



Aplicação de Machine Learning na predição do prognóstico da COVID-19 em pacientes hospitalizados: uma revisão sistemática

USE OF MACHINE LEARNING METHODS TO PREDICT COVID-19 PROGNOSIS IN HOSPITALIZED PATIENTS: A SYSTEMATIC REVIEW

Alexandre Negrão Pantaleão¹, Carolina Sant' Anna Filipin², Larissa Braga Costa³, Luíza Coimbra Teixeira⁴, Renata Araujo Avendanha⁵, Tainara Lima Fernandes⁶, Juliano Gaspar⁷, Zilma Silveira Nogueira Reis⁸

¹ Graduando. Universidade Federal de Minas Gerais.
ORCID: <https://orcid.org/0000-0002-9008-5751>
Email: ale_pantacor@hotmail.com

² Graduanda. Universidade Federal de Minas Gerais.
ORCID: <https://orcid.org/0000-0003-1191-8872>
Email: carolinafilipin@gmail.com

³ Graduanda. Universidade Federal de Minas Gerais.
ORCID: <https://orcid.org/0000-0002-2023-8097>
Email: laribage@gmail.com

⁴ Graduanda. Universidade Federal dos Vales do Jequitinhonha e Mucuri.
ORCID: <https://orcid.org/0000-0001-9708-2951>
Email: luiza.coimbrat@gmail.com

⁵ Graduanda. Universidade Federal de Minas Gerais.
ORCID: <https://orcid.org/0000-0002-4983-3527>
Email: renataavendanha@gmail.com

⁶ Graduanda. Universidade Federal de Minas Gerais.
ORCID: <https://orcid.org/0000-0002-2161-9705>
Email: tainaralimafernandes@gmail.com

⁷ Doutor. Universidade Federal de Minas Gerais.
ORCID: <https://orcid.org/0000-0003-0670-9021>
Email: julianogaspar@gmail.com

⁸ Doutora. Universidade Federal de Minas Gerais.
ORCID: <https://orcid.org/0000-0001-6374-9295>
Email: zilma.medicina@gmail.com

Correspondência: Zilma Silveira Nogueira Reis, Faculdade de Medicina, Universidade Federal de Minas Gerais, Av. Professor Alfredo Balena, 190, sala 60, Belo Horizonte, Minas Gerais, Brazil. CEP: 30.130.100.

Copyright: Esta obra está licenciada com uma Licença Creative Commons Atribuição-Não Comercial 4.0 Internacional.

Conflito de interesses: Os autores declaram que não há conflito de interesses.

Como citar este artigo

Pantaleão, AN; Filipin, CSA; Costa, LB; Teixeira, LC; Avendanha, RA; Fernandes, TL; Gaspar J; Reis, ZSN. Aplicação de Machine Learning na predição do prognóstico da COVID-19 em pacientes hospitalizados: uma revisão sistemática. Revista de Saúde Digital e Tecnologias Educacionais. [Online], volume 7, n. 1. Editor responsável: Luiz Roberto de Oliveira. Fortaleza, fluxo contínuo, de 2022, p. 01-15. Disponível em: <http://periodicos.ufc.br/resdite/index>. Acesso em "dia/mês/ano".

Data de recebimento do artigo: 13/10/2021

Data de aprovação do artigo: 11/02/2022

Data de publicação: 14/03/2022

Resumo

Introdução: A pandemia da COVID-19 tem exaurido recursos humanos e financeiros dos sistemas de saúde. Assim, a utilização de inteligência artificial no atendimento dos doentes pode ser uma estratégia eficaz na pandemia. **Objetivo:** Analisar o uso de *Machine Learning* (ML) para prever morte, internação em UTI e utilização de ventilação mecânica (VM) em pacientes hospitalizados com COVID-19. **Método:** Revisão sistemática seguindo o PRISMA. Bases consultadas: PUBMED, SCIELO, IEEE, COCHRANE, BVS e SCOPUS. Foram incluídos: estudos primários; COVID-19 confirmado por RT-PCR; pacientes hospitalizados; utilização de ML para prever um dos prognósticos pré-definidos. Foram excluídos simulações, estudos de pacientes com comorbidades específicas e estudos sem número de pacientes. **Resultados:** 18 estudos foram incluídos na revisão, sendo que alguns artigos analisaram os desfechos morte, internação em UTI e utilização de VM separadamente e outros avaliaram os desfechos combinados. Entre os artigos, obtiveram-se 22 valores de área sob a curva (AUC), sendo maior e menor valor: 1 e 0.66. As técnicas de ML utilizaram critérios clínicos, laboratoriais e/ou de imagem. **Conclusão:** Os modelos utilizados apresentaram bons resultados. O ML pode auxiliar na predição do desfecho do paciente hospitalizado com COVID-19, melhorando assistência e alocação de recursos.

Palavras-chave: COVID-19. *Machine learning*. Prognóstico.

Abstract

Introduction: The COVID-19 pandemic has depleted human and financial resources from health systems. Thus, the use of artificial intelligence in patient care can be an effective strategy in the pandemic. **Objective:** To analyze the use of *Machine Learning* (ML) to predict death, ICU admission and use of mechanical ventilation (MV) in patients hospitalized with COVID-19. **Method:** Systematic review following PRISMA. Databases consulted: PUBMED, SCIELO, IEEE, COCHRANE, BVS and SCOPUS. Were included: primary studies; COVID-19 confirmed by RT-PCR; hospitalized patients; use of ML to predict one of the predefined prognoses. Simulations, studies of patients with specific comorbidities and studies without the number of patients were excluded. **Results:** 18 studies were included in the review, with some articles analyzing the outcomes of death, ICU admission and use of MV separately, and others evaluating the combined outcomes. Among the articles, 22 values of area under the curve (AUC) were obtained, with the highest and lowest values: 1 and 0.66. The ML techniques used clinical, laboratory and/or imaging criteria. **Conclusion:** The models used showed good results. ML can help predict the outcome of patients hospitalized with COVID-19, improving care and resource allocation.

Keywords: COVID-19. *Machine learning*. Prognostic.

1. Introdução

A pandemia pelo novo coronavírus SARS-CoV-2 impactou rapidamente e profundamente populações e sistemas de saúde em todo o mundo. Até fevereiro de 2022, foram registrados mais de 410 milhões de casos e mais de 5,8 milhões de mortes em todo o mundo, com mais de 27 milhões de casos registrados somente no Brasil¹. Estima-se, no país, uma letalidade geral de 2,3%, mas atinge valores de 59% em pacientes internados em unidades de terapia intensiva (UTI) e de 80% em pacientes em uso de ventilação mecânica (VM)^{2,3}.

Diante da ausência de tratamento farmacológico efetivamente curativo e da vacinação ainda em curso, o eficiente manejo de pacientes com doença moderada e crítica representa

um desafio aos centros hospitalares. Perante a alta demanda e os recursos limitados, a triagem correta da gravidade do quadro dos pacientes com COVID-19 pode ser fator determinante para evitar o colapso dos serviços de saúde. Nesse sentido, o uso de sistemas de *Machine Learning* (ML) como possíveis ferramentas de predição de prognóstico em pacientes com infecção pelo SARS-CoV-2 tem sido proposto.

O *Machine Learning*, ou o “aprendizado de máquina”, é uma tecnologia pertencente à inteligência artificial (IA), em que os computadores são capazes de aprender sem que haja uma programação prévia destes⁴. O ML apresenta dois tipos principais de operacionalização: o aprendizado supervisionado e o aprendizado não supervisionado⁵. O primeiro necessita ser exposto a um modelo de entrada e de saída de dados, conforme os padrões que se esperam que sejam reconhecidos. Já o aprendizado não supervisionado procura por si mesmo determinados padrões nos dados fornecidos. Nesse sentido, o ML tem sido testado atualmente na pandemia do Sars-CoV-2 como ferramenta de celeridade dos processos⁶. Ao predizer possíveis prognósticos e a gravidade de pacientes com COVID-19 a ferramenta pode ajudar a direcionar os cuidados mais adequados a cada paciente e otimizar a utilização de recursos. Nesse sentido, a tecnologia poderia impactar diretamente na tomada de decisão e na otimização do manejo de recursos disponíveis.

Este estudo tem como objetivo analisar a literatura acerca do impacto da utilização de metodologias baseadas no ML na predição dos prognósticos de morte, internação em UTI e uso de ventilação mecânica em pacientes hospitalizados com COVID-19.

2. Métodos

2.1 Estratégia de Busca e Critérios de Seleção

Foi realizada uma revisão sistemática seguindo a metodologia PRISMA⁷. As bases de dados consultadas foram PUBMED, SCIELO, IEEE, COCHRANE, BVS e SCOPUS no período de 18 e 24 de março de 2021. As estratégias de busca foram: (*COVID-19 or SARS-CoV2 or coronavirus*) and (*machine learning or deep learning*) and (*prognosis or death or mechanical ventilation or Intensive Care Unit*) (Apêndice 1). Os resultados foram restringidos aos anos de 2020 e 2021, publicados nas línguas inglesa e/ou portuguesa, devidamente aprovados e publicados em sua versão final.

Os critérios de inclusão determinados foram: 1. diagnóstico de COVID-19 confirmado por RT-PCR; 2. pacientes hospitalizados; 3. uso de *Machine Learning* para predição de prognóstico; 4. análise dos prognósticos de morte e/ou internação em UTI e/ou uso de ventilação mecânica e 5. estudos primários. Os critérios de exclusão foram estudos: 1. de simulação; 2. em que a amostra era composta exclusivamente por pacientes com

comorbidades específicas; 3. que não informaram o número de pacientes. A partir disso, os estudos foram agrupados de acordo com o prognóstico avaliado para uma análise mais crítica dos achados.

Os artigos foram distribuídos aleatoriamente em três blocos e cada um deles foi analisado por pares independentes, por meio da leitura de título e resumo dos artigos. As discordâncias foram encaminhadas posteriormente a um terceiro avaliador. Os selecionados para leitura completa foram novamente divididos em três blocos aleatórios e analisados por pares. O protocolo de busca e de extração dos dados foi publicado na Base de Protocolos on-line Protocols.io em junho de 2021⁸.

2.2 Análise dos Dados

Para a análise dos dados, 3 pares de pesquisadores independentes realizaram a leitura e extração das informações previamente estabelecidas. As discrepâncias foram analisadas e solucionadas por meio de consenso pelos investigadores.

Os dados extraídos foram os seguintes: nome do Autor, ano de publicação e país de realização; Delineamento do estudo; Cenário do estudo, se é multicêntrico, o tipo de hospital e o local do hospital em que os pacientes estavam; Perfil dos pacientes da amostra, com número total, idade média, sexo e presença de comorbidades; Tecnologia de ML utilizada; Momento da história clínica em que a intervenção foi realizada; Tempo médio de *follow-up* dos pacientes; Análise do resultado das tecnologias de ML utilizadas, com extração da sensibilidade, especificidade e acurácia - ou área sob a curva ROC (AUC); Tipo de variável analisada no estudo, sendo exames de imagem e/ou parâmetros laboratoriais e/ou parâmetros clínicos e Prognóstico avaliado pelo estudo, sendo ele admissão em UTI e/ou morte e/ou necessidade de ventilação mecânica.

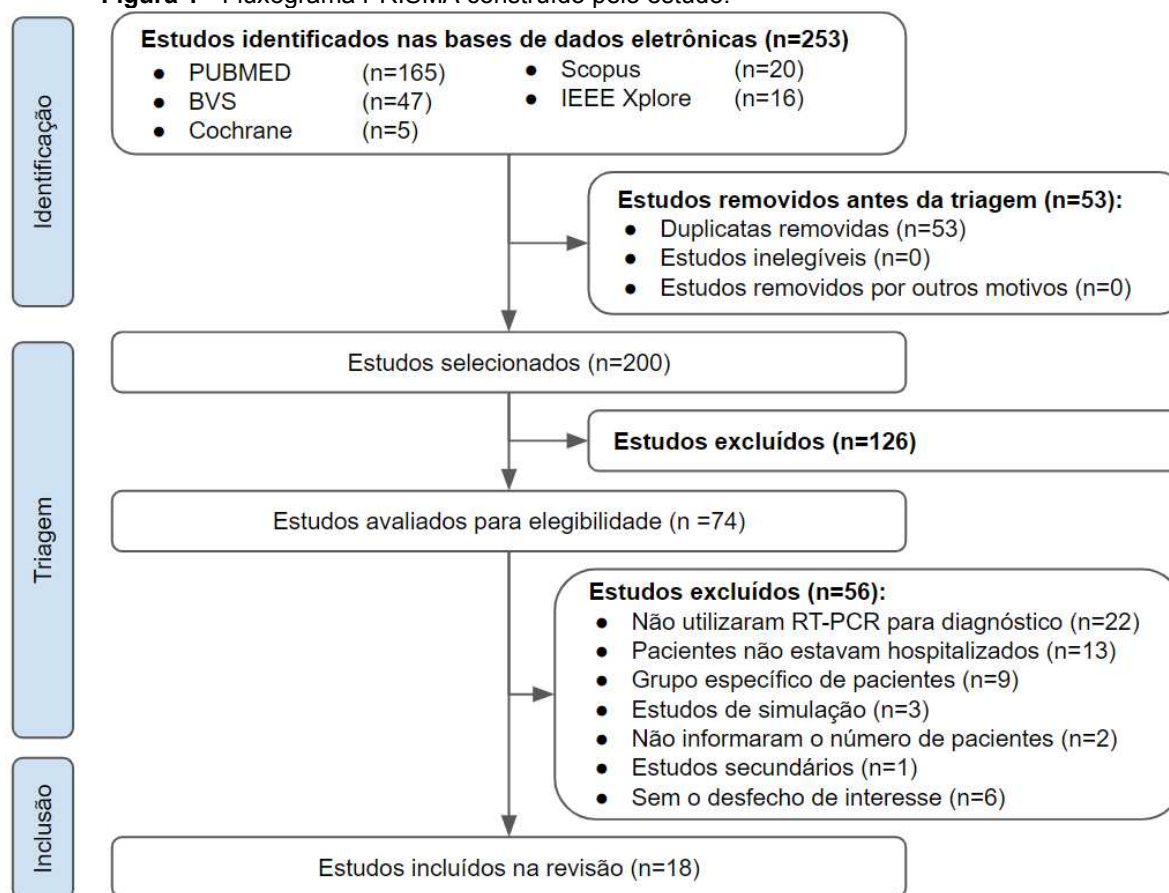
2.3 Avaliação da qualidade do estudo

A qualidade metodológica dos estudos foi avaliada por dois revisores independentes usando o *checklist de Downs and Black* proposto pela Colaboração Cochrane^{9,10}. O *checklist* utilizado foi adaptado para a revisão sistemática proposta neste estudo, conforme sugestão dos autores do *checklist Downs and Black*⁹. A ferramenta adaptada constituiu-se de 22 questões, sendo que 8 avaliaram a qualidade do relato, 3 avaliaram a validade externa, 5 avaliaram a presença de viés, 4 avaliaram o viés de seleção ou variáveis de confusão e, por fim, 2 que avaliaram o poder ou probabilidade da casualidade do resultado final. O questionário aplicado pode ser observado no apêndice 2.

3. Resultados

Inicialmente foram encontrados 253 artigos, sendo 53 duplicatas excluídas por meio da ferramenta *Mendeley*. Os 200 artigos resultantes foram submetidos à leitura de título e resumo. Fazendo uso dos critérios de inclusão e exclusão definidos, 79 artigos foram selecionados para a leitura completa. Por fim, foi realizada a seleção final e extração dos dados desejados para análise de 18 artigos. A análise dos resultados foi feita considerando os diferentes desfechos de cada estudo, conforme é apresentado no fluxograma PRISMA (Figura 1).

Figura 1 - Fluxograma PRISMA construído pelo estudo.



Fonte:

Em relação ao desfecho morte, foram encontrados 9 trabalhos, cujos dados extraídos estão na tabela 1. O estudo de Fernandes *et al.* apresentou o maior valor de sensibilidade (0.964), pela tecnologia *Extra Trees*¹¹. A melhor especificidade (0.9563) foi observada pela técnica *Deep Neural Network* em um estudo realizado por Naseem *et al.*¹². A melhor acurácia/AUC foi de 1.0, obtido no trabalho de Guan *et al.* usando a tecnologia *XGBoost*¹³. Todos os estudos aplicaram a tecnologia no momento da admissão hospitalar pela COVID-

19, excetuando-se Fernandes *et al.*, que a realizou entre as 24 horas antes e depois do diagnóstico hospitalar¹¹.

Tabela 1: Estudos que avaliam técnicas de ML para predição do prognóstico de morte em pacientes hospitalizados com COVID-19

Artigo	Seguimento	Pacientes	Técnicas de ML	Sensibilidade e Especificidade	ROC
<i>Fernandes, Fernando et al., 2021</i>	SI	1040	<i>Extra Trees e Random Forest</i> ^{a+b}	0.97 e 0.86	0.97
<i>Guan, Xin et al., 2021</i>	SI	1270	<i>XGBoost</i> ^{a+b}	0.86 e SI	1.00
<i>Heldt, Frank et al., 2021</i>	Período de internação	879	<i>Logistic regression, Random forest, XGBoost</i> ^{a+b}	0.76 e 0.67	0.77
<i>Ikemura, Kenji et al., 2021</i>	30 Dias*	4313	<i>AutoML(GBM, XGBoost, GLM, DL) e Random Forest</i> ^{a+b}	0.84 e 0.84	0.79 **
<i>Li, Xiaoran et al., 2020</i>	SI	1108	<i>Deep Neural Network (Random Forest)</i> ^{a+b+c}	0.75 e 0.87	0.84
<i>Mushtaq, Junaid et al., 2021</i>	SI	697	<i>Deep Learning com qXR baseado em CNNs</i> ^c	0.77 e 0.59	0.66
<i>Naseem, Maleeha et al., 2021</i>	24 horas	1214	<i>Deep Neural Network</i> ^{a+b}	0.89 e 0.96	0.88
<i>Shiri, Isaac et al., 2021</i>	Período de internação	152	<i>MRMR e XGBoost</i> ^{a+b+c}	0.88 e 0.89	0.95
<i>Woo, Sang et al., 2021</i>	14 dias	415	<i>Scikit-Learn</i> ^{a+b}	0.80 e SI	0.81

Legenda: ROC = área sobre a curva ROC (acurácia), SI = Sem informação, *= até a alta ou o óbito, ** = AUPRC (Area Under Precision Recall Curve),^a = Parâmetros clínicos, ^b = Parâmetros laboratoriais, ^c = Exames de imagem, SI= sem informação.

Fonte:

Considerando o desfecho de internação em UTI, todos apresentaram sensibilidade igual ou superior a 0,72, especificidade igual ou superior a 0,709 e AUC/acurácia igual ou superior a 0,728. À exceção de Fernandes *et al.*, que usou o momento da história clínica entre 24 horas antes e depois do diagnóstico hospitalar para aplicar o ML, os demais artigos utilizaram a tecnologia durante a admissão hospitalar pela COVID-19¹¹. Os resultados relacionados à internação em UTI estão descritos na tabela 2.

Tabela 2. Estudos que avaliam técnicas de ML para predição do prognóstico de internação em UTI em pacientes hospitalizados com COVID-19

Artigo	Seguimento	Pacientes	Técnicas de ML	Sensibilidade e Especificidade	ROC
<i>Burian, Egon et al., 2020</i>	SI	65	<i>Random Forest</i> ^{a+b+c}	0.72 e 0.86	0.80*
<i>Cai, Wenli et al., 2020</i>	Período de internação	99	<i>Random Forest</i> ^{a+b+c}	0.88 e 0.84	0.86*
<i>Durhan, Gamze et al., 2020*</i>	Até definição do desfecho	90	<i>Deep learning</i> (técnica não especificada) ^{a+b+c}	0.86 e 0.87	0.94
<i>Fernandes, Fernando et al., 2021</i>	SI	1040	<i>Extra Trees</i> e <i>Random forest</i> ^{a+b}	0.91 e 0.87	0.96
<i>Heldt, Frank et al., 2021</i>	Período de internação	879	<i>Logistic regression, Random forest, XGBoost</i> ^{a+b}	0.75 e 0.80	0.84
<i>Li, Xiaoran et al., 2020</i>	SI	1108	<i>Deep Neural Network (Random Forest)</i> ^{a+b+c}	0.76 e 0.71	0.73
<i>Statsenko, Yauhen et al., 2021</i>	Até 2 resultados de RT-PCR negativos	560	<i>AdaBoost, Gradient Boosting, Random Forest, Extra Trees</i> ^b	0.98 e 0.75	0.90

Legenda: *Tecnologia de ML baseada na avaliação dos exames de imagem, SI= sem informação, ^a= Parâmetros clínicos, ^b= Parâmetros laboratoriais, ^c= Exames de imagem, ROC = área sobre a curva ROC.

Fonte:

A mesma análise foi feita considerando o desfecho utilização de ventilação mecânica e seus resultados encontram-se na tabela 3. Em relação a este desfecho, também se destacou o trabalho de *Fernandes et al.*, que obteve melhor acurácia, 0,945, e como sensibilidade e especificidade, 0,906 e 0,819 respectivamente¹¹. O ML foi utilizado no momento da admissão hospitalar pela COVID-19 por *Bolourani et al.*, *Ferrari et al.*, *Heldt et al.* e *Zhong et al.*¹⁴⁻¹⁷. As primeiras 12 horas de internação foi o momento da história clínica usado por *Rodriguez* e colaboradores¹⁸.

Alguns estudos também analisaram os desfechos de maneira combinada, de forma que dois ou mais desfechos de interesse fossem considerados como apenas um prognóstico. Foram encontradas as seguintes combinações (Tabela 4): 1. morte e internação em UTI e 2. morte, internação em UTI e utilização de ventilação mecânica. Nessa categoria, a AUC variou de 0.77 a 0.85. O uso da técnica de ML para definição do prognóstico foi feita com dados clínicos, laboratoriais e/ou de imagem obtidos no momento da admissão hospitalar por *Mushtaq et al.* e por *Woo et al.*^{19,20}. Já *Schöning* considerou os dados obtidos 72 horas antes e 24 horas após o diagnóstico hospitalar²¹.

Tabela 3. Estudos que avaliam técnicas de ML para predição do prognóstico de utilização de ventilação mecânica em pacientes hospitalizados com COVID-19

Artigo	Seguimento	Pacientes	Técnica de ML	Sensibilidade e Especificidade	ROC
<i>Fernandes, Fernando et al., 2021</i>	SI	1040	<i>Extra Trees e Random forest</i> ^{a+c}	0.91 e 0.82	0.94
<i>Ferrari, Davide et al., 2020</i>	Primeiras 48h	198	<i>LightGBM suite</i> (árvore de decisão) ^{a+b}	0.73 e 0.82	0.84
<i>Heldt, Frank et al., 2021</i>	Período de internação	879	<i>Logistic regression, Random forest, XGBoost</i> ^{a+b}	0.82 e 0.85	0.87
<i>Rodriguez, Victor et al., 2021*</i>	SI	2256	<i>Regressão logística com L1-penalizado (logistic L1), Regressão logística elástica (logistic EN) e gradient boosted trees (GBT)</i> ^{a+b}	SI	0.74
<i>Zhong, Aoxiao et al., 2021</i>	Primeiras 72h	4119	<i>Rede neural</i> ^{a+b+c}	SI e SI	0.91

Legenda: *Considerando resultados estatísticos do cenário de validação, SI= sem informação, ^a= Parâmetros clínicos, ^b= Parâmetros laboratoriais, ^c= Exames de imagem, ROC = área sobre a curva ROC (acurácia).

Fonte:

Tabela 4. Estudos que avaliam técnicas de ML para predição de prognósticos combinados de morte e internação em UTI em pacientes hospitalizados com COVID-19

Artigo	Seguimento	Pacientes	Técnica de ML	Sensibilidade e Especificidade	ROC	VM
<i>Mushtaq, Junaid et al., 2021</i>	Período de internação	697	<i>Deep Learning com qXR baseado em CNNS</i> ^c	0.77 e 0.59	0.77	Não
<i>Schöning, Verena et al., 2021</i>	SI	657	<i>Support Vector Machines, Random Forest e outros</i> ^{a+b}	SI	0.85	Não
<i>Woo, Sang et al., 2021</i>	14 dias	415	<i>Scikit-Learn</i> ^{a+b}	0.83 e 0.60	0.82	Sim

Legenda: ROC= área sobre a curva ROC (acurácia), VM= Ventilação Mecânica, SI= Sem informação, ^a= dados laboratoriais; ^b= dados clínicos; ^c= exames de imagem.

Fonte:

Dos 18 estudos avaliados, 6 eram multicêntricos. Ao todo, 16 eram estudos retrospectivos e 2 coortes prospectivas. Foi possível observar que 17 dos 18 estudos têm

predomínio de pacientes do sexo masculino. Além disso, 13 estudos informaram quais comorbidades dos pacientes foram investigadas (Apêndice 3).

Quanto à análise de qualidade dos estudos, todos atingiram pelo menos 60% da pontuação total, indicando alto teor de qualidade. O melhor critério observado entre os estudos se refere ao relato do estudo, provavelmente devido à similaridade aos critérios de inclusão e exclusão da presente revisão sistemática. O pior critério dos artigos considerados foi a validação externa, provavelmente devido ao fato de que a maioria dos estudos considera pacientes com COVID-19 de um determinado hospital, não sendo uma situação representativa das populações em questão. Nesse sentido, difere-se o estudo de Bolourani *et al.*, que teve o tamanho da amostra de 11525 pacientes, uma população com uma discrepância elevada comparada aos outros estudos e que, por analisarem um banco de dados, puderam ser considerados representativos da população¹⁴. O trabalho de Ferrari *et al.* obteve um dos piores desempenho dos estudos considerados, considerando que a população utilizada não foi suficiente para executar um estudo confiável e indutivo¹⁵.

Tabela 5. Análise de qualidade dos estudos selecionados

Estudo	Relato	Validade Externa	Viés	Variável de confusão	Poder	Total
Bolourani, S., <i>et al.</i> 2021	★★★★★★	★★★	★★★★★	★★★★	★	21/22
Burian, E., <i>et al.</i> 2020	★★★★★★	★	★★	★★	★	13/22
Cai, W., <i>et al.</i> 2020	★★★★★★	★	★★★	★★★	★	16/22
Durhan, G., <i>et al.</i> 2020	★★★★★★	-	★★★★	★★★	★	16/22
Fernandes, F., <i>et al.</i> 2021	★★★★★★	★	★★★★	★★	★★	16/22
Ferrari, D., <i>et al.</i> 2020	★★★★★★	★	★★★	★★★	-	14/22
Guan, X., <i>et al.</i> 2021	★★★★★★	★	★★★★	★★★	★	17/22
Heldt, F. S., <i>et al.</i> 2021	★★★★★★	★	★★★★★	★★	★	17/22
Ikemura, K., <i>et al.</i> 2021	★★★★★★	★	★★★★	★★★	★	17/22
Li, X., <i>et al.</i> 2020	★★★★★★	★	★★★★	★★★★	★★	19/22
Mushtaq, J., <i>et al.</i> 2020	★★★★★★	★	★★★★	★★★	★	17/22
Naseem, M., <i>et al.</i> 2021	★★★★★★	★	★★★★	★★★	★★	18/22
Rodriguez, V., <i>et al.</i> 2021	★★★★★★	★	★★★★★	★★★	★★	18/22
Schöning, V., <i>et al.</i> 2021	★★★★★★	★	★★★★★	★★★	★★	18/22
Shiri, I., <i>et al.</i> 2021	★★★★★★	★	★★★★	★★★	★★	17/22
Statsenko, Y., <i>et al.</i> 2021	★★★★★★	★	★★★★	★★★	★★	17/22

Woo, S., <i>et al.</i> 2021	★★★★★★★	★	★★★	★★★	★★	17/22
Zhong, A., <i>et al.</i> 2021	★★★★★★★	★★★	★★★★★	★★	★	18/22

Fonte:

3.1 Discussão

A partir da análise dos resultados obtidos, observou-se, no geral, que os modelos de ML analisados apresentaram resultados promissores na análise das métricas de acurácia, sensibilidade e especificidade.

Em relação ao desfecho morte, Guan *et al.* utilizaram o modelo *XGBoost* e obtiveram a melhor acurácia¹³, enquanto Fernandes *et al.* demonstraram os melhores resultados na predição de utilização de ventilação mecânica e internação em UTI, separadamente, com os modelos *Extra Trees* e *Random Forest*¹¹.

As variáveis mais utilizadas para entrada de dados nos softwares de ML foram provenientes de informações laboratoriais e clínicas, combinação usada por 13 dos artigos incluídos, seguida da associação de dados laboratoriais, clínicos e exames de imagem. Dentre os artigos que combinaram achados de laboratório e da clínica, a maior sensibilidade encontrada foi de 0.964¹¹, e a maior especificidade de 0.994¹⁴. Já os estudos que utilizaram os três conjuntos de variáveis obtiveram valores de sensibilidade que variaram de 0.720 a 0.884 e valores de especificidade de 0.709 a 0.890. A maior sensibilidade encontrada dentre todos os artigos, de 0.9754 por Statsenko *et al.* para predição da necessidade de internação em UTI, usou apenas informações provenientes de testes laboratoriais²². A alta sensibilidade presente nas tecnologias que utilizaram marcadores séricos aponta para a relevância dos exames laboratoriais na avaliação prognóstica da COVID-19, sendo os mais comuns: dosagens de proteína C reativa, lactato desidrogenase, D-dímero e troponina. Apesar de usualmente serem não específicos e terem pouco valor diagnóstico, esses testes apresentam relevante valor clínico ao auxiliarem na avaliação da gravidade da doença²³. A maioria dos artigos analisados aplicaram a tecnologia de ML no momento de admissão hospitalar dos pacientes, demonstrando, assim, capacidade de predição do prognóstico de maneira precoce, com uso de testes comumente disponíveis.

Na avaliação do desfecho morte, a maior acurácia encontrada foi de 1, por Guan *et al.*, seguida de acurácia de 0.97 obtida por Fernandes *et al.*^{11,13}, sendo as suas amostras de 554 e 1040 pacientes, respectivamente. Essa diferença amostral aponta um possível interferente para o resultado de acurácia mais elevada. Apenas um artigo usou exclusivamente exames de imagem para a predição, e obteve a menor acurácia encontrada na revisão, de 0.66. Nesse caso, foram computados pelo ML somente dados da extensão

de consolidações e de opacidades pulmonares observadas em radiografias de tórax. Ademais, o estudo foi realizado na região italiana de Lombardia, com pacientes admitidos entre fevereiro e abril de 2020, contexto de alto número de pacientes graves e quantidade expressiva de óbitos, durante o início da pandemia da COVID-19 na Itália¹⁹. Possivelmente, o viés da amostra e a limitação dos dados embutidos no software do ML contribuíram para obtenção de menor acurácia.

Nos trabalhos cujo desfecho em análise foi a internação em UTI, a melhor acurácia foi de 0.959¹¹. A AUC mais baixa foi de 0.728, obtida em estudo multicêntrico, com 1108 pacientes, nos Estados Unidos²⁴. Neste último, foi também construído um sistema de ML para predição de morte, com obtenção de AUC de 0.844. Sugere-se que a diferença entre os resultados encontrados dentro do mesmo estudo está associada à variabilidade nos critérios de decisão entre os profissionais da saúde para internação de um paciente em UTI. Além disso, Li *et al.* utilizaram uma base de dados independente para a validação do modelo de ML construído, o que pode ter influenciado o resultado quando comparado a outros estudos²⁴.

Na análise da necessidade de ventilação mecânica, as acurácias encontradas variaram de 0.743, no estudo de Rodriguez *et al.*, a 0.945, no estudo de Fernandes *et al.*^{11,18}. Destaca-se o uso de dados provenientes de populações diferentes para desenvolvimento e validação do ML, em estudos não multicêntricos, o que pode ter instigado menores valores de acurácia¹⁸. É provável que, com o uso de amostras distintas, os níveis de gravidade da doença e o recebimento de cuidados de saúde sejam também díspares, influenciando na capacidade de predição da máquina. Tal fato evidencia a realização de estudos multicêntricos, em que as tecnologias são treinadas com uma amostra mais heterogênea, permitindo sua aplicação em contextos clínicos mais diversos.

3.1.1 O que as outras revisões apontam

Ao comparar o presente estudo com outros relacionados à temática da ciência de dados sobre a COVID-19, similaridades e diferenças foram encontradas. A revisão sistemática de Albahri *et al.* avaliou estudos sobre mineração de dados e outras técnicas de machine learning no diagnóstico da infecção pela família Cov²⁵. Utilizando busca em cinco bases bibliográficas, apenas dois^{26,27} dos oito estudos incluídos apresentavam medida de sensibilidade, especificidade e/ou AUC, sendo a melhor AUC foi de 94,2%. Revisões que abordaram o uso da IA para diagnóstico e prognóstico da COVID-19 concordam e divergem com os resultados que aqui apresentamos. Adamidi *et al.*, na métrica de acurácia, trouxeram os valores de precisão, valores preditivos positivo e negativo, limitações éticas e risco de

viés²⁶. Como diferencial, eles excluíram estudos que não traziam dados que não podiam ser aplicados em ambiente clínico e os que afetavam a saúde mental. A melhor AUC foi de 0,953, sensibilidade 0,923 e especificidade 0,851²⁶. A revisão de Figliozzi *et al.* foi a que mais se assemelhou a esta revisão. Teve o objetivo de estabelecer parâmetros para prognóstico de COVID-19 identificados pela presença de pelo menos um dos desfechos: óbito, infecção grave por COVID-19, internação em UTI, uso de ventilação mecânica e progressão da doença e mortalidade hospitalar como desfecho secundário. Infelizmente, não foi apresentada uma análise estatística de acurácia para comparabilidade com a presente revisão. No entanto, utilizando-se da razão de chances, a história de doença cardiovascular, lesão aguda de órgão relacionada ao vírus e aumento da procalcitonina e dímero-D foram os principais preditores de ambos os desfechos²⁷. Um grande fator em comum com todos citados é a utilização do modelo PRISMA para execução da revisão sistemática.

A presente revisão se destaca pelo rigor na padronização recomendada pela metodologia PRISMA quanto à extração de dados estatísticos fornecidos pelos estudos, bem como as informações gerais dos trabalhos incluídos⁷. Além disso, houve elaboração e publicação prévia do protocolo da revisão⁸ e uma ampla seleção de bases de dados e artigos analisados, atingindo uma maior abrangência de publicações.

3.1.2 Limitações

Os artigos analisados apresentavam limitações, uma delas é o número pequeno de pacientes utilizados nas amostras para permitir a generalização e a devida validação externa dos dados obtidos com os algoritmos. Além disso, a grande quantidade de estudos realizados em um único centro, 13 dos 18 artigos, também contribui para a dificuldade de generalização. Outro ponto chave para interpretação dos resultados de acurácia dos estudos levantados é a ausência de protocolos clínicos padronizados de manejo da COVID-19, o que pode enviesar os resultados de acordo com cada abordagem realizada. Destaca-se, ainda, a predominância de estudos retrospectivos, 16 entre os 18 selecionados, aumentando a possibilidade de viés com relação à obtenção dos dados e completude deles. Nesse contexto, dois estudos relataram a utilização de algoritmos para preencher de forma randômica os dados que tinham sido perdidos nos relatórios, fator que pode comprometer os resultados das tecnologias de ML^{14,18}. Outro ponto foi a heterogeneidade quanto às comorbidades dos pacientes na análise pelos algoritmos, variáveis que poderiam ser muito influentes na predição dos desfechos. Por fim, a dificuldade de padronização da segmentação e da análise dos exames de imagem compromete a avaliação desses dados

nos estudos que fazem uso desses parâmetros. Frente a tais discrepâncias, optamos por não realizar a síntese através de técnicas de meta análise, o que pode ser considerado uma limitação da revisão. É também relevante pontuar que o pouco conhecimento acerca da fisiopatologia da infecção por SARS-CoV-2, somada à dinamicidade de novas informações sobre a doença, fez com que, entre o período da seleção dos artigos e o atual momento, novas diretrizes de tratamento e novos estudos sobre o tema tenham surgido, não sendo incorporados na presente revisão.

3.1.3 Implicações clínicas e para o futuro

Os estudos analisados nessa revisão avaliaram positivamente a utilização das tecnologias de ML no contexto clínico e hospitalar, demonstrando que a IA pode ser uma importante estratégia para auxiliar os sistemas de saúde. Entre as principais vantagens dessa utilização, destaca-se a identificação de relações e padrões, não aparentes à percepção humana, em bancos de dados complexos e longitudinais de forma rápida e precisa²⁴. No entanto, esses métodos de aprendizagem geralmente requerem grande tamanho amostral, há potencial de sobreajuste e as relações complexas podem gerar resultados de difícil interpretação²⁴.

No contexto da COVID-19, a aplicação dos modelos de ML mostrou-se especialmente útil, visto que a progressão da doença é imprevisível e as mudanças clínicas são rápidas, sendo o tempo um fator crucial para a adoção de intervenções que previnam a admissão na UTI e a ventilação mecânica¹⁵. No entanto, espera-se que estudos prospectivos controlados e de maior validade possam comprovar esses achados. A composição de dados clínicos nos modelos de prognóstico se mostrou mais efetiva do que os modelos baseados apenas em imagens. Os hospitais com plena informatização de registro poderão se beneficiar mais rapidamente destas novas abordagens, desde que implementadas com responsabilidade e ética. Recomendações internacionais de boas práticas reforçam a necessidade de participação multidisciplinar no desenvolvimento da IA para saúde, para que possam ser compreendidas pelos profissionais de saúde, assim como os demais recursos tecnológicos de apoio à prestação de cuidados²⁸.

4. Conclusão

As tecnologias de ML analisadas apresentaram bons resultados, mostrando-se úteis para prever prognósticos de morte, requisição de ventilação mecânica e internação em UTI em pacientes hospitalizados com COVID-19. Portanto, a utilização da IA pode ser de

grande auxílio aos sistemas de saúde, sobretudo em circunstâncias de elevada demanda por recursos escassos e intervenções rápidas.

Apêndices: Suplementar a este artigo, os apêndices são disponibilizados on-line e podem ser acessados pelo endereço:

https://ftp.medicina.ufmg.br/cins/pesquisa/covid/ML_COVID_Supl.pdf

Referências

1. Hopkins J. COVID-19 Content Portal [Internet]. 2022 [citado 13 de fevereiro de 2022]. Disponível em: <https://systems.jhu.edu/research/public-health/ncov/>
2. Brasil. Painel Coronavírus Brasil - DATASUS [Internet]. 2022 [citado 13 de fevereiro de 2022]. Disponível em: <https://covid.saude.gov.br/>
3. Ranzani OT, Bastos LSL, Gelli JGM, Marchesi JF, Baião F, Hamacher S, et al. Characterisation of the first 250 000 hospital admissions for COVID-19 in Brazil: a retrospective analysis of nationwide data. *Lancet Respir Med*. 1º de abril de 2021;9(4):407–18.
4. Reis ZSN, Gaspar J. Informação e Decisão em Saúde I. Vol. 1. UFMG; 2019. 123 p.
5. Deo RC. Machine Learning in Medicine. *Circulation*. 17 de novembro de 2015;132(20):1920–30.
6. Alimadadi A, Aryal S, Manandhar I, Munroe PB, Joe B, Cheng X. Artificial intelligence and machine learning to fight COVID-19. *Physiol Genomics*. 1º de abril de 2020;52(4):200–2.
7. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JPA, et al. The PRISMA Statement for Reporting Systematic Reviews and Meta-Analyses of Studies That Evaluate Health Care Interventions: Explanation and Elaboration. *PLoS Med*. 21 de julho de 2009;6(7):e1000100.
8. Pantaleao AN. Prediction of COVID-19 prognosis using machine learning techniques: a systematic review. 2 de junho de 2021 [citado 4 de agosto de 2021]; Disponível em: <https://www.protocols.io/view/prediction-of-covid-19-prognosis-using-machine-learning-bu3enyje>
9. Brasil. Diretrizes metodológicas: elaboração de revisão sistemática e metanálise de estudos observacionais comparativos sobre fatores de risco e prognóstico [Internet]. Ministério da Saúde do Brasil; 2014 [citado 11 de agosto de 2021]. 132 p. Disponível em: https://bvsmms.saude.gov.br/bvs/ct/PDF/diretrizes_metodologias_estudos_observacionais.pdf
10. Downs SH, Black N. The feasibility of creating a checklist for the assessment of the methodological quality both of randomised and non-randomised studies of health care interventions. *J Epidemiol Community Health*. junho de 1998;52(6):377–84.
11. Fernandes FT, de Oliveira TA, Teixeira CE, Batista AF de M, Dalla Costa G, Chiavegatto Filho ADP. A multipurpose machine learning approach to predict COVID-19 negative prognosis in São Paulo, Brazil. *Sci Rep*. 8 de fevereiro de 2021;11:3343.
12. Naseem M, Arshad H, Hashimi SA, Irfan F, Ahmed FS. Predicting mortality in SARS-COV-2 (COVID-19) positive patients in the inpatient setting using a Novel Deep Neural Network [Internet]. *Infectious Diseases (except HIV/AIDS)*; 2020 dez [citado 8 de agosto de 2021]. Disponível em: <http://medrxiv.org/lookup/doi/10.1101/2020.12.13.20247254>
13. Guan X, Zhang B, Fu M, Li M, Yuan X, Zhu Y, et al. Clinical and inflammatory features based machine learning model for fatal risk prediction of hospitalized COVID-19 patients: results from a retrospective cohort study. *Ann Med*. 1º de janeiro de 2021;53(1):257–66.
14. Bolourani S, Brenner M, Wang P, McGinn T, Hirsch JS, Barnaby D, et al. A Machine Learning Prediction Model of Respiratory Failure Within 48 Hours of Patient Admission for COVID-19: Model Development and Validation. *J Med Internet Res*. 10 de fevereiro de 2021;23(2):e24246.
15. Ferrari D, Milic J, Tonelli R, Ghinelli F, Meschiari M, Volpi S, et al. Machine learning in predicting respiratory failure in patients with COVID-19 pneumonia—Challenges, strengths,

- and opportunities in a global health emergency. *PLoS ONE*. 12 de novembro de 2020;15(11):e0239172.
16. Heldt FS, Vizcaychipi MP, Peacock S, Cinelli M, McLachlan L, Andreotti F, et al. Early risk assessment for COVID-19 patients from emergency department data using machine learning. *Sci Rep*. 18 de fevereiro de 2021;11:4200.
 17. Zhong A, Li X, Wu D, Ren H, Kim K, Kim Y, et al. Deep metric learning-based image retrieval system for chest radiograph and its clinical applications in COVID-19. *Med Image Anal*. maio de 2021;70:101993.
 18. Rodriguez VA, Bhave S, Chen R, Pang C, Hripcsak G, Sengupta S, et al. Development and validation of prediction models for mechanical ventilation, renal replacement therapy, and readmission in COVID-19 patients. *J Am Med Inform Assoc JAMIA*. 11 de março de 2021;28(7):1480–8.
 19. Mushtaq J, Pennella R, Lavallo S, Colarieti A, Steidler S, Martinenghi CMA, et al. Initial chest radiographs and artificial intelligence (AI) predict clinical outcomes in COVID-19 patients: analysis of 697 Italian patients. *Eur Radiol*. 2021;31(3):1770–9.
 20. Woo SH, Rios-Diaz AJ, Kubey AA, Cheney-Peters DR, Ackermann LL, Chalikonda DM, et al. Development and Validation of a Web-Based Severe COVID-19 Risk Prediction Model. *Am J Med Sci [Internet]*. 23 de maio de 2021 [citado 8 de agosto de 2021]; Disponível em: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8141270/>
 21. Schöning V, Liakoni E, Baumgartner C, Exadaktylos AK, Hautz WE, Atkinson A, et al. Development and validation of a prognostic COVID-19 severity assessment (COSA) score and machine learning models for patient triage at a tertiary hospital. *J Transl Med*. 5 de fevereiro de 2021;19:56.
 22. Statsenko Y, Al Zahmi F, Habuza T, Gorkom KN-V, Zaki N. Prediction of COVID-19 severity using laboratory findings on admission: informative values, thresholds, ML model performance. *BMJ Open*. 26 de fevereiro de 2021;11(2):e044500.
 23. Carpenter CR, Mudd P, West CP, Wilber E, Wilber ST. Diagnosing COVID-19 in the Emergency Department: A Scoping Review of Clinical Exam, Labs, Imaging Accuracy and Biases. *Acad Emerg Med*. 16 de junho de 2020;10.1111/acem.14048.
 24. Li X, Ge P, Zhu J, Li H, Graham J, Singer A, et al. Deep learning prediction of likelihood of ICU admission and mortality in COVID-19 patients using clinical variables. *PeerJ*. 6 de novembro de 2020;8:e10337.
 25. Albahri AS, Hamid RA, Alwan J k., Al-qays ZT, Zaidan AA, Zaidan BB, et al. Role of biological Data Mining and Machine Learning Techniques in Detecting and Diagnosing the Novel Coronavirus (COVID-19): A Systematic Review. *J Med Syst*. 25 de maio de 2020;44(7):122.
 26. Adamidi ES, Mitsis K, Nikita KS. Artificial intelligence in clinical care amidst COVID-19 pandemic: A systematic review. *Comput Struct Biotechnol J*. 7 de maio de 2021;19:2833–50.
 27. Figliozzi S, Masci PG, Ahmadi N, Tondi L, Koutli E, Aimo A, et al. Predictors of adverse prognosis in COVID-19: A systematic review and meta-analysis. *Eur J Clin Invest*. 2020;e13362–e13362.
 28. WHO. Ethics and governance of artificial intelligence for health: WHO guidance [Internet]. World Health Organization; 2021 [citado 11 de agosto de 2021]. 60 p. Disponível em: <https://www.who.int/docs/default-source/documents/gS4dhdaa2a9f352b0445bafbc79ca799dce4d.pdf>