



## Previsão de óbito e importância de características clínicas em idosos com COVID-19 utilizando o Algoritmo Random Forest


Tiago Pessoa Ferreira Lima <sup>1</sup>

 <https://orcid.org/0000-0002-1148-4288>


Maria Julia Gonçalves Mello <sup>6</sup>

 <https://orcid.org/0000-0003-4645-8343>


Gabrielle Ribeiro Sena <sup>2</sup>

 <https://orcid.org/0000-0002-8430-3599>


Flávia Augusta de Orange Lins da Fonseca e Silva <sup>7</sup>

 <https://orcid.org/0000-0003-0528-4164>


Camila Soares Neves <sup>3</sup>

 <https://orcid.org/0000-0001-5477-4296>

Suely Arruda Vidal <sup>4</sup>

 <https://orcid.org/0000-0002-4268-520X>

Jurema Telles Oliveira Lima <sup>5</sup>

 <https://orcid.org/0000-0003-2671-3570>

<sup>1,2,4,5,6</sup> Instituto de Medicina Integral Prof. Fernando Figueira. Rua dos Coelhos, 300, Boa Vista. Recife, PE, Brazil. CEP: 50.070-902. E-mail: gabriellesena8@gmail.com

<sup>3,7</sup> Faculdade Pernambucana de Saúde. Recife, PE, Brazil.

### Resumo

*Objetivos:* treinar um classificador do tipo Random Forest (RF) para estimar o risco de óbito em idosos (com mais de 60 anos) diagnosticados com COVID-19 em Pernambuco. Uma "feature" deste classificador, chamada *feature\_importance*, foi usada para identificar os atributos (principais fatores de risco) relacionados com o desfecho final (cura ou óbito) através do ganho de informação.

*Métodos:* dados dos casos confirmados de COVID-19 foram obtidos entre os dias 13 de fevereiro e 19 de junho de 2020, em Pernambuco, Brasil. O algoritmo K-fold Cross Validation, com K=10, foi usado para avaliar tanto o desempenho do RF quanto a importância das características clínicas.

*Resultados:* o algoritmo RF classificou corretamente 78,33% dos idosos, com AUC de 0,839. A idade avançada é o fator que representa maior risco de evolução para óbito. Além disso, a principal comorbidade e sintoma também identificados, foram, respectivamente, doença cardiovascular e saturação de oxigênio  $\leq 95\%$ .

*Conclusão:* este trabalho se dedicou à aplicação do classificador RF para previsão de óbito e identificou as principais características clínicas relacionadas com este desfecho em idosos com COVID-19 no estado de Pernambuco.

**Palavras-chave** COVID-19, Fatores de risco, Idosos, Random Forest



## Introdução

Desde o início da pandemia pela COVID-19 (doença do coronavírus 2019) até 27 de setembro de 2020, o Brasil, maior país da América do Sul e o quinto maior do mundo, já era considerado o segundo país em número de óbitos pela doença. Até meados de outubro, pelo menos 4.717.991 de brasileiros haviam desenvolvido a infecção e destes, 141.406 evoluíram para óbito.<sup>1</sup> A taxa de letalidade em diversos estados do norte/nordeste brasileiro foi muito maior que a média nacional, especialmente em Pernambuco.<sup>1</sup> Frente a esse cenário epidemiológico, um dos desafios, além da vacina, é a necessidade de orientar as políticas de saúde pública para a vigilância e o controle da doença. Através da identificação dos principais fatores de risco, por exemplo, é possível proporcionar o monitoramento precoce dos grupos mais vulneráveis, diminuindo a chance de evolução para desfechos clínicos desfavoráveis.

Dados extraídos de pacientes com COVID-19 mostram-se como uma fonte valiosa de informações, tanto acerca da fisiopatologia da doença, como dos fatores de risco associados ao óbito. Estes dados têm sido largamente estudados, sendo atualmente consensual que a idade avançada e presença de comorbidades estão associadas ao aumento da morbimortalidade.<sup>2</sup> A disponibilidade abundante desses dados permite a construção de algoritmos de Aprendizagem de Máquina (AM) - uma vertente da Inteligência Artificial - em que é possível identificar pessoas mais suscetíveis com base em características individuais. Através de métodos denominados de Classificação, o algoritmo aprende durante um processo denominado treinamento ao receber um conjunto de entradas (características clínicas) junto com as saídas (desfecho). Finalmente, o algoritmo é capaz de prever uma saída a partir de entradas não vistas durante o treinamento.

Vários algoritmos de AM são amplamente utilizados na construção de modelos preditivos de doenças. O *Random Forest* (RF), particularmente, mostrou maior precisão quando comparados a outros algoritmos.<sup>3</sup> Este tem a capacidade de elencar quais atributos contribuem para a tomada de decisão, sendo utilizado muitas vezes como uma técnica de seleção de características. A seleção de características é considerada uma etapa essencial para análise de dados, pois pode reduzir a complexidade/dimensionalidade do problema.<sup>4</sup> Um conjunto de dados otimizado leva a um modelo mais preciso e melhora também a sua interpretabilidade.<sup>5</sup> Isso é especialmente importante no desenvolvimento de algoritmos para triagem clínica, uma vez que o custo computacional dela deve ser o mais baixo possível e os profissionais de saúde estão interessados nos mecanismos fisiopatológicos subjacentes ao modelo de AM.

cional dela deve ser o mais baixo possível e os profissionais de saúde estão interessados nos mecanismos fisiopatológicos subjacentes ao modelo de AM.

## Conceitos Básicos

Esta seção apresenta conceitos de AM essenciais para entendimento do trabalho.

### Classificador

Dado um conjunto de instâncias, composto por exemplos construídos com valores de atributos bem como a classe associada, um algoritmo de aprendizado (ou indutor) gera como saída um classificador (também denominado hipótese) de forma que, dada uma instância com a classe desconhecida, ele possa rotulá-la. Formalmente, um exemplo é um par  $\{x_i, f(x_i)\}$ , onde  $x_i$  é a entrada (conjunto de atributos) e  $f(x_i)$  é a saída (classe ou rótulo). Seja  $X = \{\{x_1, f(x_1)\}, \{x_2, f(x_2)\}, \dots, \{x_n, f(x_n)\}\}$  um conjunto de  $n$  exemplos, a tarefa do algoritmo de aprendizado é induzir uma função  $h(\cdot)$  que aproxima a função  $f(\cdot)$ . Neste sentido,  $h(\cdot)$  é chamada uma hipótese sobre a função objetivo  $f(\cdot)$ , ou seja,  $h(x_1) \approx f(x_1)$ .

### Árvores de Decisão

São construídas e representadas através de dois elementos: nós e os ramos que conectam nós. Para tomada de uma decisão, o fluxo começa no nó raiz, navega através dos ramos até chegar a um nó folha. Cada nó da árvore denota um teste de um atributo, e os ramos denotam os possíveis valores que o nó pode assumir. Durante o processo de formação da árvore, também conhecido por treinamento ou aprendizado, leva-se em consideração a homogeneidade das classes para cada divisão do nó. Basicamente, o algoritmo avalia o ganho de informação dos atributos para separação das amostras presentes no conjunto de dados destinado ao treinamento.<sup>6</sup> O Gini impurity (GI) é um índice para avaliação de atributos na separação das amostras de mesmo rótulo, ou seja, busca-se a homogeneidade das classes para compor um nó. O GI é definido da Equação 2.1, onde  $p = p_1 \dots p_c$  é a proporção das amostras da classe  $p_c$  para o nó  $m$ , respectivamente. O índice avalia todos os preditores selecionados aleatoriamente para construção da árvore e escolherá aquele que possuir maior grau de homogeneidade entre as amostras. Se o nó  $m$  é puro (homogêneo), então a proporção  $p_i(m)$  da classe  $i$  para o nó  $m$  será igual a 1 e consequentemente o índice igual a 0. O atributo para divisão é escolhido conforme o decremento de pureza apresentado na Equação 2.2, onde a divisão do nó  $m$ ,

$P_{\text{esq}}$  e  $P_{\text{dir}}$ , são as proporções das amostras no nó filho à esquerda e à direita, respectivamente.

$$I_G(m) = 1 - \sum_{i=1}^c p_i(m)^2 \quad (2.1)$$

$$I_G(m) = 1 - \sum_{i=1}^c p_i(m)^2 \quad (2.2)$$

### Algoritmo *Random Forest*

Seja  $H = \{h_1, h_2, h_3\}$  um conjunto ou *ensemble* de três classificadores  $H$ . Se os três classificadores cometem erros distintos, então quando  $h_1(x_i)$  está errado, é possível que  $h_2(x_i)$  e  $h_3(x_i)$  estejam corretos, de forma que a combinação das hipóteses por votação pode classificar corretamente  $x_i$ . O algoritmo de floresta aleatória ou RF baseia-se na estratégia de ensembles. Ele provê a diversidade, lançando mão do conceito de redistribuição aleatória dos dados. Assim, na construção de cada  $h_i \in H$ , para um dado conjunto de treinamento  $\mathcal{I}$ , é gerado um subconjunto de dados  $\mathcal{J}$ . Dessa forma, o algoritmo gera diversas árvores de decisão, cada uma treinada com uma distribuição aleatória. Uma grande qualidade do RF é a facilidade para se medir a importância relativa de cada atributo para a predição. O algoritmo implementado no Sklearn,<sup>8</sup> por exemplo, provê uma excelente ferramenta para isto, que mede a importância das características analisando quantos nós das árvores, que usam um dado atributo, reduzem a impureza geral da floresta. Ele calcula este valor automaticamente para cada atributo após o treinamento e normaliza os resultados para que a soma de todas as importâncias seja igual a 1. Quanto mais alto, mais importante é o atributo. A importância de um atributo é calculada como a redução total (normalizada) do critério trazido por esse atributo. É também conhecida como a importância Gini.<sup>8</sup>

### Validação cruzada *K-FOLD*

A validação cruzada (*K-fold cross validation*) é um método de amostragem utilizado para análise de desempenho de algoritmos de AM.<sup>9,10</sup> Consiste em dividir aleatoriamente o conjunto  $X$  em  $K$  folds (partes) mutuamente exclusivos e de tamanhos iguais. Os exemplos nos  $K-1$  folds são então usados para o treinamento do modelo e a hipótese induzida é testada no *fold* restante. Tal processo é repetido  $K$  vezes, de modo que todos os *folds* sejam usados uma única vez como conjunto de teste, conforme mostra a Figura 1 que utilizou  $K=10$ .

### Desempenho Métrico

A taxa de erro de um classificador  $h$  é denotada por  $err(h)$ , obtida por meio da Equação 2.3. Essa medida compara a classe atribuída pelo classificador de cada exemplo com a sua classe verdadeira. Se as duas classes forem iguais, ou seja,  $h(x_i) = f(x_i)$  então  $|h(x_i) - f(x_i)| = 0$ ; caso contrário,  $|h(x_i) - f(x_i)| = 1$ . A precisão ou taxa de acerto é denotada por  $acc$  e corresponde ao complemento da taxa de erro, conforme Equação 2.4.

$$err(h) = \frac{1}{\mathcal{I}} \sum_{i=1}^n |h(x_i) - f(x_i)| \quad (2.3)$$

$$acc(h) = 1 - err(h) \quad (2.4)$$

As taxas de erro e acerto podem ser obtidas por meio de uma matriz de confusão, que corresponde a uma matriz cuja dimensão é o número de classes existentes em  $X$ . Em uma matriz de confusão referente a um conjunto de exemplos com duas classes, geralmente denominadas como positiva e negativa, temos: verdadeiros positivos (VP) que correspondem ao exemplo que é positivo e foi classificado como positivo; falsos positivos (FP) que são exemplos negativos classificados como positivos; verdadeiros negativos (VN) que são exemplos negativos classificados como negativos; e finalmente os falsos negativos (FN) que são exemplos positivos que foram classificados como negativos. A partir da matriz de confusão, pode-se então obter a taxa de erro e a taxa de acerto por meio das Equações 2.5 e 2.6, respectivamente.

$$err(h) = \frac{FN + FP}{VP + FN + FP + VN} \quad (2.5)$$

$$acc(h) = \frac{VP + VN}{VP + FN + FP + VN} \quad (2.6)$$

Outra métrica de desempenho bastante utilizada, AUC (área *under the ROC curve*), é obtida por meio da geração de um gráfico de sensibilidade versus (1-especificidade), conhecido como curva ROC (*receiver operating characteristic*) e calculando-se a área sob a curva. A sensibilidade é a razão entre os verdadeiros positivos e o total de exemplos positivos, conforme mostrado na Equação 2.7. A especificidade é a razão entre os verdadeiros negativos e o total de exemplos negativos, conforme mostrado na Equação 2.8. Quanto maior o valor de AUC, melhor é o desempenho do classificador. Os valores de AUC variam em um intervalo [0,1].

$$\text{sensibilidade} = \frac{VP}{VP + FN} \quad (2.7)$$

$$\text{especificidade} = \frac{VN}{FP + VN} \quad (2.8)$$

## Métodos

Foram identificados 11.375 pacientes idosos que atendiam aos critérios de elegibilidade (idade superior a 60 anos) e que foram separados em um único banco de dados. Estes idosos foram notificados no período de 13 de fevereiro a 19 de junho de 2020 no estado de Pernambuco, Brasil. Os dados analisados são oriundos da Secretaria de Planejamento e Gestão de Pernambuco (SEPLAG-PE), baixados no dia 20 de junho em: [www.dados.seplag.pe.gov.br](http://www.dados.seplag.pe.gov.br). Foram excluídos todos os idosos que estavam em isolamento domiciliar ou hospitalizados, uma vez que estes ainda não tinham o desfecho concluído até o término do período considerado. Um total de 7486 idosos permaneceram depois disso, destes 4356 (58,19%) estavam recuperados e 3130 (41,81%) foram a óbito.

Foram considerados os atributos: sexo

(masculino, feminino), idade e características clínicas, como: tosse, dispnéia, febre, saturação de oxigênio  $\leq 95\%$ , presença de doenças cardiovasculares, respiratórias crônicas, renais crônicas, diabetes, neurológicas, neoplasias, alcoolismo, tabagismo. O objetivo foi construir um RF, com base nesses atributos, e apresentar quais são os mais importantes na predição de óbito em pacientes idosos com COVID-19 em Pernambuco. O trabalho foi implementado na linguagem Python,<sup>11</sup> utilizado o algoritmo RF, disponível no módulo Sklearn, conforme documentação disponível em: <https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html>. Uma validação cruzada com  $K=10$  foi empregada para calcular a performance e a importância dos atributos. O fluxograma da metodologia, ilustrado na Figura 1, mostra como foram calculadas as métricas que são apresentadas nos resultados.

## Resultados

A média e desvio padrão da idade foi  $72,94 \pm 9,55$  anos, com mediana de 71,0 anos. A média de idade entre os pacientes recuperados e os que vieram a

**Figura 1**

Fluxograma da metodologia de Validação Cruzada utilizando 10 *fold*s.

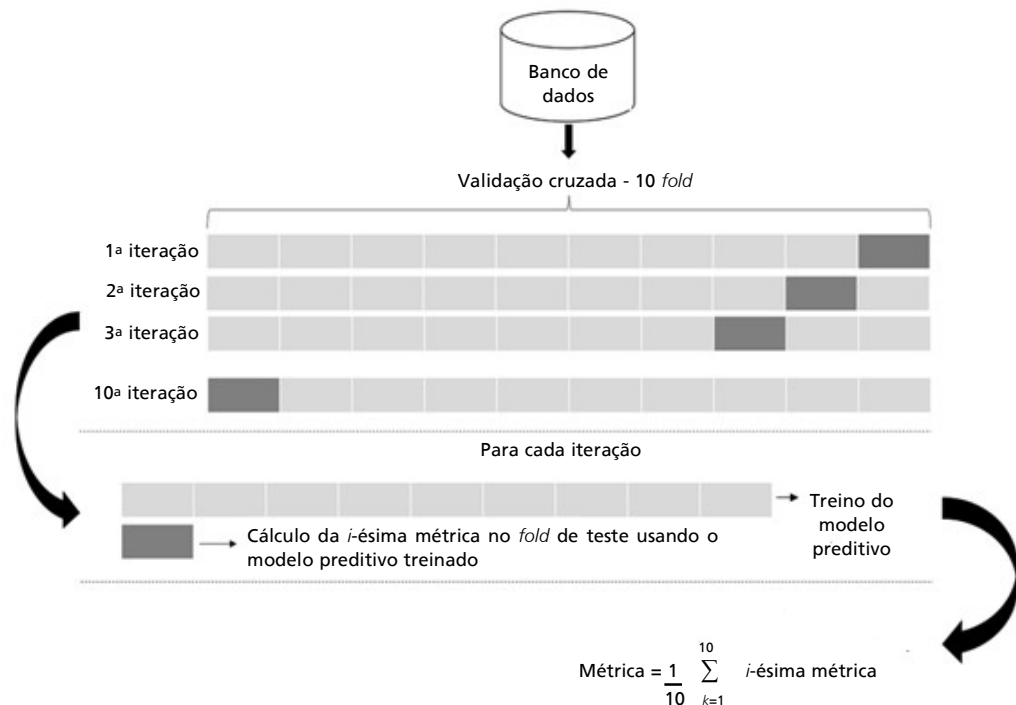


Tabela 1

Características clínicas dos idosos com COVID-19 no período de 12 de março a 19 de junho de 2020.

Casos	Total (N=7486)		Recuperados (N=4356)		Óbitos (N=3130)	
	n	%	n	%	n	%
Sexo						
Feminino	3821	51,04	2369	62	1452	38,0
Masculino	3665	48,95	1987	54,22	1678	45,78
Idade ( $\bar{X}\pm DP$ )	72,94 ( $\pm 9,55$ )		70,95 ( $\pm 9,06$ )		75,70 ( $\pm 9,52$ )	
Grupo etário (anos)						
60-69	3245	43,35	2288	52,53	957	30,58
70-79	2314	30,91	1252	28,74	1062	33,93
80 ou mais	1927	25,74	816	18,83	1111	35,50
Comorbidades						
Doenças cardiovasculares	1298	17,34	79	6,09	1219	38,95
Diabetes Mellitus	1081	14,44	426	1,81	655	20,93
Doenças respiratórias crônicas	246	3,29	30	0,69	216	6,90
Doença renal crônica	136	1,82	5	0,11	131	4,19
Doença neurológica	103	1,38	6	0,14	97	3,10
Neoplasias	93	1,24	3	0,07	90	2,88
Tabagismo	30	0,40	2	0,05	28	0,89
Etilismo	12	0,16	0	-	12	0,38
Sinais e Sintomas						
Tosse	4860	64,92	2766	63,50	2094	66,90
Febre	4403	58,82	2485	57,05	1918	61,28
Dispneia	3773	50,40	1529	35,10	2244	71,69
Saturação < 95%	2614	34,92	705	16,18	1909	60,99

Dados SEPLAG PE.

Tabela 2

Métricas para avaliação do desempenho do classificador *Random Forest*.

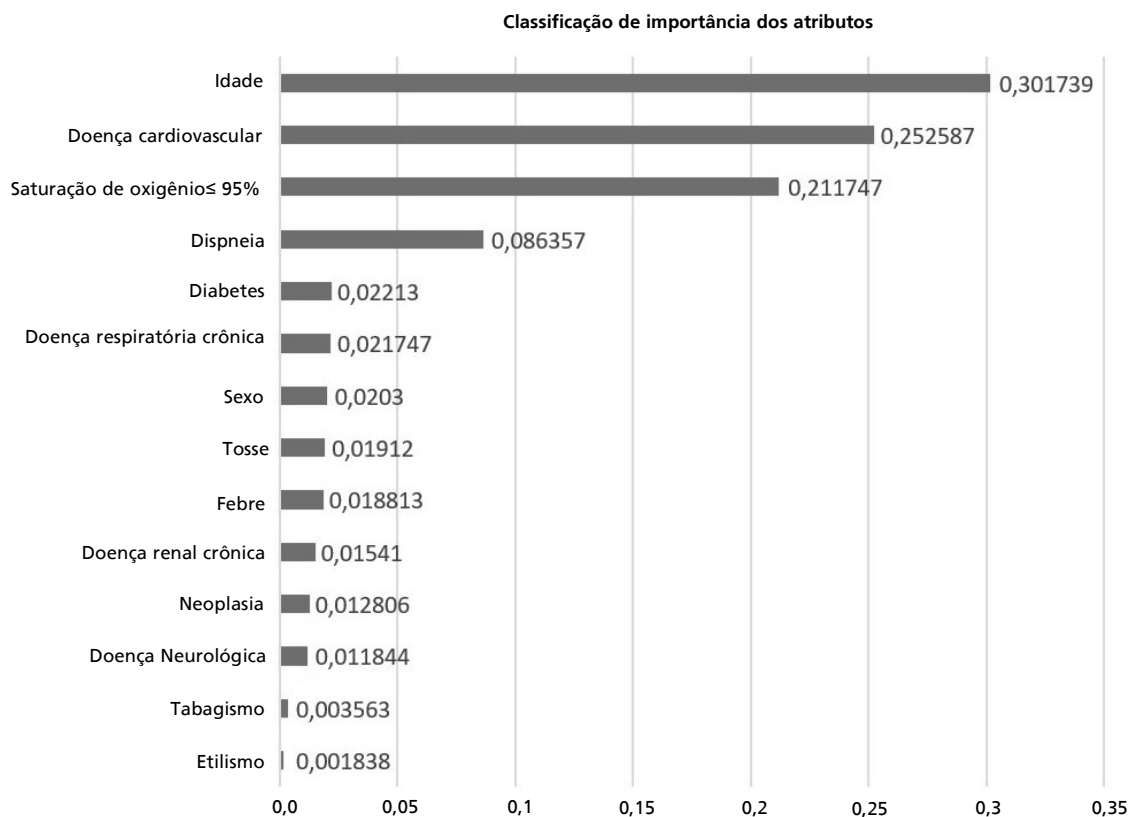
Taxa de Verdadeiro positivo	Taxa de Falso positivo	Precisão	Sensibilidade	AUC ROC	Desfecho
0,848	0,306	0,794	0,848	0,839	Recuperado
0,694	0,152	0,767	0,694	0,839	Óbito
0,784	0,241	0,783	0,784	0,839	Média ponderada

óbito foi  $70,95 \pm 9,06$  e  $75,70 \pm 9,52$  anos, respectivamente. As pacientes do sexo feminino corresponderam a 3821 (51,04%) e os do sexo masculino 3665 (48,96%). A taxa de letalidade global foi de 41,81%. A taxa de letalidade por faixa etária, sendo 29,49% entre 60-69 anos, 45,89% entre 70-79 anos 57,65% com mais de 80 anos. Com relação aos sintomas apresentados pelo grupo geral, 4860 (64,92%) apresentaram tosse, 4403 (58,82%) febre, 3773 (50,40%)

dispneia e 2614 (34,92%) saturação periférica de  $O_2 \leq 95\%$ . Contudo, no grupo dos pacientes que vieram a óbito, a manifestação clínica mais relevante foi a dispneia 2244 (71,69%). Em relação as comorbidades, as mais frequentes em toda a amostra foram Doenças Cardiovasculares 1298 (17,34%), Diabetes Mellitus 1081 (14,44%), Doença Respiratória Crônica 246 (3,29%). As neoplasias estavam presentes em 93 (1,24%) pacientes, 30 (0,40%)

Figura 2

Importância dos atributos: analisando quantos nós das árvores, que usam um dado atributo, reduzem a impureza geral da floresta.



pacientes eram tabagistas e 12 (0,16%) etilistas, conforme mostra a Tabela 1.

O classificador RF conseguiu acertar o desfecho de 78,39% dos pacientes do banco de dados. Para medir o desempenho da classificação, foi construída uma matriz de confusão e algumas métricas foram adotadas, conforme Tabela 2. É possível perceber que para prever o desfecho óbito, o algoritmo RF apresentou uma sensibilidade de 0,784 e uma taxa de precisão de 0,783, obtendo também uma Área sob a curva ROC (AUC) de 0,839. Ademais, a importância dos atributos mostrou que a idade (0,302), a presença de doença cardiovascular (0,252) e a saturação de oxigênio menor ou igual a 95% (0,212) são as três características mais importantes para a evolução do paciente idoso a óbito por COVID-19, conforme mostra Figura 2.

## Discussão

A idade foi o atributo mais importante relacionada

ao óbito, com importância de 0,302. Enquanto a taxa de letalidade global em Pernambuco, ao final dos três primeiros meses da pandemia, foi de 8,25%<sup>12</sup> a letalidade para os pacientes idosos no mesmo período foi de 41,81%. Este valor foi muito superior às taxas encontradas na literatura, que variaram entre 5,6% e 28,6%.<sup>13,14</sup> A análise da letalidade por grupo etário também mostrou taxas superiores as apresentadas na Itália, onde os casos fatais aumentaram principalmente após os 70 anos, sendo de 12,5% na faixa de 70-79 anos, 19,7% na faixa de 80-89 anos e 22,7% após 90 anos.<sup>15</sup> Vale ressaltar que as altas taxas de letalidades encontradas em Pernambuco refletem um período quando não havia ampla disponibilidade de testes.

Vários artigos também mostram que a presença de comorbidades é um fator de risco para desfechos clínicos adversos como o óbito,<sup>16-21</sup> sendo as doenças cardiovasculares sempre uma das comorbidades mais prevalentes nas amostras analisadas. Neste estudo, o algoritmo RF mostrou que as

doenças cardiovasculares foram a segunda característica mais importante para a predição de óbito em idosos com COVID-19, com valor de 0,252. Embora seja mais conhecido por causar danos ao sistema respiratório, sabe-se que a COVID-19 também pode comprometer ou agravar parâmetros cardiovasculares. Ademais, um estudo retrospectivo mostrou que 33% dos óbitos por COVID-19 foram atribuídas à insuficiência cardiorrespiratória e 7% à insuficiência cardíaca isolada.<sup>22</sup>

A terceira variável destacada para a predição de óbito, com valor de importância de 0,212, foi a saturação periférica de oxigênio  $\leq 95\%$ , concordando com a literatura atual.<sup>23</sup> O Ministério da Saúde inclusive considera o diagnóstico de Síndrome Respiratória Aguda Grave (SRAG) para todo indivíduo, de qualquer idade, com síndrome gripal e que apresente sinais de hipoxemia, como a saturação de  $O_2 \leq 95\%$  em ar ambiente.<sup>24</sup> Além disso, estudos

ressaltam que o reconhecimento precoce da hipóxia e da administração de oxigênio demonstrou reduzir a mortalidade para pacientes com COVID-19.<sup>25</sup>

Em conclusão, este estudo mostrou que o algoritmo RF foi capaz de revelar os aspectos mais importantes para predição de óbito em pacientes idosos com COVID-19, sendo os três mais importantes: a idade avançada, a presença de doença cardiovascular e a evidência de uma saturação periférica de  $O_2 < 95\%$ . Ademais, foi possível perceber que o algoritmo conseguiu prever o desfecho corretamente em 78,33% dos pacientes, obtendo uma AUC de 0,839.

### Contribuição dos autores

Todos os autores contribuíram de forma igual para a construção deste artigo.

### Referências

1. Brasil. Ministério da Saúde. Brasília, DF; 2020 [acesso 5 dez 2020] Disponível em: <https://sus.analitico.saude.gov.br>.
2. Thuler L, Melo A. Sars-CoV-2/Covid-19 em Pacientes com Câncer. *Rev Bras Cancerol*. 2020;66 (2): e-00970
3. Uddin S, Khan A, Hossain M, Moni M. Comparing different supervised machine learning algorithms for disease prediction. *BMC Med Inform Decis Mak*. 2019; 19 (1): 1-16.
4. Dash M, Liu H. Feature selection for classification. *Intelligent Data Analysis*. 1997; 1 (1-4): 131-56.
5. Guyon I, Elisseeff A. An introduction to variable and feature selection. *J Mach Learn Res*. 2003; 3: 1157-182.
6. Breiman L, Friedman J, Stone C, Olshen R. In: Chapman and Hall. *Classification and regression trees*. First edition. Wadsworth, New York: CRC Press; 1984.
7. Breiman L. Random forests. *Machine learning* 2002; 45: 5-32.
8. Scikit-learn: Machine Learning in Python, Pedregosa F, et al. *J Mach Learn Res*. 2011; 12: 2825-30. .
9. Stone M. Cross-validatory choice and assessment of statistical predictions. *J R Stat Soc Ser B Methodol*. 2018; 36 (2): 111-33.
10. Hastie T, Tibshirani R, Friedman J. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Second Edition. USA: Springer; 2009
11. Van Rossum G, Drake FL. *Python 3 Reference Manual*. Scotts Valley, CA: CreateSpace; 2009.
12. Centro de Informações Estratégicas Vigilância em Saúde Pernambuco. *Novo Coronavírus (COVID-19) Atualizações Epidemiológicas SES/PE*. Recife, Brasil; 2020. [acesso 5 dez 2020]. Disponível em: <https://www.cievspe.com/novo-coronavirus-2019-ncov>.
13. Zhang L, Zhu F, Xie L, Wang C, Wang J. Clinical characteristics of COVID-19-infected cancer patients: a retrospective case study in three hospitals within Wuhan, China. *Ann Oncol*. 2020; 31 (7): 894-901.
14. Epidemiology Working Group for NCIP Epidemic Response, Chinese Center for Disease Control and Prevention. The epidemiological characteristics of an outbreak of 2019 novel coronavirus diseases (COVID-19) in China. *Zhonghua Liu Xing Bing Xue Za Zhi*. 2020; 41 (2): 145-51.
15. Edward KB. Coronavirus Disease 2019 (COVID-19) in Italy. *JAMA*. 2020; 323 (14): 1335.
16. Huang PC, Wang Y, Li PX, Ren PL, Zhao PJ, Hu Y. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet*. 2020; 395 (10223): 497-506.
17. Chen PN, Zhou PM, Dong X, Qu PJ, Gong F, Han Y. Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. *Lancet*. 2020; 395 (10223): 507-13.
18. Liu J, Liu Y, Xiang P, Pu L, Xiong H. Neutrophil-to-Lymphocyte Ratio Predicts Severe Illness Patients with 2019 Novel Coronavirus in the Early Stage. 2020 [acesso 5 dez 2020]. Disponível em: <https://www.medrxiv.org/content/10.1101/2020.02.10.20021584v1.full.pdf>.
19. Ma J, Jing Y, Qian Y, Wu Y. Clinical characteristics and prognosis in cancer patients with COVID-19: A single center's retrospective study. *J Infect*. 2020; 81 (12): 318-56.

20. GuanW,LiangW, ZhaoY, LiangH, ChenZ. Comorbidity and its impact on 1590 patients with COVID-19 in China: a nationwide analysis. *Eur Respir J.* 2020; 55 (5): 2000547.
21. Ferreira J, Lima F, Oliveira J, Cancela M, Santos M. Covid-19 e Câncer: Atualização de Aspectos Epidemiológicos. *Rev Bras Cancerol.* 2020; 66:e-1013.
22. Ruan Q, Yang K, Wang W, Jiang L, Song J. Clinical predictors of mortality due to COVID-19 based on an analysis of data of 150 patients from Wuhan, China. *Intensive Care Med.* 2020; 46 (5): 846-8.
23. GuanW, NiZ, HuY, LiangW, Chun-quan OU. Clinical Characteristics of Coronavirus Disease 2019 in China. *N Engl J Med.* 2020; 382: 1708-20.
24. Brasil. Ministério da Saúde. Protocolo de Manejo Clínico do Coronavírus (COVID-19) na Atenção Primária à Saúde (v9). Brasília, Brasil; 2020. [acesso 5 dez 2020]. Disponível em: <https://portaldeboaspraticas.iff.fiocruz.br/biblioteca/protocolo-de-manejo-clinico-do-coronavirus-covid-19-na-atencao-primaria-a-saude/>
25. Sun Q, Qiu H, Huang M, Yang Y. Lower mortality of COVID-19 by early recognition and intervention: experience from Jiangsu Province. *Ann Intensive Care.* 2020; 10:33.

---

Recebido em 16 de Outubro de 2020

Aprovado em 8 de Fevereiro de 2021