



## OPEN Coronavirus covid-19 detection by means of explainable deep learning

Francesco Mercaldo<sup>1✉</sup>, Maria Paola Belfiore<sup>2</sup>, Alfonso Reginelli<sup>2</sup>, Luca Brunese<sup>1</sup> & Antonella Santone<sup>1</sup>

The coronavirus is caused by the infection of the SARS-CoV-2 virus: it represents a complex and new condition, considering that until the end of December 2019 this virus was totally unknown to the international scientific community. The clinical management of patients with the coronavirus disease has undergone an evolution over the months, thanks to the increasing knowledge of the virus, symptoms and efficacy of the various therapies. Currently, however, there is no specific therapy for SARS-CoV-2 virus, know also as *Coronavirus disease 19*, and treatment is based on the symptoms of the patient taking into account the overall clinical picture. Furthermore, the test to identify whether a patient is affected by the virus is generally performed on sputum and the result is generally available within a few hours or days. Researches previously found that the biomedical imaging analysis is able to show signs of pneumonia. For this reason in this paper, with the aim of providing a fully automatic and faster diagnosis, we design and implement a method adopting deep learning for the novel coronavirus disease detection, starting from computed tomography medical images. The proposed approach is aimed to detect whether a computed tomography medical images is related to an healthy patient, to a patient with a pulmonary disease or to a patient affected with Coronavirus disease 19. In case the patient is marked by the proposed method as affected by the Coronavirus disease 19, the areas symptomatic of the Coronavirus disease 19 infection are automatically highlighted in the computed tomography medical images. We perform an experimental analysis to empirically demonstrate the effectiveness of the proposed approach, by considering medical images belonging from different institutions, with an average time for Coronavirus disease 19 detection of approximately 8.9 s and an accuracy equal to 0.95.

The SARS-CoV-2 virus enters the body by binding to the angiotensin 2 converting enzyme, an enzyme involved in blood pressure regulation and found on the cells of the lung epithelium where it defends the lungs from damage from infections and inflammation. The virus, by binding to ACE2, enters the cell and prevents the enzyme from fulfilling its protective role<sup>1</sup>.

Once in the cells, SARS-CoV-2 begins to replicate and clinically this phase is typically characterized by malaise, fever and dry cough<sup>2</sup>.

In some cases the SARS-CoV-2, also know as *covid-19* (acronym of the English COronaVIRus Disease 19) can evolve into a second phase which is characterized by alterations in the lungs with interstitial pneumonia<sup>3</sup>, very often bilateral and therefore with the involvement of both lungs, associated with respiratory symptoms that may initially be limited but which may lead to progressive clinical instability with respiratory failure<sup>4</sup>.

In a restricted number of patients, the clinical picture may worsen. As a matter of fact the infection leads to a state of excessive inflammation<sup>5</sup>, with local and systemic consequences (i.e., of the whole organism), with the risk of serious and sometimes permanent lung lesions (pulmonary fibrosis)<sup>6</sup>.

A clinical picture that can worsen further and lead to severe acute respiratory distress syndrome and sometimes to disseminated intravascular coagulation phenomena, with the formation of thrombus in small vessels throughout the body and the potential interruption of the normal flow of blood<sup>7,8</sup>.

In light of these three phases of the disease and also taking into account the radiological criteria, the US National Institutes of Health (NIH) have formulated a classification of the five clinical stages of *covid-19*<sup>9,10</sup>:

<sup>1</sup>Department of Medicine and Health Sciences "Vincenzo Tiberio", University of Molise, Campobasso, Italy. <sup>2</sup>Department of Precision Medicine, University of Campania "Luigi Vanvitelli", Naples, Italy. ✉email: francesco.mercaldo@unimol.it

- *Asymptomatic or pre-symptomatic infection*: there is a diagnosis of SARS-CoV-2, but a complete absence of symptoms<sup>11</sup> (but the subject is still contagious even in the absence of symptoms);
- *Mild illness*: the patient has mild symptoms (fever, cough, altered taste, malaise, headache, myalgia or muscle aches), but there is neither dyspnoea (breathing difficulties) nor radiologically detectable changes<sup>12</sup>;
- *Moderate disease*: the patient has a saturation—that is the oxygenation of the blood that is detected with an oximeter, greater than or equal to 94% and there is clinical or radiological evidence of pneumonia<sup>13</sup>;
- *Severe disease*: where one of the parameters is the saturation is less than 94%<sup>14</sup>;
- *Critical illness*: with respiratory failure, septic shock and/ or failure of one or more organs.

People can be infected with this disease through coughing and in general with direct contacts<sup>4</sup>.

To limit transmission, precautions must be taken<sup>15</sup>, such as maintaining a safety distance of at least 1.5 m, and maintaining correct hygiene behavior (periodically washing and disinfecting hands<sup>16</sup>, sneezing or coughing into a handkerchief or with the elbow bent and where necessary wear masks and gloves)<sup>17</sup>.

Medical imaging (for example, chest computed tomography) is able to show pneumonia in patients with *covid-19*<sup>18</sup>. For this reason, the World Health Organization published several additional diagnostic protocols (<https://www.who.int/emergencies/diseases/novel-coronavirus-2019/technical-guidance/laboratory-guidance>) for increase *covid-19* detection from medical images.

Currently, diagnostic tests for the confirmation of *covid-19* infection are performed in various public and private laboratories.

The tests currently available to detect SARS-CoV-2 infection are as follows:

- molecular test, which highlights the presence of genetic material of the virus. It is performed on a rhino-pharyngeal swab;
- antigen test, which highlights the presence of components of the virus. It is performed on a rhino-pharyngeal swab;
- traditional or rapid serological test, which shows the presence of antibodies against the virus. Serological tests are performed on venous sampling and capillary blood.

The main problem of these tests is represented by the fact that their results are typically available in some hour or also in days<sup>19,20</sup>.

As discussed in current literature in this context<sup>21,22</sup> there is the possibility to better diagnose *covid-19* by exploiting radiological imaging: this is the reason in this paper we consider the possibility to diagnose the *covid-19* disease using medical images in particular, computed tomography (CT) have been used.

In last years research community demonstrate, with several studies<sup>2,23,24</sup>, that it is possible to detect pulmonary diseases by consider the analysis medical images with artificial intelligence techniques. In a nutshell artificial intelligence currently represents a research field aimed to build models starting from data. In last is adopted from both the industrial that academic researchers for the development of methods for assist experts in the medical images interpretation.

We exploit in the proposed paper the transfer learning i.e., a method aimed to adapt a model to a task different from the one for which it was initially trained. The intuition behind the transfer learning idea is the the knowledge learned in a certain context can be immediately reapplied to another context by “retuning”: in this way it is possible to avoid to retrain the model from scratch.

Transfer learning allows two results, the first one is represented by the reuse of the behavior of a network already trained to effectively extract features from input data and the second one is represented by limiting the processing to a significantly smaller number of parameters (corresponding to the last layers)<sup>25,26</sup>.

To support radiologists and pathologists in the task of screening the populations, we consider the usage of deep learning with the aim to understand whether there is the presence of *covid-19* in CT images by considering transfer learning. Moreover, one of the distinctive points of the proposed contribution is the automatically localise the infected lung areas. To do this, the network activation layers are exploited: in a nutshell we consider the CT lung areas considered by the trained model for prediction generation, with the aim to furnish a kind of prediction explainability. The idea behind the activation layers adoption is that they can be considered as an insight for the radiologist to visualise the areas of the CT exams which deserve an in-depth analysis.

Artificial intelligence for the *covid-19* identification was already explored in<sup>4</sup> for covid-19 diagnosis. With respect to the work presented in<sup>4</sup> is the adoption of CT images, which offer diagnostic potential far superior to x-rays<sup>27–29</sup>.

As a matter of fact, the difference between CT and X-rays is really significant. As a matter of fact, pathologists and radiologists typically analyse X-rays to detect bone dislocations and fractures, but also to detect tumors and pneumonia. However, CT scans are a kind of advanced X-ray devices that pathologists and radiologists use to better diagnose injury to internal organs.

X-ray machines can fail to diagnose issues with damage to muscle, and in general to soft tissue. Differently, with CT scans is possible to have evidence of these problems<sup>28</sup>. Furthermore, X-rays show a 2D representation of the tissue under analysis, while from the other side the CT shows a 3D representation. In CT exams a thin layer of the body is crossed by a highly collimated X-ray beam, produced by a tube that rotates around the patient, in a consensual manner to detectors placed beyond the patient. CT exams offer a layered visualization of the anatomical structures eliminates the problem of overlapping present in the x-ray examination (that can be reflected by noise in the data), revealing the presence of injury or disease to pathologists and radiologists. Furthermore, the approach proposed in<sup>4</sup> considers two different step to detect the *covid-19* detection, while the proposed approach in a single step is able to detect whether a patient is affected by *covid-19*, a generic pulmonary

disease or he/she is an healthy patient. Finally, while the method in<sup>4</sup> is evaluated by considering a dataset freely available for research purposes, in this paper we evaluated the proposed approach with a dataset personally gathered and labelled by authors.

The idea behind the proposed method is the proposal of a deep learning network aimed to classify a pulmonary CT exams as related to a patient affected by the covid-19 disease, by other pulmonary disease or as healthy patient. The convolutional neural network is designed by authors. Moreover the proposed method is aimed to automatically localise the region of interest i.e., the areas of the image under analysis symptomatic of the covid-19 infection (for this task we resort to the grad-cam algorithm). Through the visualization of the areas of interest we are able to understand the reasons why the proposed model obtains a satisfactory accuracy. In fact, the proposed model is able to accurately localize the signs of the covid-19 in the medical image and is able to distinguish them from those of other lung diseases which, although giving similar symptoms, are not covid-19. The proposed model is able to understand this difference thanks to the use of different convolutional layers that allow it to extract different characteristics of the image at different levels of depth. In fact, tests made by the authors with convolutional models with a lower number of layers obtain unsatisfactory results. Convolutional layer consists of a collection of digital filters to perform the convolution operation on the input data, therefore by inserting more layers more features are extracted from the images and therefore it is possible to do finer-grained analysis obtaining better results. As a matter of fact, the multiple layers in deep neural networks allow models to become more efficient at learning complex features and performing more intensive computational tasks, i.e., execute many complex operations simultaneously. This is due to the ability of deep learning models to eventually learn from own errors, considering the deep learning model ability to verify the accuracy of its predictions/outputs and make the necessary adjustments.

The paper continues in the following way: next section presents the proposed method, experiments are presented in “The evaluation” section; in “Discussion” section a state-of-the-art discussion about the adoption of artificial intelligence for pulmonary disease detection with a specific focus on *covid-19* disease is provided, and, finally, in the last section, conclusion and future research plan are drawn.

## Materials and method

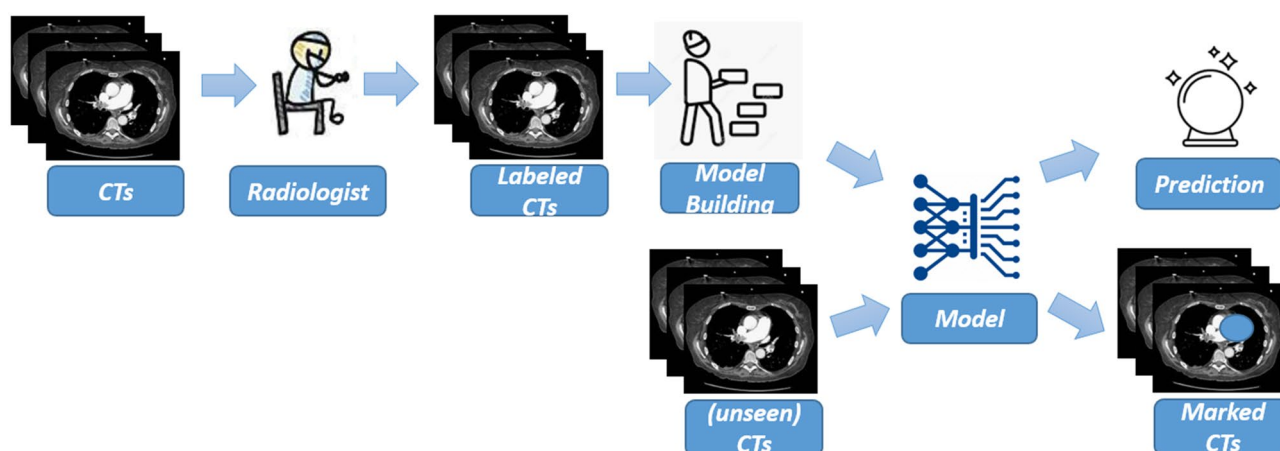
The typical image classification problem is one of the most popular tasks in deep learning. It basically consists of classifying images that contain items or generic shapes (as, for instance, typewritten letters) with the highest accuracy possible. The deep learning model completes the task by leveraging the information of a dataset of input samples. In the training phase, the deep learning models extract and memorise features and patterns peculiar to a specific output class, thus learning how to distinguish between the different input samples.

One of the most widely used deep learning models for image classification is the Convolutional Neural Network (CNN), which exploit mathematical convolutional operators on the input image to extract features. The input images pass through several layers of convolution, to combine the pixels with the neighbouring ones, and subsampling, to reduce the size of the two-dimensional matrix while preserving the most relevant information. Finally, the last part of the CNN is usually composed of dense layers, which are formed by a variable number of *perceptrons*; this last part of the model perform the classification, and it is trainable with the standard back-propagation algorithm. We refer to the literature for further information on the CNNs<sup>30,31</sup>.

Many complex CNN variants were proposed in the literature; mainly, they differ in the size of the architecture and the number of convolutional layers. In this paper, we experiment with a CNN designed by the authors.

Figure 1 shows the main architecture of the approach we propose for *covid-19* detection starting from CT analysis.

We obtained a data-set of CTs, obtained from Italian hospitals (Italy was one of the first European countries where *covid-19* has widely spread<sup>15</sup>), belonging to different patients afflicted by different pulmonary diseases (including the *covid-19* coronavirus). Moreover also healthy patients are considered as a matter of fact the idea behind the proposed method is to infer a model, directly from the patients data, aimed to automatically identify



**Figure 1.** The workflow of the proposed method.

COVID-19 patients. For this reason we need patients afflicted by *covid-19*, but also of patients afflicted from other pulmonary diseases and healthy ones, to make able to deep learning network to infer the distinctive characteristics of *covid-19* infection.

The CTs collected are scrupulously verified by expert radiologists with the aim to add a label to each patient (i.e., *COVID-19* is for a patient afflicted by the new coronavirus, *other* for patients diagnosed with other pulmonary disease and *healthy* for patient not afflicted by pulmonary diseases).

Once obtained the labelled patient CTs we input a deep learning network designed by authors, in detail we apply transfer learning.

We resort to transfer learning with fine-tuning for model training. In a nutshell we initialize the VGG-Net network by exploiting weights pre-trained on the ImageNet dataset, cutting out the head related to the fully connected layer. An additional fully-connected layer head is added. It consists of several layers: Average-Pooling2D, Flatten, Dense, Dropout and , finally, a Dense layer exploiting the “softmax” activation devoted to label prediction<sup>32</sup>. We add it above VGG16. We then freeze the VGG16’s convolutional weights in such a way that only the fully connected level head is trained and this last step completes our model configuration.

The first layer is represented by convolutional of a size fixed to  $224 \times 224$  RGB image, for this reason CT slices are resized to this dimension. The lung CT slice is passed through a series of convolutional layers: the convolution is set to 1 pixel; the spatial padding of convolutional layer input is such that the spatial resolution is preserved after convolution. Spatial pooling is performed by exploiting 4 max-pooling layers, which follow some of the convolutional layers. Max-pooling is performed considering a two pixels high and two pixels wide window.

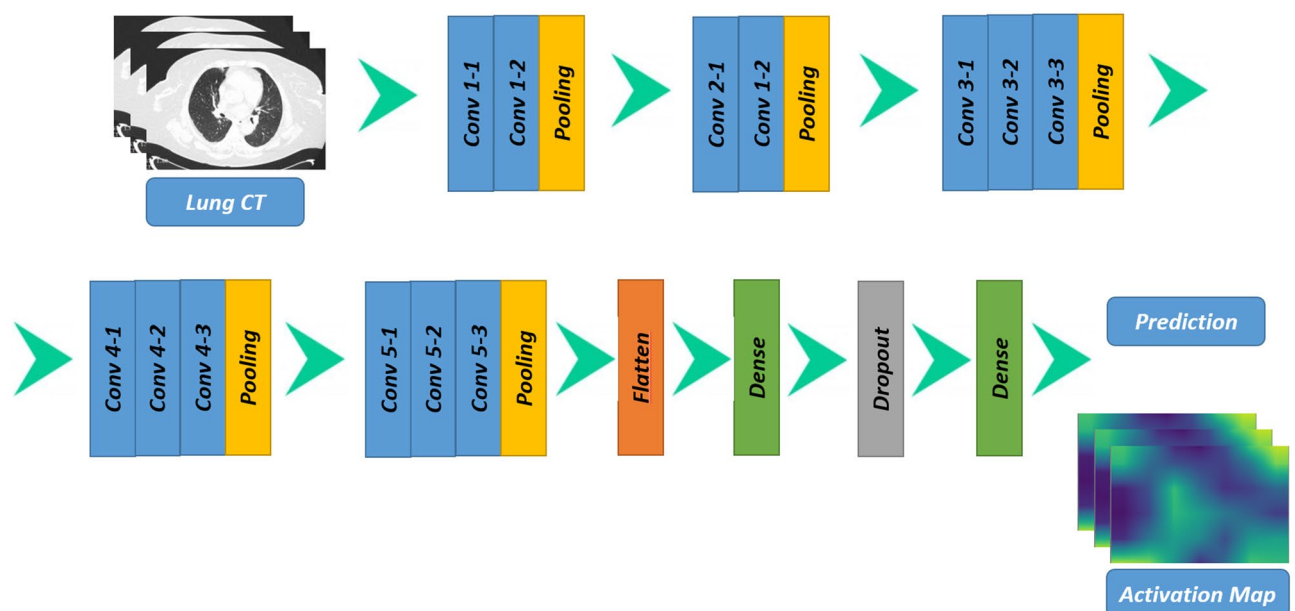
The first layer is represented by a convolutional layer of a fixed dimension on an RGB image of  $224 \times 224$ , this is the reason why the CT slices are resized at this height and width. The lung CT slice is passed through a series of convolutional layers: the convolution is set to 1 pixel; the spatial padding of the convolutional level input is such that the spatial resolution is preserved after the convolutional operation. Spatial pooling is performed using 4 max-pooling layers, which follow some of the convolutional layers. Max-pooling considers a window of two pixels high and two pixels wide.

All the considered hidden layers consider a rectification non-linearity. More details on the VGG-16 model architecture can be found in reference<sup>33</sup>.

We add to the VGG-16 architecture a set of layers (as shown by the deep learning model we depicted in Fig. 2): *AveragePooling2D*, *Flatten*, *Dense*, *Dropout* and another *Dense* layer.

In detail:

- *AveragePooling2D*: this layer is aimed to perform medium pooling operations. This level involves averaging for each patch of the feature map being analyzed. This means that each two pixels high and two pixels wide square of the feature map is sampled at the mean value;
- *Flatten*: the purpose of this layer is to flatten the input. It is considered a level of utility: it transforms an input, for example a row  $\times$  column matrix, into a simple vector output in the form of rows  $\times$  columns. Flattening transforms a two-dimensional array of features into a vector that can be inserted into a fully connected neural network classifier;
- *Dense*: is the regular deeply connected neural network layer. It is aimed to data transformation. It is most common and frequently used layer. In this case this case layer reduces the vector of height 512 to a vector of 64 elements;



**Figure 2.** The proposed deep neural network for COVID-19 explainable detection.

- **Dropout:** this layer basically works in the following way, namely by randomly selecting neurons not considered in the training. The purpose of this level is to improve generalization, in fact we are forcing the network to train the same high-level concept by exploiting different neurons. We chose to ignore the 50% of neurons. We are aware that typically exploiting this level may result in worse performance, but we want to generate a model that is less sensitive to changes in the data;
- **Dense:** the last dense layer is aimed to reduce the vector of height 64 to a vector of 2 elements (i.e., the classes to predict).

Further details on the VGG-16 model architecture can be found in reference<sup>33</sup>.

To furnish explainability<sup>4</sup>, we resort to a visualisation considering activation maps, typically exploited for deep learning network debug. For this purpose, we resort to the Gradient-weighted Class Activation Mapping (Grad-CAM) algorithm<sup>34</sup>.

The Grad-CAM is a technique to extract the gradients of the DL models convolutional layers and use them to provide graphical information on the inference step. Briefly, the gradients capture high-level visual patterns and can describe which areas of the input image have influenced the most the model output decision. Also, the convolutional layer preserves spatial information, thus, the Grad-CAM uses this data to provide a heatmap of the input image. This heatmap highlights the input image area which was used by the DL model to classify a specific input; it provides a visual “explanation” to a certain decision. The Grad-CAM adopted in this work is an implementation of the one introduced by this paper<sup>34</sup>.

Simply put, Grad-CAM uses the gradients of any target concept, flowing into the final convolutional layer to produce a coarse location map that highlights the important regions in the image to predict the concept.

By exploiting the Grad-CAM, it is possible to validate in a visual and immediate way where the network is looking at where a chest X-ray is evaluated: in this way it is possible to verify that it is actually looking at the right patterns in the image and activating around those patterns.

Grad-CAM works by observing the last network convolutional layer and then examining the gradient information flowing into that layer. The output of the Grad-CAM is a heatmap visualization for a given predicted label. We consider the generated heatmap to visually verify where the convolutional neural network is looking in the image, as shown in the experimental analysis section.

## The evaluation

In the follow we present the results we obtained by the experimental analysis.

As stated into the previous section, we gathered a dataset to validate the effectiveness of the proposed method. We obtain CTs and the radiologist diagnosis for the following 45 patients:

- 20 patients afflicted by *covid-19* disease;
- 9 patients afflicted by *other* pulmonary diseases;
- 16 patients without pulmonary diseases, labelled as *healthy*.

The proposed method marked as *covid-19* 19 patients on 20 afflicted by COVID-19. With regard to the 9 patients afflicted by other pulmonary diseases only 1 was wrongly labelled as *covid-19*, while the remaining ones were rightly labelled as *non covid-19*. All the 16 *healthy* were rightly labelled as *non covid-19* by the proposed method. For each patient, in average, we have 400 images, for an average total number equal to 18,000 images. The medical images were obtained from the Department of Precision Medicine, belonging to the University of Campania, Caserta, Italy.

The 400 central images of the CT examination were considered for each patient, thus excluding the totally black ones that did not present any organ.

We train the designed deep learning network by considering the 80% of the dataset, while the remaining 20% is exploited for the model testing. Balanced instance were considered for the three classes involved in the experiment (i.e., covid-19, other and healthy), with a cross-validation with  $k = 5$  in order to evaluate all the patients.

In  $k$ -fold cross validation, the dataset is divided into a series of equal portions of data ( $k$ -fields) and a set of images belonging to the dataset is exploited for the training, while the remaining is exploited for the testing. The cross-validation is a technique for training a model on subsets of the available input data and evaluating them against a complementary subset of the data. We resort to cross-validation to avoid overfitting, which is the failure to generalize a model. In practice, the  $k$ -fold cross-validation method divides the input data into  $k$  subsets of data (also known as folds). We train a model with  $(k-1)$  subsets of data, then we evaluate the model on the subset that was not used for training i.e., the remaining one. This process is repeated  $k$  times, each time with a different subset reserved for evaluation (and clearly excluded from training). In the experimental analysis we selected  $k = 5$ : in this way the dataset was split into 5 equal parts, where each part contains the 20% of the images belonging to the dataset. Starting from this, 5 different model training were carried out, where for each training the 80% of the dataset (therefore 4 parts out of 5) was considered for training and the remaining 20% for testing. In subsequent training, the remaining 20% was used for testing. The final performance values are therefore the average of the performance values obtained in the 5 classifications, where a 20% (different for each classification) of the dataset was used in each classification.

Table 1 shows the performance obtained by the proposed method in terms of sensitivity, specificity, f-measure and accuracy.

As shown from the experimental analysis evaluation in Table 1, a sensitivity and a specificity equal to 0.95 is reached with regard to the discrimination between patients afflicted by *covid-19* disease, patient afflicted by other pulmonary diseases and healthy ones.

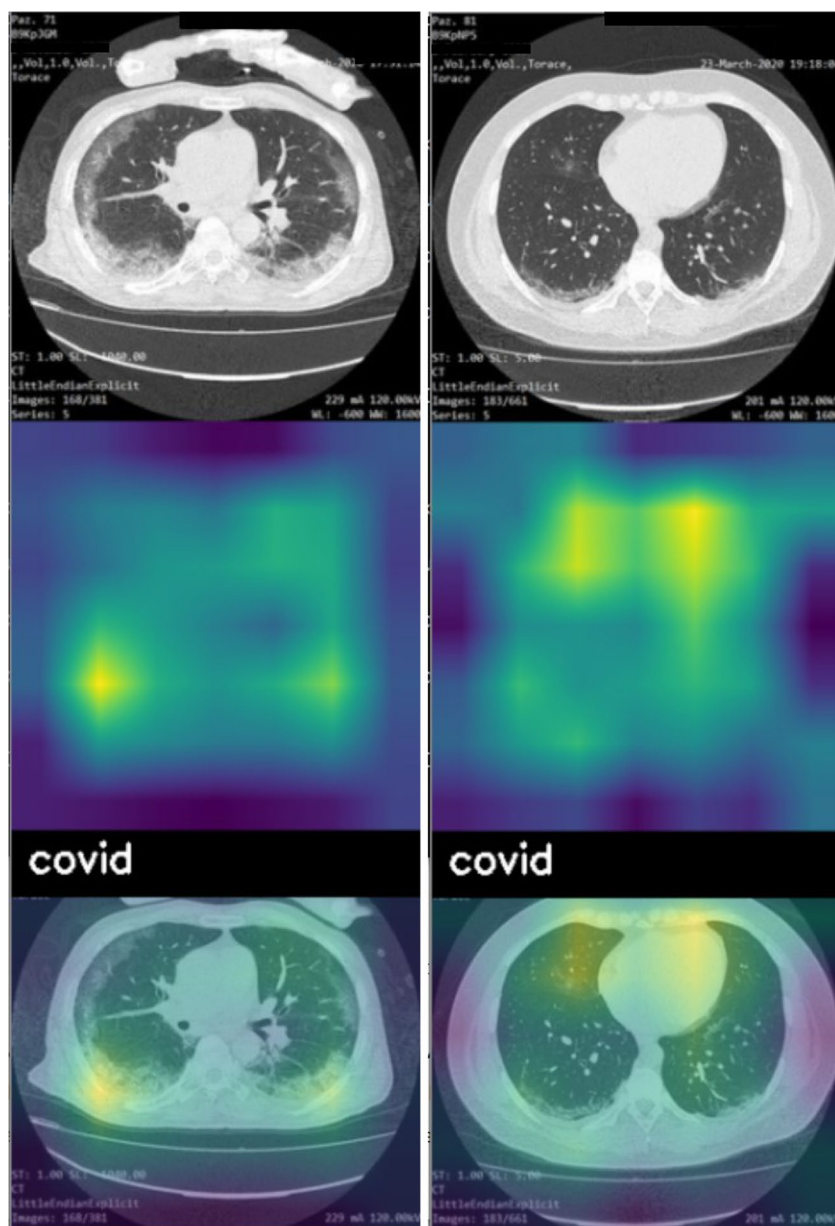


Model	Sensitivity	Specificity	F-measure	Accuracy
Our method	0.95	0.95	0.95	0.95

**Table 1.** Experimental analysis evaluation.

Figure 3 provides a couple of example of explainability automatically inferred by the proposed approach. In fact, we show two slices (belonging to two different *covid-19* patients as shown from the labels applied by the deep learning model), the activation maps and the overlay between the slices and the activation maps with the aim to show the area of interest (highlighted in yellow and green, while the blue area are not considered of interest from the model to perform the detection).

From the activation map related to the first slice (in the left part of Fig. 3) we note that the low areas of the CT are symptomatic of the *covid-19* disease infection, while in the second slice (in the right part of Fig. 3) it seems that the upper areas are more symptomatic of the *covid-19* disease infection. The overlay of the activation maps with the slices provide explainability about the relevant areas symptomatic of the *covid-19* disease. Moreover,



**Figure 3.** Two slices belonging to two different *covid-19* patients with the related activation maps: in yellow the areas symptomatic of the *covid-19* infection.

they can provide interesting insights for radiologists for the localisation of the areas interested by the *covid-19* infection. The outcomes of the activation maps were confirmed also by expert radiologists.

The analysis requires for a new CT series approximately 8.9 s to make the prediction and the visualise the activation maps. The machine used to run the experiments and to take measurements was an Intel Core i7 8th gen, equipped with 2GPU and 16Gb of RAM.

## Discussion

Current literature presents several research papers focused to *covid-19* detection. Below we discuss these papers.

Authors in<sup>35</sup> exploit deep learning obtaining an accuracy equal to 98.85%. The main difference with respect to the proposed method is that authors in<sup>35</sup> do not take into account healthy patients. Moreover they do not provide explainability about their network predictions.

Chen et al.<sup>35</sup> considering a deep network achieving an accuracy of 98.85%. The main difference from the proposed method is that authors in<sup>35</sup> do not consider patients without pulmonary pathologies. They also do not provide explainability about the deep learning predictions.

Wang et al.<sup>36</sup> design a deep learning model for detecting *covid-19*. The authors do not consider patients without pulmonary pathologies and do not provide some sort of explanation on the results of their network. Additionally, the method in<sup>35</sup> requires you to manually tag the *covid-19* region of interest. They obtained accuracy of 73.1% by evaluating a dataset obtained from two hospitals.

Xu and colleagues<sup>37</sup> propose a deep learning approach obtaining a 86.7% accuracy. Two three-dimensional convolutional neural networks are exploited: the first is the ResNet23 network, while the second network represents a variant of the first, in which the authors have added several layers.

Below, we discuss research works<sup>38,39</sup> on the application of deep learning techniques to proteins with the aim of promoting the study for new vaccines.

Zhang et al.<sup>39</sup> leverages deep learning techniques by exploiting *covid-19* RNA sequences to predict which current antivirals may help *covid-19* patients.

Beck et al.<sup>38</sup> experiment the adoption of deep models in combination with a molecule repository. The outcome of this research is that 2019-nCoV 3C proteinase is expected to bind with atazanavir, which is an antiviral drug used to treat Human Immunodeficiency Virus.

Apostolopoulos and colleagues<sup>40</sup> exploit transfer learning reaching a 0.98 detection rate in the discrimination between *covid-19* and healthy patient.

Zheng and colleagues in reference<sup>41</sup> adopt a model for detecting *covid-19* from CT images reaching a detection rate equal of 0.9, while researchers in<sup>37</sup> obtain a 0.86 detection rate in *covid-19* identification exploiting the ResNet network by analyzing CT images. Most of these articles, with respect to the proposed work, evaluate a small amount of data for model training and they do not provide explainability of the results.

Narin et al. in<sup>42</sup> propose the usage of 3 networks evaluating 50 *covid-19* chest X-ray images and 50 heartily images gathered from a Kaggle competition (<https://www.kaggle.com/paultimothymooney/chest-xray-pneumonia>). It should be noted that in the reference<sup>42</sup> researchers consider non-COVID images belonging to children, while those *covid-19* are related to adult patients.

Researchers in<sup>43</sup> achieved an accuracy equal to 0.86 analysing CT images with a deep model built on the ResNet50 model. Wang et al.<sup>36</sup> achieved an accuracy of 0.82 exploiting the modified Inception (M-Inception) deep model by analysing CT images.

Song et al.<sup>43</sup> reached a 0.86 accuracy by analyzing CT images with a deep learning network built on the ResNet50 model, while researchers in<sup>36</sup> reached a 0.82detection rate by exploiting the Inception network with CT medical images.

Ozturk and colleagues<sup>44</sup> analyse 1750 x-ray adopting the DarkCovidNet network: in detail they consider 250 *covid-19* positive patient, 500 related to generic lung diseases and 1000 obtained from healthy patients obtaining a detection rate equal to 0.98.

Ardakani et al.<sup>45</sup> propose a deep learning based approach for distinguishing from *covid-19* from non-*covid-19*. Ten different well-known convolutional neural networks were used. Using the Xception model they obtained an accuracy equal to 0.99.

Researchers in<sup>45</sup> design a method aimed to discriminate between *covid-19* and non-*covid-19* patients. Ten deep learning models are considered for this purpose and a detection rate equal to 0.99 is obtained with the Xception network.

Authors in<sup>18</sup> leveraged the COVNet model to identify the *covid-19* by learning a ResNet50 models, by reaching a detection rate of 0.96. Also this method, differently from the one we proposed, does not provide prediction explainability.

Transfer learning is exploited also in<sup>4</sup> for covid detection. Differently from the proposed method, in reference<sup>4</sup> the analysis is performed on x-ray images.

Authors in<sup>46</sup> propose a method aimed to detect finger skin. They exploit three kinds of images: 60 h after injury, 160 h after injury, 450 h after injury. Authors state that the advantage of the presented method is the automatic detection of the finger skin using a smartphone and they method can be helpful to diagnose pathologies of human skin.

Piekarski and colleagues<sup>47</sup> exploit a CNN for fault detection in time series data. In a nutshell, they propose a method aimed to detect the abnormal status of sensors in certain time steps. They consider transfer learning by examining pre-trained VGG-16, VGG-19, InceptionV3 and Xception CNN models with an adjusted densely-connected classifiers.

In Table 2 we compare the current state-of-the-art in COVID detection by means of deep learning in terms of acquisition (X-ray or CT), number of images analysis (Images column) and the obtained accuracy.

Method	Acquisition	Images	Accuracy
Apostolopoulos et al. <sup>40</sup>	X-ray	1428	0.93
Wang and Wong <sup>48</sup>	X-ray	13,645	0.92
Sethy et al. <sup>49</sup>	X-ray	50	0.95
Hemdan et al. <sup>50</sup>	X-ray	50	0.90
Narin et al. <sup>42</sup>	X-ray	100	0.98
Song et al. <sup>43</sup>	CT	1485	0.86
Wang et al. <sup>36</sup>	CT	453	0.82
Zheng et al. <sup>41</sup>	CT	542	0.90
Xu et al. <sup>37</sup>	X-ray	618	0.86
Ozturk et al. <sup>44</sup>	X-ray	1750	0.92
Ardakani et al. <sup>45</sup>	CT	1020	0.99
Li et al. <sup>18</sup>	CT	4356	0.96
Butt et al. <sup>51</sup>	CT	1710	0.86
Brunese et al. <sup>4</sup>	X-ray	6523	0.97
Our method	CT	18,000	0.95

**Table 2.** State-of-the-art comparison.

Moreover in Table 3 we compare four of the most exploited deep learning model used in covid-19 detection<sup>52</sup> with the dataset we gathered. The aim of this experiment is to perform a direct comparison between the deep learning models currently employed by the state-of-the-art and the proposed model.

As shown from the experimental results obtained with the proposed model in Table 1, the state-of-the-art models obtain an accuracy ranging from 0.71 to 0.93, while the proposed method is able to reach an accuracy equal to 0.95, thus confirming the effectiveness of the proposed method.

In Fig. 4 we report a chart aimed to present a direct visual comparison between the state-of-the-art models and the proposed deep learning model in terms of accuracy.

From Fig. 4 it emerges that the model obtaining the worst performances is the AlexNet one, the InceptionV3 model obtains slightly better performances, the ResNet50 and VGG19 models obtain a good interesting and, finally, the proposed model with an accuracy of 0.95 overcomes the performances of the remaining analysed models.

## Conclusion and future work

We design an approach focused on the detection of *covid-19* by analysing CT medical images. In particular, we consider a transfer learning model developed by authors, aimed to label a CT as *covid-19*, for patients affected with the *covid-19*; *other*, for patients affected with a pulmonary disease different from the *covid-19* one; and *healthy*, for patients with no pulmonary disease. As additional contribution, the proposed approach is aimed to highlight on the patient CT the areas symptomatic of the infection provided by *covid-19*: this represents an important characteristic because this representation can provide interesting and useful information to pathologist and radiologist. We think that the proposed method can be considered for a rapid screening and for an immediate diagnosis and visualisation of the lung areas affected by the *covid-19* infection.

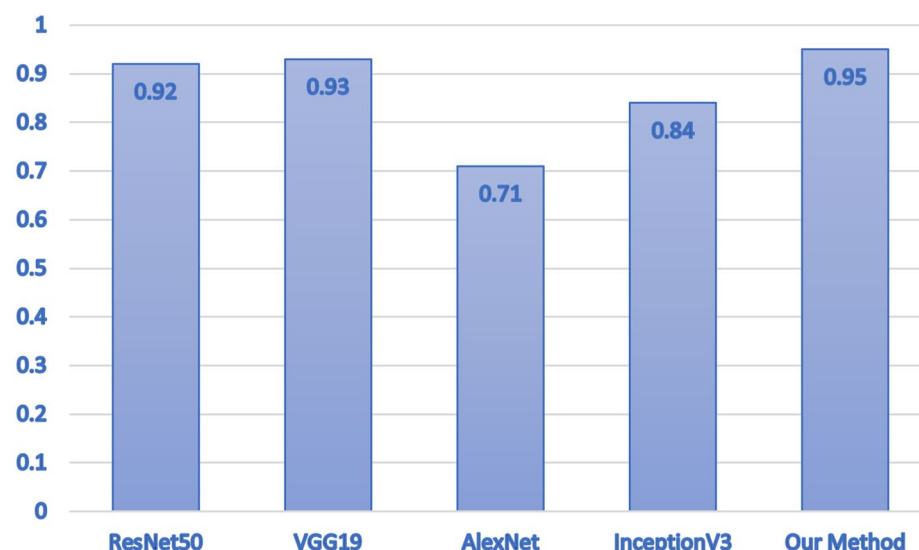
As a matter of fact, the GRAD-CAM was applied to other contexts, for instance cybersecurity<sup>53</sup> for this reason, we are confident that the proposed method can be exploited for explainable classification tasks in other contexts.

Furthermore, we will explore if model checking can be considered to increase the novel coronavirus detection accuracy obtained by the proposed method: as a matter of fact, model checking already demonstrated their effectiveness in medical context as, for instance, the detection of prostate cancer Gleason score from computed tomography images<sup>54</sup>.

Model	Sensitivity	Specificity	F-measure	Accuracy
ResNet50	0.89	0.91	0.82	0.92
VGG19	0.90	0.92	0.83	0.93
AlexNet	0.72	0.69	0.69	0.71
InceptionV3	0.84	0.82	0.81	0.84
Our method	0.95	0.95	0.95	0.95

**Table 3.** Experimental comparison.





**Figure 4.** The comparison in terms of accuracy obtained by the state-of-the-art models and the proposed deep learning model.

# Data availability

The dataset used and analysed during the current study is available from the corresponding author on reasonable request.

Received: 1 February 2022; Accepted: 5 January 2023

Published online: 10 January 2023

# References

1. Struyf, T. *et al.* Signs and symptoms to determine if a patient presenting in primary care or hospital outpatient settings has COVID-19. *Cochrane Database Syst. Rev.* **5**, 5–11 (2022).
2. Brunese, L., Martinelli, F., Mercaldo, F. & Santone, A. Machine learning for coronavirus COVID-19 detection from chest x-rays. *Procedia Comput. Sci.* **176**, 2212–2221 (2020).
3. Jeyanathan, M., Afkhami, S., Smaili, F., Miller, M.S., Lichty, B. D. & Xing, Z. Immunological considerations for COVID-19 vaccine strategies. *Nat. Rev. Immunol.* **20**, 1–18, (2020).
4. Brunese, L., Mercaldo, F., Reginelli, A., & Santone, A. Explainable deep learning for pulmonary disease and coronavirus COVID-19 detection from x-rays. *Comput. Methods Programs Biomed.* **196**, 105608, (2020).
5. Le, T. T. *et al.* The COVID-19 vaccine development landscape. *Nat. Rev. Drug Discov.* **19**(5), 305–306 (2020).
6. Li, Q. *et al.* Early transmission dynamics in Wuhan, China, of novel coronavirus-infected pneumonia. *New Engl. J. Med.* **382**, 1199–1207 (2020).
7. Gu, J., Han, B. & Wang, J. COVID-19: Gastrointestinal manifestations and potential fecal-oral transmission. *Gastroenterology* **158**(6), 1518–1519 (2020).
8. Roques, L., Klein, E. K., Papaix, J., Sar, A. & Soubeyrand, S. Using early data to estimate the actual infection fatality ratio from COVID-19 in France. *Biology* **9**(5), 97 (2020).
9. Covid, T.C., Team, R. Severe outcomes among patients with coronavirus disease 2019 (COVID-19)-United States. *MMWR Morb. Mortal. Wkly. Rep.* **69**(12), 343–346 (2020).
10. Wang, Y., Wang, Y., Chen, Y. & Qin, Q. Unique epidemiological and clinical features of the emerging 2019 novel coronavirus pneumonia (COVID-19) implicate special control measures. *J. Med. Virol.* **92**(6), 568–576 (2020).
11. Holmes, K. V. SARS-associated coronavirus. *N. Engl. J. Med.* **348**(20), 1948–1951 (2003).
12. van der Hoek, L. *et al.* Identification of a new human coronavirus. *Nat. Med.* **10**(4), 368–373 (2004).
13. Abroug, F. *et al.* Family cluster of middle east respiratory syndrome coronavirus infections, Tunisia, 2013. *Emerg. Infect. Dis.* **20**(9), 1527 (2014).
14. Jung, S.-M. *et al.* Real-time estimation of the risk of death from novel coronavirus (COVID-19) infection: Inference using exported cases. *J. Clin. Med.* **9**(2), 523 (2020).
15. Livingston, E. & Bucher, K. Coronavirus disease 2019 (COVID-19) in Italy. *Jama* **323**, 1335 (2020).
16. Wang, C., Horby, P. W., Hayden, F. G. & Gao, G. F. A novel coronavirus outbreak of global health concern. *Lancet* **395**(10223), 470–473 (2020).
17. Huang, C. *et al.* Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet* **395**(10223), 497–506 (2020).
18. Li, L. *et al.* Artificial intelligence distinguishes COVID-19 from community acquired pneumonia on chest CT. *Radiology* **296**, 200905, (2020).
19. Long, C. *et al.* Diagnosis of the coronavirus disease (COVID-19): RRT-PCR or CT? *Eur. J. Radiol.* **126**, 108961, (2020).
20. Sellers, S. A. *et al.* Burden of respiratory viral infection in persons with human immunodeficiency virus. *Influenza Other Respir. Viruses* **14**, 465–469 (2020).
21. Ai, T. *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**, 200642 (2019).
22. Fang, Y., Zhang, H., Xie, J., Lin, M., Ying, L., Pang, P., & Ji, W. Sensitivity of chest CT for COVID-19: Comparison to RT-PCR. *Radiology* **296**, 200432 (2020).

23. Alakwaa, W., Nassef, M. & Badr, A. Lung cancer detection and classification with 3D convolutional neural network (3D-CNN). *Lung Cancer* **8**(8), 409–417 (2017).
24. Bhatia, S., Sinha, Y. & Goel, L. Lung cancer detection: A deep learning approach. In *Soft Computing for Problem Solving* 699–705 (Springer, 2019).
25. Brunese, L., Mercaldo, F., Reginelli, A., & Santone, A. Neural networks for lung cancer detection through radiomic features. In *2019 International Joint Conference on Neural Networks (IJCNN)*, 1–10 (IEEE, 2019).
26. Bulten, W. *et al.* Automated deep-learning system for Gleason grading of prostate cancer using biopsies: A diagnostic study. *Lancet Oncol.* **21**, 233–241 (2020).
27. Puderbach, M. *et al.* Assessment of morphological MRI for pulmonary changes in cystic fibrosis (CF) patients: Comparison to thin-section CT and chest X-ray. *Investig. Radiol.* **42**(10), 715–724 (2007).
28. Rohde, M. *et al.* Head-to-head comparison of chest x-ray/head and neck MRI, chest CT/head and neck MRI, and 18F-FDG PET/CT for detection of distant metastases and synchronous cancer in oral, pharyngeal, and laryngeal cancer. *J. Nucl. Med.* **58**(12), 1919–1924 (2017).
29. Schaefer, O. & Langer, M. Detection of recurrent rectal cancer with CT, MRI and PET/CT. *Eur. Radiol.* **17**(8), 2044–2054 (2007).
30. Khan, S., Rahmani, H., Shah, S. A. A. & Bennamoun, M. A guide to convolutional neural networks for computer vision. *Synth. Lect. Comput. Vis.* **8**(1), 1–207 (2018).
31. LeCun, Y., Bottou, L., Bengio, Y. & Haffner, P. Gradient-based learning applied to document recognition. *Proc. IEEE* **86**(11), 2278–2324 (1998).
32. de Lima Hedayioglu, F., Coimbra, M. T., & da Silva Mattos, S. A survey of audio processing algorithms for digital stethoscopes. In *HEALTHINF* 425–429, (2009).
33. Simonyan, K. & Zisserman, A. Very deep convolutional networks for large-scale image recognition. arXiv preprint [arXiv:1409.1556](https://arxiv.org/abs/1409.1556), (2014).
34. Selvaraju, R. R. *et al.* Grad-cam: Visual explanations from deep networks via gradient-based localization. In *Proceedings of the IEEE International Conference on Computer Vision* 618–626, (2017).
35. Chen, J., Wu, L., Zhang, J. *et al.* Deep learning-based model for detecting 2019 novel coronavirus pneumonia on high-resolution computed tomography. *Sci Rep* **10**, 19196 <https://doi.org/10.1038/s41598-020-76282-0> (2020).
36. Wang, S. *et al.* A deep learning algorithm using CT images to screen for corona virus disease (COVID-19). *Eur. Radiol.* **31**(8), 6096–6104 (2021).
37. Xu, X. *et al.* A deep learning system to screen novel coronavirus disease 2019 pneumonia. *Engineering* **6**(10), 1122–1129 (2020).
38. Beck, B. R., Shin, B., Choi, Y., Park, S., & Kang, K. Predicting commercially available antiviral drugs that may act on the novel coronavirus (2019-nCoV), Wuhan, China through a drug-target interaction deep learning model. *bioRxiv* **18**, 784–790 (2020).
39. Zhang, H. *et al.* Deep learning based drug screening for novel coronavirus 2019-nCoV. (2020).
40. Apostolopoulos, I. D., & Mpesiana, T. A. Covid-19: Automatic detection from x-ray images utilizing transfer learning with convolutional neural networks. *Phys. Eng. Sci. Med.* **43**, 1 (2020).
41. Wang, X. *et al.* A weakly-supervised framework for COVID-19 classification and lesion localization from chest CT. In *IEEE Transactions on Medical Imaging*, **39**(8), 2615–2625 <https://doi.org/10.1109/TMI.2020.2995965> (2020).
42. Narin, A., Kaya, C., & Pamuk, Z. Automatic detection of coronavirus disease (COVID-19) using x-ray images and deep convolutional neural networks. arXiv preprint [arXiv:2003.10849](https://arxiv.org/abs/2003.10849), (2020).
43. Song, Y. *et al.* Deep learning enables accurate diagnosis of novel coronavirus (COVID-19) with CT images. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **18**(6), 2775–2780 (2021).
44. Ozturk, T. *et al.* Automated detection of COVID-19 cases using deep neural networks with x-ray images. *Comput. Biol. Med.* **121**, 103792, (2020).
45. Ardakani, A. A., Kanafi, A. R., Acharya, U. R., Khadem, N. & Mohammadi, A. Application of deep learning technique to manage COVID-19 in routine clinical practice using CT images: Results of 10 convolutional neural networks. *Comput. Biol. Med.* **121**, 103795 (2020).
46. Glowacz, A. & Glowacz, Z. Recognition of images of finger skin with application of histogram, image filtration and K-NN classifier. *Biocybern. Biomed. Eng.* **36**(1), 95–101 (2016).
47. Piekarski, M., Jaworek-Korjakowska, J., Wawrzyniak, A. I. & Gorgon, M. Convolutional neural network architecture for beam instabilities identification in synchrotron radiation systems as an anomaly detection problem. *Measurement* **165**, 108116 (2020).
48. Wang, L., Lin, Z. Q. & Wong, A. Covid-net: A tailored deep convolutional neural network design for detection of covid-19 cases from chest x-ray images. *Sci. Rep.* **10**(1), 1–12 (2020).
49. Sethy, P. K. & Behera, S. K. Detection of coronavirus disease (COVID-19) based on deep features. (2020).
50. Hemdan, E. E.-D., Shouman, M. A., & Karar, M. E. Covidx-net: A framework of deep learning classifiers to diagnose COVID-19 in x-ray images. arXiv preprint [arXiv:2003.11055](https://arxiv.org/abs/2003.11055), (2020).
51. Butt, C., Gill, J., Chun, D. & Babu, B. A. Deep learning system to screen coronavirus disease, pneumonia. *Appl. Intell.* **2020**, 1 (2019).
52. Peláez, E., Serrano, R., Murillo, G. & Cárdenas, W. A comparison of deep learning models for detecting covid-19 in chest x-ray images. *Ifac-papersonline* **54**(15), 358–363 (2021).
53. Iadarola, G., Martinelli, F., Mercaldo, F. & Santone, A. Towards an interpretable deep learning model for mobile malware detection and family identification. *Comput. Security* **105**, 102198 (2021).
54. Brunese, L., Mercaldo, F., Reginelli, A. & Santone, A. Prostate Gleason score detection and cancer treatment through real-time formal verification. *IEEE Access* **7**, 186236–186246 (2019).

## Acknowledgements

This work has been partially supported by MUR-REASONING: foRmal mEthods for computAtional analySis for diagnOsis and progNosis in imagING-PRIN.

## Author contributions

F.M. developed and evaluated the proposed method and wrote the paper. M.P.B. and A.R. gathered the medical image dataset. A.S. and L.B. wrote the paper.

## Competing interests

All authors confirm that there are not potential conflicts of interest include employment, consultancies, stock ownership, honoraria, paid expert testimony, patent applications/registrations, and grants or other funding.

## Additional information

**Correspondence** and requests for materials should be addressed to F.M.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

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



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023