melhor, ainda mais se houver uma variedade de tipos de árvore entre elas. A técnica destaca-se pela forma com a qual é criada a aleatoriedade a partir de um conjunto de dados padrões; primeiramente, utiliza-se o método de *bagging*, que resumidamente é criar árvores diferentes a partir de distintas amostras do dado. O passo seguinte é adicionar a aleatoriedade para limitar as escolhas que as árvores de decisão podem fazer (MARSLAND, 2015).

Em cada nó, um subconjunto aleatório de atributos é dado à árvore, e esta só pode escolher um subconjunto, em vez de todo o conjunto. Esse processo, além de aumentar a aleatoriedade no treino de cada árvore, reduzindo a variação e o efeito de vieses, aumenta a velocidade do treinamento, pois, em cada estágio, há um conjunto menor de atributos para pesquisar (MARSLAND, 2015).

Na sua implementação, um dos principais parâmetros é o número de árvores (*trees*) na floresta (*forest*); entretanto, é fácil de achar o valor ideal, uma vez que é só aumentar o número de árvores até o valor do erro parar de aumentar (MARSLAND, 2015). Uma vez que o conjunto de árvores é treinado, o resultado é composto pela maioria dos exemplos do conjunto do treinamento (MARSLAND, 2015). A regressão logística é relativamente robusta (ou insensível) à configuração de hiperparâmetros, sendo necessário identificar quais atributos e hiperparâmetros são úteis para determinado problema (ZHENG; CASARI, 2018).

2.3.6 Regressão logística (*logistic regression*)

A regressão logística obteve o terceiro lugar na classificação de desempenho de classificadores na revisão integrativa; seu melhor desempenho foi apresentado no estudo de Guillén et al. (2015), com a acurácia de 87,7%, sensibilidade de 60,3% e especificidade de 93,2%, se mostrando uma boa opção, neste caso, linear.

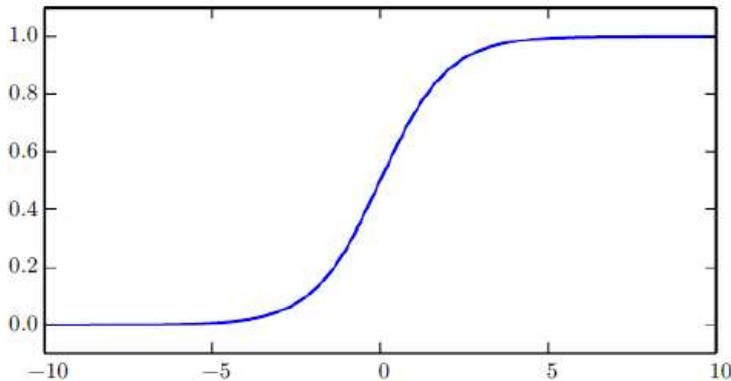
Também conhecida como *logistic regression*, *logit regression* e *logit model*, é um modelo probabilístico que foi desenvolvido por D. R. Cox, em 1958, sendo o modelo logístico utilizado para prever uma resposta binária baseada em um ou mais variáveis preditoras (atributos) (ALPAYDIN, 2014).

As probabilidades que descrevem os possíveis resultados de um único ensaio são modeladas como variáveis explicativas ou preditoras. A regressão logística é frequentemente empregada para se referir especificamente a problemas cuja variável é binária, ou seja, o número de possíveis categorias são duas; quando o problema tem mais de duas categorias, é referenciado como *multinomial logistic regression* (ALPAYDIN, 2014).

A regressão logística mede a relação entre a variável dependente categórica e uma ou mais variáveis independentes, que são contínuas e categóricas, estimando as probabilidades. Assim, trata o mesmo conjunto de problemas com a função de *probit regression*, que é a transformação do resultado pela função *logit* inversa (ALPAYDIN, 2014).

A classificação é realizada em uma combinação baseada em pesos dos atributos de entrada, passando por uma função sigmoide, Figura 6, que tem o objetivo de transformar o valor em um número real no intervalo de zero e um (ZHENG; CASARI, 2018).

Figura 6 – Gráfico de representação da função sigmoide.



Fonte: Goodfellow, Bengio e Courville (2015, p. 84).

As suposições feitas pela regressão logística sobre a distribuição e os relacionamentos em seus dados são muito parecidas com aquelas feitas na regressão linear (ALPAYDIN, 2014).

É preciso ter em mente algumas características para o uso da regressão logística, como, por exemplo, a função não levar em conta os erros no conjunto de treinamento, fazendo-se necessária a remoção de ruídos e *outliers* para diminuir o risco de alguma classificação errada. A regressão logística é um algoritmo linear (com uma transformação não linear na saída). A função assume uma relação linear entre as variáveis de entrada e de saída, sendo interessante realizar transformações de dados (funções como *log*, *root*, *Box-Cox*) nas suas variáveis de entrada que melhor exponham esse relacionamento linear; esse procedimento pode resultar em um modelo mais preciso (ALPAYDIN, 2014).

2.4 TRABALHOS RELACIONADOS

Com o objetivo de identificar as técnicas de IA adotadas para reconhecimento de sepse em ambientes hospitalares, uma revisão integrativa foi desenvolvida, resultando no artigo *Técnicas de inteligência artificial para reconhecimento de sepse em ambientes hospitalares: revisão integrativa*, submetido e aceito em 27 de agosto de 2019 pela Revista de Gestão em Sistemas de Saúde (*Health Systems Management Journal*). O artigo na sua íntegra está no Apêndice A.

3 MAPEAMENTO ENTRE O REFERENCIAL TEÓRICO E OS ENCAMINHAMENTOS METODOLÓGICOS

Nesta seção, são apresentadas as referências que apoiam a definição dos encaminhamentos metodológicos desta dissertação (Quadro 4).

Quadro 4 – Mapeamento entre o referencial teórico e os encaminhamentos metodológicos.

Referência	Conceito	Contribuição	Objetivo específico
Henkin et al. (2009)	Sepse	Conceituação, critérios e definições na identificação da sepsse	Construção do modelo
Kaukonen et al. (2015)	Critérios de identificação de sepsse	Levantamento dos critérios e definições na identificação da sepsse	Construção do modelo
Vincent, Moreno e Takala (1996) Singer et al. (2016) MacDonald et al. (2017)	Descrição dos escores utilizados para avaliação e identificação dos casos de sepsse	Utilização de escores (SOFA, qSOFA) para identificação dos pacientes com sepsse	Construção do modelo
Lipton et al. (2016) Kam e Kim (2017) Zhang et al. (2017) Mao et al. (2018)	Utilização de técnicas de ML para identificação de sepsse	Utilização do LSTM para predição de diagnóstico clínico	Construção do modelo
Lipton et al. (2016)	Definição de parâmetros para utilização do LSTM	Utilização dos parâmetros como <i>baseline</i>	Construção do modelo
Zheng e Casari (2018)	FE para ML	Utilização de técnica de <i>wrapper</i> de FE na seleção dos atributos do projeto	Construção do modelo
Guillén et al. (2015) Taylor et al. (2016) Horng et al. (2017) Lamping et al. (2018)	Utilização de técnicas de ML para identificação de sepsse	Utilização de técnicas como <i>random forest</i> e regressão logística	Construção do modelo
Westphal e Lino (2015)	Estratégias de identificação precoce em diferentes setores do hospital	Utilização de escores para definição dos critérios de identificação da sepsse	Construção do modelo
Lipton et al. (2016) Nemati et al. (2018)	Utilização dos critérios de SOFA no treinamento de um algoritmo de ML na identificação precoce de sepsse	Atributos e parâmetros utilizados para o treinamento do algoritmo	Construção do modelo
Schuh (2007)	Utilização de redes neurais para identificação de sepsse	Utilização de atributos e parâmetros no desenvolvimento da pesquisa	Construção e avaliação do modelo
Taylor et al. (2016)	Utilização de regressão logística na previsão de mortalidade e comparação dos resultados obtidos	Metodologia de avaliação dos resultados obtidos	Avaliação do modelo
Yoshihara et al. (2011)	Análise descritiva de pacientes com sepsse	Exemplo de análises uni e multivariada para diferentes escores utilizados	Construção e avaliação do modelo

Mao et al. (2018)	Desenvolvimento de algoritmo de reconhecimento de sepse utilizando apenas sinais vitais	Influência do tempo e da perda de dados no reconhecimento de sepse e utilização de apenas sinais vitais no reconhecimento	Construção e avaliação do modelo
Mitchell et al. (2016)	Reconhecimento de sepse em pacientes de enfermaria	Estudo apresentou as diferenças no reconhecimento de sepse entre os pacientes de UTI e enfermarias, utilizando regressão logística	Construção e avaliação do modelo
Horng et al. (2017)	Reconhecimento de sepse no setor de triagem	Estudo utilizou textos abertos, além dos sinais vitais e dados demográficos, para identificar pacientes com suspeita de infecção	Criação do método
Amland e Hahn-Cover (2016)	Reconhecimento precoce de sepse	Utilização de escores para definição dos critérios de identificação da sepse	Construção e avaliação do modelo

Fonte: O autor (2019).

4 ENCAMINHAMENTOS METODOLÓGICOS

Este trabalho tem como objetivo desenvolver uma ferramenta para identificação dos casos de risco de sepse nos hospitais que são gerenciados pelo GSUS do governo do estado do Paraná. Trata-se de uma pesquisa descritiva, retrospectiva e de abordagem quantitativa. Descritiva, pois buscou explorar a relação e associações entre os atributos; retrospectiva, pois explorou fatos que já ocorreram e os participantes foram selecionados a partir do desfecho; e quantitativa pois foi realizada a análise de dados e o pesquisador não teve influência na pesquisa (DYNIEWICZ, 2014).

Para o desenvolvimento deste estudo, foram utilizados os dados do GSUS do governo do estado do Paraná, sistema que gerencia os dados dos prontuários eletrônicos dos pacientes atendidos nos hospitais próprios do estado, conforme descrição na próxima subseção.

As etapas metodológicas são detalhadas na seção 4.2.

4.1 CENÁRIO DA PESQUISA

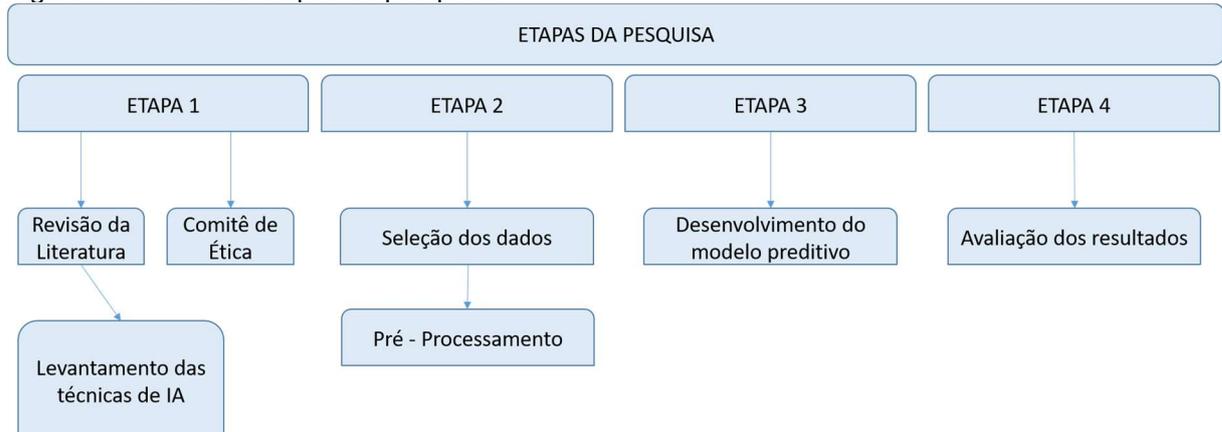
O cenário da pesquisa foi baseado nos hospitais e estabelecimentos de saúde atendidos pelo GSUS. Em 2019, o sistema atende a 11 hospitais, 15 ambulatórios, 11 laboratórios, 372 farmácias e 15 prontos-socorros.

Lembra-se de que a instalação do GSUS nas instituições é modular; a implantação dos módulos depende, entre outros fatores, da infraestrutura do local, sistemas pré-instalados, legados e, principalmente, do corpo técnico e clínico.

4.2 ETAPAS DO PROJETO

O projeto foi organizado em quatro etapas (Figura 7), descritas em mais detalhes nas subseções a seguir, de forma a deixar mais claro o método adotado.

Figura 7 – Fluxo das etapas da pesquisa.



Fonte: O autor (2018).

4.2.1 Etapa 1 – revisão de literatura e Comitê de Ética

A primeira etapa consistiu na realização de uma revisão integrativa para identificação da eficiência das técnicas de IA adotadas para reconhecimento de sepse em ambientes hospitalares, descrita na subseção 2.4. O artigo, intitulado *Técnicas de inteligência artificial para reconhecimento de sepse em ambientes hospitalares: revisão integrativa*, foi escrito pelo autor, submetido e aceito em 27 de agosto de 2019 pela Revista de Gestão em Sistemas de Saúde (*Health Systems Management Journal*).

Esta pesquisa foi aprovada pelo Comitê de Ética em Pesquisa da Pontifícia Universidade Católica do Paraná (PUCPR), sob Parecer nº 13818519.9.0000.0020 (Anexo A), e pelo Comitê de Ética em Pesquisa Envolvendo Seres Humanos CEP-SESA/PR do Centro de Estudos, Pesquisa e Desenvolvimento Humano do Hospital de Trabalhador (CEPDH-HT), sob o Parecer nº 13818519.9.3001.5225 (Anexo B).

4.2.2 Etapa 2 – seleção dos dados e pré-processamento

A segunda etapa foi composta por duas atividades: (i) seleção dos dados; (ii) pré-processamento desses dados.

Os dados de pacientes e internamentos foram selecionados conforme os critérios: pacientes internados e maiores de 18 anos. Os atributos foram selecionados por possuírem a maior relevância preditiva, conforme os estudos de Guillén et al. (2015), Kaukonen et al. (2015), Calvert et al. (2016), Lipton et al. (2016), Singer et al. (2016), Horng et al. (2017), MacDonald et al. (2017), Taylor et al. (2016), Garcia-Gallo et al. (2018), Lamping et al. (2018), Mao et al. (2018), apresentados no Quadro 5.

Quadro 5 – Atributos de pacientes e internamentos selecionados a partir de sua relevância preditiva.

Cesta de atributos iniciais	
Local de internamento (UTI ou enfermaria)	Frequência cardíaca
Idade	Frequência respiratória
Sexo	Pressão sanguínea
Raça	Temperatura
Administração de antibiótico	Concentração de lactato
Contagem de leucócitos	Data de internamento
Oxigenação do sangue	Classificação Estatística Internacional de Doenças (CID)
Procedimento cirúrgico – cinco dias	Evolução – palavras relacionadas à infecção

Fonte: O autor (2019).

Os exames selecionados possuem relevância preditiva, conforme apresentado nos trabalhos de Seymour et al. (2017) e Garcia-Gallo et al. (2018), quais sejam: cultura de sangue, lactato e bilirrubina.

Os dados dos pacientes e internamentos foram selecionados com a finalidade de criar o conjunto de dados final para treinamento, validação e testes do algoritmo. Para essa etapa, foi necessário associar os dados relativos ao paciente, como idade, sexo e raça, e os dados de internamento, como sinais vitais, exames, identificação de infecções, identificação de procedimentos cirúrgicos e prescrição de antibióticos.

O critério de identificação de sepse no internamento foi pelo cadastro do código CID A41 ao A41.9, relacionado à septicemia, nos campos: CID primário, CID secundário e CID óbito, conforme Quadro 6.

Quadro 6 – Descrição dos códigos da família A41 – Outras septicemias, conforme CID10.

A41 – Outras septicemias
A410 – Septicemia por <i>Staphylococcus aureus</i>
A411 – Septicemia por outros estafilococos especificados
A412 – Septicemia por estafilococos não especificados
A413 – Septicemia por <i>Haemophilus influenzae</i>
A414 – Septicemia por anaeróbios
A415 – Septicemia por outros microrganismos gram-negativos
A419 – Septicemia não especificada

Fonte: DATASUS (2018).

As referências utilizadas neste trabalho, como Garcia-Gallo et al. (2018), utilizaram como critério de seleção apenas registros que tivessem pelo menos 70% de preenchimento. A abordagem empregada por Lipton et al. (2016) agrupou os registros por hora, com o objetivo de ter pelo menos um registro por hora contendo todos os valores dos atributos de sinais vitais. Para este estudo, foi adotada a

abordagem de Garcia-Gallo et al. (2018), sendo possível adequar o percentual de preenchimento de acordo com a frequência em que os dados dos pacientes foram registrados. A abordagem de Lipton et al. (2016) parece ser mais adequada para os casos em que o registro dos dados dos pacientes é mais frequente, como, por exemplo, em situações em que os registros são feitos a cada minuto e nos casos em que é necessário reduzir a dimensionalidade dos dados.

A atividade 2, de pré-processamento, tem o objetivo de uniformizar os dados para obter o melhor resultado esperado com o modelo proposto, assim como a remoção dos ruídos e coleta de informações necessárias para decisão das estratégias para tratamento dos dados faltantes (FAYYAD; PIATETSKY-SHAPIRO; SMYTH, 1996; KAM; KIM, 2017).

Essa atividade é composta por seis processos, sendo:

- a) Agrupamento dos dados clínicos e laboratoriais dos pacientes.
- b) Identificação dos procedimentos clínicos e cirúrgicos.
- c) Identificação das prescrições de medicamentos contendo antibióticos.
- d) Identificação de termos clínicos em textos abertos que remetem a infecções.
- e) Preenchimento dos dados faltantes.
- f) Criação da série temporal.

O processo A, referente ao agrupamento, foi responsável por associar as informações de pacientes e internamentos com os resultados dos exames laboratoriais. Em bases como MIMIIC, utilizadas por vários estudos, como em Guillén et al. (2015), Henry et al. (2015), Calvert et al. (2016), Ghosh et al. (2017), Jiang et al. (2016), Mitchell et al. (2016), Horng et al. (2017) e Kam e Kim (2017), as coletas de dados são mais frequentes, permitindo o agrupamento em intervalos de tempo variando de minutos a horas. Neste estudo, os dados dos pacientes foram agrupados por paciente e internamento por data, conforme referência (GARCIA-GALLO et al., 2018). A situação ideal era que, para cada dia de internamento, existissem pelo menos um registro dos sinais vitais do paciente e um resultado de exame para a mesma data; caso houvesse um resultado de exame sem um registro de sinal vital para a mesma data, ele foi associado ao registro de sinais vitais mais próximo, com o limite de um dia. Se houver um resultado de exame laboratorial sem o registro de sinais vitais por mais de dois dias, os resultados seriam descartados.

No processo B, os tipos de procedimento precisaram ser classificados em procedimentos clínicos ou cirúrgicos, pois todo atendimento hospitalar é vinculado a um grupo de procedimento; essa classificação foi necessária, pois o paciente que é submetido a procedimentos cirúrgicos pode apresentar uma vulnerabilidade maior decorrente desse tipo de procedimento. O código do procedimento pode ser alterado ao longo do internamento, então foi utilizado o último código de procedimento cadastrado, que é o vigente e utilizado para fins de faturamento junto ao SUS.

No processo C, foram identificadas as prescrições contendo antibióticos. A inadequação da abordagem inicial ao agente infeccioso por meio da utilização de antibióticos está intimamente relacionada à mortalidade de pacientes com sepse, assim como existem evidências claras de que a demora no início da antibioticoterapia também aumenta o risco de morte (ILAS, 2017). Segundo Henkin et al. (2009), o tratamento com antibióticos deve ser iniciado o mais precocemente possível, visando ao controle do foco de infecção como pré-requisito para a eliminação do agressor para possibilitar a recuperação do paciente (SALOMÃO et al., 2011). Assim, pode-se utilizar a informação sobre a administração de antibióticos no decorrer do internamento, com o objetivo de avaliar se existe alguma relação entre a administração de antibióticos e o desfecho, além de possibilitar outras análises, como, nos casos identificados como sepse, se foi realizada a antibioticoterapia ou não.

No processo D, foi realizada a identificação de palavras que remetem a infecções, pois a identificação de focos de infecção no paciente faz parte do processo de reconhecimento de pacientes em risco de sepse; essa, segundo o ILAS (2017), é uma das maiores dificuldades, pois nem sempre é clara a presença do foco infeccioso. Horng et al. (2017) utilizaram texto aberto do setor de triagem de um hospital para melhorar a identificação de pacientes com suspeita de infecções no departamento de emergência, obtendo um ganho de aproximadamente 19% na identificação de infecções.

Para realizar a identificação, foi necessário buscar palavras que indicassem algum tipo de infecção em cada evolução ou descrição registrada no sistema. Nesse sentido, foram verificados os campos: Histórico de Moléstia Atual (HMA), preenchido na recepção do paciente pelo profissional da triagem; Evolução, em que consta a evolução diária do paciente, preenchido pelo médico responsável; e Exame Físico, preenchido diariamente pelos profissionais de enfermagem. As palavras-chave foram selecionadas por um profissional da área de infecção hospitalar, devendo indicar

processos de infecção; foram elas: ascite, bronquite, nefrite, peritonite, meningite, apendicite, cistite, celulite, pneumonia, sepse, septicemia, sepsis, osteomielite, conjuntivite, sinusite, otite, gengivite, laringite, faringite, endocardite, gastroenterite, erisipela, amigdalite, pielonefrite, mastoetite, abscesso, pericardite, endometrite, colecistite, pancreatite, diverticulite, colite, mastite e anexite.

Para o processo de identificação das palavras-chaves, foi utilizada a função *stringdist*, com o método de Damerau-Levenshtein do pacote *stringdist* do R (DER LOO et al., 2019), que tem o objetivo de calcular a distância entre vetores de caracteres ou entre vetores que representam sequências genéricas. Para flexibilizar um pouco o processo de identificação, foi aplicado um grau de distanciamento igual a dois, ou seja, seriam necessárias duas alterações para a palavra identificada no texto se tornar a palavra procurada; assim, palavras escritas com no máximo dois caracteres errados ainda foram identificadas.

Esse método foi empregado por levar em conta o peso das inserções e deleções junto às substituições necessárias para converter uma palavra em outra no cálculo da distância entre as palavras ou na indicação do nível de semelhança das palavras mais próximas. Para Der Loo (2014), o processo de *string match* é comumente dividido em duas situações: a primeira consiste na criação de um índice de busca, criado a partir do pré-processamento realizado pela aplicação; e a segunda é a procura de texto por aproximação, estando o interesse na localização da palavra de correspondência mais próxima em um texto, podendo a palavra ou conjunto de palavras procuradas ser uma tabela ou um dicionário (DER LOO, 2014).

No processo E, foi realizado o preenchimento dos dados faltantes referentes aos sinais vitais, quais sejam: temperatura, glicemia, pressão arterial sistólica e pressão arterial diastólica, frequência cardíaca e frequência respiratória, conforme proposto por Lipton et al. (2016) e Kam e Kim (2017); a proposta era preencher os dados faltantes com os valores próximos. Foram utilizados apenas os registros que tiverem um percentual de preenchimento adequado à qualidade dos dados, conforme o seguinte critério: Nos casos em que faltaram até duas leituras diárias, foi usada a mediana entre os dois valores mais próximos, um dia antes e um dia depois; na falta de duas a três leituras, foram utilizados os valores normais clínicos. Para os dados faltantes de temperatura, foi usada a tabela de Kelly (2006), em que se obteve a temperatura padrão por sexo, idade e hora do dia; para os dados faltantes de glicemia, foram utilizados os dados do Consenso Brasileiro sobre Diabetes (2000); para os

dados de frequência cardíaca, foram utilizados os dados de Paschoal et al. (2006); para os dados faltantes de frequência respiratória, os dados de Parreira et al. (2010); para os dados faltantes de pressão arterial sistólica e diastólica, os dados da 7ª Diretriz Brasileira de Hipertensão Arterial (MALACHIAS et al., 2016). No entanto, sabe-se que esses procedimentos não são apropriados para todas as situações (LIPTON et al., 2016).

No processo F, foi realizada a criação da série temporal, uma vez que o algoritmo LSTM utilizado para implementação faz uso de séries temporais para as previsões. Lipton et al. (2016) criaram sua série temporal usando os dados dos pacientes, com as informações das primeiras e últimas seis horas. Essa opção teve melhores resultados do que utilizando as últimas 12 horas, sendo assim, para cada evento classificado, foi agrupada a série histórica de 12 horas (*timesteps*). Para este estudo, foi utilizado o mesmo princípio, onde o *timestep* utilizado foi oito (dias), esse é o valor da mediana de dias de internamento do grupo identificado com o CID de sepse, conforme apresentado na Tabela 7. Os internamentos com oito ou mais dias correspondem a 81% do total de internamentos.

A série temporal é construída a partir de cada data de identificação da sepse, deste ponto, coleta-se os dados dos oito dias anteriores. O conjunto de dados final é o histórico dos oito dias que antecederam cada evento de sepse.

4.2.3 Etapa 3 – implementação do algoritmo LSTM

A terceira etapa foi composta por uma atividade, consistindo na implementação de um algoritmo que focasse na tarefa de previsão a partir de dados em uma série temporal. Segundo Rumelhart, Hinton e Williams (1986), a RNN é a melhor opção, pois esses modelos possuem uma memória interna que permite que a rede se “lembre” das séries de informações, tornando-as adequadas para realizar a tarefa de previsão (ZHANG et al., 2017).

A proposta deste trabalho é implementar uma rede LSTM, pois se mostrou resiliente à falta de dados, por isso uma de suas aplicações é justamente o diagnóstico médico (LIPTON et al., 2016). Em seu trabalho, Lipton et al. (2016) compararam o desempenho do LSTM com fortes referências de base, como redes *Multilayer Perceptron* (MLP) e classificadores lineares para classificação de 429 códigos de diagnósticos, utilizando 13 atributos de pacientes (sinais vitais, exames de sangue,

entre outros). Seu desempenho superou os resultados do MLP e dos classificadores lineares, mesmo o MLP sendo treinado e ajustado especificamente para esse caso e o LSTM tendo utilizado os dados sem tratamento (LIPTON et al., 2016). A única limitação computacional encontrada foi com relação ao desempenho (GERS; SCHMIDHUBER, 2000), tendo sido discutida e proposta uma solução por meio da utilização de *peephole*.

O modelo implementa *deep learning* a partir de RNN para classificar se determinados pacientes estão ou não em risco de sepse; para tal, utilizou-se como conjunto de dados de treinamento uma base selecionada conforme os códigos CID A41 cadastrados nos campos CID óbito, CID principal e CID secundário do GSUS e os critérios SOFA e qSOFA.

Os parâmetros iniciais de configuração foram os mesmos utilizados por Lipton et al. (2016), tendo sido o LSTM empregado para classificar 128 diagnósticos clínicos, pois tanto o conjunto de dados utilizado no projeto de Lipton et al. (2016) quanto neste estudo apresentam características semelhantes, tratando-se de séries temporais clínicas com conjunto de dados altamente desbalanceados.

4.2.4 Etapa 4 – avaliação dos resultados

A avaliação dos resultados foi realizada por meio dos resultados de acurácia, sensibilidade, especificidade e curva ROC, conforme apresentado em Sokolova e Lapalme (2009), como sendo as medidas para avaliação de classificações binárias. Os termos utilizados na descrição da sensibilidade, especificidade e acurácia foram: verdadeiro positivo (VP), verdadeiro negativo (VN), falso positivo (FP) e falso negativo (FN). Se determinada doença está presente no paciente e o teste de diagnóstico indica sua presença, então o resultado do teste é considerado VP; se o paciente não apresenta a doença e o teste sugere também a ausência da doença, então é considerado VN; ambos, VP e VN, sugerem um resultado consistente entre o teste de diagnóstico e a situação do paciente. Entretanto, se o teste médico apresenta a ausência da doença e o teste de diagnóstico sugere a presença, então é considerado FP; se o paciente apresenta a doença e no teste de diagnóstico é sugerida a ausência, então é considerado FN (ZHU; ZENG; WANG, 2010).

Para melhor compreensão dos conceitos de sensibilidade, especificidade e acurácia, o Quadro 7 apresenta os termos utilizados para o teste de diagnóstico.

Quadro 7 – Termos utilizados na definição da sensibilidade e especificidade em um teste de diagnóstico.

Resultado do teste	Doença (referência)	
	Presente	Ausente
Positivo	VP	FP
Negativo	FN	VN

Fonte: Adaptado de Zhu, Zeng e Wang (2010).

A sensibilidade (Equação 2) ou *recall* avalia como o modelo se comporta em detectar os casos positivos; em outras palavras, ela representa, em termos numéricos, a probabilidade de um teste de diagnóstico identificar pacientes que realmente têm determinada doença.

$$\text{Sensibilidade} = \frac{VP}{VP+FN} \quad (2)$$

A especificidade (Equação 3) é a proporção de TN corretamente identificados; representa a probabilidade de um teste diagnosticar uma doença específica sem fornecer resultados FP.

$$\text{Especificidade} = \frac{VN}{VN+FP} \quad (3)$$

Os valores de sensibilidade e especificidade produzem uma medida que consegue descrever melhor a qualidade do classificador a partir da relação entre exemplos classificados como VP, FP e FN. Essa medida é a métrica de avaliação sugerida por Sokolova e Lapalme (2009) para classificadores binários.

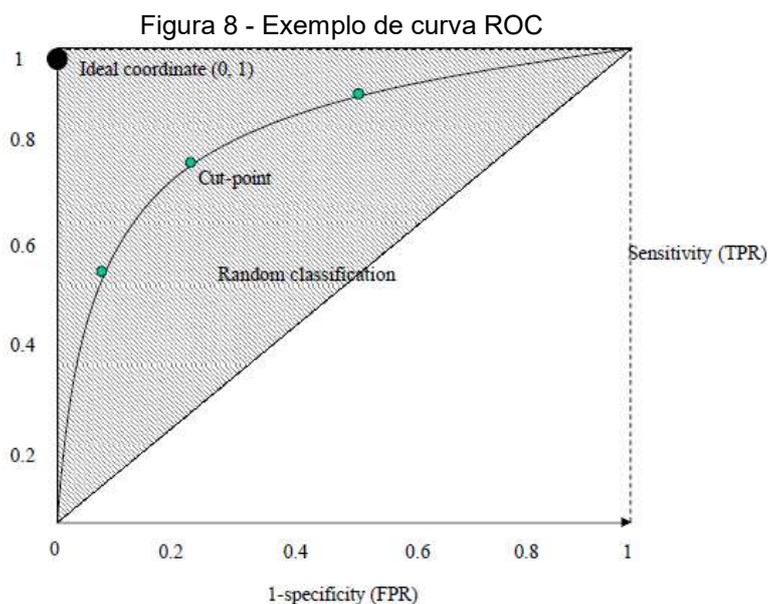
A acurácia (Equação 4) é a proporção de resultados VP ou VN em uma população. Mede o grau de veracidade de um teste diagnóstico em uma condição.

$$\text{Acurácia} = \frac{VN + VP}{VN+VP+FN+FP} \quad (4)$$

Além de avaliar o desempenho do modelo, essas medidas permitem a comparação com os resultados achados na literatura por meio da revisão integrativa (PARIKH et al., 2008; ASHAL et al., 2015).

A curva ROC foi inicialmente desenvolvida no contexto de detecção de sinais eletrônicos. Durante a Segunda Guerra Mundial, tinha o objetivo de quantificar a habilidade dos operadores dos radares em distinguir um sinal de um ruído. A partir da década de 1960, foram utilizadas em psicologia experimental e, a partir dos anos 1970, passaram a ser usadas em vários ramos da pesquisa bioética; na área de terapia intensiva, diversos estudos têm utilizado essa metodologia para avaliação da confiabilidade de modelos preditivos (ZANON et al., 2008). Ela permite expressar graficamente a relação entre a sensibilidade e a especificidade, sendo de fundamental importância quando os testes de diagnóstico produzem uma resposta sob forma de uma variável categórica (HOSMER; LEMESHOW, 2000).

A construção da curva ROC é realizada colocando os valores da sensibilidade, referente à proporção de VP (ou TPR), no eixo Y e o complemento da especificidade, referente à proporção de FP (ou FPR), no eixo X, para diferentes pontos de corte. Sua interpretação dá-se pela observação da curvatura – quanto mais próxima a curva estiver do canto superior esquerdo do gráfico, melhor será o poder discriminatório do teste de diagnóstico e, quanto mais distante estiver, pior será o poder de discriminação entre doentes e não doentes (SOKOLOVA; LAPALME, 2009; ZHU; ZENG; WANG, 2010). A Figura 8 apresenta um exemplo de curva ROC onde as possíveis combinações de sensibilidade e especificidade compõe um espaço ROC, assim, a localização do ponto em um espaço ROC, demonstra se a classificação diagnóstica é boa ou não.



Fonte: (Zhu, Zeng, & Wang, 2010)

Uma das formas de identificação do melhor ponto de corte é o índice de Youden, que é a máxima distância vertical da linha de igualdade (diagonal central) para o ponto $[x, y]$, em que x representa a especificidade e y , a sensibilidade. O ponto encontrado nesse critério é o mais distante da diagonal central, sendo a soma da sensibilidade e da especificidade máxima (PERKINS; SCHISTERMAN, 2006).

Define-se o melhor modelo como aqueles resultados que apresentam os maiores valores percentuais de acurácia, sensibilidade e especificidade. É apresentada também a matriz de confusão dos modelos treinados, com o objetivo de entender melhor os resultados em nível de classe. Além disso, utilizando essas métricas, é possível comparar o desempenho do modelo com os escores utilizados como padrão ouro, SOFA e qSOFA, para identificação e pontuação da gravidade do paciente com sepse.

4.2.5 Protocolo de testes

Tendo em vista o objetivo desta dissertação, o principal resultado esperado é um algoritmo que tenha a possibilidade de identificar precocemente pacientes com risco de sepse, tendo sido, além da proposta do desenvolvimento do modelo, realizada uma revisão integrativa na intenção de identificar os métodos mais eficientes para a identificação de pacientes em risco de sepse. Com base nesse levantamento, os modelos sugeridos para testes foram LSTM e *random forest*.

Para o primeiro teste, foi utilizado o LSTM. Para os parâmetros iniciais, foram adotados os valores utilizados no trabalho desenvolvido por Lipton et al. (2016) e Garcia-Gallo et al. (2018), devido às semelhanças nas características e objetivos do trabalho. Para o ajuste inicial do treinamento, foram usadas cem épocas por *fold* e *batchsize* de 32, com *dropout* de 0,5. O otimizador utilizado inicialmente foi o “rmsprop”, pois é indicado como uma das melhores recomendações para RNNs. Nos testes, foi aplicado o mesmo conjunto de dados, porém a identificação de sepse foi feita de três maneiras diferentes.

Como primeiro critério de identificação de sepse, foram considerados os campos de CID (principal, secundário ou óbito); como segundo critério, os critérios do escore SOFA; e, como terceiro critério, os critérios de identificação do escore qSOFA, ambos descritos na subseção 2.1. Apesar de o método de *cross-validation* ser um dos mais utilizados para avaliar os algoritmos de classificação e regressão, quando se

trata de séries temporais, não existe um padrão determinado, pois em algumas situações pode-se ter a sensação de estar utilizando dados do futuro para prever o passado (BERGMEIR; HYNDMANB; KOO, 2018). Além disso, a correlação serial nos dados, com possíveis não estacionaridades, faz com que o uso de *cross-validation* pareça problemático, pois não leva em conta esses problemas.

Ainda segundo Bergmeir, Hyndmanb e Koo (2018), os profissionais recorrem à técnica *out-of-sample*, em que uma seção do fim da série de dados é retida para avaliação. Dessa forma, apenas uma avaliação em um conjunto de testes é considerada, enquanto, com o uso da validação cruzada, várias avaliações são realizadas (BERGMEIR; BENÍTEZ, 2012). Sendo assim, para este estudo foram utilizados 70% da base para treinamento, 10% para validação e 20% para testes.

Para o segundo teste, foi utilizado a *random forest*, com o conjunto de dados utilizando o critério qSOFA. O algoritmo foi escolhido por ser o classificador que obteve os melhores resultados de desempenho na revisão integrativa.

Essa estratégia de testes está diretamente associada ao fato de a base estar desbalanceada; neste caso, o treinamento foi realizado por meio da validação cruzada, o que permitiu que toda a base fosse utilizada como treino e teste. Com essa estratégia, foi possível obter como avaliação uma média do desempenho do algoritmo.

5 RESULTADOS

Nesta seção, são apresentados os resultados dos experimentos conduzidos com os algoritmos LSTM e *random forest* das atividades relacionadas às etapas 2, 3 e 4 do método proposto para o desenvolvimento da ferramenta.

Para a seleção de dados (atividade 1 da etapa 2), foi necessário desenvolver um *data warehouse* para tratamento dos dados, uma vez que tanto a lógica de funcionamento do GSUS quanto as especificidades dos hospitais são gerenciadas pelo sistema por meio de tabelas de configuração e, em algumas situações, tratadas via implementação; além disso, existe a restrição de concorrência para a base de dados de produção. Assim, foi necessário implementar uma coleta incremental com janelas de acesso curtas. A ferramenta utilizada para a construção do *datawarehouse* foi o QlikView, versão 12.10 SR6. O processo de criação do *datawarehouse* contou com 38 rotinas que tiveram o objetivo de coletar, tratar e armazenar 19.7Gb de dados que resultou em um conjunto de dados final contendo 4.331 pacientes, 4.810 internamentos.

O *data warehouse* criado (Figura 9) tem como principal evento a internação do paciente, que difere um pouco do objetivo do sistema, que é gerenciar os serviços de assistência da SESA. O modelo desenvolvido contém 32 tabelas e trata de módulos como internamento, laboratório, farmácia, atendimento, prescrição, leitos e procedimentos.

Tabela 4 – Número total de pacientes maiores de 18 anos internados por ano, segundo hospitais.

Hospital	2011	2012	2013	2014	2015	2016	2017	2018	Total
H1	22	30	16	23	20	11	12	14	148
H2	3.053	5.585	5.487	5.713	6.147	6.300	6.670	6.865	45.820
H3	5.132	5.028	5.248	5.669	6.169	6.710	6.486	6.596	47.038
H4	-	127	1.241	1.969	3.029	5.149	7.568	7.795	26.878
H5	-	61	209	419	579	693	776	688	3.425
H6	-	-	-	228	347	404	468	439	1.886
H7	-	-	-	536	507	470	891	817	3.221
H8	-	-	-	-	-	408	3.620	4.499	8.527
H9	-	-	-	-	-	224	237	216	677
H10	-	-	-	-	-	-	800	13.803	14.603
H11	-	-	-	-	-	-	-	554	554
Total	8.207	10.831	12.201	14.557	16.798	20.369	27.528	42.286	152.777

Fonte: O autor (2019).

Na Tabela 5, são apresentadas as quantidades totais de internamentos dos hospitais do estado.

Tabela 5 – Número total de internamentos por ano, segundo hospitais.

Hospital	2011	2012	2013	2014	2015	2016	2017	2018	Total
H1	3.240	3.731	3.615	3.822	3.602	3.223	3.185	2.919	27.337
H2	4.437	8.345	8.232	8.273	8.846	9.390	9.525	10.334	67.382
H3	7.563	7.469	7.732	8.208	8.724	9.425	10.068	9.944	69.133
H4	-	149	1.487	2.685	4.209	7.984	12.458	12.858	41.830
H5	-	68	274	527	751	995	1.049	901	4.565
H6	-	-	-	239	366	434	509	465	2.013
H7	-	-	-	627	642	597	1.086	976	3.928
H8	-	-	-	-	-	634	6.280	7.684	14.598
H9	-	-	-	-	-	287	308	263	858
H10	-	-	-	-	-	-	1.027	18.891	19.918
H11	-	-	-	-	-	-	-	684	684
Total	15.240	19.762	21.340	24.381	27.140	32.969	45.495	65.919	252.246

Fonte: O autor (2019).

Na Tabela 6, são apresentadas as quantidades totais de internamentos levando em conta apenas os pacientes internados maiores de 18 anos.

Tabela 6 – Número total de pacientes maiores de 18 anos internados por ano, segundo hospitais.

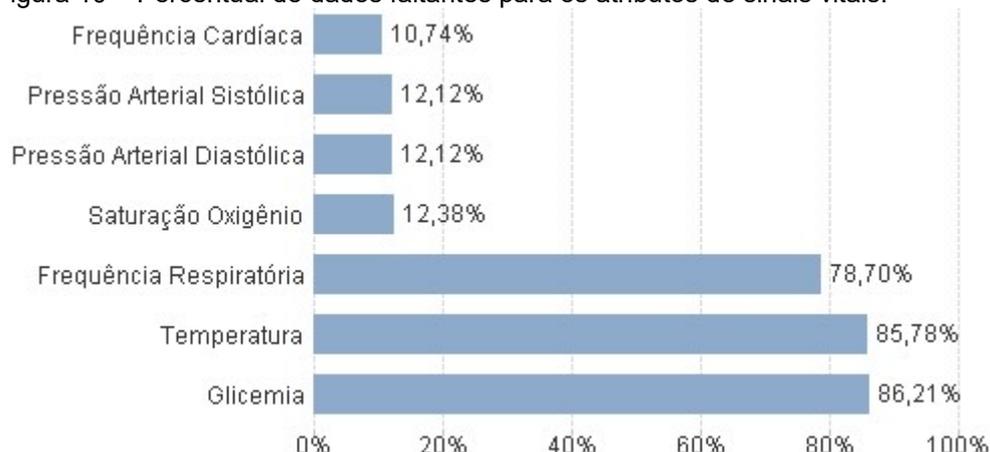
Hospital	2011	2012	2013	2014	2015	2016	2017	2018	Total
H1	23	33	17	25	28	13	14	15	168
H2	3.445	6.417	6.317	6.583	7.052	7.390	7.685	8.015	52.904
H3	6.106	5.986	6.141	6.661	7.118	7.769	7.547	7.642	54.970
H4	-	128	1.378	2.365	3.714	6.016	8.845	8.984	31.430
H5	-	65	241	457	661	892	971	830	4.117
H6	-	-	-	239	366	434	509	465	2.013
H7	-	-	-	619	616	583	1.061	958	3.837
H8	-	-	-	-	-	446	4.521	5.510	10.477
H9	-	-	-	-	-	287	307	261	855
H10	-	-	-	-	-	-	827	15.841	16.668
H11	-	-	-	-	-	-	-	672	672
Total	9.574	12.629	14.094	16.949	19.555	23.830	32.287	49.193	178.111

Fonte: O autor (2019).

O hospital escolhido para análise foi o H3, por apresentar maior quantidade de pacientes e internamentos, tendo-se optado por trabalhar com os dados de 2017 e 2018, por haver mais registros e por estarem mais completos, além do fato de ser o cenário mais próximo da situação atual do hospital.

Os atributos de sinais vitais disponíveis no GSUS são: temperatura, frequência respiratória, frequência cardíaca, pressão arterial sistólica, pressão arterial diastólica, pressão arterial média, glicemia, saturação de oxigênio, pressão intra-abdominal, pressão intracraniana e pressão venosa central. A Figura 10 apresenta o percentual de dados faltantes para cada um dos atributos de sinais vitais para o hospital H3.

Figura 10 – Percentual de dados faltantes para os atributos de sinais vitais.



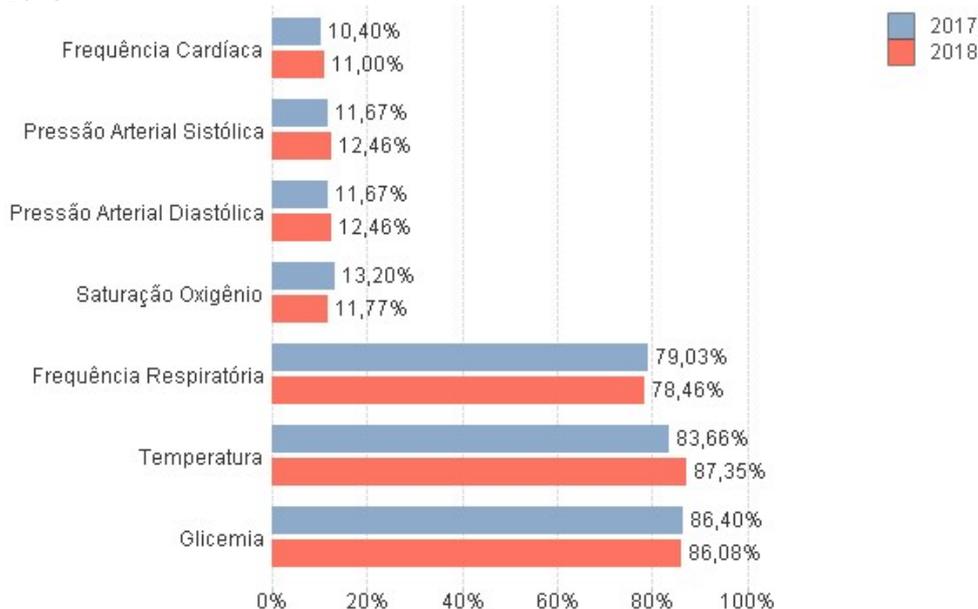
Fonte: O autor (2019).

Considerando os dados dos últimos anos, o hospital H3 apresentou melhores percentuais de preenchimento, porém ainda possui uma grande quantidade de dados faltantes, contendo 94,79% dos registros de sinais vitais incompletos. Para este

estudo, foi adotado o critério de seleção dos registros contendo pelo menos 50% de preenchimento de dados relacionados aos sinais vitais, pois esse percentual contempla 87% do conjunto total de dados.

Na Figura 11 são apresentados os atributos pressão arterial média, pressão venosa central, pressão intracraniana e pressão intra-abdominal que foram retirados das análises por apresentarem 100% de dados ausentes. No intuito de analisar o problema de dados faltantes, na Figura 11, são apresentadas as frequências dos dados faltantes para os anos de 2017 e 2018 respectivamente. É possível observar aumento de dados faltantes em todos atributos, menos na frequência respiratória sinalizando a necessidade de alguma ferramenta que permita o acompanhamento na qualidade do registro dos dados.

Figura 11 – Percentual de dados faltantes para os atributos de sinais vitais para os anos de 2017 e 2018.



Fonte: O autor (2019).

Aplicando os critérios de seleção propostos na subseção 4.2.2 – pacientes internados e maiores de 18 anos –, obteve-se um conjunto de dados de 55.590 registros. Os casos de sepse foram identificados por meio dos códigos CID relacionados à septicemia, pertencentes à família CID A41, nos campos: CID principal, CID secundário e CID óbito, em que foram identificados 164 internamentos (3,52% do total de internamentos). Na Tabela 7 é apresentado o perfil geral dos pacientes e o perfil dos pacientes com CID de sepse do hospital H3. Os exames selecionados são os considerados pelo escore SOFA, a saber: bilirrubina, lactato, coagulação e

creatinina. Para facilitar o entendimento dos próximos passos, esse conjunto de dados será tratado por “conjunto de dados inicial”.

Tabela 7 – Perfil dos internamentos nos grupos com e sem CID de sepse do hospital H3.

Variável	Todos Pacientes	Com CID A41
Número de pacientes	4.331	163
Internamentos	4.810	164
Sepse		
Sim		164 (3,52%)
Não	4.646 (96,47%)	
Sexo		
Masculino	2.421 (50,33%)	91 (55,48%)
Feminino	2.389 (49,66%)	73 (44,51%)
Idade		
Mínimo	18	21
Máximo	106	95
Média	50	66
Mediana	50	69
Cor		
Branca	3.820 (79,41%)	132 (80,48%)
Preta	259 (5,38%)	8 (4,87%)
Parda	688 (14,30%)	23 (14,02%)
Amarela	20 (0,41%)	1 (0,60%)
Indígena	13 (0,27%)	0
Sem informação	10 (0,20%)	0
Infecção		
Sim	2.983 (62,01%)	134 (81,70%)
Não	1.827 (37,98%)	30 (18,29%)
Antibióticos		
Sim	118 (2,45%)	12 (7,31%)
Não	4.692 (97,54%)	152 (92,68%)
Dias de internamento		
Média	5,93	14,2
Mediana	3	8
Máximo	203	191
Internamento		
Cirúrgico	1.762 (36,16%)	19 (11,58%)
Clínico	3.105 (63,73%)	145 (88,41%)
Diagnóstico	5 (0,10%)	0 (0,00%)

Fonte: O autor (2018).

Segundo o ILAS (2018), em seu estudo *spread* ainda não publicado, aproximadamente 30% dos leitos das UTIs brasileiras estão ocupados com pacientes com sepse; então, os 3,52% encontrados podem indicar uma subnotificação. Na intenção de identificar outros possíveis casos de risco de sepse, optou-se por utilizar os critérios do escore SOFA e qSOFA no conjunto de dados denominado “conjunto de dados inicial”; o critério SIRS não foi elencado, pois, segundo Gupta et al. (2018), as evidências sugerem que o qSOFA é um melhor critério de diagnóstico.

Aplicando os critérios qSOFA propostos por Singer et al. (2016) para seleção de pacientes, a saber: taxa de respiração maior que 22 por minuto e pressão sistólica

menor que 100 Hg, foram identificadas 216 (4,49%) internações para os dois anos. Para efeito de comparação, foram selecionados apenas os internamentos que continham taxa de respiração maior que 22, resultando em 337 internamentos; para pressão sistólica menor que 100, foram identificados 713 internamentos, indicando a possibilidade de existir mais casos de risco de sepse e, quando aplicado o processo de *input* dos valores normais para frequência cardíaca e respiratória, pressão e temperatura, esses possíveis casos se perdem. O escore qSOFA também considera escala de coma de Glasgow menor que 15, porém foi constatado que esse dado não é registrado com frequência no GSUS.

Aplicando os critérios SOFA, foram identificados 178 internamentos, equivalentes a 3,70% do número total; vale ressaltar que os critérios do escore SOFA, apresentados no Quadro 1, levam em consideração exames laboratoriais, como exame de sangue, bilirrubina e creatinina. Na Tabela 8 é apresentado a quantidade de internamentos identificados por cada critério e suas respectivas quantidades de óbitos.

Tabela 8 – Quantidade de internamentos com suspeita de sepse identificados por meio dos critérios GSUS, qSOFA e SOFA e suas respectivas quantidades de óbitos.

Critério	Internamentos	%	Óbitos	%
GSUS	164	3,52	48	29,26
qSOFA	216	4,49	37	17,88
SOFA	178	3,70	17	9,55

Fonte: O autor (2019).

O coeficiente Kappa proposto por Jacob Cohen em 1960 (KOSH; LANDIS, 1977) tem a finalidade de medir o grau de concordância entre proporções derivadas de amostras dependentes (Fleiss & Cohen, 1973), e é baseado no número de respostas concordantes, permitindo avaliar se a concordância está além do esperado tão somente pelo acaso e o grau dessa concordância.

O coeficiente Kappa é calculado por:

$$\text{Kappa} = \frac{P(O) - P(E)}{1 - P(E)} \quad (5)$$

Onde, P(O) é a proporção observada de concordâncias (soma das respostas concordantes dividida pelo total) e P(E) é a proporção esperada de concordâncias (soma dos valores esperados das respostas concordantes dividida pelo total). Portanto o Kappa foi utilizado para medir o grau de concordância entre as seleções

realizadas pelos três escores, conforme Tabela 9, obtendo o melhor índice de concordância entre os escores SOFA e qSOFA, que segundo a tabela de concordância de Landis, indica uma concordância “substancial”, confirmando que a seleção foi realizada mediante critérios.

Tabela 9 - Coeficientes Kappa para avaliação entre os escores

Escores	Kappa	Concordância
GSUS e qSOFA	0,68	Substancial
GSUS e SOFA	0,35	Razoável
SOFA e qSOFA	0,79	Substancial

Fonte: O autor (2019).

Vale ressaltar que nem todos os exames laboratoriais avaliados pelo escore SOFA estão presentes em todos os internamentos, o que pode ter influência no índice calculado.

Além dos atributos iniciais, foram criados mais dois: tempo de internação (em dias), que é a diferença entre a data de internamento e a data do evento (exame ou coleta de sinais vitais), e idade no internamento, que é a diferença entre a data de nascimento e a data do internamento.

Para o processo de identificação das palavras-chave referentes a infecções, utilizou-se a função *stringdist*, com o método Damerau-Levenshtein do R (DER LOO et al., 2019), que tem o objetivo de calcular a distância entre vetores de caracteres ou entre vetores que representam sequências genéricas, com grau de distanciamento igual a dois, ou seja, seriam necessárias duas alterações para a palavra identificada no texto se tornar a palavra procurada. Esse método foi aplicado por levar em conta o peso das inserções e deleções junto às substituições necessárias para converter uma palavra em outra no cálculo da distância entre as palavras ou na indicação do nível de semelhança das palavras mais próximas. Para Der Loo (2014), o processo de *string match* é comumente dividido em duas situações: a primeira consiste na criação de um índice de busca, criado a partir do pré-processamento realizado pela aplicação, e a segunda é a procura de texto por aproximação, estando o interesse na localização da palavra de correspondência mais próxima em um texto, podendo a palavra ou conjunto de palavras procuradas ser uma tabela ou um dicionário. Nesse processamento, foram identificadas 697 evoluções com as palavras referentes à infecção (listadas na seção 4.2.2), aproximadamente 14,49% do total eventos do hospital H3.

5.1 RESULTADO DA ETAPA 4 – DESENVOLVIMENTO DO MODELO PREDITIVO

Os parâmetros iniciais de configuração foram os mesmos utilizados por Lipton et al. (2016), pois tanto os dados de seu projeto quanto deste apresentam características semelhantes, tratando-se de séries temporais clínicas com conjunto de dados altamente desbalanceados.

Inicialmente os hiperparâmetros do LSTM foram ajustados como proposto em Lipton et al. (2016), porém foi utilizada a técnica de *grid-search* para encontrar possíveis hiperparâmetros que resultassem em previsões mais precisas possíveis (BERGSTRÄ; BENGIO, 2012; ZHANG et al., 2018). Os parâmetros utilizados para o processo de *Grid-Search* foram os indicados por Brownlee, (2017) e Sutskever, Vinyals e Le (2014), sendo esses recomendados para utilização em séries temporais. Na Tabela 10 são apresentados os parâmetros utilizados por Lipton et al., (2016) e os encontrados pela técnica de *grid-search*. A rede final utilizou três *hidden layers*, com 128 células de memórias na primeira camada e 64 nas duas seguintes, com um *dropout* de 0,5 para cada. A função de ativação utilizada foi a sigmoide, por se tratar de um problema binário (CHOLLET, 2018).

Tabela 10 – Parâmetros propostos e identificados para implementação do LSTM a partir do *grid-search*.

Parâmetros	LIPTON et al., (2016)	Grid-search
Camadas ocultas	2	3
Neurônios	64	128 e 64
<i>Dropout rate</i>	0.5	0.5
<i>Weight decay</i>	10^{-6}	10^{-8}
Épocas	100	200
<i>Batchsize</i>	-	32
Learning rate	-	0.3
<i>Momentum</i>	-	0.5
Otimizador	SGD	SGD
<i>Loss function</i>	Log Loss	MSE
Ativação	Sigmoide	Sigmoide

Fonte: O autor (2019).

A *loss function* utilizada foi o *mean squared error* (MSE), pois, apesar de ser mais indicado para problemas de regressão, obteve melhores resultados que *binary_crossentropy*, que normalmente é a melhor recomendação para problemas binários (CHOLLET, 2018). Para a função de ativação da camada de saída, foi usada a função sigmoide, indicada para essa situação, pois ela varia entre zero e um.

O valor de *timestep* utilizado para criação das séries temporais foi oito, pois esse é o valor da mediana do número de dias de internamento para as internações com CID A41, conforme Tabela 7, ou seja, para cada evento identificado como risco de sepse, tem-se uma série histórica de sete eventos anteriores.

Para o desenvolvimento, foi utilizada a linguagem R versão 3.6.0 e Python versão 3.7.3, além das bibliotecas Keras, Tensor Flow e Sci-Kit Learn, com o *framework* RStudio versão 1.2.1335, Build 1379 e Spyder.

Foram criadas três séries temporais a partir do “conjunto de dados inicial”, para treinamento dos modelos LSTM. A primeira série temporal foi criada utilizando o critério de identificação do CID A41 no GSUS, obtendo um conjunto de dados de 1.148 registros; a segunda foi originada com os critérios do escore qSOFA, obtendo um conjunto de dados de 1.599 registros; e a terceira foi criada usando os critérios do escore SOFA, obtendo um conjunto de dados de 1.476 registros. Para identificação dos casos de sepse para os critérios qSOFA e SOFA foi seguido a definição Sepsis-3 (Singer et al., 2016), ou seja, 2 ou mais pontos associados a identificação de uma infecção.

Três modelos LSTM foram treinados; para facilitar a identificação, são chamados M-qSOFA, M-SOFA e M-GSUS. O modelo M-qSOFA foi treinado com o conjunto de dados obtido pelo critério qSOFA; M-SOFA, com o conjunto de dados obtido por meio do critério SOFA; e M-GSUS, com o conjunto de dados obtido pelo critério de identificação dos casos de sepse por meio do CID. As bases foram divididas em 70% para treinamento, 10% para validação e 20% para testes.

Os resultados de acurácia, sensibilidade, especificidade de cada modelo são apresentados na Tabela 11, ordenados pelo melhor resultado de acurácia, sensibilidade e especificidade.

Tabela 11 – Acurácia, sensibilidade e especificidade obtidos a partir das bases criadas por meio dos critérios GSUS, SOFA e qSOFA, intervalo de confiança de 95%.

Modelo	Acurácia	Sensibilidade	Especificidade
M-qSOFA	87,2%	87,7%	86,9%
M-SOFA	68,7%	56,3%	76,8%
M-GSUS	31,6%	8,4%	75%

Fonte: O autor (2019).

As matrizes de confusão são apresentadas individualmente, conforme teste executado utilizando as bases criadas por meio dos critérios qSOFA, SOFA e GSUS, respectivamente (Tabelas 12 a 14).

Tabela 12 – Matriz de confusão obtida para o modelo M-qSOFA, treinado com a série temporal qSOFA.

	VP	FP
FN	157	38
VN	22	253

Fonte: O autor (2019).

Tabela 13 – Matriz de confusão obtida para o modelo M-SOFA, treinado com a série temporal SOFA.

	VP	FP
FN	98	62
VN	76	206

Fonte: O autor (2019).

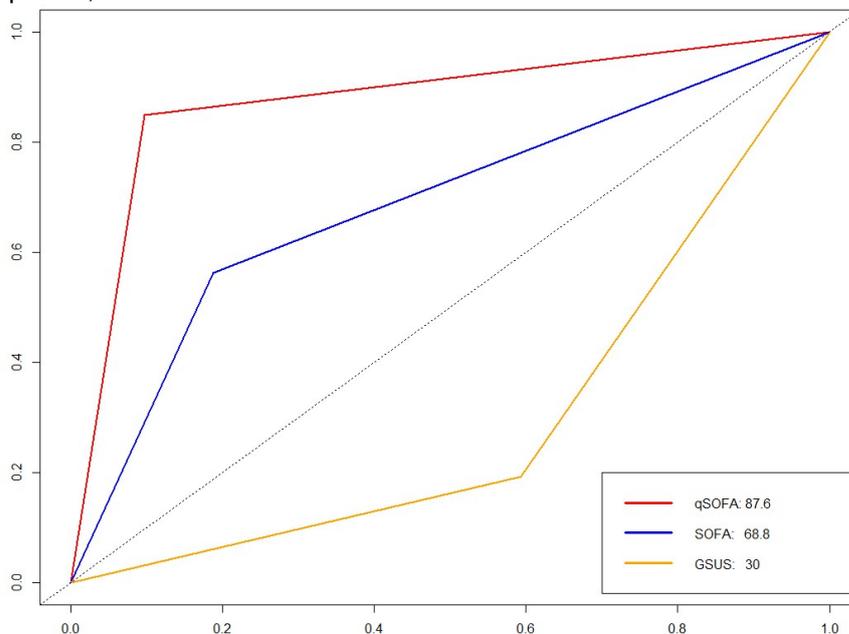
Tabela 14 – Matriz de confusão obtida para o modelo M-GSUS, treinado com a série temporal GSUS.

	VP	FP
FN	19	30
VN	205	90

Fonte: O autor (2019).

Para melhor visualizar e comparar as curvas ROC, optou-se por mostrar as três curvas obtidas no mesmo gráfico, utilizando o índice Youden disponível no pacote pROC da linguagem R, que apresenta os valores de sensibilidade e especificidade no ponto máximo, obtendo a máxima distância vertical da linha de igualdade (diagonal central) (Figura 12). A curva ROC obtida pelo modelo qSOFA foi de 87,6%, a do modelo SOFA foi de 68,8% e a do modelo GSUS foi de 30%

Figura 12 – Curvas ROC obtidas do algoritmo LSTM para as bases criadas por meio dos critérios qSOFA, SOFA e GSUS.



Fonte: O autor (2019).

Utilizando o modelo que obteve o melhor resultado de desempenho, M-qSOFA, foi realizado um segundo treinamento do algoritmo, desta vez alterando a quantidade de *timesteps* de previsão, para cada evento classificado como sepse; os valores de

timesteps foram testados para oito, cinco, três e um dia a partir da data do evento de suspeita de sepse, obtendo os valores de acurácia, sensibilidade e especificidade demonstrados na Tabela 15.

Tabela 15 – Valores de acurácia, sensibilidade e especificidade obtidos por meio do treinamento do LSTM utilizando *timesteps* de 8, 5, 3 e 1 dia de internamento, com intervalo de confiança de 95%.

Modelo	<i>Timesteps</i>	Acurácia	Sensibilidade	Especificidade
M-qSOFA	8	87,2%	87,7%	86,9%
	5	89,6%	89,9%	93,6%
	3	90,7%	91,2%	97,1%
	1	91,3%	92,2%	98,9%

Fonte: O autor (2019).

Verifica-se que, como esperado, o modelo tem um desempenho melhor quanto mais próximo do evento de suspeita de sepse, porém o teste com oito dias de antecedência também apresentou bons resultados, permitindo o acompanhamento dos casos que evoluíram gradativamente, seguindo o conceito de séries temporais, em que se obtém o máximo de valores históricos para cada valor a ser classificado (BONTEMPI; TAIEB; BORGNE, 2013).

Para testar o poder de generalização, o modelo M-qSOFA foi testado com os dados do hospital H2 para o mesmo período. O hospital H2 foi escolhido por ser o segundo hospital com melhor percentual de preenchimento de dados, porém tem um perfil de atendimento diferente. Os resultados obtidos estão apresentados na Tabela 16.

Tabela 16 – Comparativo com o teste do modelo M-qSOFA utilizando os dados do hospital H2, com intervalo de confiança de 95%.

Hospital	Acurácia	Sensibilidade	Especificidade
H3	87,5%	87,7%	93,5%
H2	82,8%	80,4%	94,1%

Fonte: O autor (2019).

Para o segundo teste, o algoritmo *random forest* foi implementado utilizando o conjunto de dados denominado “conjunto de dados inicial”, contendo 55.590 registros, onde foi aplicado o critério de identificação de pacientes com risco de sepse a partir do escore qSOFA, identificando 216 (4,49%) internamentos. O algoritmo foi treinado com os valores de *ntree* = 100 e *mtry* = 2.

Os resultados de acurácia, sensibilidade, especificidade, valor preditivo positivo e valor preditivo negativo do algoritmo são apresentados na Tabela 17.

Tabela 17 – Resultados obtidos de acurácia, sensibilidade, especificidade, valor preditivo positivo e valor preditivo negativo, utilizando o algoritmo *random forest*, com intervalo de confiança de 95%.

Modelo	Acurácia	Sensibilidade	Especificidade
<i>Random forest</i>	97,1%	61,1%	99,8%

Fonte: O autor (2019).

Os resultados obtidos pela *random forest* não são necessariamente uma surpresa, lembrando que aproximadamente 95% dos registros tinham dados ausentes e foram preenchidos por critérios preestabelecidos, como a mediana de seus valores próximos ou os valores normais clínicos de referência. Apesar de apresentar melhores valores de acurácia e especificidade, o valor de sensibilidade caiu aproximadamente 26,6%. A matriz de confusão é apresentada na Tabela 18.

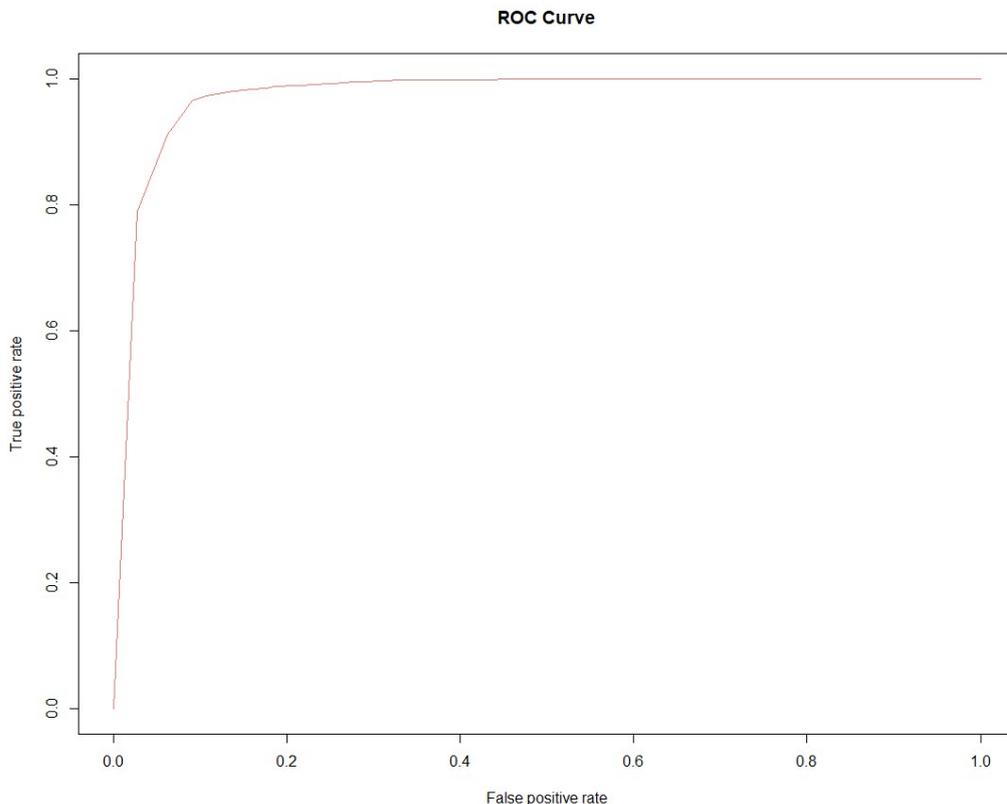
Tabela 18 – Matriz de confusão para o algoritmo *random forest*.

	VP	FP
FN	2.352	115
VN	1.497	51.626

Fonte: O autor (2019).

A curva ROC obtida pelo treinamento do algoritmo é mostrada na Figura 14.

Figura 13 – Curva ROC obtida por meio do treinamento do algoritmo *random forest*.

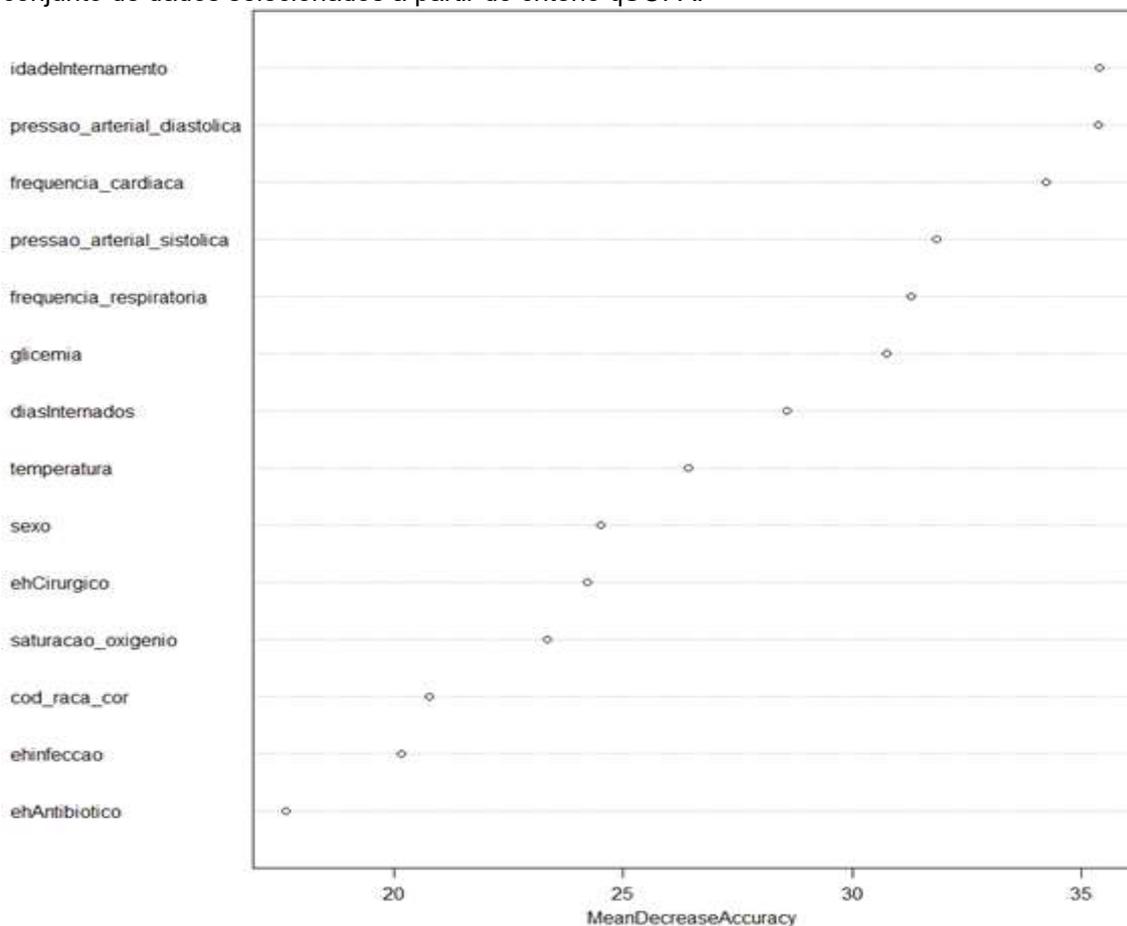


Fonte: O autor (2019).

As medidas de importância dos atributos para o treinamento supervisionado são fundamentais para a melhora do desempenho e da interpretação do aprendizado, dizendo o quão importante é cada variável para a classificação do dado. O grau de importância é dado pela diferença do erro de classificação média de todas as árvores e normalizado pelo desvio padrão dessas diferenças. Quando dois ou vários atributos têm importância semelhante, é preferível aquele com uma definição mais simples ou mais amplamente utilizada (LIAW; WIENER, 2002).

A Figura 15 apresenta os resultados da classificação dos 14 atributos presentes no conjunto de dados, tendo sido a avaliação realizada pelo método *mean decrease accuracy*, em que o erro é calculado pela média da redução da precisão de determinada variável; sendo assim, quanto mais a precisão diminui devido à exclusão de uma única variável, mais importante ela é e, conseqüentemente, as variáveis que causam uma grande redução na média de precisão são as mais importantes (LIAW; WIENER, 2002).

Figura 14 – Gráfico de importância *mean decrease accuracy*, referente aos atributos preditores para o conjunto de dados selecionados a partir do critério qSOFA.

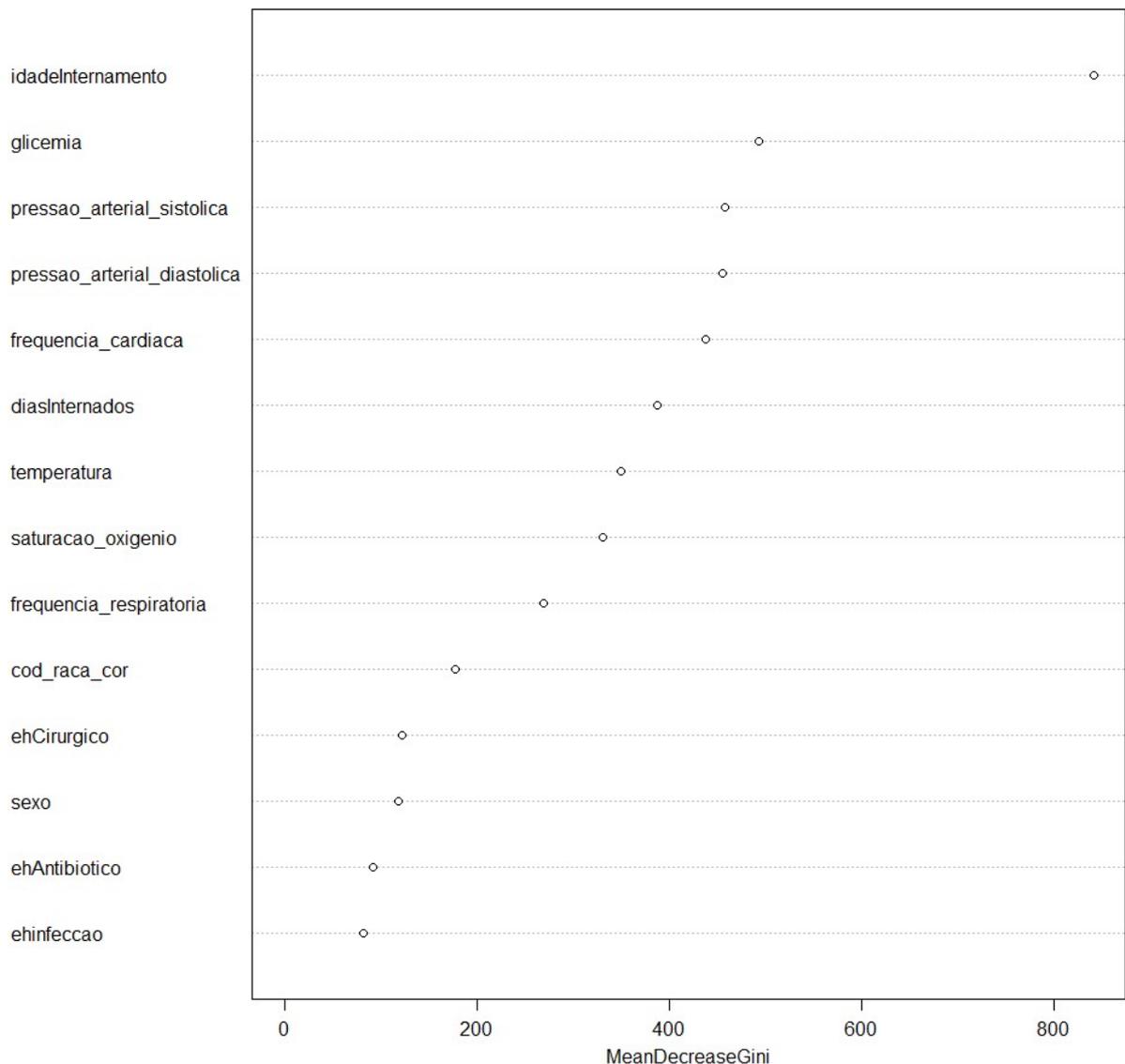


Fonte: O autor (2019).

A classificação de importância dos atributos encontrados neste estudo está de acordo com os encontrados na literatura, como em Garcia-Gallo et al. (2018), em que, por exemplo, a idade do paciente é o atributo com maior grau de importância, seguido de temperatura e frequência cardíaca.

A Figura 16 traz o gráfico relativo ao grau de importância de cada variável pelo método *mean decrease* GINI, cujo coeficiente é a medida de quanto cada variável contribui para a homogeneidade dos nós no resultado final. Assim, cada vez que determinada variável é utilizada para dividir um nó, os coeficientes dos novos nós são calculados e comparados com o nó original (LIAW; WIENER, 2002).

Figura 15 – Gráfico de importância *mean decrease* GINI, referente aos atributos preditores para o conjunto de dados selecionados a partir do critério qSOFA.

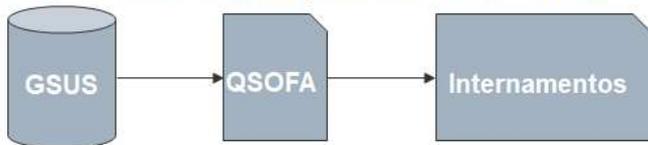


Fonte: O autor (2019).

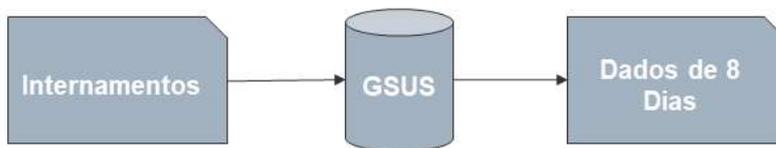
O modelo foi implantado em um ambiente de desenvolvimento virtualizado, no qual foi possível executar todos os processos, desde a seleção e tratamento dos dados até a disponibilização do modelo, sendo eles treinamento ou predição. A Figura 15 apresenta todas as etapas e processos do modelo proposto.

Figura 16 – Processo de tratamento dos dados, treinamento e predição do modelo proposto.

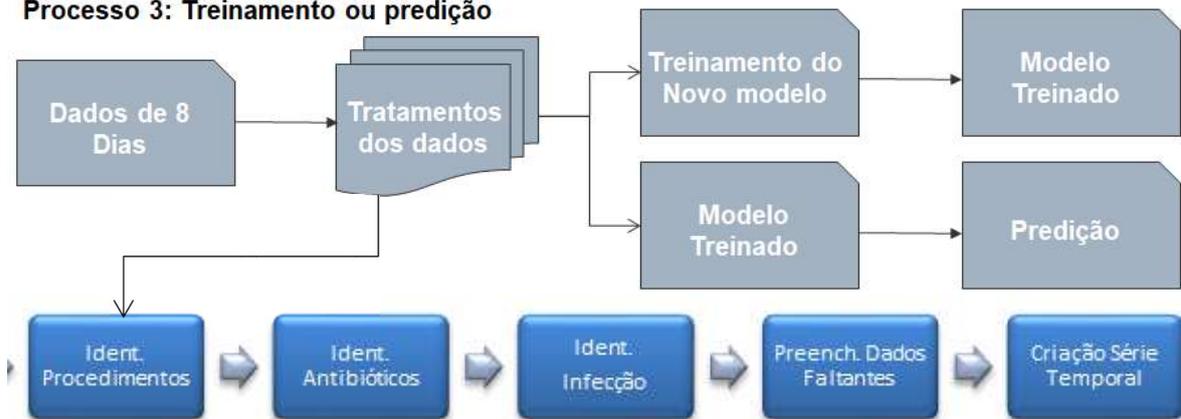
Processo 1: Identificar potenciais internamentos com risco de Sepsis



Processo 2: Recuperação das informações dos últimos 8 dias de internamento



Processo 3: Treinamento ou predição



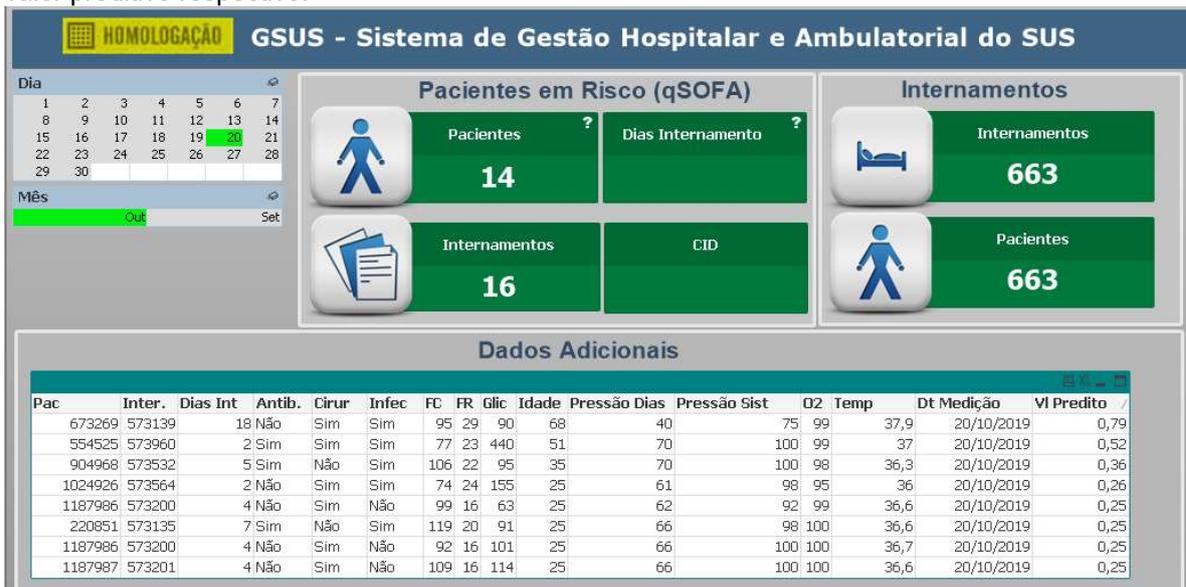
Fonte: O autor (2019).

Por restrições de *hardware*, na primeira implementação do modelo proposto, optou-se por trabalhar apenas com os pacientes que apresentaram os critérios do escore qSOFA (conforme processo 1), porém para um trabalho futuro, o objetivo deverá ser preparar o ambiente para trabalhar a ferramenta para a funcionalidade de *screening*, ou seja, testar no modelo desenvolvido todos os pacientes internados, diariamente, para identificar precocemente qualquer valor preditivo que possa indicar risco para o paciente.

A segunda etapa da implementação consistiu em desenvolver um aplicativo que utiliza o modelo criado ou treinado, o objetivo deste aplicativo é apresentar os pacientes e seus respectivos dados hospitalares, onde, na última coluna é

apresentado o “Valor Predito” (VI Predito) que é a probabilidade, ou o risco, calculado pelo modelo treinado do paciente estar com sepse. A Figura 18 apresenta informações como a quantidade de pacientes selecionados por meio do critério estabelecido, quantidade de internamento e seus respectivos dados hospitalares, na figura abaixo é apresentado para a data 16 internamentos, onde é mostrado apenas os oito maiores valores preditos.

Figura 17 – Tela de identificação dos pacientes que satisfazem os critérios do escore qSOFA e seu valor preditivo respectivo.



Fonte: O autor (2019).

A Figura 19 apresenta a seleção de apenas um paciente, o que apresentou o maior valor preditivo; nessa situação, a aplicação apresentou os valores de cada dia de internação, neste caso, os últimos oito dias (*timesteps*).

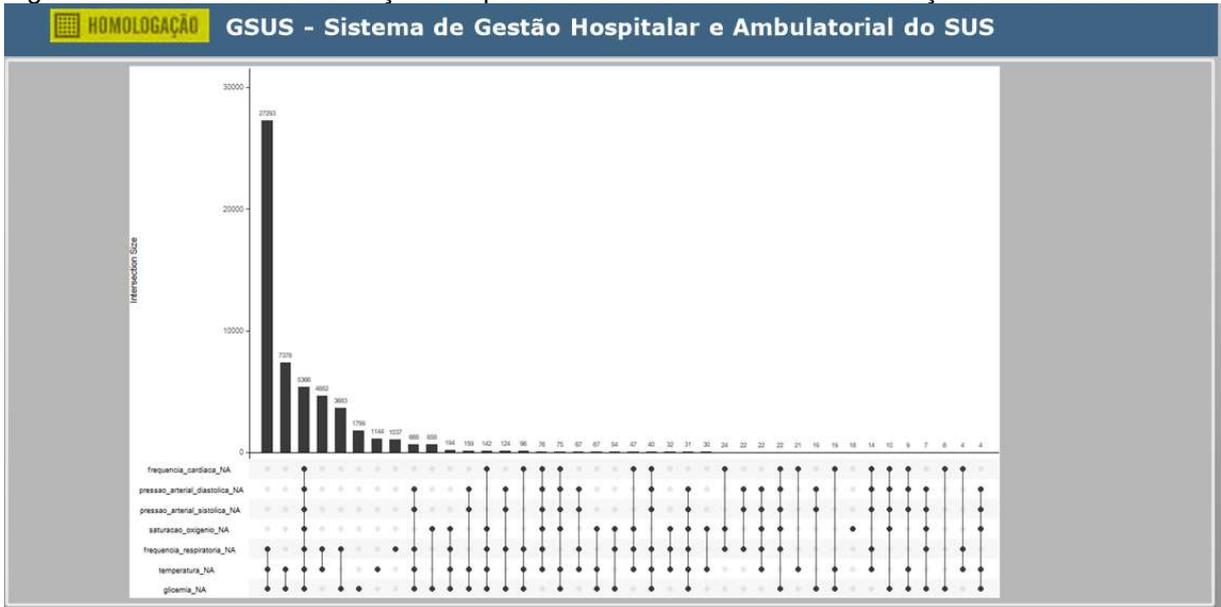
Figura 18 – Valores dos registros de sinais vitais e valores preditivos diários de um único paciente.



Fonte: O autor (2019).

A Figura 20 apresenta o acompanhamento da proporção de valores ausentes, tendo-se optado por trazer essa imagem para o QlikView, pois ele é a interface entre o modelo desenvolvido e os responsáveis pelo acompanhamento dos dados e pacientes, permitindo, assim, um ponto único de acesso e consulta aos dados. O gráfico foi gerado via *script* durante o processo de avaliação dos dados e salvo em uma pasta; durante o processo de atualização dos dados do QlikView, a tela de acompanhamento foi atualizada. Esse gráfico foi escolhido por permitir a visualização dos padrões de dados ausentes e, principalmente, das combinações de falta entre os atributos.

Figura 19 – Gráfico de visualização dos padrões de falta de dados e combinações entre os atributos.



Fonte: O autor (2019).

6 DISCUSSÃO

Para os anos de 2017 e 2018, foram identificados 164 internamentos com CID de sepse, referentes a 3,52% do total. Esse percentual está abaixo do encontrado na literatura, onde estima-se 30% de ocupação dos leitos hospitalares por pacientes com sepse (ILAS, 2018). Dos 164 internamentos, 48 pacientes (29,26%) foram a óbito, percentual igual ao apresentado por Lobo et al. (2019) para o ano de 2016 e menor que o apresentado pelo ILAS (2017), para hospitais públicos (58,5%). O perfil dos pacientes foi compatível com os achados na literatura: maioria masculina (55,48%), com média de idade de 66 anos e mediana de 69 (ZONTA et al., 2018). Para os critérios SOFA e qSOFA, foram identificados mais internamentos com suspeita de sepse (3,70% e 4,49%, respectivamente), porém a quantidade de óbitos relacionados a esses escores foi menor (9,55% e 17,88%, respectivamente).

Como o total de internamentos identificados pelo escore qSOFA ficou abaixo das estimativas apontadas na literatura, foram selecionados os internamentos que continham apenas um dos critérios de seleção – taxa de respiração maior que 22 –, que resultou em 337 internamentos. Utilizando apenas o segundo critério do escore qSOFA – pressão sistólica menor que 100 –, foram identificados 713 internamentos, o que pode indicar a possibilidade de existirem mais casos de risco de sepse. Esses casos podem ter sido perdidos durante o processo de *input* dos dados faltantes, em que foram empregados valores normais para frequência cardíaca e respiratória, pressão e temperatura. Se mapeados adequadamente, esses dados podem ajudar a melhorar o processo de identificação precoce dos pacientes em risco.

Com relação ao modelo desenvolvido para identificação precoce dos pacientes com risco de sepse, os resultados demonstraram que o modelo treinado a partir do critério de seleção qSOFA apresentou melhor desempenho frente aos critérios GSUS e SOFA, obtendo acurácia de 87,2%, sensibilidade de 87,7% e especificidade de 89,6%. Esses resultados foram obtidos utilizando como atributos apenas sete sinais vitais e informações sobre o internamento do paciente. Comparando-os com os resultados do algoritmo Insight apresentado no estudo de Mao et al. (2018), reconhecido por se tornar uma ferramenta comercial especializada em reconhecimento precoce de sepse, obteve-se uma queda no desempenho médio de 7%, o que demonstra o potencial do algoritmo, visto as limitações encontradas nos dados. Sabe-se que o critério SOFA é mais completo e robusto que o qSOFA, pois leva

em consideração exames laboratoriais para identificação dos casos de sepse, porém nesse estudo o qSOFA apresentou melhores resultados na identificação desses pacientes, esse resultado pode ter ocorrido devido à ausência de resultados de exames laboratoriais específicos para a identificação de sepse. No cenário estudado a quantidade de pacientes identificados pelo critério qSOFA (216, 4,49% do total) foi maior que os pacientes identificados pelo critério SOFA (178, 3,7% do total), e desses, nem todos contavam com os exames necessários para confirmação de sepse.

O poder preditivo do modelo foi testado por meio da redução de *timesteps* do conjunto de testes, tendo o modelo proposto apresentado resultados promissores, atingindo para o dia em que a sepse apresentou sua manifestação (dia zero) acurácia de 91,3%, sensibilidade de 92,2% e especificidade de 98,9%. Os resultados dos modelos de Calvert et al. (2016), Kam e Kim (2017) e Mao et al. (2018) foram obtidos em horas, que antecedem o evento da sepse conhecido como hora zero, para tanto, foram usados os conjuntos de dados MIMIC, que são uma fonte de acesso público e livre, compostos por dados de pacientes relacionados à saúde e informações demográficas, onde a frequência de coleta dos dados varia de minutos a horas.

A qualidade dos dados foi apontada por alguns autores como sendo a maior limitação nos estudos, causando reduções significativas nos conjuntos de dados de treinamento e testes (KAM; KIM, 2017; ROTHMAN et al., 2017; SAQIB; SHA; WANG, 2018). No GSUS, foi constatado que 95% dos registros apresentavam informações ausentes, tendo sido o preenchimento dos dados baseado em critérios preestabelecidos, levando em consideração dados próximos ou valores clinicamente normais relacionados ao sexo e idade (LIPTON et al., 2016; KALE et al., 2017; GARCIA-GALLO et al., 2018; SAQIB; SHA; WANG, 2018). Devido a essa significativa intervenção nos dados, acredita-se que o conjunto de dados deixou de apresentar um perfil de série temporal e passou a apresentar um perfil de classificação, o que justificaria os resultados do algoritmo *random forest* – valores de 97,1% de acurácia, 61,1% de sensibilidade e 99,8% de especificidade.

Comparando os valores do Random forest com o LSTM, o primeiro apresentou melhores resultados para os valores de acurácia e especificidade, porém a sensibilidade teve queda de 26,6%, o que indica que a utilização de séries temporais nesse contexto, ainda é a melhor alternativa, visto que a sensibilidade representa a habilidade do algoritmo em identificar os pacientes com risco de sepse. Em comparação com Lamping et al., (2018), segundo melhor resultado na revisão

integrativa, obteve melhores resultados, superando em 5,6% a acurácia e 5,4% a especificidade.

Como primeira oportunidade de melhoria, sugere-se aumentar a frequência de registros de sinais vitais do paciente, realizando a verificação da qualidade de preenchimento. Essas alterações permitiriam a melhoria no desempenho do algoritmo proposto, além de oportunizar outros estudos igualmente importantes dentro das instituições médicas do estado. O atual processo de registro diário impossibilita análises em que o acompanhamento é realizado de forma mais intensiva, buscando a identificação de sinais e sintomas dentro das primeiras horas de transferência para UTI ou de internamento hospitalar (KAUKONEN et al., 2015; SAQIB; SHA; WANG, 2018). Em conversas informais com o quadro clínico dos hospitais, foram apontados alguns motivos para um índice tão alto de dados faltantes, os principais foram falta de pessoal para executar todas as tarefas tanto de cuidado quanto burocráticas, falta de uma quantidade adequada de computadores e computadores lentos e defasados e por fim, a falta de incentivo à cultura de registrar os dados dos pacientes conforme orientações. Dessa maneira fica clara a necessidade de maiores investimentos em pessoal e infra-estrutura para que o objetivo de gerar dados de maior qualidade seja alcançado.

Segundo o ILAS (2017), é necessário um processo de reconhecimento em todas as áreas dos hospitais, pois de 30% a 50% dos pacientes com sepse dão entrada na instituição via unidades de urgência e emergência, outros 25% a 35% desenvolveram sepse nas unidades de internação regulares e somente 15% a 20%, durante a internação em UTI. Ainda com relação aos dados, Gupta et al., (2018) identificaram em seu estudo que, a variável mais importante do escore qSOFA é a escala de coma de Glasgow, que nesse estudo não pode ser utilizado por apresentar menos de 1% de cadastro nos internamentos.

A utilização de dados mais frequentes permitiu que Rothman et al. (2017) avaliassem todos os pacientes que deram entrada no hospital, separando-os em dois grupos distintos: os pacientes que já apresentavam sepse na admissão e o grupo de pacientes que desenvolveram sepse no hospital, permitindo a avaliação da conduta de alta, identificando possíveis situações em que o paciente pode ter tido alta já com alguns sintomas, o que geraria reinternações.

O poder de generalização do modelo foi testado utilizando dados de outro hospital, tendo sido observada uma queda de aproximadamente 10% no desempenho

médio do modelo. Esse resultado indica a necessidade de uma abordagem mais adequada se considerada a implantação do modelo para todos os hospitais do estado. Nesse cenário, é necessário considerar pelo menos duas opções: a primeira é um modelo mais genérico utilizando dados de todos hospitais, o que geraria um modelo maior e mais complexo; e a segunda é um modelo individual para cada hospital, o qual se adequaria ao perfil e especificidade de cada um. Em ambas as situações seriam necessárias um maior poder de processamento e armazenamento, em outras palavras, maior investimento financeiro.

Embora o LSTM tenha produzido resultados promissores, este é apenas o primeiro passo na utilização dessa linha de pesquisa pelo governo do estado. O reconhecimento de pacientes com risco de sepse por meio de séries temporais mais completas demonstrou que o LSTM pode identificar situações significativas, implicando diretamente a qualidade do diagnóstico e, conseqüentemente, evitando que o paciente chegue a situações mais graves.

6.1 LIMITAÇÕES

A primeira limitação encontrada foi a utilização da CID para identificação do internamento com sepse. Wyk, Khojandi e Kamaleswaran (2017) apontaram que a principal limitação em seu estudo foi o uso do código CID-9, para identificar os pacientes com e sem sepse. De modo geral, é importante lembrar que a identificação dos pacientes com sepse pode estar subestimada, pois, muitas vezes, a causa da morte pode estar atribuída à patologia de base e não propriamente à sepse (ILAS, 2017). Sendo assim, a utilização desse código se mostrou incorreta diversas vezes e pode ter afetado o desempenho do algoritmo durante a fase de treinamento e testes.

A segunda limitação foi com relação à disponibilidade dos atributos necessários para o treinamento dos algoritmos. Por se tratar de novas implantações em hospitais públicos, muitas vezes os sistemas de prontuário do paciente e os sistemas laboratoriais podem não estar integrados, prejudicando a disponibilidade de dados.

A terceira limitação deu-se por conta da qualidade dos dados, em que a frequência da coleta e o tipo dos dados influenciam o treinamento e testes dos algoritmos, podendo gerar problemas na generalização do modelo. Porém a limitação mais impactante foi com relação à frequência e a falta dos dados. Sendo ambas tratadas conforme sua variação.

Na fase de implementação do modelo, as limitações foram com relação ao equipamento disponível, onde não foi possível testar diferentes soluções para melhoria do desempenho do modelo.

6.2 TRABALHOS FUTUROS

Tendo em vista os resultados obtidos e as limitações encontradas, é possível identificar algumas sugestões para trabalhos futuros:

- a) a melhoraria na qualidade dos dados é imprescindível para o aperfeiçoamento do modelo assim como a quantidade de registros, para tornar a previsão de dias para horas;
- b) adequar o ambiente de desenvolvimento para que o modelo realize o processo de busca em todos os pacientes, não apenas nos pacientes que satisfazem os critérios dos escores avaliados;
- c) utilizar mais e mais longas séries temporais;
- d) utilização de um módulo específico de UTI, todas as avaliações levaram em conta o hospital como um todo, por mais que os leitos de UTI estejam cadastrado no sistema, os dados dos processos específicos de UTI não são devidamente registrados;
- e) desenvolvimento de modelos levando em consideração as diferentes áreas dos hospitais;
- f) melhora do processo de identificação de infecções nas evoluções dos pacientes utilizando processamento de linguagem natural;
- g) identificar e separar os processos relativos à causa inicial dos internamentos, dos processos realizados por conta dos problemas causados pela sepse;
- h) testar diferentes modelagens como: um modelo geral para todos hospitais ou cada hospital com seu modelo individual?
- i) criação de um protocolo de interoperabilidade para futura utilização intra-sistemas hospitalares.
- j) Para uma futura implementação do modelo para todos os hospitais, é preciso levar em conta o dimensionando adequado do *hardware* e o planejamento da arquitetura de *software* a ser utilizada, pois essas atitudes garantem o mínimo de concorrência possível com os servidores

dos hospitais; nesse sentido, sugere-se a utilização de um *cluster big data* para o processamento em memória dos dados que são gerados continuamente (*streaming*). A arquitetura distribuída utilizando código aberto, como o *Hadoop*, permite trabalhar com grandes volumes de dados, onde a principal característica é a escalabilidade do *cluster*, ou seja, o incremento do poder de processamento é realizado apenas adicionando computadores de baixo custo. Como principal benefício, destaca-se a remoção de todo e qualquer processamento necessário para desenvolvimento da ferramenta dos servidores dos hospitais.

7 CONSIDERAÇÕES FINAIS

Esta dissertação teve o objetivo de desenvolver um modelo preditivo utilizando algoritmo LSTM e séries temporais multivariadas para reconhecimento precoce de pacientes em risco de sepse. O LSTM foi escolhido devido à característica de memorizar as dependências temporais de longos períodos, possibilitando capturar, inclusive, as progressões mais sutis da sepse. De acordo com a proposta de contribuição social, foi possível ter uma visão das ocorrências dos casos de sepse do hospital estudado, informações que são relevantes, uma vez que os dados de sepse para o estado do Paraná são escassos.

Por meio das etapas da pesquisa, foram encontradas diversas propostas e objetivos de estudos envolvendo ML e sepse em vários contextos, como, por exemplo, avaliação de mortalidade e gravidade, simulação de tratamentos, avaliação dos escores para identificação de sepse e principalmente modelos preditivos. A implementação seguiu a metodologia proposta pelas principais referências no reconhecimento precoce de sepse. A partir dos resultados obtidos foi possível comparar o desempenho dos algoritmos implementados com os achados na literatura, permitindo identificar pontos fortes e fracos, além de possíveis melhorias no processo de registro dos dados dos pacientes, no sistema utilizado e nos modelos implementados.

Assim, para que esta ferramenta tenha uma contribuição social, é fundamental considerar os aspectos apontados nas seções anteriores, para possibilitar a identificação dos casos com suspeita de sepse mais próximo possível do tempo real, desta maneira é possível contribuir efetivamente para a redução dos óbitos. Espera-se que a ferramenta de fato apoie o corpo clínico na redução do tempo entre o início dos sintomas e o primeiro atendimento médico, pois somente desta maneira estará otimizando o planejamento das intervenções quanto à implementação dos protocolos assistenciais.

REFERÊNCIAS

- ALPAYDIN, E. **Introduction to machine learning**. Cambridge: MIT Press, 2014.
- AMLAND, R. C.; HAHN-COVER, K. Clinical decision support for early recognition of sepsis. **American Journal of Medical Quality**, v. 31, n. 2, p. 103-110, 2016.
- ANDRADE, A. **Best practices for convolutional neural networks applied to object recognition in images**. Toronto: University of Toronto, 2013.
- ANGUS, D. C. et al. Epidemiology of severe sepsis in the United States: analysis of incidence, outcome, and associated costs of care. **Critical Care of Medicine**, v. 29, n. 7, p. 1303-1310, 2001.
- ASHAL, G. E. et al. Simple definition and calculation of accuracy, sensitivity and specificity. **Emergency**, v. 3, n. 2, p. 48-49, 2015.
- BERGMEIR, C.; BENÍTEZ, J. M. On the use of cross-validation for time series predictor evaluation. **Information Sciences**, v. 191, May 2012.
- BERGMEIR, C.; HYNDMAN, R. J.; KOO, B. A note on the validity of cross-validation for evaluating autoregressive time series prediction. **Computational Statistics & Data Analysis**, v. 120, n. C, p. 70-83, 2018.
- BERGSTRA, J.; BENGIO, Y. Random search for hyper-parameter optimization. **Journal of Machine Learning Research**, v. 13, p. 281-305, Feb. 2012.
- BONE, R. C. et al. Definitions for sepsis and organ failure and guidelines for the use of innovative therapies in sepsis. **Chest**, v. 101, n. 6, p. 1644-1655, 1992.
- BONTEMPI, G.; TAIEB, S. B.; BORGNE, Y. Machine learning strategies for time series forecasting. **eBISS**, p. 62-77, 2013.
- BRASIL. Ministério da Saúde. Portaria n. 1.559, de 1º de agosto de 2008. Institui a Política Nacional de Regulação do Sistema Único de Saúde – SUS. **Diário Oficial da União**, Brasília, DF, 2 ago. 2008.
- BREIMAN, L. Random forests. **Machine Learning**, v. 45, n. 1, p. 5-32, 2001.
- BROWNLEE, J. **Long short-term memory networks with python: develop sequence prediction models with deep learning**. [S.l.], 2017.
- BUDUMA, N.; LOCASCIO, N. **Fundamentals of deep learning: designing next-generation machine intelligence algorithms**. Boston: O'Reilly Media, 2017.
- CALVERT, J. S. et al. A computational approach to early sepsis detection. **Computers in Biology and Medicine**, n. 74, p. 69-73, 2016.
- CARVALHO, D. R.; ESCOBAR, L. F. A.; TSUNODA, D. Pontos de atenção para o uso da mineração de dados na saúde. **Informação & Informação**, Londrina, v. 19, p. 249-273, 26 abr. 2014.

CHOLLET, F. **Deep learning with Python**. [S.l.]: Manning, 2018.

SOCIEDADE BRASILEIRA DE DIABETES. **Diagnóstico e classificação do diabetes mellitus e tratamento do diabetes mellitus tipo 2**. [S.l.], 2000. Disponível em: http://bvsmms.saude.gov.br/bvs/publicacoes/consenso_bras_diabetes.pdf. Acesso em: 12 jul. 2019.

COPPIN, B. **Artificial intelligence illuminated**. [S.l.]: Jones and Bartlett, 2004.

BRASIL, DATASUS. **CID10**, 1 dez. 2018. Disponível em: <http://www.datasus.gov.br/cid10/V2008/cid10.htm>. Acesso em: 1 dez. 2018.

DER LOO, M. V. The stringdist package for approximate string matching. **The R Journal**, v. 6, n. 1, June 2014.

DER LOO, M. V. et al. **Approximate string matching and string distance functions**. 2019. Disponível em: <https://cran.r-project.org/web/packages/stringdist/stringdist.pdf>. Acesso em: 10 jul. 2019.

DESAUTELS, T. et al. Prediction of sepsis in the intensive care unit with minimal electronic health record data: a machine learning approach. **JMIR Medical Informatics**, v. 4, n. 3, p. e28, 2016.

DYNIIEWICZ, A. M. **Metodologia da pesquisa em saúde para iniciantes**. 3. ed. São Caetano do Sul: Difusão, 2014.

EKE, C. et al. A survey of user profiling: state-of-the-art, challenges and solutions. **IEEE Access**, v. 7, p. 144907-144924, 2019.

FAISAL, M. et al. 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. **Critical Care Medicine**, v. 46, n. 4, p. 612-618, 2018.

FAYYAD, U.; PIATETSKY-SHAPIRO, G.; SMYTH, P. From data mining to knowledge discovery in databases. **AI Magazine**, v. 17, n. 3, p. 37, 1996.

FELLOW, K. S. F.; ROSENFELD, A. Pattern recognition and image processing. **IEEE Transactions On Computers**, v. C-25, n. 12, p. 1336-1346, 1976.

FLEISS, J. L.; COHEN, J. The equivalence of weighted Kappa and the intraclass correlation coefficient as measures of reliability. **Educational and Psychological Measurement**, n. 33, p. 613-619, 1973.

GARCIA-GALLO, J. E. et al. A machine learning-based model for 1-year mortality prediction in patients admitted to an intensive care unit with a diagnosis of sepsis. **Medicina Intensiva**, n. 8, p. 30245-30246, 2018.

GERS, F. A.; SCHMIDHUBER, J. Recurrent nets that time and count. *In*: INTERNATIONAL JOINT CONFERENCE ON NEURAL NETWORKS, 2000, Como. **Proceedings [...]**. [S.l.]: IEEE, 2000.

GHALWASH, M.; RADOSAVLJEVIC, V.; OBRADOVIC, Z. Early diagnosis and its benefits in sepsis blood purification treatment. *In: IEEE INTERNATIONAL CONFERENCE ON HEALTHCARE INFORMATICS*, 2013, Washington, D.C. **Proceedings [...]**. [S.l.]: IEEE, 2013.

GHOSH, S. et al. Septic shock prediction for ICU patients via coupled HMM walking on sequential contrast patterns. **Journal of Biomedical Informatics**, n. 66, p. 19-31, 2017.

GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. **Deep learning**. Cambridge: MIT Press, 2015.

GRAVES, A. Supervised Sequence Labelling with Recurrent Neural Networks: [s.n.], 2012.

GRAVES, A. et al. A novel connectionist system for unconstrained handwriting recognition. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 31, n. 5, p. 855-868, 2008.

GREIN, D. **Projeto GSUS**: relatório gerencial. Curitiba: [s.n.], 2018.

GUILLÉN, J. et al. Predictive models for severe sepsis in adult ICU patients. *In: IEEE SYSTEMS AND INFORMATION ENGINEERING DESIGN SYMPOSIUM*, 2015, Charlottesville. **Proceedings [...]**. [S.l.]: IEEE, 2015.

GULTEPE, E. et al. A Bayesian network for early diagnosis of sepsis patients: a basis for a clinical decision support system. *In: INTERNATIONAL CONFERENCE ON COMPUTATIONAL ADVANCES IN BIO AND MEDICAL SCIENCES*, 2., 2012, Las Vegas. **Proceedings [...]**. [S.l.]: IEEE, 2012.

GULTEPE, E. et al. From vital signs to clinical outcomes for patients with sepsis: a machine learning basis for a clinical decision support system. **Journal of the American Medical Informatics Association**, v. 21, n. 2, p. 315-325, 2014.

GUPTA, A. et al. Using statistical and machine learning methods to evaluate the prognostic accuracy of SIRS and qSOFA. **Healthcare Informatics Research**, v. 2, p. 139-147, Apr. 2018.

HENKIN, C. S. et al. Sepsis: uma visão atual. **Scientia Medica**, Porto Alegre, v. 19, n. 3, p. 135-145, jul./set. 2009.

HENRY, K. E. et al. A targeted real-time early warning score (TREWScore) for septic shock. **Science Translational Medicine**, v. 7, n. 299, 299ra122, 5 Aug. 2015.

HOCHREITER, S.; SCHMIDHUBER, J. Long short-term memory. **Neural Computation**, v. 9, n. 8, p. 1735-1780, 1997.

HORNG, S. et al. Creating an automated trigger for sepsis clinical decision support at emergency department triage using machine learning. **PLoS One**, v. 12, n. 4, p. e0174708, 2017.

HOSMER, W. D.; LEMESHOW, S. **Applied logistic regression**. 2. ed. [S.l.]: Wiley &

Sons, 2000.

INSTITUTO LATINO AMERICANO DE SEPSE (ILAS). **Sepse**: um problema de saúde pública. Brasília, DF: [s.n.], 2017.

INSTITUTO LATINO AMERICANO DE SEPSE (ILAS). **Implementação de protocolo gerenciado de sepse**: protocolo clínico. São Paulo, 2018.

JAIN, A. K.; MAO, J. Artificial neural networks: a tutorial. **IEEE Computer**, v. 29, n. 3, p. 56-63, 1996.

JIANG, Y. et al. A self-adaptively evolutionary screening approach for sepsis patient. *In*: INTERNATIONAL SYMPOSIUM ON COMPUTER-BASED MEDICAL SYSTEMS, 29., 2016, Belfast. **Proceedings [...]**. [S.l.]: IEEE, 2016.

KALCHBRENNER, N.; DANIHELKA, I.; GRAVES, A. Grid long short-term memory. *In*: INTERNATIONAL CONFERENCE ON LEARNING REPRESENTATIONS, 4., 2016, San Juan. **Proceedings [...]**. [S.l.: s.n.], 2016.

KAM, H. J.; KIM, H. Y. Learning representations for the early detection of sepsis with deep neural networks. **Computers in Biology and Medicine**, n. 89, p. 248-255, 2017.

KAUKONEN, K. M. et al. Systemic inflammatory response syndrome criteria in defining severe sepsis. **The New England Journal of Medicine**, v. 373, n. 9, p. 10-17, 2015.

KELLY, G. Body temperature variability (part 1): a review of the history of body temperature and its variability due to site selection, biological rhythms, fitness, and aging. **Alternative Medicine Review**, v. 11, n. 4, p. 278-293, Dec. 2006.

KOSH, G.; LANDIS, R. J. The measurement of observer agreement for categorical data. **Biometrics**, v. 33, n. 1, p. 159-174, Mar. 1977.

KUMAR, A. et al. Duration of hypotension before initiation of effective antimicrobial therapy is the critical determinant of survival in human septic shock. **Critical Care Medicine**, v. 34, n. 6, p. 1589-1596, 2006.

KUSHIMOTO, S. et al. Complementary role of hypothermia identification to the quick sequential organ failure assessment score in predicting patients with sepsis at high risk of mortality: a retrospective analysis from a multicenter, observational study. **Journal of Intensive Care Medicine**, p. 1-9, Jan. 2018.

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. **BMC Pediatrics**, v. 18, n. 112, 2018.

LIAW, A.; WIENER, M. Classification and regression by random forest. **R News**, v. 2/3, p. 18-21, 2002.

LIPTON, Z. C. et al. Learning to diagnose with lstm recurrent neural networks. *In*: INTERNATIONAL CONFERENCE ON LEARNING REPRESENTATIONS, 4., 2016,

San Juan. **Proceedings [...]**. [S.l.: s.n.], 2016.

LIPTON, Z. C.; BERKOWITZ, J.; ELKAN, C. A critical review of recurrent neural networks for sequence learning. **Neural and Evolutionary Computing**, p. 1-38, 2015.

LOBO, S. M. et al. Mortalidade por sepse no Brasil em um cenário real: projeto UTIs brasileiras. **Revista Brasileira de Terapia Intensiva**, São José do Rio Preto, v. 31, n. 1, p. 1-4, nov. 2019.

MACDONALD, S. P. et al. Review article: sepsis in the emergency department – part 1: definitions and outcomes. **Emergency Medicine Australasia**, v. 29, n. 6, p. 619-625, 2017.

MACHADO, F. R. et al. The epidemiology of sepsis in Brazilian intensive care units (the Sepsis PREvalence Assessment Database, SPREAD): an observational study. **The Lancet Infectious Disease**, v. 17, n. 11, p. 1180-1189, 2017.

MALACHIAS, M. V. B. et al. 7ª Diretriz brasileira de hipertensão arterial. **Arquivos Brasileiros de Cardiologia**, Rio de Janeiro, v. 107, n. 3, supl. 3, set. 2016.
Disponível em: http://www.scielo.br/pdf/abc/v107n3s3/pt_0066-782X-abc-107-03-s3-0007.pdf. Acesso em: 15 set. 2019.

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**, v. 8, n. 1, p. e017833, 2018.

MARSLAND, S. **Machine learning: an algorithmic perspective**. Boca Raton: CRC, 2015.

MITCHELL, S. et al. Optimization of sepsis risk assessment for ward patients. *In*: IEEE SYSTEMS AND INFORMATION ENGINEERING DESIGN CONFERENCE, 2016, Charlottesville. **Proceedings [...]**. [S.l.]: IEEE, 2016.

MOHRI, M.; ROSTAMIZADEH, A.; TALWALKAR, A. **Foundations of machine learning**. Cambridge: MIT Press, 2012.

NEMATI, S. et al. An interpretable machine learning model for accurate prediction of sepsis in the ICU. **Critical Care Medicine**, v. 46, n. 4, p. 547-553, 2018.

PARIKH, R. et al. Understanding and using sensitivity, specificity and predictive values. **Indian Journal of Ophthalmology**, v. 56, n. 1, p. 45-50, 2008.

PARREIRA, V. F. et al. Padrão respiratório e movimento toracoabdominal em indivíduos saudáveis: influência da idade e do sexo. **Revista Brasileira de Fisioterapia**, São Carlos, v. 14, n. 5, p. 411-416, out. 2010.

PASCHOAL, M. et al. Variabilidade da frequência cardíaca em diferentes. **Revista Brasileira de Fisioterapia**, São Carlos, v. 10, p. 413-419, dez. 2006.

PERKINS, N. J.; SCHISTERMAN, E. F. The inconsistency of “optimal” cutpoints obtained using two criteria based on the receiver operating characteristic curve.

American Journal of Epidemiology, v. 163, n. 7, p. 670-675, 2006.

RIVERS, E. et al. Early goal-directed therapy in the treatment of severe sepsis and septic shock. **New England Journal Medical**, v. 345, n. 19, p. 1368-1377, Nov. 2001.

ROTHMAN, M. et al. Sepsis as 2 problems: identifying sepsis at admission and predicting onset in the hospital using an electronic medical record–based acuity score. **Journal of Critical Care**, v. 38, p. 237-244, 2017.

RSTUDIO. Disponível em: <https://www.rstudio.com/>. Acesso em: 1 jul. 2019.

RUMELHART, D. E.; HINTON, G. E.; WILLIAMS, R. J. Learning representations by back-propagating errors. **Nature**, n. 323, p. 533-536, 1986.

RYBICKA, M. K.; SCHRAUDOLPH, N. N.; SCHMIDHUBER, J. **Unsupervised learning in LSTM recurrent neural networks**. [S.l.: s.n.], 2001.

SAK, H.; SENIOR, A.; BEAUFAYS, F. Long short-term memory recurrent neural network architectures for large scale acoustic modeling. **Computer Science**, 2014.

SALEHINEJAD, H. et al. Recent advances in recurrent neural networks. **Computer Science**, 2018.

SALES JÚNIOR, J. A. L. et al. Sepses Brasil: estudo epidemiológico da sepse em unidades de terapia intensiva brasileiras. **Revista Brasileira Terapia Intensiva**, São Paulo, v. 18, n. 1, p. 9-17, jan. 2006.

SALOMÃO, R. et al. Diretrizes para tratamento da sepse grave/choque séptico: abordagem do agente infeccioso – controle do foco infeccioso e tratamento antimicrobiano. **Revista Brasileira Terapia Intensiva**, São Paulo, v. 23, n. 2, p. 145-157, 2011.

SAMUEL, L. A. Some studies in machine learning using the game of checkers. **IBM Journal**, v. 3, n. 3, p. 210-229, 1959.

SAQIB, M.; SHA, Y.; WANG, M. D. Early prediction of sepsis in EMR records using traditional ml techniques and deep learning LSTM networks. *In: ANNUAL INTERNATIONAL CONFERENCE OF THE IEEE ENGINEERING IN MEDICINE AND BIOLOGY SOCIETY, 40.*, 2018, Honolulu. **Proceedings [...]**. [S.l.]: IEEE, 2018.

SCHUH, C. J. Sepsis and septic shock analysis using neural networks. *In: ANNUAL MEETING OF THE NORTH AMERICAN FUZZY INFORMATION PROCESSING SOCIETY, 2007, Vienna.* **Proceedings [...]**. [S.l.: s.n.], 2007.

SERAFIM, R. et al. A comparison of the Quick-SOFA and systemic inflammatory response syndrome criteria for the diagnosis of sepsis and prediction of mortality a systematic review and meta-analysis. **Chest**, v. 153, n. 3, p. 646-655, 2018.

SEYMOUR, C. W. et al. Time to treatment and mortality during mandated emergency care for sepsis. **New England Journal Medical**, v. 23, p. 2235-2244, June 2017.

- SHASHIKUMAR, P. S. et al. Multiscale network representation of physiological time series for early prediction of sepsis. **Physiological Measurement**, v. 38, n. 12, p. 2235-2248, 2017.
- SINGER, M. et al. The third international consensus definitions for sepsis and septic shock (Sepsis-3). **JAMA**, v. 315- n. 8, p. 801-810, 2016.
- SOKOLOVA, M.; LAPALME, G. A systematic analysis of performance measures for classification tasks. **Information Processing and Management**, n. 45, p. 427-437, 2009.
- SOUZA, E. S. et al. Mortalidade e riscos associados a infecção relacionada à assistência à saúde. **Texto Contexto Enfermagem**, Florianópolis, v. 24, n. 1, p. 220-228, 2015.
- SUTSKEVER, I.; VINYALS, O.; LE, V. Q. Sequence to sequence learning with neural networks. **Advances in Neural Information Processing System**, p. 3104-3112, 2014.
- TABOR, J.; SOUREK, P. Cross-entropy clustering. **Pattern Recognition**, p. 1-41, Mar. 2014.
- TAYLOR, A. R. et al. Prediction of in-hospital mortality in emergency department patients with sepsis: a local big data–driven, machine learning approach. **Academic Emergency Medicine**, v. 23, n. 3, p. 269-278, 2016.
- VINCENT, J. L.; MORENO, R.; TAKALA, J. The SOFA (Sepsis-related Organ Failure Assessment) score to describe organ dysfunction/failure. **Intensive Care Medicine**, v. 22, n. 7, p. 707-710, 1996.
- WESTPHAL, G. A.; LINO, A. S. Rastreamento sistemático é a base do diagnóstico precoce da sepse grave e choque séptico. **Revista Brasileira de Terapia Intensiva**, São Paulo, v. 27, n. 2, p. 96-101, jun. 2015.
- WYK, F. V.; KHOJANDI, A.; KAMALESWARAN, R. How much data should we collect: a case study in sepsis detection using deep learning. *In: IEEE HEALTHCARE INNOVATIONS AND POINT OF CARE TECHNOLOGIES (HI-POCT)*, 2017, Bethesda. **Proceedings [...]**. [S.l.]: IEEE, 2017.
- YOSHIHARA, J. C. et al. Análise descritiva dos pacientes com sepse grave ou choque séptico e fatores de risco para mortalidade. **Semina: Ciências Biológicas e da Saúde**, Londrina, v. 32, n. 2, p. 127-134, dez. 2011.
- ZANON, F. et al. sepse na unidade de terapia intensiva: etiologias, fatores prognósticos e mortalidade. **Revista Brasileira de Terapia Intensiva**, São Paulo, v. 20, n. 2, p. 128-134, 2008.
- ZHANG, et al. **Dive into deep learning**. [S.l.: s.n.], 2018.
- ZHANG, Y. et al. LSTM for septic shock: adding unreliable labels to reliable predictions. *In: IEEE INTERNATIONAL CONFERENCE ON BIG DATA*, 2017, Boston. **Proceedings [...]**. [S.l.]: IEEE, 2017.

ZHENG, A. **Evaluating machine learning models**. Boston: O'Reilly Media, 2015.

ZHENG, A.; CASARI, A. **Feature engineering for machine learning**. Boston: O'Reilly, 2018.

ZHU, W.; ZENG, N.; WANG, N. Sensitivity, specificity, accuracy, associated confidence interval and ROC analysis with practical SAS® implementations. **NESUG Proceedings: Health Care and Life Sciences**, p. 1-9, 2010.

ZONTA, F. N. S. et al. Características epidemiológicas e clínicas da sepse em um hospital público do Paraná. **Revista de Epidemiologia e Controle de Infecção**, Santa Cruz do Sul, v. 8, n. 3, p. 224-231, 2018.

APÊNDICE A – ARTIGO DE REVISÃO

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

Everton O. Cesario¹
 Cristiane Y. Nakamura²
 Yohan B. Gumiel³
 Deborah R. Carvalho⁴

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

¹ Mestrando pela Pontifícia Universidade Católica do Paraná (PUCPR) – Curitiba (PR) – Brasil. E-mail: evertoncesario@hotmail.com

² Mestranda pela Pontifícia Universidade Católica do Paraná (PUCPR) – Curitiba (PR) – Brasil. E-mail: cristy_nakamura@hotmail.com

³ Mestre e doutorando pela Pontifícia Universidade Católica do Paraná (PUCPR) – Curitiba (PR) – Brasil. E-mail: yohan.gumiel@hotmail.com

⁴ 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. E-mail: ribeiro.carvalho@pucpr.br

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.

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 et al., 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 (Gulpepe 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, et al., 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: *sepsis*, *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 da 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 *sepsis* 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.

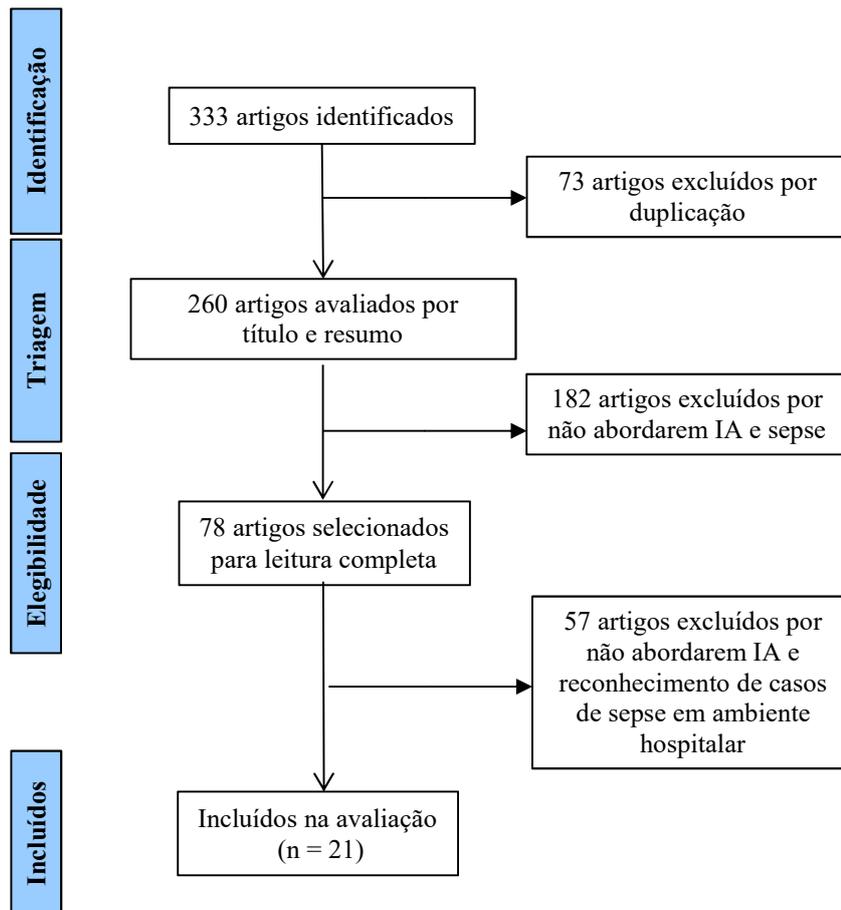


Figura 1 – Processo de seleção dos artigos.
 Fonte: Adaptado de Altman et al. (2009).

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:

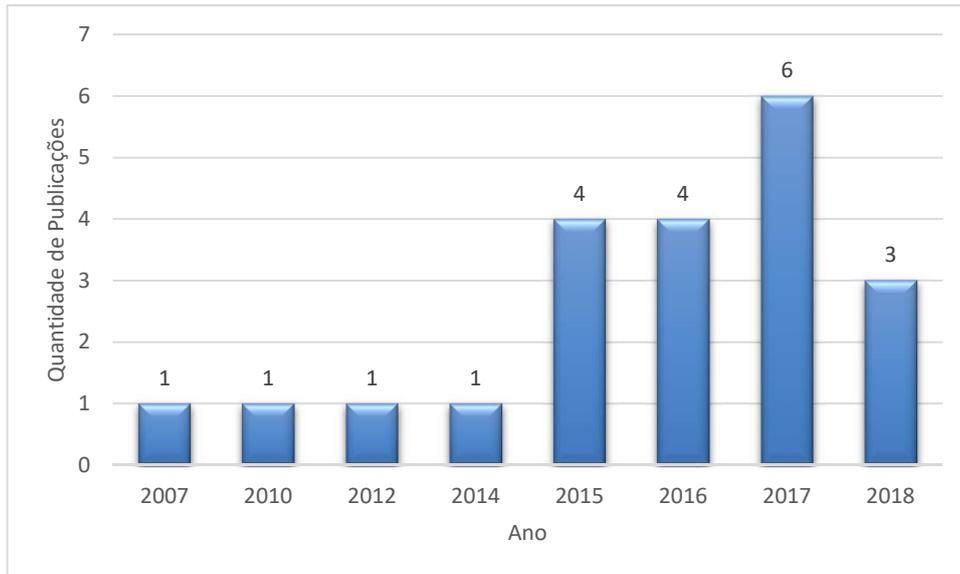


Figura 2 – Quantidade de publicações selecionadas por ano.
Fonte: os autores (2019).

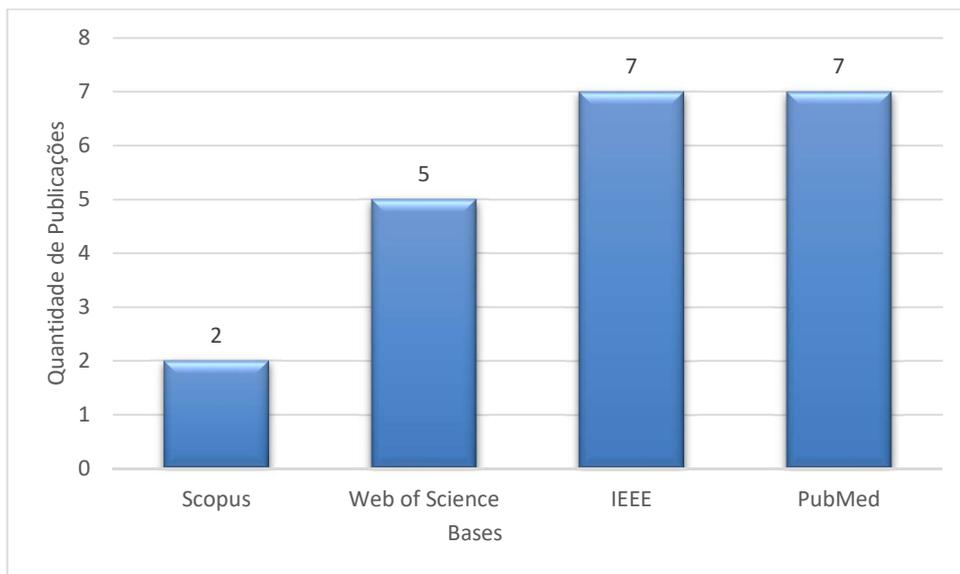


Figura 3 – Quantidade de publicações selecionadas por base.
Fonte: os autores (2019).

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, et al., 2015; Parikh, et al., 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:

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

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:

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

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%
Horng et al. (2017)	Random Forest	87%	83%	70%	12+PL N		230.936	14%
Horng et al. (2017)	Regressão Logística	86%	84%	67%	12+PL N		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	-
Gulpepe et al. (2012)	Bayesian Networks			Mapa	7		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 diminuiu 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 *desempenho* da detecção de casos de sepse cujos sintomas iniciaram dentro da hora zero, sob os valores de curva ROC (*Receiver Operating Characteristic*), sensibilidade e especificidade, ficando evidenciada a melhor *desempenho* 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
ROC	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
Sensibilidade	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
Especificidade	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
Auroc	0,90	0,82	0,79	0,76	0,75	0,72
Sensibilidade	0,80	0,80	0,80	0,80	0,80	0,80
Especificidade	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 Horneg 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.

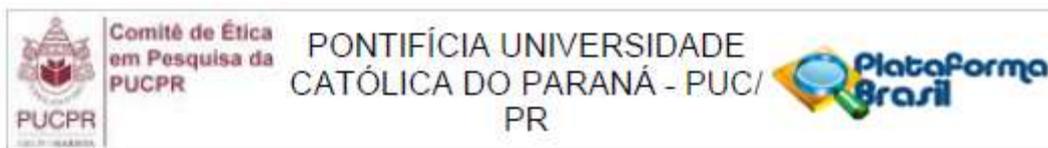
Foi possível perceber que, utilizando modelos mais simples e dados mais fáceis de serem obtidos, se podem obter alguns resultados expressivos. Os modelos mais complexos necessitam de um conjunto maior de dados e, normalmente, apenas são passíveis de serem desenvolvidos em hospitais que possuem uma infraestrutura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

ANEXO A – APROVAÇÃO DO COMITÊ DE ÉTICA DA PUCPR



PARECER CONSUBSTANCIADO DO CEP

DADOS DO PROJETO DE PESQUISA

Título da Pesquisa: IDENTIFICAÇÃO DE PACIENTES COM RISCO DE SEPSE EM AMBIENTE

Pesquisador: Deborah Ribeiro Carvalho

Área Temática:

Versão: 2

CAAE: 13818519.9.0000.0020

Instituição Proponente: Pontifícia Universidade Católica do Parana - PUCPR

Patrocinador Principal: Financiamento Próprio

DADOS DO PARECER

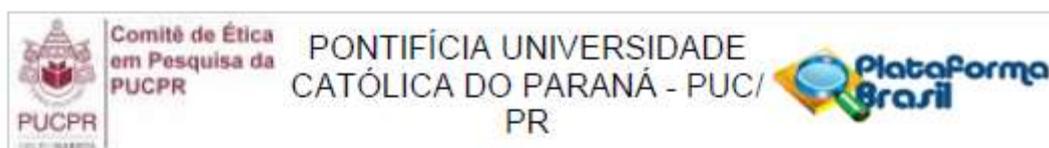
Número do Parecer: 3.388.114

Apresentação do Projeto:

Descrito pelos autores:

Sepse pode ser definida como a resposta sistêmica a uma doença infecciosa, seja ela causada por bactérias, fungos, vírus ou protozoários, manifestando-se em diferentes estágios clínicos de um mesmo processo fisiopatológico. Inicia-se como um processo inflamatório agudo frente à infecção do paciente, podendo evoluir para uma inflamação exacerbada e crônica, chamada de Síndrome da Resposta Inflamatória Sistêmica (SIRS). Nos Estados Unidos pelo menos 750 mil pessoas desenvolvem sepse severa ou choque séptico todo ano, mais da metade desses casos foram admitidos em uma unidade de terapia intensiva (UTI) e correspondem de 20% a 30% do total de mortes do hospital, com um custo anual de 15,4 milhões de dólares. No Reino Unido a sepse afeta mais de 260 mil pessoas por ano, gerando um custo direto de 2 milhões de libras. O ILAS (Instituto Latino Americano de Sepse) estima que a expectativa de ocorrência de sepse no Brasil seja de 670 mil casos por ano, com 50% destes resultando em morte. Machado (2017) publicou um estudo de prevalência em ambiente nacional, selecionando 230 UTI's em um só dia, e identificou que 30% dos leitos estavam ocupados por pacientes com sepse ou choque séptico. Não existem dados oficiais da Secretaria de Saúde do Estado do Paraná, com relação ao número de casos de sepse, taxa de mortalidade de 69% e outros a taxa de mortalidade foi de 52,4%, ficando acima da média de 46,6% constatada pelos autores. O problema se caracteriza pelo reconhecimento tardio da sepse, pois suas manifestações podem ser confundidas com as de

Endereço: Rua Imaculada Conceição 1155
Bairro: Prado Velho **CEP:** 80.215-901
UF: PR **Município:** CURITIBA
Telefone: (41)3271-2103 **Fax:** (41)3271-2103 **E-mail:** nep@pucpr.br



Continuação do Parecer: 3.388.114

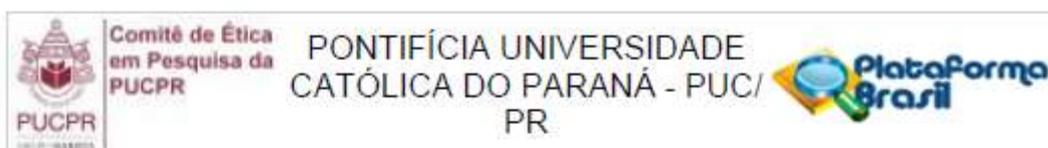
outros processos não infecciosos, ou mesmo podem passar despercebidas. As intervenções iniciais dependem do reconhecimento precoce, que é um processo contínuo e manual do profissional de saúde, sendo esta, uma das maiores dificuldades na prática clínica, pois depende diretamente da habilidade do profissional de identificar os pacientes em risco. Estas identificações são realizadas por meio de escores específicos que mensuram a gravidade e também são utilizadas para fazer o reconhecimento dos casos, porém esses critérios falham na identificação de um a cada oito pacientes com sepse severa. Kumar demonstrou que a cada hora de atraso no tratamento, a taxa de sobrevivência se reduz em até 8%, para eles um diagnóstico rápido aliado ao início precoce da terapia, administrando antimicrobianos, preferencialmente, em até 60 minutos após o reconhecimento da hipotensão, causa um impacto positivo na redução da mortalidade. O reconhecimento precoce provê benefícios para o paciente no tratamento e nos resultados, chegando a reduzir de 46,5% a 30,5% a mortalidade. Kumar ainda observou que a introdução de antibióticos eficientes, durante a primeira hora de hipotensão documentada, aumenta a sobrevida dos pacientes com choque séptico. Estudos demonstraram o potencial de métodos baseados em machine-learning para apoiar a predição de sepse, além da possibilidade de criar um modelo que mapeie as diversas progressões relacionadas a sepse, como mortalidade, falha sistêmica de órgãos e possíveis terapias para o tratamento de pacientes. Mao comparou um algoritmo de machine-learning de detecção e predição de sepse com três sistemas de escore utilizados para identificação de sepse em ambientes hospitalares, Modified Early Warning Score (MEWS), SIRS e Sequential Organ Failure Assessment (SOFA). Ainda no estudo de Mao et al., (2017), foram utilizados apenas 6 dados de sinais vitais para realização das análises, sendo: pressão sistólica, pressão diastólica, frequência cardíaca, frequência respiratória, saturação de oxigênio e temperatura. Os resultados mostraram que a acurácia, sensibilidade e especificidade do algoritmo de machine-learning obtiveram melhores resultados em relação aos outros três escores. Portanto, considerando a complexidade envolvida na detecção precoce de sepse, este projeto tem como objetivo desenvolver um modelo preditivo que identifique os pacientes em risco de sepse, nos hospitais gerenciados pelo GSUS, baseado em estratégias da Inteligência Artificial.

Objetivo da Pesquisa:

Descrito pelos autores:

Objetivo Primário: Construir um modelo preditivo que auxilie o corpo clínico na identificação dos pacientes em risco de sepse nos hospitais que são gerenciados pelo sistema do Gsus do Governo do Estado do Paraná.

Endereço:	Rua Imaculada Conceição 1155		
Bairro:	Prado Velho	CEP:	80.215-901
UF:	PR	Município:	CURITIBA
Telefone:	(41)3271-2103	Fax:	(41)3271-2103
E-mail:	nep@pucpr.br		



Continuação do Parecer: 3.388.114

Avaliação dos Riscos e Benefícios:

Descrito pelos autores:

Riscos: Existe risco de identificação dos pacientes, dos profissionais de saúde e dos estabelecimentos hospitalares. Porém este risco é minimizado a partir dos procedimentos de anonimização dos dados, bem como da adoção de datas fictícias. Para o desenvolvimento do projeto serão adotados dados secundários oriundos dos estabelecimentos hospitalares, que não serão identificados nem mesmo para as publicações oriundas do projeto. Benefícios: O reconhecimento precoce da sepse permite agilizar o início de tratamento, contribuindo diretamente com melhores resultados, podendo atingir uma redução da taxa de mortalidade de 46,5% para 30,5% (RIVERS et al., 2001). Kumar et al., (2006) também observaram que a introdução de antibióticos adequados, durante a primeira hora de hipotensão documentada, aumenta a sobrevivência dos pacientes com choque séptico.

Comentários e Considerações sobre a Pesquisa:

Os objetivos e a metodologia apresentados estão adequados e de acordo com a Resolução 466/2012.

Considerações sobre os Termos de apresentação obrigatória:

Todos os termos de apresentação obrigatória foram anexados e estão adequados e em acordo com a Resolução 466/2012.

Recomendações:

Sem recomendações.

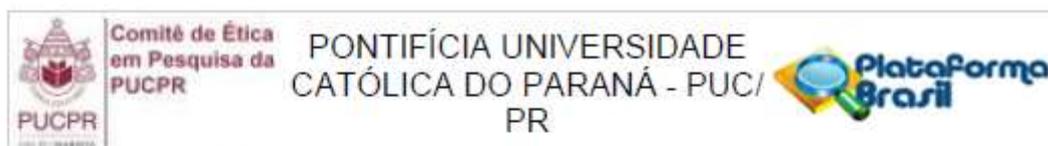
Conclusões ou Pendências e Lista de Inadequações:

Projeto aprovado.

Considerações Finais a critério do CEP:

Lembramos aos senhores pesquisadores que, no cumprimento da Resolução 466/2012, o Comitê de Ética em Pesquisa (CEP) deverá receber relatórios anuais sobre o andamento do estudo, bem como a qualquer tempo e a critério do pesquisador nos casos de relevância, além do envio dos relatos de eventos adversos, para conhecimento deste Comitê. Salientamos ainda, a necessidade de relatório completo ao final do estudo. Eventuais modificações ou emendas ao protocolo devem ser apresentadas ao CEPPUCPR de forma clara e sucinta, identificando a parte do protocolo a ser modificada e as suas justificativas. Se a pesquisa, ou parte dela for realizada em outras instituições, cabe ao pesquisador não iniciá-la antes de receber a autorização formal para a sua

Endereço: Rua Imaculada Conceição 1155
 Bairro: Prado Velho CEP: 80.215-901
 UF: PR Município: CURITIBA
 Telefone: (41)3271-2103 Fax: (41)3271-2103 E-mail: nep@pucpr.br



Continuação do Parecer: 3.388.114

realização. O documento que autoriza o início da pesquisa deve ser carimbado e assinado pelo responsável da instituição e deve ser mantido em poder do pesquisador responsável, podendo ser requerido por este CEP em qualquer tempo.

Este parecer foi elaborado baseado nos documentos abaixo relacionados:

Tipo Documento	Arquivo	Postagem	Autor	Situação
Informações Básicas do Projeto	PB_INFORMAÇÕES_BÁSICAS_DO_PROJETO_1313673.pdf	05/06/2019 16:51:28		Aceito
Declaração de Pesquisadores	TCUD.pdf	05/06/2019 16:50:51	EVERTON OSNEI CESARIO	Aceito
Projeto Detalhado / Brochura Investigador	Everton_Osnei_Cesario.docx	09/05/2019 21:24:44	EVERTON OSNEI CESARIO	Aceito
Declaração de Instituição e Infraestrutura	OF_DTI_204_2019_Celepar.pdf	09/05/2019 21:18:18	EVERTON OSNEI CESARIO	Aceito
Folha de Rosto	folhaderostoassinada.pdf	09/05/2019 19:02:34	Deborah Ribeiro Carvalho	Aceito

Situação do Parecer:

Aprovado

Necessita Apreciação da CONEP:

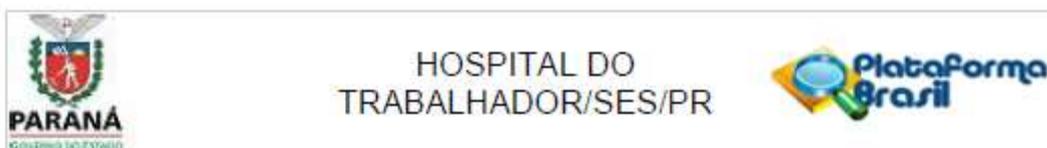
Não

CURITIBA, 12 de Junho de 2019

Assinado por:
NAIM AKEL FILHO
 (Coordenador(a))

Endereço: Rua Imaculada Conceição 1155.
 Bairro: Prado Velho CEP: 80.215-901
 UF: PR Município: CURITIBA
 Telefone: (41)3271-2103 Fax: (41)3271-2103 E-mail: nep@puopr.br

ANEXO B – APROVAÇÃO DO COMITÊ DE ÉTICA DA SESA



PARECER CONSUBSTANCIADO DO CEP

Elaborado pela Instituição Coparticipante

DADOS DO PROJETO DE PESQUISA

Título da Pesquisa: IDENTIFICAÇÃO DE PACIENTES COM RISCO DE SEPSE EM AMBIENTE

Pesquisador: Deborah Ribeiro Carvalho

Área Temática:

Versão: 2

CAAE: 13818519.9.3001.5225

Instituição Proponente: Secretaria de Estado da Saúde do Paraná

Patrocinador Principal: Financiamento Próprio

DADOS DO PARECER

Número do Parecer: 3.534.238

Apresentação do Projeto:

Trata-se de resposta às pendências levantadas no parecer 13818519.9.3001.5225.

Objetivo da Pesquisa:

Já avaliado anteriormente.

Avaliação dos Riscos e Benefícios:

Já avaliados anteriormente.

Comentários e Considerações sobre a Pesquisa:

Já avaliados anteriormente.

Considerações sobre os Termos de apresentação obrigatória:

Foi acrescentado o Termo de Dispensa do TCLE.

O cronograma está atualizado.

Conclusões ou Pendências e Lista de Inadequações:

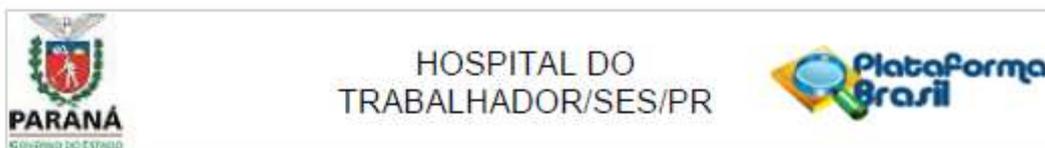
Nenhuma pendência ética significativa.

Considerações Finais a critério do CEP:

Relembramos os autores que devem ser encaminhados como NOTIFICAÇÃO os seguintes assuntos:

- Relatórios parciais (semestrais), se for o caso
- Relatório final

Endereço: Hospital do Trabalhador Avenida República Argentina, 4406 - Novo Mundo - 81.050-000 - Curitiba - PR 41
Bairro: Novo Mundo **CEP:** 81.050-000
UF: PR **Município:** CURITIBA
Telefone: (41)3212-6871 **E-mail:** cepht@sesa.pr.gov.br



Continuação do Parecer: 3.534.238

- O trabalho concluído (publicação, anais de congresso, etc)
- Comunicação de eventos adversos (se houverem)
- Comunicação de início do trabalho e término do mesmo

Alterações no projeto (inclusive em relação à equipe de pesquisa) devem ser submetidos como EMENDA.

Este parecer foi elaborado baseado nos documentos abaixo relacionados:

Tipo Documento	Arquivo	Postagem	Autor	Situação
Informações Básicas do Projeto	PB_INFORMAÇÕES_BÁSICAS_DO_PROJETO_1378374.pdf	19/08/2019 06:40:23		Aceito
Projeto Detalhado / Brochura Investigador	Projeto_Cronograma_alterado_Everton_Osnei_Cesario.docx	05/08/2019 17:37:12	Deborah Ribeiro Carvalho	Aceito
TCLE / Termos de Assentimento / Justificativa de Ausência	DispensaTCLE.pdf	05/08/2019 17:19:15	Deborah Ribeiro Carvalho	Aceito
Outros	SESA.pdf	03/07/2019 06:41:44	Deborah Ribeiro Carvalho	Aceito
Projeto Detalhado / Brochura Investigador	Everton_Osnei_Cesario.docx	09/05/2019 21:24:44	EVERTON OSNEI CESARIO	Aceito

Situação do Parecer:

Aprovado

Necessita Apreciação da CONEP:

Não

CURITIBA, 27 de Agosto de 2019

Assinado por:
FABIO TERABE
(Coordenador(a))

Endereço: Hospital do Trabalhador Avenida República Argentina, 4406 - Novo Mundo - 81.050-000 - Curitiba - PR 41
Bairro: Novo Mundo CEP: 81.050-000
UF: PR Município: CURITIBA
Telefone: (41)3212-6871 E-mail: cepht@sesa.pr.gov.br

ANEXO C – TERMO DE COMPROMISSO DE UTILIZAÇÃO DE DADOS

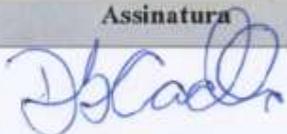
Termo de Compromisso de Utilização de Dados (TCUD)

Nós, Dra^a Deborah Ribeiro de Carvalho, Professora e orientadora do mestrado da PUCPR e Everton Osnei Cesario, funcionário da Celepar, viemos por meio deste, solicitar permissão de utilização dos dados do Sistema GSUS (Sistema de Gestão Hospitalar e Ambulatorial do SUS) para o projeto de mestrado intitulado: IDENTIFICAÇÃO PRECOCE DE PACIENTES COM RISCO DE SEPSE EM AMBIENTE HOSPITALAR, que está sendo desenvolvido em parceria com a Celepar. O objetivo do projeto é desenvolver um modelo de Inteligência Artificial para identificação precoce de pacientes com risco de sepse nos hospitais que utilizam o GSUS, sendo este um projeto sem fins lucrativos e de cunho social.

Comprometemo-nos a manter a confidencialidade sobre os dados coletados nos bancos de dados da Companhia de Tecnologia da Informação e Comunicação do Paraná (Celepar), bem como a privacidade de seus conteúdos, como preconizam os Documentos Internacionais e as Resoluções 466/12 e 510/16 do Conselho Nacional de Saúde.

Os dados coletados correspondem aos pacientes internados, maiores de 18 anos, entre os anos 2011 e 2018, sendo os seguintes: Local de Internamento sendo enfermaria ou UTI, Frequência cardíaca, Idade, Frequência respiratória, Sexo, Pressão sanguínea, Raça, Temperatura, Concentração de lactato, Contagem leucócitos, Data de internamento, Código do paciente (descaracterizado), código dos estabelecimentos de saúde (descaracterizados), Oxigenação do sangue, CID, Procedimento cirúrgico.

Envolvidos na manipulação e coleta dos dados:

Nome completo	CPF	Assinatura
Deborah Ribeiro Carvalho Prof ^a Orientadora	401.595.239-15	
Everton Osnei Cesario Pesquisador	032.947049-33	

Curitiba, 25 de Abril de 2019.

ANEXO D – TERMO DE DISPENSA DO TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO

DISPENSA DO TCLE (TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO)

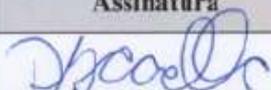
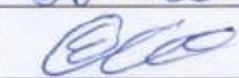
Solicito a dispensa da aplicação do Termo de Consentimento Livre e Esclarecido do projeto de pesquisa intitulado “IDENTIFICAÇÃO PRECOCE DE PACIENTES COM RISCO DE SEPSE EM AMBIENTE HOSPITALAR”, com a seguinte justificativa:

Trata-se de pesquisa retrospectiva com uso de prontuários eletrônicos obtidos por meio do sistema GSUS (Sistema de Gestão Hospitalar e Ambulatorial do SUS) utilizado como sistema de gestão hospitalar do Governo do Estado do Paraná.

Declaro:

- a) Que o acesso aos dados registrados em prontuário de pacientes ou em bases de dados para fins da pesquisa científica será feito somente após aprovação do projeto de pesquisa pelo Comitê de Ética;
- b) O acesso aos dados será supervisionado por uma pessoa que esteja plenamente informada sobre as exigências de confiabilidade;
- c) Assegurar o compromisso com a privacidade e a confidencialidade dos dados utilizados preservando integralmente o anonimato e a imagem do sujeito bem como a sua não estigmatização.
- d) Assegurar a não utilização das informações em prejuízo das pessoas e/ou das comunidades, inclusive em termos de autoestima, de prestígio e/ou econômico-financeiro;
- e) O pesquisador responsável estabeleceu salvaguardas seguras para confidencialidades dos dados de pesquisa;
- f) Os dados obtidos na pesquisa serão usados exclusivamente para finalidade prevista no protocolo;
- g) Os dados obtidos na pesquisa somente serão utilizados para o projeto vinculado;

Devido à impossibilidade de obtenção do TCLE (Termo de Consentimento Livre Esclarecido) de todos os sujeitos, assino este termo para salvaguardar seus direitos.

Nome completo	CPF	Assinatura
Pesquisadora Responsável: Deborah Ribeiro Carvalho	401.595.239-15	
Pesquisador Everton Osnei Cesario	032.947049-33	

Curitiba 27 de Abril de 2019.

**ANEXO E – PROCESSO ADMINISTRATIVO DE SOLICITAÇÃO DE PERMISSÃO
PARA UTILIZAÇÃO DOS DADOS DO GSUS PARA PESQUISA ACADÊMICA**



 ESTADO DO PARANÁ	 DIGITAL	Folha 1
Órgão Cadastro: CELEPAR Em: 29/04/2019 17:30 CNPJ Interessado 1: 76.545.011/0001-19		Protocolo: 15.738.758-8 Vol.: 1
Interessado 1: COMPANHIA DE TECNOLOGIA DA INFORMAÇÃO E COMUNICAÇÃO DO PARANÁ Interessado 2: - Assunto: DOCUMENTAÇÃO E INFORMAÇÃO Palavras chaves: PESQUISA CIENTÍFICA Nº/Ano Documento: 204/2019 Complemento: PERMISSÃO PARA UTILIZAÇÃO DOS DADOS DO SISTEMA GSUS PARA PESQUISA ACADÊMICA	Cidade: CURITIBA / PR Origem: CELEPAR	
Código TTD: -	Para informações acesse: www.eprotocolo.pr.gov.br/consultapublica	



Ofício de Solicitação

À Secretaria de Saúde do Estado do Paraná

Nós, Dra^a Deborah Ribeiro de Carvalho, Professora e orientadora do mestrado da PUCPR e Everton Osnei Cesario, funcionário da Celepar, viemos por meio deste, solicitar permissão de utilização dos dados do Sistema GSUS (Sistema de Gestão Hospitalar e Ambulatorial do SUS) para o projeto de mestrado intitulado: IDENTIFICAÇÃO PRECOCE DE PACIENTES COM RISCO DE SEPSE EM AMBIENTE HOSPITALAR, que está sendo desenvolvido em parceria com a Celepar. O objetivo do projeto é desenvolver um modelo de Inteligência Artificial para identificação precoce de pacientes com risco de sepse nos hospitais que utilizam o GSUS, sendo este um projeto sem fins lucrativos e de cunho social.

Comprometemo-nos a manter a confidencialidade sobre os dados coletados nos bancos de dados da Companhia de Tecnologia da Informação e Comunicação do Paraná (Celepar), bem como a privacidade de seus conteúdos, como preconizam os Documentos Internacionais e as Resoluções 466/12 e 510/16 do Conselho Nacional de Saúde.

Os dados coletados correspondem aos pacientes internados, maiores de 18 anos, entre os anos 2011 e 2018, sendo os seguintes: Local de Internamento sendo enfermaria ou UTI, Frequência cardíaca, Idade, Frequência respiratória, Sexo, Pressão sanguínea, Raça, Temperatura, Concentração de lactato, Contagem leucócitos, Data de internamento, Código do paciente (descaracterizado), código dos estabelecimentos de saúde (descaracterizados), Oxigenação do sangue, CID, Procedimento cirúrgico.

Envolvidos na manipulação e coleta dos dados:

Nome completo	CPF	Assinatura
Deborah Ribeiro Carvalho Prof ^a Orientadora	401.595.239-15	
Everton Osnei Cesario Pesquisador	032.947049-33	

Curitiba, 25 de Abril de 2019.



OF. 204/2019 - DTI

Curitiba, 26 de abril de 2019

Assunto: Permissão para utilização dos dados do Sistema GSUS (Sistema de Gestão Hospitalar e Ambulatorial do SUS) para pesquisa acadêmica.

À Secretaria de Saúde do Estado do Paraná

Conforme solicitação anexa, a Companhia de Tecnologia da Informação e Comunicação do Paraná – Celepar solicita a esta Secretaria permissão para utilização dos dados do Sistema GSUS (Sistema de Gestão Hospitalar e Ambulatorial do SUS), ao empregado da Celepar, Sr. Everton Osnei Cesario, para o projeto de mestrado intitulado: Identificação precoce de pacientes com risco de sepse em ambiente hospitalar.

Leandro Victorino de Moura
Diretor de Tecnologia da Informação e Comunicação

Ao Senhor
Nestor Werner Junior
Diretor-Geral
Secretaria da Saúde do Paraná
Nesta Capital



**SECRETARIA DA SAÚDE
DIRETORIA GERAL**

Protocolo: 15.738.758-8
Assunto: Permissão para utilização dos dados do sistema GSUS para pesquisa acadêmica.
Interessado: Companhia de Tecnologia da Informação e Comunicação do Paraná
Data: 08/05/2019 09:55

DESPACHO

À SUP

Para análise e manifestação quanto à solicitação, se há possibilidade da utilização dos dados para pesquisa, tendo em vista se tratar de dados de pacientes hospitalizados, a qual essa superintendência tem gerência. Após retorne-se, para deliberação do Diretor Geral.

Atenciosamente,

Ana Luiza Zaninelli Ferreira
Téc. Administrativo
Diretoria Geral



**SECRETARIA DA SAÚDE
SUPER.UNIDADES HOSPIT.PROPRIAS**

Protocolo: 15.738.758-8
Assunto: Permissão para utilização dos dados do sistema GSUS para pesquisa acadêmica.
Interessado: Companhia de Tecnologia da Informação e Comunicação do Paraná
Data: 09/05/2019 12:42

DESPACHO

Esta Superintendência entende e apoia o desenvolvimento de Pesquisas Científicas para melhoria dos processos, contudo, por se tratar de um Sistema de Gestão Hospitalar e Ambulatorial do SUS (GSUS) - Prontuário Eletrônico, somente poderá iniciar a coleta da pesquisa pretendida após exarado pelo Comitê de ética em Pesquisa.

Diante do exposto e considerando por parte dos pesquisadores o comprometimento de manter a confiabilidade dos dados coletados bem como a privacidade de seus conteúdos como preconizam as Resoluções 466/12 e 510/16 do Conselho Nacional de Saúde esta Superintendência se faz favorável a coleta dos dados após aprovação pelo Comitê.

Autorizamos os pesquisadores a coletar os dados referido às fls 03, como consta em processo.



**SECRETARIA DA SAÚDE
DIRETORIA GERAL**

Protocolo: 15.738.758-8
Assunto: Permissão para utilização dos dados do sistema GSUS para pesquisa acadêmica.
Interessado: Companhia de Tecnologia da Informação e Comunicação do Paraná
Data: 24/05/2019 11:08

DESPACHO

À CELEPAR,

Autorizado por esta Diretoria Geral, salientando a informação da Superintendência de Unidades Hospitalares Próprias, de que o projeto deve ser apresentado ao Comitê de Ética em Pesquisa do Hospital do Trabalhador (CEPSH-SESA/HT).
Atenciosamente.

Assinado eletronicamente
Nestor Werner Junior
Diretor Geral