

Modelo preditivo para classificação de risco de óbito de pacientes com COVID-19 utilizando dados abertos

Gustavo Rodrigues^{1,2}, Diego Kreutz³

¹Curso de Ciência da Computação
Universidade Federal do Pampa (UNIPAMPA)

²Combate à Fraude

³Laboratório de Estudos Avançados em Computação (LEA)
Programa de Pós-Graduação em Engenharia de Software (PPGES)
Universidade Federal do Pampa (UNIPAMPA)

`gustavo.rodrigues@combatea fraude.com.br`, `diegokreutz@unipampa.edu.br`

Resumo. *Com o intuito de mitigar a subjetividade de políticas para acesso a leitos de UTI, propomos um preditor baseado em florestas aleatórias para classificação de risco de óbito de pacientes com COVID-19. O conjunto de dados abertos utilizados engloba mais de 600 mil pacientes reportados através do Painel Coronavírus RS. No conjunto de teste, o modelo classificou a chance de óbito com uma pontuação AUC-ROC de 0,97. Estes resultados evidenciam o potencial do preditor em auxiliar na tomada de decisão no ambiente hospitalar.*

1. Introdução

Durante a pandemia de COVID-19 os sistemas de saúde do mundo inteiro enfrentaram problemas quanto a disponibilidade e alocação de recursos como respiradores e leitos de UTI [Latif et al. 2020, Ranney et al. 2020]. Políticas e decisões críticas foram tomadas quanto a priorização de pacientes com a doença e, em países como o Brasil, existem regras que definiram quem tem direito a leitos de UTI. Em alguns lugares, devido a urgência e falta de dados mais precisos, a política foi priorizar as pessoas mais jovens para a ocupação dos leitos de UTI¹.

Estudos mostram que modelos de aprendizagem de máquina (*machine learning*) conseguem prever a chance de óbito de um paciente positivo para COVID-19 com até 0,99 de pontuação AUC-ROC [Wynants et al. 2020]. Por se tratar de um contexto crítico (pandemia e risco de vida), os modelos preditivos devem ser robustos, seguros e transparentes, sem comprometer a ética. Um dos principais desafios é desenvolver e garantir a credibilidade destes modelos.

Muitas propostas de classificação de risco refletem dificuldades nas suas aplicações ou até mesmo em como podem auxiliar no contexto médico [Wynants et al. 2020]. Diversos modelos contêm alto risco de enviesamento (*i.e.*, modelo não generalista) devido ao desenvolvimento baseado em amostras pequenas da população (*e.g.*, dados de apenas um hospital), métricas de avaliação inadequadas (*e.g.*, acurácia para dados desbalanceados), ou pobreza de detalhes sobre o desenvolvimento

¹<https://www.nsctotal.com.br/colonistas/dagmara-spautz/estado-oficializa-criterio-que-da-prioridade-a-mais-jovens-e-saudaveis>

e objetivo do modelo (*e.g.*, omissão dos métodos utilizados) [Wynants et al. 2020]. Devido a alta variabilidade climática, cultural e racial entre diferentes países e populações [Assaf et al. 2020, Cheng et al. 2020, Zhao et al. 2020, Casiraghi et al. 2020] é necessário o desenvolvimento de uma solução que utilize dados de populações pouco exploradas, como é o caso do Brasil, além de utilizar métricas para avaliação do desempenho de modelos com dados desbalanceados. Por fim, para mitigar a falta de detalhes na explanação do estudo, devemos utilizar guias que auxiliem no relatório transparente do desenvolvimento do modelo (*e.g.*, STARD [*Standards for Reporting of Diagnostic Accuracy*] e o TRIPOD [*Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis*]).

Neste trabalho, propomos um modelo preditivo baseado em florestas aleatórias para classificação de risco de óbito de pacientes confirmados de COVID-19 no estado do Rio Grande do Sul/Brasil. Utilizamos a pontuação AUC-ROC para avaliar a performance do modelo. O desenvolvimento do modelo é explicado utilizando o guia TRIPOD [Collins et al. 2015] para atenuar o risco da falta de detalhes. Utilizamos também da validação cruzada *5-fold* para validar internamente o modelo e, desta forma, reduzir o risco de enviesamento.

O restante deste trabalho está organizado como segue. Na Seção 2 é apresentado o estado da arte do aprendizado de máquina aplicado ao combate da pandemia. Na Seção 3 é discutido o projeto do estudo, a aquisição dos dados e a construção do modelo de prognóstico. Na Seção 4 avaliamos o desempenho do modelo construído, bem como a sua interpretabilidade. Por fim, na Seção 5 apresentamos a discussão geral sobre as contribuições e limites do presente estudo.

2. Estado da arte

Os métodos identificados na literatura tem por objetivo auxiliar na tomada de decisão médica, incluindo prever a severidade dos casos confirmados (classificação de risco), chance de óbito ou recuperação, e chance de internação em UTI [Iwendi et al. 2020, Pourhomayoun and Shakibi 2021a, Zhao et al. 2020, Yadaw et al. 2020, Cheng et al. 2020, Assaf et al. 2020, Gao et al. 2020, Chowdhury et al. 2021, Casiraghi et al. 2020, Dun et al. 2020, Pourhomayoun and Shakibi 2021b, Schöning et al. 2021, Vepa et al. 2021, Magunia et al. 2021]. Ao analisarmos a literatura, constatamos que o método Floresta Aleatória é o mais frequentemente utilizado. Isto ocorre devido ao fato deste método ser considerado acurado e robusto pois conta com um alto número de árvores de decisão no processo de treinamento, por não sofrer com sobreajuste e pelo seu auxílio na classificação da importância das características.

As principais características utilizadas para treinamento são dados demográficos (*e.g.*, idade, sexo e país de origem); dados clínicos (*e.g.*, sintomas e comorbidades); dados laboratoriais (*e.g.*, amostras de sangue), onde foi possível identificar a mínima saturação de O₂ como característica chave. Por fim, características radiológicas (*e.g.*, imagens de radiografia) também aparecem como características de treinamento [Casiraghi et al. 2020].

Nos modelos propostos pelos trabalhos, a quantidade de características utilizadas varia de 3 [Yadaw et al. 2020] a 99 [Cheng et al. 2020]. Entretanto, foi possível observar que a idade e o sexo foram unânimes nos estudos que utilizaram

de dados demográficos. Além disso, é importante ressaltar que nos modelos que utilizam dados clínicos, em termos de sintomas e comorbidades, todos utilizaram características relacionadas a resfriados, como febre, tosse e dispnéia.

Para entender quão bom um modelo preditivo é, existem métricas para avaliar o desempenho (ou generalização) de um método de aprendizagem de máquina [Amidi and Amidi 2020]. As métricas mais encontradas nos trabalhos foram métricas obtidas através da matriz de confusão, sendo a AUC-ROC a mais utilizada, seguida por acurácia, precisão, sensibilidade, especificidade e F1 *score*.

Uma limitação comum dos trabalhos encontrados na literatura é o grupo de pacientes utilizados para treino. Um modelo de previsão aplicado em um novo sistema de saúde, configuração ou país geralmente produz previsões que estão mal calibradas e podem precisar ser atualizadas antes que possam ser aplicadas com segurança nessa nova configuração. Os mesmos métodos, com outras populações ou de outros sistemas de saúde, podem levar a resultados diferentes [Assaf et al. 2020, Cheng et al. 2020, Zhao et al. 2020, Casiraghi et al. 2020].

É importante destacar também que existem trabalhos similares ao nosso, isto é, com o objetivo de estratificar o risco de complicações por COVID-19 e auxiliar na alocação de recursos, como [Pourhomayoun and Shakibi 2021b], [Schöning et al. 2021], e [Vepa et al. 2021]. Entretanto, esses trabalhos são aplicados em outras populações e utilizam características distintas. Por exemplo, em [Vepa et al. 2021] os autores utilizam dados de um único hospital da Inglaterra e atingem uma acurácia de apenas 84,1%. Um segundo exemplo interessante é o [Pourhomayoun and Shakibi 2021b], onde os autores utilizam características muito similares (demográficas e clínicas) às nossas porém com dados de 147 países, atingindo uma acurácia de apenas 89,98% (provavelmente devido a grande quantidade de países que acarretaria numa generalidade muito grande e perda de desempenho), enquanto nós, utilizando os dados abertos do RS, atingimos uma AUC-ROC de 0,97.

3. Metodologia

Objetivando mitigar problemas relacionados ao risco de enviesamento por pobreza no relatório sobre o desenvolvimento do modelo [Wynants et al. 2020], a metodologia para o desenvolvimento e a exposição dos resultados seguem o guia TRIPOD [Collins et al. 2015]. O guia é um *checklist* sobre pontos essenciais que um estudo de diagnóstico ou prognóstico deve relatar quando estiver explanando o desenvolvimento de um modelo preditivo multivariado. O *checklist* TRIPOD deste trabalho está disponível no GitHub¹. As seções a seguir detalham os participantes, a engenharia de características, a variável resposta, as variáveis preditoras e os métodos estatísticos.

3.1. Participantes

Para o desenvolvimento, treino e validação do modelo preditivo, utilizamos dados anonimizados provenientes do Painel Coronavírus RS³. Neste estudo, utilizamos um conjunto de dados com 604.389 registros (observações) de casos confirmados de

¹https://github.com/gustavocrod/predict_death_covid/blob/main/imgs/TRIPOD.pdf

³<https://ti.saude.rs.gov.br/covid19/>

COVID-19. O conjunto de dados original contém um total de 30 variáveis em nível-paciente, incluindo dados demográficos (*e.g.*, sexo e faixa-etária), clínicos (*e.g.*, tosse e dispneia) e temporais (*e.g.*, data do início dos sintomas) e compreende o período de 01 de janeiro de 2021 à 08 de junho de 2021.

3.2. Engenharia de características

No conjunto de dados, existem registros de pacientes que vieram a óbito por causas distintas da COVID-19 (20 [$< 0,01\%$]), e também há aqueles cujo parecer evolutivo ainda não era conhecido naquele momento, ou seja, que ainda estavam em acompanhamento (17.724 [$< 3\%$]). Há também registros com informação faltante do sintoma dispneia (2.418 [$< 0,5\%$]). Optamos por remover os registros com informações faltantes e aqueles que não possuem informação relevante para o estudo (*i.e.*, óbitos por outras causas ou ainda em acompanhamento). Ao final da seleção, dos 604.389 registros iniciais, 584.228 permaneceram na base para prosseguir o estudo.

Além das características iniciais do conjunto de dados⁴, outras seis características foram extraídas do campo textual “CONDICOES” (campo pré-existente no conjunto de dados), a saber: cardiopatia, diabetes, doença respiratória, problema renal, obesidade e doença cromossômica. Como a variável “Raça/cor” possuía 19,98% (n=116.752) de dados faltantes, sendo estes 58% dos óbitos totais (n=10.889), e devido ao decréscimo da pontuação do modelo ao adicioná-la como preditora (ver Seção 4), optamos por sua remoção do conjunto de dados.

3.3. Variável resposta

O principal objetivo do trabalho é prever a chance de óbito entre pacientes positivos de COVID-19. Esta análise prognóstica é importante para a priorização dos pacientes na alocação dos recursos hospitalares. Os dados coletados possuem a informação relacionada a evolução do paciente na variável “EVOLUCAO”. Esta variável é dicotômica – “ÓBITO” (1) se o paciente faleceu, e “RECUPERADO” (0) caso contrário – e nos permite a criação de um modelo de aprendizagem de máquina para classificar os pacientes quanto à sua chance de falecimento, pois informa quais e quantos foram recuperados ou vieram a óbito por COVID-19.

3.4. Variáveis Predictoras

As variáveis candidatas para predictoras foram selecionadas baseando-se em um dos critérios comuns [Collins et al. 2015]: variáveis clínicas conhecidas de diagnóstico comum relacionado a pneumonia ou quadros gripais. Após a extração das novas características e remoção de campos que não são candidatos a variáveis predictoras (*i.e.*, datas e códigos), os dados clínicos, demográficos e categóricos resultantes e que foram utilizados no estudo podem ser vistos na Tabela 1.

⁴<https://ti.saude.rs.gov.br/covid19/api>

Nome	Descrição	Tipo
SEXO	Sexo	Binária/Demográfica
FAIXAETARIA	Faixa etária	Categórica/Demográfica
EVOLUCAO	Qual o parecer evolutivo do paciente?	Binária/Alvo
FEBRE	Sintomas de febre	Binária/Clínica
TOSSE	Sintomas de tosse	Binária/Clínica
GARGANTA	Sintomas de dor de garganta	Binária/Clínica
DISPNEIA	Sintomas de dispnéia/ falta de ar	Binária/Clínica
GESTANTE	O paciente é gestante?	Binária/Clínica
SRAG	O quadro clínico se caracteriza como Síndrome Respiratória Aguda Grave?	Binária/Clínica
CARDIOPATIA	O paciente possui alguma doença cardíaca crônica?	Binária/Clínica
DIABETES	O paciente possui diabetes?	Binária/Clínica
DOENCA_RESPIRATORIA	O paciente possui alguma doença respiratória crônica?	Binária/Clínica
PROBLEMA_RENAL	O paciente possui problema renal?	Binária/Clínica
OBESIDADE	O paciente é obeso?	Binária/Clínica
DOENCA_CROMOSSOMICA	O paciente possui doença cromossômica?	Binária/Clínica

Tabela 1. Variáveis

3.5. Métodos Estatísticos

Análise descritiva. Para entendermos sobre possíveis características que expliquem o risco de óbito, performamos análises descritivas dos preditores por cada grupo e apresentamos os resultados em quantidade e proporção de óbitos. Possíveis correlações foram analisadas utilizando o coeficiente de Pearson [Benesty et al. 2009] e a importância relativa das características foi analisada através do grau de impureza de Gini [Menze et al. 2009] (*i.e.*, a importância de cada característica para o modelo).

Análise preditiva. Aplicamos um algoritmo de aprendizagem de máquina para prever a chance de óbito entre pacientes confirmados com COVID-19. Para a implementação, análise estatística, avaliação e validação dos modelos, utilizamos a linguagem Python 3, juntamente com as bibliotecas para aprendizagem de máquina: Scikit-learn [Pedregosa et al. 2011] (floresta aleatória), e de manipulação e visualização de dados Pandas [McKinney 2010], Numpy [Harris et al. 2020] e Seaborn [Hunter 2007].

Para garantir a reprodutibilidade do experimento, definimos arbitrariamente a semente 777 para embaralhar o conjunto de dados inicial. Utilizamos uma divisão pseudo-aleatória (*random split*) de 70%/30%, a partir dos dados iniciais, sendo 70% utilizado para treinos e testes, e os 30% restante, para validação interna do modelo [James et al. 2013]. Avaliamos o desempenho/discriminação do modelo através da área sobre a curva ROC (AUC-ROC). A pontuação AUC-ROC pode assumir valores de 0 a 1, onde um valor de 0 indica um modelo que erra todas as predições e 1 reflete

um teste perfeito de predição. Em geral, um valor de 0,5 sugere que o modelo não tem capacidade de discriminação (*i.e.*, não consegue diferenciar pacientes que vieram a óbito dos recuperados); 0,7 a 0,8 é considerado aceitável; 0,8 a 0,9 é ótimo; e valores acima de 0,9 são considerados excelentes [Mandrekar 2010].

Por se tratar de um conjunto de dados desbalanceado, utilizamos da técnica de subamostragem para balancear os dados de treino, e então confrontar com o treinamento sem balanceamento para detectar melhorias. Utilizamos da pesquisa de grade por validação cruzada para encontrar os melhores hiperparâmetros para o modelo - incluindo a quantidade de árvores de decisão (`n_estimators`) e a profundidade máxima de cada árvore (`max_depth`) [Krstajic et al. 2014] - que maximizassem a pontuação AUC-ROC nos dados de treino.

Validação do modelo. Para validação interna do modelo, utilizamos da técnica *K-Fold Cross Validation*, que consiste em dividir o conjunto de dados em K dobras (*folds*). A função de previsão é aprendida usando K-1 dobras, e a dobra deixada de fora é usada para teste. Optamos por utilizar sua variação estratificada, ou seja, na adaptação para conjuntos de dados desbalanceados, pois mantém a proporção original de cada classe na etapa de teste. Geralmente, realiza-se a validação cruzada k-fold usando $k = 5$ ou $k = 10$, uma vez que estes valores produzem estimativas de taxa de erro de teste que não sofrem de viés ou variância excessivamente altos. Por padrão, a biblioteca Scikit-learn [Pedregosa et al. 2011] utiliza $K = 5$, e por isso decidimos manter este valor.

4. Resultados

4.1. Participantes

O perfil dos participantes é apresentado na Tabela 2. Do conjunto de dados utilizado no estudo, 565.649 (96,82%) casos possuem evolução conhecida como recuperação, e 18.579 (3,18%) são casos que tiveram óbito comprovado por COVID-19, onde a taxa de letalidade geral é de 3,07% ($n=18.579$). Dos casos confirmados, pouco mais da metade é do sexo feminino (53,10%, $n=310.263$), tendo o sexo masculino a maior taxa de letalidade proporcional (3,6%, $n=9.889$). Entre as faixa-etárias, a maior concentração está entre 30 e 39 anos (20,86%), sendo a faixa superior aos 70 anos a mais letal (43.74%, $n=7.909$). Dentre os sintomas, destacamos a falta de ar/dispnéia como o sintoma com a maior taxa de letalidade (17,49%, $n=15.676$). Em contra-ponto, a dor-de-garganta é o sintoma apresentado que possui a menor taxa de letalidade (1,39%, $n=2.952$).

Nome do atributo		Infectados	Óbitos	Proporção (%)
SEXO	Masculino	273.965	9889	3,60
	Feminino	310.263	8.690	2,80
FAIXA_ETARIA	<1	4.133	4	0,09
	01 a 04	7.403	4	0,05
	05 a 09	9.267	2	0,02
	10 a 14	13390	5	0,03
	15 a 19	30.292	19	0,06
	20 a 29	108.501	225	0,20
	30 a 39	121.926	758	0,62
	40 a 49	108.501	225	0,20

	40 a 49	104.102	1.716	1,64
	50 a 59	88.424	3.138	3,54
	60 a 69	57.058	4.799	8,41
	70 a 79	27.233	4.516	16,58
	80 e mais	12.490	3.393	27,16
FEBRE	Sim	198.357	8.909	4,49
	Não	385.871	9.670	2,50
TOSSE	Sim	270.337	10.873	4,02
	Não	313.891	7.706	2,45
GARGANTA	Sim	211.297	2.952	1,39
	Não	372.931	2.952	4,19
DISPNEIA	Sim	89.606	15.676	17,49
	Não	494.622	2.903	0,58
GESTANTE	Sim	3.157	47	1,48
	Não	581.071	18.532	3,18
SRAG	Sim	51.691	18.579	35,94
	Não	532.537	0	0,0
CARDIOPATIA	Sim	16.981	7.728	45,50
	Não	556.396	10.851	1,91
DIABETES	Sim	22.627	5.593	24,71
	Não	561.601	17.034	2,31
DOENCA_RESPIRATORIA	Sim	3.566	1.552	43,52
	Não	563.635	17.027	2,93
PROBLEMA_RENAL	Sim	2254	861	38,19
	Não	581.974	17.718	3,04
OBESIDADE	Sim	10.855	2.868	26,42
	Não	573.373	15.711	2,74
DOENCA_CROMOSSOMICA	Sim	141	83	41,13
	Não	584.087	18.521	3,17

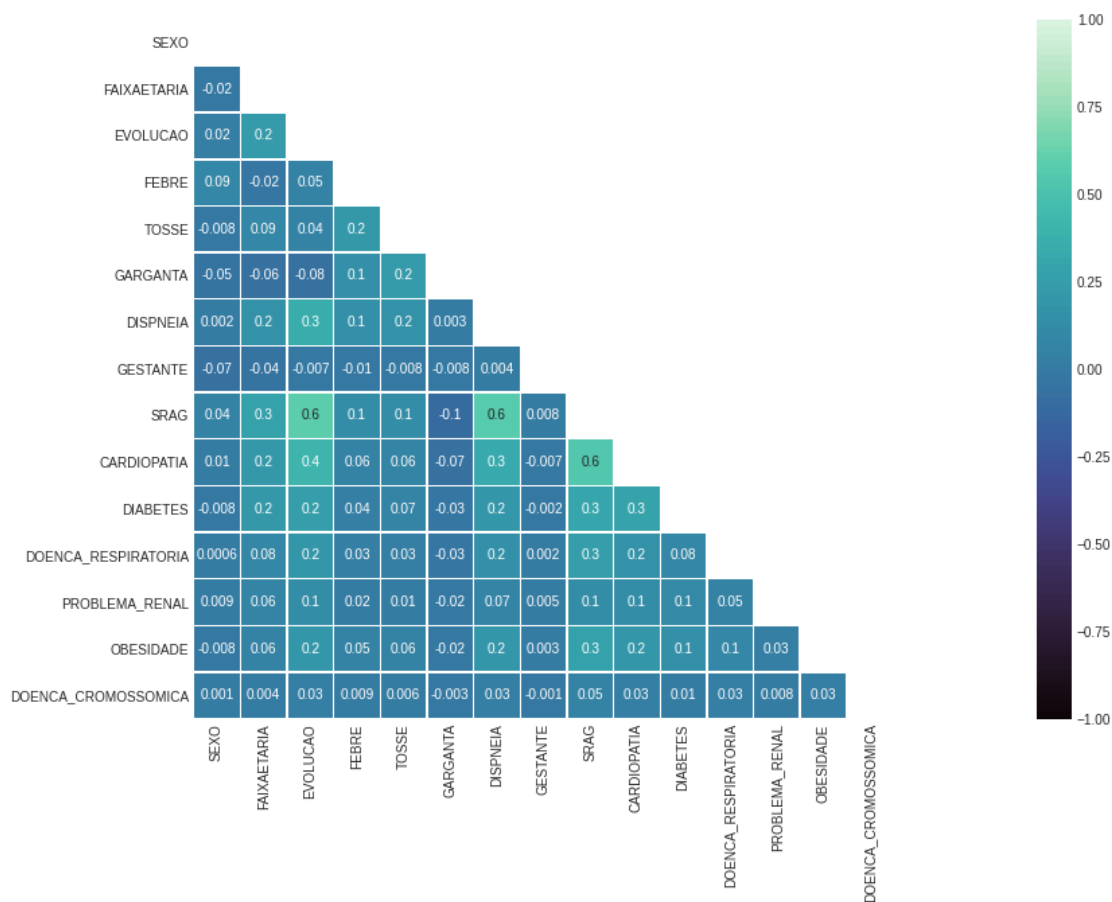
Tabela 2: Dados dos participantes

4.2. Desenvolvimento do modelo

Aplicamos o algoritmo de Floresta aleatória, utilizando 100 árvores de decisão (`n_estimators`) no processo e com a profundidade máxima (`max_depth`) definida em 4, pois foram os melhores hiperparâmetros retornados pela busca em grade com validação cruzada ($k = 5$). Como o conjunto de dados de treino é desbalanceado, possuindo 395.954 (96,82%) casos recuperados e 13.005 (3,18%) óbitos, optamos por utilizar da técnica de subamostragem aleatória (*random under sampling*) [Anand et al. 2010].

O coeficiente de correlação de Pearson entre as variáveis independentes pode ser observado na Figura 1. Podemos constatar uma variação de -0,8 a 0,6, e que as variáveis que possuem uma maior dependência ($\geq 0,2$) com a variável resposta (EVOLUCAO) são SRAG, DISPNEIA, CARDIOPATIA, DIABETES, DOENCA_RESPIRATORIA e OBESIDADE (*i.e.*, variáveis que possuem maior correlação). E de acordo com grau de impureza de Gini, constatamos a influência relativa das *top 5* características na classificação predita pelo modelo, a saber: SRAG, FAIXAETARIA, CARDIOPATIA, DISPNEIA e DIABETES.

Figura 1. Matriz de correlação de Pearson



	Pontuação AUC-ROC	
	Média (5-V.C.)	I.C. = 95 %
Floresta aleatória (com raça/cor)	0,539	(0,528 - 0,550)
Floresta aleatória (linha de base)	0,705	(0,694 - 0,716)
Floresta aleatória (subamostragem aleatória)	0,969	(0,968 - 0,970)
Floresta aleatória (subamostragem aleatória + ajuste de hiperparâmetro)	0,981	(0,981 - 0,982)

Tabela 3. Desempenho do modelo

4.3. Avaliação e Teste do modelo

O conjunto inicial dos dados possuía 584.228 registros e um total de 18.579 óbitos. O conjunto de treino e validação possui 408.959 registros, sendo 3,18% (n=13.005) óbitos. O conjunto de teste possui a mesma distribuição, sendo 5.574 óbitos. Os

valores das pontuações médias após a validação cruzada (5-V.C.) do modelo de floresta aleatória, juntamente com seus intervalos de confiança com nível de confiança de 95% (I.C. = 95%) podem ser vistos na Tabela 3. O modelo com menos registros de óbito (com raça/cor) possui o pior desempenho e foi descartado das melhorias posteriores (subamostragem e ajuste de hiperparâmetros).

Utilizamos uma matriz de confusão (Figura 2) para descrever e visualizar o desempenho do classificador nos dados de teste e também para prover *insights* sobre onde o modelo errou. É importante salientar que pontuação AUC-ROC é derivada desta matriz, que informa os falsos positivos e negativos, bem como os classificados corretamente. No nosso contexto, a nossa classe positiva são os casos de óbitos e a classe negativa são os casos recuperados. Podemos observar na matriz que apenas 5,92% (n=10.045) casos recuperados foram classificados erroneamente (falsos positivos) e que todos os casos de óbito foram classificados corretamente (*i.e.*, sem falsos negativos). Como o modelo performou 0,97 (*i.e.*, acima de 0,90) de pontuação AUC-ROC, podemos classificá-lo como excelente discriminador [Hosmer Jr et al. 2013].

Valor Real	Recuperado	94.08%	5.92%
	Óbito	0.00%	100.00%
		Recuperado	Óbito
		Valor Predito	

Figura 2. Matriz de confusão conjunto de teste

Em termos comparativos numéricos, sabendo que há nuances e particulares de cada contexto, o resultado do nosso modelo é superior (AUC-ROC de 0,97) aos trabalhos similares existentes, como [Schöning et al. 2021] (AUC-ROC de 0,96), [Vepa et al. 2021] (acurácia de 84,1%) e [Magunia et al. 2021] (AUC-ROC de 0,84). É importante destacar que o trabalho relacionado que obteve o maior desempenho ([Schöning et al. 2021]) também utilizou dados de apenas uma região, enquanto os outros utilizaram dados de várias regiões ou baixa amostragem de uma região.

5. Limitações

Este estudo possui pelo menos quatro limitações. Primeiramente, alguns dos maiores preditores laboratoriais utilizados na construção de modelos prognósticos (*i.e.*, proteína C-reativa e nível de linfócitos no sangue [Wynants et al. 2020]) não estão disponíveis no Painel Coronavírus RS, e por isso fomos impossibilitados de utilizar neste trabalho. Acreditamos que elas podem ajudar a construir um modelo com maior poder preditivo.

Segundo, a metodologia de treinamento do modelo adotada é denominada “aprendizado *offline*”, onde o nosso conjunto de dados é estático, ou seja, a quantidade de observações corresponde a um intervalo de tempo bem definido. Para mitigar o problema relacionado a possíveis mudanças devido a novas cepas e mutações do vírus, técnicas como o “aprendizado *online*” podem ser abordadas e comparadas com a solução do presente trabalho. Esta outra abordagem de aprendizado consiste em treinar o modelo à medida em que novos dados são gerados.

É importante salientar também, que analisamos os dados como sendo registros de pacientes únicos. Isto ocorre devido aos dados de reinfecção não estarem disponíveis no Painel Coronavírus RS e implica que eventuais casos de reinfecção foram ignorados. E por fim, como utilizamos apenas um algoritmo de aprendizagem de máquina no processo, em pesquisas futuras outros métodos de análise preditiva podem ser utilizados para comparar com o desempenho do algoritmo de florestas aleatórias.

6. Considerações Finais

Desenvolvemos e validamos internamente um classificador que permite estratificar com precisão o risco de óbito de pacientes com COVID-19. O modelo preditivo proposto pode contribuir para a diminuição da subjetividade na tomada de decisão, para que o profissional de saúde possa tomar suas decisões baseadas não somente no empirismo, mas também em um parâmetro técnico dentro do ambiente hospitalar. O código e os dados estão disponíveis no GitHub². Entretanto, é importante destacar também que os aspectos éticos e legais devem ser considerados no momento em que os profissionais tomam a decisão, ou seja, esta é apenas uma ferramenta adicional que pode auxiliar no processo de tomada de decisão e não substitui a expertise médica.

Referências

- Amidi, A. and Amidi, S. (2020). Machine learning tips and tricks cheatsheet. <https://tinyurl.com/ml-tips-and-tricks>.
- Anand, A., Pugalenti, G., Fogel, G. B., and Suganthan, P. (2010). An approach for classification of highly imbalanced data using weighting and undersampling. *Amino acids*, 39(5):1385–1391.
- Assaf, D., Gutman, Y., Neuman, Y., Segal, G., Amit, S., Gefen-Halevi, S., Shilo, N., Epstein, A., Mor-Cohen, R., Biber, A., et al. (2020). Utilization of machine-learning models to accurately predict the risk for critical COVID-19. *Internal and emergency medicine*, 15(8):1435–1443.
- Benesty, J., Chen, J., Huang, Y., and Cohen, I. (2009). Pearson correlation coefficient. In *Noise reduction in speech processing*, pages 1–4. Springer.
- Casiraghi, E., Malchiodi, D., Trucco, G., Frasca, M., Cappelletti, L., Fontana, T., Esposito, A. A., Avola, E., Jachetti, A., Reese, J., et al. (2020). Explainable machine learning for early assessment of COVID-19 risk prediction in emergency departments. *IEEE Access*, 8:196299–196325.

²https://github.com/gustavocrod/predict_death_covid

- Cheng, F.-Y., Joshi, H., Tandon, P., Freeman, R., Reich, D. L., Mazumdar, M., Kohli-Seth, R., Levin, M., Timsina, P., and Kia, A. (2020). Using machine learning to predict ICU transfer in hospitalized COVID-19 patients. *Journal of Clinical Medicine*, 9(6):1668.
- Chowdhury, M. E., Rahman, T., Khandakar, A., Al-Madeed, S., Zughailer, S. M., Hassen, H., and Islam, M. T. (2021). An early warning tool for predicting mortality risk of COVID-19 patients using machine learning. *Cognit Comput*, 21:1–16.
- Collins, G. S., Reitsma, J. B., Altman, D. G., and Moons, K. G. (2015). Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD) the TRIPOD statement. *Circulation*, 131(2):211–219.
- Dun, C., Walsh, C., Bae, S., Adalja, A., Toner, E., Lash, T. A., Hashim, F., Paturzo, J., Segev, D. L., and Makary, M. A. (2020). A machine learning study of 534,023 medicare beneficiaries with COVID-19: Implications for personalized risk prediction. *medRxiv*.
- Gao, Y., Cai, G.-Y., Fang, W., Li, H.-Y., Wang, S.-Y., Chen, L., Yu, Y., Liu, D., Xu, S., Cui, P.-F., et al. (2020). Machine learning based early warning system enables accurate mortality risk prediction for COVID-19. *Nature communications*, 11(1):1–10.
- Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N. J., Kern, R., Picus, M., Hoyer, S., van Kerkwijk, M. H., Brett, M., Haldane, A., Fernández del Río, J., Wiebe, M., Peterson, P., Gérard-Marchant, P., Sheppard, K., Reddy, T., Weckesser, W., Abbasi, H., Gohlke, C., and Oliphant, T. E. (2020). Array programming with NumPy. *Nature*, 585:357–362.
- Hosmer Jr, D. W., Lemeshow, S., and Sturdivant, R. X. (2013). *Applied logistic regression*, volume 398. John Wiley & Sons.
- Hunter, J. D. (2007). Matplotlib: A 2D graphics environment. *Computing in science & engineering*, 9(3):90–95.
- Iwendi, C., Bashir, A. K., Peshkar, A., Sujatha, R., Chatterjee, J. M., Pasupuleti, S., Mishra, R., Pillai, S., and Jo, O. (2020). COVID-19 patient health prediction using boosted random forest algorithm. *Frontiers in public health*, 8:357.
- James, G., Witten, D., Hastie, T., and Tibshirani, R. (2013). *An introduction to statistical learning*, volume 112. Springer.
- Krstajic, D., Buturovic, L. J., Leahy, D. E., and Thomas, S. (2014). Cross-validation pitfalls when selecting and assessing regression and classification models. *Journal of cheminformatics*, 6(1):1–15.
- Latif, S., Usman, M., Manzoor, S., Iqbal, W., Qadir, J., Tyson, G., Castro, I., Razi, A., Boulos, M. N. K., Weller, A., et al. (2020). Leveraging data science to combat COVID-19: A comprehensive review. *IEEE Transactions on Artificial Intelligence*.
- Magunia, H., Lederer, S., Verbuecheln, R., Gilot, B. J., Koeppen, M., Haeberle, H. A., Mirakaj, V., Hofmann, P., Marx, G., Bickenbach, J., et al. (2021). Machine learning identifies ICU outcome predictors in a multicenter COVID-19 cohort. *Critical Care*, 25(1):1–14.
- Mandrekar, J. N. (2010). Receiver operating characteristic curve in diagnostic test assessment. *Journal of Thoracic Oncology*, 5(9):1315–1316.
- McKinney, W. (2010). Data structures for statistical computing in Python. In van der Walt, S. and Millman, J., editors, *Proceedings of the 9th Python in Science Conference*, pages 51–56.

- Menze, B. H., Kelm, B. M., Masuch, R., Himmelreich, U., Bachert, P., Petrich, W., and Hamprecht, F. A. (2009). A comparison of random forest and its Gini importance with standard chemometric methods for the feature selection and classification of spectral data. *BMC bioinformatics*, 10(1):1–16.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., et al. (2011). Scikit-learn: Machine learning in Python. *the Journal of machine Learning research*, 12:2825–2830.
- Pourhomayoun, M. and Shakibi, M. (2021a). Predicting mortality risk in patients with COVID-19 using machine learning to help medical decision-making. *Smart Health*, 20:100178.
- Pourhomayoun, M. and Shakibi, M. (2021b). Predicting mortality risk in patients with COVID-19 using machine learning to help medical decision-making. *Smart Health*, 20:100178.
- Ranney, M. L., Griffeth, V., and Jha, A. K. (2020). Critical supply shortages—the need for ventilators and personal protective equipment during the COVID-19 pandemic. *New England Journal of Medicine*, 382(18):e41.
- Schöning, V., Liakoni, E., Baumgartner, C., Exadaktylos, A. K., Hautz, W. E., Atkinson, A., and Hammann, F. (2021). Development and validation of a prognostic COVID-19 severity assessment (COSA) score and machine learning models for patient triage at a tertiary hospital. *Journal of Translational Medicine*, 19(1):1–11.
- Vepa, A., Saleem, A., Rakhshan, K., Daneshkhah, A., Sedighi, T., Shohaimi, S., Omar, A., Salari, N., Chatrabgoun, O., Dharmaraj, D., et al. (2021). Using machine learning algorithms to develop a clinical decision-making tool for COVID-19 inpatients. *International journal of environmental research and public health*, 18(12):6228.
- Wynants, L., Van Calster, B., Collins, G. S., Riley, R. D., Heinze, G., Schuit, E., Bonten, M. M., Dahly, D. L., Damen, J. A., Debray, T. P., et al. (2020). Prediction models for diagnosis and prognosis of covid-19: systematic review and critical appraisal. *BMJ*, 369.
- Yadaw, A. S., Li, Y.-c., Bose, S., Iyengar, R., Bunyavanich, S., and Pandey, G. (2020). Clinical features of COVID-19 mortality: development and validation of a clinical prediction model. *The Lancet Digital Health*, 2(10):e516–e525.
- Zhao, Z., Chen, A., Hou, W., Graham, J. M., Li, H., Richman, P. S., Thode, H. C., Singer, A. J., and Duong, T. Q. (2020). Prediction model and risk scores of icu admission and mortality in COVID-19. *PloS one*, 15(7):e0236618.