

Machine Learning Applied in SARS-CoV-2 COVID-19 Screening using Clinical Analysis Parameters

R. F. A. P. Oliveira, C. J. A. Bastos-Filho, A. C. A. M. V. F. Medeiros, P. Buarque, D. L. Freire

Abstract—COVID-19 was considered a pandemic by the World Health Organization. Since then, world governments have coordinated information flows and issued guidelines to contain the overwhelming effects of this disease. At the same time, the scientific community is continually seeking information about transmission mechanisms, the clinical spectrum of the disease, new diagnoses, and strategies for prevention and treatment. One of the challenges is performing the tests for the diagnosis of the disease, whose technique adopted for the detection of the genetic material of COVID-19 requires equipment and specialized human resources, making it an expensive procedure. We hypothesize that machine learning techniques can be used to classify the test results for COVID-19 through the joint analysis of popular laboratory tests' clinical parameters. Machine learning techniques, such as Random Forest, Multi-Layer Perceptron, and Support Vector Machines Regression, enable the creation of disease prediction models and artificial intelligence techniques to analyze clinical parameters. Thus, we evaluated the existing correlations between laboratory parameters and the result of the COVID-19 test, and developed two classification models: the first classifies the test results for patients with suspected COVID-19, and the second classifies the hospitalization units of patients with COVID-19, both according to the laboratory parameters. The models achieved an accuracy above 96%, showing that they are promising to the classification of tests for COVID-19 and screen patients by hospitalization unit.

Index Terms—Artificial intelligence, Machine learning, Random forest, Multilayer perceptron, Support vector machine, COVID-19 diagnostic aid model

I. INTRODUÇÃO

Um novo tipo de infecção respiratória foi relatado na China, a partir de um surto de pneumonia causado por um agente posteriormente identificado, o SARS-Cov-2. A infecção pelo novo coronavírus (denominada de COVID-19) vem se mostrando um desafio para os sistemas de saúde mundiais. Dentre os desafios enfrentados no combate à pandemia, estão o custo e o tempo de processamento dos testes para detecção da doença. Para o diagnóstico, o padrão ouro é a reação em cadeia da polimerase (PCR), que analisa a presença do ácido nucléico viral. Porém, diversos fatores, como o tipo de amostra utilizada na reação, podem influenciar o resultado do exame, ocasionando variações na sensibilidade da técnica a depender da carga viral e do material utilizado [1]. Devido às diversas dificuldades inerentes ao processo de diagnóstico, como a necessidade de técnicas específicas e a complexidade

do PCR, os resultados dos testes acabam sendo demorados. O custo dos exames PCR é alto, quando comparado ao custo de exames laboratoriais, como os exames de sangue ou urina. Além disso, o PCR tem uma alta complexidade, uma vez que necessita de profissionais específicos e técnicas especializadas, além de que um único resultado não detectado com RT-PCR para SARS-CoV-2 não exclui o diagnóstico da infecção [2]. Estes aspectos dificultam o diagnóstico e acabam por aumentar as possibilidades de transmissão da doença e, conseqüentemente, as curvas de contágio. No Brasil, existem deficiências, tanto na falta de kits de teste, como na lentidão do processamento destes testes.

Diante deste cenário, médicos e hospitais estão utilizando informalmente o padrão da doença e suas manifestações laboratoriais para compor um painel para decisão a respeito das internações e também para acompanhar o monitoramento do prognóstico dos pacientes. Estudos científicos já sugerem um painel básico de exames laboratoriais para investigação inicial e avaliação prognóstica de pacientes com quadros moderados/graves de COVID-19, que estejam sob regime de internação hospitalar. Embora as características clínicas e moleculares do COVID-19 tenham sido amplamente estudadas e definidas, as alterações laboratoriais ainda não ganharam tanta atenção devido à falta de estudos e à pequena quantidade de dados disponível para mensurar o comportamento do corpo frente à infecção pelo vírus[3].

A medicina laboratorial desempenha um papel essencial na detecção, diagnóstico e tratamento precoces de muitas doenças, uma vez que o diagnóstico laboratorial ultrapassa os aspectos do diagnóstico etiológico e da vigilância epidemiológica dos testes *in vitro* [3]. Outro aspecto relevante é que a maior parte dos exames laboratoriais são acessíveis em redes públicas de saúde e, geralmente, apresentam baixo custo para a população em laboratórios clínicos da rede privada. O monitoramento de diversas alterações hematológicas e bioquímicas podem diferenciar a infecção do SARS-Cov-2 de outras possíveis infecções virais, bacterianas ou até de outros distúrbios metabólicos. Pesquisas realizadas com pacientes da China, primeiro epicentro da pandemia, revelaram que os pacientes positivos tinham padrões de alterações laboratoriais diretamente relacionados com a evolução da doença [4]. Embora diversas alterações já tenham sido descritas e relacionadas com a infecção do COVID-19, os resultados ainda são preliminares, uma vez que os estudos contam com poucos pacientes (120 amostras em média) e as pesquisas foram realizadas com um recorte populacional no país de origem (China) [4], não se sabendo ainda se os padrões se repetirão em outros países, como no Brasil, onde as alterações ainda precisam de mais

R. F. A. P. Oliveira, C. J. A. Bastos-Filho e P. Buarque, Universidade de Pernambuco, Recife, Pernambuco, Brasil

A. C. A. M. V. F. Medeiros, Universidade Federal de Pernambuco, Recife, Pernambuco, Brasil

D. L. Freire, Universidade de São Paulo, Instituto de Ciências Matemática e de Computação (ICMC), São Carlos, São Paulo, Brasil

estudos e validações para serem confirmadas.

A bioinformática é um campo de conhecimento que pode manipular grandes volumes de dados clínicos, por meio de técnicas de aprendizado de máquina. Ferramentas modernas proporcionam melhor visualização dos dados, favorecendo a extração de correlações e padrões a partir da análise dos dados. Além disso, as técnicas de aprendizado de máquina proporcionam a criação de modelos que permitem previsões [5], [6], [7]. Dentre as aplicações da bioinformática, estão o diagnóstico de doenças e a simulação de prognósticos por meio de modelos preditivos [8]. Floresta Aleatória (*Random Forest, RF*), Perceptron Multicamadas (*Multi-Layer Perceptron, MLP*) e Máquina de Vetores de Suporte (*Support Vector Machine, SVM*) são exemplos de algoritmos de aprendizado de máquina que são capazes modelar processos complexos [9], [10].

Este artigo tem como objetivo utilizar técnicas de aprendizado de máquina para encontrar padrões de comportamento dos parâmetros laboratoriais, pelos quais se possa prever casos positivos e negativos de COVID-19. Além disso, buscase a utilização destas técnicas para o apoio à triagem dos pacientes para unidades de atendimento, de acordo com a gravidade da doença, usando os padrões laboratoriais encontrados nas análises das bases de dados. Os modelos preditivos desenvolvidos alcançaram uma acurácia acima de 99%, quando comparada com os rótulos obtidos a partir dos exames PCR. Os resultados obtidos foram comprovados com testes estatísticos, assim como, validados por biomédicos especialistas em exames clínicos.

O resto do artigo está organizado da seguinte forma. Seção II fornece informações gerais sobre técnicas de aprendizagem de máquina. A Seção III discute trabalhos relacionados. A Seção IV reporta os experimentos e os resultados obtidos. A Seção V apresenta as conclusões.

II. REFERENCIAL TEÓRICO

A aprendizagem de máquina lida com técnicas computacionais capazes de se adaptar a partir das experiências [11]. Este aprendizado pode ser usado para diversas tarefas, entre elas, classificação e regressão. Dentre os algoritmos de aprendizagem de máquina mais utilizados estão RF, MLP e SVM.

RF é formado por várias árvores de decisão, criadas de forma aleatória, onde cada árvore estabelece regras para tomada de decisão, que são usadas na escolha do resultado final [12]. RF tem mostrado bom desempenho em diversos problemas de otimização e apresenta a característica relevante de expressar a importância relativa de cada característica de entrada no processo. RF usa a média para melhorar a precisão preditiva e controlar o *overfitting*. O *overfitting* é um comportamento característico de um modelo que se adaptou demais aos dados do treinamento, mas não é capaz de generalizar a predição para novos dados.

MLP é uma das arquiteturas de redes neurais artificiais com uma ou mais camadas ocultas, contendo um número pré-determinado de neurônios. MLPs podem ser usados na classificação de conjuntos de dados que não são linearmente separáveis. Em um MLP, usa-se primeiramente uma transformação não-linear que projeta os dados de entrada em um espaço onde eles se tornam linearmente separáveis [13].

SVM é uma abordagem fundamentada na teoria de aprendizado estatístico, baseadas em hiperplanos que separam os dados em espaços de alta dimensionalidade a partir do cálculo das matrizes inversas de Mel Penrose [14]. SVM pode ser usado para desafios de classificação ou regressão, sendo mais usado para classificação.

III. TRABALHOS RELACIONADOS

A urgência de encontrar ferramentas que auxiliem o combate à COVID-19 tem motivado a comunidade científica a produzir muitos trabalhos no campo da bioinformática [15]. Nestes trabalhos, técnicas de aprendizagem de máquinas têm sido utilizadas na classificação taxonômica dos genomas COVID-19 [16], no ensaio de detecção de COVID-19 baseado no CRISPR (do inglês *Clustered Regularly Interspaced Short Palindromic Repeats*) [17], na previsão de sobrevivência de pacientes com COVID-19 grave [18] e na descoberta de potenciais candidatos a medicamentos para tratamento da COVID-19 [19].

Nesta seção estão listados os trabalhos que, assim como este, utilizaram a inteligência artificial para análise de parâmetros de exames clínicos, com o propósito de auxiliar na classificação dos pacientes diagnosticados com e sem COVID-19, assim como na triagem dos pacientes com COVID-19 para unidades de tratamento, dependendo da gravidade dos sintomas da doença. Os trabalhos relacionados incluíram sintomas e imagens computadorizadas de exames clínicos para a classificação dos resultados dos exames de COVID-19, enquanto que, a proposta deste trabalho se concentrou apenas na análise de parâmetros clínicos de exames laboratoriais populares, como exames de sangue e urina.

Butt et al. [20] classificaram as imagens de tomografia computadorizada do tórax em três classes: COVID-19, Influenza-A pneumonia viral, e casos saudáveis. O conjunto de dados utilizado continha 618 imagens, que incluía 219 imagens de 110 pacientes com COVID-19, 224 imagens de 224 pacientes com pneumonia viral por Influenza-A e 175 imagens de 175 pacientes saudáveis. Os autores classificaram as imagens com modelo de aprendizagem profunda tridimensional e alcançaram uma acurácia de classificação geral de 86,7%. Ghoshal e Tucker [21] classificaram as imagens radiográficas de tórax em quatro classes: COVID-19, Pneumonia viral, Pneumonia bacteriana, e casos saudáveis. O conjunto de dados utilizado continha 5941 imagens, que incluía 68 imagens de pacientes com COVID-19, 1504 imagens de pacientes com pneumonia viral, 2786 imagens de pacientes com pneumonia bacteriana e 1583 pacientes saudáveis. Os autores usaram Redes Neurais Convolucionais Bayesianas para estimar a incerteza em soluções de aprendizagem profunda e alcançaram uma acurácia de classificação geral de 88,1%. Feng et al. [22] desenvolveram um modelo de suporte ao diagnóstico de COVID-19 considerando pacientes adultos com febre, utilizando seus históricos epidemiológicos (tais como resultados de exames para influenza A e B), sinais e sintomas clínicos, resultados de testes laboratoriais rotineiros (tais como linfopenia, valores da rotina para exame de sangue, biomarcadores relacionados a infecções) e achados positivos

da tomografia computadorizada do tórax. O conjunto de dados utilizado continha dados de 164 pacientes, sendo 106 sem suspeita de COVID-19 e 26 com suspeita. A técnica de regressão linear com LASSO foi utilizada para selecionar as características (*features*) importantes para treinar o modelo. Os resultados de desempenho do modelo para o conjunto de dados de teste e de validação, respectivamente, foram: AUC (*Area Under The Curve*) 0,841 e 0,938, *score F-1* de 0,571 e 0,667, *recall* de 1,000 e 1,000, especificidade de 0,727 e 0,778, e precisão de 0,400 e 0,500.

Yan et al. [18] analisaram dados de históricos epidemiológicos e manifestações clínicas, tais como febre e/ou sintomas respiratórios, contagem normal ou diminuída de leucócitos ou linfócitos, e características de exames de imagem de 375 pacientes com COVID-19. O objetivo foi identificar pacientes de alto risco, a fim de melhorar o prognóstico e reduzir a mortalidade de pacientes críticos. Os autores utilizaram o método de aprendizagem da máquina XGBoost para a identificação precoce de pacientes críticos. Este modelo permitiu a extração de três principais características clínicas (lactíc dehidrogenase, proteína C reativa de alta sensibilidade e linfócitos), dentre mais de 300 características. A precisão do modelo na predição da sobrevivência dos pacientes foi superior a 90%. Jiang et al [23] objetivou a predição do agravamento do COVID-19 em pacientes com os sintomas preliminares da doença. Este agravamento refere-se à propensão para o desenvolvimento da síndrome do desconforto respiratório agudo (do inglês *Acute respiratory distress syndrome*). O conjunto de dados foi composto por informações de 53 pacientes hospitalizados com COVID-19. As informações referiam-se a sintomas comuns (febre e tosse, diarreia, dispnéia, mialgias e sibilos, congestão nasal e dor de garganta), parâmetros de exames clínicos, exames de imagem computadorizada de tórax, histórico de outras doenças (hipertensão arterial, hepatite, diabetes) e tratamentos (oxigenoterapia e medicação). Seis algoritmos de predição foram testados: *Logistic Regression* (acurácia de 50%), *KNN* ($k=5$) (80%), *Decision Tree (based on Gain Ratio)* (acurácia de 70%), *Decision Tree (based on Gini Index)* (acurácia de 70%), *Random Forests* (acurácia de 70%) e *Support Vector Machine* (acurácia de 80%).

IV. EXPERIMENTO

A. Materiais e Métodos

O conjunto de dados da pesquisa, intitulado “*Diagnosis of COVID-19 and its clinical spectrum*”, foi disponibilizado na plataforma Kaggle [24], pelo Hospital Israelita Albert Einstein, contendo dados anônimos de pacientes atendidos neste hospital, localizado em São Paulo, Brasil. Os pacientes que compõem o conjunto de dados tiveram amostras coletadas para realizar o teste para SARS-CoV-2 e testes laboratoriais adicionais durante uma visita ao hospital. Todos os exames clínicos foram realizados pelos mesmos métodos e pelo mesmo grupo laboratorial. O conjunto de dados contém 111 variáveis referentes a parâmetros laboratoriais de exames clínicos dos pacientes, além da idade, resultados de exames para alguns vírus, resultados do exame para SARS-Cov-2 e tipo de unidade de internamento dos pacientes com COVID-19

(enfermaria, unidade semi-intensiva e unidade de atendimento intensiva). Há um total de 5644 amostras, uma para cada paciente. Deste total, há 558 pacientes com o resultado positivo e 5086 pacientes com o resultado negativo do teste para SARS-Cov-2.

Inicialmente, *RF* foi usado para selecionar as características (*features*) mais importantes para treinar os modelos. Depois, três algoritmos de aprendizado de máquina foram usados para o treinamento: *RF*, *MLP* e *SVM*. Os modelos desenvolvidos foram otimizados usando o método de busca em grade (*Grid-searching*) combinado ao método de validação cruzada *k-fold* (*KFoldCrossValidation*), a fim de se obter melhor desempenho em termos de tempo de resposta e custo computacional. A busca em grade é o processo de verificação dos dados para configurar os melhores parâmetros para um determinado modelo de aprendizado de máquina. Enquanto que, validação cruzada *k-fold* é um método para avaliar modelos de aprendizado de máquina, a partir de subconjuntos (*fold*) do conjunto de dados de entrada, onde *k* é o número de subconjuntos. A métrica usada para avaliar o desempenho foi a acurácia (*accuracy*), definida pela divisão do número de acertos pelo número total de amostras. Os resultados obtidos pelos modelos desenvolvidos foram avaliados e interpretados por especialistas, a fim de validá-los também no âmbito da biomedicina e possibilitar o uso efetivo dos modelos no meio científico.

B. Preparação dos Dados

A primeira etapa de preparação é a análise da natureza e distribuição dos dados e a limpeza dos dados. Como o conjunto de dados possui muitos dados ausentes, foi realizada uma limpeza vertical, eliminando as colunas pelas seguintes regras: (i) excluir coluna que tenha apenas valores ausentes; (ii) excluir coluna que tenham um valor que representa menos de 8% do total de amostras com valores presentes para a respectiva coluna; (iii) excluir coluna que cobre apenas um espectro da variável de saída. A segunda etapa de preparação é a imputação de valores nas amostras onde eles estão ausentes. Primeiro, todas as variáveis categóricas foram transformadas em variáveis numéricas. Depois, foi realizada uma simples imputação de dados numéricos pelo valor da média de cada coluna e imputação de dados categóricos pelo valor mais frequente de cada coluna. Este procedimento não muda significativamente as distribuições de dados, dado que as colunas e linhas eliminadas foram as de maior percentual de dados faltantes. A terceira etapa de preparação é o balanceamento de valores nas amostras. Para balancear o conjunto de dados, optou-se por igualar a quantidade de amostras das duas classes. Assim, foram escolhidas aleatoriamente 558 amostras de pacientes da classe de pacientes com o resultado negativo.

C. Seleção de Características

Após a preparação dos dados, o conjunto de dados resultante foi dividido aleatoriamente, com 80% das amostras para treinamento e 20% para testes. O ranking de importância das *features* foi realizado no conjunto de treinamento, usando o *Random Forest*, com 100 árvores de decisão e o tamanho dos

subconjuntos igual ao do conjunto de dados de entrada. A Fig. 1 mostra as 15 características mais importantes usadas no treinamento do modelo.

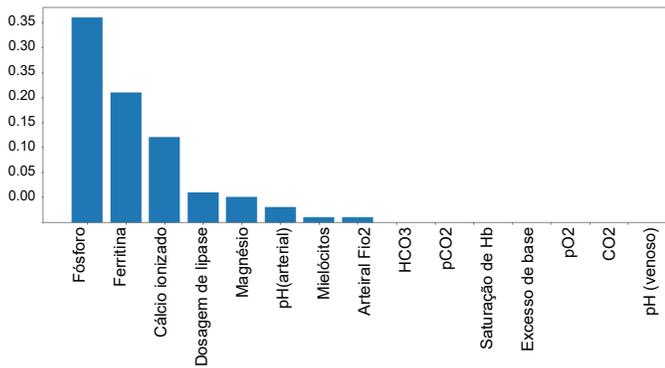


Fig. 1. Ranking das 15 características mais importantes.

D. Primeira Tarefa - Classificar Resultados

A primeira tarefa foi classificar uma nova amostra na classe de resultado negativo (Classe 0) ou na classe de resultado positivo (Classe 1) para COVID-19. Foram utilizadas apenas as 15 características mais importantes pré-selecionadas pelo RF.

1) *Desenvolvimento dos modelos*: O *KFoldCrossValidation* foi utilizado para avaliar os modelos de aprendizagem de máquinas usando o conjunto de dados resultante. Foi definido o parâmetro $k = 10$, que representa a quantidade na qual a base de dados foi dividida, para uma validação cruzada em 10 vezes. Um dos principais objetivos da validação cruzada é estimar a capacidade de um modelo de aprendizagem de máquina, quando os dados não são aqueles utilizados durante o treinamento do modelo. O *GridSearchCV* foi usado para automatizar o processo de ajuste de parâmetros, criando combinações destes parâmetros e posteriormente avaliando aquela que traz os melhores resultados. Os parâmetros testados e o melhores estimadores encontrados são mostrados na Tabela I para cada modelo treinado. *KFoldCrossValidation* e *GridSearchCV* pertencem a biblioteca *scikit-learn* [25] do Python.

2) *Resultados*: O modelo treinado apresentou uma *accuracy* de 100% com o algoritmo RF, de 99,40% com MLP e de 99,89% com SVM. O resumo das métricas aferidas para os modelos são mostradas na Tabela II. A *precision* é a capacidade do classificador de não rotular como positiva uma amostra negativa. O *recall* é a capacidade do classificador de encontrar todas as amostras positivas.

A *f1-score* é uma média harmônica ponderada da *precision* e do *recall*. O valor de *f1-score* varia de 0 a 1, sendo 0 o pior valor e 1 o melhor. O suporte é o número de ocorrências de cada classe. As médias relatadas incluem *accuracy*, média macro (ou média não ponderada) e média ponderada.

A matriz de confusão da Fig. 2 resume a precisão das previsões dos modelos treinados com cada uma das três algoritmos. A matriz (A) mostra a *precision* do modelo treinado com o algoritmo RF, na qual o modelo previu assertivamente

TABELA I
PARÂMETROS TESTADOS NA TAREFA 1.

Parâmetros RF	Valores testados	Melhor estimador
Critério	gini, entropy	gini
Nº mínimo de amostras para folhas	1, 5, 10, 25, 50, 70	1
Nº mínimo de amostras para dividir um nó	2, 4, 10, 12, 16, 18, 25, 35	2
Nº de estimadores (<i>trees</i>)	100, 400, 700, 1000, 1500	100
Parâmetros MLP	Valores testados	Melhor estimador
Nº de camadas ocultas	64	18
Nº de número de neurônios em cada camada	30	30
Número de iterações	5000	5000
Funções de Ativação	tanh, relu, logistic	tahn
Algoritmos de otimização	lbfgs, sgd, adam	lbfgs
Função de inicialização	constant	constant
Taxa de aprendizado inicial	0,001; 0,01; 0,1	0,01
Métrica do estimador	roc_auc	-
Taxa de aprendizado	-	0,0001
Parâmetros SVM	Valores testados	Melhor estimador
Tipo específico de kernel	rbf, linear, poly, sigmoid	linear
Parâmetro de regularização (C)	linspace(1, 50, 10)	1
Coefficiente Kernel(γ)	linspace(1, 100, 50)	1

todos os casos negativos e todos positivos para COVID-19. A matriz (B) mostra a precisão do modelo treinado com o algoritmo MLP e a matriz (C) com o algoritmo SVM, nas quais os modelos previram assertivamente dos 83 dos 84 casos negativos para COVID-19 e todos os casos positivos para a doença.

0	84	0	0	83	1	0	83	1
1	0	84	1	0	84	1	0	84
	0	1		0	1		0	1
	RF (A)			MLP (B)			SVM (C)	

Fig. 2. Matrizes de confusão da tarefa 1

E. Segunda Tarefa - Classificar Unidades de Tratamento

A segunda tarefa foi classificar uma nova amostra em uma das classes referente à unidade de internação para qual um paciente deverá ser alocado, de acordo seus parâmetros clínicos. As classes são: pacientes de enfermaria (Classe 0), pacientes de unidade semi-intensiva (Classe 1) e pacientes de unidade de atendimento intensiva (Classe 2). Para cada classe, o valor 1 é atribuído para pacientes admitidos e o valor 0 para pacientes não admitidos na unidade de internação correspondente. Para esta tarefa, apenas as amostras dos casos positivos para COVID-19 foram utilizadas, considerando o conjunto de dados resultante da limpeza e utilizando as 15 mais importantes características.

TABELA II
MÉTRICAS DO MODELO TREINADO NA TAREFA 1.

	Algoritmo de aprendizado de máquina									
	RF			MLP			SVM			suporte
	precision	recall	f-score	precision	recall	f-score	precision	recall	f-score	
Classe 0	1,00	1,00	1,00	1,00	0,99	0,99	1,00	0,99	0,99	84
Classe 1	1,00	1,00	1,00	0,99	1,00	0,99	1,00	0,99	0,99	84
accuracy			1,00			0,99			0,99	168
média macro	1,00	1,00	1,00	0,99	0,99	0,99	0,99	0,99	0,99	168
média ponderada	1,00	1,00	1,00	0,99	0,99	0,99	0,99	0,99	0,99	168

1) *Desenvolvimento dos modelos*: Os algoritmos usados no treinamento dos modelos para esta tarefa, foram os mesmos usados na primeira tarefa: RF, MLP e SVM. Os parâmetros testados e os melhores estimadores encontrados são mostrados na Tabela III para os modelos treinados com RF, MLP e SVM. Na implementação dos modelos também foram utilizadas as bibliotecas *KFoldCrossValidation* e *GridSearchCV* para otimização dos parâmetros. Também foi utilizada a biblioteca *OneVsRestClassifier* para o algoritmo SVM com múltiplas classes.

TABELA III
PARÂMETROS TESTADOS NA TAREFA 2.

Parâmetros RF	Valores testados	Melhor estimador
Critério	gini, entropy	entropy
Nº mínimo de amostras para folhas	1, 5, 10, 25, 50, 70	1
Nº mínimo de amostras para dividir um nó	2, 4, 10, 12, 16, 18, 25, 35	2
Nº de estimadores (<i>trees</i>)	100, 400, 700, 1000, 1500	100
Parâmetros MLP	Valores testados	Melhor estimador
Nº de camadas ocultas	64	14
Nº de número de neurônios em cada camada	30	30
Número de iterações	5000	5000
Funções de Ativação	tanh, relu, logistic	tahn
Algoritmos de otimização	lbfgs, sgd, adam	lbfgs
Função de inicialização	constant	constant
Taxa de aprendizado inicial	0,001, 0,01, 0,1	0,1
Métrica do estimador	f1_micro	-
Taxa de aprendizado	-	0,0001
Parâmetros SVM	Valores testados	Melhor estimador
Tipo específico de kernel	rbf, linear, poly, sig-moid	linear
Parâmetro de regularização (<i>C</i>)	linspace(1, 50, 10)	11,88

2) *Resultados*: Os modelos treinados com os algoritmos RF e MLP apresentaram uma *accuracy* de 92,86% e o modelo treinado com SVM apresentou uma *accuracy* de 100%. O resumo das métricas aferidas para o modelo são mostradas na Tabela IV. Neste caso, como se trata de classificação de três classes (múltiplas), além da média macro e da média ponderada, tem-se a média micro e a média das amostras. A média micro é a média do total de verdadeiros positivos, falsos negativos e falsos positivos.

As matrizes de confusão da Fig. 3 (A) resumem a precisão das previsões do modelo treinado com o algoritmo RF. A

matriz da classe 0 mostra que o modelo previu assertivamente dos 76 dos 78 casos negativos para admissão em enfermaria e 3 dos 6 casos positivos. A matriz da classe 1 mostra que o modelo previu assertivamente todos os 83 casos negativos para admissão em semi-UTI e errou o caso positivo. A matriz da classe 2 mostra que o modelo previu assertivamente dos 80 dos 83 casos negativos para admissão em UTI e errou o caso positivos.

As matrizes de confusão da Fig. 3 (B) resumem a precisão das previsões do modelo treinado com o algoritmo MLP. A matriz da classe 0 mostra que o modelo previu assertivamente todos os casos negativos para admissão em enfermaria e 2 dos 6 casos positivos. A matriz da classe 1 mostra que o modelo previu assertivamente todos os 83 casos negativos para admissão em semi-UTI e errou o caso positivo. A matriz da classe 2 mostra que o modelo previu assertivamente todos os 83 casos negativos para admissão em UTI e errou o caso positivo.

As matrizes de confusão da Fig. 3 (C) resumem a precisão das previsões do modelo treinado com o algoritmo SVM. A matriz da classe 0 mostra que o modelo previu assertivamente 75 dos 78 casos negativos para admissão em enfermaria e 4 dos 6 casos positivos. A matriz da classe 1 mostra que o modelo previu assertivamente 82 dos 83 casos negativos para admissão em semi-UTI e errou o caso positivo. A matriz da classe 2 mostra que o modelo previu assertivamente 82 dos 83 casos negativos para admissão em UTI e o caso positivo.

F. Análise Estatística

Nesta secção, descreve-se o experimento usado para comparar o desempenho dos algoritmos de aprendizagem de máquina RF, MLP e SVM, avaliando a acurácia dos modelos na classificação dos resultados dos exames de COVID-19 e na classificação das unidades de internação para pacientes infectados pelo vírus SARS-Cov-2. O protocolo aplicado foi baseado nos trabalhos de [26], [27] e [28], que indicam os procedimentos para experimentos controlados no campo de estudos de engenharia.

1) *Variáveis*: A variável independente é o algoritmo de aprendizagem de máquina. Os valores testados para esta variável foram: RF, MLP e SVM. A variável dependente é a *accuracy*, referente a porcentagem de acerto do algoritmo de aprendizagem de máquina usada para classificação.

2) *Execução*: Neste experimento, cada um dos três algoritmos de aprendizagem de máquina foi executada 30 vezes, resultando em 90 diferentes cenários.

TABELA IV
MÉTRICAS DO MODELO TREINADO NA TAREFA 2.

	Algoritmo de aprendizado de máquina									
	RF			MLP			SVM			suporte
	precision	recall	f-score	precision	recall	f-score	precision	recall	f-score	
Classe 0	1,00	0,33	0,50	0,56	0,83	0,67	0,40	0,33	0,36	6
Classe 1	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	1
Classe 2	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	1
média micro	1,00	0,25	0,40	0,50	0,62	0,56	0,29	0,25	0,27	8
média macro	0,33	0,11	0,17	0,19	0,28	0,22	0,13	0,11	0,12	8
média ponderada	0,75	0,25	0,38	0,42	0,62	0,50	0,30	0,25	0,27	8
média das amostras	0,02	0,02	0,02	0,06	0,06	0,06	0,02	0,02	0,02	8

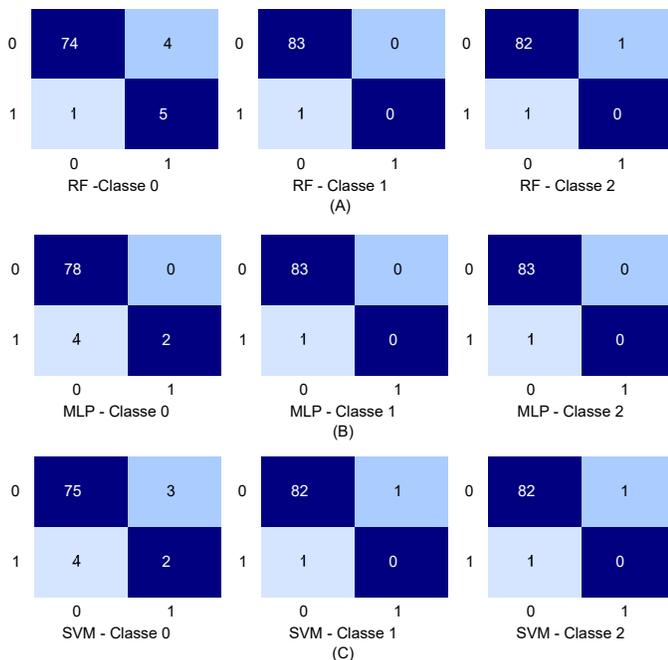


Fig. 3. Matrices de confusão da tarefa 2.

A técnica estatística de análise de variância (ANOVA) [29] foi utilizada para diferenciar, entre as variações encontradas em um conjunto de resultados, aquelas que são derivadas de fatores aleatórios, chamados erros, daquelas que são influenciadas pelos tratamentos. Neste experimento, os tratamentos são os algoritmos de aprendizado de máquina: RF, MLP e SVM. Como houve uma diferença estatística nos resultados da ANOVA, foi usada a técnica de Tukey HSD (do inglês *honestly significant difference*) [30] para descobrir se havia diferença entre os tratamentos.

A Tabela V mostra a ANOVA [29] para a tarefa 1 e a Tabela VI mostra as comparações par a par dos tratamentos com intervalos de confiança Tukey HSD [30] com nível de significância do teste igual a 0,05. Para a tarefa 1, os valores resultantes para *f-value* e *p-value* foram, respectivamente, 3,931627 e 0,023191. As colunas da Tabela VI são os pares de tratamentos (Pares), diferença entre as médias (Dif.), *p-value* ajustados para Tukey HSD (*p-adj*), a mínima diferença entre as médias (Mín.), a máxima diferença (Máx.) e a coluna “Rejeita”, que pode ter o valores “sim” ou “não”, Se o valor desta coluna for “não”, há diferença estatística entre os

tratamentos, caso contrário, não há. Para a tarefa 2, os valores resultantes para *f-value* e *p-value* foram, respectivamente, 588,7374 e 2,7289 e-51. Para a tarefa 2, a ANOVA é mostrada na Tabela VII e o teste Tukey HSD é mostrado na Tabela VIII.

TABELA V
ANÁLISE DE VARIÂNCIA PARA TAREFA 1.

Fontes de Variação	QM	GL	<i>f-value</i>	<i>p-value</i>
Tratamentos	0.198	2	3.931627	0.023191
Erro	2.1987	87		

SQM - Quadrado médio; GL - Grau de liberdade.

TABELA VI
TESTE DE TUKEY HSD PARA TAREFA 1.

Pares	Dif.	<i>p-adj</i>	Mín.	Máx.	Rejeita
MLP x RF	0.0997	0.045	0.0018	0.1975	Sim
MLP x SVM	0.0997	0.045	0.0018	0.1975	Sim
RF x SVM	0.0	0.9	-0.0979	0.0979	Não

TABELA VII
ANÁLISE DE VARIÂNCIA PARA TAREFA 2.

Fontes de Variação	QM	GL	<i>f-value</i>	<i>p-value</i>
Tratamentos	1639.9021	2	588.737	2.728 e-51
Erro	121.1673	87		

SQM - Quadrado médio; GL - Grau de liberdade.

TABELA VIII
TESTE DE TUKEY HSD PARA TAREFA 2.

Pares	Dif.	<i>p-adj</i>	Mín.	Máx.	Rejeita
MLP x RF	7.1057	0.001	6.3791	7.8323	Sim
MLP x SVM	10.1957	0.001	9.4691	10.9223	Sim
RF x SVM	3.0900	0.001	2.3634	3.8166	Sim

G. Discussões

Na primeira tarefa, percebe-se que os modelos, treinados com os algoritmos RF, MLP e SVM, foram assertivos tanto para os casos negativos, como para os casos positivos na previsão dos pacientes com COVID-19. A *accuracy* foi acima de 99% nos três algoritmos. A análise de variância resultou em um *f-value* < 0,05, indicando que houve diferença estatística e o teste Tukey HSD, indicou que não houve diferença entre os pares de tratamentos MLP e RF, nem entre os pares MLP

e SVM, mas, houve diferença entre os pares de tratamentos RF e SVM.

Na segunda tarefa, percebe-se que o desbalanceamento entre casos positivos e negativos de admissão nas unidades de atendimento fez com que os modelos acertassem mais de 94% dos casos negativos de admissão e errassem os casos positivos de admissão. Tal resultado permite inferir que a assertividade da não admissão dos pacientes nas unidades impede a ocupação desnecessária nas unidades hospitalares, mitigando o problema de falta de leitos. O baixo número de casos positivos de admissão não permite generalizar a predição do modelo para essa tarefa. Porém, pode-se inferir que o paciente será prejudicado, a medida que o mesmo receberá a indicação de internamento, podendo ser, então, melhor avaliado pelos médicos, os quais decidiram sobre a permanência do paciente na unidade de internação. A análise de variância resultou em um $f\text{-value} < 0,05$, indicando que houve diferença estatística e o teste Tukey HSD, indicou que não houve diferença entre nenhum dos pares de tratamentos.

A validade de construção discute se o planejamento e a execução do estudo estão adequados para responder às perguntas da pesquisa. Para isto, o experimento foi planejado de acordo com os procedimentos da engenharia de software empírica [26], [28], [27].

A validade interna visa garantir que o tratamento tenha causado o resultado, atenuando os efeitos de outros fatores incertos ou não medidos [31]. O experimento foi realizado no mesmo ambiente, a fim de evitar a interferência do ambiente na execução dos algoritmos. Além disso, os pesquisadores e especializada da biomedicina inspecionaram com profundidade os procedimentos e usaram testes estatísticos para validar as medidas. A validade externa concentra-se na generalização dos resultados fora do escopo de nosso estudo [31]. Destaca-se também que, a mudança no método usado para avaliar os parâmetros clínicos pode alterar a relevância de algumas características e acurácia do modelo. Logo, para que se possa ter um modelo preditivo genérico, é necessário treiná-lo com um conjunto de dados mais abrangente, em relação aos métodos de coleta e análise dos parâmetros clínicos. O experimento é válido para testar outros algoritmos de aprendizado de máquina. Como trabalho futuro, serão realizados experimentos com um conjunto de dados maior e mais balanceado para avaliar a generalização dos resultados. Também pretende-se testar outras técnicas de imputação de dados e de aprendizado de máquina.

V. CONCLUSÕES

Neste artigo, algoritmos de aprendizado de máquinas foram usados com sucesso para treinar modelos para duas tarefas de classificação para auxílio ao combate ao COVID-19. A primeira tarefa foi classificar os resultados positivos e negativos de exames para COVID-19 e a segunda tarefa foi classificar a unidade de internação do paciente com COVID-19, sendo possíveis as unidades: enfermagem, unidade semi-intensiva ou de unidade de atendimento intensiva. Foi utilizando um conjunto de dados contendo parâmetros laboratoriais de resultados de exames clínicos de pacientes com suspeita de COVID-19.

Por meio do algoritmo RF, foram encontrados os parâmetros laboratoriais mais importantes para embasar o diagnóstico da doença. Para as tarefas de classificação, os algoritmos de aprendizado de máquinas utilizados foram RF, MLP e SVM. Com os três algoritmos, a acurácia dos modelos foi superior a 99% para a primeira tarefa de classificação e superior a 92% para a segunda.

Os resultados dos desempenhos dos modelos foram analisados estatisticamente, usando os testes ANOVA e Tukey HSD, e validados por uma especialista em biomedicina. Como trabalhos futuros, planeja-se analisar um conjunto de dados com menos dados faltantes e aplicar outras técnicas de aprendizado de máquina, tanto para imputação de dados, como para o treinamento dos modelos.

O artigo contribuiu para o combate à pandemia do COVID-19, visto que, a partir dos modelos preditivos desenvolvidos com aprendizagem de máquina, é possível se obter uma indicação positiva ou negativa da doença, assim como, pode-se obter uma indicação da unidade de internação para o paciente. Logo, os modelos provaram ser uma boa alternativa para o suporte à tomada de decisão pelos profissionais de saúde. Além disso, o diagnóstico da doença pode se tornar mais barato e acessível para a população, visto que exames laboratoriais dos quais são extraídos os parâmetros laboratoriais analisados pelos modelos, costumam ser menos onerosos que os atuais testes para COVID-19.

REFERÊNCIAS

- [1] S.-Y. Xiao, Y. Wu, and H. Liu, "Evolving status of the 2019 novel coronavirus infection: Proposal of conventional serologic assays for disease diagnosis and infection monitoring," *Journal of medical virology*, vol. 92, no. 5, pp. 464–467, 2020.
- [2] W. Wang, Y. Xu, R. Gao, R. Lu, K. Han, G. Wu, and W. Tan, "Detection of sars-cov-2 in different types of clinical specimens," *Jama*, vol. 323, no. 18, pp. 1843–1844, 2020.
- [3] G. Lippi and M. Plebani, "Laboratory abnormalities in patients with covid-2019 infection," *Clinical Chemistry and Laboratory Medicine (CCLM)*, vol. 58, no. 7, pp. 1131–1134, 2020.
- [4] D. Wang, B. Hu, C. Hu, F. Zhu, X. Liu, J. Zhang, B. Wang, H. Xiang, Z. Cheng, Y. Xiong *et al.*, "Clinical characteristics of 138 hospitalized patients with 2019 novel coronavirus-infected pneumonia in Wuhan, China," *Jama*, vol. 323, no. 11, pp. 1061–1069, 2020.
- [5] C. C. M. Davi, D. S. Silveira, and F. B. L. Neto, "A framework using computational intelligence techniques for decision support systems in medicine," *IEEE Latin America Transactions*, vol. 12, no. 2, pp. 205–211, 2014.
- [6] F. E. S. Alencar, D. C. Lopes, and F. M. Mendes Neto, "Development of a system classification of images dermoscopic for mobile devices," *IEEE Latin America Transactions*, vol. 14, no. 1, pp. 325–330, 2016.
- [7] A. Urrutia, E. Chavez, R. Motz, and R. Gajardo, "An ontology to assess data quality domains. a case study applied to a health care entity," *IEEE Latin America Transactions*, vol. 15, no. 8, pp. 1506–1512, 2017.
- [8] J.-j. Zhang, X. Dong, Y.-y. Cao, Y.-d. Yuan, Y.-b. Yang, Y.-q. Yan, C. A. Akdis, and Y.-d. Gao, "Clinical characteristics of 140 patients infected with sars-cov-2 in wuhan, china," *Allergy*, 2020.
- [9] T. Ma and A. Zhang, "Multi-view factorization autoencoder with network constraints for multi-omic integrative analysis," in *IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. IEEE, 2018, pp. 702–707.
- [10] D. Sun, M. Wang, and A. Li, "A multimodal deep neural network for human breast cancer prognosis prediction by integrating multi-dimensional data," *IEEE/ACM transactions on computational biology and bioinformatics*, vol. 16, no. 3, pp. 841–850, 2018.
- [11] B. Mirza, W. Wang, J. Wang, H. Choi, N. C. Chung, and P. Ping, "Machine learning and integrative analysis of biomedical big data," *Genes*, vol. 10, no. 2, p. 87, 2019.
- [12] L. Breiman, *Machine Learning*. Dordrecht, 1997, vol. 45, no. 15–32.

- [13] F. Murtagh, "Multilayer perceptrons for classification and regression," *Neurocomputing*, vol. 2, no. 5-6, pp. 183–197, 1991.
- [14] V. Vapnik, *The nature of statistical learning theory*. Springer science & business media, 2013.
- [15] A. Alimadadi, S. Aryal, I. Manandhar, P. Munroe, B. Joe, and X. Cheng, "Artificial intelligence and machine learning to fight covid-19." *Physiological genomics*, vol. 52, no. 4, pp. 200–202, 2020.
- [16] G. S. Randhawa, M. P. Soltysiak, H. El Roz, C. P. de Souza, K. A. Hill, and L. Kari, "Machine learning using intrinsic genomic signatures for rapid classification of novel pathogens: Covid-19 case study," *PLoS One*, vol. 15, no. 4, pp. 1–24, 2020.
- [17] H. C. Metsky, C. A. Freije, T.-S. F. Kosoko-Thoroddsen, P. C. Sabeti, and C. Myhrvold, "Crispr-based covid-19 surveillance using a genomically-comprehensive machine learning approach," *bioRxiv*, 2020.
- [18] L. Yan, H.-T. Zhang, Y. Xiao, M. Wang, C. Sun, J. Liang, S. Li, M. Zhang, Y. Guo, Y. Xiao *et al.*, "Prediction of survival for severe covid-19 patients with three clinical features: development of a machine learning-based prognostic model with clinical data in wuhan," *medRxiv*, 2020.
- [19] Y. Ge, T. Tian, S. Huang, F. Wan, J. Li, S. Li, H. Yang, L. Hong, N. Wu, E. Yuan *et al.*, "A data-driven drug repositioning framework discovered a potential therapeutic agent targeting covid-19," *bioRxiv*, 2020.
- [20] C. Butt, J. Gill, D. Chun, and B. A. Babu, "Deep learning system to screen coronavirus disease 2019 pneumonia," *Applied Intelligence*, vol. 1, no. 1, pp. 1–7, 2020.
- [21] B. Ghoshal and A. Tucker, "Estimating uncertainty and interpretability in deep learning for coronavirus (covid-19) detection," *arXiv*, pp. arXiv–2003, 2020.
- [22] C. Feng, Z. Huang, L. Wang, X. Chen, Y. Zhai, F. Zhu, H. Chen, Y. Wang, X. Su, S. Huang *et al.*, "A novel triage tool of artificial intelligence assisted diagnosis aid system for suspected covid-19 pneumonia in fever clinics," *medRxiv*, 2020.
- [23] X. Jiang, M. Coffee, A. Bari, J. Wang, X. Jiang, J. Shi, J. Dai, J. Cai, T. Zhang, Z. Wu *et al.*, "Towards an artificial intelligence framework for data-driven prediction of coronavirus clinical severity," *Computers, Materials & Continua*, vol. 63, no. 1, pp. 537–551, 2020.
- [24] "Kaggle," kaggle.com/pedrojb/diagnosis-of-covid-19-a-careful-analysis-v10-0, Acessado em: 2020-05-28.
- [25] "scikit," scikit-learn.org, Acessado em: 2020-05-28.
- [26] A. Jedlitschka and D. Pfahl, "Reporting guidelines for controlled experiments in software engineering," in *Int. Sym. Empirical Soft. Engineering (ESEM)*, 2005, pp. 95–104.
- [27] C. Wohlin, P. Runeson, M. Höst, M. C. Ohlsson, B. Regnell, and A. Wesslén, *Experimentation in software engineering*. Springer Science & Business Media, 2012.
- [28] V. R. Basili, D. Rombach, K. S. B. Kitchenham, D. Selby, and R. W. Pfahl, *Empirical Software Engineering Issues*. Springer Berlin/Heidelberg, 2007.
- [29] G. N. Wilkinson and C. E. Rogers, "Symbolic description of factorial models for analysis of variance," *Journal of the Royal Statistical Society Series C*, vol. 22, no. 3, pp. 392–399, 1973.
- [30] J. Tukey, "Multiple comparisons," *Journal of the American Statistical Association*, vol. 48, no. 264, pp. 624–25, 1953.
- [31] R. Feldt and A. Magazinius, "Validity threats in empirical software engineering research-an initial survey," in *Int. Conf. on Software Engineering and Knowledge Engineering*, 2010, pp. 374–379.



C. J. A. Bastos-Filho é, atualmente, Diretor de ambientes tecnológicos e formação superior da Secretaria de Ciência, Tecnologia e Inovação de Pernambuco, professor associado do quadro efetivo da Escola Politécnica da Universidade de Pernambuco (UPE) e livre docente. É conselheiro da Sociedade Brasileira de Inteligência Computacional, membro sênior do IEEE, e bolsista de Produtividade em Pesquisa do CNPq desde 2008, sendo promovido a bolsista nível 1D em 2018.



A. C. A. M. V. F. Medeiros é graduada em Biomedicina pela Universidade Federal de Pernambuco e atualmente participa do programa de pós graduação da Faculdade Global, cursando Biologia Molecular. É Pós graduada pela Famart em Análises Clínicas e Microbiologia. É integrante do grupo de pesquisa internacional ALiLife em projetos de Biotecnologia e atua como Biomédica na Pickcells em projetos de pesquisa utilizando tecnologia aplicada às análises clínicas.



P. Buarque é bacharel e mestrando em Engenharia da Computação pela Universidade de Pernambuco (UPE) com foco em uso de Inteligência Artificial aplicada na saúde e Inteligência de Enxames em aplicações de engenharia. Participou do programa de graduação sanduíche em Washington University e do *summer research* em Illinois Institute of Technology, onde fez pesquisa na área de algoritmos para suporte a infraestrutura Big Data.



D. L. Freire é pós-doutoranda em Inteligência Artificial, no Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo (ICMC-USP). Doutora em Modelagem Matemática pela Universidade Regional do Noroeste do Estado do Rio Grande do Sul (UNIJUÍ), com bolsa CAPES e Doutora em Ciências da Informação e Tecnologias pelo Instituto Universitário de Lisboa (ISCTE-IUL).



R. F. A. P. Oliveira atua na área de Tecnologia da Informação com ênfase em Segurança da Informação, Inteligência Computacional e Computação em Nuvem. É doutorando e mestre em Engenharia da Computação pela UPE, além de Especialista em Engenharia de Software pela UPE, graduado em Licenciatura da Computação pela UFRPE. Cofundador e diretor de tecnologia da PickCells, startup de detecção de doenças com Inteligência Artificial.