



COVID-19 classification in X-ray chest images using a new convolutional neural network: CNN-COVID

Pedro Moisés de Sousa¹ · Pedro Cunha Carneiro¹ · Mariane Modesto Oliveira¹ · Gabrielle Macedo Pereira¹ · Carlos Alberto da Costa Junior¹ · Luis Vinicius de Moura² · Christian Mattjie² · Ana Maria Marques da Silva² · Ana Claudia Patrocínio¹

Received: 3 July 2020 / Accepted: 22 December 2020 / Published online: 4 January 2021

© Sociedade Brasileira de Engenharia Biomedica 2021

Abstract

Purpose COVID-19 causes lung inflammation and lesions, and chest X-ray and computed tomography images are remarkably suitable for differentiating the new disease from patients with other lung diseases. In this paper, we propose a computer model to classify X-ray images of patients diagnosed with COVID-19. Chest X-ray exams were chosen over computed tomography scans because they are low cost, results are quickly obtained, and X-ray equipment is readily available.

Methods A new CNN network, called CNN-COVID, has been developed to classify X-ray patient's images. Images from two different datasets were used. The images of Dataset I is originated from the COVID-19 image data collection and the ChestXray14 repository, and the images of Dataset II belong to the BIMCV COVID-19+ repository. To assess the accuracy of the network, 10 training and testing sessions were performed in both datasets. A confusion matrix was generated to evaluate the model's performance and calculate the following metrics: accuracy (ACC), sensitivity (SE), and specificity (SP). In addition, Receiver Operating Characteristic (ROC) curves and Areas Under the Curve (AUCs) were also considered.

Results After running 10 tests, the average accuracy for Dataset I and Dataset II was 0.9787 and 0.9839, respectively. Since the weights of the best test results were applied in the validation, it was obtained the accuracy of 0.9722 for Dataset I and 0.9884 for Dataset II.

Conclusions The results showed that the CNN-COVID is a promising tool to help physicians classify chest images with pneumonia, considering pneumonia caused by COVID-19 and pneumonia due to other causes.

Keywords Chest X-ray images · CNN · CNN-COVID · Convolutional neural network · Coronavirus · COVID-19 · Deep learning

✉ Pedro Moisés de Sousa
profpedromoises@gmail.com

Pedro Cunha Carneiro
pedrocameiro@ufu.br

Mariane Modesto Oliveira
modestomariane@gmail.com

Gabrielle Macedo Pereira
gabriellem@ufu.br

Carlos Alberto da Costa Junior
carlosjunior@ufu.br

Luis Vinicius de Moura
luis.moura@edu.pucrs.br

Christian Mattjie
christian.mattjie@gmail.com

Ana Maria Marques da Silva
ana.marques@pucrs.br

Ana Claudia Patrocínio
ana.patrocinio@ufu.br

¹ Biomedical Lab, Faculty of Electrical Engineering, Federal University of Uberlândia, Campus Sta Mônica, Av. João Naves de Ávila, 2121, Bloco 1E, CEP, Uberlândia, MG 38400-000, Brazil

² Medical Image Computing Laboratory, Pontifical Catholic University of Rio Grande do Sul, Av. Ipiranga, 6681 Partenon, CEP, Porto Alegre, RS 90619-900, Brazil

Introduction

In December 2019, a group of patients with atypical pneumonia, cause unknown, was associated to the consumption of bat meat bought at an exotic animal meat market in Wuhan, Hubei, China. The disease has quickly spread to other corners of the world and, on March 11, 2020, the World Health Organization declared the COVID-19 global pandemic outbreak, which is still ongoing (Zhu et al. 2020; Yang et al. 2020). By using impartial sequencing of samples from patients, it was possible to identify a new type of beta coronavirus. This novel coronavirus, named 2019-nCoV, was compared to Severe Acute Respiratory Syndrome (SARS) and Middle East Respiratory Syndrome (MERS), presented a lower mortality rate and a higher transmission capacity (Zhu et al. 2020; Yang et al. 2020).

Coronaviruses are RNA viruses with a lipid envelope able to cause respiratory, enteric, liver, and neurological diseases in domestic animals and human beings (Zhu et al. 2020; Weiss and Leibowitz 2011). The existence of wild viruses in nature is well known by researchers. They are highly prevalent, broadly distributed, genetically diverse, and frequently genetically recombined. There is a higher probability of natural hosts disseminating these viruses to human beings as the man-animal interface activity is increasing.

Both SARS and MERS were associated with a zoonotic origin and transmitted by civets and camels, respectively. The appearance of diseases such as SARS and MERS, supposedly caused by the new coronavirus, tends to become more frequent events when no barriers are imposed between human society and wild nature (Zhu et al. 2020; Cui et al. 2019).

As there is no specific drug or vaccine to treat the new coronavirus, early detection is crucial so that the patient is isolated from the healthy population as soon as possible (Ai et al. 2020). Thus, the research for methods of early detection has become essential to fight this pandemic outbreak.

Currently, the gold standard of the COVID-19 diagnosis is made by viral nucleic acid detection using reverse transcription-polymerase chain reaction (RT-PCR) in real-time, although its effective accuracy is 30 to 50% (Ai et al. 2020; Zhang et al. 2020; Ozturk et al. 2020).

An underlying problem with this method is how unavailable it is in several regions and countries impacted; this generates logistic and political issues for providing enough test kits for the increasing number of patients suspected of having the disease (Zhang et al. 2020). Moreover, the delay in processing and getting results and the significant number of false negatives urged researchers from all over the world to try to find a solution for this problem in various areas of knowledge (Ozturk et al. 2020).

Medical image processing is one of the areas which has been contributing to promising studies. Research in this area is being done to clinically aid to diagnose the disease for patients who develop lung atypical pneumonia, using chest X-ray images or

computed tomography (CT) scans (Ozturk et al. 2020). As several patients with COVID-19 develop lung infection, CT scans are useful to detect lung impairment, as well as to classify its progression (Zhang et al. 2020; Dai et al. 2020).

Radiological images from COVID-19 patients may present similarities with those from patients with bacterial or viral pneumonia, specifically the ones caused by SARS and MERS. Thus, the ability to accurately differentiate diseases by analyzing medical images has become a vital challenge, and overcoming it means helping healthcare professionals with detecting the disease early and isolating affected patients as soon as possible (Ozturk et al. 2020; Chung et al. 2020).

Medical image processing area has several studies aimed at developing machine learning methods able to help to diagnose COVID-19, either using CT scan images or chest X-ray ones (Ozturk et al. 2020). Problems using CT scans rather than chest X-rays include low availability of equipment, radiologists, and physicians; higher cost; and longer time to obtain images (Zhang et al. 2020).

A machine learning technique used in research is called deep learning. It allows computer models with several layers of processing to learn how to represent data in various levels of abstraction (Zhang et al. 2020; Lecun et al. 2015; Martin et al. 2020). This technique allows the design of applications which can perform recognition, such as speech recognition, visual recognition, and object detection.

Medical images of patients with COVID-19 presents common features that might show a pattern. Deep learning is an effective technique used by researchers to help healthcare professionals to analyze vast volumes of data generated by chest X-ray images, for instance (Zhang et al. 2020; Martin et al. 2020).

When it comes to applying artificial intelligence and machine learning to the medical field, convolutional neural networks (CNNs) stand out. CNNs are deep artificial neural networks which can be used to classify images, group them by similarity, and run object recognition. These networks are inspired in the human visual cortex processing and used for medical images where irregularities in tissue morphology may be used to classify tumors. CNNs can detect patterns, which are hard to be found by human specialists, for instance, initial stages of disease in tissue samples (Balas et al. 2019).

This paper proposes a new Convolutional Neural Network model called CNN-COVID, to classify images of COVID-19 patients and differentiate them from those who do not have COVID-19, with a focus in analysis, classification, and high accuracy. The CNN-COVID model and its related work are presented in the following sections.

Related work

In the study from Zhang et al. (2020), a model to detect anomalies was developed by using deep learning. The goal is to do a

quick and trustworthy triage of patients with COVID-19. One hundred chest X-ray images from 70 patients with COVID-19 were sourced from a GitHub repository to assess the performance of the model, and 1431 chest X-ray images from 1008 patients with other types of pneumonia were sourced from ChestX-ray14, a public data pool. The Zhang model in (Zhang et al. 2020) is composed of three components, namely, a backbone network, a classification head, and an anomaly detection head. Thus, the first component extracts the high-level features from a chest X-ray which will be the input data for the classification head, and then, the input data for the anomaly detection head. So, with this paper, it resulted 90.00% sensitivity, 87.84% specificity (when parameter T from the study was 0.25), or 96.00% sensitivity and 70.65% specificity (when parameter T was 0.15). Nonetheless, this model presented some limitations, such as missing 4% of COVID-19 cases and having an approximately 30% false positives.

In a similar study, Sethy and Behera (2020) suggested a model to detect COVID-19 from chest X-rays using deep learning. Support vector machine (SVM) classified images of patients suffering from this disease differentiating from images of patients who suffer from other diseases. A subset of 25 COVID-19 images were sourced from the GitHub repository, and a subset of 25 pneumonia images was sourced from the Kaggle repository. The ResNet50 deep neural network model with SVM classification has proven to be the best approach to detect COVID-19, with 95.38% accuracy, 97.2% sensitivity, and 93.4% specificity.

In Wang and Wong (2020), a CNN was adapted to detect COVID-19 cases. It was called COVID-Net, and it used chest X-rays images from open source code and available to the general audience. Thus, COVIDX was created, a database sourcing samples from five different databases (Wang and Wong 2020): 100 samples of chest X-rays of healthy patients, 100 samples of pneumonia patients, and 100 images of COVID-19 patients. COVID-NET presented the following results: 93.3% accuracy, 91.0% sensitivity, and 99.9% specificity when detecting COVID-19 X-ray images out of healthy cases and severe acute respiratory syndrome cases.

The study from Abbas et al. (2020) uses the transfer learning technique, an effective mechanism to provide a promising solution by transferring knowledge from generic tasks for object recognition to domain-specific tasks. Therefore, DeTraC, a deep CNN was adapted with functions to decompose, transfer, and compose samples to classify chest X-ray images as COVID-19. They sourced 80 chest X-ray images (4020×4892 pixels) of healthy patients from the Japanese Society of Radiological Technology (JSRT), 105 images (4248×3480 pixels) of COVID-19 patients, and 11 images (4248×3480 pixels) of SARS patients from GitHub. DeTraC presented the following results: 95.12% accuracy, 97.91% sensitivity, and 91.87% specificity when detecting COVID-19 X-ray

images out of healthy cases and severe acute respiratory syndrome cases.

Methods

CNN-COVID was trained and tested using images from two different databases (public domain) containing chest X-ray who tested positive for COVID-19. The 434 images from Dataset I was obtained at the beginning of the pandemic, and the 4030 images from Dataset II, in October 2020. Due to the limited availability of public images at the beginning of the pandemic, Dataset I was expanded using data augmentation techniques. For Dataset II, a data augmentation was not necessary, and the images were randomly selected for each phase of training and testing. Figure 1 represents the proposed methodology for image classification.

Dataset I

Two databases compose Dataset I. The first one has 217 chest X-ray images from the COVID-19 image data collection (Cohen et al. 2020). In this database, the images are from 141 patients who tested positive for COVID-19 was labeled as COVID. From the second database, ChestXray14 (Wang et al. 2019), 1126 images were used. All these images correspond to chest X-ray images labeled for the presence of 14 common chest radiographic observations that, in this paper, were labeled as NON-COVID.

A subset of 166 images, out of the 217 images from the COVID-19 set were selected randomly (Cohen et al. 2020). Out of these 166 selected images, around 75% were for the training phase and 25% were for the testing phase. From the NON-COVID set, 1000 images were selected randomly, 80% were for the training phase and 20% were for the testing phase.

For the validation phase, 126 images were selected from the COVID-19 set (out of the 217 total images) and 126 images from the NON-COVID set (out of the 1126 total images), as shown in Table 1.

Data augmentation

As the state of pandemic was declared on March 2020, one of the challenges of this study was to find suitable COVID-19 images to work with. So, we had a limited number of X-rays images from COVID-19 patients for training the deep learning models. To overcome that issue, we used the *ImageDataGenerator*, that provided us with new images to be generated for the training phase. The new images were generated from digital processing, using geometric transformations of the original images.

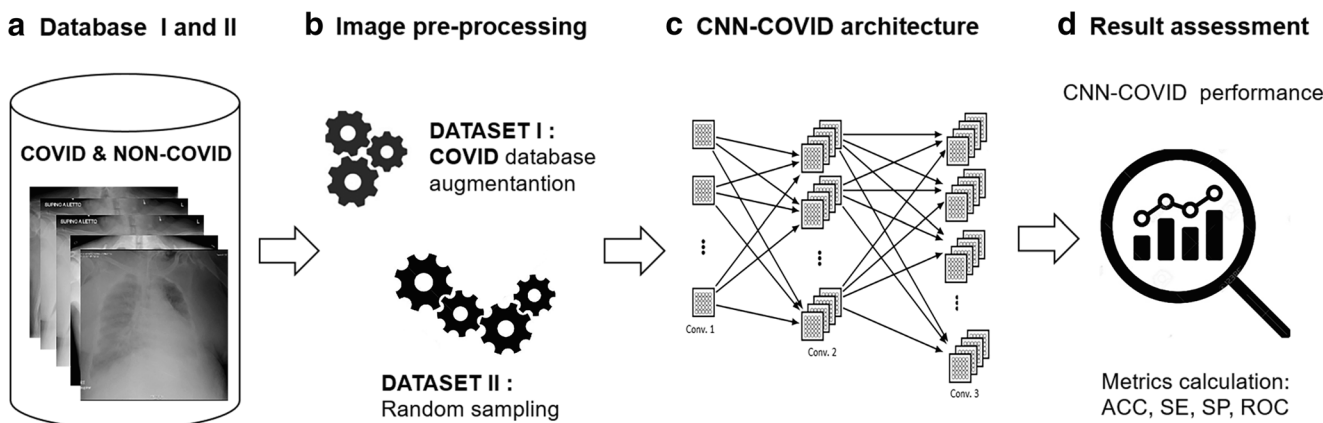


Fig. 1 Diagram of the proposed methodology for image classification (COVID and NON-COVID). Created by the author

These geometric transformations, such as translation, rotation, patch extraction, and reflection, do not make changes to the image object properties, making “data augmentation” possible. The positive side of this technique is that it increases the ability to generalize CNN-COVID when trained with an augmented dataset (Aggarwal et al. 2018; Chollet 2016). Thus, overfitting, which is when the network is no longer able to generalize when presented with new data, can be reduced.

The following common methods were used for dataset augmentation:

- range rotation
- width shift range
- height shift range
- zoom range
- horizontal flip
- vertical flip

After these changes, it was possible to balance the Dataset I from both COVID and NON-COVID classification on the testing and training sets. This database augmentation happens during run time, when chest X-ray images are presented as an input for CNN-COVID.

Dataset II

Dataset II is composed of images from the Valencian Region Medical ImageBank (BIMCV) repository (Vayá et al. 2020) containing chest X-ray and computed tomography (CT) imaging of COVID-19+ patients. From this repository, two

Table 1 Number of images in Dataset I used for CNN-COVID

Dataset I	Database total	Training	Test	Validation
COVID	217	126	40	126
NON-COVID	1126	800	200	126

Created by the author

databases were created: COVID and NON-COVID. The COVID has 2025 chest X-ray images from patients who tested positive for COVID-19, and NON-COVID database, 2025 chest X-ray images from patients who tested negative for the disease.

Among the 2025 images, chosen from the COVID and NON-COVID databases, approximately 75% of them made up the training phase, 15% the testing phase, and 15% the validation phase, as shown in Table 2.

Created by the author

CNN-COVID creation

A CNN is composed of two stages: feature extraction stage and classification stage. In the CNN, the pooling and convolution layers act as a stage of feature extraction. In contrast, the classification stage is made of one or more fully connected layers followed by a sigmoid function layer (Wani et al. 2020), which are presented below. For our proposed CNN-COVID method, we used Python programming language (using Keras library) (Chollet, 2016) to create and train CNN-COVID. The work was developed using a I7-8750H Intel processor, 2.21 GHz CPU, 16.0 GB RAM, and a GeForce GTX 1060 graphic card with Max-Q Design.

Convolution layer

In the proposed CNN-COVID, a new convolution operation is established for the convolutional layer, in which a kernel is

Table 2 Number of images in Dataset II used for CNN-COVID

Dataset II	Database total	Training	Test	Validation
COVID	2025	1419	303	303
NON-COVID	2025	1419	303	303

used to map the activations from one layer into the next. The convolution operation places the kernel in each possible position in the image (or hidden layer) so that the kernel overlaps the entire image and executes a dot product between the kernel parameters and its corresponding receptive field—to which a kernel is applied—in the image. The convolution operation is executed in every region of the image (or hidden layer) to define the next layer (in which activations keep their spatial relations in the previous layer) (Ponti and Da Costa, 2018; Aggarwal et al. 2018; Lecun et al. 2015).

There may be several kernels in the convolutional layer. Every kernel has a feature, such as an edge or a corner. During the forward pass, every kernel is slid to the image width and height (or hidden layer), thus, generating the feature map (Ponti and Da Costa, 2018; Aggarwal et al. 2018; Lecun et al. 2015; Balas et al. 2019

Adaptive moment estimation (ADAM)

CNN-COVID uses Adaptive Moment Estimation (ADAM), an adaptive optimization technique that saves an exponentially decaying average of previously squared gradients v_t . Besides, ADAM also computes the average of the second moments of the gradients m_t (Wani et al. 2020; Kingma and Ba 2014).

Average and non-centered variance values m_t are presented in Eqs. 1 and 2, respectively:

$$m_t = \beta_1 m_{t-1} + (1-\beta_1)gt \tag{1}$$

$$v_t = \beta_2 v_{t-1} + (1-\beta_2)gt^2 \tag{2}$$

ADAM updates exponential moving averages of the gradient and the squared gradient where the hyperparameters $\beta_1, \beta_2 \in [0, 1]$ control the decay rates of these moving averages (Eqs. 3 and 4):

$$\hat{m}_t = \frac{m_t}{1-\beta_1^t} \tag{3}$$

$$\hat{v}_t = \frac{v_t}{1-\beta_2^t} \tag{4}$$

The final formula for the update is presented in Eq. (5):

$$w_{t+1} = w_t - \frac{\alpha \cdot \hat{m}_t}{\sqrt{\hat{v}_t + \epsilon}} \tag{5}$$

Where α is the learning rate and ϵ is a constant added to the denominator for quick conversion methods in order to avoid the division by 0 (Wani et al. 2020; Kingma and Ba 2014).

Dropout technique

CNN-COVID uses the Dropout technique, the most popular technique to reduce overfitting. Dropout refers to dropping out

neurons in a neural network during training. Dropping out a neuron means temporarily disconnecting it, as well as all its internal and external connections, from the network.

Dropped-out neurons neither contribute to the forward pass nor add to the backward pass. By using the dropout technique, the network is forced to learn the most robust features as the network architecture changes with every input (Wani et al. 2020; Balas et al. 2019).

Activation functions

An activation function feeds the output of every convolutional layer. The activation function layer consists of an activation function which uses the resource map produced by the convolutional layer and generates the activation map as the output. The activation function is used to change a neuron activation level in an output signal. Thus, it performs a mathematical operation and generates the neuron activation level at a specific interval, for instance, 0 to 1 or -1 to 1 (Wani et al. 2020). The functions used were the following:

- Sigmoid/Logistic activation function: The sigmoid function $\sigma(x) = \frac{1}{1+e^{-x}}$ is a curve shaped like an S (Ponti and Da Costa, 2018).
- The activation function $f(x) = \max(0, x)$ is called Rectified Linear Unit–ReLU (Ponti and Da Costa, 2018) and generates a non-linear activation map.

Pooling layer

In the CNN-COVID pooling layer, or downsampling layer, is used to reduce the receptive field spatial size, thus, reducing the number of network parameters. The pooling layer selects each convolutional layer feature map and creates a reduced sample. Max-pooling was the technique used for this work. It generates the maximum value in the receptive field. The receptive field is 2×2 ; therefore, max-pooling will issue the maximum of the four input values (Wani et al. 2020).

Fully connected layer

After the convolution and pooling processes, the next step is to decide based on the features detected. This is done by adding one or more fully connected layers to the end. In the fully connected layer, each neuron from the previous layer is connected to each neuron from the following one. All values contribute to predict how strong a value correlates to a given class (Wani et al. 2020). The fully connected layers may be layered on top of one another to learn even more sophisticated feature combinations. The output of the last fully connected layer is fed by an activation function which generates the class

scores. The sigmoid activation function is the one used for CNN-COVID. It produces class scores, and the class with the highest score is treated as the correct one (Wani et al. 2020).

CNN-COVID structure development

Convolutional Neural Networks (CNNs) were proposed to assess image data. The name comes from the convolution operator, a simple way of doing complex operations using the convolution kernel (Ravi et al. 2017).

Many variations of CNN were proposed, such as AlexNet (Krizhevsky et al. 2012), Clarifai (Zeiler and Fergus 2014), and GoogleNet (Szegedy et al. 2015). CNN-COVID structure is also a variation of CNN with the following architecture: an input layer, a convolutional layer, a dense layer, and an output layer, as per Fig. 2.

CNN-COVID detailed architecture is illustrated in Table 3. Our proposed network consists of conventional layers, including input, convolution layer, max-pooling layer, and fully-connected layers.

Besides, a rectified linear unit (ReLU) activation function is used after each convolution layer (1st, 3rd, 5th, and 7th) and dense layers (9th, 10th, 11th, and 12th). To reduce the possibility of overfitting, a dropout rate of 20% was implemented to the first four fully connected layers (9th, 10th, 11th, and 12th).

CNN-COVID training

In the training phase, weights are initialized randomly. The network was trained as per the ADAM model (Wani et al. 2020). Standard parameters $\beta_1 = 0.9$ and $\beta_2 = 0.999$ were used (Kingma and Ba, 2014), as well as the initial learning rate $\alpha = 0.001$ reduced by a factor of 10.

The ADAM training model (Wani et al. 2020) has a better performance compared with other adaptive techniques; it has a quick conversion rate; thus, it reduces the chances of error and increases accuracy. It also overcomes problems faced by other

Table 3 CNN-COVID architecture. The network contains the input (I), the convolution (C), the max-pooling (M) layers, and the fully connected network (F). Created by the author

Layer	CNN-COVID		
Filter	Dimensions	Input/Output	Dimensions
0	I		300 × 300
1	C	5 × 5 × 256	296 × 296 × 256
2	M	2 × 2	148 × 148 × 256
3	C	3 × 3 × 128	146 × 146 × 128
4	M	2 × 2	73 × 73 × 128
5	C	3 × 3 × 64	71 × 71 × 64
6	M	2 × 2	35 × 35 × 64
7	C	3 × 3 × 32	33 × 33 × 32
8	M	2 × 2	16 × 16 × 32
9	F	16 × 16 × 32 × 256	1 × 256
10	F	1 × 1 × 256 × 128	1 × 128
11	F	1 × 1 × 128 × 64	1 × 64
12	F	1 × 1 × 64 × 32	1 × 32
13	F	1 × 132 × 2	1 × 2

optimization techniques, such as decaying learning rate, high variance in updates, and slow convergence (Wani et al. 2020).

CNN-COVID input parameters

Several options were tested when choosing the input parameter values and the CNN-COVID batch size, considering the performance capacity of the available hardware. For input sizes 200 × 200 and 220 × 220, the batch size was 20. For larger input sizes, the batch size was 10. The input parameter tests were run in 500 epochs. The test with better accuracy results was the one with input size 300 × 300 and batch size 10, as per Fig. 3.

CNN-COVID training, testing, and validation

For Dataset I and II, the training and testing phases were performed as follows:

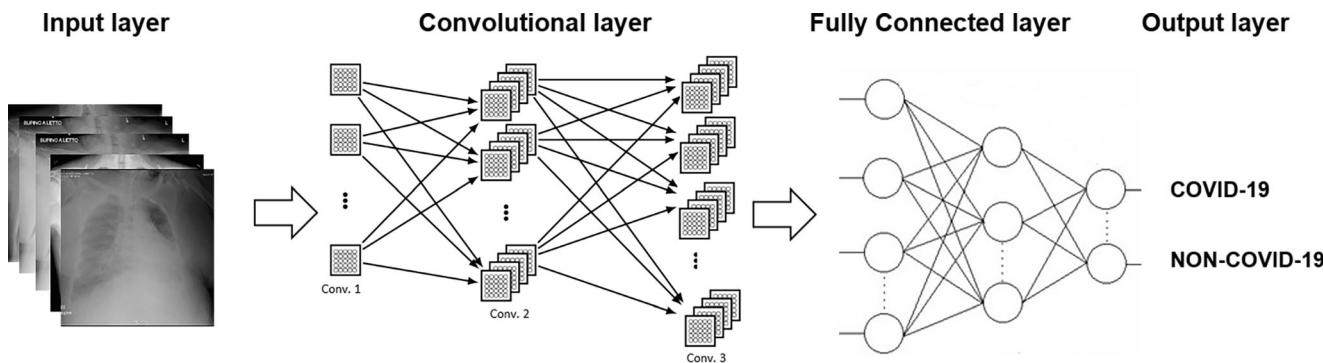
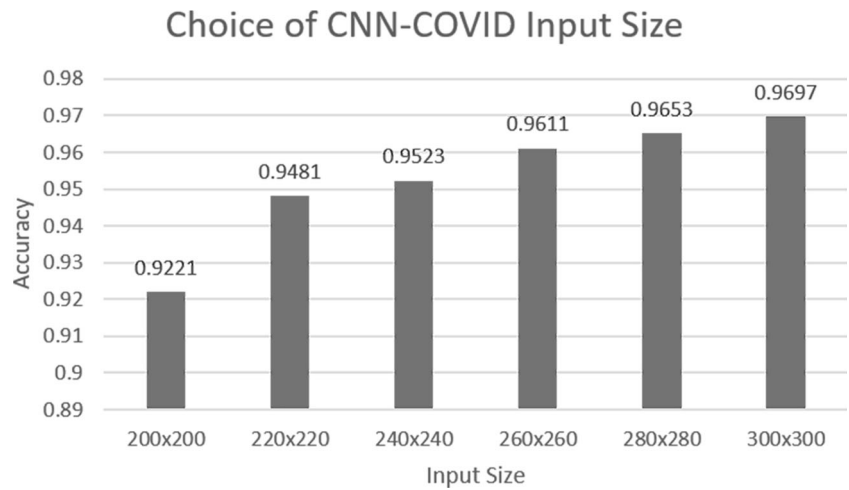


Fig. 2 Deep neural network classification scheme. Created by the author

Fig. 3 Choice of CNN-COVID Input Size (200 × 200 300 × 300). Created by the author



- Regarding Dataset I, the network was tested 10 times varying the parameters used by the ImageDataGenerator class to generate new samples on the image databases for both COVID-19 and NON-COVID classes. For the validation phase, the network with the best accuracy in the test phase (10 tests) was considered.
- Furthermore, for Dataset II, the network was also tested 10 times, and the training and testing database images were randomly selected. For the validation phase, the network with the best accuracy in the test phase (during 10 tests) was considered.

Performance metrics evaluation

The following metrics were used to validate the CNN-COVID system:

- Accuracy (ACC): accurate classification rate as per the total number of elements.
- Precision (P): ratio of true positive classifieds.
- Recall/Sensitivity (SE): true positive rate.
- Specificity (SP): true negative rate.
- F1-Score: relationship between precision and recall/sensitivity.

They are commonly used to assess the performance of classification algorithms (Ruuska et al. 2018; Skansi 2018; Khatami et al. 2017). There is a standard way to show the number of true positives (TP), false positives (FP), true

negatives (TN), and false negatives (FN) to be more visual. This method is called the confusion matrix. For a classification of two classes, the confusion matrix is presented in Table 4.

The confusion matrix allows us to determine the following metrics (Narin et al. 2020; Ruuska et al. 2018; Skansi 2018; Khatami et al. 2017):

$$\text{Accuracy : ACC} = \frac{\text{TN} + \text{TP}}{\text{TN} + \text{TP} + \text{FN} + \text{FP}} \tag{6}$$

$$\text{Precision : P} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{7}$$

$$\text{Recall/Sensitivity : SE} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{8}$$

$$\text{Specificity : SP} = \frac{\text{TN}}{\text{TN} + \text{FP}} \tag{9}$$

$$\text{F1 Score : F1-Score} = \frac{2 \times (\text{Precision} \times \text{Recall})}{(\text{Precision} + \text{Recall})} \tag{10}$$

Performance metrics help assess the CNN-COVID network and must be interpreted as follows: accuracy is the proportion of correct classification, considering the samples classified; precision, in turn, shows the odds of a positive anomaly be confirmed, recall is the ability of the model to identify all positive examples, and F-score is the harmonic average between precision and recall, best results being closer to 1 and worst, close to 0.

It is also possible to get the Receiver Operating Characteristic Curve (ROC curve). ROC analysis is often called the ROC accuracy ratio, a common technique for judging the accuracy of default probability models (Shirazi et al. 2018).

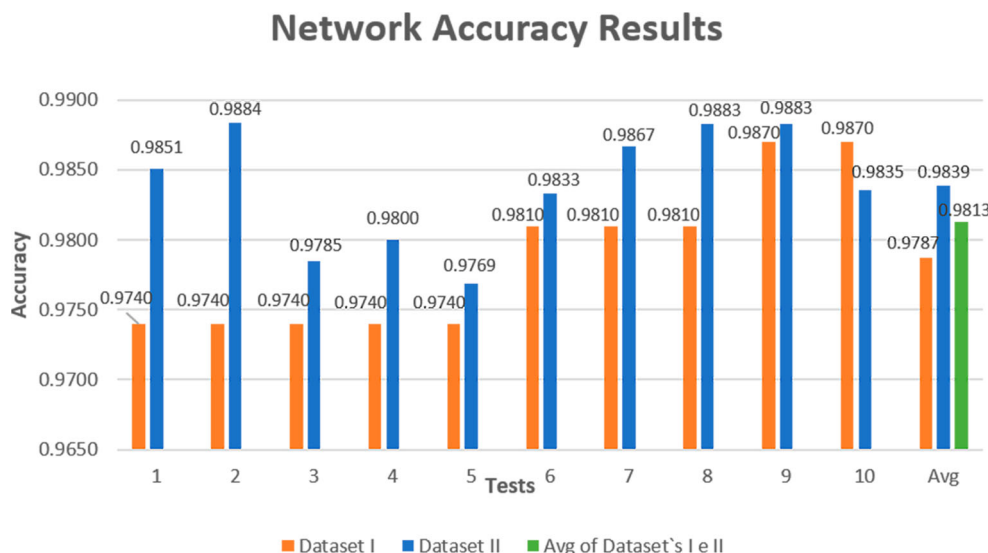
Table 4 Example of a confusion matrix. Adapted from (Skansi 2018)

	Classifier says YES	Classifier says NO
In reality YES	True positives	False positives
In reality NO	False negatives	True negatives

Results

A total of 10 trainings and tests were done with CNN-COVID. A total of 2000 epochs were applied to each training and test

Fig. 4 Dataset I and II: CNN-COVID results of each test and overall average (Avg) of all 10 tests. Source: Created by the author



in both Dataset I and II. The results are shown in Fig. 4. The average accuracy of Dataset I and II was 0.9787 and 0.9839, respectively, and the overall average between the two datasets was 0.9813.

To perform the validation phase, we take the weights of the best accuracy from the 10 tests obtained in CNN-COVID. So, the confusion matrix for Dataset I was generated with 126 COVID-19 images and 126 NON-COVID images which resulted in 252 images for validating the model. Otherwise, the confusion matrix for Dataset II was generated with 303 COVID-19 images and 303 NON-COVID images which resulted in 606 images for validating the model. Figure 5 shows the confusion matrix generated for Dataset I (Fig. 5a) and Dataset II (Fig. 5b), respectively.

Using the TP, TN, FP, and FN parameters, the accuracy, sensitivity, and specificity were calculated for both Dataset I and Dataset II, as registered in Table 5.

According to the results presented in Table 5, the ROC curve was calculated for both datasets as (1-SP) and SE, as x

and y, respectively. As per the ROC assessment, the area (AUC) is shown in Fig. 6.

From the results of Table 5, we calculated the average of the ACC, SE, and SP metrics for the two datasets, which was compared with the results of the state-of-the-art methods. Our method (CNN-COVID) presents better results for SE (98.83%) and ACC (98.03%) metrics, in comparison with other studies as shown in bold text in Table 6. In addition, the methodology applied at CNN-COVID is broader than that of the studies analyzed, as it uses the average of 10 tests, that is, the average accuracy of 10 tests made with CNN-COVID was considered.

Discussion

For the chest X-ray images that have various sizes over 1000 × 1000 pixels, a study was needed to decide which image sizes should be remodeled to have the best input size for

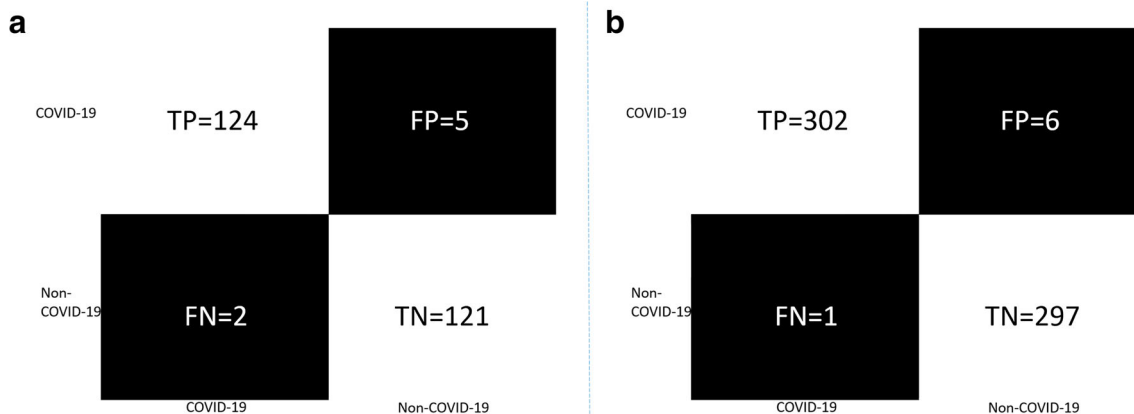


Fig. 5 CNN-COVID confusion matrix: **a** Dataset I; **b** Dataset II. Source: Created by the author

Table 5 Metrics results. Source: Created by the author

Class	Dataset	Accuracy	Recall/ Sensitivity	Specificity
COVID-19	I	0.9722	0.9800	0.9600
COVID-19	II	0.9884	0.9966	0.9801
AVG	I e II	0.9803	0.9883	0.9700

CNN-COVID. The initial input size was 200×200 , and then, it was increased by 20 until it reached 300×300 . CNN-COVID processed every input size in 500 epochs. This study showed that the higher the input size, the more accurate the network was. As per Fig. 2, when the input size was 300×300 , the accuracy was 0.9697.

The algorithm for gradient-based optimization of stochastic objective functions chosen was the Adaptive Moment Estimation (ADAM). It is an adaptive optimization technique which leverages both AdaGrad (the ability to deal with sparse gradients) and RMSProp (the ability to deal with non-stationary objectives) (Wani et al. 2020). Besides, the method is straightforward to implement and requires little memory.

For Dataset I, which presented a reduced number of COVID-19 images, augmenting the database was needed, to increase the CNN-COVID generalization. The database augmentation was done during run time. In the training phase, 126 COVID-19 images generated 252,000 new images, and 800 NON-COVID images generated 1,600,000 new images bringing it all to a total of 1,852,000 images. In the test phase, 40 COVID-19 images generated 80,000 new images, and 200 NON-COVID images generated 400,000 new images bringing it all to a total of 480,000 images. In the validation phase, the data generator was not used.

In this work, we trained and tested CNN-COVID 10 times, in order to obtain a more reliable value for ACC. The average accuracy

of the 10 tests was 0.9787 for Dataset I and 0.9839 for Dataset II (according to Fig. 4). As for the validation, the accuracy obtained was 0.9722 for Dataset I and 0.9884 for Dataset II (according to Table 5). With this result, it was found that the accuracy of the validation of Dataset I was close to the average accuracy of the tests with a percentage error less than 1%.

In Dataset II, the validation accuracy was higher than the average of the tests, showing that the average accuracy is actually more reliable. Thus, these results indicate that the investment of time, human, financial, and computational resources in the creation and improvement of techniques based on machine learning is a promising approach to assist professionals in the prognosis of the new coronavirus through chest X-ray images.

Conclusion and future work

This paper proposed a deep neural network, called CNN-COVID, for prognosis of the COVID-19 virus. Two different datasets were used: Dataset I is formed by a COVID-19 image data repository (Cohen et al. 2020), as well as the ChestXray14 data set (Wang et al. 2019), and Dataset II, on the other hand, is formed by the BIMCV COVID-19 + repository (Vayá et al. 2020). After completing the 10 tests, the average accuracy of Dataset I and Dataset II was 0.9787 and 0.9839, respectively. The weights of the best test results were applied in the validation, obtaining values of accuracy 0.9722 for Dataset I and 0.9884 for Dataset II.

The results proved that the CNN-COVID model is a promising tool to help physicians classify chest images with pneumonia, considering atypical pneumonia caused by COVID-19 and pneumonia due to other causes. We hope this technology enhances the provision of healthcare services, contributing to the disease prognosis through straightforward exams, such as

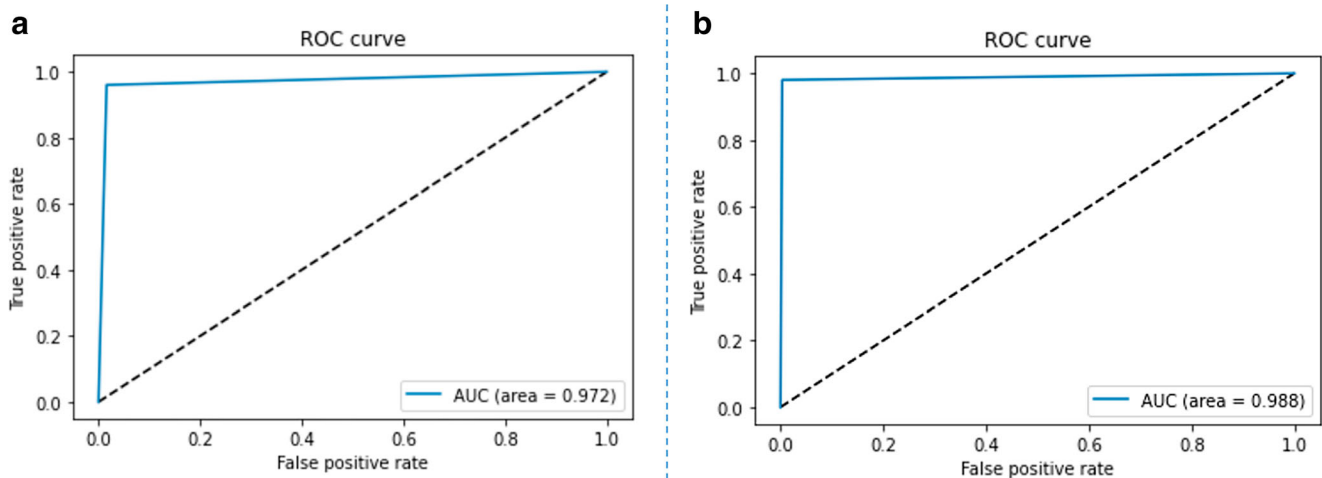


Fig. 6 CNN-COVID ROC curve: **a** Dataset I/AUC = 0.972; **b** Dataset II/AUC = 0.988. Source: Created by the author

Table 6 Comparison of CNN-COVID with state-of-the-art methods. Created by the author

Class	Methods	Size of the dataset COVID	Accuracy	Recall	Specificity
(Zhang et al. 2020)	Deep Anomaly Detection	100	0.9518	0.9600	0.7060
(Sethy and Behera 2020)-ResNet50	ResNet50	213	0.9538	0.9727	0.9347
(Sethy and Behera 2020)-VGG16	VGG16	213	0.9276	0.9747	0.8805
(Wang and Wong 2020)	COVID-NET	358	0.9333	0.9100	0.9900
(Abbas et al. 2020)	DeTrac	105	0.9512	0.9791	0.9187
Proposed method (AVG of Dataset I e II)	CNN-COVID	2025	0.9803	0.9883	0.9700

chest X-ray, and broadening the access to information through tools that help with using images for diagnosing.

For future work, we plan to improve the CNN-COVID accuracy as new COVID-19 data is collected.

Acknowledgements This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior—Brasil (CAPES)—Finance Code 001.

Compliance with ethical standards

Conflict of interest Author Pedro Moises de Sousa declares that he has no conflict of interest. Author Mariane Modesto Oliveira declares that she has no conflict of interest. Author Gabrielle Macedo Pereira declares that she has no conflict of interest. Author Carlos Alberto da Costa Junior declares that he has no conflict of interest. Author Luis Vinicius de Moura declares that he has no conflict of interest. Author Christian Mattjie declares that he has no conflict of interest. Author Pedro Cunha Carneiro declares that he has no conflict of interest. Author Ana Maria Marques da Silva declares that she has no conflict of interest. Author Ana Cláudia Patrocínio declares that she has no conflict of interest.

Ethical approval This article does not contain any studies with human participants or animals performed by any of the authors.

References

- Abbas A, Abdelsamea M, Gaber M. Classification of COVID-19 in chest X-ray images using DeTraC deep convolutional neural network. medRxiv. 2020. <https://doi.org/10.1101/2020.03.30.20047456>.
- Aggarwal CC, et al. Neural networks and deep learning. Springer. 2018;10:978–3. <https://doi.org/10.1007/978-3-319-94463-0>.
- Ai T, Yang Z, Hou H, Zhan C, Chen C, Lv W, et al. Correlation of chest CT and RT-PCR testing in coronavirus disease 2019 (COVID-19) in China: a report of 1014 cases. Radiology. 2020;10:200642–E40. <https://doi.org/10.1148/radiol.2020200642>.
- Balas VE, et al. (Ed.). Handbook of deep learning applications. Springer, 2019; <https://doi.org/10.1007/978-3-030-11479-4>
- Chollet F. Building powerful image classification models using very little data. Keras Blog, 2016. <https://blog.keras.io/building-powerful-image-classification-models-using-very-little-data.html>. Acessado em 01 de junho de 2020.
- Chung M, Bernheim A, Mei X, Zhang N, Huang M, Zeng X, et al. CT imaging features of 2019 novel coronavirus (2019-nCoV). Radiology. 2020;295(1):202–7. <https://doi.org/10.1148/radiol.2020200230>.

- Cui J, Li F, Shi Z-L. Origin and evolution of pathogenic coronaviruses. Nat Rev Microbiol. 2019;17(3):181–92. <https://doi.org/10.1038/s41579-018-0118-9>.
- Cohen, J.P., Morrison, P., Dao, L., Covid-19 image data collection. arXiv 2003.11597, 2020. URL: <https://github.com/ieee8023/covid-chestxray-dataset>. Acessado em 1 de junho de 2020.
- Dai W-C, Zhang HW, Yu J, Xu HJ, Chen H, Luo SP, et al. CT imaging and differential diagnosis of COVID-19. Can Assoc Radiol J. 2020;71(2):195–200. <https://doi.org/10.1177/0846537120913033>.
- Khatami A, Khosravi A, Nguyen T, Lim CP, Nahavandi S. Medical image analysis using wavelet transform and deep belief networks. Expert Syst Appl. 2017;86:190–8. <https://doi.org/10.1016/j.eswa.2017.05.073>.
- Kingma D and Ba J. Adam: a method for stochastic optimization. arXiv preprint arXiv:1412.6980, 2014.
- Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. In: Advances in neural information processing systems. 2012. p. 1097–1105.
- Lecun Y, Bengio Y, Hinton G. Deep learning. Nature. 2015;521(7553):436–44. <https://doi.org/10.1038/nature14539>.
- Martin DR, Hanson JA, Gullapalli RR, Schultz FA, Sethi A, Clark DP. A deep learning convolutional neural network can recognize common patterns of injury in gastric pathology. Archives of Pathology & Laboratory Medicine. 2020;144(3):370–8. <https://doi.org/10.5858/arpa.2019-0004-OA>.
- Narin A, Kaya C, Pamuk Z. Automatic detection of coronavirus disease (covid-19) using X-ray images and deep convolutional neural networks. arXiv, p. arXiv: 2003.10849, 2020.
- Ozturk S, Ozkya U, Barstugan M. Classification of coronavirus images using shrunken features. medRxiv, p. 2020.04.03.20048868, 2020; <https://doi.org/10.1101/2020.04.03.20048868>
- Ponti MA, Da Costa GBP. Como funciona o deep learning. arXiv preprint arXiv:1806.07908, 2018.
- Ravi D, Wong C, Deligianni F, Berthelot M, Andreu-Perez J, Lo B, et al. Deep learning for health informatics. IEEE Journal of biomedical and health informatics. 2017;21(1):4–21. <https://doi.org/10.1109/JBHI.2016.2636665>.
- Ruuska S, Hämäläinen W, Kajava S, Mughal M, Matilainen P, Mononen J. Evaluation of the confusion matrix method in the validation of an automated system for measuring feeding behaviour of cattle. Behav Process. 2018;148:56–62. <https://doi.org/10.1016/j.beproc.2018.01.004>.
- Sethy PK, Behera SK. Detection of coronavirus disease (COVID-19) based on deep features. Preprints. 2020;2020030300:2020.
- Shirazi AZ, Chabok SJS, Mohammadi Z. A novel and reliable computational intelligence system for breast cancer detection. Medical & Biological Engineering & Computing. 2018;56(5):721–32. <https://doi.org/10.1007/s11517-017-1721-z>.
- Skansi S. Introduction to deep learning: from logical calculus to artificial intelligence. Springer. 2018. <https://doi.org/10.1007/978-3-319-73004-2>.

- Szegedy C, et al. Going deeper with convolutions. In: Proceedings of the IEEE conference on computer vision and pattern recognition p 1-9, 2015; <https://doi.org/10.1109/CVPR.2015.7298594>.
- Vayá MDLI, et al. BIMCV COVID-19+: a large annotated dataset of RX and CT images from COVID-19 patients. arXiv preprint arXiv:2006.01174, 2020.
- Wang L, Wong A. COVID-Net: a tailored deep convolutional neural network design for detection of COVID-19 cases from chest radiography images. arXiv preprint arXiv:2003.09871, 2020.
- Wang X, et al. ChestX-ray: Hospital-Scale Chest X-ray Database and Benchmarks on Weakly. Deep Learning and Convolutional Neural Networks for Medical Imaging and Clinical Informatics, p. 369, 2019; https://doi.org/10.1007/978-3-030-13969-8_18
- Wani MA, et al. Advances in deep learning. Springer. 2020. <https://doi.org/10.1007/978-981-13-6794-6>.
- Weiss SR and Leibowitz JL. Coronavirus pathogenesis. In: Advances in virus research. Academic Press, v.81, n. 1, p. 85–164, 2011; <https://doi.org/10.1016/B978-0-12-385885-6.00009-2>
- Yang W, Cao Q, Qin L, Wang X, Cheng Z, Pan A, et al. Clinical characteristics and imaging manifestations of the 2019 novel coronavirus disease (COVID-19): a multi-center study in Wenzhou city, Zhejiang. China Journal of Infection. 2020;80:388–93. <https://doi.org/10.1016/j.jinf.2020.02.016>.
- Zeiler MD and Fergus R. Visualizing and understanding convolutional networks. In: European conference on computer vision. Springer, Cham, 2014. p. 818–833; https://doi.org/10.1007/978-3-319-10590-1_53.
- Zhang J, et al. Covid-19 screening on chest X-ray images using deep learning based anomaly detection. arXiv preprint arXiv:2003.12338, 2020.
- Zhu N, Zhang D, Wang W, Li X, Yang B, Song J, et al. A novel coronavirus from patients with pneumonia in China, 2019. N Engl J Med. 2020;382(8):727–33. <https://doi.org/10.1056/NEJMoa2001017>.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.