

# Few-shot image classification for automatic COVID-19 diagnosis

Daniel Cores<sup>1</sup>[0000–0002–5548–4837], Nicolás Vila-Blanco<sup>1,2,3</sup>[0000–0001–5865–9973], Manuel Mucientes<sup>1,2</sup>[0000–0003–1735–3585],  
and María J. Carreira<sup>1,2,3</sup>[0000–0003–0532–2351]

<sup>1</sup> Centro Singular de Investigación en Tecnoloxías Intelixentes (CiTIUS),  
Universidade de Santiago de Compostela, Santiago de Compostela, Spain  
{daniel.cores, nicolas.vila, manuel.mucientes, mariajose.carreira}@usc.es

<sup>2</sup> Departamento de Electrónica e Computación, Escola Técnica Superior de  
Enxeñaría, Universidade de Santiago de Compostela, Santiago de Compostela, Spain

<sup>3</sup> Instituto de Investigación Sanitaria de Santiago de Compostela (IDIS), Santiago de  
Compostela, Spain

**Abstract.** Developing robust and performant methods for diagnosing COVID-19, particularly for triaging processes, is crucial. This study introduces a completely automated system to detect COVID-19 by means of the analysis of Chest X-Ray scans (CXR). The proposed methodology is based on few-shot techniques, enabling to work on small image datasets. Moreover, a set of additions have been done to enhance the diagnostic capabilities. First, a network to extract the lung region to rely only on the most relevant image area. Second, a new cost function to penalize each misclassification according to the clinical consequences. Third, a system to combine different predictions from the same image to increase the robustness of the diagnoses. The proposed approach was validated on the public dataset COVIDGR-1.0, yielding a classification accuracy of  $79.10\% \pm 3.41\%$  and, thus, outperforming other state-of-the-art methods. In conclusion, the proposed methodology has proven to be suitable for the diagnosis of COVID-19.

**Keywords:** Chest X-Ray · COVID-19 · deep neural networks · few-shot classification

## 1 Introduction

The assessment of radiological images such as Computerized Tomography (CT) or Chest X-Ray (CXR) scans has demonstrated to be a reliable method for the screening and diagnosis of COVID-19 [9]. Different studies pointed out a set of visual indicators that can be used in this regard, including bilateral and/or interstitial abnormalities, among others [6]. In this regard, automated methods powered by machine learning techniques are a useful tool to improve the diagnosis workflow by reducing the time needed to analyze each scan [5]. In particular, convolutional neural networks (CNN) stand out as the most used technique due

to their ability to extract high-level image characteristics that are useful in the diagnosis.

The particularity of CNN-based systems is that very large datasets are needed for the training process. This led the research groups to collect their own sets of images [3], which turned out to be mostly unbalanced due to the high number of patients with severe condition. This, in addition to the heterogeneous sources used to acquire the images, caused the developed systems to report suspiciously high levels of sensitivity [27]. In the end, this conditioned the applicability of these models, as the detection of patients with low or moderate severity showed very poor performance.

Although more than three years have passed since the start of the pandemic, there are still few high-quality CXR datasets that can be used to build COVID-19 detection systems [17]. Furthermore, the limited size of the available datasets makes it difficult to carry out large-scale studies, and forces the researchers to apply techniques to artificially increase the number of images or to use specific learning algorithms that take into account the lack of data. Among the latter, the use of few-shot techniques is particularly relevant [7].

In a traditional setup, a CNN is fed with a large number of images of each output category during training. The model parameters are modified iteratively to better detect the relevant features in the images and ultimately improve the classification performance. If few images are available during the training process, the model is prone to overfitting the data, that is, learning specific features of each training image instead of general enough features associated with each category. In this regard, the few-shot frameworks address this issue through a meta-learning and fine-tuning approach. The main task in this case is no longer the extraction of relevant characteristics that identify each category, but those that allow to know whether two images belong to the same or to a different category. The strategy consists in comparing template images of each category (which will be referred to as the support set) with each input image and choosing the category that yields the higher affinity. An advantage of this methodology is that the categories used to train the model do not have to be the same as those used during the test phase.

In this paper, we introduce a new approach for the screening of patients with COVID-19 from CXR images. The proposal relies on few-shot techniques, so it is prepared to be used in reduced sets of images. The main contributions of this work are: 1) a system that focuses only on lung regions to classify the images, discarding other meaningless structures that may hinder the results; 2) a novel cost function that penalizes each misclassification according to the specific clinical cost; 3) an ensemble technique that combines different support sets to increase the robustness of the classification; and 4) a validation setup in a public dataset [22] that proves the suitability of the proposed approach regarding the classification performance.

## 2 Related work

Automatic COVID-19 diagnosis from CXR images quickly emerged as a very active research area, addressing the automatic triage problem applying image classification techniques. Thus, the initial trend was to rely on well-known traditional architectures for general image classification such as VGG [13], Xception [11], or CapsNets [1]. Due to the promising results achieved with these simple methods, more complex models specifically designed to detect the COVID-19 disease were proposed. COVID-Net [25] seeks execution efficiency while maintaining a high classification performance with the definition of lightweight residual blocks. CVDNet [15] focused on extracting local and global features, including two interconnected paths with different kernel sizes. Also, a lung segmentation network was introduced as a preprocessing stage to force the network to focus only on the lung region [22] or to guide the learning algorithm [14].

Many datasets with CXR COVID-19 images have been released since the pandemic outbreak. Due to the lack of available data, the initial approach followed by early work on automatic COVID-19 diagnosis was to aggregate multiple datasets [2,3]. Unfortunately, most of these datasets are biased towards patients with severe conditions, and generally do not provide a balanced set of positive and negative samples. The severity level bias might explain the abnormal accuracy level reported by some works by not considering the most challenging cases, i.e., patients with mild and moderate conditions [17]. Also, the variability in the data acquisition process caused by integrating different datasets from different sources may cause the model to learn features specific to each device [23].

COVIDGR [22] was presented as a more complete dataset, including high-quality annotated CXR images for both COVID and non-COVID patients extracted under the same conditions. Positive patients are also classified into four severity levels according to the radiological findings. Therefore, this dataset defines a realistic environment to evaluate solutions for automatic COVID-19 triage systems. However, the number of training samples per severity level remains low, which hinders the training of traditional models.

To address the problem of data scarcity, simple data augmentation transformations have been successfully applied to increase the variability of the training set and prevent the overfitting of the model [13]. This overfitting problem was also addressed by including handcrafted features [10]. As a step forward, the generation of synthetic images with generative adversary networks (GAN) was explored in [12]. Alternatively, transfer learning techniques have also proven to be highly effective in adapting a model pretrained in a large dataset to a new domain with limited training samples [12,26].

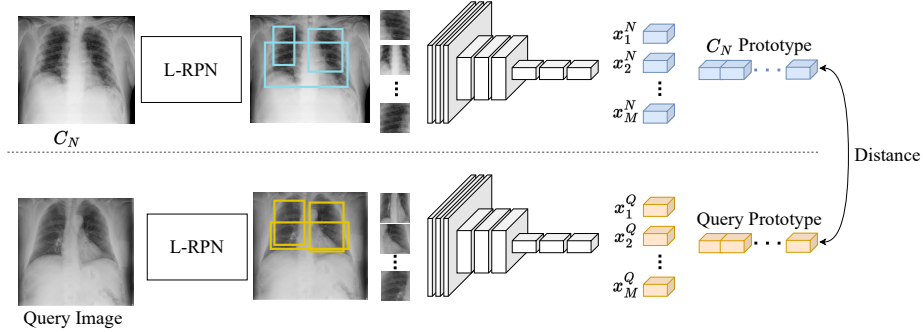
Few-shot techniques have recently emerged as a popular solution to solve the general image classification problem, dealing with very limited training sets [24]. One of the main lines of research in few-shot learning is the definition of meta-learners. Meta-learning algorithms redefine the classification problem into a similarity metric learning that can identify if two images belong to the same category. This property makes meta-learners more generalizable to novel classes with limited or even no training samples — typically up to 30 labeled samples

per category. These techniques have been effectively applied to general medical image processing [4], and to solve the COVID-19 classification problem [7,21].

As data availability in current COVID-19 datasets such as COVID-GR exceeds the usual few-shot setting but does not suffice to properly train a traditional classifier, an intermediate solution is desirable. Therefore, we propose a meta-learning architecture that fully exploits the available labeled information by performing a model fine-tuning in the complete training set and implement a support set ensemble at test time.

### 3 Methodology

The illustration in Fig. 1 depicts a novel classification framework for diagnosing COVID-19 using CXR scans. Our approach is ideal for cases where a limited number of images are available for training, as it is the situation with many new diseases. Here, few-shot classification methods stand out, as they are specifically developed to handle scenarios where conventional image classification methods would not perform well enough due to the scarcity of data. In the few-shot approach, every input image (query) is categorized based on its similarity to the images of the support set. To accomplish this, the prototypes for each category are computed and compared with the query prototype. The prototypes are derived from the deep feature maps generated by a CNN backbone.



**Fig. 1.** The proposed workflow for categorizing images using one support set.

We introduce a lung-focused network as the first step of our pipeline. This system aims at proposing regions of interest located in the lung area, both for the query and support images (as seen in Fig. 1 as output of L-RPN). This network design allows the subsequent classification system to process only the most relevant regions for the diagnosis of COVID-19, leading to improved performance. Additional information regarding this component is provided in subsection 3.1.

Next, a CNN network is used as the backbone to generate rich representations of the regions proposed in both the query and the support images. The

obtained per-region feature maps are combined to create a per-image feature representation. Then, the feature maps corresponding to the support images of each category are averaged to obtain the per-category feature map. In a further step, these maps are refined through gradient descent to obtain robust prototypes which can be utilized to evaluate the similarity of each query image to each category. This evaluation of affinity is performed using Earth Mover’s Distance (EMD) [19], which determines the pairwise distance between the prototypes of the query and those belonging to support regions. Therefore, the pair of regions that show the minimum EMD value is deemed to be the most similar.

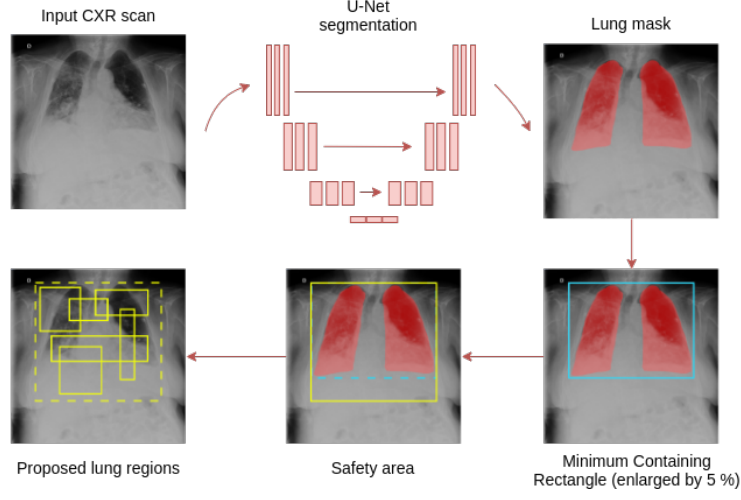
In typical image classification using few-shot methods, a minimal amount of support images is utilized, with the most common cases ranging from 1-shot to 10-shot, meaning between one and ten images per category. Nevertheless, these conditions are not applicable in this specific scenario, as more images per category are available in the dataset. Hence, we suggest a different approach employing a group of  $S$  sets of  $k$  support images each, rather than merely selecting a single set randomly, to mitigate the impact of each support image. This concept is outlined in subsection 3.2.

For training purposes, four categories of COVID-19 affection are considered, namely negative, mild, moderate, and severe. Since the magnitude of classification errors should vary depending on the proximity between the correct and the predicted categories, we propose a novel cost function that incorporates expert knowledge in this regard. This technique is explained in more detail in subsection 3.3.

### 3.1 L-RPN: Lung-aware region proposal network

It is commonly agreed that the lung area is the most affected by COVID-19, and so it has been widely used as the main radiological indicator for diagnosis. However, the position of the patient during a CXR scan is not homogeneous, as it depends on his/her severity and the presence of ventilation or monitoring devices that may not be removable. This situation can lead to a high degree of variability in the CXR scans in terms of the position and scale of the lungs, which inevitably makes diagnosis even more difficult. In this work, a network to propose lung-aware regions of interest is utilized to make the image classification task focus only on the lung area and discard other meaningless regions in the image.

The process of generating region proposals, depicted in Fig. 1, is performed at the beginning of the proposed pipeline. It is presented in Fig. 2, and it has three main steps. First, a U-Net segmentation network [18] is applied to the input CXR image to extract the lung mask. Second, the minimum rectangle which encloses the lung masks is obtained and enlarged by 5% in all dimensions to avoid losing any pixel on the lung border. Given that the lower part of the lungs can often be difficult to see, especially in severe cases, the box is further enlarged on the lower side to a maximum height/width ratio of 1. Third, a set of  $Q$  rectangular regions of interest is randomly generated within the limits of the box.



**Fig. 2.** L-RPN pipeline to focus on the pulmonary zone only.

### 3.2 Support set ensemble

When the data available to train a machine learning model is scarce, an effective way to improve performance is by generating new synthetic training data by applying random transformations. This image augmentation can be used during both training and inference. In the first case, the variability of training data is increased, and so the model is able to learn richer image features. In the second case, different versions of a given image can be used to obtain different predictions, which can be combined for a more robust prediction.

In this work, an inference-time augmentation is performed, so  $S$  support sets are generated and kept static throughout the test step. In this regard, a given query image can be compared with all the support sets to obtain the similarity vectors  $a_i^s$ , being  $i = \{1, \dots, N\}$  the category index, and  $s = \{1, \dots, S\}$  the index of the corresponding support set. Then, the similarity vectors are aggregated so that the chances that an image has to be assigned to a specific category can be obtained by a voting formula:

$$p_i = \prod_{s=1}^S \frac{e^{a_i^s}}{\sum_{j=1}^N e^{a_j^s}} \quad (1)$$

where  $p_i$  is the probability of belonging to the  $i$ -th class, and  $a_i^s$  is the similarity value obtained for the  $i$ -th class through the  $s$ -th support set. Values of  $p_i$  are normalized to ensure that  $\sum_{i=0}^{|C|} p_i = 1$ , being  $C$  the set of categories.

### 3.3 Misdiagnosis-sensitive learning

The common clinical approach for COVID-19 triaging is the same as in other diseases. First, a cheap test is used in a large sample of the population. As these

tests are characterized by a high number of false positives —negative patients tested positive—, a more accurate and expensive test is used to refine the diagnostic of the patients tested positive.

The aim of this work is to develop a binary classifier to differentiate between COVID-19 positive and COVID-19 negative CXR images. To make the model robust enough, we used the extra categories provided by the COVIDGR-1.0 database. In particular, the images in this dataset are categorized into five classes: negative patients without visible lung abnormalities and positive patients with no, mild, moderate, and severe visual affection. In the following, we will refer to these classes as N\_NORMAL, P\_NORMAL, P\_MILD, P\_MODERATE and P\_SEVERE.

With the main goal of enhancing the current triaging methods, the followed approach in this work is targeted at reducing the number of positive patients tested negative. To do so, the cost function proposed to train the model is provided with a novel penalty term which is based not only on the distance between classes in terms of visual affection, but also on the clinical implication of each misclassification. Thus, the penalty term follows the principle that the penalty is supposed to be maximum when a severe patient is classified as negative, as the clinical cost is very high. In contrast, a positive patient with moderate affection should lead to a low penalty when he/she is classified as severe, as the clinical cost is negligible and the visual affection can be very similar in both classes.

The addition of this cross-penalties is carried out through a cost matrix  $M$  which indicates the penalty of every possible misclassification scenario. This matrix, developed by an expert radiologist, is shown in Table 1, and is included in the modified cross-entropy equation (2). Each matrix cell  $M_{ij}$  denotes the penalty of a patient of the  $i$ -th class when is predicted to belong to the  $j$ -th class. As mentioned above, the highest penalties are associated with severe and moderate patients which are assigned to the negative class. It should be noted that the P\_NORMAL class is not included in the cost matrix  $M$  because, as will be explained in Section 4.1, this class is not used in the training process.

**Table 1.** Cost matrix  $M$  included in the new loss function in equation (2).

		<b>Predicted</b>			
		<b>Negative</b>	<b>Mild</b>	<b>Moderate</b>	<b>Severe</b>
<b>Real</b>	<b>Negative</b>	0	0.1	0.2	0.3
	<b>Mild</b>	0.3	0	0.1	0.2
	<b>Moderate</b>	0.4	0.1	0	0.1
	<b>Severe</b>	0.5	0.075	0.025	0

The cost of making a prediction on a patient belonging to the  $i$ -th class is calculated via

$$\mathcal{L}_i = -\log(p_i) + \sum_{j=0}^{|C'|} [-\log(1 - p_j)M_{ij}] \quad , \quad C' = C \setminus \{C_i\} \quad (2)$$

where  $C'$  is the set consisting of all classes except the class  $C_i$ .

## 4 Experiments

We performed all experiments in the COVIDGR-1.0 dataset [22] to compare the results of our method with previous state-of-the-art approaches for automatic COVID-19 diagnosis with CXR images. This dataset provides a balanced set with 426 positive samples and 426 negative samples. It includes the level of severity for positive patients, including RT-PCR positive cases identified as negatives by expert radiologists.

We report both the overall classification performance and the detailed results for each level of severity. We follow the same experimental setting established by [22], calculating the mean and standard deviation for five 5-fold cross-validation. For a fairer comparison, the same data splits were used for all the methods.

### 4.1 Implementation details

The L-RPN module described in subsection 3.1 follows a U-net model trained on the Montgomery and SCR datasets [8,20]. The number of regions extracted from the lung area  $Q$  is set to 9. All input images are resized to  $128 \times 128$  pixels.

The backbone of the few-shot classifier is implemented as ResNet-12 pre-trained on the miniImageNet dataset [24]. The support set contains 10 randomly selected images for each training episode—each training iteration of the meta-learner—, while we select  $S = 5$  static support sets with 10 images each for validation and test. These support images for validation and test are extracted from the training set, performing inference on the complete test set. Positive patients with no radiological findings—category P\_NORMAL in COVIDGR-1.0—are not considered for training, as their strong similarity to negative samples could hinder the training process. However, the test set must include images from this category to maintain a more realistic evaluation setting.

The initial learning rate is  $0.5 \times 10^{-3}$ , applying a reduction factor of 0.5 after each 500 iterations. The maximum number of training iterations is set to 5,000 with a validation stage every 50 iterations. Then, the best performing model on the validation set is selected as the final classifier. As the goal is to implement a triage system, both the global classification accuracy and the sensitivity are taken into account to establish the validation iteration. Therefore, the validation performance  $VP$  is defined as:

$$VP = \alpha * sensitivity + \beta * accuracy \quad (3)$$

where  $\alpha$  and  $\beta$  are hyperparameters heuristically set to 0.3 and 0.7 respectively.

### 4.2 Results

We conducted a series of experiments and compared our model with the state-of-the-art using binary classification metrics. These include accuracy, specificity,

sensitivity, precision, and F1 (the last two are reported for both positive and negative classes). Accuracy measures the global classification performance by means of the percentage of images correctly classified, and so it is a useful metric to compare different approaches. Sensitivity is essential for detecting positive cases in a triage system, but a maximum sensitivity could be obtained if every patient is classified as positive, which in the end would make no sense. Therefore, a good balance between sensitivity and specificity is needed to minimize false positives. The F1 considers this balance, utilizing recall and precision metrics for each class.

Table 2 shows a detailed comparison of our method with state-of-the-art approaches assessed on the COVIDGR-1.0 dataset [22]. The reported results comprise a set of methods developed and/or tested by other authors, including a general-purpose classifier —ResNet-50— and methodologies specifically designed for COVID-19 CXR images. The results in this table do not include the P\_NORMAL class for a better comparability, as no previous study uses this category in their experiments. Overall, our method achieves the best results in all metric except for specificity and precision in the positive class. However, the best-performing method in these metrics —COVIDNet-CXR— achieves very poor results in the other metrics, lacking more than 37% behind our method in terms of sensitivity. In fact, our approach outperforms the second-best result by around 11 points in this key metric for any triage system. Also, regarding the overall accuracy, we set a new state-of-the-art result on this dataset, improving previous methods by 2.9% accuracy.

**Table 2.** Results on COVIDGR-1.0.

	Specificity	N_Precision	N_F1	Sensitivity	P_Precision	P_F1	Accuracy
COVIDNet [25]	<b>88.8±0.9</b>	3.4±6.2	73.3±3.8	46.8±17.6	<b>81.7±6.0</b>	56.9±15.1	67.8±6.1
CAPS [1]	65.7±9.9	65.6±4.0	65.2±5.0	64.9±9.7	66.1±4.5	64.9±4.9	65.3±3.3
ResNet-50 [22]	79.9±8.9	71.9±3.1	75.4±4.9	68.6±6.1	78.8±6.3	72.7±3.5	74.3±3.6
FuCiTNet [16]	80.8±7.0	72.0±4.5	75.8±3.2	67.9±8.6	78.5±5.0	72.4±4.8	74.4±3.3
SDNet [22]	79.8±6.2	74.7±3.9	76.9±2.8	72.6±6.8	78.7±4.7	75.7±3.4	76.2±2.7
ours	75.1±6.5	<b>85.3±3.7</b>	<b>79.7±3.9</b>	<b>84.0±4.9</b>	73.8±4.9	<b>78.4±3.1</b>	<b>79.1±3.4</b>

Moreover, we compare our method with the best previous approach —COVID-SDNet— for all severity levels in COVIDGR-1.0, including the P\_NORMAL category. As seen in Table 3, our method achieves the best results in every subset, achieving the largest margins in the most demanding subsets. Thus, COVID-SDNet underperforms a random classifier in the P\_MILD subset, tending to classify these patients as positive cases. Our approach is capable of increasing the classification accuracy in this demanding subset by more than 16%. This proves that our method is more robust against images with minimal radiological findings.

**Table 3.** Obtained accuracy by severity level in COVIDGR-1.0.

	COVID-SDNet[22]	ours
P_NORMAL	-	41.6±13.1
P_MILD	46.0±7.1	62.4±9.9
P_MODERATE	85.4±1.9	89.4±6.2
P_SEVERE	97.2±1.9	99.5±1.7

## 5 Conclusions

We have defined a new few-shot classifier for CXR images, able to correctly identify positive COVID-19 cases. Although the proposed method follows a few-shot architecture, the combination of support sets leverages all the information available, boosting the performance. Furthermore, the L-RPN module ensures that the diagnosis is made only from lung information, which makes the model robust against poorly framed CXR images due to different patient positions or the use of external monitoring devices. Experimental results show that our method scores the best result in the COVIDGR data set, improving the binary classification of COVID-19 cases by around 3 accuracy points. Regarding the four severity levels defined in COVIDGR, our method outperforms the second best approach in all cases, especially in the challenging subset that contains patients with mild condition with an improvement of more than 16%.

**Acknowledgements** This work has received financial support from the Spanish Ministry of Science and Innovation under grant PID2020-112623GB-I00, Consellería de Cultura, Educación e Ordenación Universitaria under grants ED431C 2021/48, ED431G-2019/04, ED481A-2018 and ED431C 2018/29 and the European Regional Development Fund (ERDF), which acknowledges the CiTIUS-Research Center on Intelligent Technologies of the University of Santiago de Compostela as a Research Center of the Galician University System.

## References

1. Afshar, P., Heidarian, S., Naderkhani, F., Oikonomou, A., Plataniotis, K.N., Mohammadi, A.: COVID-CAPS: A capsule network-based framework for identification of COVID-19 cases from X-ray images. *Pattern Recogn Lett* **138**, 638–43 (2020)
2. Chowdhury, M.E., Rahman, T., Khandakar, A., Mazhar, R., Kadir, M.A., Mahbub, Z.B., Islam, K.R., Khan, M.S., Iqbal, A., Al Emadi, N., et al.: Can AI help in screening viral and COVID-19 pneumonia? *IEEE Access* **8**, 132665–76 (2020)
3. Cohen, J.P., Morrison, P., Dao, L., Roth, K., Duong, T.Q., Ghassemi, M.: COVID-19 image data collection: Prospective predictions are the future. *arXiv 2006.11988* (2020), <https://github.com/ieee8023/covid-chestxray-dataset>
4. Cui, H., Wei, D., Ma, K., Gu, S., Zheng, Y.: A unified framework for generalized low-shot medical image segmentation with scarce data. *IEEE T Med Imaging* (2020)

5. Fernández-Miranda, P.M., Bellón, P.S., Del Barrio, A.P., Iglesias, L.L., García, P.S., Aguilar-Gómez, F., González, D.R., Vega, J.A.: Developing a training web application for improving the COVID-19 diagnostic accuracy on chest X-ray. *J Digit Imaging* pp. 1–15 (2021)
6. Huang, C., Wang, Y., Li, X., Ren, L., Zhao, J., Hu, Y., Zhang, L., Fan, G., Xu, J., Gu, X., et al.: Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet* **395**(10223), 497–506 (2020)
7. Jadon, S.: COVID-19 detection from scarce chest X-ray image data using few-shot deep learning approach. In: *Proc SPIE*. vol. 11601, p. 116010X (2021)
8. Jaeger, S., Candemir, S., Antani, S., Wang, Y.X.J., Lu, P.X., Thoma, G.: Two public chest X-ray datasets for computer-aided screening of pulmonary diseases. *Quant Imag Med Surg* **4**(6), 475 (2014)
9. Jin, K.N., Do, K.H., Da Nam, B., Hwang, S.H., Choi, M., Yong, H.S.: Korean clinical imaging guidelines for justification of diagnostic imaging study for COVID-19. *J Korean Soc Radiol* **83**(2), 265–283 (2022)
10. Kang, H., Xia, L., Yan, F., Wan, Z., Shi, F., Yuan, H., Jiang, H., Wu, D., Sui, H., Zhang, C., et al.: Diagnosis of coronavirus disease 2019 (COVID-19) with structured latent multi-view representation learning. *IEEE T Med Imaging* **39**(8), 2606–2614 (2020)
11. Khan, A.I., Shah, J.L., Bhat, M.M.: CoroNet: A deep neural network for detection and diagnosis of COVID-19 from chest X-ray images. *Comput Meth Prog Bio* **196**, 105581 (2020)
12. Loey, M., Smarandache, F., M Khalifa, N.E.: Within the lack of chest COVID-19 X-ray dataset: a novel detection model based on GAN and deep transfer learning. *Symmetry* **12**(4), 651 (2020)
13. Nishio, M., Noguchi, S., Matsuo, H., Murakami, T.: Automatic classification between COVID-19 pneumonia, non-COVID-19 pneumonia, and the healthy on chest X-ray image: combination of data augmentation methods. *Sci Rep* **10**(1), 1–6 (2020)
14. Oh, Y., Park, S., Ye, J.C.: Deep learning COVID-19 features on CXR using limited training data sets. *IEEE T Med Imaging* **39**(8), 2688–700 (2020)
15. Ouchicha, C., Ammor, O., Meknassi, M.: CVDNet: A novel deep learning architecture for detection of coronavirus (COVID-19) from chest X-ray images. *Chaos Solitons Fractals* **140**, 110245 (2020)
16. Rey-Area, M., Guirado, E., Tabik, S., Ruiz-Hidalgo, J.: FuCiTNet: Improving the generalization of deep learning networks by the fusion of learned class-inherent transformations. *Inform Fusion* **63**, 188–95 (2020)
17. Roberts, M., Driggs, D., Thorpe, M., Gilbey, J., Yeung, M., Ursprung, S., Aviles-Rivero, A.I., Etmann, C., McCague, C., Beer, L., et al.: Common pitfalls and recommendations for using machine learning to detect and prognosticate for COVID-19 using chest radiographs and CT scans. *Nat Mach Intell* **3**(3), 199–217 (2021)
18. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: *Proc MICCAI*. pp. 234–41 (2015)
19. Rubner, Y., Tomasi, C., Guibas, L.J.: The earth mover’s distance as a metric for image retrieval. *Int J Comput Vision* **40**(2), 99–121 (2000)
20. Shiraishi, J., Katsuragawa, S., Ikezoe, J., Matsumoto, T., Kobayashi, T., Komatsu, K.i., Matsui, M., Fujita, H., Kodera, Y., Doi, K.: Development of a digital image database for chest radiographs with and without a lung nodule: receiver operating characteristic analysis of radiologists’ detection of pulmonary nodules. *Am J Roentgenol* **174**(1), 71–74 (2000)

21. Shorfuzzaman, M., Hossain, M.S.: MetaCOVID: A siamese neural network framework with contrastive loss for n-shot diagnosis of COVID-19 patients. *Pattern Recogn* **113**, 107700 (2021)
22. Tabik, S., Gómez-Ríos, A., Martín-Rodríguez, J.L., Sevillano-García, I., Rey-Area, M., Charte, D., Guirado, E., Suárez, J.L., Luengo, J., Valero-González, M., et al.: COVIDGR dataset and COVID-SDNet methodology for predicting COVID-19 based on chest X-ray images. *IEEE J Biomed Health* **24**(12), 3595–605 (2020)
23. Teixeira, L.O., Pereira, R.M., Bertolini, D., Oliveira, L.S., Nanni, L., Cavalcanti, G.D., Costa, Y.M.: Impact of lung segmentation on the diagnosis and explanation of COVID-19 in chest X-ray images. *Sensors* **21**(21), 7116 (2021)
24. Vinyals, O., Blundell, C., Lillicrap, T., Wierstra, D., et al.: Matching networks for one shot learning. *Adv Neur In* **29**, 3630–8 (2016)
25. Wang, L., Lin, Z.Q., Wong, A.: COVID-Net: A tailored deep convolutional neural network design for detection of COVID-19 cases from chest X-ray images. *Sci Rep* **10**(1), 1–12 (2020)
26. Wang, N., Liu, H., Xu, C.: Deep learning for the detection of COVID-19 using transfer learning and model integration. In: *IEEE Int Conf Electr.* pp. 281–4. IEEE (2020)
27. Wong, H.Y.F., Lam, H.Y.S., Fong, A.H.T., Leung, S.T., Chin, T.W.Y., Lo, C.S.Y., Lui, M.M.S., Lee, J.C.Y., Chiu, K.W.H., Chung, T., et al.: Frequency and distribution of chest radiographic findings in COVID-19 positive patients. *Radiology* p. 201160 (2020)