

dx.doi.org/10.17488/RMIB.43.1.1

E-LOCATION ID: 1208

Detection of COVID-19 Lung Lesions in Computed Tomography Images Using Deep Learning

Detección de lesiones pulmonares por COVID-19 en imágenes de tomografía computarizada mediante aprendizaje profundo

Joy Ingrid Arreola Minjarez , José David Díaz Román , Boris Jesús Mederos Madrazo , José Manuel Mejía Muñoz , Lidia Hortencia Rascón Madrigal , Juan de Dios Cota Ruiz 

Universidad Autónoma de Ciudad Juárez

ABSTRACT

The novel coronavirus (COVID-19) is a disease that mainly affects the lung tissue. The detection of lesions caused by this disease can help to provide an adequate treatment and monitoring its evolution. This research focuses on the binary classification of lung lesions caused by COVID-19 in images of computed tomography (CT) using deep learning. The database used in the experiments comes from two independent repositories, which contains tomographic scans of patients with a positive diagnosis of COVID-19. The output layers of four pre-trained convolutional networks were adapted to the proposed task and re-trained using the fine-tuning technique. The models were validated with test images from the two database's repositories. The model VGG19, considering one of the repositories, showed the best performance with 88% and 90.2% of accuracy and recall, respectively. The model combination using the soft voting technique presented the highest accuracy (84.4%), with a recall of 94.4% employing the data from the other repository. The area under the receiver operating characteristic curve was 0.92 at best. The proposed method based on deep learning represents a valuable tool to automatically classify COVID-19 lesions on CT images and could also be used to assess the extent of lung infection.

KEYWORDS: Lung Lesions, Classification, Deep Learning, Computed Tomography

Corresponding author

TO: José David Díaz Román

INSTITUTION: Universidad Autónoma de Ciudad Juárez

ADDRESS: Av. Plutarco Elías Calles #1210,

Col. Fovissste Chamizal, C.P. 32310, Ciudad Juárez,

Chihuahua, México

CORREO ELECTRÓNICO: david.roman@uacj.mx

Received:

5 September 2021

Accepted:

17 January 2022

INTRODUCTION

Coronavirus disease 2019 (COVID-19) is caused by the severe acute respiratory syndrome coronavirus type 2 (SARS-CoV-2). It primarily affects the human respiratory system and represents the seventh member of the coronavirus family that infects humans [1]. The first case, identified as viral pneumonia until then, appeared in late December 2019 in Wuhan, China. According to the records issued by the World Health Organization, until September 1st, 3,341,264 cases have been registered in Mexico and 217,558,771 around the world [1] [2] [3]. It is known that the COVID-19 infection has an incubation period from 1 to 14 days, which varies depending on some human characteristics like the status of the immune system and the age [1]. In Mexico, coronavirus cases are classified by stages according to their severity, clinical stage and signs presented: stage 1 (early infection), stage 2 (pulmonary stage) and stage 3 (hyperinflammatory stage) [4].

The reverse transcription - polymerase chain reaction (RT-PCR) tests represent the main method to detect COVID-19, providing results with a specificity close to 100% [5]; however, when using this standard test as a reference, some drawbacks must be considered. For example, a low sensitivity (59% - 79%) has been observed during the early phase of the disease [5] [6] [7]. Due to the continuous evolution and genetic diversity that the new coronavirus has presented, the results of clinical tests can be affected by the variation in the viral ribonucleic acid (RNA) sequence [8]. Also, it is important to remark that the diagnostic period can vary from 5 to 72 hours [9].

The study presented by Uysal *et al.* [10] found that 25% of asymptomatic patients, diagnosed with an RT-PCR test, did not show signs of lesions on their computed tomography (CT) scans, while the rest showed abnormal findings associated with lesions similar to those in patients with symptoms. The most common signs were ground glass opacity (GGO), pure or with consolidation or crazy-paving patterns. Thus, due those find-

ings some authors emphasize over the importance of performing RT-PCR tests in conjunction with imaging procedures such as CT to increase the accuracy of the diagnosis, injury identification, and in this way provide an adequate patient management [11].

To confirm the coronavirus disease, the chest CT in conjunction with clinical manifestations and the epidemiological evidence have become a fundamental diagnostic tool. However, discrepancies have been reported between the results of laboratory tests and the characteristics observed in diagnostic images [12].

Recently, some studies have shown that the CT scan of patients (asymptomatic or those in whom the result of RT-PCR test was negative) depicts abnormal signs that can be useful for the disease detection, where these studies have reported a sensitivity between 88% and 98% [5] [6] [13] [14]. The advantage of CT diagnosis lies in its short exploration time and the high resolution of the acquired image, useful for detecting and classifying lung lesions.

At present, most of the expert researchers in the clinical applications of Artificial Intelligence (AI) have focused on the diagnosis of patients with COVID-19 through the processing of medical images, addressing the analysis of findings observed in chest x-rays and/or CT scans [15] [16]. There are several approaches that aim to take advantage of machine learning (ML), especially deep learning, to diagnose CT scans using binary pathway convolutional neural networks (CNN) (positive vs. negative) or multiple classification (healthy vs. COVID-19 versus other types of pneumonia) [16]. An example of this is the COVNet architecture performed by Li *et al.*, which classifies positive results for COVID-19, community acquired pneumonia or negative for any lung disease through a three-dimensional CNN constituted by the ResNet50 architecture, resulting in 90% of sensitivity and a specificity of 96% [17]. Similarly, Yang *et al.* in [13], built a publicly available database of

CT scans of COVID-19 patients that could be used to train deep learning models. This database was subsequently used to develop an algorithm to classify COVID-19 patients in a binary way, obtaining an accuracy of 83% and an area under the receiver operating characteristic curve (AUC-ROC) of 0.95. Other work that uses deep learning techniques developed a model called CTnet-10 obtaining an accuracy of 82.1%. The authors also tested models such as DenseNet169, VGG16, ResNet50, InceptionV3 and VGG19, obtaining an accuracy of 94.52% with the latest network [18]. On the other hand, in [19] the authors attempted to segment lung lesions associated with COVID-19, reaching specificity values of up to 100% in specific tasks and models tested, but with a very low sensitivity (between 1.2% and 64.8%).

As mentioned before, a large percentage of asymptomatic patients already have abnormal findings on their CT scan images whose lesion patterns are similar to those found in symptomatic patients. In this sense, it is very important to detect these patterns in CT images to allow physicians to know if a patient has lung lesions and thus guide their treatment.

The purpose of this investigation is to detect the presence or absence (i.e., a binary classification) of lung lesions due to COVID-19 in images originated from chest CT studies using deep learning. It could be useful when it is desired to identify whether the lesions are disseminated in a large part of the lung tissue, indicating that the lesions occur in many slices of the CT study; this detection can even be valuable in assessing the evolution of lung tissue damage, and thus provide adequate treatment to the patients.

MATERIALS AND METHODS

The database used in this research corresponds to “COVID-19 CT Lung and Infection Segmentation Dataset” [20]. The images are in NIfTI (Neuroimaging Informatics Technology Initiative) format and were

prepared through the collection of 20 public CT scans of patients with COVID-19 belonging to the Coronacases Initiative and Radiopaedia repositories. All cases present COVID-19 infection in the lungs; however, the percentage of slices per patient showing abnormal findings (related to infection) ranges from 0.01% to 59%. Abnormal findings on the chest CTs are: GGO, lung consolidation, pleural effusion, and mixed GGO with crazy-paving pattern or consolidation.

Figure 1 shows different patterns of abnormal findings present in the images of the database: a) GGO, b) consolidation, c) pleural effusion, d) GGO with crazy-paving pattern, and e) GGO with consolidation, where GGO is indicated with green arrows, consolidation is surrounded by segmented red ovals, pleural effusion is pointed with a yellow arrow, and crazy-paving pattern is enclosed by a blue line (also indicated by the blue arrow). In Figure 1, the images (a, b, c) belong to the Coronacases Initiative repository, and the images (d, e) correspond to the Radiopaedia repository.

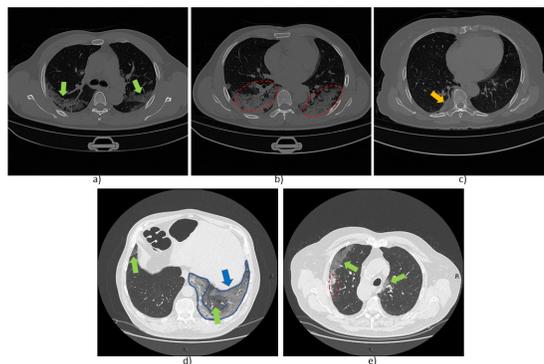


FIGURE 1. Images with different abnormal findings from CT scans of the database. GGO is indicated with green arrows in a), d) and e); consolidation is enclosed by segmented red ovals in b) and e); pleural effusion is indicated with yellow arrow in c); and crazy-paving pattern is enclosed by a blue line in d).

There are images that present inconspicuous abnormalities that could be challenging for both an inexperienced radiologist and an automatic detection model.

For instance, Figure 2 shows an example of two images from a CT scan of the same patient. The slice in a) shows slight evidence of GGO, while in b) no abnormalities or lesions are observed. Thus, the detection system must be able to identify negligible lesions that commonly appear at the early stage of the disease.

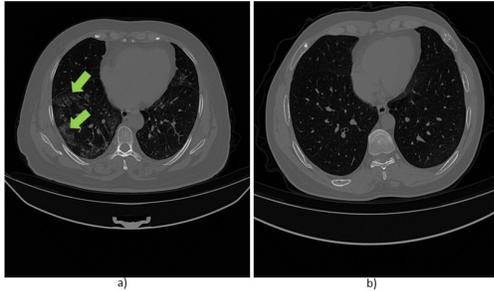


FIGURE 2. Images obtained from a CT scan of a patient of the database. a) Slice with lesion (GGO indicated with green arrows), b) slice without lesion.

The resolution for the x and y axes is 512x512 pixels for the scans obtained from the Coronacases Initiative repository and 630x630 pixels for the Radiopaedia repository, except for case 5 "radiopaedia_14_85914_0" with 630x401 pixels. The CT scans have between 39 and 418 slices, with a total of 3,520 images. The database was grouped by counting images with lesions due to COVID-19 infection and without lesions, obtaining a total of 1,844 and 1,676, respectively.

The database of 20 patient scans was divided into training (80%, N = 16) and test (20%, N = 4) sets. The data were partitioned in such a way that there were the same number of cases from the Coronacases Initiative and Radiopaedia repositories in the training and test sets. The total of images (slices) was 3020 for training and 500 for testing. The purpose of this division was to have a balanced number of images between slices with and without lesions in both sets. A 15% (N = 483) of the training data was considered for internal validation during the training phase of the models used. Table 1 shows the division of the data set into subsets: training, validation, and testing.

TABLE 1. Number of images for the training, validation, and test subsets (Coronacases + Radiopaedia repositories).

Images	With lesions	Without lesions	Total
Training	1339	1198	2537
Internal validation	255	228	483
Test	250	250	500
Total	1844	1676	3520

The CT volumes belonging to the Radiopaedia database were previously pre-processed with a pulmonary window [-1250, 250] ^[19]. The image format was converted from NIFTI to 8-bit grayscale png (Portable Network Graphics). The pixel values were normalized from [0-255] to [0-1]. After normalization, a resizing was applied to finally have images of size 128x128 pixels (or 331x331 in the case of one of the networks used).

Implementation of convolutional neural networks

The algorithm was developed in Python. The implementation of the network models was carried out by means of transfer learning and subsequent fine-tuning. Transfer learning is a technique that takes advantage of existing knowledge to solve problems from a source domain to a destination domain in which, although the same task is not performed, both tasks have a certain similarity. Thus, the purpose is to solve a learning problem using the knowledge acquired by solving similar tasks ^[21]. On the other hand, the fine-tuning process applied in the context of deep learning model training is a way of applying learning transfer, but especially it consists of fine-tuning the weights of the pretrained model to fit to new observations. Transfer learning and fine-tuning techniques have been used in other investigations to identify and / or differentiate patients with COVID-19 from patients without pulmonary pathology or with pneumonia using chest x-ray images, where this methodology has

provided accuracy values between 89% and 99% [22] [23] [24]. Also, Perumal *et al.* used the learning transfer technique with pre-trained models of the ResNet50, VGG16 and InceptionV3 networks to differentiate patients with COVID-19, viral and bacterial pneumonia, and healthy patients. In their models, they combined CT images and chest x-ray images where the best performance was achieved with the VGG16 model with an accuracy of 93% [25].

In this work, the transfer learning technique was implemented using four pretrained models belonging to the following networks: ResNet50 (RN50) [26], VGG16 [27], InceptionResNetV2 (IRNV2) [28], and NASNetLarge (NNL) [29]. These networks were chosen due the well performance in large scale image recognition tasks, and their architectures and weights of pretrained networks are publicly available. Likewise, these networks have been used in numerous medical image classification applications [18] [22] [23] [24] [30]. For each of these networks, they employed weights obtained from training using data from the ImageNet repository [31]. ImageNet corresponds to a dataset widely used for object recognition purposes. Figure 3 shows the general configuration of the architectures used for the construction of each model. The last fully connected layer from each base model (used for ImageNet data classification) was excluded, and the top of the architecture was configured to classify only two classes as follows: An average subsampling layer (GlobalAveragePooling2D) was included, followed by a fully connected dense layer (Dense) of size 1024 with a ReLu activation function and, finally, a Dense layer with two neurons (one for each class) with a Softmax activation function. The input dimension was set to 128x128 pixels, except for the IRNV2 architecture that used 331x331 pixels.

Initially, only the last added Dense layers of each model (the layers enclosed in the red segmented box in Figure 3) were trained for 100 epochs using the ADAM optimizer with a learning rate of 0.001. The

categorical crossentropy loss was used as cost function. The weights of the trained models were saved for testing and subsequent training.

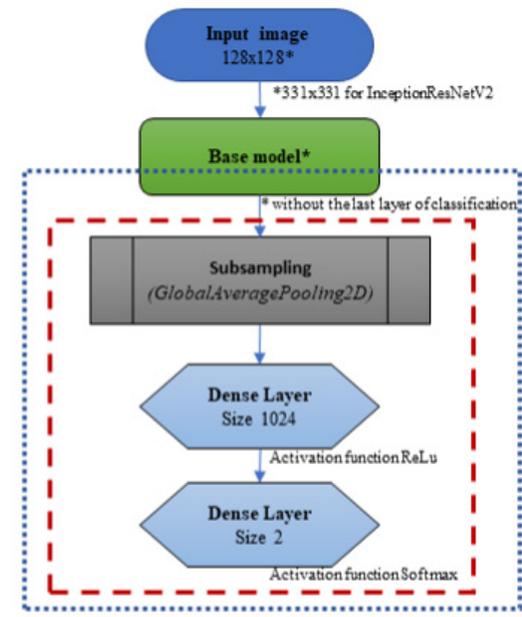


FIGURE 3. General diagram of the architecture of the models used in the training process.

Once the training process considering the last layers of the models was carried out, the fine-tuning (F-T) technique was applied, unfreezing a certain number of layers at the end of the base model (of each network) for training together with the layers trained in the previous stage; In Figure 3, the layers enclosed by the dotted blue box are those involved in the fine-tuning of the models. For the RN50 architecture, the layers were unfrozen from the fifth convolutional block onwards; for the VGG16, the unfrozen of the layers started from the fourth block; in the IRNV2, it started from the layer 547th; and in the NNL architecture, it started from the layer 902nd. For training with F-T, the learning rate of the optimizer was decreased to 0.0001 using the ADAM optimizer. In summary, eighth models were built, two for each network (models without and with F-T). Also, with the goal to observe the effect caused by a different optimizer in the learning process, the stochastic descending gradient (SGD) opti-

mizer was applied to the VGG16 network during the fine-tuning phase. Finally, the models were evaluated using five metrics: accuracy (Acc), recall (RE), specificity (SP), F1-Score (F1) and AUC-ROC.

Majority voting ensemble

In terms of classification, majority voting (hard voting) is an ensemble machine learning model that combines the predictions of multiple models. It seeks to optimize the performance of the classification based on consensus, which takes into account the sum of the votes of independent models. The hard voting ensemble used the five models that had the highest accuracy in the validation set: RN50, VGG16, IRNV2, IRNV2 with fine-tuning and NNL without fine-tuning.

A variant of the hard voting arrangement is the ensemble of soft voting, which seeks to get a classification based on the probability values of belonging to a class given by the classifiers used. The labeling (0 or 1) is done after all the probabilities of the models have been considered. The models used for this ensemble are the same as those ones used in the hard voting scheme.

Creation and evaluation of models with separate data repositories

As an additional experimentation, the division of the data by sources (Coronacases Initiative and Radiopaedia repositories) was proposed in order to assess the performance when evaluating the models with the data from the repositories separately. For these tests, both repositories were inspected with the intention of finding and removing low-quality images. In this process, 77 slices with high opacity were excluded, possibly due to an inadequate reconstruction of the tomographic image. Finally, there were 3443 images where 2504 belong to Coronacases and 939 to Radiopaedia. The organization of the training and test sets for the case of Coronacases consisted of using eight scans (2,080 slices) and two scans (424 slices) for training and testing, respectively. In the

case of Radiopaedia, seven scans were used for training (792 slices) and three for testing (147 slices). For both cases, 15% of the training data was considered for internal validation. Tables 2 and 3 specify the number of cases for training, validation and testing of the models with the separate data repositories.

TABLE 2. Number of images for the training, validation, and test subsets for the Coronacases repository.

Images	With lesions	Without lesions	Total
Training	921	847	1768
Internal validation	162	150	312
Test	205	219	424
Total	1288	1216	2504

TABLE 3. Number of images for the training, validation, and test subsets for the Radiopaedia repository.

Images	With lesions	Without lesions	Total
Training	354	312	666
Internal validation	67	59	126
Test	72	75	147
Total	493	446	939

The same criteria (the five models that had the highest accuracy in the validation set) were used in the selection of models for the hard voting and soft voting ensembles. For Coronacases, the best models were RN50, VGG16, VGG16 with fine-tune, IRNV2 with fine-tune and NNL. In the case of Radiopaedia, RN50, IRNV2, NNL, VGG16 (all the above with fine-tune) and VGG16 with fine-tune and the SGD optimizer were used.

RESULTS AND DISCUSSION

Table 4 shows the performance of the trained architectures in their different stages (without and with F-T). They were evaluated with the test set of the "COVID-19 CT Lung and Infection Segmentation" data-

base (including both repositories). It is observed that the VGG16 architecture presents the highest values of accuracy and recall, however, the specificity metric is slightly lower than other models, and the Inception-ResNetV2 network presents the best performance in this metric. The VGG16 model has the highest F1-score value above 81%, followed by the model implemented with soft voting. The highest value of the AUC-ROC (0.880) is also obtained by the VGG16 model.

TABLE 4. Performance of the models evaluated with the test set (Coronacases + Radiopaedia), where the best performances of the models are bold remarked.

Model	Acc (%)	RE (%)	SP (%)	F1 (%)	AUC-ROC
RN50	70.8	77.6	64.0	72.7	0.749
RN50 + F-T	68.5	72	64.8	69.5	0.733
VGG16	79.6	93.6	63.6	81.4	0.880
VGG16 + F-T	74.8	81.6	68.0	76.4	0.863
VGG16 + F-T/SGD	77.4	86.0	68.8	79.2	0.872
IRNV2	79.0	77.6	80.4	78.7	0.862
IRNV2 + F-T	76.6	81.6	71.6	77.7	0.826
NNL	73.0	78.0	68.0	74.3	0.818
NNL + F-T	77.0	81.6	72.4	78.0	0.822
Hard Voting	78.0	84.0	72.0	79.2	--
Soft Voting	78.6	84.4	72.8	79.8	0.867

A good classification of CT slices with COVID-19 lesions is observed, with an accuracy equal to or greater than 78% in four of the nine models evaluated. It is also important to mention that only one of the models presents a SP greater than 80%, which indicates that in most models, there is a tendency to misclassify the negative class (images without lesions).

Figure 4 shows the accuracy performance of the nine models, which were trained with the training set (that includes data from both repositories) but evaluated

with the test sets of each repository independently. For purposes of better identification, data from the Coronacases Initiative repository is named as DB1 and data from Radiopaedia named as DB2. As it is observed in Figure 4, the accuracy in the classification of the DB1 images was superior in six of the nine models evaluated.

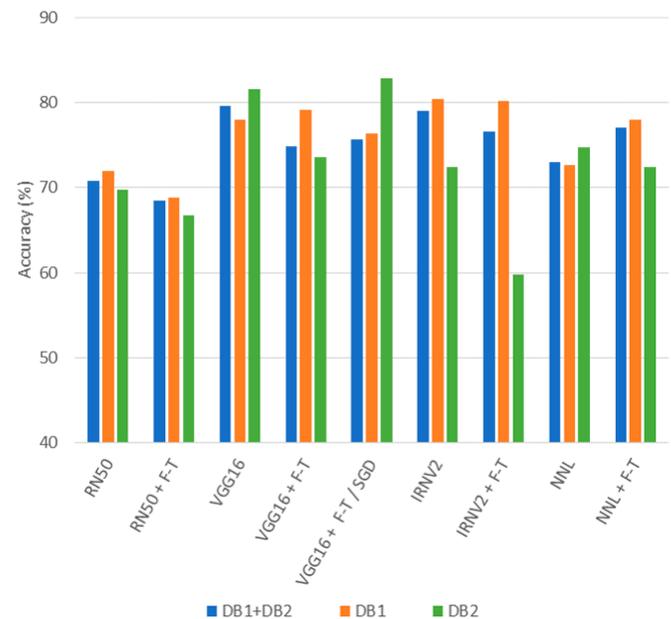


FIGURE 4. Accuracy of the models evaluated with a mixed set of data (DB1+DB2), and independent data sets (DB1 and DB2).

It can be observed from Figure 4 that the models do not show a consistent fit to the data from both repositories separately; this may be due to the lung window preprocessing previously applied to the images of the Radiopaedia repository.

Table 5 presents the performance evaluation of the models trained only using data from the Coronacases repository. The VGG16 architecture (without F-T) shows the best performance with 88% of accuracy, 90.2% of recall, a specificity greater than 85% and an F1-score above 87% (AUC-ROC of 0.929). The IRNV2 + F-T network presents a good assessment in all the metrics evaluated, just below, in average, to the VGG16

model. On the other hand, even though the ResNet50 network obtains an excellent recall of 98.5%, its specificity is around 57% making it unreliable to classify cases without lesions.

TABLE 5. Evaluation of the models with data from the Coronacases Initiative repository.

Model	Acc (%)	RE (%)	SP (%)	F1 (%)	AUC-ROC
RN50	77.4	98.5	57.5	80.8	0.912
RN50 + F-T	76.4	86.8	66.7	78.0	0.806
VGG16	88.0	90.2	85.8	87.8	0.929
VGG16 + F-T	80.4	93.7	68.0	82.3	0.936
VGG16 + F-T/SGD	80.2	96.6	64.8	82.5	0.925
IRNV2	75.9	69.3	82.2	73.6	0.861
IRNV2 + F-T	85.9	92.7	79.5	86.4	0.899
NNL	75.7	87.8	64.4	77.8	0.871
NNL + F-T	65.1	60.0	69.9	62.4	0.782
Hard Voting	81.6	93.7	70.3	83.1	--
Soft Voting	80.2	92.7	68.5	81.9	0.925

Finally, Table 6 shows the performance evaluation of the models trained only using data from the Radiopaedia repository. As observed, the model built with the soft voting ensemble presents the best performance, reaching an accuracy of 84.4%, a high recall of 94.4% and the highest F1-score of 85.5%, with moderate specificity higher than 74%. The models VGG16 and VGG16+F-T (using the ADAM optimizer), obtain the greatest specificity compared with the rest of the networks; however, they present a low recall making such models not appropriate to detect cases with lung lesions in CT images.

In general terms, training and testing with separated data repositories show a better performance in the models evaluated in this study, which is evidenced by the maximum accuracy values obtained with the

Coronacases repository (88% in the VGG16 model), and Radiopaedia repository (84.4% in the soft voting model) when compared with the models trained using data from both repositories together (79.6% for the VGG16 model).

TABLE 6. Evaluation of the models using data from the Radiopaedia repository.

Model	Acc (%)	RE (%)	SP (%)	F1 (%)	AUC-ROC
RN50	73.5	91.7	56.0	77.2	0.873
RN50 + F-T	79.6	87.5	72.0	80.8	0.914
VGG16	78.2	66.7	89.3	75.0	0.892
VGG16 + F-T	80.3	73.6	86.6	78.5	0.883
VGG16 + F-T/SGD	79.6	83.3	76.0	80.0	0.887
IRNV2	77.6	88.9	66.7	79.5	0.841
IRNV2 + F-T	78.9	94.4	64.0	81.4	0.874
NNL	73.5	83.3	64.0	75.5	0.845
NNL + F-T	74.8	84.7	65.3	76.7	0.840
Hard Voting	83.0	91.7	74.7	84.1	--
Soft Voting	84.4	94.4	74.7	85.5	0.920

Other investigations that seek to identify the presence of lesions on CT images using transfer learning have reported an accuracy of 99%. Such is the case of Ahuja *et al.* [32], who used different versions of the ResNet and the SqueezeNet networks; They worked with a data set of 746 images of which 349 showed signs of COVID-19 lesions, obtained from 216 patients. However, unlike to our research, where all images from the CT studies were used and all patients had the disease, they did not use the full CT study for their experiments, just a few selected images of patients with the infection. In a similar task, Dey *et al.* [33] used an algorithm based on a segmentation and feature extraction scheme in CT images to detect COVID-19 lesions. Testing different classifiers, its algorithm reached a maximum accuracy of 87.75%.

It is important to mention that the studied networks presented a high classification error or misclassification in slices that were located at the beginning or at the end of the scans (at the cephalocaudal ends). It could be due to the fact that these images show a reduced area of lung tissue while the rest of the tissue can generate structures similar to abnormality findings suggesting a pulmonary lesion. An example of this issue can be seen in Figure 5. Here, two cases of slices located in the apex region of the upper lobes of the lungs are observed. The image in a) shows signs of consolidation in the left lung (enclosed with a segmented red oval), and the image in b) does not show signs of abnormality, however, in both cases the models classify the images with the presence of lesions.

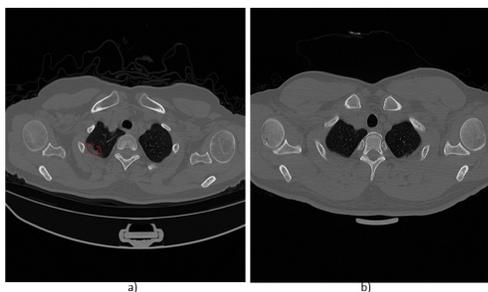


FIGURE 5. CT images from the apex region of the upper lobes of the lungs. a) CT slice that presents consolidation in the right lung (enclosed with a red segmented red oval), b) CT slice that does not show abnormal signs.

CONCLUSIONS

The objective of the present work was to detect the presence or absence of lung lesions in chest computed tomography images of patients with COVID-19 infection using deep learning models. In our study, the VGG16 model using the Coronacases Initiative repository presented the best results with an accuracy of 88%, AUC-ROC of 0.929 and F1-score of 87.8%. On the other hand, the soft voting ensemble, using the Radiopaedia repository, reached an accuracy of 84.4%, AUC-ROC of 0.92 and F1-score of 85.5%. The results of both models represent a good trade-off between the recall, specific-

ity and precision of the classifiers. It should be remarked that the management of the repositories, used independently of each other, improved the adjustment of the models, showing a greater generalization.

The model VGG16 with FT reached an accuracy of 80.3% using the Radiopaedia repository, however this performance was improved using combination models such as the soft voting and hard voting ensembles, with 84.4% and 83% of accuracy, respectively (both models with a high recall). It must be noted that this combination scheme was only satisfactory for this repository.

This research demonstrates that deep learning models can be useful to detect lung lesions of COVID-19 with high sensitivity and specificity for diagnosis; it can be valuable when considering the possible high false positive rate of clinical tests. In this way, an automatic detection model can serve as reference in radiology, allowing a quick localization of the lesion from a CT study with greater precision.

We must emphasize that in the present research, all the CT scans of the database included patients with a positive diagnosis of COVID-19, so the abnormality patterns found in the images are assumed to be indicative of lesions due to this disease. This represents a limitation in the present study since certainly other lung diseases such as interstitial pneumonia, sarcoidosis, alveolar proteinosis, carcinoma, etc., can produce similar patterns in CT scans to those found in patients with COVID-19 [10] [34]. Therefore, as future work, it is necessary to advance in this research to include patients with different lung diseases and classify the lesions according to their pathology of origin.

AUTHOR CONTRIBUTIONS

J.A.M. contributed to the writing of the original draft of the manuscript, performed data curation, organization, and annotation, and performed the experiments. J.D.D. conceptualized the project, designed, and devel-

oped the methodology, participated in the design of specialized software, carried out statistical analysis and writing of the manuscript. B.M.M. contributed to the design and development of the experiments and participated in the programming of the software for the implementation of deep learning models. J.M.M. contributed to the implementation of the computer algorithms, tested the codes for the reproducibility of the results and verified the organization of the data for the training, validation and testing of the imple-

mented models. L.R.M. participated in all the writing stages of the manuscript (preparation of the original draft, review, and edition of the final version), elaborated the images, carried out statistical analysis and contributed providing material and computer resources. J.C.R. supervised the development of the methodology, contributed to writing the draft and final version of the manuscript, designed the statistical tests and analysis. All authors reviewed and approved the final version of the manuscript.

REFERENCES

- [1] Rothan H, Byrareddy S. The epidemiology and pathogenesis of coronavirus disease (COVID-19) outbreak. *J Autoimmun* [Internet]. 2020;109:102433. Available from: <https://doi.org/10.1016/j.jaut.2020.102433>
- [2] Palacios Cruz M, Santos E, Velázquez Cervantes MA, León Juárez M. COVID-19, a worldwide public health emergency. *Rev Clin Esp* [Internet]. 2021;221(1):55-61. Available from: <https://doi.org/10.1016/j.rce.2020.03.001>
- [3] World Health Organization. Coronavirus Disease (COVID-19) Dashboard [Internet]. WHO Coronavirus Disease (COVID-19) Dashboard; 2021. Available from: <https://covid19.who.int/>
- [4] Instituto Mexicano del Seguro Social. Algoritmos interinos para la atención del COVID-19. Gobierno de México [Internet]. 2020; 1-31. Available from: http://educacionensalud.imss.gob.mx/es/system/files/Algoritmos_interinos_COVID19_CTEC.pdf
- [5] He JL, Luo L, Luo ZD, Lyu JX, et al. Diagnostic performance between CT and initial real-time RT-PCR for clinically suspected 2019 coronavirus disease (COVID-19) patients outside Wuhan, China. *Respir Med* [Internet]. 2020;168:105980. Available from: <https://doi.org/10.1016/j.rmed.2020.105980>
- [6] Ai T, Yang Z, Hou H, Zhan C, et al. Correlation of Chest CT and RT-PCR Testing for Coronavirus Disease 2019 (COVID-19) in China: A Report of 1014 Cases. *Radiology* [Internet]. 2020;296(2):E32-40. Available from: <https://doi.org/10.1148/radiol.2020200642>
- [7] Rubin GD, Ryerson CJ, Haramati LB, Sverzellati N, et al. The Role of Chest Imaging in Patient Management During the COVID-19 Pandemic: A Multinational Consensus Statement From the Fleischner Society. *Radiology* [Internet]. 2020;296(1):172-180. Available from: <https://doi.org/10.1148/radiol.2020201365>
- [8] Shen M, Zhou Y, Ye J, AL-maskri AAA, et al. Recent advances and perspectives of nucleic acid detection for coronavirus. *J Pharm Anal* [Internet]. 2020;10(2):97-101. Available from: <https://doi.org/10.1016/j.jpha.2020.02.010>
- [9] Araujo Oliveira B, Campos de Oliveira L, Cerdeira Sabino E, Okay TS. SARS-CoV-2 and the COVID-19 disease: A mini review on diagnostic methods. *Rev Inst Med Trop Sao Paulo* [Internet]. 2020;62:e44. Available from: <https://doi.org/10.1590/S1678-9946202062044>
- [10] Uysal E, Kiliçer A, Cebeci H, Özer H, et al. Chest CT findings in RT-PCR positive asymptomatic COVID-19 patients. *Clin Imaging* [Internet]. 2021;77:37-42. Available from: <https://doi.org/10.1016/j.clinimag.2021.01.030>
- [11] Tahamtan A, Ardebili A. Real-time RT-PCR in COVID-19 detection: issues affecting the results. *Expert Rev Mol Diagn* [Internet]. 2020;20(5):453-4. Available from: <https://doi.org/10.1080/14737159.2020.1757437>
- [12] Li X, Zeng W, Li X, Chen H, et al. CT imaging changes of coronavirus disease 2019(COVID-19): A multi-center study in Southwest China. *J Transl Med* [Internet]. 2020;18:154. Available from: <https://doi.org/10.1186/s12967-020-02324-w>
- [13] Yang X, He X, Zhao J, Zhang Y, et al. COVID-CT-Dataset: A CT Scan Dataset about COVID-19. arXiv:2003.13865 [Preprint]. 2020. Available from: <https://arxiv.org/abs/2003.13865>
- [14] Bernheim A, Mei X, Huang M, Yang Y, et al. Chest CT findings in coronavirus disease 2019 (COVID-19): Relationship to duration of infection. *Radiology* [Internet]. 2020;295(3):685-91. Available from: <https://doi.org/10.1148/radiol.2020200463>
- [15] Naudé W. Artificial intelligence vs COVID-19: limitations, constraints and pitfalls. *AI Soc* [Internet]. 2020;35(3):761-765. Available from: <https://doi.org/10.1007/s00146-020-00978-0>
- [16] Bullock J, Luccioni A, Pham KH, Lam C, et al. Mapping the Landscape of Artificial Intelligence Applications against COVID-19. arXiv:2003.11336 [Preprint]. 2020;1-32. Available from: <http://arxiv.org/abs/2003.11336>
- [17] Li L, Qin L, Xu Z, Yin Y, et al. Using Artificial Intelligence to Detect COVID-19 and Community-acquired Pneumonia Based on Pulmonary CT: Evaluation of the Diagnostic Accuracy. *Radiology* [Internet]. 2020;296(2):E65-71. Available from: <https://doi.org/10.1148/radiol.2020200905>
- [18] Shah V, Keniya R, Shridharani A, Punjabi M, et al. Diagnosis of COVID-19 using CT scan images and deep learning techniques. *Emerg Radiol* [Internet]. 2021;28: 497-505. Available from: <https://doi.org/10.1007/s10140-020-01886-y>
- [19] Ma J, Wang Y, An X, Ge C, et al. Toward data-efficient learning: A benchmark for COVID-19 CT lung and infection segmentation. *Med Phys* [Internet]. 2021;48(3):1197-1210. Available from: <https://doi.org/10.1002/mp.14676>
- [20] Jun M, Cheng G, Yixin W, Xingle A, et al. COVID-19 CT Lung and Infection Segmentation Dataset [Data set]. Zenodo. 2020. Available from: <https://doi.org/10.5281/zenodo.3757476>
- [21] Karimpanal TG, Bouffanais R. Self-organizing maps for storage and transfer of knowledge in reinforcement learning. *Adapt Behav* [Internet]. 2019;27(2):111-126. Available from: <https://doi.org/10.1177%2F1059712318818568>
- [22] Nefoussi S, Amamra A, Amarouche IA. A Comparative Study of Deep Learning Networks for COVID-19 Recognition in Chest X-ray Images. In 2020 2nd International Workshop on Human-Centric Smart Environments for Health and Well-being (IHSH) [Internet]. Boumerdes: IEEE; 2021:237-41. Available from: <https://doi.org/10.1109/IHSH51661.2021.9378703>
- [23] Shazia A, Xuan ZT, Chuah JH, Usman J, et al. A comparative study of multiple neural network for detection of COVID-19 on chest X-ray. *EURASIP J Adv Signal Process* [Internet]. 2021;2021(1):50. Available from: <https://doi.org/10.1186/s13634-021-00755-1>
- [24] Perumal, V, Narayanan V, Rajasekar SJS. Detection of COVID-19 using CXR and CT images using Transfer Learning and Haralick features. *Appl Intell* [Internet]. 2021;51:341-358. Available from: <https://doi.org/10.1007/s10489-020-01831-z>
- [25] Rahaman MM, Li C, Yao Y, Kulwa F, et al. Identification of COVID-19 samples from chest X-Ray images using deep learning: A comparison of transfer learning approaches. *J Xray Sci Technol* [Internet]. 2020;28(5):821-39. Available from: <https://doi.org/10.3233/xst-200715>

- [26] He K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition. In 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) [Internet]. Las Vegas: IEEE; 2016:770-778. Available from: <https://doi.org/10.1109/CVPR.2016.90>
- [27] Simonyan K, Zisserman A. Very Deep Convolutional Networks For Large-Scale Image Recognition. arXiv:1409.1556 [Internet]. 2015. Available from: <https://arxiv.org/abs/1409.1556>
- [28] Szegedy C, Ioffe S, Vanhoucke V, Alemi AA. Inception-v4, inception-ResNet and the impact of residual connections on learning. In Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence [Internet]. San Francisco: AAAI Pres; 2017:4278-4284. Available from: <https://dl.acm.org/doi/10.5555/3298023.3298188>
- [29] Zoph B, Brain G, Vasudevan V, Shlens J, Le Google Brain Q V. Learning Transferable Architectures for Scalable Image Recognition. In 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition [Internet]. Salt Lake City :IEEE; 2018:8697-8710. Available from: <https://doi.org/10.1109/CVPR.2018.00907>
- [30] Sahlol AT, Kollmannsberger P, Ewees AA. Efficient Classification of White Blood Cell Leukemia with Improved Swarm Optimization of Deep Features. Sci Rep [Internet]. 2020;10(1):2536. Available from: <https://doi.org/10.1038/s41598-020-59215-9>
- [31] Russakovsky O, Deng J, Su H, Krause J, et al. ImageNet Large Scale Visual Recognition Challenge. Int J Comput Vis [Internet]. 2015;115:211-52. Available from: <https://doi.org/10.1007/s11263-015-0816-y>
- [32] Ahuja S, Panigrahi BK, Dey N, Rajinikanth V, et al. Deep transfer learning-based automated detection of COVID-19 from lung CT scan slices. Appl Intell [Internet]. 2021;51:571-585. Available from: <https://doi.org/10.1007/s10489-020-01826-w>
- [33] Dey N, Rajinikanth V, Fong SJ, Kaiser MS, et al. Social Group Optimization-Assisted Kapur's Entropy and Morphological Segmentation for Automated Detection of COVID-19 Infection from Computed Tomography Images. Cogn Comput [Internet]. 2020;12:1011-1023. Available from: <https://doi.org/10.1007/s12559-020-09751-3>
- [34] Franquet, T. Diagnóstico por imagen de las enfermedades Pulmonares difusas: Signos y patrones diagnósticos básicos. Med respir. 2012;5(3):49-67