ai-corona: Radiologist-Assistant Deep Learning Framework for COVID-19 Diagnosis in Chest CT Scans

M. Yousefzadeh[1,2,3*], P. Esfahanian[1,2*], S. M. S. Movahed[4,3], S. Gorgin[1,5], R. Lashgari[2], D. Rahmati[6,1], A. Kiani[7], S. Kahkouee[8], S. A. Nadji[9], S. Haseli[8], M. Hoseinyazdi[10], J. Roshandel[8], N. Bandegani[8], A. Danesh[8], M. Bakhshayesh Karam[8†], A. Abedini[8†]

[1] School of Computer Science, Institute for Research in Fundamental Sciences (IPM), Tehran, Iran
[2] Brain Engineering Research Center, Institute for Research in Fundamental Sciences (IPM), Tehran, Iran
[3] Ibn-Sina Multidisciplinary Laboratory, Department of Physics, Shahid Beheshti University, Tehran, Iran
[4] Department of Physics, Shahid Beheshti University, Tehran, Iran
[5] Department of Electrical Engineering and Information Technology, Iranian Research Organization for Science and Technology (IROST), Tehran, Iran
[6] Department of Computer Science and Engineering, Shahid Beheshti University, Tehran, Iran
[7] Tracheal Diseases Research Center, National Research Institute of Tuberculosis and Lung Diseases (NRITLD), Shahid Beheshti University of Medical Sciences and Health Services, Tehran, Iran
[8] Chronic Respiratory Diseases Research Center, National Research Institute of Tuberculosis and Lung Diseases (NRITLD), Shahid Beheshti University of Medical Sciences and Health Services, Tehran, Iran
[9] Virology Research Center, National Research Institute of Tuberculosis and Lung Diseases (NRITLD), Shahid Beheshti University of Medical Sciences and Health Services, Tehran, Iran
[10] Medical Imaging Research Center, Department of Radiology, Shiraz University of Medical Sciences, Shiraz, Iran

Abstract

Background: With the global outbreak of COVID-19 epidemic since early 2020, there has been considerable attention on CT-based diagnosis as an effective and reliable method. Recently, the advent of deep learning in medical diagnosis has been well proven. Convolutional Neural Networks (CNN) can be used to detect the COVID-19 infection imaging features in a chest CT scan. We introduce *ai-corona*, a radiologist-assistant deep learning framework for COVID-19 infection diagnosis using the chest CT scans.
Method: Our dataset comprises 2121 cases of axial spiral chest CT scans in three classes; COVID-19 abnormal, non COVID-19 abnormal, and normal, from which 1764 cases were used for training and 357 cases for validation. The training set was annotated using the reports of two experienced radiologists. The COVID-19 abnormal class validation set was annotated using the general consensus of a collective of criteria that indicate COVID-19 infection. Moreover, the validation sets for the non COVID-19 abnormal and the normal classes were annotated by a different experienced radiologist. *ai-corona* constitutes a CNN-based feature extractor conjoined with an average pooling and a fully-connected layer to classify a given chest CT scan into the three aforementioned classes.
Results: We compare the diagnosis performance of *ai-corona*, radiologists, and model-assisted radiologists for six combinations of distinguishing between the three mentioned classes, including COVID-19 abnormal vs. others, COVID-19 abnormal vs. normal, COVID-19 abnormal vs. non COVID-19 abnormal, non COVID-19 abnormal vs. others, normal vs. others, and normal vs. abnormal. *ai-corona* achieves an AUC score of 0.989 (95% CI: 0.984, 0.994), 0.997 (95% CI: 0.995, 0.999), 0.986 (95% CI: 0.981, 0.991), 0.959 (95% CI: 0.944, 0.974), 0.978 (95% CI: 0.968, 0.988), and 0.961 (95% CI: 0.951, 0.971) in each combination, respectively. By employing Bayesian statistics to calculate the accuracies at a 95% confidence interval, *ai-corona* surpasses the radiologists in distinguishing between the COVID-19 abnormal class and the other two classes (especially the non COVID-19 abnormal class). Our results show that radiologists' diagnosis performance improves when incorporating *ai-corona*'s prediction. In addition, we also show that RT-PCR's diagnosis has a much lower sensitivity compared to all the other methods.
Conclusion: *ai-corona* is a radiologist-assistant deep learning framework for fast and accurate COVID-19 diagnosis in chest CT scans. Our results ascertain that our framework, as a reliable detection tool, also improves experts' diagnosis performance and helps especially in diagnosing non-typical COVID-19 cases or non COVID-19 abnormal cases that manifest COVID-19 imaging features in chest CT scan.
Our framework is available at: ai-corona.com

Keywords— COVID-19 . Computed Tomography . Deep Learning . Convolutional Neural Networks

*Main authors. mah.yousefzadeh@mail.sbu.ac.ir, parsa.esfahanian@ipm.ir
†Corresponding authors.

1 Introduction

Since the beginning of 2020, novel Coronavirus Disease 2019 (COVID-19) has widely spread all over the world. As of April 29th, 2020, there have been 3,188,596 confirmed cases and 225,615 deaths reported worldwide, with a mortality rate of 19% in the closed cases [1]. Patients with the COVID-19 infection commonly display symptoms such as fever, cough, tiredness, breathing difficulties, and muscle ache [2, 3, 4].

The spread of COVID-19 throughout the world demonstrates a serious epidemic disease, and consequently threatens various aspects of communities such as economics, social managements, and even security. Any mathematical modeling to show the epidemic behaviour of COVID-19 is essentially encountered with incorrect data or lack of proper information which ultimately results a failure in having an impactful assistance in controlling the epidemic [5, 6, 7, 8]. Meanwhile, different strategies to reduce the impact of COVID-19 have been carried out. One of the prerequisites of such endeavour is procuring fast and reliable methods for detecting the infection that is very accurate and robust.

COVID-19 diagnosis consists of several standard methods, one of which is Real-Time Polymerase Chain Reaction (RT-PCR) to detect viral nucleotides from upper respiratory specimen obtained by nasopharyngeal swab, oropharyngeal swab, or nasal mid-turbinate swab [9]. Yet, it has been demonstrated that RT-PCR might have a low sensitivity in COVID-19 detection [10, 11]. Reports suggest that oropharyngeal swabs tend to detect COVID-19 less frequently than nasopharyngeal swabs. Apart from improper clinical sampling, reasons for the low efficiency of viral nucleic acid detection may include immature development of nucleic acid detection technology, variation in detection rate by using different gene region targets, and a low patient viral load [12]. In addition, the extended time period for the test completion contribute in ruling out RT-PCR as a reliable early detection and screening method [10, 11].

In contrast to RT-PCR, diagnosis from chest Computed Tomography (CT) has been shown to be an effective early detection and screening method with high sensitivity [13]. The Chest CT scan of a COVID-19 infected patient reveals bilateral peripheral involvement in multiple lobes with areas of consolidation and ground-glass opacity that progresses to "crazy-paving" patterns as the disease develops [13]. Asymmetric bilateral subpleural patchy ground glass opacities and consolidation with a peripheral or posterior distribution, mainly in middle and lower lobes, are described as the most common image finding of COVID-19 [14]. To elaborate more, additional common findings include interlobular septal thickening, air bronchogram and crazy paving pattern in the intermediate stages of the disease [13]. The most common pattern in advanced stage is subpleural parenchymal bands, fibrous stripes, and subpleural resolution. Nodules, cystic change, pleural effusion, pericardial effusion, lymphadenopathy, cavitation, CT halo sign, and pneumothorax are some of the uncommon but possible findings [13, 15]. Recent studies indicate that the organizing pneumonia, which occurs in the course of viral infection, is pathologically responsible for the main clinical and radiological manifestation of Coronavirus pneumonia [14].

Deep learning, as a subset of Artificial Intelligence (AI), has demonstrated tremendous capabilities in image feature extraction and has been recognized as a successful tool in medical imaging based diagnosis, performing exceptionally with single-image cases, such as X-Ray, and multi-image cases, such as Magnetic Resonance Imaging (MRI) and CT [16, 17, 18, 19]. Recently, the research of AI-assisted respiratory diagnosis, especially pneumonia, has gained a lot of attention, with competitions being held for its further development [20]. One of the well established standards in this research is the comparison of AI with expert medical and radiology professionals. As a pioneering work in this filed, [21] introduced a radiologist-level deep learning framework trained and validated on the ChestX-ray8 dataset [22] for the detection of 14 abnormalities, including pneumonia, in chest X-Ray images, which was further developed [23] to propose a deep learning framework with pneumonia detection capabilities equivalent to that of expert radiologists, with an Area Under the Receiver Operating Characteristic Curve (AUC) score of 0.851 (99.6%CI: 0.781, 0.911). [24] introduced a novel dataset of chest X-Ray images annotated with 14 abnormalities (7 the same as ChestX-ray8) and a state-of-the-art deep learning framework with a 0.90 AUC score in consolidation detection. Working with multi-image cases, [25] proposed a deep learning framework consisting of a feature extractor based on AlexNet [26], joined with some dense layers to create a model that is capable of accurate diagnosis on a MRI image.

In the COVID-19 related research, [10] has reported a sensitivity of 0.59 for RT-PCR test kit and 0.88 for CT-based diagnosis for patients with COVID-19 infection, and a radiologist sensitivity of 0.97 in diagnosing COVID-19 infected patients with a RT-PCR confirmation. Furthermore, [27] introduces a deep learning framework with a 0.96 (95%CI: 0.94, 0.99) AUC score in diagnosis of RT-PCR confirmed COVID-19 infected patients, achieving a sensitivity of 0.90 (95%CI: 0.83, 0.94) and a specificity of 0.96 (95%CI: 0.93, 0.98). A complete survey of integrating deep learning with COVID-19 research can be found at [28].

In this paper, we present *ai-corona*, a novel radiologist-level deep learning framework for the detection of COVID-19 in chest CT scans. Our framework was trained and tested on an all-inclusive dataset of over 2000 cases. A comprehensive and accurate methodology was carried out in order to annotate the validation set, on which we evaluate and compare the performance of *ai-corona*, radiologists, and RT-PCR in COVID-19 diagnosis and demonstrate the superiority of model-assisted radiologist diagnosis. Automated and early detection of COVID-19 infection would certainly prove invaluable and life saving in the global health-care battle against the COVID-19 epidemic. In short, the main advantages and novelties of our work are as follows:

- Introducing a comprehensive and authentic methodology for annotating the dataset cases for such work, especially the COVID-19 infection
- Proposing a deep learning framework which is capable of accurately diagnosing chest CT scans for COVID-19,

while being robust to the number of slices in the scan and having a low computational load

- Thorough evaluation of the diagnosis performance of *ai-corona*, radiologists, and RT-PRC in six distinct combinations of comparisons
- Evaluating and elucidating the impact of *ai-corona* on radiologists' diagnosis performance

The rest of this paper is organized as follows; section 2 provides the complete details of our dataset description, data pre-processing, and deep learning model. Our proposed strategy for COVID-19 detection based on deep learning and the statistical inference are also given in section 2. In addition, we will elucidate our results for different evaluation criteria and compare the performance of *ai-corona*, radiologists, RT-PCR, and model-assisted radiologists and assess the model's impact on radiologist diagnosis in section 3. Finally, we conclude with the discussion and a brief overview of the results and propose some future research directions in section 4.

2 Data and Method

2.1 Data Description

The cascade-like structure of our dataset utilized in this study is represented in Figure 1. A preliminary dataset was selected from a pre-existing repository of 2510 chest CT scans accompanying proper exclusive reports made by two practicing board-certified radiologists each with more than 10 years of experience. According to the radiologists' advice, the preliminary dataset of 2124 spiral CT scans was obtained by removing High-Resolution (HR) and abdominal CTs, which are not ideal for COVID-19 diagnosis. Based on the reports, the preliminary dataset was split into two categories; those that were suspicious for COVID-19 infection and those that were not.

Cases not reporting suspicious for COVID-19 infection at all were split into two classes; normal and non COVID-19 abnormal. The normal class holds patients that reported no pulmonary abnormalities, while the non COVID-19 abnormal class includes patients that the presence of at least one respiratory abnormality was seen or noted in their chest CT scan report. These abnormalities include atelectasis, cardiomegaly, lung emphysematous, hydropneumothorax, pneumothorax, cardiopulmonary edema, cavity, fibrocavitary changes, fibrobronchiectatic, mass, and nodule.

It is also worth noting that certain imaging characteristic manifestations and features in the lung associated with COVID-19 might be similar with other pathogens. Therefore it is crucial for our deep learning framework to distinguish between COVID-19 infection and other pathogens as the cause for a detected image feature, which will also

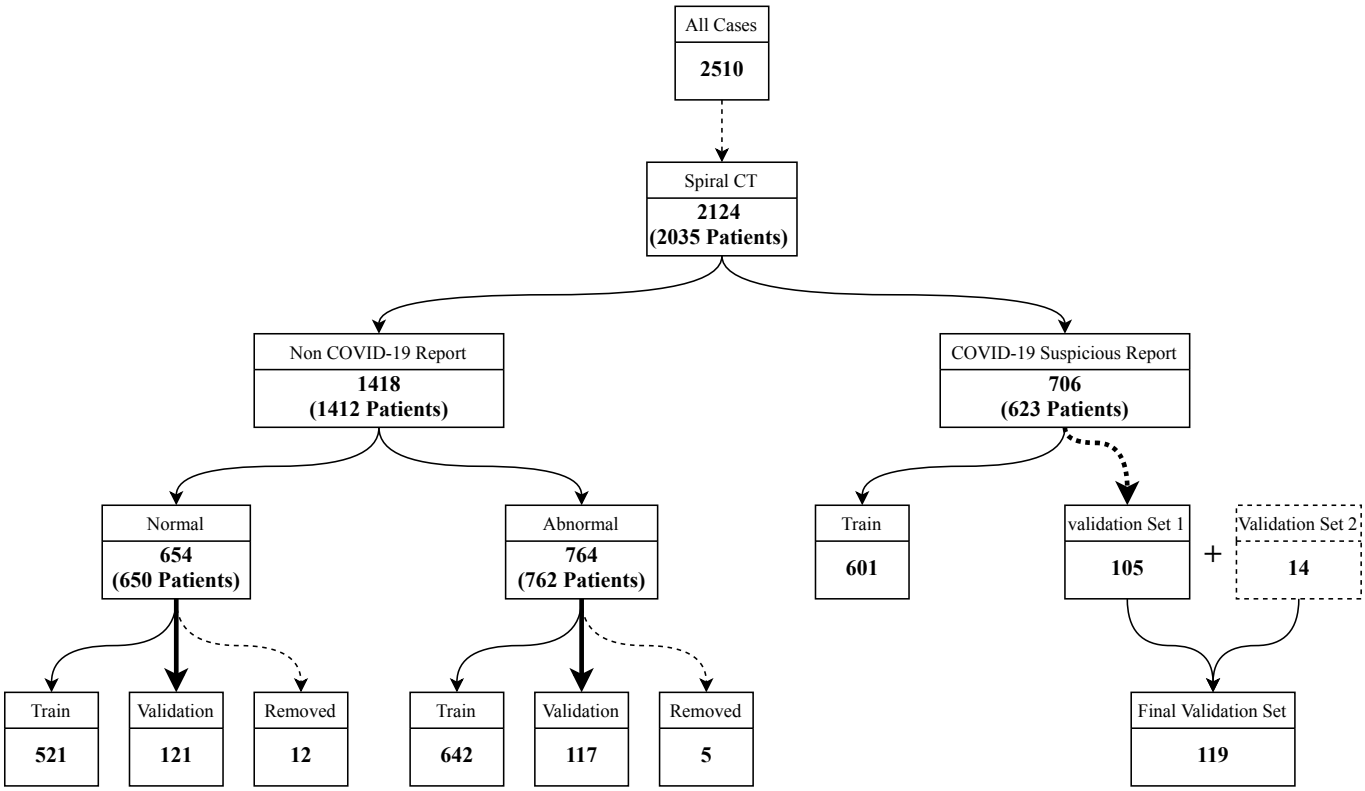


Figure 1: The structure of our dataset. Numbers indicate the amount of cases. Normal connections indicate normal split or merge. Thick or dashed connections indicate a validation set split that was re-annotated by another practicing board-certified radiologist, thick for accepted and dashed for removed cases. The thick dashed connection in the right indicates the special re-annotation process for the COVID-19 infected validation set. The addendum validation set in the far right of the diagram indicated with the dashed box and denoted "validation set 2" was not included in the initial 2510 cases and was later added to the set.

subsequently lead to an increase in the framework's specificity. Henceforth, CT scans with reports comprising conditions such as ground glass, consolidation, infiltration, and especially, non-COVID viral pneumonia, that might also appear in COVID-19 infection, are also included in the non COVID-19 abnormal class.

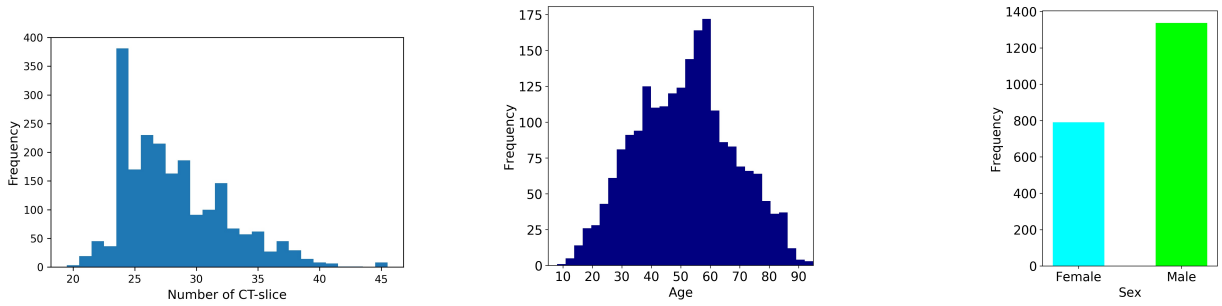
A subset of the cases in both aforementioned classes were randomly selected for the validation set and the model's training process was carried out using the rest. In order to introduce more fidelity and confidence in our framework's evaluation, the validation subset of the cases not reporting suspicious for COVID-19 infection was re-annotated by another practicing board-certified radiologist that has not seen or encountered any of the cases before. After the re-annotation, those that received the same annotation as before would make way into the validation set, and those with a new annotation that contradicted their original were removed.

As for cases reporting suspicious for COVID-19 infection, we selected a subset that met a certain harsh criteria as the COVID-19 abnormal class validation set. Such criteria are the collective consensus of a number of metrics that indicate the infection. These metrics include the report of at least one radiologist on the chest CT scan, confirmation of the infection by two pulmonologists, clinical presentation, RT-PCR report, and the fact that a patient checking-in with such infection claim has indeed been hospitalized for more than 3 days. The most important clinical features of COVID-19 are fever, dry cough, dyspnea, and myalgia or fatigue. Although sputum production, sore throat, rhinorrhoea, chest pain, headache, haemoptysis and diarrhea are seen as less common symptoms. [2, 3]

Since RT-PCR alone does not have a high sensitivity in COVID-19 diagnosis (as is explained more in section 3), we do not rely solely on its diagnosis, which is in contrary of what has been done in this research topic such as [10, 27], and in addition, we include the other mentioned metrics for truth annotation. These metrics insure the very high-accuracy annotations of our validation set, which will contribute to better model performance.

Respiratory samples including pharyngeal swabs/washing were obtained from February 20th till April 3rd, 2020, from the hospitalized patients. Nucleic acid was extracted from the samples with the QiaSymphony system (QIAGEN, Hilden, Germany) and SARS-CoV-2 RNA was detected using primer and probe sequences for screening and conformation on the basis of the sequence described by [29].

The validation set selection process for COVID-19 abnormal class is indicated by the thick dashed connection in the diagram in Figure 1. Furthermore, a small addendum set of 14 confirmed COVID-19 infected cases, that were not included in the initial 2510 cases but underwent the same selection process as the other validation set, was later joined with the rest of the COVID-19 validation set. Cases not included in final the validation set were used for training.



(a) (b) (c)

Figure 2: The left panel corresponds to the distribution of image slices for cases in our dataset, middle panel shows the distribution of Age, while the right panel illustrates the sex distribution of cases in our dataset.

All the scans in our dataset are in DICOM format and contain between 21 to 46 slices, taken in the axial plane. The slice thickness in each case varies between 8 and 10 *mm*. The CT scan machine used for producing our dataset cases operated on 110 kV and 50-60 mA. Also, the histogram representation for the number of scan slices has been indicated in Figure 2a, while Figure 2b and Figure 2c illustrate the age and sex distribution of cases in our dataset.

In conclusion, our dataset consisted of 521 cases in the training set and 121 cases in the validation set for the normal class, 642 cases in the training set and 117 cases in the validation set for the non COVID-19 abnormal class, and 601 cases in the training set and 119 cases in the validation set for the COVID-19 abnormal class. In our approach, we eventually try to classify our samples into three classes; COVID-19 abnormal, non COVID-19 abnormal, and normal.

2.2 Data Pre-Processing

Our dataset cases had to be processed before being fed to our deep learning pipeline. For all the cases and for each image slice, the top 0.5% of pixels with the highest value were selected and their values were clipped to the lowest one in the range. Next, a simple transformation is made to bring all the pixel values to the range [0, 255]. Since we utilize models pre-trained on the ImageNet dataset [30], an additional ImageNet normalization was also carried out.

2.3 Method

Inspired by [25], *ai-corona*'s deep learning model consists of two main blocks; a feature extractor and a classifier. The main challenge is mapping a 3-dimensional CT scan, which is a series of multiple image slices, to three probability

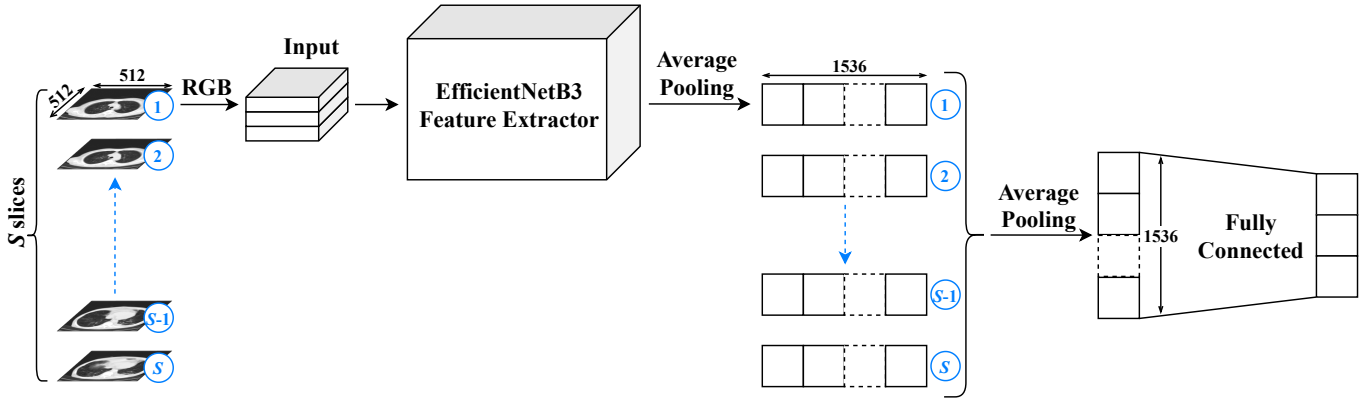


Figure 3: The schematic structure of *ai-corona*'s deep learning model. The total number of utilized slices is labeled by S. Each selected slice is fed to the feature extractor block pipeline one by one so that we end up with S vectors, which is then transformed to a single vector via an average pooling function. Afterwards, the result is passed through a fully connected network to reach the three output neurons, corresponding to our three classes; COVID-19 abnormal, non COVID-19 abnormal, and normal.

values. Another challenge is that all the scans do not have the same number of slices and not all the slices are useful for diagnosis. To address this, we take the middle 50% image slices in each scan and denote the number of selected slices from each scan with S. As shown in Figure 3, the feature extractor block is a pipeline, receiving each slice with dimensions $512 \times 512 \times 3$ (3 represents the number of color channels, but with all templates being exactly the same as for each image) and outputting a vector of length 1536 through an average pooling function. After all the slices have passed through the feature extractor block, we ended up with S vectors. After all the S slices have passed through the feature extractor block, an average pooling is applied to all that results in a single vector of length 1536.

This pipeline manner ensures that our framework is independent of the number of slices in a CT scan, as we always end up with a single vector of length 1536 at the end of the feature extractor block. The pipeline works like a machine. It receives any number of slices, extracts their features, and finally outputs a single vector of known length. Moreover, the use of only a single feature extractor at a time significantly reduces the computational load of our framework, resulting in a much faster training and prediction time.

Convolutional Neural Networks (CNN) were used for the feature extraction block. We experimented with different CNN models, such as DenseNet, ResNet, Xception, and EfficientNetB0 through EfficientNetB5 [31, 32, 33, 34], taking into account their accuracy and accuracy density on the ImageNet dataset [35]. All of these models were initialized with their respective pre-trained weights on the ImageNet dataset. At the end, the EfficientNetB3 model stripped of its last dense layers was chosen as the primary feature extractor for our deep learning framework. The vector output of the EfficientNetB5 feature extraction block is then passed through the classifier block, which contains yet another average pooling layer that is connected to the model's three output neurons corresponding to our three classes via a dense network of connections. *ai-corona* is implemented with Python 3.7 [36] and Keras 2.3 [37] framework and was trained on NVIDIA GeForce RTX 2080 Ti for 60 epochs in a total of three hours. The Pydicom [38] package was used to read the DICOM file of our dataset cases.
ai-corona has received the ethical license of IR.SBMU.NRITLD.REC.1399.024 from the Iranian National Committee for Ethics in Biomedical Research.

2.4 Statistical Inference

In order to quantify the reliability of our findings and the performance of our results based on *ai-corona* detection of COVID-19 in chest CT scans, we provide a thorough comparison with expert practicing radiologists diagnosis. To achieve a more conservative discrimination strategy, we compute the following evaluation criteria ranging from sensitivity (true positive rate), specificity (true negative rate), accuracy, F1-score Cohen's kappa, and finally to AUC.

We set the presence of the underlying class with positive label and the rest of the classes assigned by negative label. Incorporating error propagation and using the Bayesian statistics, we calculate the marginalized confidence region at 95% level for each computed quantity. The significance of diagnostic results is examined by computing the p-value statistics systematically. To achieve a conservative decision, the 3σ significance level is usually considered.

We take into account all the possible combinations of distinguishing between the three classes, including COVID-19 abnormal versus others, COVID-19 abnormal versus normal, COVID-19 abnormal versus non COVID-19 abnormal, non COVID-19 abnormal versus others, normal versus others, and normal versus abnormal, for evaluating the diagnosis performance of *ai-corona*, radiologists, and model-assisted radiologists.

Since the radiologists diagnosis is given by "Yes" or "No" statements for each class, it is necessary to convert the probability values computed by *ai-corona* to binary values. Hence, we change the threshold for distinguishing a given case among others and compute the true positive rate (sensitivity) versus false positive rate (1-specificity). To make more sense, as well as the other mentioned evaluation criteria, the Receiver Operating Characteristic (ROC) diagram is also estimated for all the various combinations. All of our criteria were calculated using the scikit-learn [39] package.

3 Results

As explained in subsection 2.1, our dataset of 2121 spiral CT cases (more precisely, we had 2124 cases originally, but removed 17 according to the re-annotation methodology and included 14 additional cases in the end) was used to train *ai-corona* on a 1764 case training set and then evaluate both the framework and radiologists on a 357 case validation set with all the evaluation criteria mentioned in subsection 2.4. We will also elucidate the increase in radiologists' diagnosis performance when *ai-corona*'s prediction is taken into consideration. Moreover, the sensitivity of *ai-corona* and radiologists was compared to RT-PCR. Our radiologists included two practicing academic board-certified radiologists, one practicing non board-certified radiologist, and one radiology resident, all of whom different from the previous aforementioned ones.

For the first part of our results, the sensitivity of RT-PCR for COVID-19 diagnosis was examined on a daily basis between February 24th and March 19th, 2020. RT-PCR's sensitivity at each day was determined by averaging the sensitivity of a 7-day period, centered around that day, and was calculated as the the ratio of COVID-19 positive cases in all the patients that were admitted for COVID-19 were hospitalized for more than three days while displaying clinical symptoms discussed in subsection 2.1. The patient specimen sampling was done in the early days of the hospitalization. As shown in Figure 4, RT-PCR's sensitivity starts at 0.362 (95% CI: 0.315, 0.409) and peaks at 0.579 (95% CI: 0.537, 0.621). The peak is an upper bound, because if instead of testing patients hospitalized for more than three days, every COVID-19 admitted patient was tested, RT-PCR's sensitivity would be much lower than 0.579. This result is comparable to the sensitivity of RT-PCR on specimens collected from patients deemed suspicious with COVID-19 infection by radiologists reported in [10]. We must point that, with more time and experience, the clinical respiratory samplings and adaptation and conducting the molecular testings were made more efficient and reliable, which would explain the improvement in RT-PCR's sensitivity during the examination period. Hereafter, we take the maximum value of RT-PCR's sensitivity as it's best and move forward from there.

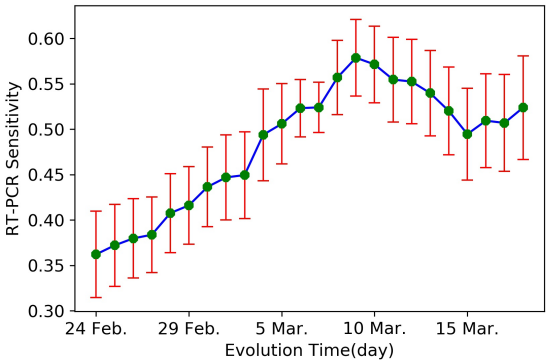


Figure 4: Evolution of RT-PCR's sensitivity in a 7-day period. The horizontal axis denotes the date at which the tests were taken.

Various combinations of performance comparison are made in our evaluation. In our work, RT-PCR's specificity was not available to us. Hence, its performance is represented as a solid horizontal line in Figure 5a, with a fixed sensitivity of 0.579 and a specificity in the range of [0, 1]. The shaded area around the solid horizontal line indicates the 95% confidence interval. A summery of *ai-corona*'s AUC scores for all the comparisons is brought in Table 1. Furthermore, Table 2 shows the diagnosis time for *ai-corona* and the radiologists for all the 357 cases in the validation set, which points out one of the obvious advantages of AI over the humans; time.

	AUC (95% CI)
COVID-19 abnormal vs. others	0.989 (95% CI: 0.984, 0.994)
COVID-19 abnormal vs. normal	0.997 (95% CI: 0.995, 0.999)
COVID-19 abnormal vs. non COVID-19 abnormal	0.986 (95% CI: 0.981, 0.991)
non COVID-19 abnormal vs. others	0.959 (95% CI: 0.944, 0.974)
normal vs. others	0.978 (95% CI: 0.968, 0.988)
normal vs. abnormal	0.961 (95% CI: 0.951, 0.971)

Table 1: AUC score of *ai-corona* for our six comparisons at a 95% confidence interval.

	ai-corona	Radiologist 1	Radiologist 2	Radiologist 3	Radiology resident
Diagnosis Time	12 *min.*	360 *min.*	300 *min.*	320 *min.*	400 *min.*

Table 2: Diagnosis time comparison for *ai-corona* and radiologists on the 357 case validation set.

The first comparison is associated with distinguishing between the COVID-19 abnormal class and the other two classes. Figure 5a shows the ROC diagram of this comparison. The inset plot in the figure magnifies the highest part

of sensitivity and specificity, to make more sense. *ai-corona* achieves a sensitivity of 0.924 (95% CI: 0.895, 0.953) and a specificity of 0.983 (95% CI: 0.959, 1.000) on the validation set, both of which better than the three radiologists and the radiology resident (the filled triangle symbols are below the ROC curve in the inset plot of the figure). The average sensitivity of 0.851 for the three radiologists and the radiology resident is comparable with the radiologists sensitivity of 0.88 reported in [10]. Incorporating *ai-corona*'s prediction in radiologists' diagnosis results an increase in their sensitivity (except Radiologist 2) and specificity. Such effect is labeled by adding the notation of +*ai-corona* to each diagnostic and it is indicated by the filled circle symbols in the figure. The brief notations such as Rad.# and R_res. refer to the radiologists' and the radiology resident's diagnosis, respectively. Both *ai-corona*, which gained an AUC score of 0.989 (95% CI: 0.984, 0.994), and radiologists had a better diagnosis performance than RT-PCR. The complete quantitative results can be found in Table 3. In this comparison, 93.2% of COVID-19 abnormal cases in the validation set (110 of 119) were diagnosed as infected by at least one radiologist. Out of the other 9 that were not, *ai-corona* managed to report one and RT-PCR reported two as infected. If RT-PCR was the only criteria for the truth annotation in our validation set, the overall sensitivity of radiologists would improve to 97%, confirming [10].

Since not all the patients admitted for COVID-19 had a previous history of respiratory diseases, the next important comparison made was in distinguishing between the COVID-19 abnormal class and the normal class, in which *ai-corona* performed pretty much the same as the radiologists and only slightly improved their diagnosis when assisting. This comparison is elaborated more in Table 4. Additionally, the ROC diagram of this comparison is shown in Figure 5b, which *ai-corona* gained and AUC score of 0.997 (95% CI: 0.995, 0.999).

Distinguishing between the COVID-19 abnormal class and the non COVID-19 abnormal class, *i.e.* diagnosing correctly between COVID-19 and other abnormality as the cause of a respiratory abnormalities is our third comparison. As some patients in the latter class might display imaging features in their chest CT scan similar to the COVID-19 infection, this comparison is extremely important. The ROC diagram for this comparison in Figure 5c shows that *ai-corona*, with an AUC score of 0.986 (95% CI: 0.981, 0.991), had the biggest improvement impact in radiologists' diagnosis in this particular comparison. Here, radiologists decided to diagnose suspicious cases as non COVID-19 abnormal, which led to a decrease in their sensitivity, but an increase in specificity. Complete details are in Table 5.

In the fourth and fifth comparison, diagnosis performance evaluation in distinguishing between the non COVID-19 abnormal class and the normal class with the other classes were made, respectively, which *ai-corona* gained an AUC score of 0.959 (95% CI: 0.944, 0.974) and 0.978 (95% CI: 0.968, 0.988). In the forth comparison, our deep learning framework managed to only outperform the radiology resident, while improving everyone's diagnosis performance, as presented in Table 6 and Figure 5d. The fifth comparison, exhibited in Table 7 and Figure 5e, plays out pretty much the same as forth, showcasing *ai-corona*'s slight disadvantage in distinguishing between the non COVID-19 abnormal and normal classes, which is investigated further and proven in the sixth comparison; as it only evaluates the diagnosis performance in distinguishing between the non COVID-19 abnormal and the normal class. More details for the sixth comparison and its ROC diagram can be found in Table 8 and Figure 5f.

Since there are many types of image features recognized for all our accounted abnormalities in our non COVID-19 abnormal class and due to confining all of them as one, *ai-corona* slightly under-performs the radiologists in distinguishing between all of the non COVID-19 abnormalities. Yet, for detecting the COVID-19 abnormal case, we use a distinct class, which consequently yields the overall performance of the *ai-corona* for this purpose being better than the other diagnosis methods. To improve the deep learning model's diagnosis performance in the forth, fifth, and sixth comparison, we need to take into account more precisely annotated cases for all different types of pulmonary abnormalities and this is beyond the scope of the current study.

4 Conclusion and Discussion

We introduce *ai-corona*, a radiologist-assistant deep learning framework capable of accurate COVID-19 diagnosis in chest CT scans. Our deep learning framework was trained on an all-inclusive dataset of 1764 axial spiral CT cases in order to learn to diagnose patients infected with COVID-19, as well as patients with non COVID-19 abnormalities and normal patients. The trained model was then evaluated on a 357 case validation set (Figure 1).

COVID-19 annotations in the validation set did not rely solely on RT-PCR's result. Instead, a collection of metrics, including the report of at least one radiologist on the chest CT scan, confirmation of COVID-19 infection by two pulmonologists, clinical presentation, and at least 3 days of hospitalization for COVID-19 infected patients, in companion with RT-PCR's result, decided the more confident and accurate annotation of COVID-19 cases.

We employed an EfficientNetB3-based feature extractor in *ai-corona* to address the issue of variable slice size in the dataset cases. We dynamically select the middle 50% slices in each case and feed it to the feature extractor, which will result in a single vector that will be classified. This method ensures that our framework is independent of the number of slices in a given scan (Figure 3). Moreover, the pipeline fashion of the feature extractor, and hence having only one of it, will reduce the computational overhead, resulting in a faster training and prediction time (Table 2).

In this literature, an AI's diagnosis performance is usually compared to that of expert professionals and other means of diagnosis in order to achieve a comprehensive and sensible image of the AI's abilities. Therefore in this work, we also followed such guidelines, and compared *ai-corona*'s diagnosis performance to a team of expert radiologists and RT-PCR. In the end, *ai-corona* triumphed over the others.

Since there were three classes of cases in our datase, COVID-19 abnormal, non COVID-19 abnormal, and normal, we make various combinations of performance comparisons in our evaluation. These comparisons include COVID-19 abnormal versus others, COVID-19 abnormal versus normal, COVID-19 abnormal versus non COVID-19 abnormal, non COVID-19 abnormal versus others, normal versus others, and normal versus abnormal, which their results based on sensitivity, specificity, accuracy, F1-score, Cohen's kappa, and AUC have been demonstrated in Table 3 through Table 8. The ROC diagrams of all the comparisons are also showcased in Figure 5, with the AUC score of *ai-corona* in each comparison summarized in Table 1. In the most important evaluation, *ai-corona* achieved an AUC score of 0.989 (95% CI: 0.984, 0.994) for distinguishing between the COVID-19 abnormal class and the other two classes.

At last, *ai-corona*'s impact on assisting radiologists' diagnosis was evaluated, which in the six comparisons, mostly indicates a positive impact over radiologists' sensitivity and specificity by several percentages.

On having a positive impact on radiologists diagnosis, two cases are discussed here to showcase how *ai-corona* made the radiologists change their mind for good in suspicious cases. At least one radiologist misdiagnosed Figure 6a's case as non COVID-19 abnormal at first, but upon seeing *ai-corona*'s diagnosis, corrected their diagnosis to COVID-19 abnormal. Radiologists cited "Peribrochovascular distribution was seen, which is not common in COVID-19 (no subpleural distribution)." as the reason for their misdiagnosis. In addition, Figure 6b's case was initially misdiagnosed as COVID-19 abnormal by at least one radiologists, yet was changed correctly to non COVID-19 abnormal when seeing *ai-corona*'s correct diagnosis. They cited "Cavity, centrilobular nodule, mass, and mass like consolidations are not commonly seen in COVID-19 pneumonia and implicate other diagnosis."

On the other hand, the existence of error in CT-based diagnosis, both for *ai-corona* and radiologists, encourages us to study the cause for such errors, which might lead to better and more accurate predictions, or point out any if existing fundamental flaws in CT-based diagnosis. Figure 6c's case was misdiagnosed as COVID-19 abnormal by all the radiologists. And *ai-corona*, while diagnosing correctly for non COVID-19 abnormal itself, was not able to change their minds. The final report was presented by radiologists in consensus:

> "Mediastinal and bilateral hilar adenopathies were seen. Diffuse bilateral interstitial infiltrations are detected with crazy paving pattern, ground glass, and traction bronchiectasis, mainly in the right lung and partial volume loss of the right lung. Anterior mediastinal soft-tissue density is seen. The position of central venous catheter tip is seen in the left brachiocephalic vein."

Finally, Figure 6d's case was misdiagnosed by all the radiologists and *ai-corona*. Complicated imaging manifestations of bacterial infection or emphysematous changes or pulmonary edema can obscure the typical imaging findings of COVID-19. For instance, architectural distortion in cases with emphysema or consolidation opacities in bacterial infections make it difficult to diagnose COVID-19 based on CT images.

ai-corona is an AI-based radiologist assistant tool that results an increase in expert's diagnosis accuracy and due to its very fast prediction time, leads to a much faster detection in suspicious patients. *ai-corona* is capable of detecting COVID-19 imaging features in a chest CT scan with very high accuracy. As the other forms of pulmonary abnormalities are all bundled up in only one additional class, *ai-corona* is in a slight disadvantage in distinguishing between COVID-19 and non COVID-19 abnormalities, which will open up a possible research direction for future. With accurate abnormality annotations, besides COVID-19, a capable deep learning framework would certainly perform better. Additionally, evaluating *ai-corona* on external validation data is most necessary. In conclusion, with the individual drawbacks of diagnosing based on clinical representation, RT-PCR, and CT-based diagnosis, a method comprised of all three would definitely yield the most accurate diagnosis of COVID-19.

5 Acknowledgement

Our framework is available to expert professionals and the public health-care via the website at ai-corona.com for free and unlimited use, where they can upload a chest CT scan and have it diagnosed for COVID-19 infection. The authors would like to express their gratitude to Masih Daneshvari Hospital and Zahra Yousefi for all their hard work and assistance in this project. The computational part of this work was carried out on the High-Performance Computing Cluster of the Institute for Research in Fundamental Sciences (IPM). Our project has received the ethical license of IR.SBMU.NRITLD.REC.1399.024 from the Iranian National Committee for Ethics in Biomedical Research.

	Sensitivity (95% CI)	Specificity (95% CI)	Accuracy (95% CI)	F1-Score (95% CI)	Cohen's kappa (95% CI)
ai-corona	0.924 (0.895, 0.953)	0.983 (0.959, 1.000)	0.964 (0.951, 0.977)	0.953 (0.934, 0.972)	0.917 (0.887, 0.947)
Rad.1	0.857 (0.833, 0.881)	0.979 (0.972, 0.986)	0.938 (0.929, 0.947)	0.903 (0.886, 0.920)	0.858 (0.838, 0.878)
Rad.1+*ai-corona*	0.908 (0.887, 0.929)	0.987 (0.982, 0.992)	0.961 (0.953, 0.969)	0.939 (0.927, 0.951)	0.910 (0.892, 0.928)
Rad.2	0.899 (0.874, 0.924)	0.979 (0.973, 0.985)	0.952 (0.944, 0.960)	0.926 (0.912, 0.940)	0.891 (0.868, 0.914)
Rad.2+*ai-corona*	0.891 (0.869, 0.913)	0.992 (0.987, 0.997)	0.961 (0.954, 0.968)	0.938 (0.927, 0.949)	0.909 (0.893, 0.925)
Rad.3	0.765 (0.738, 0.792)	0.992 (0.987, 0.997)	0.916 (0.905, 0.927)	0.858 (0.838, 0.878)	0.800 (0.775, 0.825)
Rad.3+*ai-corona*	0.857 (0.833, 0.881)	1.000 (1.000, 1.000)	0.952 (0.945, 0.959)	0.923 (0.908, 0.938)	0.889 (0.869, 0.909)
Res_r	0.882 (0.858, 0.906)	0.920 (0.907, 0.933)	0.908 (0.896, 0.920)	0.864 (0.846, 0.882)	0.794 (0.766, 0.822)
R_res+*ai-corona*	0.899 (0.877, 0.921)	0.966 (0.958, 0.974)	0.944 (0.934, 0.954)	0.915 (0.901, 0.929)	0.873 (0.853, 0.893)

Table 3: The quantitative evaluation of *ai-corona*, radiologists, and model-assisted radiologists performance results for distinguishing between the COVID-19 abnormal class and the other two classes at a 95% confidence interval.

	Sensitivity (95% CI)	Specificity (95% CI)	Accuracy (95% CI)	F1-Score (95% CI)	Cohen's kappa (95% CI)
ai-corona	0.983 (0.969, 0.997)	0.967 (0.951, 0.983)	0.975 (0.964, 0.986)	0.975 (0.965, 0.985)	0.950 (0.929, 0.971)
Radiologist 1	0.958 (0.947, 0.969)	0.992 (0.987, 0.997)	0.975 (0.969, 0.981)	0.974 (0.968, 0.980)	0.950 (0.939, 0.961)
Rad.1+*ai-corona*	0.992 (0.987, 0.997)	0.983 (0.976, 0.990)	0.988 (0.984, 0.992)	0.987 (0.983, 0.991)	0.975 (0.966, 0.984)
Radiologist 2	0.966 (0.956, 0.976)	0.942 (0.929, 0.955)	0.954 (0.946, 0.962)	0.954 (0.947, 0.961)	0.908 (0.894, 0.922)
Rad.2+*ai-corona*	0.975 (0.967, 0.983)	0.975 (0.967, 0.983)	0.975 (0.969, 0.981)	0.975 (0.969, 0.981)	0.950 (0.938, 0.962)
Radiologist 3	0.983 (0.976, 0.99)	0.959 (0.949, 0.969)	0.971 (0.965, 0.977)	0.971 (0.964, 0.978)	0.942 (0.929, 0.955)
Rad.3+*ai-corona*	0.983 (0.976, 0.990)	0.950 (0.939, 0.961)	0.967 (0.961, 0.973)	0.967 (0.960, 0.974)	0.933 (0.918, 0.948)
Radiology resident	0.966 (0.957, 0.975)	0.917 (0.903, 0.931)	0.942 (0.933, 0.951)	0.943 (0.933, 0.953)	0.883 (0.866, 0.900)
Rad.res+*ai-corona*	0.966 (0.956, 0.976)	0.967 (0.958, 0.976)	0.967 (0.960, 0.974)	0.966 (0.959, 0.973)	0.933 (0.919, 0.947)

Table 4: The quantitative evaluation of *ai-corona*, radiologists, and model-assisted radiologists performance results for distinguishing between the COVID-19 abnormal class and the normal class at a 95% confidence interval.

	Sensitivity (95% CI)	Specificity (95% CI)	Accuracy (95% CI)	F1-Score (95% CI)	Cohen's kappa (95% CI)
ai-corona	0.924 (0.896, 0.952)	0.974 (0.959, 0.989)	0.949 (0.934, 0.964)	0.949 (0.934, 0.964)	0.898 (0.872, 0.924)
Radiologist 1	0.857 (0.836, 0.878)	0.957 (0.945, 0.969)	0.907 (0.896, 0.918)	0.903 (0.891, 0.915)	0.814 (0.794, 0.834)
Rad.1+*ai-corona*	0.908 (0.893, 0.923)	0.974 (0.966, 0.982)	0.941 (0.933, 0.949)	0.939 (0.929, 0.949)	0.881 (0.863, 0.899)
Radiologist 2	0.899 (0.882, 0.916)	0.957 (0.947, 0.967)	0.928 (0.920, 0.936)	0.926 (0.915, 0.937)	0.856 (0.835, 0.877)
Rad.2+*ai-corona*	0.899 (0.883, 0.915)	0.983 (0.976, 0.990)	0.941 (0.931, 0.951)	0.939 (0.929, 0.949)	0.881 (0.864, 0.898)
Radiologist 3	0.765 (0.742, 0.788)	0.983 (0.976, 0.990)	0.873 (0.861, 0.885)	0.858 (0.844, 0.872)	0.746 (0.720, 0.772)
Rad.3+*ai-corona*	0.857 (0.837, 0.877)	1.000 (1.000, 1.000)	0.928 (0.918, 0.938)	0.923 (0.913, 0.933)	0.856 (0.836, 0.876)
Radiology resident	0.882 (0.867, 0.897)	0.855 (0.838, 0.872)	0.869 (0.855, 0.883)	0.871 (0.858, 0.884)	0.737 (0.713, 0.761)
Rad.res+*ai-corona*	0.899 (0.883, 0.915)	0.940 (0.928, 0.952)	0.919 (0.909, 0.929)	0.918 (0.907, 0.929)	0.839 (0.818, 0.860)

Table 5: The quantitative evaluation of *ai-corona*, radiologists, and model-assisted radiologists performance results for distinguishing between the COVID-19 abnormal class and the non COVID-19 abnormal class at a 95% confidence interval.

	Sensitivity (95% CI)	Specificity (95% CI)	Accuracy (95% CI)	F1-Score (95% CI)	Cohen's kappa (95% CI)
ai-corona	0.915 (0.883, 0.947)	0.929 (0.893, 0.965)	0.924 (0.908, 0.940)	0.922 (0.894, 0.950)	0.831 (0.793, 0.869)
Radiologist 1	0.897 (0.876, 0.918)	0.946 (0.936, 0.956)	0.930 (0.920, 0.94)	0.894 (0.877, 0.911)	0.841 (0.817, 0.865)
Rad.1+*ai-corona*	0.949 (0.934, 0.964)	0.950 (0.939, 0.961)	0.950 (0.940, 0.960)	0.925 (0.913, 0.937)	0.887 (0.867, 0.907)
Radiologist 2	0.949 (0.934, 0.964)	0.938 (0.925, 0.951)	0.941 (0.932, 0.950)	0.914 (0.900, 0.928)	0.869 (0.851, 0.887)
Rad.2+*ai-corona*	0.974 (0.963, 0.985)	0.950 (0.940, 0.960)	0.958 (0.951, 0.965)	0.938 (0.926, 0.950)	0.906 (0.886, 0.926)
Radiologist 3	0.923 (0.901, 0.945)	0.871 (0.855, 0.887)	0.888 (0.874, 0.902)	0.844 (0.829, 0.859)	0.757 (0.729, 0.785)
Rad.3+*ai-corona*	0.983 (0.974, 0.992)	0.912 (0.896, 0.928)	0.936 (0.927, 0.945)	0.909 (0.894, 0.924)	0.860 (0.840, 0.880)
Radiology resident	0.821 (0.793, 0.849)	0.925 (0.913, 0.937)	0.891 (0.878, 0.904)	0.831 (0.813, 0.849)	0.750 (0.720, 0.780)
R_res.+*ai-corona*	0.923 (0.904, 0.942)	0.954 (0.945, 0.963)	0.944 (0.935, 0.953)	0.915 (0.901, 0.929)	0.873 (0.851, 0.895)

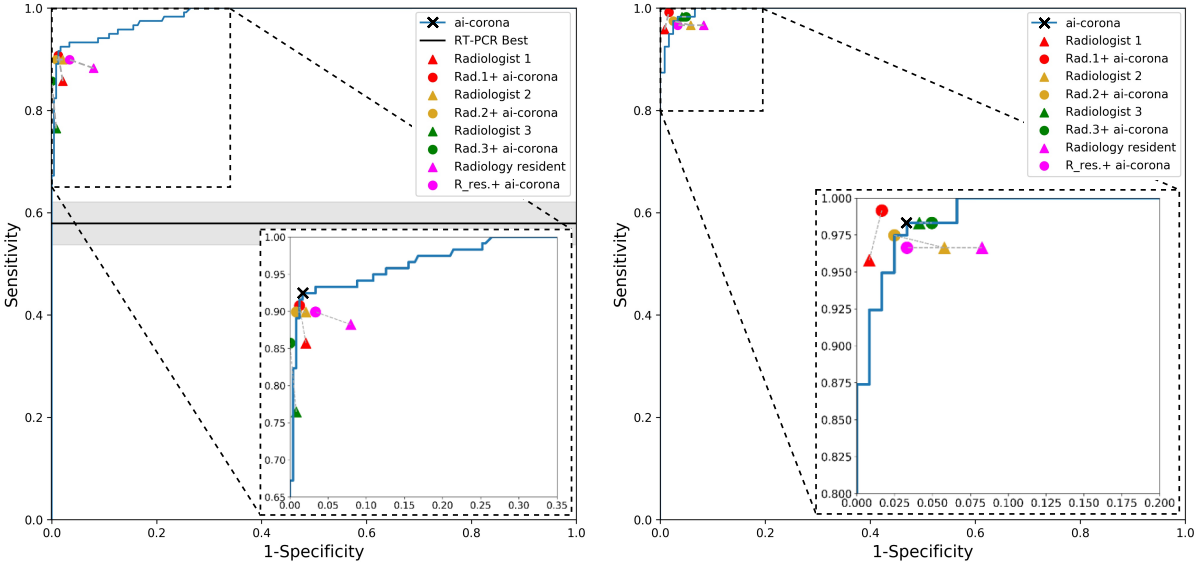
Table 6: The quantitative evaluation of *ai-corona*, radiologists, and model-assisted radiologists performance results for distinguishing between the non COVID-19 abnormal class and the other two classes at a 95% confidence interval.

	Sensitivity (95% CI)	Specificity (95% CI)	Accuracy (95% CI)	F1-Score (95% CI)	Cohen's kappa (95% CI)
ai-corona	0.942 (0.917, 0.967)	0.919 (0.883, 0.955)	0.927 (0.909, 0.945)	0.931 (0.907, 0.955)	0.841 (0.801, 0.881)
Radiologist 1	0.992 (0.986, 0.998)	0.949 (0.938, 0.960)	0.964 (0.957, 0.971)	0.949 (0.939, 0.959)	0.920 (0.904, 0.936)
Rad.1+*ai-corona*	0.983 (0.975, 0.991)	0.983 (0.977, 0.989)	0.983 (0.978, 0.988)	0.975 (0.967, 0.983)	0.963 (0.950, 0.976)
Radiologist 2	0.942 (0.925, 0.959)	0.979 (0.971, 0.987)	0.966 (0.958, 0.974)	0.950 (0.938, 0.962)	0.925 (0.908, 0.942)
Rad.2+*ai-corona*	0.975 (0.963, 0.987)	0.983 (0.977, 0.989)	0.980 (0.975, 0.985)	0.971 (0.963, 0.979)	0.956 (0.944, 0.968)
Radiologist 3	0.959 (0.945, 0.973)	0.962 (0.953, 0.971)	0.961 (0.953, 0.969)	0.943 (0.931, 0.955)	0.913 (0.895, 0.931)
Rad.3+*ai-corona*	0.950 (0.936, 0.964)	0.983 (0.976, 0.990)	0.972 (0.965, 0.979)	0.958 (0.948, 0.968)	0.937 (0.924, 0.950)
Radiology resident	0.917 (0.900, 0.934)	0.966 (0.959, 0.973)	0.950 (0.940, 0.960)	0.925 (0.911, 0.939)	0.887 (0.869, 0.905)
R_res.+*ai-corona*	0.967 (0.955, 0.979)	0.975 (0.968, 0.982)	0.972 (0.966, 0.978)	0.959 (0.949, 0.969)	0.938 (0.923, 0.953)

Table 7: The quantitative evaluation of *ai-corona*, radiologists, and model-assisted radiologists performance results for distinguishing between the normal class and the other two classes at a 95% confidence interval.

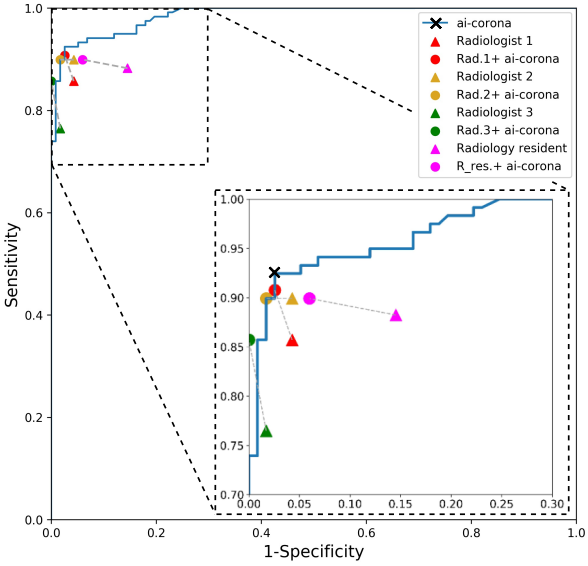
	Sensitivity (95% CI)	Specificity (95% CI)	Accuracy (95% CI)	F1-Score (95% CI)	Cohen's kappa (95% CI)
ai-corona	0.906 (0.878, 0.934)	0.917 (0.891, 0.943)	0.912 (0.896, 0.928)	0.912 (0.893, 0.931)	0.823 (0.789, 0.857)
Radiologist 1	0.940 (0.928, 0.952)	0.992 (0.988, 0.996)	0.966 (0.959, 0.973)	0.965 (0.958, 0.972)	0.933 (0.917, 0.949)
Rad.1+*ai-corona*	0.974 (0.967, 0.981)	0.983 (0.976, 0.990)	0.979 (0.974, 0.984)	0.979 (0.974, 0.984)	0.958 (0.947, 0.969)
Radiologist 2	0.991 (0.986, 0.996)	0.942 (0.931, 0.953)	0.966 (0.960, 0.972)	0.967 (0.960, 0.974)	0.933 (0.919, 0.947)
Rad.2+*ai-corona*	0.991 (0.987, 0.995)	0.975 (0.967, 0.983)	0.983 (0.978, 0.988)	0.983 (0.978, 0.988)	0.966 (0.956, 0.976)
Radiologist 3	0.940 (0.928, 0.952)	0.959 (0.948, 0.970)	0.950 (0.942, 0.958)	0.948 (0.938, 0.958)	0.899 (0.883, 0.915)
Rad.3+*ai-corona*	0.983 (0.975, 0.991)	0.950 (0.939, 0.961)	0.966 (0.959, 0.973)	0.966 (0.959, 0.973)	0.933 (0.918, 0.948)
Radiology resident	0.966 (0.955, 0.977)	0.917 (0.901, 0.933)	0.941 (0.932, 0.950)	0.942 (0.934, 0.950)	0.882 (0.863, 0.901)
R_res.+*ai-corona*	0.983 (0.976, 0.990)	0.967 (0.957, 0.977)	0.975 (0.969, 0.981)	0.975 (0.969, 0.981)	0.950 (0.937, 0.963)

Table 8: The quantitative evaluation of *ai-corona*, radiologists, and model-assisted radiologists performance results for distinguishing between the non COVID-19 abnormal class and the normal class at a 95% confidence interval.

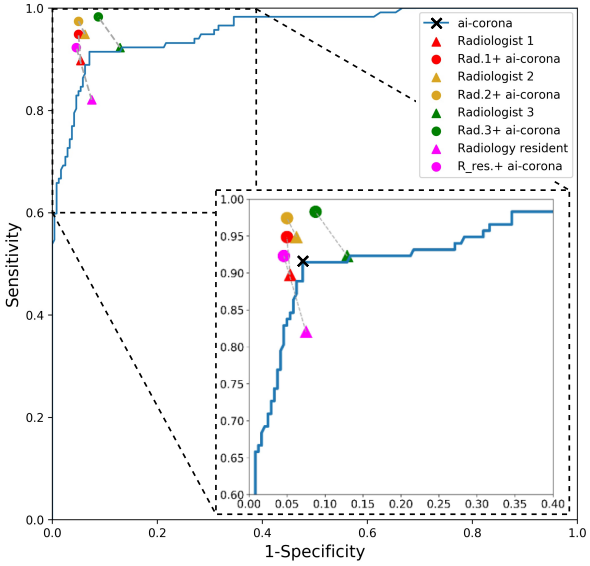


(a) COVID-19 abnormal vs. others

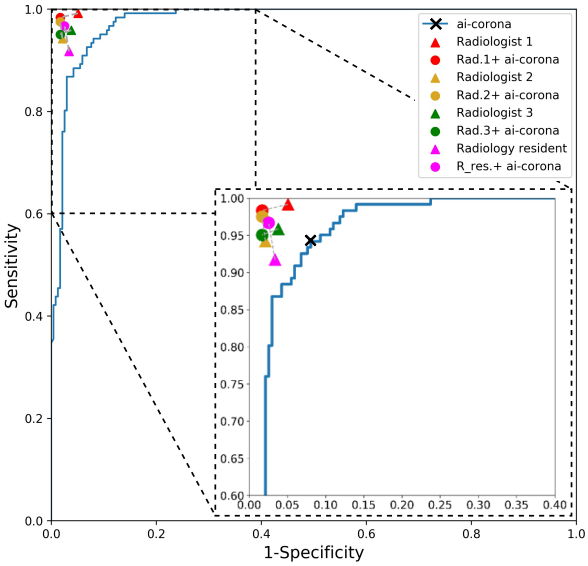
(b) COVID-19 abnormal vs. normal



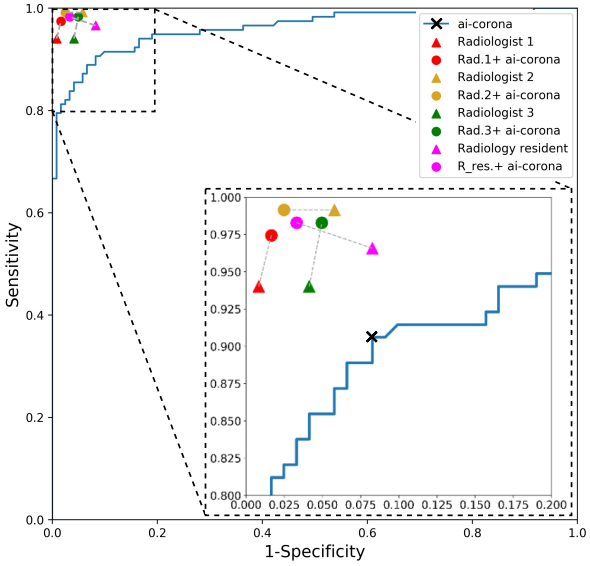
(c) COVID-19 abnormal vs. non COVID-19 abnormal



(d) non COVID-19 abnormal vs. others



(e) normal vs. others



(f) normal vs. abnormal

Figure 5: The ROC diagram representing the performance of various pipelines for the different combinations of comparison. The Solid blue line is for *ai-corona* by adapting different discrimination threshold value which is used to convert the continuous probability to binary "Yes" or "No" results. The filled triangle symbols are the (1-specificity, sensitivity) for the individual clinical expert, while the filled circle symbols are for the model-assisted radiologist. The inset plots magnify the highest part of sensitivity and specificity.

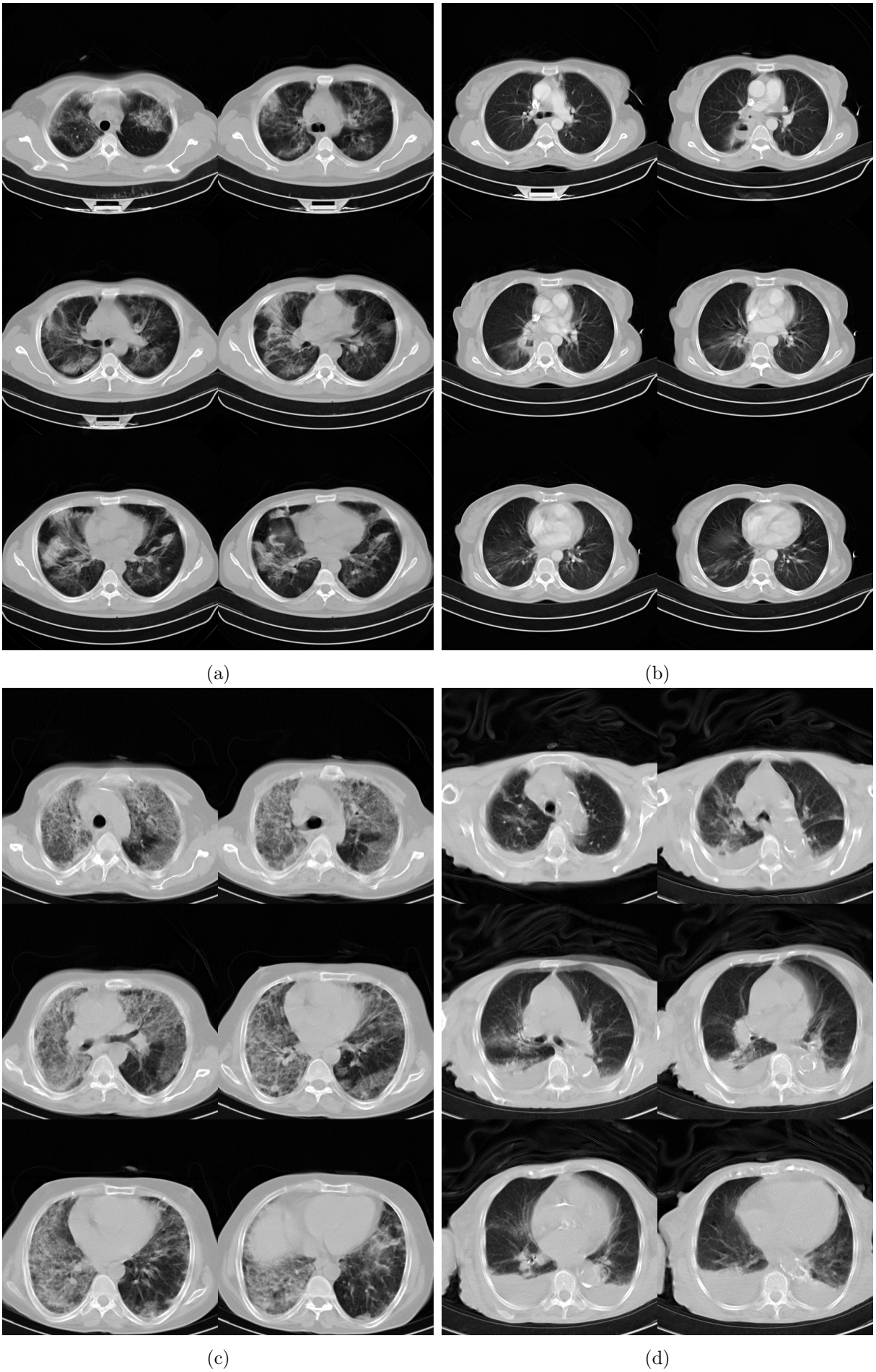


Figure 6: Panels (a), (b), and (c), are the chest CT scans of patients who were initially misdiagnosed by at least one radiologist but were then diagnosed correctly upon incorporating *ai-corona*'s correct prediction. Panel (d) shows the chest CT scans of patient that was misdiagnosed by *ai-corona* and radiologists.

References

[1] www.worldometers.info/coronavirus

[2] Huang, Chaolin, et al. "Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China." The Lancet 395.10223 (2020): 497-506.

[3] Chen, Nanshan, et al. "Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study." The Lancet 395.10223 (2020): 507-513.

[4] Kujawski, Stephanie A., et al. "characteristics of the first 12 patients with coronavirus disease 2019 (COVID-19) in the United States." Nature Medicine (2020): 1-7.

[5] Giordano, Giulia, et al. "Modelling the COVID-19 epidemic and implementation of population-wide interventions in Italy." Nature Medicine (2020): 1-6.

[6] Peng, Liangrong, et al. "Epidemic analysis of COVID-19 in China by dynamical modeling." arXiv preprint arXiv:2002.06563 (2020).

[7] Chen, Tian-Mu, et al. "A mathematical model for simulating the phase-based transmissibility of a novel coronavirus." Infectious diseases of poverty 9.1 (2020): 1-8.

[8] Fernandes, Nuno. "Economic effects of coronavirus outbreak (COVID-19) on the world economy." Available at SSRN 3557504 (2020).

[9] Centers for Disease Control and Prevention. Interim Guidelines for Collecting, Handling, and Testing Clinical Specimens from Persons Under Investigation (PUIs) for Coronavirus Disease 2019 (COVID-19). 2020. www.cdc.gov/coronavirus/2019-ncov/lab/guidelines-clinical-specimens.html. Published February 14, 2020. Accessed April 14, 2020.

[10] Ai, Tao, 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.

[11] Fang, Yicheng, et al. "Sensitivity of chest CT for COVID-19: comparison to RT-PCR." Radiology (2020): 200432.

[12] Wang, Wenling, et al. "Detection of SARS-CoV-2 in different types of clinical specimens." Jama (2020).

[13] Shi, Heshui, et al. "Radiological findings from 81 patients with COVID-19 pneumonia in Wuhan, China: a descriptive study." The Lancet Infectious Diseases (2020).

[14] Revel, Marie-Pierre, et al. "COVID-19 patients and the Radiology department–advice from the European Society of Radiology (ESR) and the European Society of Thoracic Imaging (ESTI)."

[15] Bernheim, Adam, et al. "Chest CT findings in coronavirus disease-19 (COVID-19): relationship to duration of infection." Radiology (2020): 200463.

[16] Liu, Xiaoxuan, et al. "A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging: a systematic review and meta-analysis." The lancet digital health 1.6 (2019): e271-e297.

[17] Ardila, Diego, et al. "End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography." Nature medicine 25.6 (2019): 954-961.

[18] Coudray, Nicolas, et al. "Classification and mutation prediction from non–small cell lung cancer histopathology images using deep learning." Nature medicine 24.10 (2018): 1559-1567.

[19] Fourcade, A., and R. H. Khonsari. "Deep learning in medical image analysis: A third eye for doctors." Journal of stomatology, oral and maxillofacial surgery 120.4 (2019): 279-288.

[20] www.kaggle.com/c/rsna-pneumonia-detection-challenge/overview/description

[21] Rajpurkar, Pranav, et al. "Chexnet: Radiologist-level pneumonia detection on chest x-rays with deep learning." arXiv preprint arXiv:1711.05225 (2017).

[22] Wang, Xiaosong, et al. "Chestx-ray8: Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases." Proceedings of the IEEE conference on computer vision and pattern recognition. 2017.

[23] Rajpurkar, Pranav, et al. "Deep learning for chest radiograph diagnosis: A retrospective comparison of the CheXNeXt algorithm to practicing radiologists." PLoS medicine 15.11 (2018): e1002686.

[24] Irvin, Jeremy, et al. "Chexpert: A large chest radiograph dataset with uncertainty labels and expert comparison." Proceedings of the AAAI Conference on Artificial Intelligence. Vol. 33. 2019.

[25] Bien, Nicholas, et al. "Deep-learning-assisted diagnosis for knee magnetic resonance imaging: development and retrospective validation of MRNet." PLoS medicine 15.11 (2018): e1002699.

[26] Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. "Imagenet classification with deep convolutional neural networks." Advances in neural information processing systems. 2012.

[27] Li, Lin, et al. "Artificial Intelligence Distinguishes COVID-19 from Community Acquired Pneumonia on Chest CT." Radiology (2020): 200905.

[28] Shi, Feng, et al. "Review of artificial intelligence techniques in imaging data acquisition, segmentation and diagnosis for covid-19." IEEE Reviews in Biomedical Engineering (2020).

[29] Corman, Victor M., et al. "Detection of 2019 novel coronavirus (2019-nCoV) by real-time RT-PCR." Eurosurveillance 25.3 (2020): 2000045.

[30] Deng, Jia, et al. "Imagenet: A large-scale hierarchical image database." 2009 IEEE conference on computer vision and pattern recognition. Ieee, 2009.

[31] Huang, Gao, et al. "Densely connected convolutional networks." Proceedings of the IEEE conference on computer vision and pattern recognition. 2017.

[32] He, Kaiming, et al. "Deep residual learning for image recognition." Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.

[33] Chollet, François. "Xception: Deep learning with depthwise separable convolutions." Proceedings of the IEEE conference on computer vision and pattern recognition. 2017.

[34] Tan, Mingxing, and Quoc V. Le. "Efficientnet: Rethinking model scaling for convolutional neural networks." arXiv preprint arXiv:1905.11946. (2019).

[35] Bianco, Simone, et al. "Benchmark analysis of representative deep neural network architectures." IEEE Access 6 (2018): 64270-64277.

[36] Van Rossum, Guido, and Fred L. Drake. "PYTHON 3 Reference Manual." (2009).

[37] Chollet, François. "keras." (2015).

[38] Mason, Darcy. "SU-E-T-33: pydicom: an open source DICOM library." Medical Physics 38.6Part10 (2011): 3493-3493.

[39] Scikit-learn: Machine Learning in Python, Pedregosa et al., JMLR 12, pp. 2825-2830, 2011.