Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Applied Soft Computing 132 (2023) 109851

Contents lists available at ScienceDirect

Applied Soft Computing

journal homepage: www.elsevier.com/locate/asoc

Automatic detection of Covid-19 from chest X-ray and lung computed tomography images using deep neural networks and transfer learning

Linh T. Duong [a], Phuong T. Nguyen [b,*], Ludovico Iovino [c], Michele Flammini [c]

[a] Institute of Research and Development, Duy Tan University, Viet Nam
[b] Department of Information Engineering, Computer Science and Mathematics University of L'Aquila, Italy
[c] Gran Sasso Science Institute, Italy

ARTICLE INFO

Article history:
Received 29 October 2020
Received in revised form 2 October 2022
Accepted 14 November 2022
Available online 24 November 2022

Keywords:
COVID-19
AI Diagnosis systems
Expert systems
Chest X-ray image
Lung CT images

ABSTRACT

The world has been undergoing the most ever unprecedented circumstances caused by the coronavirus pandemic, which is having a devastating global effect in different aspects of life. Since there are not effective antiviral treatments for Covid-19 yet, it is crucial to early detect and monitor the progression of the disease, thereby helping to reduce mortality. While different measures are being used to combat the virus, medical imaging techniques have been examined to support doctors in diagnosing the disease. In this paper, we present a practical solution for the detection of Covid-19 from chest X-ray (CXR) and lung computed tomography (LCT) images, exploiting cutting-edge Machine Learning techniques. As the main classification engine, we make use of EfficientNet and MixNet, two recently developed families of deep neural networks. Furthermore, to make the training more effective and efficient, we apply three transfer learning algorithms. The ultimate aim is to build a reliable expert system to detect Covid-19 from different sources of images, making it be a multi-purpose AI diagnosing system. We validated our proposed approach using four real-world datasets. The first two are CXR datasets consist of 15,000 and 17,905 images, respectively. The other two are LCT datasets with 2,482 and 411,528 images, respectively. The five-fold cross-validation methodology was used to evaluate the approach, where the dataset is split into five parts, and accordingly the evaluation is conducted in five rounds. By each evaluation, four parts are combined to form the training data, and the remaining one is used for testing. We obtained an encouraging prediction performance for all the considered datasets. In all the configurations, the obtained accuracy is always larger than 95.0%. Compared to various existing studies, our approach yields a substantial performance gain. Moreover, such an improvement is statistically significant.

1. Introduction

Covid-19 is a coronavirus-induced infection that can be associated with a coagulopathy and infection-induced inflammatory changes [1]. The disease poses a serious threat to public health, and thus in March 2020, the World Health Organization (WHO) declared Covid-19 a pandemic. At the time of writing, the virus has infected more than four hundreds millions of people, and has claimed over six million peoples' lives worldwide.[1] The clinical spectrum of the disease is very wide, ranging from fever, dry cough and diarrhea, but can be combined with mild pneumonia and mild dyspnoea. In some cases, the infection can evolve to severe pneumonia, causing approximately 5% of the infected patients to severe lung dysfunction. Given the circumstances, patients need ventilation as they are highly exposed to multiple extra pulmonary organ failure.

Since so far there have been no effective antiviral vaccines for Covid-19, it is crucial to reduce mortality by early detecting and monitoring the progression of the disease [2], so as to effectively personalize patient's treatment. Radiology is part of a fundamental process to detect whether or not the radiological outcomes are consistent with the infection and radiologists should expedite as much as possible the exploration, and provide accurate reports of their findings. Chest X-ray (CXR) and lung computed tomography (LCT) images of Covid-19 patients usually show multifocal, bilateral and peripheral lesions, but in the early phase of the disease they may present a unifocal lesion, most commonly located in the inferior lobe of the right lung. Considering the fact that the number of false positives obtained by swab results is considerably large, an expert system able to provide doctors with a preliminary diagnosis by automated recognition of Covid-19 from CXR or LCT images would be of crucial importance.

* Corresponding author.
E-mail addresses: duongtuanlinh@duytan.edu.vn (L.T. Duong), phuong.nguyen@univaq.it (P.T. Nguyen), ludovico.iovino@gssi.it (L. Iovino), michele.flammini@gssi.it (M. Flammini).

[1] https://www.worldometers.info/coronavirus/.

https://doi.org/10.1016/j.asoc.2022.109851
1568-4946/

The proliferation of disruptive Machine Learning (ML) and especially Deep Learning (DL) algorithms in recent years has enabled a plethora of applications across several domains [3,4]. Such techniques work on the basis of complex artificial neural networks, which are capable of effectively learning from data by means of a large number of hyper-parameters distributed in different network layers. In this way, they are able to simulate humans' cognitive functions, aiming to acquire real-world knowledge autonomously [5]. In a nutshell, ML/DL techniques are an advanced paradigm that brings in substantial improvement in performance compared to conventional learning algorithms. In the Health care sector, the potential of ML/DL to allow for rapid diagnosis of diseases has also been proven by various research work [6].

Aiming to assist the clinical care, this paper presents a practical solution for the detection of Covid-19 from CXR and LCT images exploiting two cutting-edge deep neural network families, i.e., EfficientNet [7] and MixNet [8]. Moreover, we empower the learning process by means of three different transfer learning strategies, namely **ImageNet** [9], **AdvProp** [10] and **Noisy Student** [11]. The evaluation on four large CXR and LCT images datasets demonstrate that our proposed models obtain a superior performance compared to the existing studies that we are aware of. In this respect, our work makes the following contributions:

- A system for detection of Covid-19 from CXR and LCT images exploiting cutting-edge deep learning algorithms;
- A successful empirical evaluation on four large datasets consisting of CXR and LCT images;
- A software prototype in the form of a mobile app ready to be downloaded.

The paper is structured as follows. Section 2 briefly introduces EfficientNet and MixNet as well as the transfer learning methods. In Section 3, we present the dataset and metrics used for our evaluation, while in Section 4 we analyze the experimental results. The related work is reviewed in Section 5. Finally, Section 6 discusses future work and concludes the paper.

2. Background

As a base for our presentation, Section 2.1 provides a background on two families of deep neural networks, i.e., EfficientNet and MixNet, which are used as the classification engine in our work. Afterwards, a brief introduction to transfer learning is given in Section 2.2.

2.1. EfficientNet and MixNet

Based on the observation that a better accuracy and efficiency can be obtained by imposing a balance between all network dimensions, EfficientNet [7] has been proposed by scaling in three dimensions, i.e., width, depth, and resolution, using a set of fixed scaling coefficients that meet some specific constraints. By the most compact configuration, i.e., EfficientNet-B0, there are 18 convolution layers in total, i.e., $D = 18$, and each layer is equipped with a kernel k(3,3) or k(5,5). The input image contains three color channels R, G, B, each of size 224×224. The next layers are scaled down in resolution to reduce the feature map size, but scaled up in width to increase accuracy. For instance, the second convolution layer consists of $W = 16$ filters, and the number of filters in the next convolution layer is $W = 24$. The maximum number of filters is $D = 1280$ in correspondence of the last layer, which is fed to the final fully connected layer. The other configurations of the EfficientNet family are generated from EfficientNet-B0 by means of different scaling values [7]. EfficientNet-B7 outperforms a CNN by achieving a better accuracy, while considerably reducing the number of parameters.

Generally, kernels of size k(3,3) [12,13], k(5,5) [14], or k(7,7) [15] are used as filters for deep neural networks. However, larger kernels can potentially improve a model's accuracy and efficiency. Furthermore, large kernels help to capture high-resolution patterns, while small kernels allow us to better extract low-resolution ones. To maintain a balance between accuracy and efficiency, the MixNet [8] family has been built based on the MobileNets architectures [13,16]. This network family also aims to reduce the number of parameters as well as FLOPs, i.e., the metric used to measure the computational complexity counted as the number of float-point operations (in billions). The most simple architecture of the MixNet family is MixNet-Small, which consists of a large number of layers and channels. Furthermore, the size of the filters varies depending on the layers. Similar to the EfficientNet family, other configurations of the MixNet family, such as MixNet-Medium or MixNet-Large, are derived from MixNet-S with different scaling values.

2.2. Transfer learning

To fine tune the hyper-parameters, i.e., the internal weights and biases, deep neural networks need a huge number of labeled data. Moreover, the deeper/wider a network is, the more parameters it possesses. As a result, deeper/wider networks require more data to guide the learning, with the aim of avoiding overfitting and being effective. Thus, it is crucial to train them with *enough* data, in order to facilitate the learning. Nevertheless, such a requirement is hard to come by in real-world settings, since the labeling process usually is done manually, thus being time consuming and subject to error [17]. To this end, transfer learning has been conceptualized as an effective way to extract and transfer the knowledge from a well-defined source domain to a novice target domain [18,19]. In other words, transfer learning facilitates the export of existing convolution weights from a model trained using large datasets to create new accurate models exploiting a relatively lower number of labeled images. As it has been shown in various studies [20,21], transfer learning remains helpful even when the target domain is quite different from the one in which the original weights have been obtained. In this work, we consider the following learning methods:

- **ImageNet** [9]: The ImageNet dataset has been widely exploited to apply transfer learning by several studies, since it contains more than 14 million images, covering miscellaneous categories;
- **AdvProp** [10]: adversarial propagation has been proposed as an improved training scheme, with the ultimate aim of avoiding overfitting. The method treats adversarial examples as additional examples, and uses a separate auxiliary batch norm for adversarial examples;
- **NS** [11]: the Noisy Student learning method attempts to improve ImageNet classification Noisy Student Training by: *(i)* enlarging the trainee/student equal to or larger than the trainer/teacher, aiming to make the trainee learn better on a large dataset, and *(ii)* adding noise to the student, thus forcing him to learn more.

To assist doctors in early detecting Covid-19 from CXR and LCT images, we develop an expert system that makes use of EfficientNet and MixNet as the classification engine. Moreover, in order to accelerate the learning process and to achieve a higher accuracy, we propose the three different learning strategies for obtaining network weights mentioned above, i.e., **ImageNet**, **AdvProp**, and **NS**. The succeeding section introduces the evaluation settings used to study the performance of our approach.

Table 1
Chest X-ray datasets.

Dataset	Type	Categories			Total
		Covid-19	Normal	Pneumonia	
$\mathbf{D}_1$	Train	98	7,966	5,447	13,511
	Test	10	885	594	1,489
	Total	**108**	**8,851**	**6,041**	**15,000**
$\mathbf{D}_2$	Train	261	8,154	5,909	14,324
	Test	66	2,038	1,477	3,581
	Total	**327**	**10,192**	**7,386**	**17,905**

Table 2
Lung CT datasets.

Dataset	Type	Categories			Total
		Covid-19	Normal	Pneumonia	
$\mathbf{D}_3$ [24]	Train	1,001	–	984	1,985
	Test	125	–	123	248
	Validation	126	–	123	249
	Total	**1,252**	–	**1,230**	**2,482**
$\mathbf{D}_4$ [25]	Train	109,250	67,029	111,791	288,070
	Test	23,410	14,364	23,955	61,729
	Validation	23,411	14,364	23,955	61,730
	Total	**156,071**	**95,756**	**159,702**	**411,528**

Table 3
Hardware and software configurations.

Name	Description
RAM	24 GB
CPU	Intel® Core™ i5-2400 CPU @ 3.10 GHz ×4
GPU	GeForce GTX 1080 Ti
OS	Ubuntu 18.04
Python	3.7.5
Pytorch	1.5
Torchvision	0.5.0
Numpy	1.15.4
Git	2.0
Timm	0.1.26

3. Evaluation

We explain in detail the datasets and methods used to study the performance of our proposed solution. Four real-world datasets have been used in the evaluation. Moreover, we make use of recent implementations[2] of EfficientNet and MixNet, which have been built atop the PyTorch framework.[3] We imported pre-trained weights from various sources to speed up the training. We publish the tool developed through this paper in GitHub to facilitate future research.[4]

3.1. Research objectives

Given an automated classifier, it is of crucial importance to get accurate predictions with respect to various quality metrics. In the evaluation, we focus on the following performance characteristics: *(i)* We identify which neural network family obtains the best prediction performance for the CXR datasets; *(ii)* Afterwards, we determine which deep neural network is the most suitable one for classifying LCT images; *(iii)* We ascertain which transfer learning method among **ImageNet**, **AdvProp**, and **NS** is beneficial to EfficientNet or MixNet; *(iv)* We measure the average recognition speed to see if the proposed model is feasible in practice in terms of timing efficiency; Finally, *(v)* A comparison with well-established baselines is performed to validate if our approach brings advancements in the field.

3.2. Datasets

To investigate the performance of our approach, with CXR we exploited two datasets used in some previous studies [22,23]. Their characteristics are summarized in Tables 1 and 2. In all of them, except $\mathbf{D}_3$, there are three categories, i.e., *Covid-19*, *Normal*, and *Pneumonia*. Only the category *Normal* exhibits no symptoms, while the other two categories, *Covid-19* and *Pneumonia*, correspond to different levels of infection-induced inflammatory changes.

Dataset $\mathbf{D}_1$ consists of 13,511 images for training and 1489 images for testing. It can be seen that there is an imbalance among the categories. In particular, both categories *Normal* and *Pneumonia* contain a large number of images, while there are only 108 images in the *Covid-19* category. We use $\mathbf{D}_1$ as a means to compare the performance of our approach with some existing studies that exploited $\mathbf{D}_1$ in their evaluation [22,23]. Compared to $\mathbf{D}_1$, $\mathbf{D}_2$ is newly updated with more data for training and testing. This means that there are overlaps between the two datasets. In particular, $\mathbf{D}_2$ consists of 14,324 and 3581 images for training and testing, respectively. We made use of $\mathbf{D}_2$ to validate the performance of our approach on a larger amount of data, proving its feasibility in real-world settings.

Table 2 summarizes the two LCT datasets used in our evaluation. Dataset $\mathbf{D}_3$ contains a small number of images in two categories, i.e., *Covid-19* and *Pneumonia*. Finally, $\mathbf{D}_4$ is much larger with 156,017 images for the *Covid-19* category. In total, $\mathbf{D}_4$ sums up to 411,528 images for training and testing. We believe that such a large dataset resembles a real scenario, and thus the evaluation on it helps to shed light on the performance of the approach in practice.

Concerning the settings, at the beginning of all experiments the datasets were randomly split using a random seeds function of the Python numpy library. With respect to the strategy it ensures that the test sets are independent and all the same for the other evaluations of each architecture. In other words, we can measure and compare prediction performance on the same criterion. This is done by following many existing studies [4,26–28], which measured the prediction performance on independent test sets.

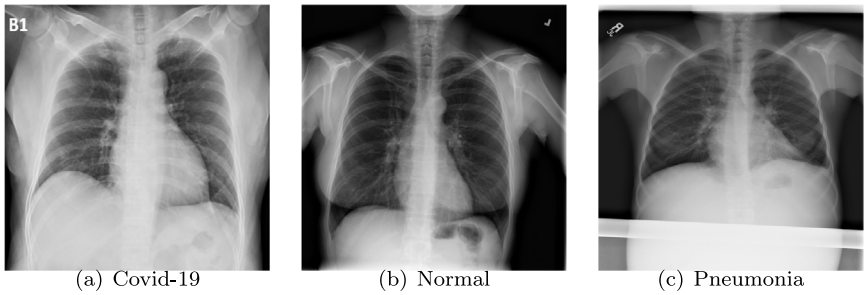
Figs. 1 and 2 give examples of CXR and LCT images for the three categories mentioned above. By plain eyes, it is quite difficult to notice the difference between them, especially for non-expert users. In this respect, we assume that the application of deep learning will facilitate the recognition and effectively assist doctors in diagnosing the disease.

3.3. Settings

Deep neural networks such as EfficientNet and MixNet require a platform with powerful computational performance. To perform the evaluation, we used a server with the hardware and software configurations listed in Table 3.

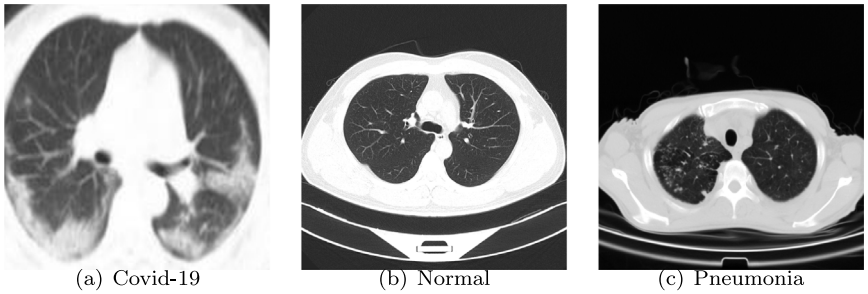
By means of an empirical study, we found out that two configurations, EfficientNet-B0 and EfficientNet-B3, are more effective than the others for the EfficientNet family, and thus we selected them for our evaluation. Weights for EfficientNet are obtained by all the transfer learning techniques mentioned in Section 2.2. In the MixNet family, we considered four different configurations, i.e., MixNet-Small, MixNet-Medium, MixNet-Large and MixNet-XL. We obtained weights coming from the available ImageNet dataset. At the end, by combining two network families with

[2] https://github.com/rwightman/gen-efficientnet-pytorch.
[3] https://pytorch.org.
[4] https://github.com/linhduongtuan/Covid-19_Xray_Classifier/.



(a) Covid-19 (b) Normal (c) Pneumonia

Fig. 1. Examples of CXR images.



(a) Covid-19 (b) Normal (c) Pneumonia

Fig. 2. Examples of LCT images.

Table 4
Experimental configurations.

Conf.	Network	Batch size	# of Params	Learning method	Size (MB)
C_1	EfficientNet-B0	110	7,919,391	ImageNet	53.1
C_2	EfficientNet-B0	110	7,919,391	AdvProp	53.1
C_3	EfficientNet-B0	110	7,919,391	NS	53.1
C_4	EfficientNet-B3	64	**14,352,075**	ImageNet	**106.9**
C_5	EfficientNet-B3	64	**14,352,075**	AdvProp	**106.9**
C_6	EfficientNet-B3	64	**14,352,075**	NS	**106.9**
C_7	MixNet-Small	110	6,253,449	ImageNet	41.8
C_8	MixNet-Medium	90	7,133,225	ImageNet	48.9
C_9	MixNet-Large	60	9,448,095	ImageNet	67.5
C_{10}	MixNet-XL	60	14,015,611	ImageNet	104.2

three learning strategies mentioned in Section 2.2, we got 10 experimental configurations $C_1, \ldots, C_{10}$, as shown in Table 4. *Batch size* corresponds to the number of items used for each training step; *# of Params* specifies the number of parameters used by each network; and finally *Size* is the file size needed to store the parameters. It is evident that EfficientNet-B3 is the largest network with respect to the number of parameters as well as the file size to store them. To be concrete, i.e., C_4, C_5, and C_6 have more than 14 millions of parameters, accounting for more than 100MB of storage space each. In the evaluation, we employed the five-fold cross-validation methodology on the datasets. Namely, each dataset is split into five equal parts, and in each validation round one part is used for testing and the other four ones for the training.

3.4. Evaluation metrics

Each image in all the datasets has been manually labeled, i.e., either *Normal* or *Pneumonia* or *Covid-19*, resulting in three independent groups, i.e., $G = (G_1, G_2, G_3)$, called ground-truth data. Using either EfficientNet or MixNet as the classifier on a test set, three predicted sets, i.e., $S = (S_1, S_2, S_3)$ of images are obtained. We measured the classification performance by evaluating the similarity of the classified categories with the ground-truth ones. To this end, three metrics, namely *accuracy*, *precision* and *recall*, and F_1 *score* are used [21]. We selected these metrics for the following reason: precision, recall and F_1 are useful, given that the number of positive images accounts for a very small percentage of all the items in the dataset. In fact, in this case a classifier always providing a negative prediction would have a very high accuracy.

True positive $TP_i = |G_i \cap S_i|$, $i = 1, 2, 3$ is defined as the number of items that appear both in the results and ground-truth data of class *i*. The metrics are defined as follows.

Accuracy: It is the fraction of correctly classified items with respect to the total number of images in the test set.

$$accuracy = \frac{\sum_i^3 TP_i}{\sum_i^3 |G_i|} \times 100\% \tag{1}$$

Precision and Recall: *Precision* is the fraction of classified images for S_i being found in the ground-truth data G_i, while *Recall* the fraction of true positives being found in the ground-truth data.

$$precision_i = \frac{TP_i}{|S_i|} \tag{2}$$

$$recall_i = \frac{TP_i}{|G_i|} \tag{3}$$

F_1 score (F-Measure): It is calculated as the average of precision and recall using the following formula:

$$F_1 = \frac{2 \cdot precision_i \cdot recall_i}{precision_i + recall_i} \tag{4}$$

Table 5
Experimental results on dataset $\mathbf{D}_1$.

Configuration		C_1	C_2	C_3	C_4	C_5	C_6	C_7	C_8	C_9	C_{10}
Accuracy (%)		95.64	95.77	95.30	96.17	96.64	95.90	95.30	95.98	96.11	96.37
Precision	Covid-19	1.000	1.000	1.000	1.000	0.875	0.857	1.000	1.000	1.000	1.000
	Normal	0.952	0.960	0.950	0.957	0.968	0.957	0.953	0.966	0.961	0.964
	Pneu.	0.961	0.954	0.956	0.968	0.964	0.963	0.951	0.950	0.960	0.962
Recall	Covid-19	0.300	0.300	0.300	0.300	0.700	0.600	0.300	0.400	0.400	0.600
	Normal	0.981	0.975	0.977	0.985	0.978	0.978	0.974	0.971	0.978	0.978
	Pneu.	0.929	0.942	0.927	0.937	0.952	0.936	0.932	0.951	0.944	0.947
F_1-score	Covid-19	0.461	0.461	0.461	0.461	0.778	0.705	0.461	0.571	0.571	0.750
	Normal	0.967	0.967	0.963	0.971	0.973	0.967	0.963	0.969	0.969	0.971
	Pneu.	0.945	0.948	0.941	0.953	0.958	0.949	0.942	0.954	0.952	0.955

While accuracy is important, we are also interested in efficiency, considering the fact that the model needs to have a high recognition speed in practice.

Speed: The system presented in Table 3 is used to benchmark the processing time, i.e., the average number of predicted items in a second.

In the next section, we present in detail the experimental results by referring to the research questions introduced in Section 3.1.

4. Results

This section reports and analyzes the results obtained from the conducted experiments by referring to the research objectives presented in Section 3.1.

4.1. Performance on the CXR datasets

The results obtained by performing the experiments on $\mathbf{D}_1$ are shown in Table 5. The accuracy for all configurations is always larger than 95%, and the maximum accuracy is 96.64% obtained by C_5, i.e., EfficientNet-B3 using pre-trained weights with **AdvProp**. With respect to Precision, eight among ten configurations get 1.000 as precision for the *Covid-19* category. This means that all images classified as *Covid-19* by the classifiers are actually *Covid-19*. For the other two categories, i.e., *Normal* and *Pneumonia*, the maximum precision is 0.968, achieved also by $\mathbf{C}_5$ for Category *Normal*, and by $\mathbf{C}_4$ for Category *Pneumonia*. Altogether, we see that all the classifiers are able to recognize the testing images, obtaining high precision.

Concerning recall, all the configurations get a considerably low score for the *Covid-19* category: the highest recall is 0.700, obtained by C_5. This means that the models are not able to retrieve all the items in the ground-truth data, though they can yield good predictions for the category. We assume that this happens due to the limited data available in the training set. Referring back to Table 1, for the *Covid-19* category there are only 98 images and 10 images for training and testing, respectively. Meanwhile, for other two image categories, the recall scores are substantially improved. The best performance is seen by category *Normal*, i.e., 0.985; while by Pneumonia, recall is 0.952. As depicted in Table 1, these categories consist of a larger number of training and testing images compared to the *Covid-19*.

For what concerns the F_1 scores, for the *Covid-19* category the maximum F_1 is 0.778, obtained by C_5. The classifiers obtain a low F_1 in the other configurations, and this happens due to the low recall scores as shown above. For the other two categories *Normal* and *Pneumonia*, the F_1 scores are improved considerably compared to *Covid-19*. C_5 is the most suitable configuration as it obtains the best F_1 scores for all the categories, i.e., also 0.973 for *Normal* and 0.958 for *Pneumonia*.

Similarly, let us analyze the results on dataset $\mathbf{D}_2$ presented in Table 6 to ascertain the best network configuration. It is evident that C_1 achieves most of the best scores with respect to various metrics as well as categories. For instance, C_1 get 95.82% as accuracy, together with C_8, which is the best over all the configurations.

Altogether, through Tables 5 and 6 we can see that C_1 and C_5 are the configurations among the others that bring the best prediction performance.

Compared to existing work that performs evaluation on the same dataset [22,29], our approach achieves a better performance with respect to accuracy, precision, recall, and F_1-score. For instance, the work by Wang et al. [22], the maximum accuracy is 93.0% with similar experimental settings. In this respect, we conclude that application of the two network families EfficientNet and MixNet as well as the different transfer learning techniques brings a good prediction performance on the considered dataset.

We conclude that EfficientNet and MixNet can successfully predict Covid-19 from CXR images, obtaining a high accuracy and precision. Among others, EfficientNet yields the best prediction performance.

4.2. Performance on the LCT datasets

The experimental results for the two LCT datasets, i.e., $\mathbf{D}_3$ and $\mathbf{D}_4$ are shown in Tables 7 and 8, respectively. Table 7 demonstrates that EfficientNet-B0 trained with ImageNet is the most effective configuration, i.e., it obtains 97.99% as accuracy. Moreover, this configuration also gets the best F_1 score for both categories.

Concerning $\mathbf{D}_4$, the dataset consists of a large number of images (cf. Table 2). This resembles a real-world scenario where images are collected and fed to the system on a daily basis, resulting in a big database. We see that both network families get a high performance with respect to different metrics. All the configurations obtain an accuracy larger than 98% with 99.66% as maximum accuracy.

In comparison to existing studies that performed evaluation on same datasets, we can see that our proposed framework achieves a better performance. For instance, He et al. [30] obtain 87.93% as maximum accuracy, while we get 97.99% and 99.66% for $\mathbf{D}_3$ and $\mathbf{D}_4$, respectively.

On the LCT datasets, EfficientNet is also the network family that helps to obtain the best prediction performance.

4.3. The benefit of transfer learning

We conducted experiments following the five-fold cross-validation technique. Moreover, to further investigate the applicability of the proposed approach, we made use of $\mathbf{D}_2$ and $\mathbf{D}_4$, which contain more images than $\mathbf{D}_1$ and $\mathbf{D}_3$ (cf. Table 1). Figs. 3(a), 3(b), and 3(c) depict the confusion matrices for EfficientNet-B0 using the three different transfer learning techniques mentioned

Table 6
Experimental results on dataset $\mathbf{D}_2$.

Configuration		C_1	C_2	C_3	C_4	C_5	C_6	C_7	C_8	C_9	C_{10}
Accuracy (%)		95.82	94.39	93.30	95.05	95.59	95.00	95.79	95.82	95.68	95.53
Precision	Covid-19	0.968	0.948	0.889	0.950	0.968	0.978	0.982	0.966	0.983	0.967
	Normal	0.958	0.942	0.932	0.942	0.955	0.953	0.957	0.957	0.951	0.951
	Pneu.	0.957	0.946	0.935	0.964	0.955	0.944	0.958	0.959	0.963	0.960
Recall	Covid-19	0.924	0.560	0.363	0.863	0.924	0.667	0.833	0.863	0.909	0.909
	Normal	0.958	0.942	0.932	0.942	0.955	0.944	0.974	0.973	0.975	0.960
	Pneu.	0.942	0.922	0.906	0.912	0.939	0.937	0.941	0.941	0.933	0.932
F_1-score	Covid-19	0.945	0.704	0.616	0.905	0.945	0.792	0.901	0.922	0.945	0.937
	Normal	0.644	0.956	0.951	0.959	0.962	0.960	0.965	0.965	0.963	0.962
	Pneu.	0.950	0.934	0.920	0.940	0.947	0.940	0.950	0.950	0.947	0.946

Table 7
Experimental results on dataset $\mathbf{D}_3$.

Configuration		C_1	C_2	C_3	C_4	C_5	C_6	C_7	C_8	C_9	C_{10}
Accuracy (%)		97.99	96.24	94.09	97.31	97.58	97.31	96.77	97.31	96.51	97.81
Prec.	Covid-19	0.991	0.972	0.982	0.978	0.994	0.973	0.973	0.978	0.978	0.957
	Pneu.	0.968	0.952	0.905	0.967	0.991	0.973	0.962	0.967	0.958	0.957
Rec.	Covid-19	0.968	0.951	0.898	0.967	0.958	0.972	0.962	0.967	0.951	0.978
	Pneu.	0.991	0.973	0.983	0.978	0.994	0.973	0.973	0.978	0.978	0.967
F_1	Covid-19	0.979	0.962	0.938	0.973	0.975	0.973	0.967	0.973	0.964	0.967
	Pneu.	0.979	0.962	0.943	0.973	0.975	0.973	0.967	0.973	0.965	0.967

Table 8
Experimental results on dataset $\mathbf{D}_4$.

Configuration		C_1	C_2	C_3	C_4	C_5	C_6	C_7	C_8	C_9	C_{10}
Accuracy (%)		99.66	99.52	98.85	99.59	99.62	99.47	99.52	98.41	99.59	99.53
Precision	Covid-19	0.997	0.995	0.991	0.997	0.996	0.995	0.994	0.990	0.996	0.996
	Normal	0.990	0.989	0.977	0.988	0.989	0.987	0.990	0.962	0.989	0.986
	Pneu.	0.999	0.998	0.992	0.999	0.999	0.998	0.999	0.991	0.999	0.998
Recall	Covid-19	0.994	0.994	0.988	0.993	0.994	0.992	0.994	0.983	0.994	0.992
	Normal	0.996	0.991	0.983	0.995	0.995	0.991	0.990	0.978	0.994	0.995
	Pneu.	0.992	0.998	0.991	0.999	0.998	0.998	0.998	0.988	0.998	0.999
F_1-score	Covid-19	0.995	0.994	0.989	0.995	0.995	0.994	0.994	0.986	0.995	0.994
	Normal	0.993	0.990	0.980	0.991	0.992	0.989	0.990	0.970	0.991	0.990
	Pneu.	0.999	0.998	0.992	0.999	0.999	0.998	0.998	0.989	0.999	0.998

in Section 2.2. The metrics for all the confusion matrices are shown in Table 6.

As we can see, each transfer learning method may have different effects on the different categories. For example, using EfficientNet-B0 with weights pre-trained by ImageNet is beneficial to *Covid-19* and *Pneumonia*, but not to *Normal*. As shown in Fig. 3(a), 61 out of 66 images in *Covid-19* are correctly classified, while for *Pneumonia* 1,392 out of 1,477. However, for the *Normal* category, only 1978 images are correctly classified over a total of 2,038 images, accounting for 97.05%. On the other hand, transfer learning with **AdvProp** (cf. Fig. 3(b)) induces a better performance for *Normal*, i.e., 1,981 among 2,038 images are classified to the correct categories. In Fig. 3(c), we see that compared to the other learning methods, **NS** has an adverse effect on the recognition of all the categories. Altogether, we come to the conclusion that training EfficientNet-B0 with weights from ImageNet yields the best prediction performance.

For EfficientNet-B3, we see that weights pre-trained with ImageNet are beneficial to the *Normal* category (cf. Fig. 3(d)). At the same time, **AdvProp** is the transfer learning method that is suitable for recognition of *Pneumonia*, i.e., it helps to detect 1,388 out of 1,477 pneumonia images, which is the best among the others.

Among the configurations, C_9 is the best one for the *Pneumonia* category. Other MixNet configurations do not outperform the ones of EfficientNet-B0 and EfficientNet-B3. MixNet-XL obtains a considerably good performance with Category *Normal*, correctly classifying 1,984 images among 2,038 images, while it suffers of a low precision and recall for the other categories. For instance, with *Pneumonia*, only 1,377 out of 1,477 images are properly recognized by MixNet-XL with weights pre-trained with ImageNet. The experimental results demonstrate that, depending on the network family, each transfer learning technique has a diverse influence on the final outcomes. By considering the results in Table 6, it is evident C_1, corresponding to training EfficientNet-B0 with weights by ImageNet, is the most effective configuration with respect to accuracy, precision, recall, and F_1 for almost all categories. Moreover, together with the results obtained from Section 4.1, *we conclude that* ***ImageNet*** *is the best transfer learning strategy for both network families on the two datasets* $\mathbf{D}_1$ *and* $\mathbf{D}_2$.

Next, we consider the results obtained for the LCT dataset in Figs. 4(a)–4(j). In general, training the deep neural networks with ImageNet helps to obtain a better performance compared to the other transfer learning techniques in all the three categories. As an example, with C_1 we get 23,281 correctly classified Covid-19 images, while the corresponding numbers for AdvProp and NS are 23,270 and 23,137, respectively. For other categories, training EfficientNet-B0 with ImageNet weights also gets a good performance.

The experimental results show that weights pre-trained from the ImageNet dataset contribute to the best prediction performance on all the CXR and LCT datasets.

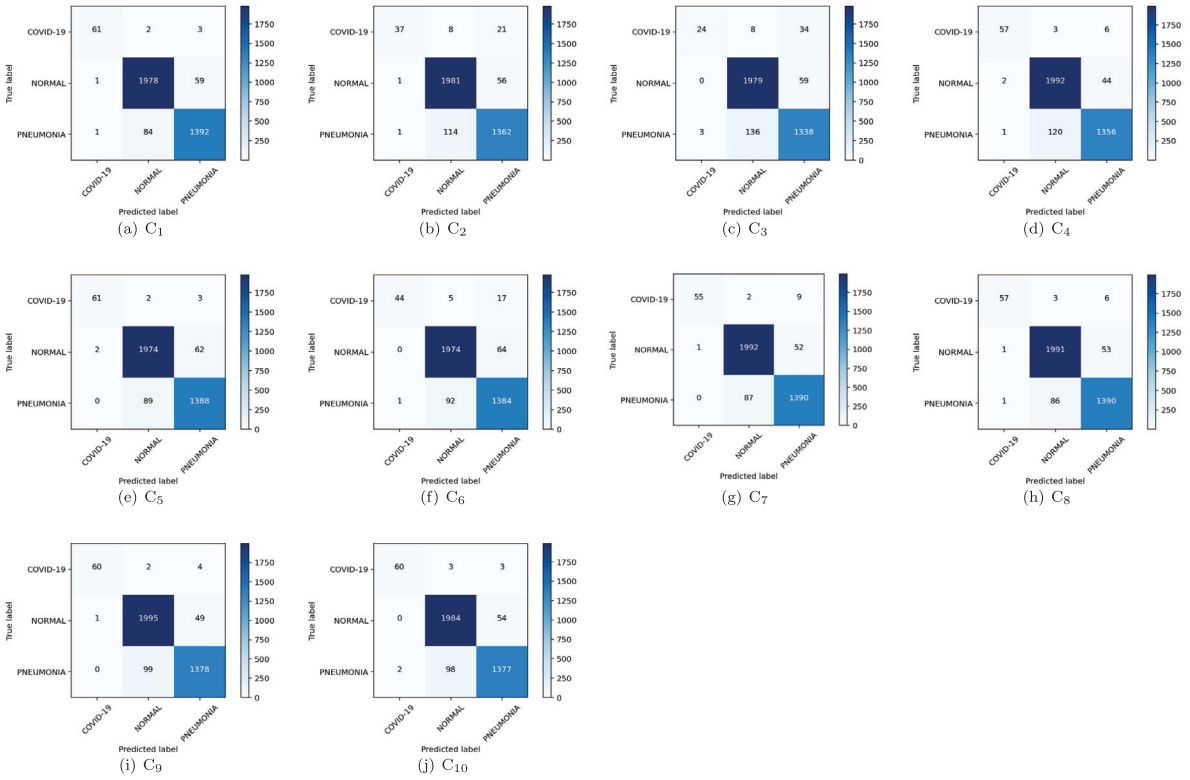


Fig. 3. Confusion matrices of EfficientNet and MixNet using different transfer learning techniques on $\mathbf{D}_2$.

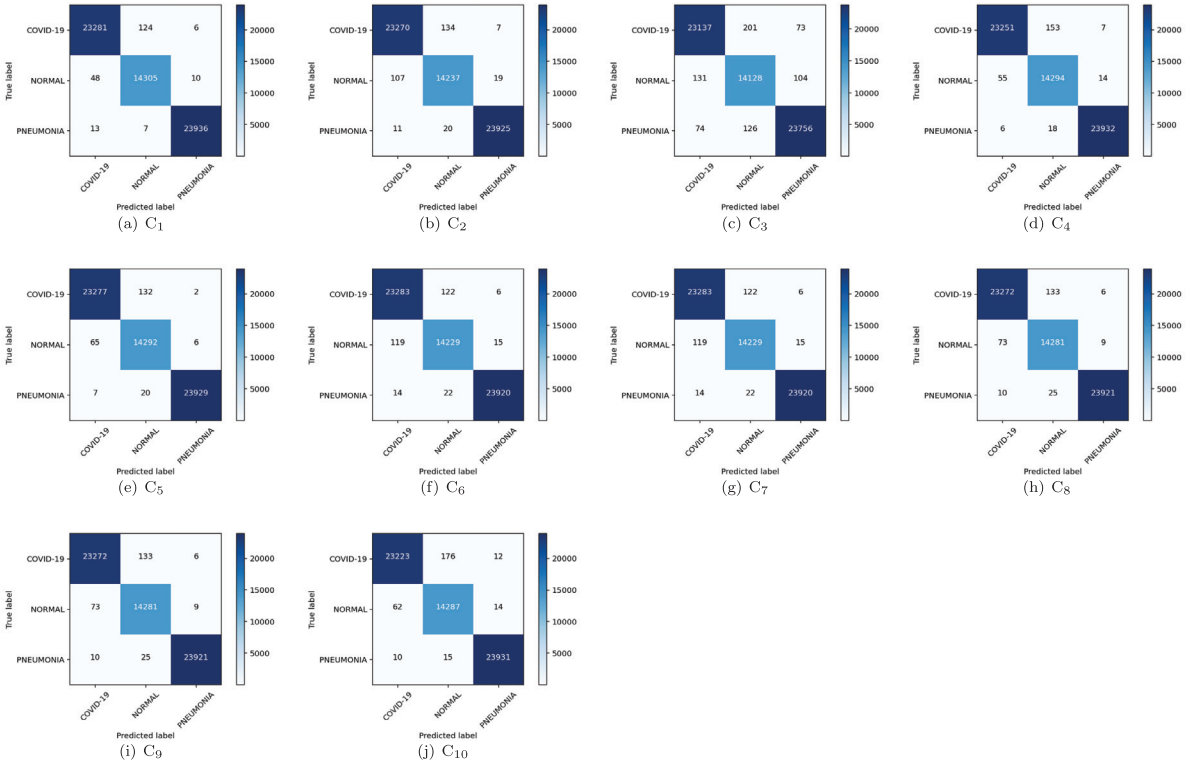


Fig. 4. Confusion matrices of EfficientNet-B0 and EfficientNet-B3 using different transfer learning techniques on $\mathbf{D}_4$.

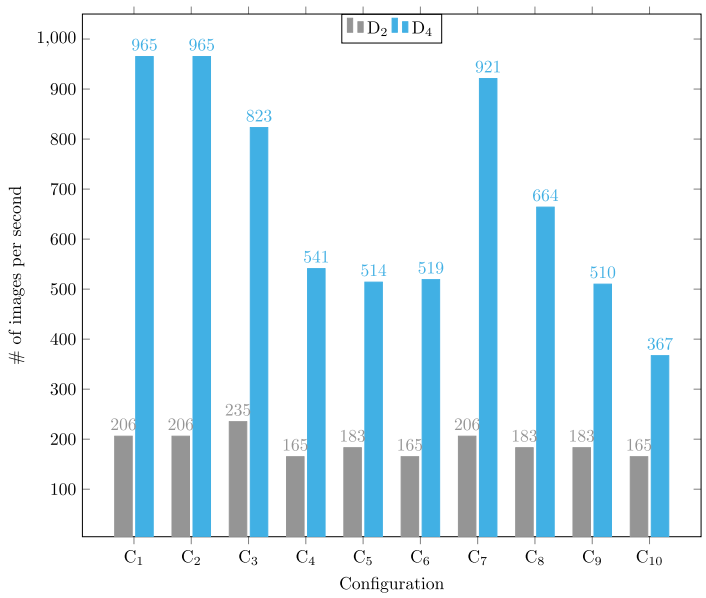


Fig. 5. Recognition speed for the configurations on dataset $\mathbf{D}_2$ and $\mathbf{D}_4$.

4.4. Timing efficiency

We investigate whether the model is feasible in practice by measuring the time efficiency. The system presented in Table 3 is used to benchmark the processing time, i.e., the average number of predicted items in a second for the cross-validation experiments. For these experiments, $\mathbf{D}_2$ and $\mathbf{D}_4$ have been selected as they contain a larger number of images compared to $\mathbf{D}_1$ and $\mathbf{D}_3$. We counted the number of predictions returned by the classifiers in a second for the two datasets. The results are depicted in Fig. 5.

Concerning $\mathbf{D}_2$, from the figure it is clear that the considered configurations have a comparable recognition speed. For instance, C_1, C_2, C_3, and C_7 corresponding to using EfficientNet-B0 and MixNet-S as the classification engine, are the most efficient configurations, as they return 206 and 235 images per second in average. EfficientNet-B3 also yields a good time performance, i.e., using C_4, C_5, or C_6 as the experimental configuration, the system generates 165 and 183 predictions per second.

The recognition speed is improved a lot by $\mathbf{D}_4$, i.e., on the LCT dataset with more images for training, compared to $\mathbf{D}_2$. The maximum speed is 965 images per second and reached by C_1, C_2, i.e., using EfficientNet-B0 together with **ImageNet** and **AdvProp**. Among the MixNet configurations, MixNet-S is the most efficient one, as it classifies 921 images per second, which is much higher compared to 664, 510 and 367 by MixNet-M, MixNet-L, and MixNet-XL, respectively.

Together with the results obtained from Section 4.1 and Section 4.2, the results demonstrate that the proposed framework is feasible in real-world settings, i.e., it can be used to effectively and efficiently predict Covid-19 from CXR and LCT images. Moreover, EfficientNet-B0 is the most efficient model on both datasets, while the MixNet family yields a low timing efficiency.

4.5. Comparison with a baseline

ResNet is a modern deep neural network, and it has been widely used in different classification tasks [31]. For the detection of Covid, ResNet50 has been applied in various work [32,33], obtaining a promising prediction performance. In this section, we run experiments to compare our approach with two baselines built on top of the ResNet50 architecture. For the comparison, we use the larger dataset among either the Chest X-ray and Lung Computed Tomography categories, i.e., $\mathbf{D}_2$ and $\mathbf{D}_4$, aiming to study the generalizability.

Table 9
Results obtained with ResNet50.

Configuration		$\mathbf{D}_2$	$\mathbf{D}_4$
Accuracy (%)		87.53	91.93
Precision	Covid-19	0.880	0.909
	Normal	0.843	0.922
	Pneu.	0.900	**0.923**
Recall	Covid-19	0.846	0.926
	Normal	0.902	0.813
	Pneu.	0.874	**0.952**
F_1-score	Covid-19	0.863	0.918
	Normal	0.872	0.864
	Pneu.	0.887	**0.938**

The experimental results obtained using ResNet50 are shown in Table 9. The approach gets an accuracy of 87.53% and 91.93% on $\mathbf{D}_2$ and $\mathbf{D}_4$, respectively. ResNet50 performs the best on the Pneumonia, i.e., it always gets a maximum value for all the three metrics (Precision, Recall, and F1-score). Concerning Precision, ResNet50 gets the maximum value of 0.923 by the Pneumonia category on $\mathbf{D}_4$. The corresponding scores for Recall and F1-score are 0.952 and 0.938, respectively.

While ResNet50 obtains a good prediction, compared to our approach, its performance is considerably lower. For instance, as shown in Table 8, on the $\mathbf{D}_4$ dataset, with Configuration $\mathbf{C}_1$ our approach gets 99.66% as accuracy. The maximum Precision, Recall, and F1 scores are 0.999, 0992, and 0.999, respectively. Finally, we investigate if the gain in performance is statistically significant. To this end, we perform a Wilcoxon signed-rank test [34] with a significance level of $\alpha = 5\%$ on the obtained Precision, Recall, and F1 scores from both approaches. The final p-value is 1.903×10^{-7}, which means that the difference is statistically significant. Altogether, this demonstrates that compared to the baselines, we always achieve a better performance.

The original EfficientNet-B0 configuration contains 5.29M params, 385.89 MMac, and it obtains an accuracy of 76.3% on the ImageNet-1K dataset [9]. However, we noticed that it cannot distinguish well between Covid-19 with other pneumonia. Thus, we attempted to improve this by modifying the original configuration by adding 12 layers to boost up the prediction accuracy

for the medical image classification for Sars-CoV-2 based on Chest X-ray (CXR) and Computed Tomography (CT). The new proposed model EfficientNet-B0 contains 7.92 M params, but still keeps the same number of MMac, i.e., 385.89 MMac. In summary, the modification of the original backbones is our key contribution for the classification task.

4.6. Threats to validity

We discuss the threats that may affect the internal, external, construct, and conclusion validity of our findings.

Internal validity. This concerns the internal factors that might have an adverse influence on the findings. A possible threat here could come from the results for the *Covid-19* category, since they are obtained with a considerably low number of items for training and testing, i.e., $\mathbf{D}_1$ with 98 and 10 images and $\mathbf{D}_2$ with 327 and 98 images for training and testing, respectively. This threat is mitigated by the other two categories in the datasets, as they contain a considerably large number of items. To the best of our knowledge, there exists no CXR dataset with more images for the *Covid-19* category. As a matter of fact, research in medical imaging on Covid-19 suffers a general lack of data. For this reason, unfortunately we are not able to study the models on a larger scale.

External validity. The main threat to *external validity* is due to the factors that might hamper the generalizability of our results to other scenarios outside the scope of this work, e.g., in practice we may encounter a limited amount of training data. The threat is moderated by evaluating EfficientNet and MixNet using the experimental settings following the five-fold cross-validation methodology. In particular, the dataset is split into five parts, and accordingly the evaluation is conducted in five rounds. By each round, four parts are combined to form the training data, and the remaining one is used for testing.

Construction validity. This is related to the experimental settings presented in the paper, concerning the simulation performed to evaluate the system. To mitigate the threat, the evaluation has been conducted on a training set and a test set, attempting to simulate a real usage where training data is already available for feeding the system, while testing data is the part that needs to be predicted. Depending on the given settings, we ran several rounds of training/testing, to make sure that the final results are generalizable, i.e., they are not obtained by chance. In the paper, we reported the most stable experimental results.

Conclusion validity. This is related to the factors that might have an impact on the obtained outcome. The evaluation metrics accuracy, precision, recall, F_1 and execution time might induce a conclusion threat. To face the issue, we adopted such measures as recommended by the previous scientific literature related to our setting, and employed the same metrics for evaluating all the classifiers.

5. Related work

The recent months have witnessed a large number of studies related to the topic Covid-19 and Machine Learning, and multiple Covid-19/ML applications have been proposed. We summarize in Table 10 some of the most notable research, providing the number of considered images for each category as well as the prediction accuracy. In this work, since we support the recognition of Covid-19 from CXR and LCT images, in the remaining of this section we concentrate on analyzing these studies.

Deep learning techniques have been exploited to predict which current antivirals might be more effective in patients infected with coronavirus [48,49]. Similarly, a specific model has been developed to forecast if a Covid-19 patient has the chance to survive based on his personal data as well as other risk factors [50]. Jiang et al. [51] propose an approach to the identification of clinical characteristics of Covid-19, and develop an AI tool to recognize patients at risk of a more severe impact of the disease.

Ozturk et al. [52] proposed an approach to deliver accurate diagnostics for binary classification, i.e., Covid-19 vs. No-Findings and multi-nominal classification, i.e., Covid-19 and No-Findings and Pneumonia. The DarkNet model has been exploited as the classification engine, consisting of 17 convolutional layers and a different filtering in each layer. The proposed model has a prediction accuracy of 98.08% and 87.02% for binary and multi-nominal classification, respectively.

Various studies have demonstrated the usefulness of CXR exams in detecting Covid-19. Hall et al. [53] analyzed 135 CXR images confirmed as Covid-19 and 320 images of viral and bacterial pneumonia. A Resnet50 DNN was trained on 102 Covid-19 and 102 pneumonia cases by means of the ten-fold cross-validation technique. The experimental results showed an overall accuracy of 89.2% with a Covid-19 true positive rate of 0.804 and an area under the curve (AUC) of 0.95. As a matter of fact, the dataset used by Hall et al. [53] is quite small, and it is necessary to investigate the proposed model on a larger amount of data to see if it still achieves such a good performance.

Narin et al. [37] developed a system for the detection of coronavirus patients from CXR images. Three different CNN-based models have been exploited, i.e., ResNet50, InceptionV3 and Inception-ResNetV2. The results show that the ResNet50 model achieves the best prediction performance with 98.0% as the final accuracy. Although the approach obtains a good classification performance, it has been studied on a considerably small dataset. It is our belief that performance may substantially change on larger datasets like the ones used in our evaluation.

Apostolopoulos et al. [38] evaluated their solution to automatic detection of Covid-19, making use of a dataset of CXR images from patients with common bacterial pneumonia, confirmed Covid-19, and normal incidents. The datasets consists of 1427 CXR images including 224 images with confirmed Covid-19 cases, 700 images with common bacterial pneumonia, and 504 images of normal situations. The experimental outcomes demonstrate that deep neural networks can be exploited to extract important biomarkers related Covid-19. Nevertheless, like some other existing studies et al. [37], again the approach has been studied by means of a small amount of data. It is our assumption that such a good performance might considerably change with larger datasets.

COVID-Net [54] is a deep convolutional neural network design tailored for the detection of Covid-19 cases from CXR images. COVID-Net achieves an accuracy of 93.3%, with 98.9% positive predictive values that is related to the detection of false positives.

A deep learning model has been proposed [55] to detect Covid-19 and differentiate it from common acquired pneumonia and other lung diseases. The analyzed dataset consists of 4356 chest CT exams collected from 3322 patients. The per-exam sensitivity and specificity for detecting COVID-19 in the independent test set was 114 of 127 (90.0%) and 294 of 307 (96.0%), respectively, with an area under the receiver operating characteristic curve (AUC) of 0.96 (p-value < 0.001). The per-exam sensitivity and specificity for detecting community acquired pneumonia in the independent test set was 87% (152 of 175) and 92% (239 of 259), respectively.

Abbas et al. [36] introduced Decompose, Transfer, and Compose (DeTraC), a deep neural network for automated recognition of Covid-19 from CXR images. An accuracy of 95% was achieved in the detection of Covid-19 CXR images from normal, and severe acute respiratory syndrome cases. COVID-CAPS [56] is a capsule Network-based Framework for Identification of Covid-19 cases from CXR Images. The approach yielded a good accuracy when working with small datasets.

Table 10
A summary of the related studies on CXR datasets.

Study	Number of images			Network	Acc. (%)
	Covid-19	Normal	Pneumonia		
Ghoshal et al. [35]	68	1583	2786	ResNet	89.82
Abbas et al. [36]	105	80	11	DeTraC based on ResNet-18	95.12
Nari et al. [37]	50	50	–	ResNet-50	98.00
Apostolopous et al. [38]	224	504	700	VGG19	93.48
Luz et al. [39]	183	–	–	EfficientNet-B3	93.90
Zhang et al. [40]	100	1431	1531	ResNet18	96.00
Hemdan et al. [41]	25	25	–	VGG19, DenseNet121	90.00

Table 11
A summary of the related studies on LCT datasets.

Study	Number of images			Network	Acc. (%)
	Covid-19	Normal	Pneumonia		
Rahimzadeh et al. [42]	465	7878	–	ResNet50V2	98.49
Anwar et al. [43]	98	–	105	EfficientNet	89.70
Gunraj et al. [44]	4346	9450	7395	COVID-Net-CT	99.10
Mobiny et al. [45]	47	58	–	DECAPS	87.60
He et al. [30]	25,442	14,471	28,160	ResNet3D34	95.90
Ardakanu et al. [46]	510	510	–	ResNet-101	99.63
Bai et al. [47]	5030	–	9152	EfficientNet-B4	96.00

Concerning LCT datasets, we review the most relevant studies to our work in Table 11. Overall, most of the studies use a small dataset for their evaluation. Only Gunraj et al. [44] made use of a considerably large dataset, however the obtained performance is much lower than our proposed approach, i.e., 95.90% compared to 99.66%. In this respect, we conclude that by combining different deep neural networks with transfer learning strategies, we can obtain a high prediction accuracy even on a very large dataset.

To the best of our knowledge, compared to different existing studies [36,38], our work is the first one that deals with big datasets. In particular, for the CXR datasets, there are 15,000 and 17,905 images in $\mathbf{D}_1$ and $\mathbf{D}_2$, respectively [57]. While for the LCT datasets, especially with $\mathbf{D}_4$, there are more than 400K of images. This suggests that with the proposed model we can work with a huge number of images and still obtain a good timing efficiency. More importantly, the application of the two deep neural network families, i.e., EfficientNet and MixNet, allows us to build an expert system being capable of working with images coming from different sources.

6. Conclusions

We presented a workable solution for the detection of Covid-19 from chest X-ray and lung computed tomography images. We designed and implemented the models based on two building blocks: EfficientNet and MixNet as the prediction engine and effective transfer learning algorithms. The proposed models have been studied by means of four datasets which have been widely used in various papers. The experimental results show that our proposed approach obtains a better prediction performance compared to some relevant previous studies. To the best of our knowledge, our work is the first one that deals with images coming from different sources. As future work, we are going to refine and evaluate the approach by taking into consideration more datasets and tuning other deep neural network configurations.

CRediT authorship contribution statement

Linh T. Duong: Conceptualization, Software, Writing – original draft. **Phuong T. Nguyen:** Conceptualization, Methodology, Writing – original draft, Writing – review & editing. **Ludovico Iovino:** Validation, Visualization, Writing – review & editing. **Michele Flammini:** Writing – review & editing, Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The authors do not have permission to share data.

References

[1] J.M. Connors, J.H. Levy, COVID-19 and its implications for thrombosis and anticoagulation, Blood J. Am. Soc. Hematol. 135 (23) (2020) 2033–2040.
[2] Q. Sun, H. Qiu, M. Huang, Y. Yang, Lower mortality of COVID-19 by early recognition and intervention: experience from Jiangsu Province, Ann. Intensive Care 10 (1) (2020) 1–4.
[3] L. Iovino, P.T. Nguyen, A.D. Salle, F. Gallo, M. Flammini, Unavailable transit feed specification: Making it available with recurrent neural networks, IEEE Trans. Intell. Transp. Syst. 22 (4) (2021) 2111–2122, http://dx.doi.org/10.1109/TITS.2021.3053373.
[4] L.T. Duong, N.H. Le, T.B. Tran, V.M. Ngo, P.T. Nguyen, Detection of tuberculosis from chest X-ray images: Boosting the performance with vision transformer and transfer learning, Expert Syst. Appl. 184 (2021) 115519, http://dx.doi.org/10.1016/j.eswa.2021.115519, URL https://www.sciencedirect.com/science/article/pii/S0957417421009295.
[5] I. Portugal, P. Alencar, D. Cowan, The use of machine learning algorithms in recommender systems: A systematic review, Expert Syst. Appl. 97 (2018) 205–227, http://dx.doi.org/10.1016/j.eswa.2017.12.020, URL http://www.sciencedirect.com/science/article/pii/S0957417417308333.
[6] X. Mei, H.-C. Lee, K.-y. Diao, M. Huang, B. Lin, C. Liu, Z. Xie, Y. Ma, P. Robson, M. Chung, A. Bernheim, V. Mani, C. Calcagno, K. Li, S. Li, H. Shan, J. Lv, T. Zhao, J. Xia, Y. Yang, Artificial intelligence–enabled rapid diagnosis of patients with COVID-19, Nat. Med. (2020) 1–5, http://dx.doi.org/10.1038/s41591-020-0931-3.
[7] M. Tan, Q. Le, EfficientNet: Rethinking model scaling for convolutional neural networks, in: K. Chaudhuri, R. Salakhutdinov (Eds.), Proceedings of the 36th International Conference on Machine Learning, in: Proceedings of Machine Learning Research, vol. 97, PMLR, Long Beach, California, USA, 2019, pp. 6105–6114, URL http://proceedings.mlr.press/v97/tan19a.html.
[8] M. Tan, Q.V. Le, MixConv: Mixed depthwise convolutional kernels, 2019, CoRR abs/1907.09595. arXiv:1907.09595. URL http://arxiv.org/abs/1907.09595.
[9] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A.C. Berg, L. Fei-Fei, ImageNet large scale visual recognition challenge, Int. J. Comput. Vis. 115 (3) (2015) 211–252, http://dx.doi.org/10.1007/s11263-015-0816-y.
[10] C. Xie, M. Tan, B. Gong, J. Wang, A. Yuille, Q.V. Le, Adversarial examples improve image recognition, 2019, arXiv e-prints arXiv:1911.09665.

[11] Q. Xie, E. Hovy, M.-T. Luong, Q.V. Le, Self-training with Noisy Student improves ImageNet classification, 2019, cite arxiv:1911.04252. URL http://arxiv.org/abs/1911.04252.

[12] F. Chollet, Xception: Deep learning with depthwise separable convolutions, 2016, cite arxiv:1610.02357. URL http://arxiv.org/abs/1610.02357.

[13] A.G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, H. Adam, MobileNets: Efficient convolutional neural networks for mobile vision applications, 2017, cite arxiv:1704.04861. URL http://arxiv.org/abs/1704.04861.

[14] M. Tan, B. Chen, R. Pang, V. Vasudevan, Q.V. Le, MnasNet: Platform-aware neural architecture search for mobile, 2018, CoRR abs/1807.11626. arXiv:1807.11626. URL http://arxiv.org/abs/1807.11626.

[15] H. Cai, L. Zhu, S. Han, ProxylessNAS: Direct neural architecture search on target task and hardware, in: International Conference on Learning Representations, 2019, URL https://openreview.net/forum?id=HylVB3AqYm.

[16] M. Sandler, A.G. Howard, M. Zhu, A. Zhmoginov, L. Chen, Inverted residuals and linear bottlenecks: Mobile networks for classification, detection and segmentation, 2018, CoRR abs/1801.04381. arXiv:1801.04381. URL http://arxiv.org/abs/1801.04381.

[17] A. Kamilaris, F.X. Prenafeta-Boldú, Deep learning in agriculture: A survey, Comput. Electron. Agric. 147 (2018) 70–90, http://dx.doi.org/10.1016/j.compag.2018.02.016.

[18] K. Weiss, T. Khoshgoftaar, D. Wang, A survey of transfer learning, J. Big Data 3 (2016) http://dx.doi.org/10.1186/s40537-016-0043-6.

[19] L. Torrey, T. Walker, J. Shavlik, R. Maclin, Using advice to transfer knowledge acquired in one reinforcement learning task to another, in: Proceedings of the 16th European Conference on Machine Learning, ECML '05, Springer-Verlag, Berlin, Heidelberg, 2005, pp. 412–424, http://dx.doi.org/10.1007/11564096_40.

[20] Z. Huang, Z. Pan, B. Lei, Transfer learning with deep convolutional neural network for SAR target classification with limited labeled data, Remote Sens. 9 (9) (2017) http://dx.doi.org/10.3390/rs9090907.

[21] L.T. Duong, P.T. Nguyen, C. Di Sipio, D. Di Ruscio, Automated fruit recognition using EfficientNet and MixNet, Comput. Electron. Agric. 171 (2020) 105326, http://dx.doi.org/10.1016/j.compag.2020.105326, URL http://www.sciencedirect.com/science/article/pii/S0168169919319787.

[22] Z.Q.L. Linda Wang, A. Wong, COVID-Net: A tailored deep convolutional neural network design for detection of COVID-19 cases from chest radiography images, 2020, arXiv:2003.09871.

[23] J.P. Cohen, P. Morrison, L. Dao, COVID-19 image data collection, 2020, arXiv:2003.11597. URL https://github.com/ieee8023/covid-chestxray-dataset.

[24] E. Soares, P. Angelov, S. Biaso, M. Higa Froes, D. Kanda Abe, SARS-CoV-2 CT-scan dataset: A large dataset of real patients CT scans for SARS-CoV-2 identification, MedRxiv (2020) http://dx.doi.org/10.1101/2020.04.24.20078584, arXiv:https://www.medrxiv.org/content/early/2020/05/14/2020.04.24.20078584.full.pdf. URL https://www.medrxiv.org/content/early/2020/05/14/2020.04.24.20078584.

[25] K. Zhang, X. Liu, J. Shen, Z. Li, Y. Sang, X. Wu, Y. Zha, W. Liang, C. Wang, K. Wang, L. Ye, M. Gao, Z. Zhou, L. Li, J. Wang, Z. Yang, H. Cai, J. Xu, L. Yang, W. Cai, W. Xu, S. Wu, W. Zhang, S. Jiang, L. Zheng, X. Zhang, L. Wang, L. Lu, J. Li, H. Yin, W. Wang, O. Li, C. Zhang, L. Liang, T. Wu, R. Deng, K. Wei, Y. Zhou, T. Chen, J. Lau, M. Fok, J. He, T. Lin, W. Li, G. Wang, Clinically applicable AI system for accurate diagnosis, quantitative measurements, and prognosis of COVID-19 pneumonia using computed tomography, Cell 181 (6) (2020) http://dx.doi.org/10.1016/j.cell.2020.04.045.

[26] G. Marques, D. Agarwal, I. de la Torre Díez, Automated medical diagnosis of COVID-19 through EfficientNet convolutional neural network, Appl. Soft Comput. 96 (2020) 106691, http://dx.doi.org/10.1016/j.asoc.2020.106691, URL http://www.sciencedirect.com/science/article/pii/S1568494620306293.

[27] A. Iqbal, M. Usman, Z. Ahmed, An efficient deep learning-based framework for tuberculosis detection using chest X-ray images, Tuberculosis 136 (2022) 102234, http://dx.doi.org/10.1016/j.tube.2022.102234, URL https://www.sciencedirect.com/science/article/pii/S1472979222000713.

[28] A. Saygılı, A new approach for computer-aided detection of coronavirus (COVID-19) from CT and X-ray images using machine learning methods, Appl. Soft Comput. 105 (2021) 107323, http://dx.doi.org/10.1016/j.asoc.2021.107323, URL https://www.sciencedirect.com/science/article/pii/S1568494621002465.

[29] T. Ozturk, M. Talo, E.A. Yildirim, U.B. Baloglu, O. Yildirim, U.R. Acharya, Automated detection of COVID-19 cases using deep neural networks with X-ray images, Comput. Biol. Med. 121 (2020) 103792, http://dx.doi.org/10.1016/j.compbiomed.2020.103792, URL http://www.sciencedirect.com/science/article/pii/S0010482520301621.

[30] X. He, S. Wang, S. Shi, X. Chu, J. Tang, X. Liu, C. Yan, J. Zhang, G. Ding, Benchmarking deep learning models and automated model design for COVID-19 detection with chest CT scans, MedRxiv (2020) http://dx.doi.org/10.1101/2020.06.08.20125963.

[31] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: 2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR, 2016, pp. 770–778, http://dx.doi.org/10.1109/CVPR.2016.90.

[32] M. Elpeltagy, H. Sallam, Automatic prediction of COVID-19 from chest images using modified ResNet50, Multimedia Tools Appl. 80 (17) (2021) 26451–26463, http://dx.doi.org/10.1007/s11042-021-10783-6.

[33] D. Yang, C. Martinez, L. Visuña, H. Khandhar, C. Bhatt, J. Carretero, Detection and analysis of COVID-19 in medical images using deep learning techniques, 11 (1) 19638. http://dx.doi.org/10.1038/s41598-021-99015-3.

[34] F. Wilcoxon, Individual comparisons by ranking methods, in: S. Kotz, N.L. Johnson (Eds.), Breakthroughs in Statistics: Methodology and Distribution, Springer New York, New York, NY, 1992, pp. 196–202, http://dx.doi.org/10.1007/978-1-4612-4380-9_16.

[35] B. Ghoshal, A. Tucker, Estimating uncertainty and interpretability in deep learning for coronavirus (COVID-19) detection, 2020, arXiv abs/2003.10769.

[36] A. Abbas, M.M. Abdelsamea, M.M. Gaber, Classification of COVID-19 in chest X-ray images using DeTraC deep convolutional neural network, 2020, arXiv preprint arXiv:2003.13815.

[37] A. Narin, C. Kaya, Z. Pamuk, Automatic detection of coronavirus disease (COVID-19) using X-ray images and deep convolutional neural networks, 2020, arXiv:2003.10849.

[38] I.D. Apostolopoulos, T.A. Mpesiana, Covid-19: automatic detection from x-ray images utilizing transfer learning with convolutional neural networks, Phys. Eng. Sci. Med. (2020) 1.

[39] E. Luz, P.L. Silva, R. Silva, G. Moreira, Towards an efficient deep learning model for covid-19 patterns detection in x-ray images, 2020, arXiv preprint arXiv:2004.05717.

[40] J. Zhang, Y. Xie, Y. Li, C. Shen, Y. Xia, Covid-19 screening on chest x-ray images using deep learning based anomaly detection, 2020, arXiv preprint arXiv:2003.12338.

[41] E.E.-D. Hemdan, M.A. Shouman, M.E. Karar, Covidx-net: A framework of deep learning classifiers to diagnose covid-19 in x-ray images, 2020, arXiv preprint arXiv:2003.11055.

[42] M. Rahimzadeh, A. Attar, S.M. Sakhaei, A fully automated deep learning-based network for detecting COVID-19 from a new and large lung CT scan dataset, MedRxiv (2020) http://dx.doi.org/10.1101/2020.06.08.20121541, arXiv:https://www.medrxiv.org/content/early/2020/09/01/2020.06.08.20121541.full.pdf.

[43] T. Anwar, S. Zakir, Deep learning based diagnosis of COVID-19 using chest CT-scan images, 2020, http://dx.doi.org/10.36227/techrxiv.12328061.

[44] H. Gunraj, L. Wang, A. Wong, COVIDNet-CT: A tailored deep convolutional neural network design for detection of COVID-19 cases from chest CT images, 2020, arXiv:2009.05383.

[45] A. Mobiny, P.A. Cicalese, S. Zare, P. Yuan, M. Abavisani, C.C. Wu, J. Ahuja, P.M. de Groot, H.V. Nguyen, Radiologist-level COVID-19 detection using CT scans with detail-oriented capsule networks, 2020, arXiv:2004.07407.

[46] A.A. Ardakani, A.R. Kanafi, U.R. Acharya, N. Khadem, A. Mohammadi, Application of deep learning technique to manage COVID-19 in routine clinical practice using CT images: Results of 10 convolutional neural networks, Comput. Biol. Med. 121 (2020) 103795, http://dx.doi.org/10.1016/j.compbiomed.2020.103795, URL http://www.sciencedirect.com/science/article/pii/S0010482520301645.

[47] H.X. Bai, R. Wang, Z. Xiong, B. Hsieh, K. Chang, K. Halsey, T.M.L. Tran, J.W. Choi, D.-C. Wang, L.-B. Shi, J. Mei, X.-L. Jiang, I. Pan, Q.-H. Zeng, P.-F. Hu, Y.-H. Li, F.-X. Fu, R.Y. Huang, R. Sebro, Q.-Z. Yu, M.K. Atalay, W.-H. Liao, Artificial intelligence augmentation of radiologist performance in distinguishing COVID-19 from pneumonia of other origin at chest CT, Radiology 296 (3) (2020) E156–E165, http://dx.doi.org/10.1148/radiol.2020201491, PMID: 32339081. arXiv:https://doi.org/10.1148/radiol.2020201491.

[48] H. Zhang, K.M. Saravanan, Y. Yang, M.T. Hossain, J. Li, X. Ren, Y. Pan, Y. Wei, Deep learning based drug screening for novel coronavirus 2019-nCov, Interdiscip. Sci. Comput. Life Sci. (2020) 1.

[49] B.R. Beck, B. Shin, Y. Choi, S. Park, K. Kang, Predicting commercially available antiviral drugs that may act on the novel coronavirus (SARS-CoV-2) through a drug-target interaction deep learning model, Comput. Struct. Biotechnol. J. (2020).

[50] L. Yan, H.-T. Zhang, Y. Xiao, M. Wang, C. Sun, J. Liang, S. Li, M. Zhang, Y. Guo, Y. Xiao, et al., Prediction of criticality in patients with severe Covid-19 infection using three clinical features: a machine learning-based prognostic model with clinical data in Wuhan, MedRxiv (2020).

[51] X. Jiang, M. Coffee, A. Bari, J. Wang, X. Jiang, J. Huang, J. Shi, J. Dai, J. Cai, T. Zhang, et al., Towards an artificial intelligence framework for data-driven prediction of coronavirus clinical severity, CMC: Comput. Mater. Contin. 63 (2020) 537–551.

[52] T. Ozturk, M. Talo, E.A. Yildirim, U.B. Baloglu, O. Yildirim, U.R. Acharya, Automated detection of COVID-19 cases using deep neural networks with X-ray images, Comput. Biol. Med. (2020) 103792.

[53] L.O. Hall, R. Paul, D.B. Goldgof, G.M. Goldgof, Finding covid-19 from chest x-rays using deep learning on a small dataset, 2020, arXiv preprint arXiv:2004.02060.

[54] L. Wang, A. Wong, COVID-net: A tailored deep convolutional neural network design for detection of COVID-19 cases from chest X-Ray images, 2020, arXiv preprint arXiv:2003.09871.

[55] L. Li, L. Qin, Z. Xu, Y. Yin, X. Wang, B. Kong, J. Bai, Y. Lu, Z. Fang, Q. Song, et al., Artificial intelligence distinguishes COVID-19 from community acquired pneumonia on chest CT, Radiology (2020) 200905.

[56] P. Afshar, S. Heidarian, F. Naderkhani, A. Oikonomou, K.N. Plataniotis, A. Mohammadi, Covid-caps: A capsule network-based framework for identification of covid-19 cases from x-ray images, 2020, arXiv preprint arXiv:2004.02696.

[57] L.T. Duong, P.T. Nguyen, L. Iovino, M. Flammini, Deep learning for automated recognition of Covid-19 from chest X-ray images, MedRxiv (2020) http://dx.doi.org/10.1101/2020.08.13.20173997, arXiv:https://www.medrxiv.org/content/early/2020/08/14/2020.08.13.20173997.full.pdf. URL https://www.medrxiv.org/content/early/2020/08/14/2020.08.13.20173997.