



Explainable Deep Learning for Pulmonary Disease and Coronavirus COVID-19 Detection from X-rays

Luca Brunese^a, Francesco Mercaldo^{a,b,*}, Alfonso Reginelli^c, Antonella Santone^d

^a Department of Medicine and Health Sciences "Vincenzo Tiberio", University of Molise, Campobasso, Italy

^b Institute for Informatics and Telematics, National Research Council of Italy (CNR), Pisa, Italy

^c Department of Precision Medicine, University of Campania "Luigi Vanvitelli", Napoli, Italy

^d Department of Biosciences and Territory, University of Molise, Pesche (IS), Italy

ARTICLE INFO

Article history:

Received 9 April 2020

Accepted 9 June 2020

Keywords:

Coronavirus

COVID-19

Chest

Deep learning

Transfer learning

Artificial intelligence

ABSTRACT

Background and Objective: Coronavirus disease (COVID-19) is an infectious disease caused by a new virus never identified before in humans. This virus causes respiratory disease (for instance, flu) with symptoms such as cough, fever and, in severe cases, pneumonia. The test to detect the presence of this virus in humans is performed on sputum or blood samples and the outcome is generally available within a few hours or, at most, days. Analysing biomedical imaging the patient shows signs of pneumonia. In this paper, with the aim of providing a fully automatic and faster diagnosis, we propose the adoption of deep learning for COVID-19 detection from X-rays.

Method: In particular, we propose an approach composed by three phases: the first one to detect if in a chest X-ray there is the presence of a pneumonia. The second one to discern between COVID-19 and pneumonia. The last step is aimed to localise the areas in the X-ray symptomatic of the COVID-19 presence.

Results and Conclusion: Experimental analysis on 6,523 chest X-rays belonging to different institutions demonstrated the effectiveness of the proposed approach, with an average time for COVID-19 detection of approximately 2.5 seconds and an average accuracy equal to 0.97.

© 2020 Elsevier B.V. All rights reserved.

1. Introduction

Coronaviruses represent an extended family of respiratory viruses able to cause diseases, for instance the cold to most important respiratory syndromes such as MERS (Middle East respiratory syndrome) and SARS (Severe acute respiratory syndrome) [1]. They are called in this way due to the crown-shaped tips on their surface [2].

These viruses are really common in a plethora of animal species but in several cases they are able to evolve and to infect humans and, consequently, rapidly and easily spread among the population [1,3]. This is what happened at the end of 2019, when the 2019 coronavirus (COVID-19, acronym of *CO*rona*V*irus *D*isease 19)¹, a new kind of coronavirus never previously detected in humans, appeared. The first cases were found during the pandemic of 2019–2020 [4], which probably started at the end of December 2019 in

the city of Wuhan [5], the capital of the Chinese province of Hubei, and subsequently spread to several countries around the world. As confirmation of the rapid spread of the disease in January 28 2020 there were more than 4,600 COVID-19 confirmed cases in a plethora of countries and 106 deaths while on February 15 (in a time-window less than a month) these data had already risen to 49,053 cases and 1,381 deaths². As of January 23 2020, the city of Wuhan has been quarantined with the suspension of all public transports into and out of the city. These measures were extended the following day to the neighboring cities of Huanggang, Ezhou, Chibi, Jingzhou and Zhijiang. Moreover, further limitations and controls have been adopted in many areas of the world, also in Europe where several cases have also been recorded. The country most affected in Europe in the month of March 2020 is Italy [6]. In fact, in April 5 2020, approximately 15.9 thousand deaths were reported by the Italian authorities, of which 8.9 thousand in the region of Lombardia, 2.1 thousand in Emilia-Romagna the region, and 1.2 thousand in the region of Piedmont, the regions

* Corresponding author.

¹ <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports>.

² <https://www.worldometers.info/coronavirus/>.

mostly hit. In March 19 2020, the number of coronavirus deaths in Italy exceeded that registered in China³. From a clinical point of view the COVID-19 infection caused clusters of fatal pneumonia with clinical presentation greatly resembling SARS-CoV. In fact, patients experience flu-like symptoms such as fever, dry cough, tiredness, difficulty breathing. In more severe cases, often found in subjects already burdened by previous diseases, pneumonia develops, acute renal failure, up to even death [4], but this new coronavirus presents several unique features [7].

However, many patients are infected with COVID-19 but are asymptomatic. In other words, there is the possibility that people infected by COVID-19 but without any symptoms could transmit the disease [8]. As a matter of fact in Italy, in the country of Vo' Euganeo, 50 km west of Venice, was closed off by authorities in mid-February: it was discovered that of all the inhabitants of the country who underwent the pharyngeal swab, 50-75% of the inhabitants were swab positive, that is, infected with the virus, but did not present any symptoms [9]. Currently the only way to find out about COVID-19 infection is to undergo the swab and then analyze the biological material taken from the patient through the polymerase chain reaction [10].

The problem is that the swab is however carried out only to those who show symptoms therefore currently people infected with COVID-19 but asymptomatic cannot be discovered unless there are particular cases [11].

While the diagnosis is confirmed using the polymerase chain reaction, infected patients with pneumonia may present on chest X-ray and computed tomography (CT) images with a pattern that is only moderately characteristic for the human eye as demonstrated by researchers in [12]. The rate of transmission of COVID-19 depends on the capacity to reliably identify infected patients with a low rate of false negatives. In addition, a low rate of false positives is required to avoid further increasing the burden on the health-care system by unnecessarily exposing patients to quarantine if that is not required. Along with proper infection control, it is evident that timely detection of the disease would enable the implementation of all the supportive care required by patients affected by COVID-19.

In late January 2020, Chinese researchers discussed the clinical and paraclinical features COVID-19 specific [13]. They reported that patients present abnormalities in chest CT images with most having bilateral involvement [13]. Bilateral multiple lobular and subsegmental areas of consolidation constitute the typical findings in chest CT images of intensive care unit (ICU) patients on admission [13]. In comparison, non-ICU patients show bilateral ground-glass opacity and subsegmental areas of consolidation in their chest CT images [13]. In these patients, later chest CT images display bilateral ground-glass opacity with resolved consolidation [13].

Biomedical imaging (radiography or computed tomography of the chest) shows signs of pneumonia. Subsequently, the World Health Organization published several additional diagnostic protocols⁴. Diagnosis is performed by performing a real-time reverse polymerase chain reaction (rRT-PCR) test on biological samples taken from the patient. The test can be performed on sputum or blood samples [14,15]: results are generally available within a few hours or, at most, days [16,17].

As stated in [18,19] COVID-19 is possibly better diagnosed using radiological imaging, for this reason, in this paper, we evaluate the possibility to detect the COVID-19 disease directly from medical images; X-Rays have been used.

Several studies discuss about the detection of pulmonary disease by analysing medical images by exploiting artificial intelligence. Artificial intelligence represents an emerging field devoted to create models from data and in last years its adoption is emerging in the development of methods for assist experts in the medical images interpretation.

In particular, the transfer learning, a deep learning method where a model developed for a task is reused as the starting point for a model on a second task, is emerging.

It represents a popular approach where pre-trained models are used as the starting point on computer vision and natural language processing tasks given the vast compute and time resources required to develop neural network models on these problems and from the huge jumps in skill that they provide on related problems.

Recent efforts have shown promise in improving detection in several medical field, for instance from lung cancer detection [20] to prostate cancer grading [21].

Starting from these considerations, in this paper, we propose the adoption of deep learning to detect whether there is the COVID-19 presence in X-ray images by exploiting transfer learning. As additional contribution, we show the network activation layers i.e., the areas of the chest X-ray that the model considered to generate the prediction, to provide explainability about the prediction. This can represent a suggestion for the radiologist to immediately localise the X-ray areas that can be of interest.

The paper proceeds as follows: Section 2 describes the proposed method, the experimental analysis is presented in Section 3; in Section 4 an overview of the literature related to the pulmonary disease detection with particular regard to COVID-19 disease is provided, and, finally, in the last section, conclusion and future research plan is presented.

2. The Method

In this paper, we propose a three-fold method aimed to: (i) detect if a chest X-ray is related to an healthy patient or to a patient with generic pulmonary disease; (ii) to discriminate between generic pulmonary diseases and COVID-19 and, once detected the COVID-19 disease, (iii) to highlight the areas in the chest X-ray symptomatic of the COVID-19 disease.

In detail, considering as input a chest X-ray the proposed method is aimed to:

- detect if the chest X-ray is related to an healthy patient or to a patient affected by a pulmonary disease;
- detect, using the second model, if the pulmonary disease is pneumonia or COVID-19, when the chest X-ray is marked with the pulmonary disease label;
- provide, when the previous model marked the chest X-ray as COVID-19, a visualisation of the chest X-ray where the areas potentially interested by the COVID-19 disease are highlighted.

To build the (first and the second) model, we consider a deep learning network, based on the VGG-16 (i.e., *Visual Geometry Group*) model [22] by exploiting transfer learning.

Transfer learning typically is related to a process where a model trained on one problem is exploited with the aim to predict labels related to a second problem [23]. The main benefit in the adoption of transfer learning is represented by the decreasing time for the training task for a neural network model. Moreover, it can result in lower generalization error. Clearly, the first problem should be related to the second problem to solve: in this case, we transfer the knowledge of a model trained for generic image detection to solve the problem of pulmonary disease identification.

VGG-16 is a convolutional neural network with 16 layers. It is possible to load a pre-trained version of the the VGG-16 network

³ <https://www.statista.com/statistics/1099389/coronavirus-deaths-by-region-in-italy/>.

⁴ <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/technical-guidance/laboratory-guidance>.

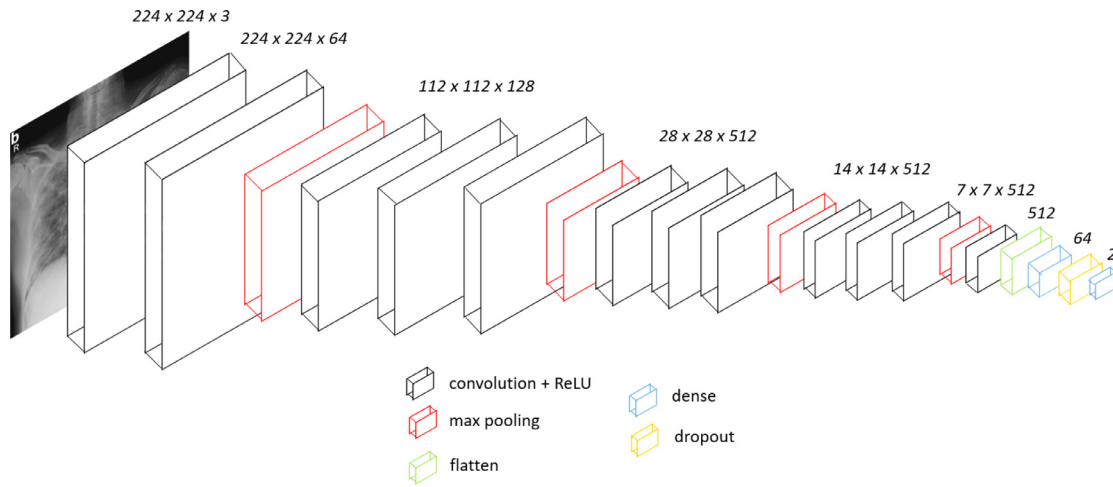


Fig. 1. The network.

trained on the ImageNet database [24], which is a data-set of over 14 million images: the pre-trained network can classify images into 1,000 different classes, for instance as cat, pen and other similar objects. This model obtains 92.7% top-5 test accuracy in ImageNet database. VGG-16 was trained for weeks and was using NVIDIA Titan Black GPU's.

We apply transfer learning employing fine-tuning. We initialize the VGGNet model and we set it up for fine-tuning: basically, we instantiate the VGG16 network with weights pre-trained on ImageNet, leaving off the fully connected layer head. From there, we build a new fully-connected layer head consisting of following layers AveragePooling2D, Flatten, Dense, Dropout and a last Dense with the “softmax” activation to predict the classes. We append it on top of VGG16. We then freeze the convolutional weights of VGG16 such that only the fully connected layer head will be trained: this completes our fine-tuning setup.

Fig. 1 shows the exploited architecture we modified for our purposes.

To ensure that the proposed models generalize, we consider data augmentation by setting the random image rotation setting to 15 degrees clockwise or counterclockwise.

The first layer is represented by convolutional of a size fixed to 224 x 224 RGB image, for this reason chest X-ray are resized to this dimension. The chest X-ray 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 2x2 pixel window.

All the considered hidden layers consider a rectification non-linearity. More details on the VGG-16 model architecture can be found in reference [22]

In detail to the VGG-16 architecture we added the following layers (as shown by the network depicted in Fig. 1):

- **AveragePooling2D**: it performs average pooling operation. This layer involves computing the average for each patch of the feature map under analysis. This means that each 2x2 square of the feature map is down sampled to the average value;
- **Flatten**: the aim of this layer is to flat the input. It is a utility layer, it flats an input, for instance a rows x columns matrix, to a simple vector output of rows * columns shape. Flattening transforms a two-dimensional matrix of features into a vector that can be fed into a fully connected neural network classifier.

- **Dense**: is the regular deeply connected neural network layer. The purpose of the layer is to transform the data. It is most common and frequently used layer. In this case this layer reduces the vector of height 512 to a vector of 64 elements;
- **Dropout**: this layer basically works in the following way i.e., by randomly select neurons not considered in the training. The aim of this layer is to ameliorate generalization in fact, we are coercing the network to train the same high-level concept by exploiting different neurons. We chose to ignore the 50% of neurons. Clearly, we are aware that typically exploiting this layer this we can reach worse performances, but we want to generate a model less sensitive to data variations;
- **Dense**: the last dense layer is aimed to reduce the vector of height 64 to a vector of 2 elements (i.e., the two classes to predict).

Moreover, to provide explainability, we propose to visualize class activation maps, a technique usually considered for debugging deep neural networks. To do this, we resort to the Gradient-weighted Class Activation Mapping (Grad-CAM) algorithm [25].

In fact, as demonstrated by recent literature, while deep learning has facilitated unprecedented accuracy in image classification [26], object detection [27], and image segmentation [28], one of their biggest problems is model interpretability, a fundamental component in model understanding and model debugging currently limiting the application of these techniques in critical contexts (for instance, the medical one).

As a matter of fact, deep learning models are considered as “black box” methods, and several times researchers and practitioners do not have reasonable idea as to: (i) where the network is “looking” in the input image; (ii) which series of neurons are activated in the forward-pass during inference/prediction; (iii) how the network arrived at its final output.

The above issues can be summarized in the following main question: “how can we trust the decisions of a model if we are not able properly validate how it arrived there?” Clearly, it is crucial to answer this question, with particular regards to biomedical contexts.

For this reason, we exploit the Grad-CAM algorithm, to visually debug the models and properly understand where the proposed network is “looking” into chest X-ray images to predict the COVID-19 class.

In a nutshell, Grad-CAM uses the gradients of any target concept, flowing into the final convolutional layer to produce a coarse localization map highlighting the important regions in the image for predicting the concept.

Table 1
Chest X-ray labels.

Label	Number
ARDS	4
Chlamydia	2
COVID-19	250
Escherichia Coli	4
Klebsiella	1
Legionella	2
Pneumocystis	14
SARS	15
Streptococcus	16
Atelectasis	508
Cardiomegaly	125
Consolidation	160
Edema	93
Effusion	392
Emphysema	84
Fibrosis	56
Hernia	7
Infiltration	623
Mass	147
Pneumonia	500
Healthy	3,520

Using Grad-CAM, it is possible to visually validate where the network is looking where a chest X-ray is evaluated, verifying that it is indeed looking at the correct patterns in the image and activating around those patterns.

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

3. Experimental Analysis

In this section we present the experiment we performed to assess the performance of the proposed methods in terms of:

- (i) discrimination between a healthy and a chest X-ray related to pulmonary diseases;
- (ii) discrimination between a chest X-ray related to pneumonia and COVID-19;
- (iii) model explainability, by providing samples of chest X-ray highlighting the fundamental regions in the X-ray for COVID-19 prediction.

3.1. Data-set

To assess the proposed method, different data-sets belonging to multiple institutions are considered. All the involved data-sets are collections of chest X-rays. In detail, we consider three different x-ray datasets: the first one obtained from the “COVID-19 image data collection” [29] freely available⁵, the second one is the data-set⁶ considered by authors in the “Automated Detection of COVID-19 Cases Using Deep Neural Networks with X-ray Images” [30] research paper, while the third one was obtained from the “National Institutes of Health Chest X-Ray” [31] and it is freely available for research purposes⁷.

Combining the data-sets, we obtain chest X-rays related to several classes.

As shown from Table 1, in the study we consider a total of 6,523 chest X-rays: 250 related to patients afflicted by COVID-19,

2,753 related to patients with other pulmonary diseases and 3,520 related to healthy patients. All the diagnosis were confirmed by expert radiologists.

Fig. 2 shows an example of chest X-ray belonging to the data-set we analyse with the relative label.

As shown from Table 5, the medical images are obtained from different institutions: from China (where COVID-19 disease started to manifest) from Italy (the European area in which it was found to be most widespread), but also from Australia and USA.

3.2. Results

Below we provide the details about the chest X-rays we considered to build the two models.

The aim of the first model is the discrimination between healthy and pulmonary disease X-ray. In this case we consider with the *health* label the healthy 3,520 X-rays in Table 1 while, with the *disease* label we consider the 3,003 remaining X-rays with pulmonary diseases (COVID-19 included).

With regard to the second model, aimed to discriminate between COVID-19 and other pulmonary disease, we consider the 250 COVID-19 X-rays labelled by radiologists as *COVID-19* and the other 2,753 X-rays related to pulmonary disease with the *disease* label.

We split the data-set in three parts: training, testing and evaluation. Table 3 shows the number of X-ray considered for the training, testing and validation tasks. For the model building a cross-validation is considered.

Fig. 3 shows the confusion matrix for the testing of the model 1, aimed to discriminate between healthy and X-rays afflicted by pulmonary disease.

On the 2000 considered X-ray for the first model testing, 18 disease X-rays were marked as *Health* by the first model, while 43 health X-rays were marked with the *Disease* label.

Fig. 4 shows the confusion matrix for the testing of the model 2, aimed to discriminate between COVID-19 and X-rays afflicted by other pulmonary disease.

On the 2000 considered X-rays for the second model testing, 9 disease X-rays were marked as *COVID-19* by the second model, while 4 *COVID-19* X-rays were marked with the *Disease* label.

Fig. 5 shows the confusion matrix for the validation of the model 1, aimed to discriminate between healthy and X-rays afflicted by pulmonary disease.

On the 2,523 considered X-rays for the first model evaluation, 19 disease X-rays were marked as *Health* by the first model, while 33 *Health* X-rays were marked with the *Disease* label.

Fig. 6 shows the confusion matrix for the evaluation of the model 2, aimed to discriminate between COVID-19 and X-rays afflicted by other pulmonary disease.

On the 803 considered X-rays for the second model evaluation, 3 *COVID-19* X-rays were marked as pulmonary *Disease* by the second model, while 7 *Disease* X-rays were marked with the *COVID-19* label.

Following four metrics are considered to evaluate the performance of the classifiers: Sensitivity, Specificity, F-Measure and Accuracy.

The sensitivity of a test is the proportion of people who test positive among all those who actually have the disease and it is defined as:

$$\text{Sensitivity} = \frac{tp}{tp + fn}$$

where *tp* indicates the number of true positives and *fn* indicates the number of false negatives

The specificity of a test is the proportion of people who test negative among all those who actually do not have that disease

⁵ <https://github.com/ieee8023/covid-chestxray-dataset>.

⁶ <https://github.com/muhammedtaloo/COVID-19>.

⁷ <https://www.kaggle.com/nih-chest-xrays/sample>.

Table 2
Institutions involved the study.

Hospital	Country	X-ray machine brand
Ospedale Santo Spirito, Rome	Italy	Siemens Healthcare
Riccione	Italy	Philips Healthcare
Myongji Hospital, Goyang	Korea	Hitachi Medical Corporation
Jönköping	Sweden	Siemens Healthcare
Melbourne	Australia	GE Healthcare
Royal Brisbane and Women's Hospital, Brisbane	Australia	GE Healthcare
Sichuan Provincial People's Hospital, Chengdu	China	Shimadzu Corporation
Tongji Medical College, Wuhan, Hubei Province	China	Fujifilm Holdings
Taoyuan General Hospital, Taoyuan	Taiwan	Fujifilm Holdings
Snohomish County, Washington	USA	GE Healthcare

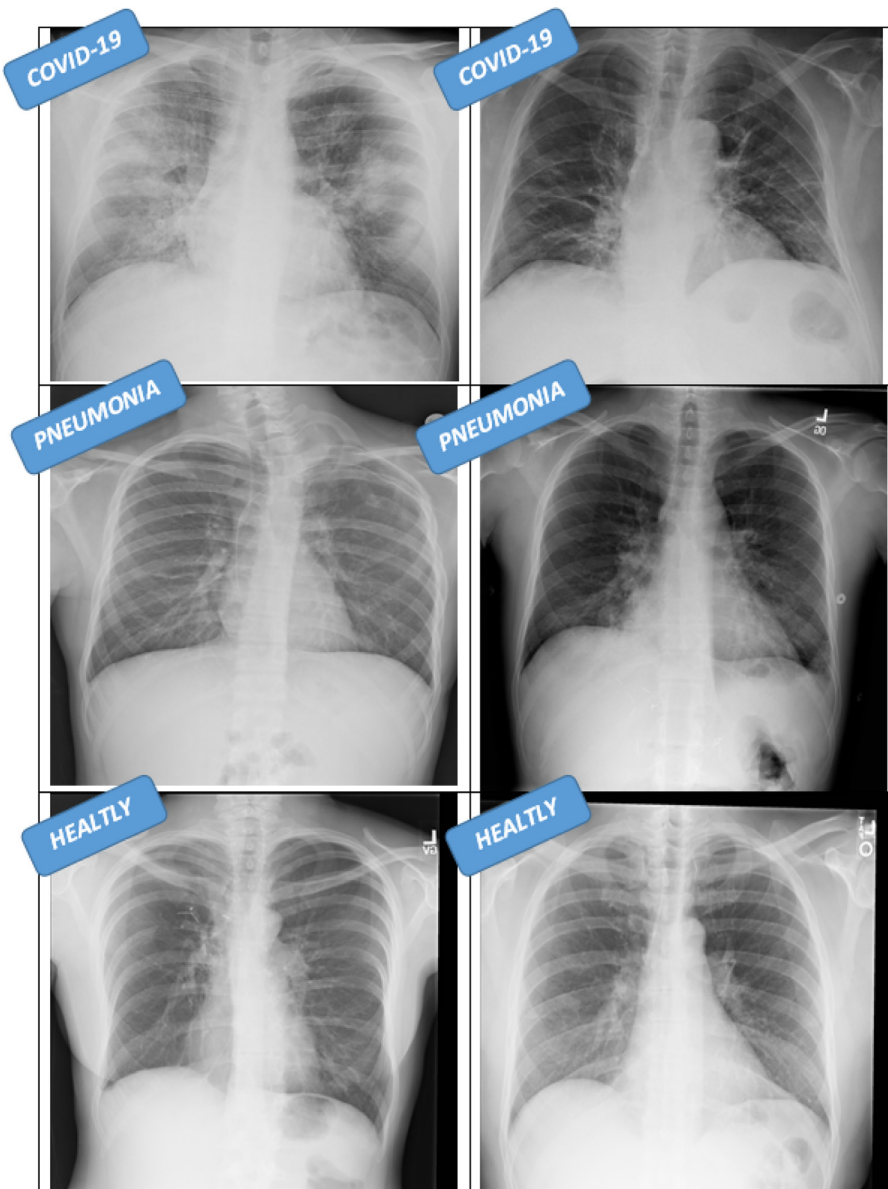


Fig. 2. Example of chest X-rays with the related label.

and it is defined as:

$$Specificity = \frac{tn}{tn + fp}$$

where tn indicates the number of true negatives.

The F-Measure is a measure of a test's accuracy. This score can be interpreted as a weighted average of the precision and recall:

$$F - Measure = 2 * \frac{Sensitivity * Specificity}{Sensitivity + Specificity}$$

The accuracy of a measurement system is the degree of closeness of measurements of a quantity to that quantity's true value:

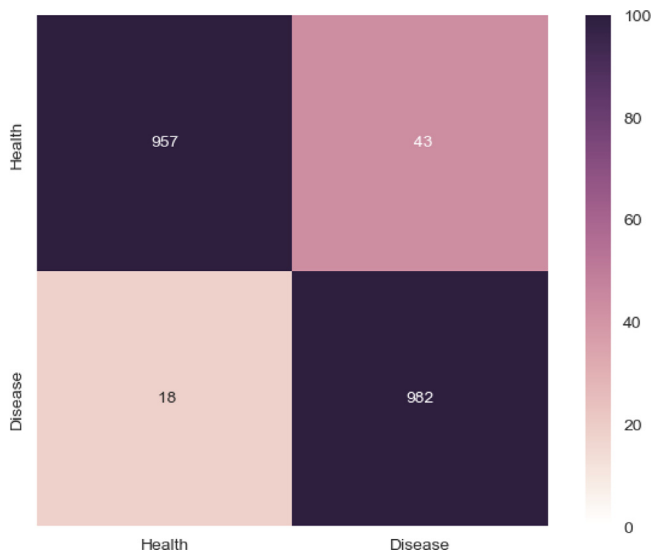


Fig. 3. Model 1 Testing Confusion Matrix.



Fig. 4. Model 2 Testing Confusion Matrix.

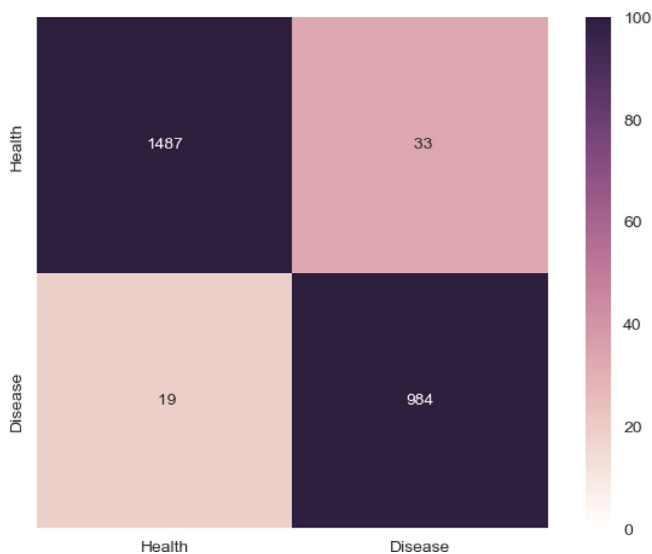


Fig. 5. Model 1 Validation Confusion Matrix.

Table 3
Training, testing and evaluation sets.

Model	Label	Training	Testing	Evaluation
1	Health	1,000	1,000	1,520
	Disease	1,000	1,000	1,003
2	Disease	1,000	1,000	753
	COVID-19	1,000	100	50

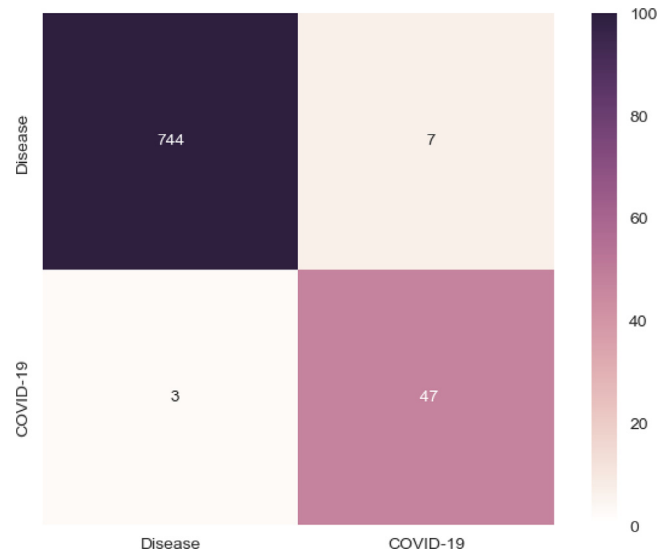


Fig. 6. Model 2 Validation Confusion Matrix.

Table 4
Results.

Model	Sensitivity	Specificity	F-Measure	Accuracy	Time
1	0.96	0.98	0.94	0.96	2.569s
2	0.87	0.94	0.89	0.98	2.498s

it is the fraction of the classifications that are correct and it is computed as the sum of true positives and negatives divided all the evaluated instances:

$$Accuracy = \frac{tp + tn}{tp + fn + fp + tn}$$

where fn indicates the number of false negatives.

Table 4 shows the classification results for both the models. We indicate in the *Model* column the first model with the *Disease* label, while the COVID-19 detection model are indicated with the COVID-19 label. In the *time* column we indicate the average time to obtain a prediction respectively from the *Disease* and the *COVID-19* models.

As shows by results highlighted by Table 4, the model for discriminating between healthy and generic pulmonary diseases obtains a sensitivity equal to 0.96 and a specificity of 0.98.

The model for discerning between generic pulmonary diseases and COVID-19 exhibits a sensitivity of 0.87 and a specificity equal to 0.94.

With regard to the accuracy, the first model reaches an accuracy equal to 0.96, while the second model obtains a value of 0.98.

The proposed method is able to output (in average) in 2.569 seconds whether the chest X-ray is related to a healthy patient or to a patient with a pulmonary disease. Moreover, the chest X-ray is marked with the COVID-19 label in 2.498 seconds.

In Figs. 7 and 8, we report the performances of the proposed network for the disease and for the COVID-19 model in terms of accuracy and loss.

Table 5
State-of-the-art comparison.

Method	Images	COVID-19	Other	Healthy	Total	Accuracy
Apostolopoulos et al. [42]	X-ray	224	700	504	1428	0.93
Wang and Wong [50]	X-ray	53	5526	8066	13645	0.92
Sethy et al. [44]	X-ray	25	n.a.	25	50	0.95
Hemdan et al. [41]	X-ray	25	n.a.	25	50	0.90
Narin et al. [43]	X-ray	50	n.a.	50	100	0.98
Song et al. [45]	CT	777	n.a.	708	1485	0.86
Wang et al. [40]	CT	195	n.a.	258	453	0.82
Zheng et al. [46]	CT	313	n.a.	229	542	0.90
Xu et al. [51]	X-ray	219	224	175	618	0.86
Ozturk et al. [30]	X-ray	250	500	1000	1750	0.92
Ardakani et al. [47]	CT	510	510	n.a.	1020	0.99
Li et al. [48]	CT	1,296	1,735	1,325	4,356	0.96
Butt et al. [49]	CT	357	1,353	n.a.	1,710	0.86
<i>Our Method</i>	X-ray	250	2,753	3,520	6,523	0.97

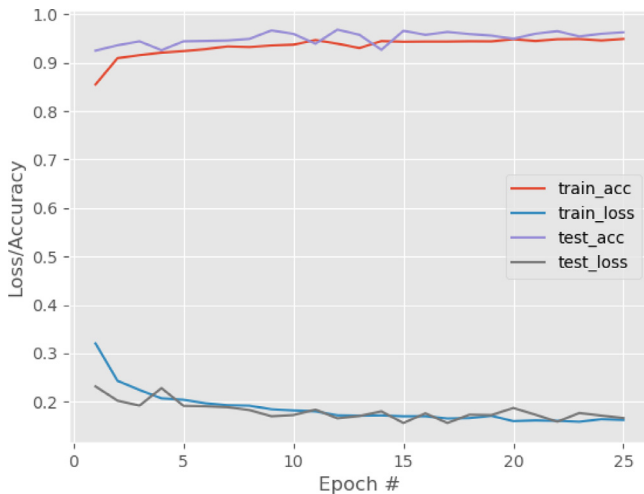


Fig. 7. Fist model performances.

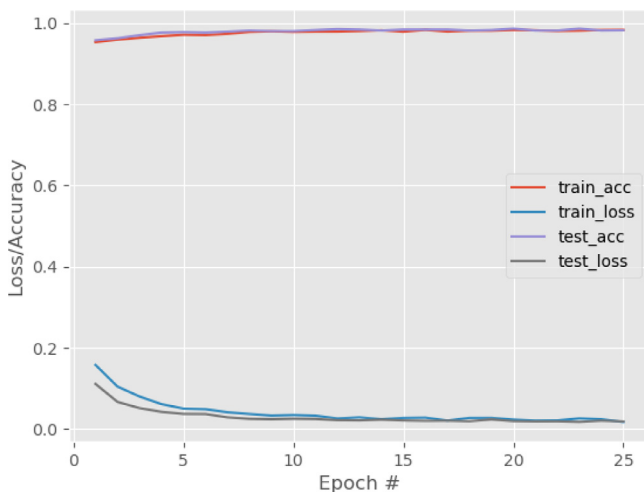


Fig. 8. Second model performances.

The Loss represents a quantitative measure of how much the predictions differ from the actual output (i.e., the label). Loss is inversely proportional to the correctness of the model.

The loss is calculated on training and validation and its interpretation is how well the model is doing for these two sets: it is a summation of the errors made for each example in training or validation sets.

From the Accuracy and Loss definitions, it is expected that Accuracy and Loss should be inversely proportional: for high values of accuracy, low loss values are expected (and the opposite). Furthermore, considering that the weights and bias are initially random selected, the accuracy trend should start by exhibiting low values (and high loss value, symptomatic that the network is performing wrong predictions), but whether the network during the several “epochs” (i.e., one forward pass and one backward pass of all the training examples) is able to learn (i.e., it is able to solve the driver prediction problem), the accuracy should start to exhibit higher values in the next iterations (and consequently the loss should exhibit low values). The epoch is a parameter chosen by the network designer, usually the number of epochs chosen is such that the loss is at least and it does not get worse in the immediately succeeding epochs and, consequently, the accuracy value reached is the maximum and in the immediately succeeding epochs is not improving, symptomatic that the network has reached the stability and that further epochs would not improve performances. We set the number of epochs equal to 25, because both the models reach the stability with a number of epochs less or equal to 25.

In particular the performance plots consider:

- *train_loss*: the loss for the training;
- *val_loss*: the loss for the validation;
- *train_acc*: the accuracy for the training;
- *val_acc*: the accuracy for the validation.

Fig. 7 shows the performance plots for the 25 epochs of the first model.

As shows by Fig. 7, initially the *train_acc* shows values slightly higher to 0.8 of accuracy, but starting from the second epoch this value is increasing. The *value_acc* value roughly remains

With regards to the loss, it shows for both the training that for the validation a trend opposed to the ones of the training and validation accuracy.

Fig. 8 shows the performance plots for the models discriminating between COVID-19 and generic pulmonary diseases.

As shows from Fig. 8, the performance trends are more linear if compared to the one of the first model (shown in Fig. 7): symptomatic that there is an evident difference between COVID-19 and the other pulmonary diseases.

To understand which areas of the chest X-ray was highlighted by the network for the COVID-19 detection, we shows examples of GRAD-CAM activation maps in Fig. 9. We show the chest X-ray used to input the model, the activation maps generated by the GRAD-CAM algorithm and the overlapping of the chest X-ray and the activation maps. In particular, we think that this last visualization can be useful for the radiologist and pathologists for localise the X-ray areas to investigate. The areas considered by the model

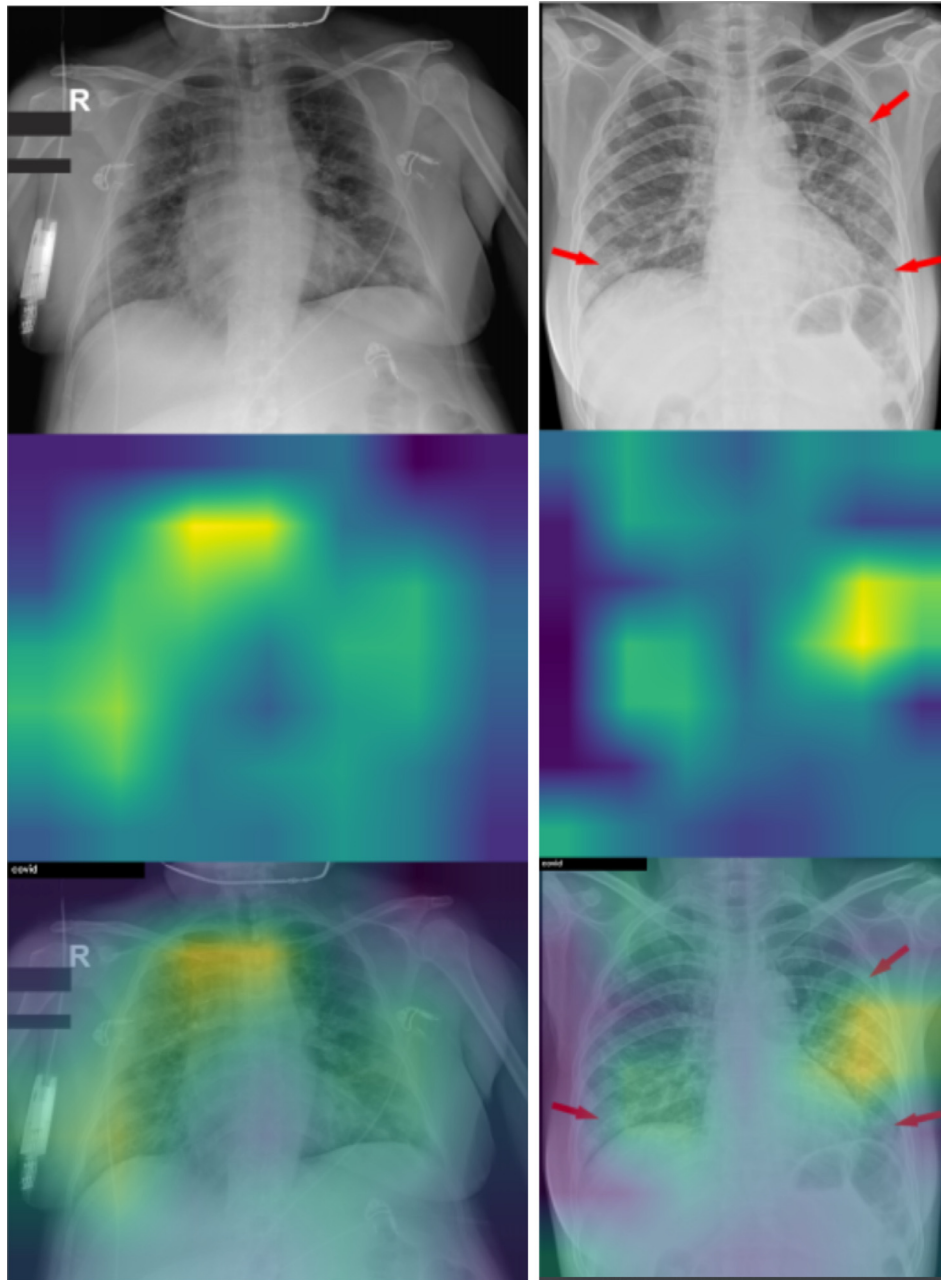


Fig. 9. Examples of COVID-19 model activation maps.

to output the prediction are in yellow with a more marked yellow where the probability of the predicted label is higher.

In particular, in all the X-ray shown in Fig. 9 the activation maps are related in areas inside the chest (regardless of the different chest inclination). This is interesting because we can state that the proposed model does not consider areas outside the chest. The second consideration is that all areas are within the lungs, for this reason we can state that the models effectively consider the lungs for discriminating between COVID-19 and generic pulmonary diseases. Another interesting result is the one provided by the right X-ray shown in Fig. 9: with red arrow the radiologists marked the areas where the COVID-19 disease is manifested, as is shown from the activation maps, the proposed models rightly highlighted the areas marked by the radiologist.

Below we provide also the radiologist diagnosis: with regard to the left X-ray in Fig. 9 the radiologist confirmed the COVID-19 diagnosis adding that the patient is afflicted by bronchial wall thickening with small peripheral patchy infiltrates. While with regard to the right X-ray in Fig. 9, the radiologist also confirmed the COVID-19 diagnosis adding that the patient exhibits multifocal patchy opacities that can be seen in both lungs, confirming the red arrow the area of interest in the X-ray that are symptomatic of the COVID-19.

The analysis requires for a new chest X-ray approximately 2.5 seconds to make the predictions 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.

4. Related Work

In this section we review the current state of the art literature about the application of artificial intelligence techniques to chest area cancer detection detection.

4.1. Methods for generic pulmonary disease detection

A detection rate equal to 92.4% is reported by researchers in [32] where a method to discriminate between healthily and pathological crackles is proposed. The consider supervised machine learning obtaining the best performance using the kNN classification algorithms.

The SVM algorithm is exploited by authors in [33] to distinguish between pneumonia and congestive heart failure. A data-set of 257 patient is considered obtaining a detection rate ranging from 82% to 87%.

An average f-measure of 0.9 is reached by the method proposed by Lang and colleagues [34]. They propose a graph-based semi-supervised one class support vector machine to detect normal lung sounds and detect the abnormal ones.

A method to detect the presence of wheeze sounds in breath recordings is proposed by Torre-Cruz et al. [35]: authors obtain a detection ratio of 95.5% for classifying presence/absence of wheezing in respiratory sounds (i.e., in the discrimination of healthy and lung disease affected patients).

The main difference between the cited works and the one we proposed is that our work is focused on the COVID-19 detection. Moreover, this represents the first tentative to provide explainability by exploiting activation areas in chest biomedical images.

4.2. Methods focused on COVID-19 disease

Current literature presents also papers focused to COVID-19 detection. In details, the current research lines are related to the COVID-19 detection from medical images [5,36,37] (this is direction of the method we propose) and to protein analysis to boost research in COVID-19 vaccines [38,39].

Below we discuss these papers.

Authors in [36] exploit deep learning by considering the UNet++ model obtaining an accuracy equal to 98.85%. The consider 51 patients confirmed as affected by the COVID-19 disease belonging to one institution (i.e., the Renmin Hospital of Wuhan University). The main difference with respect to the proposed method is that authors in [36] do not take into account healthy patients. Moreover they do not provide explainability about their network predictions.

Researchers in [40] propose also a deep learning model for COVID-19 detection. Similarly to the method proposed by [36], authors do not consider healthy patients and do not provide a kind of explainable about the results of the proposed method. Moreover, the method in [36] requires to manually mark the region of interest of the COVID-19 disease. They obtain an accuracy equal to 73.1% by evaluating a data-set belonging from two different institutions. They evaluate medical images related to 99 patients, 55 cases of pneumonia and the other 44 cases affected by the COVID-19 disease.

Xu and colleagues [37] design an approach deep learning based by obtaining an accuracy equal to 86.7% by evaluating a dataset composed by 618 medical images. They consider two three-dimensional convolutional neural networks: the first one is the ResNet23 network, while the second network represent a variant of the first one, where authors added several layers.

Below, we discuss the papers [38,39] applying deep learning techniques to proteins with the aim to boost the research in new vaccines.

Zhang et al. [38] consider the DenseNet network with the COVID-19 RNA sequences with the aim to predict which current antivirals can help patients COVID-19 affected.

Beck et al. [39] experiment deep learning with the SMILES dataset, a repository of molecules as text to encode and decode each molecule. Their main outcome is that the 2019-nCoV 3C-like proteinase is predicted to bind with atazanavir i.e., an anti-viral medication used to treat HIV/AIDS.

Hemdan et al. [41] proposed COVIDX-Net for COVID-19 detection by analysing X-ray images. They reach an accuracy equal to 0.90 exploiting using 25 COVID-19 positive and 25 healthy images. Wang and Wong [40] designed COVID-Net, a deep learning network for COVID19 detection. They obtain an accuracy equal of 0.92 exploiting medical images obtained from different repositories. Apostolopoulos and colleagues apply transfer learning [42] analysing 224 COVID-19 medical images, 700 pneumonias, and 504 normal images. They reach an accuracy equal to 0.98 in the COVID-19 and healthy discrimination. Narin et al. [43] proposed a method considering three different networks (i.e., ResNet50, InceptionV3, and Inception-ResNetV2) by considering 50 COVID-19 chest X-ray images and 50 normal images obtained from a Kaggle repository⁸. We highlight that in authors consider non-COVID images belonging to children, while the COVID-19 one are related to adult patients. Sethy and Behera [44] consider deep learning models to gather features from medical images and then classified them by exploiting the SVM classifier. Researchers obtain an accuracy of 0.95 by considering ResNet50 and SVM model by analysing 50 medical images. Researchers in [45] achieved an accuracy equal to 0.86 analysing CT images with a deep model built on the ResNet50 model. Wang et al. [40] achieved an accuracy of 0.82 exploiting the modified Inception (M-Inception) deep model by analysing CT images. Zheng et al. [46] proposed a model to detect COVID-19 from CT images by reaching an accuracy equal to 0.9, while Xu et al. [37] achieved an accuracy of 0.86 in detecting COVID-19 considering the ResNet by analysing CT images. Most of these papers consider a few amount of data to develop the predictive models. Researchers in [30] consider a total of 1750 X-ray images with the DarkCovidNet deep learning model, in particular 250 COVID-19 positive, 500 related to other pulmonary pathologies and 1000 obtained from healthy patients obtaining an accuracy of 0.98 and 0.87 evaluating two dataset.

Ardakani et al. [47] propose a deep learning based approach for distinguishing from COVID-19 from non-COVID-19. As a matter of fact, ten well-known convolutional neural networks were used: AlexNet, VGG-16, VGG-19, SqueezeNet, GoogleNet, MobileNet-V2, ResNet-18, ResNet-50, ResNet-101, and Xception. Authors state that the best accuracy was achieved by the ResNet-101 and Xception models. ResNet-101 could distinguish COVID-19 from non-COVID-19 cases with an accuracy equal to accuracy, while the Xception model obtained an accuracy equal to accuracy 0.99. The several networks were evaluated with a dataset composed by 1020 CT slices obtained from 108 patients with afflicted by COVID-19 and 86 patients with other atypical and viral pneumonia diseases.

Li et al. [48] designed the COVNet neural network for COVID-19 detection to extract visual features from CT images. Their model adopt the ResNet50 network, with an input CT images. They evaluate 4,356 chest CT exams from 3,322 patients (1296 COVID-19 CT exams, 1,735 CAP and 1,325 non-pneumonia) reaching an accuracy equal to 0.96.

Butt et al. [49] proposed two different classification models for discriminating between COVID-19 images, with an overall accuracy equal to 0.86. The first model was a ResNet23-based network while the second model was designed on the first network

⁸ <https://www.kaggle.com/paultimothymooney/chest-xray-pneumonia>.

structure by concatenating the location-attention mechanism in the full-connection layer, with the aim to improve the overall accuracy rate. Their dataset is composed by 1,710 image patches acquired from 90 CT, including 357 COVID-19, 390 Influenza-A-viral-pneumonia, and 963 irrelevant-to-infection.

From the discussions it emerges that the proposed method obtain interesting performances if compared to the methods currently proposed by the state-of-the-art.

Moreover, we consider also healthy X-ray, for this reason our method is also able to discriminate between healthy and patients affected by pulmonary diseases. As a last point, we highlight that the proposed is first one, at the authors knowledge, providing explainability by showing the activation maps.

5. Conclusion and Future Work

Considered the time windows currently requested to obtain a COVID-19 diagnosis, in this paper we propose an approach aimed to drastically reduce this time window approximately to 2.5 seconds. We propose and evaluate an approach based on transfer learning by exploiting the VGG-16 model: we built two models, the first one aimed to detect whether a chest X-ray is related to a healthy patient or to a patient with generic pulmonary disease. In case the X-ray is marked with generic pulmonary disease, we input the X-ray to a second model, aimed to detect whether the pulmonary disease is COVID-19. Moreover, we highlight the areas that the models identified as discriminating for COVID-19, to provide explainability. As shows by the experimental results, obtained considering two different data-sets for a total of 6,523 chest X-rays, show an accuracy of 0.96 for the discrimination between healthy and generic pulmonary disease patients, and an accuracy of 0.98 for the COVID-19 detection. As future work, we plan to evaluate the proposed method on a wider set of pulmonary diseases. Moreover we, will investigate if formal verification techniques [52,53] can be helpful to obtain better results in terms of accuracy.

Declaration of Competing Interest

All authors have participated in (a) conception and design, or analysis and /interpretation of the data; (b) drafting the article or revising it critically for important intellectual content; and (c) approval of the final version. This manuscript has not been submitted to, nor is under review at, another "journal or other publishing venue. The authors have no affiliation with any organization with a direct or indirect financial interest in the subject matter discussed in the manuscript.

References

- [1] K.V. Holmes, Sars-associated coronavirus, *New England Journal of Medicine* 348 (20) (2003) 1948–1951.
- [2] L. van der Hoek, K. Pyrc, M.F. Jebbink, W. Vermeulen-Oost, R.J. Berkhout, K.C. Wolthers, P.M. Wertheim-van Dillen, J. Kaandorp, J. Spaargaren, B. Berkhout, Identification of a new human coronavirus, *Nature medicine* 10 (4) (2004) 368–373.
- [3] F. Abroug, A. Slim, L. Ouane-Besbes, M.-A.H. Kacem, F. Dachraoui, I. Ouane, X. Lu, Y. Tao, C. Paden, H. Caidi, et al., Family cluster of middle east respiratory syndrome coronavirus infections, tunisia, 2013, *Emerging infectious diseases* 20 (9) (2014) 1527.
- [4] Q. Li, X. Guan, P. Wu, X. Wang, L. Zhou, Y. Tong, R. Ren, K.S. Leung, E.H. Lau, J.Y. Wong, et al., Early transmission dynamics in wuhan, china, of novel coronavirus-infected pneumonia, *New England Journal of Medicine* (2020).
- [5] C. Wang, P.W. Horby, F.G. Hayden, G.F. Gao, A novel coronavirus outbreak of global health concern, *The Lancet* 395 (10223) (2020) 470–473.
- [6] E. Livingston, K. Bucher, Coronavirus disease 2019 (covid-19) in italy, *Jama* (2020).
- [7] Y. Wang, Y. Wang, Y. Chen, Q. Qin, Unique epidemiological and clinical features of the emerging 2019 novel coronavirus pneumonia (covid-19) implicate special control measures, *Journal of medical virology* 92 (6) (2020) 568–576.
- [8] H.V. Huff, A. Singh, Asymptomatic transmission during the covid-19 pandemic and implications for public health strategies, *Clinical Infectious Diseases* (2020).
- [9] M. Day, Covid-19: identifying and isolating asymptomatic people helped eliminate virus in italian village, *Bmj* 368 (187) (2020) m1165.
- [10] P. Romagnani, G. Gnone, F. Guzzi, S. Negrini, A. Guastalla, F. Annunziato, S. Romagnani, R. De Palma, The covid-19 infection: lessons from the italian experience (2020).
- [11] U. Bastolla, How lethal is the novel coronavirus, and how many undetected cases there are? the importance of being tested., *medRxiv* (2020).
- [12] M.-Y. Ng, E.Y. Lee, J. Yang, F. Yang, X. Li, H. Wang, M.M.-s. Lui, C.S.-Y. Lo, B. Leung, P.-L. Khong, et al., Imaging profile of the covid-19 infection: radiologic findings and literature review, *Radiology: Cardiothoracic Imaging* 2 (1) (2020) e200034.
- [13] C. Huang, Y. Wang, X. Li, L. Ren, J. Zhao, Y. Hu, L. Zhang, G. Fan, J. Xu, X. Gu, et al., Clinical features of patients infected with 2019 novel coronavirus in wuhan, china, *The Lancet* 395 (10223) (2020) 497–506.
- [14] J. Cohen, Wuhan seafood market may not be source of novel virus spreading globally, *Science* 10 (2020).
- [15] D.S. Hui, E. I Azhar, T.A. Madani, F. Ntoumi, R. Kock, O. Dar, G. Ippolito, T.D. Mchugh, Z.A. Memish, C. Drosten, et al., The continuing 2019-nCoV epidemic threat of novel coronaviruses to global health—the latest 2019 novel coronavirus outbreak in wuhan, china, *International Journal of Infectious Diseases* 91 (2020) 264–266.
- [16] S.A. Sellers, K.L. Dover, A.G. Bailey, A. Cheves, A.B. Eason, E.B. Popowitch, M.B. Miller, D.A. Wohl, D.P. Dittmer, W.A. Fischer, Burden of respiratory viral infection in persons with human immunodeficiency virus, *Influenza and Other Respiratory Viruses* (2020).
- [17] C. Long, H. Xu, Q. Shen, X. Zhang, B. Fan, C. Wang, B. Zeng, Z. Li, X. Li, H. Li, Diagnosis of the coronavirus disease (covid-19): rrt-pcr or ct? *European Journal of Radiology* (2020) 108961.
- [18] Y. Fang, H. Zhang, J. Xie, M. Lin, L. Ying, P. Pang, W. Ji, Sensitivity of chest ct for covid-19: comparison to rt-pcr, *Radiology* (2020) 200432.
- [19] T. Ai, Z. Yang, H. Hou, C. Zhan, C. Chen, W. Lv, Q. Tao, Z. Sun, L. Xia, Correlation of chest ct and rt-pcr testing in coronavirus disease 2019 (covid-19) in china: a report of 1014 cases, *Radiology* (2020) 200642.
- [20] L. Brunese, F. Mercaldo, A. Reginelli, A. Santone, Neural networks for lung cancer detection through radiomic features, in: 2019 International Joint Conference on Neural Networks (IJCNN), IEEE, 2019, pp. 1–10.
- [21] W. Bulten, H. Pinckaers, H. van Boven, R. Vink, T. de Bel, B. van Ginneken, J. van der Laak, C. Hulsbergen-van de Kaa, G. Litjens, Automated deep-learning system for gleason grading of prostate cancer using biopsies: a diagnostic study, *The Lancet Oncology* (2020).
- [22] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, *arXiv preprint arXiv:1409.1556* (2014).
- [23] S.J. Pan, Q. Yang, A survey on transfer learning, *IEEE Transactions on knowledge and data engineering* 22 (10) (2009) 1345–1359.
- [24] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, L. Fei-Fei, Imagenet: A large-scale hierarchical image database, in: 2009 IEEE conference on computer vision and pattern recognition, IEEE, 2009, pp. 248–255.
- [25] R.R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, D. Batra, Grad-cam: Visual explanations from deep networks via gradient-based localization, in: Proceedings of the IEEE international conference on computer vision, 2017, pp. 618–626.
- [26] F. Mercaldo, A. Santone, Deep learning for image-based mobile malware detection, *Journal of Computer Virology and Hacking Techniques* (2020) 1–15.
- [27] C.R. Qi, H. Su, K. Mo, L.J. Guibas, Pointnet: Deep learning on point sets for 3d classification and segmentation, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 652–660.
- [28] G. Litjens, T. Kooi, B.E. Bejnordi, A.A.A. Setio, F. Ciompi, M. Ghafoorian, J.A. Van Der Laak, B. Van Ginneken, C.I. Sánchez, A survey on deep learning in medical image analysis, *Medical image analysis* 42 (2017) 60–88.
- [29] J.P. Cohen, P. Morrison, L. Dao, Covid-19 image data collection, *arXiv* 2003.11597 (2020).
- [30] T. Ozturk, M. Talo, E.A. Yildirim, U.B. Baloglu, O. Yildirim, U.R. Acharya, Automated detection of covid-19 cases using deep neural networks with x-ray images, *Computers in Biology and Medicine* (2020) 103792.
- [31] X. Wang, Y. Peng, L. Lu, Z. Lu, M. Bagheri, R.M. Summers, Chestx-ray8: Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 2097–2106.
- [32] F. Jin, S. Krishnan, F. Sattar, Adventitious sounds identification and extraction using temporal-spectral dominance-based features, *IEEE Transactions on Biomedical Engineering* 58 (11) (2011) 3078–3087.
- [33] B. Flietstra, N. Markuzon, A. Vyshedskiy, R. Murphy, Automated analysis of crackles in patients with interstitial pulmonary fibrosis, *Pulmonary medicine* 2011 (2011).
- [34] R. Lang, R. Lu, C. Zhao, H. Qin, G. Liu, Graph-based semi-supervised one class support vector machine for detecting abnormal lung sounds, *Applied Mathematics and Computation* 364 (2020) 124487.
- [35] J. Torre-Cruz, F. Canadas-Quesada, S. García-Galán, N. Ruiz-Reyes, P. Vera-Can-deas, J. Carabias-Orti, A constrained tonal semi-supervised non-negative matrix factorization to classify presence/absence of wheezing in respiratory sounds, *Applied Acoustics* 161 (2020) 107188.

- [36] J. Chen, L. Wu, J. Zhang, L. Zhang, D. Gong, Y. Zhao, S. Hu, Y. Wang, X. Hu, B. Zheng, et al., Deep learning-based model for detecting 2019 novel coronavirus pneumonia on high-resolution computed tomography: a prospective study, *medRxiv* (2020).
- [37] X. Xu, X. Jiang, C. Ma, P. Du, X. Li, S. Lv, L. Yu, Y. Chen, J. Su, G. Lang, et al., Deep learning system to screen coronavirus disease 2019 pneumonia, *arXiv preprint arXiv:2002.09334* (2020).
- [38] H. Zhang, K.M. Saravanan, Y. Yang, M.T. Hossain, J. Li, X. Ren, Y. Wei, Deep learning based drug screening for novel coronavirus 2019-ncov (2020).
- [39] B.R. Beck, B. Shin, Y. Choi, S. Park, K. Kang, 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* (2020).
- [40] S. Wang, B. Kang, J. Ma, X. Zeng, M. Xiao, J. Guo, M. Cai, J. Yang, Y. Li, X. Meng, et al., A deep learning algorithm using ct images to screen for corona virus disease (covid-19), *medRxiv* (2020).
- [41] E.E.-D. Hemdan, M.A. Shouman, M.E. Karar, Covidx-net: A framework of deep learning classifiers to diagnose covid-19 in x-ray images, *arXiv preprint arXiv:2003.11055* (2020).
- [42] I.D. Apostolopoulos, T.A. Mpesiana, Covid-19: automatic detection from x-ray images utilizing transfer learning with convolutional neural networks, *Physical and Engineering Sciences in Medicine* (2020) 1.
- [43] A. Narin, C. Kaya, Z. Pamuk, Automatic detection of coronavirus disease (covid-19) using x-ray images and deep convolutional neural networks, *arXiv preprint arXiv:2003.10849* (2020).
- [44] P.K. Sethy, S.K. Behera, Detection of coronavirus disease (covid-19) based on deep features, *Preprints* 2020030300 (2020) 2020.
- [45] Y. Song, S. Zheng, L. Li, X. Zhang, X. Zhang, Z. Huang, J. Chen, H. Zhao, Y. Jie, R. Wang, et al., Deep learning enables accurate diagnosis of novel coronavirus (covid-19) with ct images, *medRxiv* (2020).
- [46] C. Zheng, X. Deng, Q. Fu, Q. Zhou, J. Feng, H. Ma, W. Liu, X. Wang, Deep learning-based detection for covid-19 from chest ct using weak label, *medRxiv* (2020).
- [47] A.A. Ardakani, A.R. Kanafi, U.R. Acharya, N. Khadem, A. Mohammadi, Application of deep learning technique to manage covid-19 in routine clinical practice using ct images: Results of 10 convolutional neural networks, *Computers in Biology and Medicine* (2020) 103795.
- [48] L. Li, L. Qin, Z. Xu, Y. Yin, X. Wang, B. Kong, J. Bai, Y. Lu, Z. Fang, Q. Song, et al., Artificial intelligence distinguishes covid-19 from community acquired pneumonia on chest ct, *Radiology* (2020) 200905.
- [49] C. Butt, J. Gill, D. Chun, B.A. Babu, Deep learning system to screen coronavirus disease 2019 pneumonia, *Applied Intelligence* (2020) 1.
- [50] L. Wang, A. Wong, 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).
- [51] X. Xu, X. Jiang, C. Ma, et al., Deep learning system to screen coronavirus disease 2019 pneumonia. *arXiv preprint arXiv:2002.09334*, (2020)
- [52] M.G. Cimino, N. De Francesco, F. Mercaldo, A. Santone, G. Vaglini, Model checking for malicious family detection and phylogenetic analysis in mobile environment, *Computers & Security* 90 (2020) 101691.
- [53] L. Brunese, F. Mercaldo, A. Reginelli, A. Santone, An ensemble learning approach for brain cancer detection exploiting radiomic features, *Computer methods and programs in biomedicine* 185 (2020) 105134.