

Article

Deep Learning for COVID-19 Diagnosis from CT Images

Andrea Loddo * , Fabio Pili and Cecilia Di Ruberto 

Department of Mathematics and Computer Science, University of Cagliari, Via Ospedale 72, 09124 Cagliari, Italy; f.pili17@studenti.unica.it (F.P.); dirubert@unica.it (C.D.R.)

* Correspondence: andrea.loddo@unica.it

Featured Application: The investigation proposed in this work is a preliminary step towards the realisation of an automated COVID-19 detection system from CT images. It can be crucial in this medical task to aid clinicians in obtaining valuable information about CT images for their classification or obtaining general information about the status of the disease.

Abstract: COVID-19, an infectious coronavirus disease, caused a pandemic with countless deaths. From the outset, clinical institutes have explored computed tomography as an effective and complementary screening tool alongside the reverse transcriptase-polymerase chain reaction. Deep learning techniques have shown promising results in similar medical tasks and, hence, may provide solutions to COVID-19 based on medical images of patients. We aim to contribute to the research in this field by: (i) Comparing different architectures on a public and extended reference dataset to find the most suitable; (ii) Proposing a patient-oriented investigation of the best performing networks; and (iii) Evaluating their robustness in a real-world scenario, represented by cross-dataset experiments. We exploited ten well-known convolutional neural networks on two public datasets. The results show that, on the reference dataset, the most suitable architecture is VGG19, which (i) Achieved 98.87% accuracy in the network comparison; (ii) Obtained 95.91% accuracy on the patient status classification, even though it misclassifies some patients that other networks classify correctly; and (iii) The cross-dataset experiments exhibit the limitations of deep learning approaches in a real-world scenario with 70.15% accuracy, which need further investigation to improve the robustness. Thus, VGG19 architecture showed promising performance in the classification of COVID-19 cases. Nonetheless, this architecture enables extensive improvements based on its modification, or even with preprocessing step in addition to it. Finally, the cross-dataset experiments exposed the critical weakness of classifying images from heterogeneous data sources, compatible with a real-world scenario.



Citation: Loddo, A.; Pili, F.; Di Ruberto, C. Deep Learning for COVID-19 Diagnosis from CT Images. *Appl. Sci.* **2021**, *11*, 8227. <https://doi.org/10.3390/app11178227>

Academic Editor: Alessandro Stefano

Received: 12 August 2021

Accepted: 1 September 2021

Published: 4 September 2021

Keywords: COVID-19 detection; convolutional neural network; deep learning; lung CT analysis; image classification; SARS-CoV-2

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

COVID-19 is a disease caused by the SARS-CoV-2 virus, declared a pandemic by the World Health Organisation on 11 March 2020. At the time of writing, COVID-19 has more than one hundred and eighty million confirmed cases and has caused more than three million deaths, with a mortality rate of 2.1% [1]. As hospitals have been shown to have limited availability of adequate equipment, a rapid diagnosis would have been and still is essential to control the spread of the disease, increase the effectiveness of medical treatment, and, consequently, the chances of survival without intensive care. Basically, the polymerase chain reaction and reverse transcriptase (RT-PCR) method is the primary screening tool for COVID-19, in which SARS-CoV-2 ribonucleic acid (RNA) is detected within an upper respiratory tract sputum sample [2]. However, many countries are unable to provide sufficient testing, and, in any case, only people with apparent symptoms are tested, and it takes hours to provide an accurate result.

Therefore, there is a need for faster and more reliable screening techniques that could further confirm the PCR test or replace it entirely, such as imaging-based methods. They may complement its use to achieve greater diagnostic certainty or even substitute in some countries where RT-PCR is not readily available. In some cases, chest X-ray (CXR) abnormalities are seen in patients who initially had a negative RT-PCR test, and several studies have shown that chest computed tomography (CT) has greater sensitivity for COVID-19 than RT-PCR and could be considered a primary tool for diagnosis [3–6]. In response to the pandemic, researchers have rushed to develop models using artificial intelligence (AI), particularly machine learning, to support clinicians [7].

Computed tomography is already a widely explored medical imaging technique that allows non-invasive visualisation of the interior of an object [8–13] and is widely used in many applications, such as medical imaging for clinical purposes [14–18]. For this reason, clinical institutions have used CT as an effective and complementary screening tool alongside RT-PCR [5,6] with a higher sensitivity of up to 98% compared to 71% for RT-PCR [19,20]. In particular, several studies have shown that CT has excellent utility in detecting COVID-19 infections during routine CT examinations for reasons unrelated to COVID-19, such as monitoring of elective surgical procedures and neurological examinations [21]. Other scenarios where CT imaging has been exploited include cases where patients have worsening respiratory complications and cases where patients with negative RT-PCR test results are suspected to be COVID-19 positive due to other factors. Early studies have shown that chest CT images of patients may contain some potential indicators for COVID-19 [2,5,6,22] infections, but may also be contained in non-COVID-19 infections. This issue can lead to challenges for radiologists in distinguishing COVID-19 infections from non-COVID-19 infections using chest CT [23,24]. However, the duration of diagnosis is the main limitation of CT scan tests: even experienced radiologists need about 21.5 min to analyse the test results of each case [25], and during the emergency, a large number of CT images have to be evaluated in a very short time, thus increasing the probability of misclassification. For this reason, intelligent automatic diagnosis systems that automatically classify chest CT images can help to improve speed and to rapidly confirm the test result.

In recent years, deep learning workflows have emerged from the proposed AlexNet convolutional neural network (CNN) in 2012 [26]. CNNs do not follow the typical workflow of image analysis because they can extract features independently without the need for feature descriptors or specific feature extraction techniques. Therefore, they differ from conventional machine learning methods because they require little or no image pre-processing and can automatically infer an optimal data representation from raw images without requiring prior feature selection, resulting in a more objective and less biased process. Furthermore, they achieved optimal results in many domains, such as computer vision devoted to medical analysis, with images coming from magnetic resonance imaging (MRI) [27], microscopy [28], CT [29], ultrasound [30], X-ray [31], and mammography [32]. They have been successfully applied to various different problems, like classification or segmentation [33–36]. Deep learning-based methods have also made significant progress in the analysis of lung diseases, which is a comparable scenario to COVID-19 [37–39]. However, the scenario of CT images of lungs referred to COVID-19 and non-COVID-19 patients can be particularly problematic to classify, especially when the damage due to pneumonia of different causes is present simultaneously. The main findings of chest CT scans of COVID-19 positive patients indicate traces of ground-glass opacity (GGO) [40]. Two CT scans of COVID-19 and non-COVID-19 are shown in Figure 1.

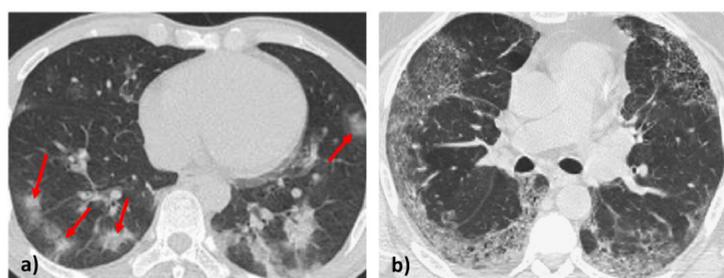


Figure 1. (a) represents a CT of the lungs of a patient with COVID-19, in which there are clear traces of GGO indicated by red arrows. (b) shows a CT of the lungs of a non-COVID-19 patient with diffuse opacity in the outer parts of the lungs. These images are courtesy of [41].

The overall objective of this study is to investigate the behaviour of the main existing off-the-shelf CNNs for the classification of patients' CTs. This work is a preliminary investigation for the future development of a tool that provides confirmation of the viral test result or provides more details about the ongoing infection, also considering that according to the Center for Disease Control (CDC), even if a chest CT or X-ray suggests COVID-19, the viral test is the only specific method for diagnosis [42]. Specifically, we propose a comprehensive investigation of the problem of COVID-19 classification from chest CT images from different perspectives:

1. We present a comparative study of several off-the-shelf CNN architectures in order to select a suitable deep learning model to perform a three-class classification on the public COVIDx CT-2A dataset, specifically divided into COVID-19, pneumonia and healthy cases;
2. On the same dataset, we performed a patient-oriented experiment by grouping all the CT images of the patients, in which the aim was to provide a diagnosis;
3. We investigated the robustness of the methods by performing two cross-dataset experiments and evaluating the performance of CNNs previously trained on COVIDx CT-2A. In particular, we performed a two-class classification between COVID-19 and healthy cases, on the COVID-CT dataset, without fine-tuning;
4. We repeated the experiment just described, by fine-tuning the most promising CNNs, demonstrating that it is still problematic to integrate automatic methods in the clinical diagnosis of COVID-19.

We both demonstrate how off-the-shelf deep learning architectures can be utilised to classify CT images representing COVID-19 affected patients and how transfer learning capabilities are still far from offering a concrete contribution in a real-world scenario, as demonstrated by our cross-dataset experiments, without addressing it with different techniques. The experiments are not intended to provide an exhaustive comparison of the performance of these methods; instead, we wanted to select the most suitable for our classification of CT images without, for the time being, investigating possible parametric improvements. The purpose is to create a concrete baseline with the potential to be modified and developed further. Moreover, several works in the context of COVID-19 diagnostics have considered small or private datasets or lacked rigorous experimental methods, potentially leading to over-fitting and overestimation of performance [7,43]. For this reason, we:

1. Carefully selected the two datasets on which to conduct the experiments described. In fact, Roberts et al. [7] have recently shown that most of the datasets used in the literature for the diagnosis or prognosis of COVID-19 suffer from duplication and quality problems;
2. Selected COVIDx CT-2A, a public reference dataset specifically proposed for COVID-19 detection from CT imaging, because of the high risks of bias due to source problems and datasets created from unsupervised public online repositories. It has already been provided with train, validation, and testing splits.

We verified the robustness of the solution on both the public COVIDx CT-2A and COVID-CT datasets. Our proposed approach achieves promising results on COVID-19 identification, although it does not show satisfactory performance on cross-dataset experiments.

The rest of the article is organised as follows. The following paragraph presents a review of deep learning approaches for COVID-19 detection. Section 2 describes the datasets used in our experiments and presents the metrics adopted to evaluate the experimental results illustrated in Section 3. In Section 4, we analyse and discuss the experimental results and give a comparison with the state of the art. Finally, conclusions and future directions are drawn in Section 5.

Related Work

Here, we briefly describe some works that have addressed tasks related to COVID-19. Although the research is still evolving, the automatic classification of COVID-19 has gained wide attention from researchers around the world [7,42,44,45]. In this context, we can broadly distinguish the proposed methods into those based on 2D and 3D images. Among the most recent ones [46,47], 3D images could be handy to avoid losing the interstitial information of the lungs. However, several works have exploited 2D images showing the property of extracting representative features of COVID-19 lesions for disease detection [48–56]. They are all CNN-based and used CT [48–55] or CXR [43,50,56] images. We particularly focused this study on deep learning-based classification methods for COVID-19 detection.

Among the CT-based methods, Jin et al. [48] proposed a deep learning-based system for COVID-19 diagnosis, performing lung segmentation, COVID-19 diagnosis, and COVID-infectious slices location. In contrast, Hu et al. [51] proposed a weakly supervised multiscale deep learning framework for COVID-19 detection, inspired by the VGG architecture [57], which assimilates different scales of lesion information using CT data of the chest. Polsinelli et al. [52] implemented a lightweight CNN, based on the SqueezeNet model [58] for efficient discrimination of COVID-19 CT images against other community-acquired pneumonia or healthy CT images. Biswas et al. [53] used a transfer learning strategy on the three pretrained models of VGG-16 [57], ResNet50 [59], and Xception [60], combining them with the ensemble stacking strategy and tested the method on CT images of the chest. Zhao et al. [55] adopted the ResNet-v2, a modified version of ResNet [59]. Moreover, they added group normalisation instead of batch normalisation and conducted a weight standardisation for all convolutional layers. Lastly, they also incorporated the pre-training data from CIFAR-10 [61], ILSVRC-2012 [62], and ImageNet-21k [63] as the parameters for initialisation.

On the subject of CXR-based works, Minaee et al. [50] employed four pretrained models (ResNet18 [59], ResNet50 [59], SqueezeNet [58] and DenseNet-121 [64]) on CXR data, and analysed their performance for COVID-19 detection. On the other hand, Signoroni et al. [43] developed BS-Net, a multi-block deep learning-based architecture designed for the assessment of pneumonia severity on CXRs. More recently, Oyelade et al. [56] proposed CovFrameNet, a novel deep learning-based framework based on a substantial image pre-processing step and a CNN architecture for detecting the presence of COVID-19 on CXRs.

Thanks to the powerful discriminative ability of CNNs, several authors tried to propose CNN-based frameworks for the diagnosis or prognosis of COVID-19, even though CNNs typically require large scale datasets to perform a correct classification. However, most of the existing CT scan datasets for COVID-19 contain at most hundreds of CT images [65–67]. Therefore, we exploited COVIDx CT-2A [68], composed of 194,922 CT images, as described in Section 2.1.1 to propose a baseline classification approach and we evaluated it on the external dataset COVID-CT, described in Section 2.1.2 to assess generalisability of the proposal. In general, we aim to avoid the following drawbacks:

1. Using small scale datasets;

2. Using not robust or multiple unsupervised source datasets;
3. Testing the method without external validation.

Regarding the works that employed the datasets used in our study, Zhao et al. [41] worked on COVID-CT, while Gunraj et al. [54] on COVIDx CT-2A. The former is based on a transfer learning approach on the DenseNet network, while the latter proposed COVID-Net CT [54], a deep convolutional neural network tailored for detection of COVID-19 cases from chest CT images.

This work differs from those described above because:

- (i) We propose an extensive comparison between different off-the-shelf CNN architectures, in order to obtain the most suitable for the task, using a large and public dataset;
- (ii) We avoid the high risks of errors due to datasets created from unsupervised online public repositories, using two public reference datasets, to try to validate our approach;
- (iii) We introduce a preliminary solution based on learning by sampling, showing how CNNs need further improvements to generalise the detection of COVID-19 in heterogeneous datasets.

2. Materials and Methods

In this work, we exploited two publicly available datasets, as described in Section 2.1. Then, in Section 2.2 we give a detailed description of the metrics adopted to evaluate the experimental results.

2.1. Datasets

The datasets exploited in this work are COVIDx CT-2A and COVID-CT, both of which are publicly available. We describe them as follows.

2.1.1. COVIDx CT-2A

COVIDx CT-2A [68] is an open-access dataset. At the time of writing, it is composed of 194,922 CT images from 3745 patients from 15 different countries, between 0 and 93 years old (median age of 51), with strongly clinically verified findings. Every image belongs to a particular class verified by expert pathologists. In particular, the classes are COVID-19, indicating CT images of COVID-19 positive patients, pneumonia indicating CT images of patients with pneumonia not caused by COVID-19, and normal, indicating CT images of patients in normal conditions.

The countries involved are part of a multinational cohort that consists of patient cases collected by the following organisations and initiatives from around the world:

1. China National Center for Bioinformation (CNCB) [49] (China);
2. National Institutes of Health Intramural Targeted Anti-COVID-19 (ITAC) Program (hosted by TCIA [69], countries unknown);
3. Negin Radiology Medical Center [70] (Iran);
4. Union Hospital and Liyuan Hospital of the Huazhong University of Science and Technology [71] (China);
5. COVID-19 CT Lung and Infection Segmentation initiative annotated and verified by Nanjing Drum Tower Hospital [72] (Iran, Italy, Turkey, Ukraine, Belgium, some countries unknown);
6. Lung Image Database Consortium (LIDC) and Image Database Resource Initiative (IDRI) [73] (countries unknown);
7. Radiopaedia collection [74] (Iran, Italy, Australia, Afghanistan, Scotland, Lebanon, England, Algeria, Peru, Azerbaijan, some countries unknown).

Figure 2 shows some sample images taken from the dataset.

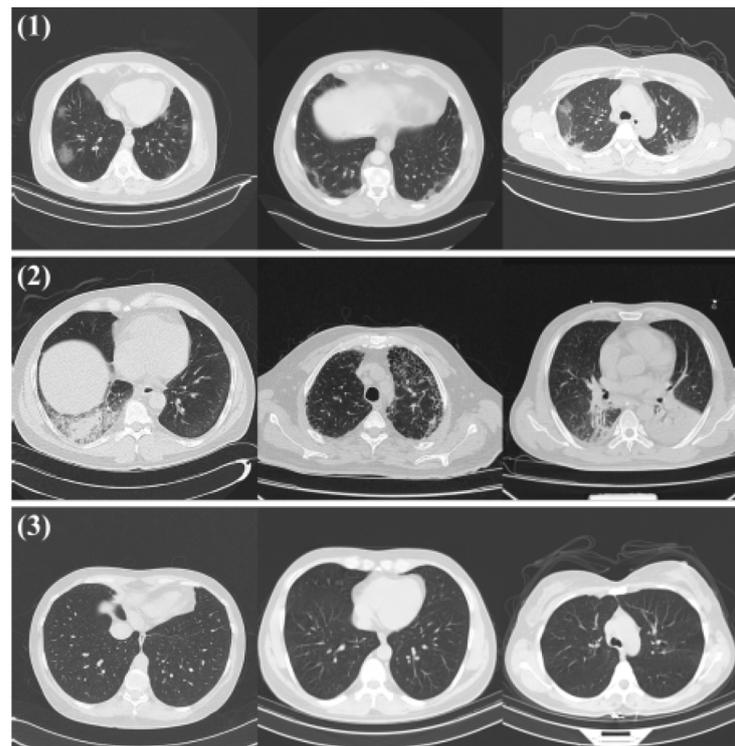


Figure 2. Sample CT images from the COVIDx CT dataset. From top to bottom, the images in (1) represent coronavirus pneumonia due to SARS-CoV-2 infection (NCP), (2) are common pneumonia (CP), and (3) are healthy lungs. Images are courtesy of [54].

2.1.2. COVID-CT Dataset

COVID-CT-dataset was created by Zhao et al. [41] to be used for future research and works related to the diagnosis of COVID-19 using CT images. The images were collected between 19 January 2020 and 25 March 2020 for a total of 470 CT images of different patients. Of these images, 275 are chest CTs of COVID-19 positive patients, while the remaining 195 are non-positive patients. The classes are COVID-19 which indicates CT images of COVID-19 positive patients, as in the previous dataset, and non-COVID-19 which indicates CT images of non-COVID-19 patients, grouping CT images of patients with pneumonia not caused by COVID-19 and of patients in normal condition. One sample per class is shown in Figure 3. A senior radiologist confirmed the usefulness of this dataset at Tongji Hospital, Wuhan, China, who diagnosed and treated a large number of COVID-19 patients during the COVID-19 outbreak between January and April 2020 [41].

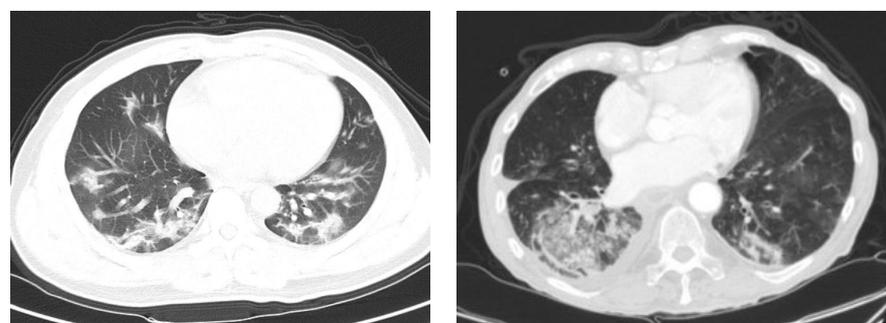


Figure 3. Sample CT images from the COVID-CT dataset. On the left, an image that represents a coronavirus pneumonia due to SARS-CoV-2 infection. On the right, image belonging to the non-COVID-19 class. Images are courtesy of [41].

2.2. Metrics

The performance measures have been evaluated by averaging five different simulations for all the networks. The measures used to quantify the performance of a network are the accuracy (Acc), precision (Pre), specificity (Spec), recall (Rec), F1-score (F1) as following defined:

$$\text{accuracy} = \frac{TP + TN}{TP + TF + FP + FN'}$$

$$\text{precision} = \frac{TP}{TP + FP'}$$

$$\text{specificity} = \frac{TN}{FP + TN'}$$

$$\text{recall} = \frac{TP}{TP + FN'}$$

$$F1 = \frac{2 * \text{precision} * \text{recall}}{\text{precision} + \text{recall}} = \frac{2 * TP}{2 * TP + FP + FN'}$$

Precision measures the number of correctly labelled items belonging to the positive class divided by the items correctly or incorrectly labelled as belonging to the same class. Specificity measures the proportion of correctly identified negatives (also called the true negative rate), while sensitivity measures the proportion of correctly identified positives (also called the true positive rate). The fourth measure is accuracy, defined as the ratio of correctly labelled instances to the entire pool of instances. The last is the F1 score, which conveys the balance between accuracy and recall.

Furthermore, since we are facing a multiclass imbalance problem, we adopted two global metrics for multiclass imbalance learning to evaluate the performance of the networks [75]. They are the macro geometric average (MAvG), defined as the geometric mean of the partial accuracy of each class, and the macro arithmetic average (MAvA), defined as the arithmetic mean of the partial accuracies of each class.

We describe them as follows:

$$MAvG = \left(\prod_{i=1}^J Acc_i \right)^{\frac{1}{J}},$$

$$MAvA = \frac{\sum_{i=1}^J Acc_i}{J}.$$

3. Results

We now describe the experimentation conducted in this work. In detail, in Section 3.1 we first describe the experimental setup adopted for the classification tasks. Then, in Section 3.2 we report the results of the experiments performed on both datasets.

3.1. Experimental Setup

The images to be classified are lung CTs. They are organised into classes, as described below. Considering this work as a baseline for further investigation, the images are not subject to any preprocessing or augmentation process. In order to make the experiments reproducible, we kept the dataset splits provided by the authors and did not apply any randomisation strategy. We employed two different training strategies:

- (i) From scratch;
- (ii) Fine-tuning the previously trained networks.

The tests were carried out on several popular CNNs to find the best architecture for our purpose. The tested networks are AlexNet [26], the Residual Networks [59] ResNet18,

ResNet50, ResNet101, GoogLeNet [76], ShuffleNet [77], MobileNetV2 [78], InceptionV3 [79], and VGG16 [57].

The experiments were performed using the hyperparameters setting described in Table 1 for all networks to assess potential performance variations. In particular, after empirical evaluation, we adopted Adam, which performed better than the other solvers. In addition, the maximum number of epochs was set to 20 due to a large number of images.

Table 1. Hyperparameters setting for training CNNs from scratch and fine-tuning.

Params	Value
Solver	Adam
Max Epochs	20
Mini Batch Size	8
Initial Learn Rate	1×10^{-4}
Learn Rate Drop Period	10
Learn Rate Drop Factor	0.1
L ₂ Regularisation	0.1
Validation Frequency	8000

Since COVIDx CT-2A is the largest dataset, we employed it for model training. Its images were divided by the authors according to the following percentages: 70%, 20%, and 10% for training, validation, and testing, respectively. As for the COVID-CT dataset, we used it in two ways: the first time, it was taken as a whole as a test set, while the second time, it was divided in the same way as COVIDx CT-2A to be used for a fine-tuning strategy.

3.2. CT Image Classification via Deep Learning

Several types of experiments were designed in this work in order to assess the feasibility of the deep learning approach and its robustness. In particular, on the COVIDx CT-2A dataset, we performed:

1. A three-class classification as reported in Section 3.2.1;
2. A patient-oriented classification, described in Section 3.2.2.

On the other hand, on the COVID-CT dataset, we realised:

1. A two-class classification using the four best-performing networks from the previous experiments on the entire dataset;
2. A two-class classification using the same four networks, fine-tuning them on this dataset.

Both are reported in Section 3.2.3.

3.2.1. Three-Class Classification on COVIDx CT-2A

In this experiment, we trained each network used in this work, using the split proposed by the authors, in order to obtain a baseline result. Table 2 shows the results obtained with each architecture employed, while Figure 4 shows the relationship between the MAvG metric and the three classes included in the dataset.

3.2.2. Patient-Oriented Classification on COVIDx CT-2A

For this experiment, the models obtained from the experiments described in Section 3.2.1 were used. We proceeded as follows: the test set (consisting of 25,658 images) was used, according to the subdivision provided by the authors, to have one set of images for each of the 426 different patients in the test set. We ensured that each patient only had images belonging to one class because otherwise, this would invalidate the test. Once the images of each patient had been examined, the model would produce results similar to those of the ternary classification. With this in mind, it was decided to use class accuracy as a critical

metric: if it was above 50%, the patient belonged to the class. Otherwise, it would be classified as incorrect. In this way, it was possible to see how each model behaved with each class, and finally, average accuracy was calculated to describe the level of accuracy of the network, as shown in Table 3.

Table 2. Macro-average performance for testing on COVIDx CT-2A.

Net	Pre	Rec	Spe	Acc	F1	MAvA	MAvG
AlexNet	91.88%	92.50%	96.64%	95.38%	92.17%	91.75%	91.88%
GoogLeNet	89.46%	88.96%	95.13%	93.59%	89.08%	89.46%	89.39%
InceptionV3	95.48%	94.34%	97.54%	96.97%	94.84%	95.48%	95.48%
VGG16	96.65%	96.57%	98.44%	97.93%	96.58%	96.65%	96.63%
VGG19	97.85%	97.87%	99.08%	98.87%	97.86%	97.85%	97.84%
ShuffleNet	95.36%	94.92%	97.94%	97.28%	95.13%	95.36%	95.34%
MobileNetV2	94.24%	93.05%	97.31%	96.38%	93.38%	94.24%	94.14%
ResNet18	96.41%	96.67%	98.22%	97.71%	95.98%	96.41%	96.40%
ResNet50	92.19%	89.53%	95.61%	94.62%	90.45%	92.19%	92.16%
ResNet101	94.99%	93.06%	97.16%	96.53%	93.08%	94.99%	94.97%

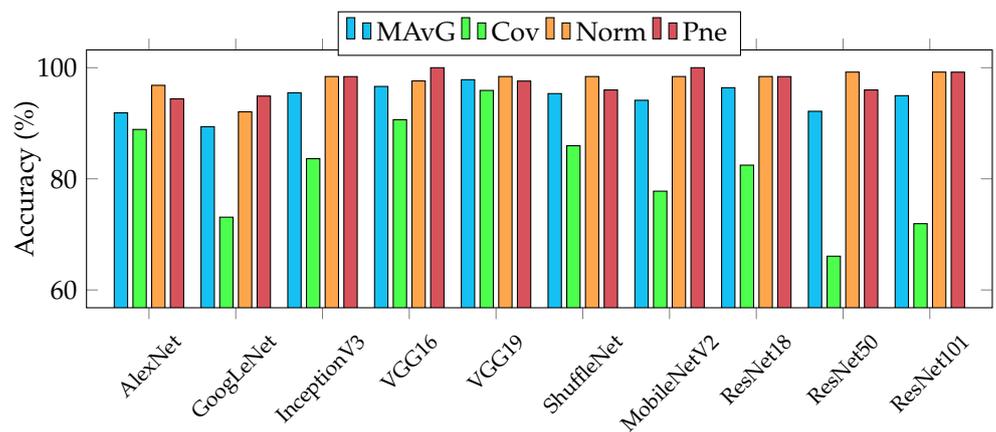


Figure 4. MAvG and classwise accuracy trends with the different classifiers adopted. Cov, Norm, Pne stands for the COVID-19, normal, and pneumonia classes, respectively.

Table 3. Patient-oriented testing with COVIDx CT-2A.

Net	COVID-19	Normal	Pneumonia	AVG
AlexNet	88.89%	96.83%	94.40%	93.97%
GoogLeNet	73.10%	92.06%	94.90%	86.52%
Inception V3	83.63%	98.41%	98.40%	93.48%
VGG16	90.64%	97.62%	100.00%	96.09%
VGG19	95.91%	98.41%	97.60%	97.31%
ShuffleNet	85.96%	98.41%	96.00%	93.46%
MobileNet V2	77.78%	98.41%	100.00%	92.06%
ResNet18	82.46%	98.41%	98.40%	93.09%
ResNet50	66.08%	99.21%	96.00%	87.10%
ResNet101	71.93%	99.21%	99.20%	90.11%

3.2.3. Two-Class Classification on COVID-CT

For this experiment, we proceeded in two steps: initially, we used the entire COVID-CT dataset as a test dataset for the top four models obtained from the Section 3.2.1 experiments. Subsequently, it was decided to perform a fine-tuning strategy on the same models. In

particular, we chose VGG19, given its results in both previous experiments, MobileNetV2, one of the most superficial networks with good results in classifying patients of the *normal* and *pneumonia* classes, and, finally, VGG16 and ResNet18, being the two networks with the best results after VGG19. The dataset was then divided into training, validation, and testing, according to the percentages provided by the authors. Table 4 shows the results on the whole dataset, while Table 5 shows the results after the fine-tuning strategy.

Table 4. Macro-average performance for testing on COVID-CT without fine-tuning.

Net	Pre	Rec	Spe	Acc	F1
VGG19	52.23%	52.13%	52.13%	52.82%	51.88%
VGG16	45.28%	45.94%	45.94%	47.18%	44.43%
ResNet18	49.70%	49.81%	49.81%	51.74%	45.75%
MobileNetV2	49.11%	49.33%	49.33%	50.94%	46.72%

Table 5. Macro-average performance for testing on COVID-CT with fine-tuning.

Net	Pre	Rec	Spe	Acc	F1
VGG19	70.19%	68.70%	68.70%	69.15%	68.40%
VGG16	61.19%	60.94%	60.94%	61.19%	60.84%
ResNet18	70.16%	70.01%	70.01%	70.15%	70.02%
MobileNetV2	67.13%	67.05%	67.05%	67.16%	67.07%

4. Discussion

In this section, we give a detailed discussion of the results obtained, shown in Section 3.

4.1. On the Three-Class Classification on COVIDx CT-2A

As Table 2 shows, it can be seen that GoogLeNet, still the worst performing network, scored consistently below 90%, except for specificity and accuracy. Even ResNet50, which reports quite different values, drops to 89.53% in recall, indicating that the network had some difficulty in accurately distinguishing true positives. As for the rest of the models, they all show consistent results, as was the case during the validation phase. VGG19 presents exceptional results, reaching 99.08% in specificity, while the other metrics exceed 97%. VGG16 and ResNet18 obtained similar results. In general, it can be seen that the network that produced the best results was VGG19, having consistently high values in every metric. It is followed by VGG16 and ResNet18, with excellent results narrowly below those reached by VGG19. The fine-tuning strategy, which will be carried out to prepare the models for the classification on the COVID-CT dataset, will concern these three models, plus MobileNet V2, to deduce possible adaptability of specific networks in the mobile environment for this task.

Some specific insights come from Figure 4, which shows the relationship between the MAVG metric and the class accuracies for each classifier used. In particular, it can be seen that in addition to having by far the highest accuracy, VGG19 is also the only one to have a uniform and acceptable score for all three classes. Concerning the problem of unbalanced and multiclass classification, it is fundamental to underline that all the networks manage to achieve high accuracies essentially thanks to the scores obtained on the *normal* and *pneumonia* classes. For example, ResNet50 obtained 99.21% accuracy on the *normal* class and only 66.08% on the *COVID-19* class, being the best and the worst, respectively. Furthermore, MobileNetV2 obtained 100.00% accuracy on the *pneumonia* class, but only 77.78% on the *COVID-19* class, being the best and the fourth-worst, respectively.

4.1.1. On the Patient-Oriented Classification on COVIDx CT-2A

The peculiarity of misclassification of CT images of patients belonging to the *COVID-19* class affects every network. In some cases in a more pronounced form and in others

less so. A deeper analysis of the results showed that many of the misclassified patients were classified as belonging to the class *pneumonia*. This issue is due to the similarity of the images, as both classes represent pneumonia, although of different origins.

On the other hand, in the case of AlexNet, for the *normal* class, a relatively high result was obtained, a sign that few patients were misclassified; a similar situation for the *pneumonia* class where some patients were classified as belonging to both the *COVID-19* and the *normal* class. In general, the model obtained a balanced average accuracy.

Additionally, GoogLeNet's results obtained for patients of the class *COVID-19* are much lower than AlexNet, and even if the accuracy of the other classes is average, we can see how it affects the average accuracy of the network.

As for InceptionV3, the accuracy of the *COVID-19* class is higher than GoogLeNet but remains at a high level when compared to the results of the other classes. On the other hand, the average accuracy of the network obtains a positive result.

VGG16 achieves 100% in the classification of patients in the class *pneumonia*, the highest result so far and the optimal one. The class *normal* also obtained a very high result, while the class *COVID-19*, although obtaining outstanding results, had some minor difficulties in classification. As for VGG19, although it did not reach 100% in the *pneumonia* class compared to VGG16, this model achieves high and uniform results: 95.91% in the *COVID-19* class is the highest achieved so far. The other classes also produced satisfactory results, and the average accuracy of the network is 97.31%.

ShuffleNet performs as well as the average, leading to a relatively low result for the *COVID-19* class compared to the other two classes, which perform very well. However, network accuracy is still average.

MobileNetV2 has excellent performance in the *pneumonia* and *normal* classes, while the poor result obtained in the *COVID-19* class affects the average accuracy of the network.

With regard to the Residual networks, ResNet18 obtained high results in the *normal* and *pneumonia* classes and remained average in the *COVID-19* class. The network's final accuracy is also average; although both ResNet50 and ResNet101 achieved very high results in the two classes *pneumonia* and *normal*, as ResNet18 did, those obtained in the *COVID-19* class are drastically low, the lowest to date. This makes ResNet50 the network with the lowest average accuracy, caused mainly by the results obtained from the classification of patients in the *COVID-19* class; ResNet101 is slightly higher, but still with unsatisfactory results.

To sum up, considering the results obtained with VGG19, the network with the lowest number of misclassified *COVID-19* patients, we went to see what they were to try to draw specific conclusions. VGG19 misclassified seven patients and, of these, not all were misclassified by the other networks, indicating that a hybrid approach could improve results.

4.1.2. On the Two-Class Classification on COVID-CT

Although VGG19 was the best in the previous tests, it did not produce the same results with this dataset, as it can be seen from Table 4. As for VGG16, it scored lower than VGG19 in the previous tests and the binary classification of COVID-CT. In this case, ResNet18 scores higher and closer to VGG19. The same applies to MobileNetV2, which achieves results just below ResNet18.

As a result of the fine-tuning, the results have improved significantly, although still below the results obtained with the previous dataset, as shown in Table 5. These lower-than-average results may be due to the fact that the original COVID-CT dataset, dating back to the early 2020s, has been slowly modified over the months with the addition of new CT images of poor quality or compromised by overlaps. This fact explains why the networks cannot classify the images correctly, as having been trained with the high-quality images from the COVIDx CT-2A dataset, they struggle to accurately classify these new elements.

To sum up, the results of the two tables do not differ as much as one would expect from a fine-tuning strategy. Finally, it can be said that fine-tuned ResNet18 was able to outperform the other CNNs, with metric values that always hover around 70%.

4.1.3. Comparison with the State-of-the-Art

Table 6 shows a brief but effective comparison with the state of the art works on COVIDx CT-2A. As it can be seen, the works of Gunraj et al. [54] produced significant early results with their proposed COVID-Net CT convolutional neural network. However, recently, the work of Zhao et al. [55] outperformed state of the art, reaching 99.2% using ResNet-v2, a modified version of ResNet and several improvements, such as the group normalisation or the weight standardisation for all convolutional layers. Finally, our baseline work demonstrates that even pre-existing architectures can reach outstanding results. As reported in Table 6, we reached 98.87% accuracy with VGG-19 without any improvement, such as preprocessing, data standardisation, group normalisation, etc. This opens the field to further investigations and improvements, as detailed in the following section.

Table 6. Comparison of our work with the state of the art on COVIDx CT-2A.

Net	Acc
COVID-Net CT-1 [54]	94.5%
COVID-Net CT-2 L [54]	98.1%
COVID-Net CT-2 S [54]	97.9%
Bit-M [55]	99.2%
VGG-19 (this work)	98.87%

4.1.4. Limitations of This Work

Although interesting results have been shown, our work suffers from some limitations. First, the most performing solution on COVID-19 class relies entirely on the VGG19 architecture, even though other networks showed excellent results in the other two classes. Considering the properties of these networks, combining their features could improve the results, particularly in increasing the capacity to distinguish the different classes more specifically. Second, every experimental condition assumed no preprocessing step. However, in the context of proposing a complete framework in the future, preprocess the images (e.g., with denoising) could be crucial. Third, the patient-oriented experiments confirmed the excellent results obtained by VGG19, even though certain patients have been misclassified in contrast to other CNNs. Efforts should be made in this sense in order to understand more clearly the sections of the CT scan that are discriminative in this critical scenario. Fourth, as represented by the classwise performance, the COVID-19 class is generally harder to distinguish with respect to the others because of their structure. For this reason, handcrafted and, potentially, the combination of heterogeneous descriptors could help recognise the most challenging cases, as already shown in similar tasks [80–82].

5. Conclusions

The objective of this work was to propose a classification methodology for the diagnosis of COVID-19 through deep learning techniques applied on CT images. To achieve this goal, an extensive comparative study of the main existing CNN architectures was carried out.

The tests carried out on the two datasets showed very different results. Those obtained with the COVIDx CT-2A dataset are excellent for all the models used; in particular, VGG19 stands out for the high values obtained in the specificity metric and precision and recall. No other network has achieved these results. However, it is important to say that networks such as VGG16 and ResNet18 also achieved more than satisfactory results. As far as the other networks are concerned, GoogLeNet and ResNet50 seem the least suitable, as they

always deviated considerably from the average values obtained. In addition, the results obtained with VGG-19 are comparable with the results of the networks currently existing at the state of the art that works on COVIDx CT-2A.

The patient-oriented classification also brought outstanding results, with high accuracy values for the class *COVID-19*, and, in some cases, 100% accuracy for the class *pneumonia*. The best network remains, in any case, VGG19, being the one with the highest average accuracy and, therefore, misclassifying a few patients compared to the other networks. Through the analysis of the misclassified patients, it was deduced that it is probably necessary to create an ad hoc network exploiting the existing CNNs to improve the results.

About the COVID-CT dataset, however, the results do not match the previous ones and, on the contrary, there was a drop in performance of almost 50%. Only fine-tuning was able to remedy this, increasing the values obtained by 20%. Nevertheless, this does not compensate for the difference in performance. The problem could be mainly due to the quality of the images of the COVID-CT dataset, which are often compromised or of very poor quality.

This work highlighted some limitations. First of all, cross-dataset experiments showed that existing CNNs, even after a fine-tuning procedure, really suffer from limited dataset scenarios. Second, patient-oriented experiments show that some networks misclassified some COVID-19 patients as normal pneumonia cases, while others did not. This clearly motivates further investigation on the models and, also, possible modifications. Third, the absence of defined standards in the acquisition of these images and, in addition, the problem of building affordable COVID-19 datasets from heterogeneous sources, especially during the early months of the pandemic [7] can be considered a limitation and also a future direction, as it clearly appears that the distinctive COVID-19 features need to be further studied.

The indications emerging from this work are that:

- (i) In addition to fine-tuning, some preprocessing steps oriented to the enhancement of CT images could be helpful for the networks to produce more discriminative features; and
- (ii) Considering the results of the patient-oriented experiments, a hybrid approach, even involving ad hoc handcrafted features, could improve the results.

In future directions, we certainly aim to discover other valuable features from CT images to recognise COVID-19, extending the investigation to include handcrafted features and even combining them with deep features. In addition, we also want to consider assessing the severity of COVID-19.

We will conduct further experiments to identify key features in CT images and facilitate screening by medical doctors. We want to stress again that this work is still at the stage of theoretical research, and the models have not been validated in real clinical routines. Our contribution is to offer a baseline with some public benchmark datasets to be extended with new investigations. Therefore, we would like to test our system in the clinical routine and communicate with doctors to understand how such a system can be integrated into the clinical routine.

Therefore, we would like to:

1. Modify VGG19 to investigate the best accuracy density (accuracy divided by the number of parameters) and the best inference time;
2. Optimise the hyperparameters, for example with Bayesian method;
3. Use class activation map (CAM) to understand which parts of the image are relevant in the misclassification cases obtained by VGG19 but not from the other networks;
4. Test our system in the clinical routine and communicate with doctors to understand how such a system can be integrated into the clinical routine.

Author Contributions: Conceptualisation, A.L. and C.D.R.; Methodology, A.L. and C.D.R.; Investigation, A.L.; software, A.L. and F.P.; writing—original draft, A.L. and F.P.; writing—review and editing, A.L. and C.D.R. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: All the codes and the data used in this study are available at the following url: [GitHub repository](#) (accessed on 12 August 2021).

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

Acc	Accuracy
Pre	Precision
Spe	Specificity
Rec	Recall
F1	F1-score
MAvG	Macro geometric average
MAvA	Macro arithmetic mean
CT	Computed tomography
CNN	Convolutional neural network

References

- University of Oxford. Coronavirus Pandemic (COVID-19)—The Data. 2021. Available online: <https://ourworldindata.org/coronavirus-data> (accessed on 30 June 2021).
- Wang, W.; Xu, Y.; Gao, R.; Lu, R.; Han, K.; Wu, G.; Tan, W. Detection of SARS-CoV-2 in different types of clinical specimens. *JAMA* **2020**, *323*, 1843–1844. [[CrossRef](#)]
- Huang, E.P.C.; Sung, C.W.; Chen, C.H.; Fan, C.Y.; Lai, P.C.; Huang, Y.T. Can computed tomography be a primary tool for COVID-19 detection? Evidence appraisal through meta-analysis. *Critical Care* **2020**, *24*, 193. [[CrossRef](#)] [[PubMed](#)]
- Long, C.; Xu, H.; Shen, Q.; Zhang, X.; Fan, B.; Wang, C.; Zeng, B.; Li, Z.; Li, X.; Li, H. Diagnosis of the Coronavirus disease (COVID-19): rRT-PCR or CT? *Eur. J. Radiol.* **2020**, *126*, 108961. [[CrossRef](#)] [[PubMed](#)]
- Fang, Y.; Zhang, H.; Xie, J.; Lin, M.; Ying, L.; Pang, P.; Ji, W. Sensitivity of chest CT for COVID-19: comparison to RT-PCR. *Radiology* **2020**, *296*, E115–E117. [[CrossRef](#)] [[PubMed](#)]
- Ai, T.; Yang, Z.; Hou, H.; Zhan, C.; Chen, C.; Lv, W.; Tao, Q.; Sun, Z.; Xia, L. Correlation of chest CT and RT-PCR testing for coronavirus disease 2019 (COVID-19) in China: A report of 1014 cases. *Radiology* **2020**, *296*, E32–E40. [[CrossRef](#)] [[PubMed](#)]
- Roberts, M.; Driggs, D.; Thorpe, M.; Gilbey, J.; Yeung, M.; Ursprung, S.; Aviles-Rivero, A.I.; Etmann, C.; McCague, C.; Beer, L.; et al. Common pitfalls and recommendations for using machine learning to detect and prognosticate for COVID-19 using chest radiographs and CT scans. *Nat. Mach. Intell.* **2021**, *3*, 199–217. [[CrossRef](#)]
- Wilson, G.T.; Gopalakrishnan, P.; Tak, T. Noninvasive cardiac imaging with computed tomography. *Clin. Med. Res.* **2007**, *5*, 165–171. [[CrossRef](#)]
- Furqan Qadri, S.; Ai, D.; Hu, G.; Ahmad, M.; Huang, Y.; Wang, Y.; Yang, J. Automatic deep feature learning via patch-based deep belief network for vertebrae segmentation in CT images. *Appl. Sci.* **2019**, *9*, 69. [[CrossRef](#)]
- Ahmad, M.; Ai, D.; Xie, G.; Qadri, S.F.; Song, H.; Huang, Y.; Wang, Y.; Yang, J. Deep belief network modeling for automatic liver segmentation. *IEEE Access* **2019**, *7*, 20585–20595. [[CrossRef](#)]
- Zhang, B.; Qi, S.; Monkam, P.; Li, C.; Yang, F.; Yao, Y.D.; Qian, W. Ensemble learners of multiple deep CNNs for pulmonary nodules classification using CT images. *IEEE Access* **2019**, *7*, 110358–110371. [[CrossRef](#)]
- Tu, X.; Xie, M.; Gao, J.; Ma, Z.; Chen, D.; Wang, Q.; Finlayson, S.G.; Ou, Y.; Cheng, J.Z. Automatic categorization and scoring of solid, part-solid and non-solid pulmonary nodules in CT images with convolutional neural network. *Sci. Rep.* **2017**, *7*, 8533. [[CrossRef](#)]
- Liu, M.; Dong, J.; Dong, X.; Yu, H.; Qi, L. Segmentation of lung nodule in CT images based on mask R-CNN. In Proceedings of the 2018 9th International Conference on Awareness Science and Technology (iCAST), Fukuoka, Japan, 19–21 September 2018; pp. 1–6.
- Roth, H.R.; Lu, L.; Seff, A.; Cherry, K.M.; Hoffman, J.; Wang, S.; Liu, J.; Turkbey, E.; Summers, R.M. A New 2.5D Representation for Lymph Node Detection Using Random Sets of Deep Convolutional Neural Network Observations. In Proceedings of the Medical Image Computing and Computer-Assisted Intervention—MICCAI 2014, 17th International Conference, Boston, MA, USA, 14–18 September 2014; Golland, P., Hata, N., Barillot, C., Hornegger, J., Howe, R., Eds.; Springer International Publishing: Basel, Switzerland, 2014; pp. 520–527.

15. Ma, L.; Liu, X.; Gao, Y.; Zhao, Y.; Zhao, X.; Zhou, C. A new method of content based medical image retrieval and its applications to CT imaging sign retrieval. *J. Biomed. Inform.* **2017**, *66*, 148–158. [[CrossRef](#)] [[PubMed](#)]
16. Sivaranjini, S.; Sujatha, C. Deep learning based diagnosis of Parkinson's disease using convolutional neural network. *Multimed. Tools Appl.* **2020**, *79*, 15467–15479. [[CrossRef](#)]
17. Isaac, A.; Nehemiah, H.K.; Isaac, A.; Kannan, A. Computer-Aided Diagnosis system for diagnosis of pulmonary emphysema using bio-inspired algorithms. *Comput. Biol. Med.* **2020**, *124*, 103940. [[CrossRef](#)]
18. Oulefki, A.; Agaian, S.; Trongtirakul, T.; Laouar, A.K. Automatic COVID-19 lung infected region segmentation and measurement using CT-scans images. *Pattern Recognit.* **2021**, *114*, 107747. [[CrossRef](#)]
19. Thompson, D.; Lei, Y. Mini review: Recent progress in RT-LAMP enabled COVID-19 detection. *Sens. Actuators Rep.* **2020**, *2*, 100017. [[CrossRef](#)]
20. Thi, V.L.D.; Herbst, K.; Boerner, K.; Meurer, M.; Kremer, L.P.; Kirrmaier, D.; Freistaedter, A.; Papagiannidis, D.; Galmozzi, C.; Stanifer, M.L.; et al. A colorimetric RT-LAMP assay and LAMP-sequencing for detecting SARS-CoV-2 RNA in clinical samples. *Sci. Transl. Med.* **2020**, *12*, eabc7075.
21. Shatri, J.; Tafilaj, L.; Turkaj, A.; Dedushi, K.; Shatri, M.; Bexheti, S.; Mucaj, S.K. The role of chest computed tomography in asymptomatic patients of positive coronavirus disease 2019: A case and literature review. *J. Clin. Imaging Sci.* **2020**, *35*. [[CrossRef](#)]
22. Guan, W.j.; Ni, Z.y.; Hu, Y.; Liang, W.h.; Ou, C.q.; He, J.x.; Liu, L.; Shan, H.; Lei, C.l.; Hui, D.S.; et al. Clinical characteristics of coronavirus disease 2019 in China. *N. Engl. J. Med.* **2020**, *382*, 1708–1720. [[CrossRef](#)]
23. Bai, H.X.; Hsieh, B.; Xiong, Z.; Halsey, K.; Choi, J.W.; Tran, T.M.L.; Pan, I.; Shi, L.B.; Wang, D.C.; Mei, J.; et al. Performance of radiologists in differentiating COVID-19 from non-COVID-19 viral pneumonia at chest CT. *Radiology* **2020**, *296*, E46–E54. [[CrossRef](#)]
24. Mei, X.; Lee, H.C.; Diao, K.y.; Huang, M.; Lin, B.; Liu, C.; Xie, Z.; Ma, Y.; Robson, P.M.; Chung, M.; et al. Artificial intelligence—Enabled rapid diagnosis of patients with COVID-19. *Nat. Med.* **2020**, *26*, 1224–1228. [[CrossRef](#)]
25. Huang, Z.; Zhao, S.; Li, Z.; Chen, W.; Zhao, L.; Deng, L.; Song, B. The battle against coronavirus disease 2019 (COVID-19): emergency management and infection control in a radiology department. *J. Am. Coll. Radiol.* **2020**, *17*, 710–716. [[CrossRef](#)]
26. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet classification with deep convolutional neural networks. *Commun. ACM* **2012**, *60*, 84–90. [[CrossRef](#)]
27. Lundervold, A.S.; Lundervold, A. An overview of deep learning in medical imaging focusing on MRI. *Z. Für Med. Phys.* **2019**, *29*, 102–127. [[CrossRef](#)]
28. Di Ruberto, C.; Loddo, A.; Puglisi, G. Blob detection and deep learning for leukemic blood image analysis. *Appl. Sci.* **2020**, *10*, 1176. [[CrossRef](#)]
29. Hu, K.; Huang, Y.; Huang, W.; Tan, H.; Chen, Z.; Zhong, Z.; Li, X.; Zhang, Y.; Gao, X. Deep supervised learning using self-adaptive auxiliary loss for COVID-19 diagnosis from imbalanced CT images. *Neurocomputing* **2021**, *458*, 232–245. [[CrossRef](#)]
30. van Sloun, R.J.G.; Cohen, R.; Eldar, Y.C. Deep Learning in Ultrasound Imaging. *Proc. IEEE* **2020**, *108*, 11–29. [[CrossRef](#)]
31. Çallı, E.; Sogancıoğlu, E.; van Ginneken, B.; van Leeuwen, K.G.; Murphy, K. Deep learning for chest X-ray analysis: A survey. *Med. Image Anal.* **2021**, *72*, 102125. [[CrossRef](#)]
32. Shen, L.; Margolies, L.R.; Rothstein, J.H.; Fluder, E.; McBride, R.; Sieh, W. Deep learning to improve breast cancer detection on screening mammography. *Sci. Rep.* **2019**, *9*, 12495. [[CrossRef](#)]
33. Liu, M.; Zhang, J.; Adeli, E.; Shen, D. Landmark-based deep multi-instance learning for brain disease diagnosis. *Med. Image Anal.* **2018**, *43*, 157–168. [[CrossRef](#)] [[PubMed](#)]
34. Dadar, M.; Pascoal, T.A.; Manitsirikul, S.; Misquitta, K.; Fonov, V.S.; Tartaglia, M.C.; Breitner, J.; Rosa-Neto, P.; Carmichael, O.T.; Decarli, C.; et al. Validation of a regression technique for segmentation of white matter hyperintensities in Alzheimer's disease. *IEEE Trans. Med. Imaging* **2017**, *36*, 1758–1768. [[CrossRef](#)]
35. Padilla, P.; Lopez, M.; Gorrioz, J.M.; Ramirez, J.; Salas-Gonzalez, D.; Alvarez, I. NMF-SVM Based CAD Tool Applied to Functional Brain Images for the Diagnosis of Alzheimer's Disease. *IEEE Trans. Med. Imaging* **2012**, *31*, 207–216. [[CrossRef](#)] [[PubMed](#)]
36. Zhong, A.; Li, X.; Wu, D.; Ren, H.; Kim, K.; Kim, Y.; Buch, V.; Neumark, N.; Bizzo, B.; Tak, W.Y.; et al. Deep metric learning-based image retrieval system for chest radiograph and its clinical applications in COVID-19. *Med. Image Anal.* **2021**, *70*, 101993. [[CrossRef](#)]
37. Guan, Q.; Huang, Y.; Zhong, Z.; Zheng, Z.; Zheng, L.; Yang, Y. Thorax disease classification with attention guided convolutional neural network. *Pattern Recognit. Lett.* **2020**, *131*, 38–45. [[CrossRef](#)]
38. Wang, H.; Jia, H.; Lu, L.; Xia, Y. Thorax-net: an attention regularized deep neural network for classification of thoracic diseases on chest radiography. *IEEE J. Biomed. Health Inform.* **2019**, *24*, 475–485. [[CrossRef](#)]
39. Liu, X.; Wang, K.; Wang, K.; Chen, T.; Zhang, K.; Wang, G. KISEG: A Three-Stage Segmentation Framework for Multi-level Acceleration of Chest CT Scans from COVID-19 Patients. In *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention*; Springer: Berlin, Germany, 2020; pp. 25–34.
40. Parekh, M.; Donuru, A.; Balasubramanya, R.; Kapur, S. Review of the chest CT differential diagnosis of ground-glass opacities in the COVID era. *Radiology* **2020**, *297*, E289–E302. [[CrossRef](#)] [[PubMed](#)]
41. Zhao, J.; Zhang, Y.; He, X.; Xie, P. COVID-CT-Dataset: a CT scan dataset about COVID-19. *arXiv* **2020**, arXiv:2003.13865.
42. Chiroma, H.; Ezugwu, A.E.; Jauro, F.; Al-garadi, M.A.; Abdullahi, I.N.; Shuib, L. Early survey with bibliometric analysis on machine learning approaches in controlling COVID-19 outbreaks. *PeerJ Comput. Sci.* **2020**, *6*, e313. [[CrossRef](#)]

43. Signoroni, A.; Savardi, M.; Benini, S.; Adami, N.; Leonardi, R.; Gibellini, P.; Vaccher, F.; Ravanelli, M.; Borghesi, A.; Maroldi, R.; et al. BS-Net: Learning COVID-19 pneumonia severity on a large chest X-ray dataset. *Med. Image Anal.* **2021**, *71*, 102046. [CrossRef]
44. Born, J.; Beymer, D.; Rajan, D.; Coy, A.; Mukherjee, V.V.; Manica, M.; Prasanna, P.; Ballah, D.; Guindy, M.; Shaham, D.; et al. On the role of artificial intelligence in medical imaging of COVID-19. *Patterns* **2021**, *2*, 100269. [CrossRef]
45. Dong, D.; Tang, Z.; Wang, S.; Hui, H.; Gong, L.; Lu, Y.; Xue, Z.; Liao, H.; Chen, F.; Yang, F.; et al. The Role of Imaging in the Detection and Management of COVID-19: A Review. *IEEE Rev. Biomed. Eng.* **2021**, *14*, 16–29. [CrossRef]
46. Ouyang, X.; Huo, J.; Xia, L.; Shan, F.; Liu, J.; Mo, Z.; Yan, F.; Ding, Z.; Yang, Q.; Song, B.; et al. Dual-sampling attention network for diagnosis of COVID-19 from community acquired pneumonia. *IEEE Trans. Med. Imaging* **2020**, *39*, 2595–2605. [CrossRef]
47. Wang, X.; Deng, X.; Fu, Q.; Zhou, Q.; Feng, J.; Ma, H.; Liu, W.; Zheng, C. A weakly-supervised framework for COVID-19 classification and lesion localization from chest CT. *IEEE Trans. Med. Imaging* **2020**, *39*, 2615–2625. [CrossRef]
48. Jin, C.; Chen, W.; Cao, Y.; Xu, Z.; Tan, Z.; Zhang, X.; Deng, L.; Zheng, C.; Zhou, J.; Shi, H.; et al. Development and evaluation of an artificial intelligence system for COVID-19 diagnosis. *Nat. Commun.* **2020**, *11*, 5088. [CrossRef] [PubMed]
49. Zhang, K.; Liu, X.; Shen, J.; Li, Z.; Sang, Y.; Wu, X.; Zha, Y.; Liang, W.; Wang, C.; Wang, K.; et al. Clinically applicable AI system for accurate diagnosis, quantitative measurements, and prognosis of COVID-19 pneumonia using computed tomography. *Cell* **2020**, *181*, 1423–1433. [CrossRef] [PubMed]
50. Minaee, S.; Kafieh, R.; Sonka, M.; Yazdani, S.; Soufi, G.J. Deep-covid: Predicting covid-19 from chest x-ray images using deep transfer learning. *Med. Image Anal.* **2020**, *65*, 101794. [CrossRef] [PubMed]
51. Hu, S.; Gao, Y.; Niu, Z.; Jiang, Y.; Li, L.; Xiao, X.; Wang, M.; Fang, E.F.; Menpes-Smith, W.; Xia, J.; et al. Weakly supervised deep learning for covid-19 infection detection and classification from ct images. *IEEE Access* **2020**, *8*, 118869–118883. [CrossRef]
52. Polsinelli, M.; Cinque, L.; Placidi, G. A light CNN for detecting COVID-19 from CT scans of the chest. *Pattern Recognit. Lett.* **2020**, *140*, 95–100. [CrossRef]
53. Biswas, S.; Chatterjee, S.; Majee, A.; Sen, S.; Schwenker, F.; Sarkar, R. Prediction of COVID-19 from Chest CT Images Using an Ensemble of Deep Learning Models. *Appl. Sci.* **2021**, *11*, 7004. [CrossRef]
54. Gunraj, H.; Sabri, A.; Koff, D.; Wong, A. COVID-Net CT-2: Enhanced Deep Neural Networks for Detection of COVID-19 from Chest CT Images Through Bigger, More Diverse Learning. *arXiv* **2021**, arXiv:eess.IV/2101.07433.
55. Zhao, W.; Jiang, W.; Qiu, X. Deep learning for COVID-19 detection based on CT images. *Sci. Rep.* **2021**, *11*, 14353. [CrossRef] [PubMed]
56. Oyelade, O.N.; Ezugwu, A.E.; Chiroma, H. CovFrameNet: An Enhanced Deep Learning Framework for COVID-19 Detection. *IEEE Access* **2021**, *9*, 77905–77919. [CrossRef]
57. Simonyan, K.; Zisserman, A. Very Deep Convolutional Networks for Large-Scale Image Recognition. In Proceedings of the 3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, 7–9 May 2015.
58. Iandola, F.N.; Moskewicz, M.W.; Ashraf, K.; Han, S.; Dally, W.J.; Keutzer, K. SqueezeNet: AlexNet-level accuracy with 50× fewer parameters and <1 MB model size. *arXiv* **2016**, arXiv:1602.07360.
59. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2016, Las Vegas, NV, USA, 27–30 June 2016; IEEE Computer Society: Washington, DC, USA, 2016; pp. 770–778. [CrossRef]
60. Chollet, F. Xception: Deep learning with depthwise separable convolutions. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 1800–1807.
61. Krizhevsky, A.; *Learning Multiple Layers of Features from Tiny Images*; Technical Report; University of Toronto: Toronto, ON, Canada, 2009.
62. Jia, D.; Wei, D.; Socher, R.; Li, J.L.; Kai, L.; Li, F.-F. Imagenet: A large-scale hierarchical image database. In Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, USA, 20–25 June 2009; pp. 248–255.
63. Russakovsky, O.; Deng, J.; Su, H.; Krause, J.; Satheesh, S.; Ma, S.; Huang, Z.; Karpathy, A.; Khosla, A.; Bernstein, M.; et al. Imagenet large scale visual recognition challenge. *Int. J. Comput. Vis.* **2015**, *115*, 211–252. [CrossRef]
64. Huang, G.; Liu, Z.; Van Der Maaten, L.; Weinberger, K.Q. Densely connected convolutional networks. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 1–26 July 2017; pp. 4700–4708.
65. Cohen, J.P.; Morrison, P.; Dao, L. COVID-19 Image Data Collection. *arXiv* **2020**, arXiv:eess.IV/2003.11597.
66. Sajid, N. COVID-19 Patients Lungs X-ray Images 10000. 2021. Available online: <https://www.kaggle.com/nabeelsajid917/covid-19-x-ray-10000-images> (accessed on 28 August 2021).
67. MedSeg. COVID-19 CT Segmentation Dataset. 2021. Available online: <http://medicalsegmentation.com/covid19/> (accessed on 28 August 2021).
68. Gunraj, H. COVID-Net Open Source Initiative-COVIDx CT-2 Dataset. 2020. Available online: <https://www.kaggle.com/hgunraj/covidxct> (accessed on 30 June 2021).
69. Harmon, S.A.; Sanford, T.H.; Xu, S.; Turkbey, E.B.; Roth, H.; Xu, Z.; Yang, D.; Myronenko, A.; Anderson, V.; Amalou, A.; et al. Artificial intelligence for the detection of COVID-19 pneumonia on chest CT using multinational datasets. *Nat. Commun.* **2020**, *11*, 4080. [CrossRef] [PubMed]
70. Rahimzadeh, M.; Attar, A.; Sakhaei, S.M. A fully automated deep learning-based network for detecting covid-19 from a new and large lung ct scan dataset. *Biomed. Signal Process. Control* **2021**, *68*, 102588. [CrossRef]

71. Ning, W.; Lei, S.; Yang, J.; Cao, Y.; Jiang, P.; Yang, Q.; Zhang, J.; Wang, X.; Chen, F.; Geng, Z.; et al. Open resource of clinical data from patients with pneumonia for the prediction of COVID-19 outcomes via deep learning. *Nat. Biomed. Eng.* **2020**, *4*, 1197–1207. [[CrossRef](#)]
72. Ma, J.; Wang, Y.; An, X.; Ge, C.; Yu, Z.; Chen, J.; Zhu, Q.; Dong, G.; He, J.; He, Z.; et al. Towards efficient covid-19 ct annotation: A benchmark for lung and infection segmentation. *arXiv* **2020**, arXiv:2004.12537v2.
73. Armato III, S.G.; McLennan, G.; Bidaut, L.; McNitt-Gray, M.F.; Meyer, C.R.; Reeves, A.P.; Zhao, B.; Aberle, D.R.; Henschke, C.I.; Hoffman, E.A.; et al. The lung image database consortium (LIDC) and image database resource initiative (IDRI): a completed reference database of lung nodules on CT scans. *Med. Phys.* **2011**, *38*, 915–931. [[CrossRef](#)] [[PubMed](#)]
74. Bell, D.J. COVID-19. 2021. Available online: <https://radiopaedia.org/articles/covid-19-4> (accessed on 9 August 2021).
75. Alejo, R.; Antonio, J.A.; Valdovinos, R.M.; Pacheco-Sanchez, J.H. Assessments Metrics for Multi-class Imbalance Learning: A Preliminary Study. In Proceedings of the Pattern Recognition-5th Mexican Conference, MCP R 2013, Querétaro, Mexico, 26–29 June 2013; Springer: Berlin, Germany, 2013; Volume 7914, pp. 335–343. [[CrossRef](#)]
76. Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S.E.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; Rabinovich, A. Going deeper with convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2015, Boston, MA, USA, 7–12 June 2015; pp. 1–9.
77. Zhang, X.; Zhou, X.; Lin, M.; Sun, J. ShuffleNet: An Extremely Efficient Convolutional Neural Network for Mobile Devices. In Proceedings of the 2018 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2018, Salt Lake City, UT, USA, 18–22 June 2018; pp. 6848–6856. [[CrossRef](#)]
78. Sandler, M.; Howard, A.G.; Zhu, M.; Zhmoginov, A.; Chen, L. MobileNetV2: Inverted Residuals and Linear Bottlenecks. In Proceedings of the 2018 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2018, Salt Lake City, UT, USA, 18–22 June 2018; pp. 4510–4520. [[CrossRef](#)]
79. Szegedy, C.; Vanhoucke, V.; Ioffe, S.; Shlens, J.; Wojna, Z. Rethinking the Inception Architecture for Computer Vision. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2016, Las Vegas, NV, USA, 27–30 June 2016; pp. 2818–2826.
80. Jarrett, K.; Kavukcuoglu, K.; Ranzato, M.; LeCun, Y. What is the best multi-stage architecture for object recognition? In Proceedings of the IEEE 12th International Conference on Computer Vision, ICCV 2009, Kyoto, Japan, 27 September–4 October 2009; pp. 2146–2153. [[CrossRef](#)]
81. Richard, A.; Gall, J. A BoW-equivalent Recurrent Neural Network for Action Recognition. In Proceedings of the British Machine Vision Conference 2015, BMVC 2015, Swansea, UK, 7–10 September 2015; pp. 57.1–57.13. [[CrossRef](#)]
82. Zhang, J.; Xia, Y.; Xie, Y.; Fulham, M.J.; Feng, D.D. Classification of Medical Images in the Biomedical Literature by Jointly Using Deep and Handcrafted Visual Features. *IEEE J. Biomed. Health Inform.* **2018**, *22*, 1521–1530. [[CrossRef](#)]