

CENTRO UNIVERSITÁRIO SENAC  
SANTO AMARO

Eduardo Albino Gonelli  
Matheus Kiss  
Rafael Marcos Tomé  
Renato Lins da Palma

**Machine Learning - Detecção de Sars-Cov2: Estudo com Florestas  
Randômicas e Redes Neurais Convolucionais**

Trabalho apresentado como exigência  
parcial para conclusão do módulo Machine  
Learning - Técnicas e Cases, do Curso de  
Pós-graduação em Inteligência Artificial.

São Paulo  
2023

## RESUMO

A recente epidemia do vírus denominado "Corona Vírus", Sars-CoV-2 ou COVID-19, vitimou centenas de milhares de pessoas ao redor do mundo. Conforme relatório da Organização Mundial da Saúde (OMS), estima-se que, no primeiro ano da pandemia, houve mais de 4,5 milhões de mortes (OPAS, 2022). A semelhança dos sintomas da COVID-19 com os da influenza e outras doenças respiratórias causa preocupação às autoridades públicas, pois pessoas contaminadas podem deixar de procurar ajuda por acreditarem estar com um resfriado comum. A OMS afirma que "Ambas, influenza e COVID-19, são doenças respiratórias com transmissão similar e possuem sintomas similares, com exceção da perda do paladar e do olfato" e também que "apenas o teste pode diferenciar a influenza da COVID-19" (OPAS, 2020 - tradução nossa). Dada relevância no contexto de saúde pública e possibilidades trazidas pelos avanços nas técnicas de inteligência artificial e redes neurais profundas, propõe-se, neste trabalho, o estudo de técnicas de inteligência artificial para classificação de imagens de tomografia com o objetivo de identificar de COVID-19 e sua diferenciá-los de outras doenças.

**Palavras-chave:** COVID-19, Diagnóstico de COVID-19, Redes Neurais Profundas, Aprendizado de Máquina, Classificação de Imagens.

## **ABSTRACT**

The recent epidemic of the virus known as "Corona Virus", Sars-CoV-2, or COVID-19, has claimed the lives of hundreds of thousands of people around the world. According to a report from the World Health Organization (WHO), it is estimated that there were more than 4.5 million deaths in the first year of the pandemic (PAHO, 2022). The similarity of COVID-19 symptoms with those of influenza and other respiratory diseases is a cause for concern among public authorities, as infected individuals may refrain from seeking help, believing they have a common cold. The WHO states that "Both influenza and COVID-19 are respiratory diseases with similar transmission and have similar symptoms, with the exception of the loss of taste and smell" and also that "only testing can differentiate influenza from COVID-19" (PAHO, 2020 - our translation). Given the relevance in the public health context and the possibilities brought about by advances in artificial intelligence techniques and deep neural networks, this work proposes the study of artificial intelligence techniques for the classification of tomography images with the objective of identifying COVID-19 and differentiating it from other diseases.

**Keywords:** COVID-19, COVID-19 Diagnosis, Deep Neural Networks, Machine Learning, Image Classification.

## SUMÁRIO

1. INTRODUÇÃO .....	5
2. DESENVOLVIMENTO .....	6
2.1. ANÁLISE DO <i>DATASET</i> .....	6
2.2. ESTRUTURAÇÃO DA SOLUÇÃO E PADRONIZAÇÃO DO DATASET .....	6
2.3. ESCOLHA DOS MODELOS .....	7
2.4. TREINAMENTO DO MODELO COM FLORESTAS RANDÔMICAS .....	8
2.4.1. PRIMEIRO TREINAMENTO .....	8
2.4.2. SEGUNDO TREINAMENTO - SELEÇÃO AUTOMATIZADA DE HIPERPARÂMETROS .....	9
2.4.3. DATA AUGMENTATION .....	10
2.4.4. COMPARAÇÃO DOS RESULTADOS COM OUTRAS MÉTRICAS .....	12
2.5. TREINAMENTO UTILIZANDO REDES NEURAIIS COM TENSORFLOW E KERAS .....	13
2.5.1. DEFINIÇÃO DO MODELO E PARÂMETROS .....	13
2.5.2. TREINAMENTOS .....	14
3.0. CONCLUSÃO .....	22
REFERÊNCIAS .....	23

## 1. INTRODUÇÃO

A epidemia de COVID-19 foi marcada por uma série de eventos importantes na área da tecnologia como, por exemplo, o desenvolvimento em tempo recorde de várias vacinas dado o contexto avançado da tecnologia na produção de vacinas e o grande investimento dos governos em vários países (GOVERNO DO ESTADO DO CEARÁ, 2021).

Outro tema importante da área da tecnologia, que foi utilizado extensivamente, foi o aprendizado de máquina para diagnósticação de COVID-19. Soares et. al. (2020) realizaram um excelente trabalho de coleta de imagens de tomografias de pessoas diagnosticadas com COVID-19 e de pessoas com suspeita de COVID-19 (que foram negativados nos testes clínicos). Esses dados foram disponibilizados publicamente e, concomitantemente, incentivaram o estudo de técnicas de inteligência artificial para a classificação de imagens para identificação de casos positivos de COVID-19, bem como implementaram uma solução com uma pontuação *F1 score*<sup>1</sup> de 97,31%, o qual consideraram excelente. (SOARES et. al., 2020).

Este *dataset*<sup>2</sup> foi fornecido na plataforma de treinamento e competições de inteligência artificial Kaggle, disponibilizada gratuitamente pelo Google para a comunidade. Além de fornecer o *dataset*, a plataforma fornece também o ambiente de desenvolvimento com as ferramentas necessárias para os estudos (GOOGLE, 2023).

O *dataset* está sob licença Creative Commons (CC BY-NC-ND 4-0), descrita no link: <https://creativecommons.org/licenses/by-nc-nd/4.0/>, e disponibilizado no Kaggle, no link: <https://www.kaggle.com/datasets/plameneduardo/sarscov2-ctscan-dataset>.

---

<sup>1</sup> "Essa métrica combina precisão e recall de modo a trazer um número único que indique a qualidade geral do seu modelo e trabalha bem até com conjuntos de dados que possuem classes desproporcionais." (LEAL, 2017)

<sup>2</sup> Conjunto de dados compilados sobre determinado assunto.

Aproveitando-se da disponibilização deste *dataset*, por meio de estudos acadêmicos, documentações e tutoriais disponibilizados na internet sobre inteligência artificial e aprendizado de máquina, este trabalho estuda técnicas sobre classificação de imagens utilizando ferramentas como Scikit-learn (SCIKIT LEARN, 2023), TensorFlow (TENSORFLOW, 2023) e Redes Neurais Convolucionais.

## **2. DESENVOLVIMENTO**

### **2.1. ANÁLISE DO DATASET**

O *dataset* é composto por 1.252 imagens de pessoas diagnosticadas positivamente para COVID-19 e 1.229 imagens de pessoas com sintomas semelhantes mas que não foi detectada a presença da doença.

Observou-se que as imagens não possuem padronização no tamanho ou modo de cor, sendo necessário realizar um pré-processamento das imagens.

### **2.2. ESTRUTURAÇÃO DA SOLUÇÃO E PADRONIZAÇÃO DO DATASET**

Ao se trabalhar com modelos diferentes utilizando o mesmo conjunto de dados, para tirar proveito de definições de funções em códigos Python, optou-se por criar funções especializadas para carregamento de imagens, aumento sintético dos dados, pré-processar as imagens etc.

Escolheu-se, após testes iniciais, padronizar as imagens em tamanho e cores, reduzindo as cores para escala de cinza e redimensionando sua escala para 128 por 128 pixels no treinamento com Florestas Randômicas e, com modo de cor RGB no treinamento com Redes Neurais Convolucionais.

Como boa prática, dividiu-se o dataset em dois conjuntos de dados, sendo 80% deles utilizado para treinamento e os outros 20% para validação. Na divisão dos dados optou-se por manter um padrão no hiperparâmetro "*random\_state*", assim é possível comparar resultados com diferentes hiperparâmetros sem perder a consistência do teste.

### 2.3. ESCOLHA DOS MODELOS

Os autores, através de estudos acadêmicos em matérias como Redes Neurais Artificiais, Machine Learning e Sistemas de Apoio a Decisão, tiveram contato com alguns modelos que chamaram atenção, como Redes Neurais Convolucionais, Florestas Randômicas, Regressão Logística, SVM, Árvores de Decisão e outros. Através destes estudos, criou-se certa familiaridade com sintaxes e formas de exploração o que, por esse motivo, para treinamento e classificação de imagens, o primeiro modelo escolhido foi o de Florestas Randômicas.

Sabe-se que Árvores de Decisões tendem a apresentar um ajustamento aos dados de treinamento (*Overfitting*), levando a classificações equivocadas quando se inserem dados diferentes dos utilizados para treinamento.

"O fenômeno de ajuste mais próximo aos nossos dados de treinamento do que à distribuição subjacente é chamado de *overfitting*, e as técnicas para combater o *overfitting* são frequentemente chamadas de métodos de regularização." (ZHANG et. al., 2021. Tradução nossa.)

As Florestas Randômicas (*Random Forests*) auxiliam a reduzir o *overfitting* ao utilizar várias árvores de decisão para extrair delas uma média, que resulta em um modelo mais robusto. "Isso significa que um conjunto diversificado de classificadores é criado pela introdução de aleatoriedade na construção do classificador. A previsão do conjunto é dada como a previsão média dos classificadores individuais. (SCIKIT-LEARN, 2023. Tradução nossa)

Como segundo modelo os autores escolheram Redes Neurais Convolucionais, o que, após pesquisas em fontes da internet e no *Kaggle*, descobriu-se que funcionam de maneira excepcional para classificação de imagens. Para trabalhar com Redes Neurais Convolucionais escolheu-se a biblioteca Tensorflow com Keras.

## 2.4. TREINAMENTO DO MODELO COM FLORESTAS RANDÔMICAS

Utilizou-se o modelo de Florestas Randômicas da biblioteca Scikit-Learn, que possui vasta documentação e é muito utilizado por especialistas da área de Inteligência Artificial.

Na configuração inicial, utilizaram-se os parâmetros "*n\_estimator*", que se refere ao número de árvores criadas, igual a 100, e "*random\_state*", que gera uma semente de aleatoriedade, com o valor 42. Este valor foi escolhido sem nenhum estudo prévio, apenas para padronizar que a mesma semente de aleatoriedade seja utilizada em todos os resultados, para fins de comparação de treinamentos.

As métricas iniciais utilizadas são: acurácia, precisão, recall e *F1 score*.

### 2.4.1. PRIMEIRO TREINAMENTO

O entendimento de como os dados se comportam com o modelo pode ser obtido treinando o modelo com os dados da forma como são fornecidos e, como os dados possuem certo balanceamento, ou seja, a quantidade de imagens positivas para COVID-19 é semelhante à quantidade de imagens de pessoas sem a doença, o treinamento apresentou bons resultados iniciais.

A acurácia de treinamento ficou em 0.879 (suprimindo as demais casas decimais) e, abaixo, o relatório de classificação e acurácia:

TABELA 1 - Relatório de Classificação e Acurácia do Primeiro Treinamento

	<b>precision</b>	<b>recall</b>	<b>F1 score</b>	<b>support</b>
COVID	0.92	0.82	0.87	239
non-COVID	0.85	0.94	0.89	258
accuracy	-	-	0.88	497
macro avg	0.89	0.88	0.88	497
weighted avg	0.88	0.88	0.88	497

Fonte: dos autores.



Observa-se que, para um modelo em que os ajustes dos hiperparâmetros não foram realizados, a pontuação foi consideravelmente alta, com *F1 score* próxima de 0.87. Quanto mais próxima de 1 esta métrica estiver, melhores serão os resultados alcançados pelo modelo. Recorda-se que, no início deste trabalho, foi informado que Soares et. al. (2019) reportaram o resultado de 97.31% de *F1 score*.

*F1 score* é uma métrica de avaliação de desempenho do modelo com o melhor resultado em 1 e o pior em 0. É dada pela formula:  $F1 = 2 * (precision * recall) / (precision + recall)$ . (SCIKIT-LEARN, 2023)

#### 2.4.2. SEGUNDO TREINAMENTO - SELEÇÃO AUTOMATIZADA DE HIPERPARÂMETROS

O ajuste correto dos hiperparâmetros permite um melhor rendimento do modelo, mas, muitas vezes, este ajuste é manual e consome muito tempo. Dado o limite de prazo dos autores para a finalização deste trabalho, a etapa de experimentação do modelo foi reduzida.

Para auxiliar na obtenção de valores de hiperparâmetros, que funcionem em conjunto com os dados, utilizou-se a biblioteca *GridSearchCV*, que é capaz de encontrar alguns hiperparâmetros que melhoram o resultado do modelo. É definida como uma "pesquisa exaustiva sobre valores de parâmetros especificados para um estimador." (SCIKIT-LEARN, 2023)

Dos parâmetros inseridos para testes no *GridSearchCV*, *n\_estimators* (50, 100, 150, 200 e 250), *max\_depth* (*None*, 10, 20, 30) e *min\_samples\_split* (2, 5 e 10), utilizando-se de três dobras, *job*<sup>3</sup> 1 e *verbose*<sup>4</sup> 2, o treinamento, que durou aproximadamente 32 minutos, encontrou os seguintes parâmetros:

---

<sup>3</sup> Neste contexto, o termo "job" refere-se ao número de tarefas de CPU que são executadas em paralelo durante o treinamento. Um valor de 1 significa que uma única CPU está sendo usada, enquanto um valor maior pode ser usado para acelerar o treinamento, aproveitando vários núcleos de CPU. (SCIKIT-LEARN, 2023)

<sup>4</sup> É um parâmetro que controla a saída de mensagens de log durante o treinamento. Um valor de 0 significa que nenhuma mensagem será exibida, enquanto um valor maior (como 2) indica que mais detalhes serão exibidos à medida que o treinamento avança. (SCIKIT-LEARN, 2023)

- *n\_estimators*: 200;
- *max\_depth*: None;
- *min\_samples\_split*: 10.

Observou-se que a acurácia, após o treinamento com os hiperparâmetros selecionados pelo algoritmo *GridSearchCV*, teve uma leve melhora, passando de 0.879 para 0.887. Abaixo se visualiza a tabela do relatório de classificação e acurácia:

TABELA 2 - Relatório de Classificação e Acurácia do segundo treinamento com Florestas Randômicas.

	<b>precision</b>	<b>recall</b>	<b>F1 score</b>	<b>support</b>
COVID	0.95	0.81	0.87	239
non-COVID	0.85	0.96	0.90	258
acuraccy	-	-	0.89	497
macro avg	0.90	0.88	0.89	497
weighted avg	0.88	0.89	0.89	497

Fonte: dos autores.

#### 2.4.3. DATA AUGMENTATION

O aumento artificial dos dados permite, em muitos casos, melhorar a qualidade do treinamento. Esse aumento, quando se trata de imagens, é feito alterando-se as características da imagem, como manipulação de *zoom*, rotação, cores, espelhamentos etc. O aumento da imagem gera exemplos de treinamento semelhantes, mas distintos, após uma série de alterações aleatórias nas imagens de treinamento, expandindo assim o tamanho do conjunto de treinamento. (ZHANG et. al., 2021. Tradução nossa)

Os parâmetros utilizados foram os mesmos encontrados pelo algoritmo *GridSearchCV* anterior. As transformações aplicadas às imagens foram *zoom* e rotação das imagens, porém, apresentou uma leve queda no desempenho se comparado aos resultados anteriores. A acurácia passou de 0.887 para 0.869 e abaixo está o relatório de classificação e acurácia deste treinamento.

TABELA 3 - Relatório de classificação e acurácia após aumento de dados sintéticos com aplicação de zoom e rotação nas imagens.

	<b>precision</b>	<b>recall</b>	<b>F1 score</b>	<b>support</b>
COVID	0.91	0.81	0.86	239
non-COVID	0.84	0.92	0.88	258
acuraccy	-	-	0.87	497
macro avg	0.87	0.87	0.87	497
weighted avg	0.87	0.87	0.87	497

Fonte: dos autores.

Dada a queda no resultado, descartou-se o aumento artificial dos dados com aplicação de zoom e rotação e aplicou-se as alterações de espelhamento de imagem horizontal/vertical e ajuste de brilho/contraste.

A melhora, após o novo treinamento, foi significativa em todas as métricas citadas até o momento. A acurácia geral passou para 0.891 e *F1 score* para 0.88. Abaixo o relatório de classificação e acurácia.

TABELA 4 - Relatório de classificação e acurácia após aplicação aumento sintético de dados com espelhamento e ajuste de brilho/contraste das imagens.

	<b>precision</b>	<b>recall</b>	<b>F1 score</b>	<b>support</b>
COVID	0.95	0.82	0.88	239
non-COVID	0.85	0.96	0.90	258
acuraccy	-	-	0.89	497
macro avg	0.90	0.89	0.89	497
weighted avg	0.90	0.89	0.89	497

Fonte: dos autores.

Como o conjunto de dados foi atualizado, entendeu-se que uma nova avaliação dos hiperparâmetros pelo *GridSeachCV* seria válida, porém, com o triplo de imagens na avaliação, o tempo de treinamento chegou a 67 minutos. Como esperado, os valores dos hiperparâmetros foram diferentes. O resultado compensou a espera com aumento da acurácia para 0.899 do *F1 score* para 0.89, os melhores resultados até o momento. O relatório de classificação e acurácia segue abaixo.

TABELA 5 - Relatório de classificação e acurácia com aumento sintético de dados com espelhamento e ajuste de brilho/contraste, com os novos parâmetros encontrados pelo GridSearchCV.

	<b>precision</b>	<b>recall</b>	<b>F1 score</b>	<b>support</b>
COVID	0.95	0.83	0.89	239
non-COVID	0.86	0.96	0.91	258
acuraccy	-	-	0.90	497
macro avg	0.91	0.90	0.90	497
weighted avg	0.90	0.90	0.90	497

Fonte: dos autores.

#### 2.4.4. COMPARAÇÃO DOS RESULTADOS COM OUTRAS MÉTRICAS

Para analisar melhor como os modelos se comportaram até agora, extraiu-se as métricas de:

- Matriz de confusão;
- Acurácia média na validação cruzada.

Os resultados foram compilados em uma tabela para melhor compreensão, sendo representados os cinco treinamentos realizados até o momento.

TABELA 6 - Comparativo entre os treinamentos utilizando acurácia média na validação cruzada e matriz de confusão.

	Treino 1	Treino 2	Treino 3	Treino 4	Treino 5
Acurácia	0.879	0.887	0.861	0.891	0.899
Acur. média na val. cruz.	0.825	0.816	0.816	0.816	0.829
Desvio padrão	0.061	0.067	0.067	0.067	0.061
True Positives	242	247	237	247	248
False Positives	44	45	48	43	40
True Negatives	195	194	191	196	199
False Negatives	16	11	21	11	10

Fonte: dos autores.

Observa-se, claramente, que o modelo do treino de número 5 apresenta melhora com relação aos modelos anteriores e os autores acreditam que, com tempo para mais experimentações, é possível melhorar os resultados com Florestas Randômicas.

## 2.5. TREINAMENTO UTILIZANDO REDES NEURAIIS COM TENSORFLOW E KERAS

### 2.5.1. DEFINIÇÃO DO MODELO E PARÂMETROS

Neste modelo de treinamento, utilizam-se redes neurais convolucionais para a análise do desempenho do modelo e ajuste dos pesos em cada época do treinamento.

"Uma **rede neural convolucional (RNC)** é aquela que contém conexões espacialmente locais, pelo menos nas camadas iniciais, e conta com padrões de pesos que são replicados nas unidades em cada camada." (RUSSEL; NORVIG, 2022).

A esses padrões de pesos, que são replicados em várias regiões locais, dá-se o nome de **kernel**, e "o processo de aplicação do kernel aos pixels da imagem (ou a unidades organizadas espacialmente em uma camada subsequente) é chamado **convolução**." (RUSSEL; NORVIG, 2022)

Redes neurais convolucionais tendem a apresentar excelentes resultados na classificação de imagens, entretanto, o número de hiperparâmetros é maior do que os utilizados nas florestas randômicas, motivo pelo qual os autores definiram algumas configurações iniciais para exploração dos modelos:

- Divisão do conjunto de dados em 80% para treinamento e 20% para teste com uma *seed*<sup>5</sup> de *random\_state* igual a 42;
- Modelo com uma camada de entrada, quatro camadas ocultas com função de ativação "*relu*", quatro camadas ocultas de *Max Pooling*, uma camada de achatamento, uma camada densamente conectada e uma camada de saída com função de ativação do tipo "*sigmoid*".
- Compilador do tipo "*adam*".

---

<sup>5</sup> Seed é um valor utilizado na equação de geração de números pseudoaleatórios que precisa de um valor inicial para poder gerar novos números. (COSTA, 2020)

A escolha do tipo sigmoid para a camada de saída se deu porque, nessa atividade de classificação, têm-se apenas dois resultados possíveis, 'COVID' e 'non-COVID'.

Outro modelo foi criado, mantendo as configurações, adicionando uma camada de "*Dropout*" para desligar, aleatoriamente, alguns neurônios, na tentativa de melhorar a generalização do modelo, evitando o fenômeno de *overfitting*.

As métricas utilizadas para o treinamento com redes neurais foram:

- Acurácia de treinamento e validação;
- Perda de treinamento e validação;
- Matriz de confusão;
- F1 score; e
- Curva ROC (AUC)

## 2.5.2. TREINAMENTOS

Vários treinamentos com configurações diferentes foram realizados e, para não estender muito este documento, foram agrupados em blocos. Cada bloco procura agrupar os treinamentos pelas características ou informações que se pretende observar de cada um deles. Salienta-se que este estudo não explora todas as possibilidades de hiperparâmetros, concentrando-se apenas em extrair informações dos treinamentos para encontrar o melhor modelo possível dentro das limitações apresentadas.

### 2.5.2.1. TREINAMENTO COM OS DADOS PUROS

Antes de aplicar técnicas como aumento sintético de imagens e ajustes nos parâmetros dos modelos, explorou-se o treinamento com os dois modelos criados (com e sem *drop* de neurônios) e ajustes das épocas.

Os resultados e parâmetros do primeiro bloco de treinamentos podem ser observados na tabela abaixo:

TABELA 7: Resultados do primeiro bloco de treinamentos

Training	Treino 01	Treino 02	Treino 03
Model	Sem drop	Sem drop	Com drop
Epochs	10	20	20
Training Accuracy	0.97	0.99	0.96
Validation Accuracy	0.92	0.93	0.88
Training Loss	0.08	0.03	0.09
Validation Loss	0.22	0.32	0.33
True Positives	227	234	190
False Positives	27	31	13
True Negatives	231	227	245
False Negatives	12	4	49
General Accuracy	0.92	0.92	0.87
F1 score	0.92	0.93	0.85
ROC AUC	0.97	0.98	0.96

Fonte: dos autores

#### Considerações das observações realizadas:

- No primeiro treinamento, no decorrer das épocas, perde-se acurácia na validação, o que pode indicar o início de overfitting. A perda na validação e na acurácia estabilizaram por volta da 6ª época, mas ficaram bem distantes no final.
- Para o segundo treinamento, aumentou-se a quantidade de épocas para confirmar se o modelo realmente estava tendendo a overfitting. Observação que foi confirmada ao verificar que a perda do treinamento abriu muita distância da perda de validação, entretanto, o *F1 score* obteve o resultado de 0.93, o que não é ruim.
- No terceiro treinamento, manteve-se 20 épocas e realizou-se o drop aleatório de alguns neurônios utilizando o segundo modelo configurado. Observou-se que a acurácia de treinamento alcançou o platô por volta da 13ª época e a partição de validação não obteve melhoras como pode se observar na perda de treinamento e validação.

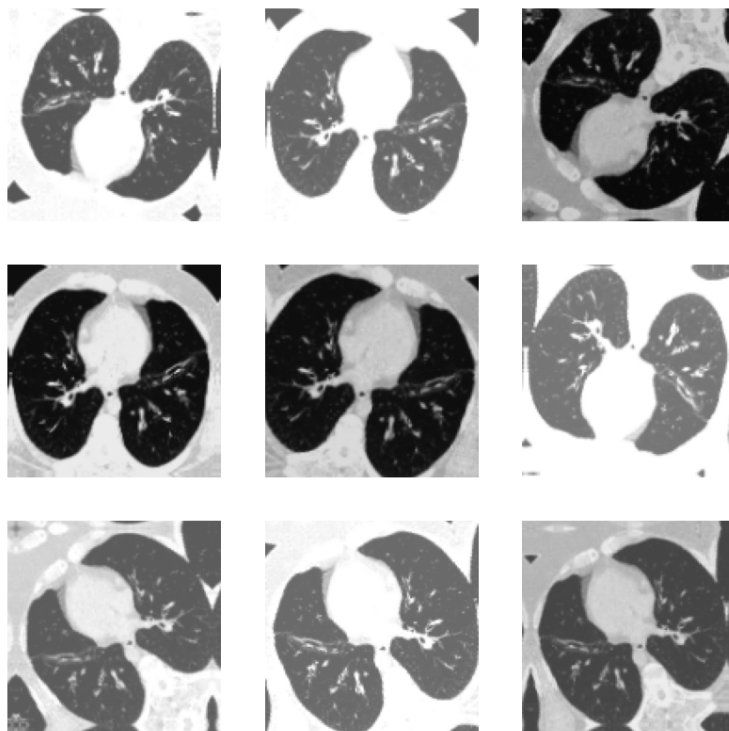
### 2.5.2.2. DATA AUGMENTATION

Para testes, considerou-se utilizar aumento sintético das imagens, que podem ser realizados de diversas formas. Os aumentos de imagens aplicados pelos autores consistem nas técnicas:

- Alteração aleatória da rotação;
- Alterações no filtro de canais RGB (HUE);
- Alterações no brilho e contraste das imagens;
- Alterações no espelhamento horizontal e vertical;
- Alterações na ampliação da imagem.

A imagem abaixo exemplifica uma das imagens com a aplicação de todas as técnicas combinadas:

Figura 1: Imagem com aplicação das técnicas de aumento sintético das imagens



Fonte: Imagem original de SOARES et. al. (2020), alterada pelos autores



As alterações foram testadas nos modelos com e sem drop de neurônios, isoladas e combinadas. Os resultados deste treinamento podem ser observados abaixo:

TABELA 8: Resultados dos treinos com zoom e flip

Training	Treino 4	Treino 5	Treino 6	Treino 7
Model	Sem drop	Com drop	Sem drop	Com drop
Epochs	20	20	20	20
Data Augmentation	Zoom	Zoom	Flip	Flip
Training Accuracy	0.98	0.96	0.93	0.93
Validation Accuracy	0.96	0.95	0.86	0.90
Training Loss	0.07	0.10	0.15	0.18
Validation Loss	0.11	0.12	0.31	0.22
True Positives	229	222	183	216
False Positives	12	8	12	26
True Negatives	246	250	246	232
False Negatives	10	7	56	23
General Accuracy	0.95	0.95	0.86	0.90
F1 score	0.95	0.94	0.84	0.89
ROC AUC	0.99	0.99	0.96	0.97

Fonte: dos autores

Considerações sobre este bloco:

- O treinamento de número 4 apresentou uma qualidade interessante, tendo o *F1 score* de 0.95 e a queda na perda da validação acompanhou a queda na perda do treinamento até o final, sendo um excelente resultado.
- Adicionando o drop de neurônios, o treinamento 5 manteve o a acurácia de treinamento e validação, assim como a perda de treinamento e validação muito próximas, sendo considerado também um bom resultado.
- A aplicação de espelhamento no treino 6 obteve um rendimento baixo com o *F1 score* de 0.84 e isso se manteve para o treino, com um *F1 score* de 0.89.

O próximo bloco de treinamento considera os ajustes das cores, brilho e contraste.

TABELA 9: Resultado dos testes com alterações no hue, brilho e contraste

Training	Treino 8	Treino 9	Treino 10	Treino 11	Treino 12	Treino 13
Model	Sem drop	Com drop	Sem drop	Com drop	Sem drop	Com drop
Epochs	20	20	20	20	20	20
Data Augmentation	Hue, brilho, contraste	Hue, brilho, contraste	Brilho e contraste	Brilho e contraste	Hue	Hue
Training Accuracy	0.98	0.98	0.99	0.99	1.00	0.97
Validation Accuracy	0.86	0.86	0.93	0.92	0.93	0.93
Training Loss	0.06	0.05	0.03	0.03	0.00	0.08
Validation Loss	0.44	0.45	0.25	0.34	0.28	0.21
True Positives	174	176	227	229	225	216
False Positives	7	6	24	31	21	11
True Negatives	251	252	234	227	237	247
False Negatives	65	63	12	10	14	23
General Accuracy	0.85	0.86	0.92	0.91	0.93	0.93
F1 score	0.82	0.83	0.92	0.91	0.92	0.92
ROC AUC	0.96	0.97	0.97	0.97	0.98	0.98

Fonte: dos autores

Considerações sobre o bloco de treino:

- O treinamento com ajuste das cores, brilho e contraste não apresentou um bom rendimento, motivo pelo qual talvez seja necessário pensar que alguma das cores pode estar atrapalhando. Na acurácia de treino, por volta da 13<sup>a</sup> época, manteve o platô de 0.98 e o mesmo aconteceu com a perda de treino, com 0.06.
- Tentou-se reverter o overtiffing com o drop de neurônios no treinamento 9 e teve uma leve melhora, mas insignificante.

- Considerando apenas o brilho e contraste, o modelo teve uma leve melhora, com *F1 score* chegando a 0.92, mas ainda com muita perda na validação, com ou sem o drop.
- Treinando com base apenas no ajuste de cores (hue), o modelo também apresentou 0.92 de *F1 score*, ou seja, brilho/contraste e hue separados obtém maiores rendimentos do que combinados.

Após a verificação dos elementos de image augmentation isoladamente, decidiu-se combinar todos os elementos e verificar como se comportavam. A tabela abaixo contém os resultados dos treinos 14 a 18.

TABELA 10: Resultados dos treinos com as técnicas de image augmentation combinadas

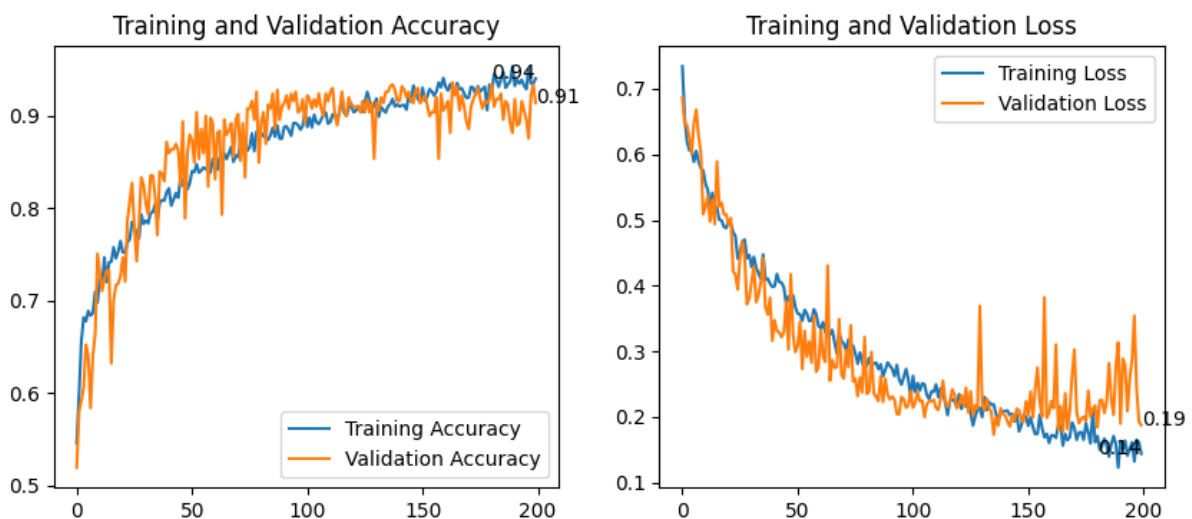
Training	Treino 14	Treino 15	Treino 16	Treino 17	Treino 18
Model	Sem drop	Com drop	Sem drop	Com drop	Sem drop
Epochs	20	20	100	200	200
Data Augmentation	Todos	Todos	Todos	Todos	Todos
Training Accuracy	0.76	0.77	0.90	0.94	0.94
Validation Accuracy	0.77	0.80	0.91	0.92	0.91
Training Loss	0.50	0.49	0.23	0.16	0.14
Validation Loss	0.47	0.45	0.21	0.26	0.19
True Positives	134	173	222	233	220
False Positives	8	33	30	35	24
True Negatives	250	225	228	223	234
False Negatives	105	66	17	6	19
General Accuracy	0.77	0.80	0.90	0.91	0.91
F1 score	0.70	0.77	0.90	0.92	0.91
ROC AUC	0.90	0.88	0.97	0.98	0.98

Fonte: dos autores

### Considerações sobre o bloco de treino:

- No treino de número 14, a acurácia de treinamento e validação, assim como a perda de treino e validação, seguiram muito próximas, o que pode indicar que o modelo está generalizando corretamente os dados, por isso entendeu-se que, talvez, com mais épocas, não apresentasse sinais de *overfitting*. O mesmo resultado foi apresentado com o *drop* de neurônios.
- Com 100 épocas o modelo mostrou uma evolução consistente e conseguiu generalizar bem, mas não chegou a atingir a melhor marca de treinamento isolado (só com *zoom*). Aumentou-se então as épocas para 200, na tentativa de reduzir a perda no erro e validação.
- Com o treinamento com 200 épocas, com e sem drop de neurônios nas camadas ocultas, chega-se a um *F1 score* de 0.92 (em outros testes chegou a 0.95), sendo um valor interessante. Observa-se, entretanto, que a partir da época 150 iniciou-se uma leve diferenciação entre a camada de treinamento e validação, o que indica que, talvez, com mais épocas, o modelo não melhorasse. Para ilustrar segue o gráfico da acurácia de validação e treinamento e perda de validação e treinamento.

Figura 2: Gráfico de acurácia de treino e validação e perda de treino e validação



Fonte: dos autores

Considerando que os autores não propuseram uma exploração exaustiva e tiveram tempo limitado para a conclusão deste trabalho, explorou-se mais algumas alternativas no treinamento, como ajustes no tamanho da imagem para mais (256 x 256 pixels), para menos (64 x 64 pixels) e os resultados do *F1 score*, considerando no máximo 30 épocas, permaneceram próximos de 0.94.

Tentou-se então deixar a padronização do tamanho para próximo das menores imagens encontradas, cerca de 200 x 150 pixels, ajustando o modelo para 200 x 200 pixels. Aplicou-se também a técnica de image augmentation zoom, que foi a que obteve melhor resultado até o momento.

Após 30 épocas, obteve-se um excelente resultado com o *F1 score* em 0.96. A tabela abaixo detalha os resultados deste treinamento:

TABELA 11: Resultado do treinamento com zoom

Training	Treino 01
Model	Sem drop
Epochs	30
Data Augmentation	Zoom
Training Accuracy	0.99
Validation Accuracy	0.96
Training Loss	0.02
Validation Loss	0.16
True Positives	231
False Positives	10
True Negatives	248
False Negatives	8
General Accuracy	0.96
F1 score	0.96
ROC AUC	0.99

Fonte: dos autores

### 3.0. CONCLUSÃO

Apesar de não terem alcançado a marca de 0.97 de *F1 score*, como reportado no artigo "**SARS-CoV-2 CT-scan dataset: A large dataset of real patients CT scans for SARS-CoV-2 identification**", os autores deste estudo obtiveram um *F1 score* promissor de 0.96. É importante destacar que esse resultado foi alcançado em um período relativamente curto e com uma exploração limitada das variações do modelo de redes neurais.

Os modelos foram padronizados e as variações nos hiperparâmetros pouco exploradas. Com base nisso, acredita-se que há espaço para melhorias no desempenho do modelo por meio de estratégias como adicionar ou remover camadas ocultas, ajustar os filtros de ativação e otimizar outros hiperparâmetros. Os autores se comprometem a investigar essas possibilidades em futuras versões deste trabalho.

## REFERÊNCIAS

COSTA, Guilherme Augusto Duque e Silva. **SEED como funciona?** Resposta 1. Disponível em: <https://cursos.alura.com.br/forum/topico-seed-como-funciona-125024>. Acesso em 27 jun. 2023.

GOVERNO DO ESTADO DO CEARÁ. **Por que as vacinas contra Covid-19 ficaram prontas tão rápido?** 2021. Disponível em: <https://coronavirus.ceara.gov.br/por-que-as-vacinas-contr-covid-19-ficaram-prontas-tao-rapido/>. Acesso em 14 jun. 2023.

KAGGLE. **Kaggle**. Disponível em: <https://www.kaggle.com/>. Acesso em 14 jun. 2023.

LEAL, Renato dos Santos. **Métricas Comuns em Machine Learning: como analisar a qualidade de chat bots inteligentes — métricas (3 de 4)**. MEDIUM, 2017. Disponível em: <https://medium.com/as-m%C3%A1quinas-que-pensam/m%C3%A9tricas-comuns-em-machine-learning-como-analisar-a-qualidade-de-chat-bots-inteligentes-m%C3%A9tricas-1ba580d7cc96>. Acesso em: 14 jun. 2023.

OPAS. Organização Pan-Americana de Saúde. **OMS divulga novas estatísticas mundiais de saúde**. Disponível em: <https://www.paho.org/pt/noticias/20-5-2022-oms-divulga-novas-estatisticas-mundiais-saude>. Acesso em 31 mai. 2023.

OPAS. Organização Pan-Americana de Saúde. **What we know about COVID-19 and Influenza**. Disponível em: <https://www.paho.org/en/documents/what-we-know-about-covid-19-and-influenza>. Acesso em: 07 jun. 2023.

RUSSEL, Stuart; NORVIG, Peter. **Inteligência Artificial: Uma Abordagem Moderna**. 4ª Edição. Tradução Daniel Vieira; Flávio Soares Corrêa da Silva. Rio de Janeiro. LTC | Livros Técnicos e Científicos Ltda., 2022.

SCIKIT-LEARN. **Ensemble methods**. Disponível em: <https://scikit-learn.org/stable/modules/ensemble.html>. Acesso em 21 jun. 2023.

SCIKIT-LEARN. **Machine Learning in Python**. Disponível em: <https://scikit-learn.org/stable/>. Acesso em 14 jun. 2023.

SCIKIT-LEARN. **sklearn.metrics.f1\_score**. Disponível em: [https://scikit-learn.org/stable/modules/generated/sklearn.metrics.f1\\_score.html#sklearn.metrics.f1\\_score](https://scikit-learn.org/stable/modules/generated/sklearn.metrics.f1_score.html#sklearn.metrics.f1_score). Acesso em 21 jun. 2023.

SCIKIT-LEARN. **sklearn.model\_selection.GridSearchCV**. Disponível em: [https://scikit-learn.org/stable/modules/generated/sklearn.model\\_selection.GridSearchCV.html](https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.GridSearchCV.html). Acesso em 21 jun. 2023.

SOARES, Eduardo; ANGELOV, Plamen; BIASCO, Sarah; FROES, Michele Higa; ABE, Daniel Kanda. **SARS-CoV-2 CT-scan dataset: A large dataset of real patients CT scans for SARS-CoV-2 identification**. MEDrXIV, 2020. DOI: <https://doi.org/10.1101/2020.04.24.20078584>. Disponível em: <https://www.medrxiv.org/content/10.1101/2020.04.24.20078584v3>. Acesso em: 14 jun. 2023.

TENSORFLOW. **Tensorflow**. Disponível em: <https://www.tensorflow.org/?hl=pt-br>. Acesso em 14 jun. 2023.

ZHANG, Aston; LIPTON, Zachary C.; LI, Mu; SMOLA, Alexander J. **Dive into Deep Learning**. 2021. arXiv preprint arXiv:2106.11342. Disponível em: <https://www.d2l.ai/index.html>. Acesso em 21 jun. 2023.