



Towards an effective and efficient deep learning model for COVID-19 patterns detection in X-ray images

Eduardo Luz¹ · Pedro Silva¹ · Rodrigo Silva¹ · Ludmila Silva² · João Guimarães³ · Gustavo Miozzo³ · Gladston Moreira¹ · David Menotti⁴

Received: 3 July 2020 / Accepted: 14 April 2021 / Published online: 20 April 2021
© Sociedade Brasileira de Engenharia Biomedica 2021

Abstract

Purpose Confronting the pandemic of COVID-19 is nowadays one of the most prominent challenges of the human species. A key factor in slowing down the virus propagation is the rapid diagnosis and isolation of infected patients. The standard method for COVID-19 identification, the Reverse transcription polymerase chain reaction method, is time-consuming and in short supply due to the pandemic. Thus, researchers have been looking for alternative screening methods, and deep learning applied to chest X-rays of patients has been showing promising results. Despite their success, the computational cost of these methods remains high, which imposes difficulties to their accessibility and availability. Thus, the main goal of this work is to propose an accurate yet efficient method in terms of memory and processing time for the problem of COVID-19 screening in chest X-rays.

Methods To achieve the defined objective, we propose a new family of models based on the EfficientNet family of deep artificial neural networks which are known for their high accuracy and low footprints. We also exploit the underlying taxonomy of the problem with a hierarchical classifier. A dataset of 13,569 X-ray images divided into healthy, non-COVID-19 pneumonia, and COVID-19 patients is used to train the proposed approaches and other 5 competing architectures. We also propose a cross-dataset evaluation with a second dataset to evaluate the method generalization power.

Results The results show that the proposed approach was able to produce a high-quality model, with an overall accuracy of 93.9%, COVID-19 sensitivity of 96.8%, and positive prediction of 100% while having from 5 to 30 times fewer parameters than the other tested architectures. Larger and more heterogeneous databases are still needed for validation before claiming that deep learning can assist physicians in the task of detecting COVID-19 in X-ray images, since the cross-dataset evaluation shows that even state-of-the-art models suffer from a lack of generalization power.

Conclusions We believe the reported figures represent state-of-the-art results, both in terms of efficiency and effectiveness, for the COVIDx database, a database of 13,800 X-ray images, 183 of which are from patients affected by COVID-19. The current proposal is a promising candidate for embedding in medical equipment or even physicians' mobile phones.

Keywords COVID-19 · Deep learning · EfficientNet · Pneumonia · Chest (X-ray) radiography

✉ Eduardo Luz
eduluz@ufop.edu.br

¹ Department of Computer Science, Federal University of Ouro Preto, Ouro Preto, MG, Brazil

² Hospital e Maternidade Otaviano Neves, Belo Horizonte, MG, Brazil

³ Hospital Cruz Vermelha, Curitiba, PR, Brazil

⁴ Department of Informatics, Federal University of Paraná, Centro Politécnico, Curitiba, PR, Brazil

Introduction

The COVID-19 is an infection caused by the SARS-CoV-2 virus and may manifest itself as a flu-like illness potentially progressing to an acute respiratory distress syndrome. The disease severity resulted in a global public health effort to contain person-to-person viral spread by early detection (Davarpanah et al. 2020).

The Reverse-Transcriptase Polymerase Chain Reaction (RT-PCR) is currently the gold standard for a definitive diagnosis of COVID-19. However, false negatives have been reported (due to insufficient cellular content in the sample or inadequate detection and extraction techniques) in the

presence of positive radiological findings (Araujo-Filho et al. 2020). Therefore, effective exclusion of the COVID-19 infection requires multiple negative tests, possibly exacerbating test kit shortage (American College of Radiology 2020).

As COVID-19 spreads in the world, there is growing interest in the role and suitability of chest X-Rays (CXR) for screening, diagnosis, and management of patients with suspected or known COVID-19 infection (Huang et al. 2020; Ai et al. 2020; Ng et al. 2020). Besides, there have been a growing number of publications describing the CXR appearance in patients with COVID-19 (American College of Radiology 2020).

The accuracy of CXR diagnosis of COVID-19 infection strongly relies on radiological expertise due to the complex morphological patterns of lung involvement which can change in extent and appearance over time. The limited number of sub-specialty trained thoracic radiologists hampers reliable interpretation of complex chest examinations, especially in developing countries, where general radiologists and occasionally clinicians interpret chest imaging (Davaranah et al. 2020).

Deep learning is a subset of machine learning in artificial intelligence (AI) concerned with algorithms inspired by the structure and function of the brain called artificial neural networks. Since deep learning techniques, in particular convolutional neural networks (CNNs), have been beating humans in various computer vision tasks (LeCun et al. 2015; Touvron et al. 2020; Rajpurkar et al. 2017), it becomes a natural candidate for the analysis of chest radiography images.

Deep learning has already been explored for the detection and classification of pneumonia and other diseases on radiography (Rajpurkar et al. 2017; Wang et al. 2017; Jaiswal et al. 2019). In this context, this work aims to investigate deep learning models that are capable of finding patterns of X-ray images of the chest, even if the patterns are imperceptible to the human eye, and to advance on a fundamental issue: computational cost.

Finding a model of low computational cost is important because it allows the exploitation of input images of much higher resolutions without making the processing time prohibitive. Besides, it becomes easier and cheaper to embed these models in equipment with more restrictive settings such as smartphones. We believe that a mobile application that integrates deep learning models for the task of recognizing patterns in X-rays must be easily accessible and readily available to the medical staff. For such aim, the models must have a low footprint and low latency, that is, the models must require little memory and perform inference quickly to allow use on embedded devices and large scale, enabling integration with smartphones and medical equipment.

To find such cost-efficient models, in this work, the family of EfficientNets, recently proposed in Tan and Le (2019), is investigated. These models have shown high performance in

the classic ImageNet dataset (Deng et al. 2009) while presenting only a small fraction of the cost of other popular architectures such as the ResNets and VGGs. We also exploit the natural taxonomy of the problem and investigate the use of hierarchical classification. In this case, low computational cost is even more critical since multiple models have to be built.

The results show that it is indeed possible to build much smaller models without compromising accuracy. Thus, embedding the proposed neural network model in a mobile device to make fast inferences becomes more feasible. Despite its low computational cost, the proposed model achieves high accuracy (93.9%) and detects infection caused by COVID-19 on chest X-rays with a sensitivity of 96.8% and positivity prediction of 100% (without false positives). Regarding the hierarchical model, in addition to consuming more computational resources than the flat classification, it showed to be less effective for minority classes, which is the case for the COVID-19 class in this work. However, we believe that the hierarchical method is very suitable for the application. It may suffer less from bias in the evaluation protocols (Maguolo and Nanni 2020), since images from different sources (datasets) are mixed to build the superclasses according to the taxonomy of the problem (Silla and Freitas 2011).

The development of this work may allow the future construction of an application for use by the medical team, through a camera on a regular cell phone. The source codes as the pre-trained models are available in <https://github.com/fopcsilab/EfficientNet-C19>.

In summary, the contributions of this work are:

- An effective yet compact new family of models, based on the state-of-the-art EfficientNet architecture, for COVID-19 screening in chest X-ray images. In particular, we highlight the model which we call B3-X, due to its compact size and high accuracy, which can favor the use of the model in systems with low computational power, such as cell phones and medical equipment.
- A hierarchical approach study. We believe that this study allows the evaluation of methods with less interference regarding the problems of mixed datasets.
- A cross-dataset study. To the best of our knowledge, this work is the first to conduct a cross-dataset analysis of the problem. We believe a cross-dataset study is of paramount importance to assess the generalization power of the models.

The remainder of this work consists of six sections. Section 2 presents a review of related works. Section 3 defines the problem tackled in this paper. The methodology and the dataset are described in Section 4. In Section 5, the results of a comprehensive set of computational experiments are presented. In Section 6, propositions for future research in the area are addressed. Finally, conclusions are pointed out in Section 7.

Related works

Several works have been proposed for the task of COVID-19 classification in X-ray images to date. In this section, we list some seminal works that, in our opinion, have a robust and reproducible methodology.

Addressing the COVID-19, in Hemdan et al. (2020), a comparison among seven different well-known deep learning neural network architectures was presented. In the experiments, they use a small dataset with only 50 images in which 25 samples are from healthy patients and 25 from COVID-19-positive patients. The models were pre-trained with the ImageNet dataset (Deng et al. 2009), which is a generic image dataset with over 14 million images of all sorts, and only the classifier is trained with the radiography. In their experiments, the VGG19 (Simonyan and Zisserman 2014) and the DenseNET201 (Huang et al. 2019) were the best performing architectures. Following a similar approach, in Al-Bawi et al. (2020), an extension of the VGG architecture was proposed, with the addition of the convolutional COVID block (CCBlock). The model was evaluated using a mixed dataset, composed of images from two public datasets, totaling 1887 images. Among the images, 300 of them belong to the COVID-19 class, 864 to the pneumonia class, and 654 images are of the normal class. The authors reported accuracy of 95.34% for the three-class classification problem.

In Wang et al. (2020), a new architecture of CNN, called COVID-net, is created to classify CXR images into normal, pneumonia, and COVID-19. Differently from the previous work, they use a much larger dataset consisting of 13,800 CXR images across 13,645 patient cases from which 182 images belong to COVID-19 patients. The authors report an accuracy of 92.4% overall and sensitivity of 80% for COVID-19.

In Farooq and Hafeez (2020), the ResNet50 (Szegedy et al. 2017) is fine-tuned for the problem of classifying CXRs into normal, COVID-19, bacterial-pneumonia, and viral pneumonia. The authors report better results when compared with the COVID-net, 96.23% accuracy overall, and 100% sensitivity for COVID-19. Nevertheless, it is important to highlight that the problem in Farooq and Hafeez (2020) has an extra class and that its dataset is a subset of the dataset used in Wang et al. (2020). In Farooq and Hafeez (2020), the dataset consists of 68 COVID-19 radiographs from 45 COVID-19 patients, 1203 healthy patients, 931 patients with bacterial pneumonia, and 660 patients with nonCOVID-19 viral pneumonia.

In Pereira et al. (2020), the authors also performed a hierarchical analysis for the task of detecting COVID-19 patterns on CXR images. A dataset was built, from other public datasets, containing 1144 X-ray images, of which only 90 were related to COVID-19 and the remaining belonging to six other classes: five types of pneumonia and one normal (healthy) type. Several techniques were used to extract

features from the images, including one based on deep convolutional networks (Inception-V3 (Szegedy et al. 2016)). For classification, the authors explored classifiers such as SVM, Random Forest, KNNs, MLPs, and Decision Trees. A F1-Score of 0.89 for the COVID-19 class is reported. In spite of having a strong relation to the present work, we emphasize that a direct comparison is not possible, due to the different nature of the datasets employed on both works.

In (Khan et al. 2020), the authors propose a convolutional neural network-based model to automate the detection of COVID-19 infection from chest X-ray images, named CoroNet. The proposed model uses the Xception CNN architecture (Chollet 2017), pre-trained on ImageNet dataset (Deng et al. 2009). CoroNet was trained and tested on the prepared dataset from two different publically available image databases (available at <https://github.com/ieee8023/covid-chestxray-dataset> and <https://www.kaggle.com/paultimothymooney/chest-xray-pneumonia>). The CoroNet model achieved an accuracy of 89.6%, with precision and recall rate for COVID-19 cases of 93 and 98.2% for 4-class cases (COVID vs Pneumonia bacterial vs Pneumonia viral vs Normal) with a fourfold cross-validation scheme. Also, the authors evaluate their model on a second dataset, though this second dataset apparently contains the same COVID-19 images used during training.

Problem setting

The problem addressed by the proposed approach can be defined as follows: given an chest X-ray, determine if it belongs to a healthy patient, a patient with COVID-19, or a patient with other forms of pneumonia. Figure 1 shows typical chest X-ray samples in COVIDx dataset (Wang et al. 2020). As can be seen, the model should not make assumptions regarding the view in which the X-ray was taken.

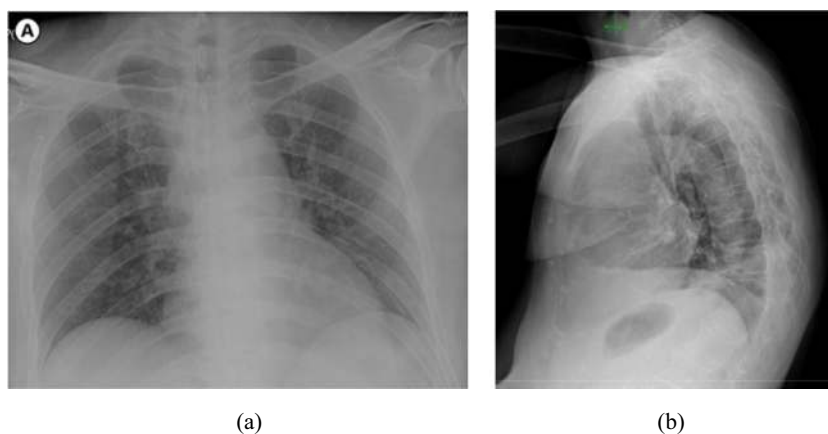
Thus, given an image similar to these ones, a model must output one of the following three possible labels:

- Normal—for healthy patients
- COVID-19—for patients with COVID-19
- Pneumonia—for patients with non-COVID-19 pneumonia

Following the rationale in Wang et al. (2020), choosing these three possible predictions can help clinicians in deciding who should be prioritized for PCR testing for COVID-19 case confirmation. Moreover, it might also help in treatment selection since COVID-19 and non-COVID-19 infections require different treatment plans.

We analyze the problem from two perspectives: (1) the traditional flat classification, in which we disregard the relationship between the classes, and (2) the hierarchical

Fig. 1 Radiograph example of images from COVID-19 image data collection (Cohen et al. 2020). **(a)** X-ray of a 54-year-old male, infected with COVID-19 (Cohen et al. 2020). **(b)** X-ray of a 70-year-old female, infected with COVID-19 (Cohen et al. 2020)



classification approach, in which we assume the classes to be predicted are naturally organized into a taxonomy.

Methodology

In this section, we present the methodology for COVID-19 detection by means of a chest X-ray image. We detail the main datasets and briefly describe the COVID-Net (Wang et al. 2020), our baseline method. Also, we describe the employed deep learning techniques as well as the learning methodology and evaluation.

Datasets

RSNA pneumonia detection challenge dataset

The RSNA Pneumonia Detection Challenge (Radiological Society of North America 2020) is a competition that aims to locate lung opacities on chest radiographs. Pneumonia is associated with opacity in the lung, and some conditions such as pulmonary edema, bleeding, volume loss, and lung cancer can also lead to opacity in lung radiography. Finding patterns associated with pneumonia is a hard task. In that sense, the Radiological Society of North America (RSNA) has promoted the challenge, providing a rich dataset. Although The RSNA challenge is a segmentation challenge, here we are using the dataset for a classification problem. The dataset offers images for two classes: Normal and Pneumonia (non-normal). We are using a total of 16,680 images of this dataset, of which 8066 are from normal class and 8614 from the pneumonia class.

COVID-19 image data collection

The “COVID-19 Image Data Collection” (Cohen et al. 2020) is a collection of anonymized COVID-19 images, acquired from websites of medical and scientific associations (Giovagnoni 2020; Società Italiana di Radiologia Medica e

Interventistica 2020) and research papers. The dataset was created by researchers from the University of Montreal with the help of the international research community to assure that it will be continuously updated. Nowadays, the dataset includes more than 183 X-ray images of patients who were affected by COVID-19 and other diseases, such as MERS, SARS, and ARDS. The dataset is public and also includes CT scan images. According to the authors, the dataset can be used to assess the advancement of COVID-19 in infected individuals and also allow the identification of patterns related to COVID-19 helping in differentiating it from other types of pneumonia. Besides, CXR images can be used as an initial screening for the COVID-19 diagnostic processes. So far, most of the images are from male individuals (approx. 60/40% of males and females, respectively), and the age group that concentrates most cases is from 50 to 80 years old. The dataset has four views: the posteroanterior (PA), anteroposterior (AP), supine (AP supine), and lateral (L). There are images from the same subject with different views and different acquisition sessions.

COVIDx dataset

In Wang et al. (2020), a new dataset is proposed by merging two other public datasets: “RSNA Pneumonia Detection Challenge dataset” and “COVID-19 Image Data Collection.” The new dataset, called COVIDx, is designed for a classification problem and contemplates three classes: normal, pneumonia, and COVID-19. Most instances of the normal and pneumonia classes come from the “RSNA Pneumonia Detection Challenge dataset,” and all instances of the COVID-19 class come from the “COVID-19 Image Data Collection.” The dataset has a total of 13,800 images from 13,645 individuals and is split into two partitions, one for training purposes and one for testing (model evaluation). The distribution of images between the partitions is shown in Table 1, and the source code to reproduce the dataset is publicly available (<https://github.com/lindawang/COVID->

Table 1 COVIDx Images distribution among classes and partitions. The dataset is proposed in (Wang et al. 2020)

Type	Normal	Pneumonia	COVID-19	Total
Train	7966	5421	152	13,569
Test	100	100	31	231

Net). The resolution of images ranges from 156×157 to 4032×3024 pixels.

HCV-UFPR COVID-19 dataset

Brazil is one of the countries most affected by covid-19, with over 6 million confirmed cases, to date. The *Hospital da Cruz Vermelha from Curitiba*, located in the state of Paraná, southern Brazil, received and documented some of those cases (See Fig. 2). The data collection consists of 281 X-ray images of people infected with COVID-19 and 232 of people who obtained negative results that are not infected. All images have 3 eight-bit color channels (RGB) and image resolution ranges from 2974×2612 to 4248×3480 pixels. The images are labeled in two classes, COVID-19 and non-COVID, and there are no annotations regarding the image angle view. The dataset is private, but it can be made available upon request.

EfficientNet

The EfficientNet (Tan and Le 2019) is in fact a family of models defined on the baseline network described in Table 2. This base architecture (B0) is found with the aid of a network architecture search (NAS) method.

Its main component (or block) is known as the Mobile Inverted Bottleneck Conv (MBconv) Block introduced in (Sandler et al. 2018) and depicted in Fig. 3.

The rationale behind the EfficientNet family is to start from a high-quality yet compact baseline model and uniformly scale each of its dimensions systematically with a fixed set of scaling coefficients. Formally, an EfficientNet is defined

Table 2 EfficientNet baseline network: B0 architecture

Stage	Operator	Resolution	#channels	#layers
1	Conv 3×3	224×224	32	1
2	MBConv1,k3x3	112×112	16	1
3	MBConv6,k3x3	112×112	24	2
4	MBConv6,k5x5	56×56	40	2
5	MBConv6,k3x3	28×28	80	3
6	MBConv6,k5x5	14×14	112	3
7	MBConv6,k5x5	14×14	192	4
8	MBConv6,k3x3	7×7	320	1
9	Conv1x1/Pooling/FC	7×7	1280	1

by three dimensions: (i) depth, (ii) width, and (iii) resolutions as illustrated in Fig. 4.

Starting from the baseline model in Table 2, each dimension is scaled by the parameter according to Eq. 1

$$\begin{aligned}
 \text{depth} &= \alpha^\phi \\
 \text{width} &= \beta^\phi \\
 \text{resolution} &= \gamma^\phi \\
 \text{s.t. } &\alpha \cdot \beta^2 \cdot \gamma^2 \approx 2 \\
 &\alpha \geq 1, \beta \geq 1, \gamma \geq 1
 \end{aligned} \tag{1}$$

where α , β , and γ are constants obtained by a grid search experiment conducted in Tan and Le (2019). As stated in Tan and Le (2019), Eq. 1 provides a nice balance between performance and computational cost. The coefficient controls the available resources. Equation 1 determines the increase or decrease of model FLOPS when depth, width, and resolution are modified.

Architectures B1 to B7 are derived from architecture B0. Using the same methodology (network architecture search), more blocks were included at the top of the B0 model, making it deeper and wider. Thus, new efficient models were found and labeled (B1 to B7) during the search.

Notably, in Tan and Le (2019), a model from EfficientNet family was able to beat the powerful GPipe Network (Huang et al. 2019) on the ImageNet dataset (Russakovsky et al. 2015) running with 8.4x fewer parameters and being 6.1x faster.

Fig. 2 Radiograph example of images from *HCV-UFPR COVID-19 Dataset*. (a) X-ray of a subject infected with COVID-19. (b) X-ray of a subject without COVID-19



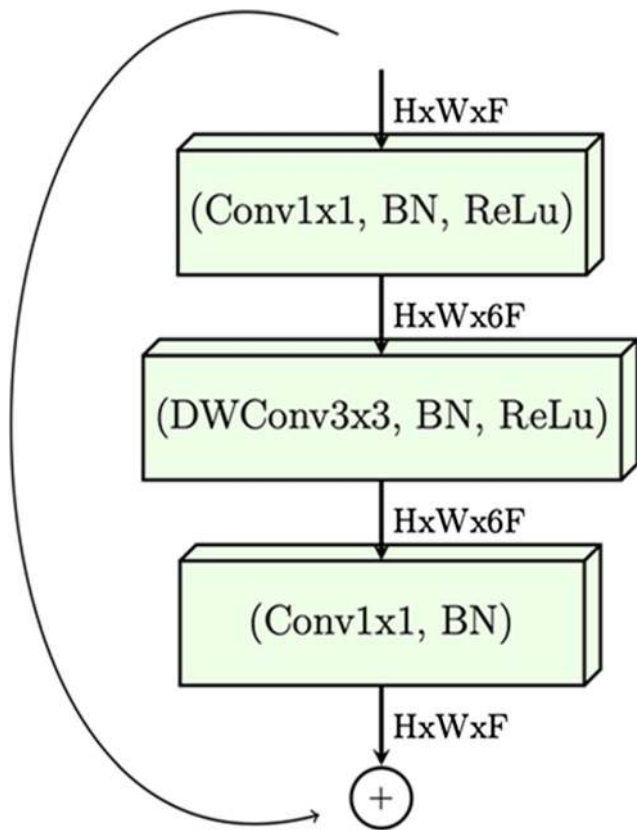


Fig. 3 MBCConv Block (Sandler et al. 2018). DWConv stands for depthwise conv, $k3 \times 3/k5 \times 5$ defines the kernel size, BN is batch norm, $H \times W \times F$ means tensor shape (height, width, depth), and $\times 1/2/3/4$ is the multiplier for number of repeated layers (Figure created by the authors)

Hierarchical classification

In classification problems, it is common to have some sort of relationship among classes. Very often, on real problems, the classes (the category of an instance) are organized hierarchically, like a tree structure. According to Silla and Freitas (2011), one can have three types of classification: flat classification, which ignores the hierarchy of the tree; local classification, in which there is a set of classifiers for each level of the tree (one classifier per node or level); and finally, global classification, in which one single classifier is built with the ability to classify any node in the tree, besides the leaves.

Fig. 4 Efficient net compound scaling on three parameters (Adapted from (Tan and Le 2019))

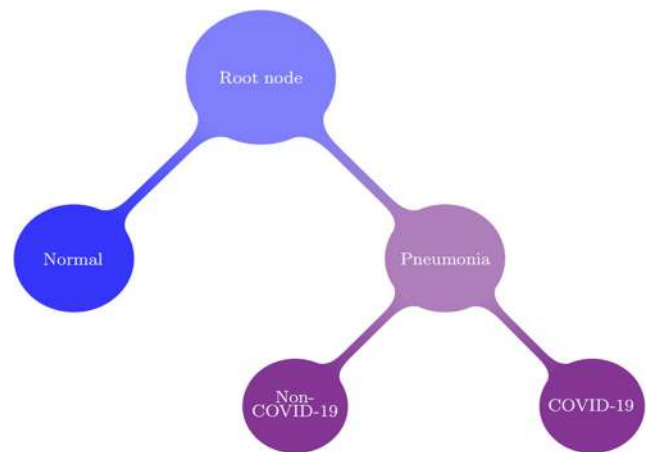
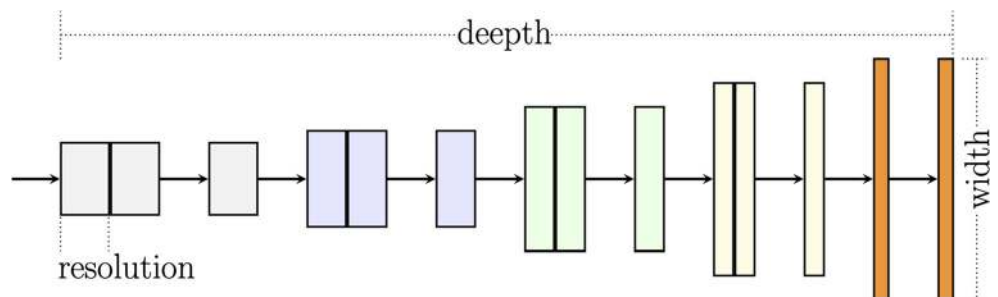


Fig. 5 Natural topology of the classes: normal, pneumonia, and COVID-19. It illustrates the Local-Per-Node hierarchical approach, in which there is a classifier on each parent node (Figure created by the authors)

The most popular type of classification in the literature is the flat one. However, here we propose the use of local classification, which we call hierarchical classification. Thus, the target classes are located in the leaves of the tree, and in the intermediate nodes, we have classifiers. In this work, we need two classifiers, one at the root node, dedicated to discriminate between the Normal and Pneumonia classes and another one in the next level dedicated to discriminate between pneumonia types. The problem addressed here can be mapped as the topology depicted in Fig. 5 in which there are two levels of classification. To make the class inference for a new instance, first, the instance is presented to the first classifier (in the root node). If it is predicted as “Normal,” the inference ends there. If the instance is considered “Pneumonia,” it is then presented to the second classifier, which will discern whether it is a pneumonia caused by “COVID-19” or “Not.”

Training

Deep learning models are complex and therefore require a large number of instances to avoid overfitting, i.e., when the learned network performs well on the training set but underperforms on the test set. Unfortunately, for most problems in real-world situations, data is not abundant. In fact, there are few scenarios in which there is an abundance of

training data, such as the ImageNet (Russakovsky et al. 2015), in which there are more than 14 million images of 21,841 classes/categories. To overcome this issue, researchers rely on two techniques: data augmentation and transfer learning. We also detail here the proposed models, based on EfficientNet.

Image pre-processing and data augmentation

Several pre-processing techniques may be used for image cleaning, noise removal, outlier removal, etc. The only pre-processing applied in this work is a simple intensity normalization of the image pixels to the range [0; 1]. In this manner, we rely on the filters of the convolutional network itself to perform possible data cleaning. Also, all images are resized according to the architecture resolution parameter (See Table 2).

Data augmentation consists of expanding the training set with transformations of the images in the dataset (Goodfellow et al. 2016) provided that the semantic information is not lost. In this work, we applied three transformations to the images: rotation, horizontal flip, and scaling, as such transformations would not hinder, for example, a physician to interpret the radiography. Figure 6 presents an example of the applied data augmentation. The transformations applied to the images are rotation (range 0 to 15 degrees clockwise or anticlockwise), Zoom with a range of 0–20%, and horizontal flipping (50% chance). All or none changes may be applied/combined according to a probability.

Fig. 6 Data augmentation applied using the Augmentor python package. The transformations applied to the images are rotation (0 to 15 degrees clockwise or anticlockwise), 20% zoom, or horizontal flipping. All or none changes may be applied/combined according to a probability (Figure created by the authors)

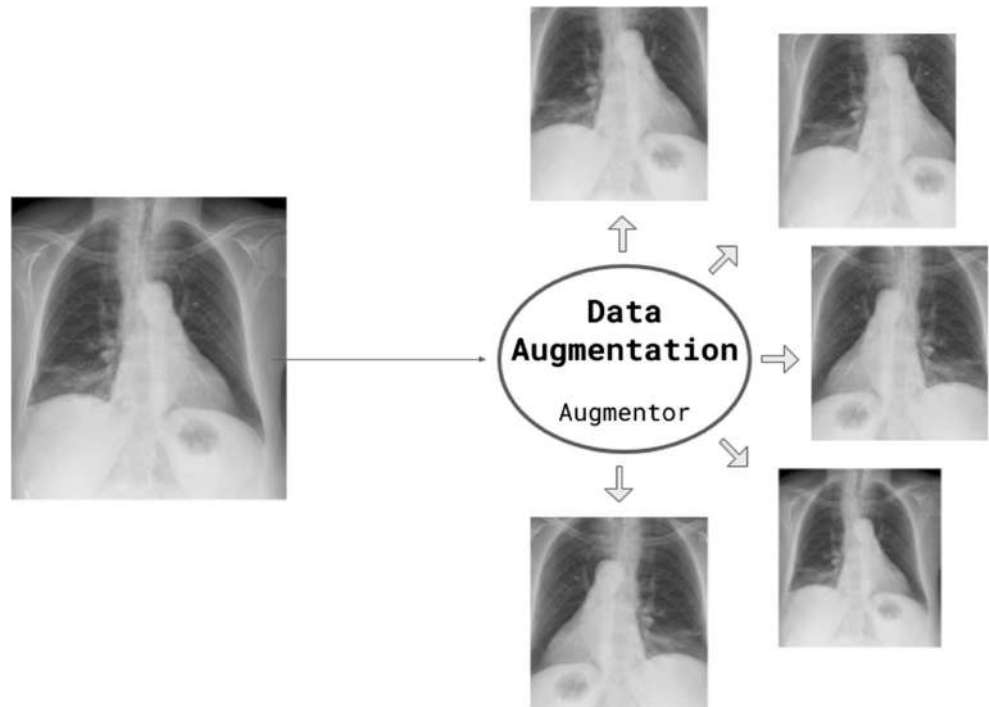


Table 3 Proposed family architectures, considering one EfficientNet model as the base model. (NC = Number of Classes)

Stage	Operator	Resolution	#channels	#layers
1–9	EfficientNet BASE (B0 to B5)	224×224	32	1
10	BN/Dropout	7×7	1280	1
11	FC/BN/Swich/Dropout	1	512	1
12	FC/BN/Swich	1	128	1
13	FC/Softmax	1	NC	1

Proposed family of models

The EfficientNet family has models of high performance and low computational cost. Since this research aims to find efficient models capable of being embedded in conventional smartphones, the EfficientNet family is a natural choice. We explore the EfficientNets by adding more operator blocks atop of it. More specifically, we add four new blocks, as detailed in Table 3. Thus, we proposed six new architectures, varying the base model. We add the suffix “X” to differentiate the proposed architectures from original EfficientNet base architectures.

Since the original EfficientNets were built to work on a different classification problem, we add new fully connected layers (FC) responsible for the last steps of the classification process. We also use batch normalization (BN), dropout, and swish activation functions for the following reasons.

The batch normalization constrains the output of the last layer in a range, forcing zero mean and standard deviation one. That acts as regularization, increasing the stability of the neural network, and accelerating the training (Ioffe and Szegedy 2015).

The Dropout (Srivastava et al. 2014) is perhaps the most powerful method of regularization. The practical effect of dropout operation is to emulate a bagged ensemble of multiple neural networks by inhibiting a few neurons, at random, for each mini-batch during training. The number of inhibited neuronal units is defined by the dropout parameter, which ranges between 0 and 100%.

The most popular activation function is the Rectified Linear Unit (ReLU), which can be formally defined as $f(x) = \max(0; x)$. However, in the added block, we have opted for the switch activation function (Ramachandran et al. 2017) defined as:

$$f(x) = x \cdot (1 + e^{-x})^{-1} \quad (2)$$

Differently from the ReLU the swish activation produces a smooth curve during the minimization loss process when a gradient descent algorithm is used. Another advantage of the swish activation regarding the ReLU is that it does not zero out small negative values which may still be relevant for capturing patterns underlying the data (Ramachandran et al. 2017).

Transfer learning

Instead of training a model from scratch, one can take advantage of using the weights from a pre-trained network and accelerate or enhance the learning process. As discussed in Oquab et al. (2014), the initial layers of a model can be seen as feature descriptors for image representation, and the latter ones are related to instance categories. Thus, in many applications, several layers can be reused. The task of transfer learning then defines how and what layers of a pre-trained model should be used. This technique has proved to be effective in several computer vision tasks, even when transferring weights from completely different domains (Goodfellow et al. 2016; Luz et al. 2018).

The steps for transfer of learning are:

1. Copying the weights from a pre-trained model to a new model
2. Modifying the architecture of the new model to adapt it to the new problem, possibly including new layers
3. Initializing the new layers
4. Defining which layers will pass through a new the learning process

5. Training (updating the weights according to the loss function) with a suitable optimization algorithm

We apply transfer learning to the EfficientNets pre-trained on the ImageNet dataset (Russakovsky et al. 2015). It is clear that the ImageNet domain is much broader than the chest X-rays that will be presented to the models in this work. Thus, the imported network weights are taken just as an initial solution and are all fine-tuned (i.e., the weights from all layers) by the optimizer over the new training phase. The rationale is that the imported models already have a lot of knowledge about all sorts of objects. By permitting all the weights to get fine-tuned, we allow the model to specialize to the problem in hands. In the training phase, the weights are updated with the Adam Optimizer and a schedule rule decreasing the learning rate by a factor of 10 in the event of stagnation (“patience = 2”). The learning rate started with 10^{-4} , and the number of epochs fixed at 20.

Model evaluation and metrics

The final evaluation is carried out with the COVIDx dataset, and since the COVIDx comprises a combination of two other public datasets, we follow the script (<https://github.com/lindawangg/COVID-Net>) provided in Wang et al. (2020) to load the training and test sets. The data is then distributed according to Table 1.

In this work, three metrics are used to evaluate models: accuracy (Acc), COVID-19 sensitivity (Se_C), and COVID-19 positive prediction ($+P_C$), i.e.,

$$\begin{aligned} Acc &= \frac{TP_N + TP_P + TP_C}{\#samples} \\ Se_C &= \frac{TP_C}{TP_C + FN_C} \\ +P_C &= \frac{TP_C}{TP_C + FP_C} \end{aligned} \quad (3)$$

wherein TP_N , TP_P , TP_C , FN_C , and FP_C stand for the normal samples correctly classified, non-COVID-19 samples correctly classified, the COVID-19 samples correctly classified, the COVID-19 samples classified as normal or non-COVID-19, the non-COVID-19, and normal samples classified as COVID-19. The number of multiply-accumulate (MAC) operations are used to measure the computational cost.

Experiments and discussion

In this section, we present the dataset setup, experimental settings, and results, which can be divided into three-fold: (i) flat vs hierarchical approaches, (ii) ablation study, and (iii) cross-dataset evaluation. Finally, we discuss the results. The execution environment of the computational experiments was

conducted on an Intel(R) Core(TM) i7-5820K CPU @ 3.30GHz, 64Gb Ram, two Titan X with 12Gb, and the TensorFlow/Keras framework for Python.

Dataset setup 1

Three different training set configurations were analyzed with the COVIDx dataset: (i) (Raw Dataset), the raw dataset without any pre-processing; (ii) (Raw Dataset + Data Augmentation), the raw dataset with a data augmentation of 1000 new images on COVID-19 samples and a limitation of 4000 images for the two remaining classes; and (iii) (Balanced Dataset), the dataset with a 1000 images per class achieved by data augmentation on COVID-19 samples and under-sampling the other two classes to 1000 samples each one. Learning with an unbalanced dataset could bias the prediction model towards the classes with more samples, leading to inferior classification models.

In this work, we evaluate two scenarios: flat and hierarchical. Regardless of the scenarios, the three training sets remain the same (Raw, Raw + Data Augmentation, and Balanced). However, for the hierarchical case, there is an extra process to split the sets into two parts: the first part, the instances of pneumonia and COVID-19 classes are joined and receive the same label (Pneumonia). In the second part, the instances related to the normal class are removed, leaving in the set only instances related to Pneumonia and COVID-19. Thus, two classifiers are built for the hierarchical case, and each one works with a different set of data (see Section 4.3 for more details). The ablation study also uses this dataset setup.

Dataset setup 2

In order to assess the impact of learning a model in one data distribution and evaluate on another one, the COVIDx dataset is only used for training/validation, and the HCV-UFPR COVID-19 Dataset is entirely reserved for testing. This scenario, the cross-dataset evaluation, is closer to reality since in a real-world situation, the models should face samples acquired from different sensors, individuals, and environment.

Experimental settings and results

Flat vs hierarchical

We evaluate four families of convolutional neural networks: EfficientNet, MobileNet, VGG and, ResNet on Dataset Setup 1. Our method includes EfficientNet architectures as base building blocks (B0-B5) with the insertion of 4 custom blocks at the top, as detailed in the Methodology section. These new architectures, we call B0-X, B1-X, B2-X, B4-X, and B5-X. Their features are summarized in Table 4. Among the

Table 4 Base models footprint details. (Mb = Megabytes)

Model	Input shape	#params	Memory usage(MB)
EfficientNet B0-X	224, 224, 3	5,330,564	21
EfficientNet B1-X	240, 240, 3	7,856,232	31
EfficientNet B2-X	260, 260,3	9,177,562	36
EfficientNet B3-X	300, 300, 3	12,320,528	48
EfficientNet B4-X	380, 380, 3	19,466,816	76
EfficientNet B5-X	456, 456, 3	30,562,520	118
MobileNet	224, 224, 3	4,253,864	17
MobileNet V2	224, 224, 3	3,538,984	14
RESNET 50	224, 224, 3	25,636,712	99
VGG-16	224, 224, 3	138,357,544	528
VGG-19	224, 224, 3	143,667,240	549

presented models, we highlight the low footprint of MobileNet and EfficientNet based models.

Regarding the base models (B0-B5 models of the EfficientNet family), the simplest one is the EfficientNet-B0. Thus, we assess the impact of the different training sets and the two forms of classification (flat and hierarchical) for our models derived from B0 (B0-X). The results are shown in Table 5.

Since there are more pneumonia, and normal x-ray samples than COVID-19, the neural network learning process tends to improve the classification of the majoritarian classes because they will have more weight in the loss calculation. This may justify the results obtained by balancing the data. As described in Section 4.3, the hierarchical approach is also evaluated here. First, classes of COVID-19 and common pneumonia are combined and presented to the first level of classification (normal vs pneumonia). At the second level, another model classifies instances into pneumonia caused by COVID-19 or other causes.

It is possible to see on Table 5 that better results are achieved with the flat approach on balanced data. This

Table 5 EfficientNet B0-X results over the three proposed training sets. (Acc = Accuracy; Se_C = COVID-19 Sensitivity; +P_C = COVID-19 positive prediction)

Approach	Training dataset	Acc	Se _C	+P _C
Flat	Raw dataset	92.2%	67.7%	100.0%
	Raw dataset + data augmentation	93.0%	83.8%	100.0%
	Balanced dataset	90.0%	93.5%	100.0%
Hierarchical	Raw dataset	54.1%	83.8%	100.0%
	Raw dataset + data augmentation	90.4%	70.9%	91.6%
	Balanced dataset	85.7%	93.5%	82.8%

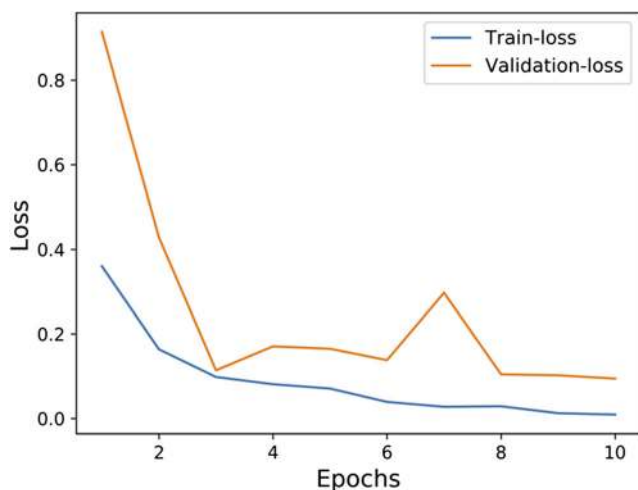


Fig. 7 Loss during training-time, EfficientNet-B0, and balanced data. Epochs vs Loss

scenario is used to evaluate the remaining network base architectures. The training loss for this scenario is presented in Fig. 7.

The results of all evaluated architectures are summarized in Table 6. We stress that we adapted all architectures by placing the same four blocks on top. It can be seen that all the networks have comparable performances in terms of accuracy. However, the more complex the model is, the worse is the performance for the minority class, the COVID-19 class.

The cost of a model is related to the number of parameters. The higher the number of parameters, the higher the amount of data the model needs to adjust them. Thus, we hypothesized that the lack of a bigger dataset may explain the difficulties faced by the more complex models.

Table 7 presents a comparison of the proposed approach and the one proposed by Wang et al. (2020) (COVID-net)

Table 6 Results on different network architectures as base model. Best scenario for COVID-19: all experiments with a balanced training set and flat classification. (Acc: = accuracy; Se_C = COVID-19 sensitivity; $+P_C$ = COVID-19 positive prediction)

Model	Acc	Se_C	$+P_C$
EfficientNet B0-X	90.0%	93.5%	100.0%
EfficientNet B1-X	91.8%	87.1%	100.0%
EfficientNet B2-X	90.0%	77.4%	100.0%
EfficientNet B3-X	93.9%	96.8%	100.0%
EfficientNet B4-X	93.0%	90.3%	93.3%
EfficientNet B5-X	92.2%	93.5%	90.6%
MobileNet	90.4%	83.8%	100.0%
MobileNet V2	90.0%	87.1%	96.4%
RESNET 50	83.5%	70.9%	81.4%
VGG-16	77.0%	67.7%	63.64%
VGG-19	75.3%	77.4%	50.0%

under the same evaluation protocol. Even though the accuracy is comparable, the proposed approach presents an improvement on positive prediction without losing sensitivity. Besides, a significant reduction both in terms of memory (our model is 15 times smaller) and latency is observed. It is worth highlighting that Wang et al. (2020) apply data augmentation to the dataset, but it is not clear in their manuscript how many new images are created.

The COVID-Net (Wang et al. 2020) is a very complex network, which demands a memory of 2.1GB (for the smaller model) and performs over 3.5 billion MAC operations implying three main drawbacks: computation-cost, time-consumption, and infrastructure costs. A 3.59 billion MAC operations model takes much more time and computations than a 11.5 million MAC model (in the order of almost 300 times), and the same GPU necessary to run one COVID-Net model can run more than 15 models of the proposed approach (B3-X flat approach) keeping a comparable (or even better) figures. The improvements, in terms of efficiency, are even greater using the B0-X - with a small trade-off in terms of the sensitivity metric. The complexity can hinder the use of the model in the future, for instance, on mobile phones or common desktop computers (without GPU).

Ablation study

In order to customize the network architectures to best suit the problem, in this work, we propose the addition of new blocks on top of the networks. To assess the effectiveness of the proposal, we performed an ablation study, training the B3-based architecture with and without the proposed blocks under the same conditions (same set of batch data, same hyperparameters, and same random seed). The study showed that the inclusion of the 4 proposed blocks improves the total accuracy of the model by 2.3% (from 91.77 to 93.0%). Also, the inclusion of the blocks allows a better trade-off between the metrics Se_C and $+P_C$. With the addition of the proposed blocks, the $+P_C$ increased from 74.19 to 100%; however, Se_C dropped from 100 to 96.8%.

Cross-dataset evaluation

Cross-evaluation between databases is of paramount importance to ascertain the power of generalization of the model regarding variations in images (due to different equipment and sensors). Thus, the Dataset Setup 2 is used for that aim. Table 8 summarizes the experimental results. The proposed approach (B3-X) overcomes the COVID-Net (version CXR Large) proving to be more robust than the other approaches evaluated.

Table 7 Comparison of the proposed approach against SOTA. (Acc = Accuracy; Se_C = COVID-19 Sensitivity; $+P_C$ = COVID-19 Positive Prediction)

Method	Acc	Se_C	$+P_C$	#Params (millions)	MACs (millions)	Memory required
Approach Flat EfficientNet B3-X	93.9%	96.8%	100.0%	11.6	11.5	134 Mb
Approach Hierarchical EfficientNet B3-X	93.5%	80.6%	100.0%	23.2	23.0	268 Mb
COVID-Net (Wang et al. 2020)	94.3%	96.8%	90.9%	126.6	3500.0	2.1Gb

Discussion

In Fig. 8, we present two X-ray images of COVID-19-infected individuals. Those images are from the test set and, therefore, were not seen by the model during training. According to studies (Ng et al. 2020), the presence of the COVID-19 infection can be observed through some opacity (white spots) on chest radiography imaging. In the first row of Fig. 8, one can see the corrected classified image and its respective activation maps generated by our model. The activation map corresponds to opaque points in the image, which may correspond to the presence of the disease. For the second row images, it is observed that the model failed to find opaque points in the image and the activation map highlights non-opaque regions.

In Fig. 9, the confusion matrices of flat and hierarchical approaches are presented. It is possible to observe that the hierarchical model classifies the normal class better, though it also shows a noticeable reduction in terms of sensitivity and positive prediction for the COVID-19 class. One hypothesis is that both Pneumonia and COVID-19 classes are similar (both kinds of pneumonia) and share key features. Thus, the lack of normal images on second classification level reduces the diversity of the training set, interfering with model training. Besides, the computational cost is twice as higher than flat classification since two models are required. However, we believe that the hierarchical approach has a key aspect: it suffers less from bias in the dataset/protocol. In Maguolo and Nanni (2020), a critical evaluation of the test protocols and databases for methods aiming at classifying COVID-19 in X-ray images is presented. According to Maguolo and Nanni (2020), the considered datasets are mostly composed of images from different distributions and different databases, and

this may favor the deep learning models to learn patterns related to the image acquisition process, instead of focusing only on disease patterns.

In the first stage of the hierarchical classification, images related to COVID-19 and non-COVID pneumonia are given the same classification label. Thus images from different datasets are combined which forces the method to disregard patterns related to the acquisition process or sensors at the first classification stage. An example of the hierarchical model application can be seen in Fig. 10. It can be seen from the confusion matrix of the first stage (Fig. 10 (left)) that the model is able to classify most instances correctly, and for that, we believe it has focused on the patterns that may help discriminate among different types of pneumonia.

Results of the ablation study showed that the inclusion of the additional blocks significantly improved the trade-off between Se_C and $+P_C$, increasing the total accuracy of the model, which justifies the increase in computational cost.

Regarding cross-dataset evaluation, the results showed that even the models considered state-of-the-art suffer from variations caused by the difference in sensors, equipment and acquisition protocols. These findings reveal that, in order to have a model able to work in the field, one must train (or adjust) the model with representative local data.

Findings and future direction

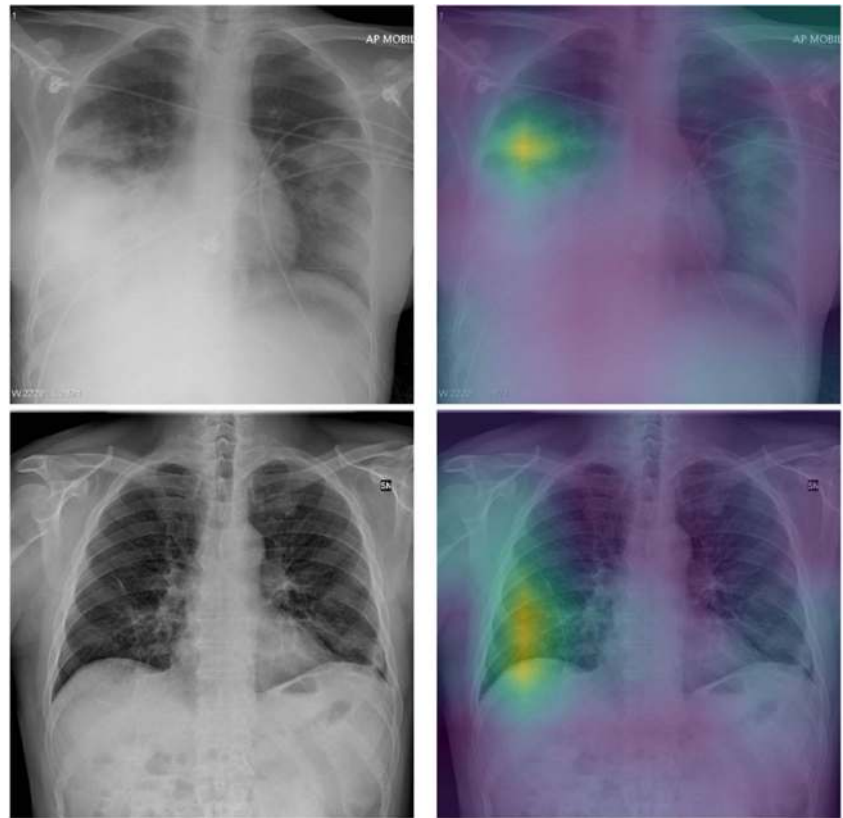
We summarize our findings as follows.

- An efficient and low computational family of models was proposed to detect COVID-19 patients from chest X-ray images. Even with only a few images of the COVID-19

Table 8 Comparison of the proposed approach against SOTA on Dataset 2 Setup: A cross-dataset evaluation on HCV-UFPR COVID-19 Dataset. (Acc = Accuracy; Se_C = COVID-19 Sensitivity; $+P_C$ = COVID-19 Positive Prediction)

Method	Acc	Se_C	$+P_C$	#Params (millions)	MACs (millions)	Memory required
Approach Flat EfficientNet B3-X	52.63%	55.43%	69.04%	11.6	11.5	134 Mb
DenseNet201 (Huang et al. 2019)	45.22%	50.00%	20.49%	18.3	18.1	212 Mb
COVID-Net(Wang et al. 2020)	51.27%	41.64%	57.64%	127.4	3500.0	3.5Gb

Fig. 8 Original images and their activation maps according to the proposed approach. First row presents a patient with COVID-19 (a corrected classified image and its respective activation maps generated by our model), the second, from a healthy chest x-ray sample (the model failed to find opaque points in the image and the activation map highlights non-opaque regions)



class, insightful results with a sensitivity of 90% and a positive prediction of 100% were obtained, with the evaluation protocol proposed in (Wang et al. 2020).

- Regarding the hierarchical analysis, we conclude that there are significant gains that justify the use of the present task. We believe that it suffers less from the bias present in the evaluation protocols, already discussed in (Maguolo and Nanni 2020).
- The proposed network blocks, put on top of the base models, showed to be very effective for the CRX detection problem, in particular, CRX related to COVID-19.
- The evaluation protocol proposed in Wang et al. (2020) is based on the public dataset “COVID-19 Image Data Collection” (Cohen et al. 2020), which is being expanded by the scientific community. With more images from the COVID-19 class, it will be possible to improve the training. However, the test partition tends to become more challenging. For sake of reproducibility and future comparisons of results, our code is available at https://github.com/ufo_pcsilab/EfficientNet-C19.
- The cross-dataset evaluation showed that even models considered state-of-the-art to detect COVID-19 in X-ray have their performance severely deteriorated by variation in images caused by differences in sensors or acquisition protocols. To overcome such an issue and increase generalization power, models should be re-trained or fine-tuned on more representative data.
- The Internet of Medical Things (IOMT) (Joyia et al. 2017) is now a hot topic in industry. However, the Internet can be a major limitation for medical equipment,

Fig. 9 Confusion matrix of flat (left) and hierarchical (right) approaches respectively with balanced training set. Class zero is the normal images (1) pneumonia non-COVID-19 and (2) COVID-19

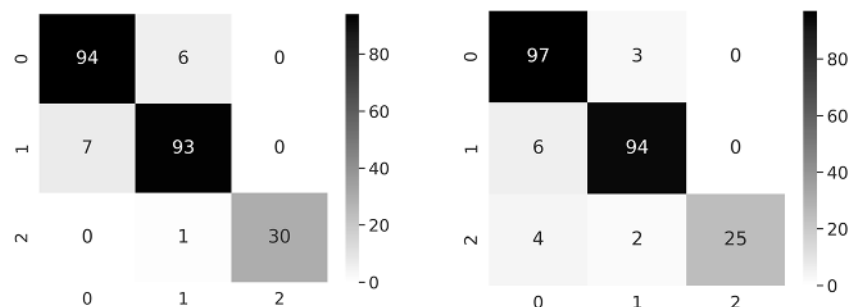
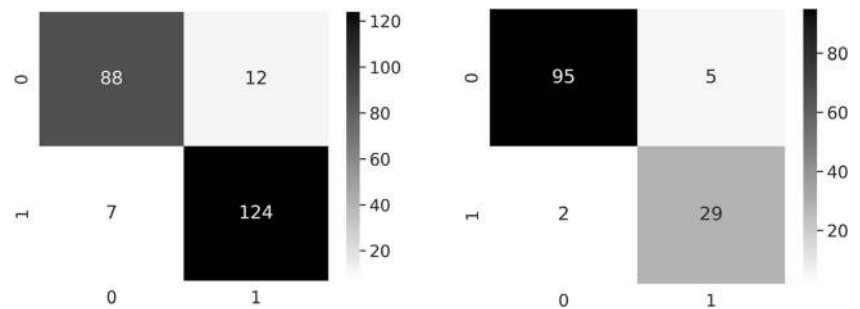


Fig. 10 Confusion matrix hierarchical approach respectively with balanced training set. On the first hierarchical stage (left), class zero is for the normal images and class 1 all Pneumonias (non-covid and COVID-19 pneumonias). On the second stage (right) class 0 non-COVID-19 pneumonia, and, 2, COVID-19



especially in poor countries. Our proposal is to move towards a model that can be fully embedded in conventional smartphones (edge computing), eliminating the use of the Internet or cloud services. In that sense, the model achieved in this work requires only 55 Mb of memory and has a viable inference time for a conventional cell phone processor.

Conclusion

In this paper, we exploit an efficient convolutional network architecture for detecting any abnormality caused by COVID-19 through chest radiography images. Experiments were conducted to evaluate the neural network performance on the COVIDx dataset, using two approaches: flat classification and hierarchical classification. Although the datasets are still incipient and, therefore, limited in the number of COVID-19-related images, effective training of the deep neural networks has been made possible with the application of transfer learning and data augmentation techniques.

Concerning evaluation, the proposed approach brought improvements compared with baseline work, with an accuracy of 93.9%, COVID-19 sensitivity of 96.8%, and positivity prediction of 100% with a computational efficiency more than 30 times higher.

We believe that the current proposal is a promising candidate for embedding in medical equipment or even physicians' mobile phones. However, larger and more heterogeneous databases are still needed to validate the methods before claiming that deep learning can assist physicians in the task of detecting COVID-19 in X-ray images.

Acknowledgments The authors would like to thank UFOP, FAPEMIG, CAPES, and CNPq for the financial support. The authors also would like NVIDIA for the donation of one GPU Titan X and two GPU Titan Xp.

Declaration

Conflict of interest On behalf of all authors, the corresponding author states that there is no conflict of interest.

References

- Ai T, Yang Z, Hou H, Zhan C, Chen C, Lv W, et al. Correlation of chest ct and rt-pcr testing in coronavirus disease 2019 (covid-19) in china: a report of 1014 cases. *Radiology*. 2020;200642.
- Al-Bawi A, Al-Kaabi K, Jeryo M, Al-Fatlawi A. CCBBlock: an effective use of deep learning for automatic diagnosis of COVID-19 using X-ray images. *Research on Biomedical Engineering*. 2020;1–10.
- American College of Radiology. ACR recommendations for the use of chest radiography and computed tomography (CT) for suspected COVID-19 infection. *ACR website. Advocacy-and Economics/ACR-Position-Statements/Recommendations-for-Chest-Radiography-and-CTfor-Suspected-COVID19-Infection..2020; Updated March, 22*.
- Araujo-Filho JAB, Sawamura MVY, Costa AN, Cerri GG, Nomura CH. COVID-19 pneumonia: what is the role of imaging in diagnosis? *J Bras Pneumol*. 2020;46.
- Chollet F. Xception: Deep learning with depthwise separable convolutions. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2017. p. 1800–7.
- Cohen JP, Morrison P, Dao L. Covid-19 image data collection. *arXiv*. 2020; preprint arXiv:200311597.
- Davaranah AH, Mahdavi A, Sabri A, Langroudi TF, Kahkouee S, Haseli S, et al. Novel screening and triage strategy in iran during deadly coronavirus disease 2019 (covid-19) epidemic: value of humanitarian teleconsultation service. *Journal of the American College of Radiology*. 2020; *JACR*;17(6):734–8.
- Deng J, Dong W, Socher R, Li L, Li K, Fei-Fei L. Imagenet: A large-scale hierarchical image database. In: 2009 IEEE Conference on Computer Vision and Pattern Recognition; 2009. p. 248–55.
- Farooq M, Hafeez A. Covid-resnet: A deep learning framework for screening of covid19 from radiographs. *arXiv*. 2020; preprint arXiv:200314395.
- Giovagnoni A. Facing the covid-19 emergency: we can and we do. *Radiol Med*. 2020;125(4):337–8.
- Goodfellow I, Bengio Y, Courville A. *Deep learning*: MIT Press; 2016.
- Hemdan EED, Shouman MA, Karar ME. Covidx-net: A framework of deep learning classifiers to diagnose covid-19 in x-ray images. *arXiv*. 2020; preprint arXiv:200311055.
- Huang Y, Cheng Y, Bapna A, Firat O, Chen D, Chen M, et al. Gpipe: Efficient training of giant neural networks using pipeline parallelism. In: *Advances in Neural Information Processing Systems*; 2019. p. 103–12.
- Huang P, Liu T, Huang L, Liu H, Lei M, Xu W, et al. Use of chest ct in combination with negative rt-pcr assay for the 2019 novel coronavirus but high clinical suspicion. *Radiology*. 2020;295(1):22–3.
- Ioffe S, Szegedy C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. *arXiv*. 2015; preprint arXiv:150203167.
- Jaiswal AK, Tiwari P, Kumar S, Gupta D, Khanna A, Rodrigues JJ. Identifying pneumonia in chest x-rays: a deep learning approach. *Measurement*. 2019;145:511–8.

- Joyia GJ, Liaqat RM, Farooq A, Rehman S. Internet of medical things (IOMT): applications, benefits and future challenges in healthcare domain. *J Commun*. 2017;12(4):240–7.
- Khan AI, Shah JL, Bhat MM. Coronet: a deep neural network for detection and diagnosis of covid-19 from chest x-ray images. *Comput Methods Prog Biomed*. 2020;196:105581.
- LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature*. 2015;521(7553):436–44.
- Luz E, Moreira G, Junior LAZ, Menotti D. Deep periocular representation aiming video surveillance. *Pattern Recogn Lett*. 2018;114:2–12.
- Maguolo G, Nanni L. A critic evaluation of methods for covid-19 automatic detection from x-ray images. *arXiv*. 2020; preprint arXiv:200412823.
- Ng MY, Lee EY, Yang J, Yang F, Li X, Wang H, et al. Imaging profile of the covid-19 infection: radiologic findings and literature review. *Radiology: Cardiothoracic Imaging*. 2020;2(1):e200034.
- Oquab M, Bottou L, Laptev I, Sivic J. Learning and transferring mid-level image representations using convolutional neural networks. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*; 2014. p. 1717–24.
- Pereira RM, Bertolini D, Teixeira LO, Silla CN Jr, Costa YM. Covid-19 identification in chest x-ray images on flat and hierarchical classification scenarios. *Computer Methods and Programs in Biomedicine*. 2020:105532.
- Radiological Society of North America. Rsn pneumonia detection challenge. <https://www.kaggle.com/c/rsna-pneumonia-detection-challenge/data>. 2020; accessed: 2020-04-01.
- Rajpurkar P, Irvin J, Zhu K, Yang B, Mehta H, Duan T, et al. Chexnet: Radiologist-level pneumonia detection on chest x-rays with deep learning. *arXiv*. 2017; preprint arXiv:171105225.
- Ramachandran P, Zoph B, Le QV. Searching for activation functions. *arXiv*. 2017; preprint arXiv:171005941.
- Russakovsky O, Deng J, Su H, Krause J, Satheesh S, Ma S, et al. Imagenet large scale visual recognition challenge. *Int J Comput Vis*. 2015;115(3):211–52.
- Sandler M, Howard A, Zhu M, Zhmoginov A, Chen LC. Mobilenetv2: Inverted residuals and linear bottlenecks. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*; 2018. p. 4510–20.
- Silla CN, Freitas AA. A survey of hierarchical classification across different application domains. *Data Min Knowl Disc*. 2011;22(1–2):31–72.
- Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. *arXiv*. 2014; preprint arXiv:14091556.
- Società Italiana di Radiologia Medica e Interventistica. COVID-19. database. <https://www.sirm.org/en/category/articles/covid-19-database/>. 2020; accessed: 2020-04-12.
- Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R. Dropout: a simple way to prevent neural networks from overfitting. *The journal of machine learning research*. 2014;15(1):1929–58.
- Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z. Rethinking the inception architecture for computer vision. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*; 2016. p. 2818–26.
- Szegedy C, Ioffe S, Vanhoucke V, Alemi AA. Inception-v4, inception-resnet and the impact of residual connections on learning. In: *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence*; 2017. p. 4278–84.
- Tan M, Le QV. Efficientnet: Rethinking model scaling for convolutional neural networks. *arXiv*. 2019; preprint arXiv:190511946.
- Touvron H, Vedaldi A, Douze M, Jégou H. Fixing the train-test resolution discrepancy: FixEfficientNet. *arXiv*. 2020; preprint arXiv:200308237.
- Wang X, Peng Y, Lu L, Lu Z, Bagheri M, Summers RM. Chestx-ray8: Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases. In: *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*; 2017. p. 3462–71.
- Wang L, Lin ZQ, Wong A. Covid-net: a tailored deep convolutional neural network design for detection of covid-19 cases from chest x-ray images. *Sci Rep*. 2020;10(1):1–12.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.