Web-based efficient dual attention networks to detect COVID-19 from X-ray images

Md. Mostafa Kamal Sarker✉, Yasmine Makhlouf, Syeda Furruka Banu, Sylvie Chambon, Petia Radeva and Domenec Puig

Rapid and accurate detection of COVID-19 is a crucial step to control the virus. For this purpose, the authors designed a web-based COVID-19 detector using efficient dual attention networks, called 'EDANet'. The EDANet architecture is based on inverted residual structures to reduce the model complexity and dual attention mechanism with position and channel attention blocks to enhance the discriminant features from the different layers of the network. Although the EDANet has only 4.1 million parameters, the experimental results demonstrate that it achieves the state-of-the-art results on the COVIDx data set in terms of accuracy and sensitivity of 96 and 94%. The web application is available at the following link: https://covid19detector-cxr.herokuapp.com/.

Introduction: Recently, a crucial challenge of human life has begun with the COVID-19 pandemic which is caused by the new type of coronavirus, the SARS-CoV-2. An important step to slow down this virus is rapid detection and isolation of infected people. However, the gold standard identification methods of COVID-19 is real-time reverse-transcriptase polymerase chain reaction [1] which is time-taking and not commonly available due to the pandemic. Thus, an alternative screening method required for analyzing COVID-19 infections in chest radiography imaging (e.g. X-ray or computed tomography (CT) imaging) [2]. In this case, radiologists need to examine the chest radiography for finding the visual indicators related to COVID-19 viral infection. In [2], the authors explain about the consistent patterns of CT and chest X-ray (CXR) images of COVID-19 patient. Nevertheless, they state that humans are incapable to distinguish these patterns from CT and CXR images for the most asymptomatic patients. Several researchers have been trying to find different quick screening solutions. In this regard, deep learning has been giving promising outcomes using CXR images to identify the COVID-19. Deep learning has already been widely investigated for the identification of pneumonia and other diseases on CXR images. A ten-layer convolutional neural network (CNN) is proposed in [3] to analyse seven different interstitial lung diseases. In their work, the total of 120 CT scans is split into 14,696 image patches that are used to classify every patch into every pattern. Their method has achieved an accuracy of 85%. In [4] another CNN with a branch for predicting a segmentation mask is used to identify pneumonia (negative and positive) from the CXR images and create a bounding box around the pneumonia positive lung opacities. In the detection of COVID-19, a comparative study between seven distinct famous deep learning CNN architectures is presented in [5]. The data set of only 50 images (25 healthy and 25 COVID-19 infected patients) are used in their experiment and obtained good performance of f1-scores of 89 and 91% with the dense convolutional network for normal and COVID-19, respectively. Recently, in COVID-net [6], a new CNN architecture and an open-access benchmark data set, COVIDx, is introduced for COVID-19 detection from CXR images. The COVID-net yields an accuracy of 90.3% and sensitivity of 91.0% for the class of the COVID-19. However, all these recent deep learning-based models for identifying COVID-19 infection from the CXR images are with heavy computational cost (the number of model parameters is very high). On the other hand, their implementation details are not publicly available. Instead, COVID-net [6] make their source code and data set publicly available. Therefore, we claim to design a lightweight model with a low computational complexity that deploys with a web-based publicly available application to help clinicians for identifying COVID-19. Moreover, up to our knowledge, there is no other web-based application for detecting COVID-19 from CXR images. This is the first web-based application for detecting COVID-19 from CXR images with the very lightweight deep model. It is easily accessible through any devices (e.g. personal computer, mobile phones, embedded devices etc.) for all clinicians around the world.

Methodology: In this section, we explain the construction of the proposed efficient dual attention network (EDANet) model and the loss function used to train the model. Fig. 1 illustrates the architecture of the proposed 'EDANet' model. It consists of two convolutional (Conv) layers, 16 mobile inverted bottleneck convolutional (MBConv) layers, and four dual attention (DUA) modules. We convert the input image of the network from three to one channel (originally CXR images are grey images and contain their information in one channel) in order to avoid the complexity of feeding three-channel input to the initial Conv layer for the feature extraction. The initial Conv layer extracts low-level feature maps using a 3×3 kernel size with stride 2. Afterwards, a sequence of one MBConv1 with 3×3, two MBConv6 with 3×3, two MBConv6 with 5×5, three MBConv6 with 3×3, seven MBConv6 with 5×5 and one MBConv6 with 3×3 are used to extract low-level to high-level feature maps to classify the input CXR images. The MBConv layers are constructed by the concept of compound model scaling that was initially proposed in EfficientNet [7]. Finally, another sequence consisting of a Conv layer with 1x1 convolution kernel, an adaptive average pooling layer, a linear fully connected (FC) and a softmax layer is used. The FC layer maps the last feature map channels into the number of classes, and softmax layer generates class probabilities for every class.

Fig. 1 *Architecture of the proposed EDANet*

To enhance the discriminant ability in between each class feature representations, we introduced another important module in our architecture, which is the DUA mechanism. We integrate four DUA modules after layer 5, 10, 13, and 17, respectively, where the feature maps are down-sampled in the network. The DUA modules are the combination of position and channel attention block (CAB) inspired by [8]. Fig. 2 shows the construction of the DUA module.

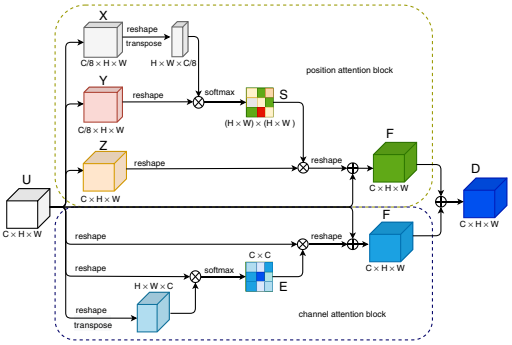

Fig. 2 *Architecture of the proposed DUA module*

The position attention block (PAB) can yield a global contextual description and selectively aggregate the context according to a spatial attention map by creating relevant semantic features that can produce mutual gains and enhance the intra-class semantic consistency. On the other hand, CAB can highlight class-dependent feature maps and discriminatively supports a feature boost that can not be generated by the convolution layers. Therefore, the combination of these two attention mechanisms can enhance the feature representation of intra-class variations between channel maps.

To illustrate better the workflow of the proposed DUA module, let us consider the input feature $U \in \mathbb{R}^{C \times H \times W}$, where C, H and W are channel, height and width dimensions, respectively, shown in Fig. 2. In the upper PAB section, U is fed into three convolution layers to produce new feature maps X, Y and Z, respectively. The generated feature maps from the first two convolutional blocks are X^p, $Y^p \in \mathbb{R}^{C' \times H \times W}$, where the channel C' is equal to $C/8$ (superscript p is for PAB). Afterwards, X^p and Y^p are reshaped into $(H \times W) \times C'$. Finally, a matrix multiplication between the transpose of $\boldsymbol{Y^p}$ and $\boldsymbol{X^p}$ is performed, and a softmax function is used to generate the spatial attention map $S^p \in \mathbb{R}^{(H \times W) \times (H \times W)}$

$$s^p_{i,j} = \frac{\exp\left(Xi^p \cdot Y\,j^p\right)}{\sum_{i=1}^{H \times W} \exp\left(Xi^p \cdot Y\,j^p\right)}, \tag{1}$$

where $s_{i,j}$ denotes the ith position's contact on jth position. Thus, we can assume that a softmax function S^p tries to determine the correlation between two spatial positions in the input feature maps. Consequently, the output of the third convolutional block Z is $Z^p \in \mathbb{R}^{C \times (H \times W)}$ that has

the same shape of the input feature map U. Afterwards, Z^p is reshaped and multiplied by a permuted form of the spatial attention map S^p and reshaped the output to a $\mathbb{R}^{C\times(H\times W)}$. The final PAB feature map F can be estimated as

$$F_{\mathrm{PAB},\,j} = \alpha_p \sum_{i=1}^{H\times W} s_{ij}^{p} Z j^{p} + U_j, \tag{2}$$

where α_p is determined as 0 as described in [8]. Based on (2), the resulting feature $\boldsymbol{F}$ at each position is a weighted sum of the features of the complete neighbours of original features. Thus, it is apparent that the PAB can yield a global contextual representation and selectively aggregate the context according to a spatial attention map by creating related semantic features that can produce mutual gains and enhance the intra-class semantic consistency.

In the lower CAB section, the input feature map $U \in \mathbb{R}^{C\times H\times W}$ is reshaped in the first two parts of the CAB, and permuted in the second part, leading to $U_0^c \in \mathbb{R}^{(H\times W)\times C}$ and $U_1^c \in \mathbb{R}^{C\times(H\times W)}$, respectively (superscript c is for CAB). Then, a matrix multiplication between U_0^c and U_1^c is performed, and the channel attention map $E^c \in \mathbb{R}^{C\times C}$ is defined as

$$e_{i,\,j}^{c} = \frac{\exp\left(U_{0,\,i}^{c} \cdot U_{1,\,j}^{c}\right)}{\sum_{i=1}^{C} \exp\left(U_{0,\,i}^{c} \cdot U_{1,\,j}^{c}\right)}, \tag{3}$$

where the result of the ith channel on the jth is yielded by $e_{i,\,j}^{c}$. This is then multiplied by a transposed version of the input U, i.e. U_2^c, whose outcome is reshaped to $\mathbb{R}^{C\times(H\times W)}$. Besides, the final channel attention map is obtained as

$$F_{\mathrm{CAB},\,j} = \alpha_{\mathrm{c}} \sum_{i=1}^{C} e_{ij}^{c} U_{2,\,j}^{c} + U_j, \tag{4}$$

where α_{c} measures the weight of the channel attention map over the input feature map U. Likewise to α_p, α_c is originally set to 0 and gradually learned. This process sums weighted versions of the features of all the channels into the initial features, emphasising class-dependent feature maps and boosting feature discrimination between classes. Finally, at the end of both attention blocks, the new generated features are performing an element-wise sum operation to generate the final DUA features as follows:

$$D_{\mathrm{DUA},\,j} = F_{\mathrm{PAB},\,j} + F_{\mathrm{CAB},\,j}. \tag{5}$$

Moreover, selecting a proper loss function is very important to train the deep models since the COVIDx data set [6] is imbalanced. Therefore, we used class-balanced (CB) focal loss (FL) function [9] to train our proposed model. In [9], the CB loss is presenting a weighting factor to solve the difficulty of training deep networks with imbalanced data. On the other hand, the FL [10] combines a scaling factor to the sigmoid cross-entropy loss to decrease the corresponding loss for correctly classified examples and focus on difficult examples. For an input image $\boldsymbol{x}$ with ground-truth $y \in \{1, 2, \ldots, C\}$, where C is the total number of classes, let the model calculated class probabilities $\boldsymbol{p} = [p_1, p_2, \ldots, p_C]^{\top}$, where $p_i \in [0, 1]\ \forall\ i$, denote $p_i^t = \mathrm{sigmoid}(x_i^t) = 1/(1 + \exp(-x_i^t))$, the FL is defined as follows:

$$\mathrm{FL}(\boldsymbol{x}, y) = -\sum_{i=1}^{C} (1 - p_i^t)^{\gamma}\ \log(p_i^t). \tag{6}$$

The final CB FL is described as follows:

$$\mathrm{CB}_{\mathrm{focal}}(\boldsymbol{x}, y) = -\frac{1-\beta}{1-\beta^{n_y}} \sum_{i=1}^{C} (1 - p_i^t)^{\gamma}\ \log(p_i^t), \tag{7}$$

where n_y is the number of images in the ground-truth class y and $(1-\beta)/(1-\beta^{n_y})$ weighting factor of the loss function with hyperparameter $\beta \in [0, 1]$ and $\gamma \in [0.5, 2]$.

Experimental results: The efficacy of the proposed model, EDANet, is assessed on COVIDx data set [6], a benchmark data set for detecting COVID-19 infection from CXR images. The data set has three different classes based on the infection type (i) normal (no infection), (ii) non-COVID19 (pneumonia) and (iii) COVID-19. It is divided into train and test sets. The number of images in the train set is 279, 5451 and 7966 for COVID-19, pneumonia and normal cases, respectively, and every class has 100 images for the test set.

For the evaluation metrics, we calculated the test accuracy, sensitivity and positive predictive value (PPV) to compare the performance of the proposed model with the COVID-net and other models in [6]. However, the test accuracy also compared with the architectural complexity (number of model parameters) and computational complexity (number of model multiply-accumulate (MAC) operations) is shown in Table 1.

Table 1: Evaluation of the tested deep models on COVIDx [6] test data set

Models	Params (M)	MACs (G)	ACC
VGG-19	20.37	89.63	83.0
ResNet-50	24.97	17.75	90.6
COVID-Net	11.75	7.50	93.5
proposed (EDANet)	**4.20**	**0.03**	**96.0**

Best outcomes presented in bold.

The proposed model is implemented on the PyTorch framework [11]. Besides, the stochastic gradient descent [12] optimiser with the momentum of 0.9 and weight decay of 0.0005 is used. The learning rate is set to 0.0001, the batch size is 32 and the number of total training epochs is 100. We used a step learning policy to reduce the learning rate after every 30 epochs. The weighting factors of CB FL $\beta = 0.9999$ and $\gamma = 0.5$ is used to train the model. Although, all the existing models are using data augmentation to fix the class imbalance issue for identifying COVID-19 infection on COVIDx data set. Notice that no data augmentation method is applied to train our network. We solve the data set class imbalance problem with only using the CB FL.

The quantitative analysis of the proposed EDANet is shown in Tables 1–3. In Table 1, we compared the EDANet with three different deep models VGG-19, ResNet-50 and COVID-Net in terms of test accuracy, model complexity (number of parameters) and computational complexity (number of MAC operations). It shows that EDANet outperforms the all tested models on the COVIDx [6] test data set. EDANet yields the accuracy of 96.00% which is 13.0, 5.4 and 2.5% higher than VGG-19, ResNet-50 and COVID-Net [6], respectively. Moreover, EDANet has only 4.20 millions of parameters with 0.03 MACs(G) which is significantly faster and less complex compared with the other models shown in Table 1. It has $5\times$, $6\times$, $3\times$ lower parameters than the VGG-19, ResNet-50 and COVID-Net models, respectively.

Table 2: Sensitivity for every infection class

Sensitivity, %			
Models	Normal	Non-COVID19	COVID-19
VGG-19	**98.0**	90.0	58.7
ResNet-50	97.0	92.0	83.0
COVID-Net	95.0	94.0	91.0
proposed (EDANet)	**98.0**	**96.0**	**94.0**

Best outcomes presented in bold.

Table 3: PPV for every infection class

PPV, %			
Models	Normal	Non-COVID19	COVID-19
VGG-19	83.1	75.0	98.4
ResNet-50	88.2	86.8	98.8
COVID-Net	90.5	91.3	**98.9**
proposed (EDANet)	**96.0**	**96.1**	95.9

Best outcomes presented in bold.

In Table 2, the sensitivity of EDANet is compared to VGG-19, ResNet-50 and COVID-Net [6] for every infection class. The proposed EDANet achieved a sensitivity of 98, 96 and 94% for normal, non-COVID19 and COVID-19 classes, respectively. The proposed model obtains 1 and 3% increment of sensitivity compared with ResNet-50, and COVID-Net, but is comparative to VGG-19 for the normal class. Consequently, it yields an improvement of 6, 4 and 2% of sensitivity compared with VGG-19, ResNet-50 and COVID-Net, respectively, in the Non-COVID19 class. Moreover, it accomplished 35.3, 11 and 3% higher sensitivity than VGG-19, ResNet-50 and COVID-Net, respectively, in the COVID-19 class.

In Table 3, the PPV of EDANet is compared to VGG-19, ResNet-50 and COVID-Net, [6]. The proposed EDANet yields 96, 96.1 and 95.9% PPV in normal, non-COVID19 and COVID-19 class, respectively. The EDANet achieved 12.9, 7.8, 5.5 and 21.1, 9.3, 4.8% higher PPV compared with VGG-19, ResNet-50 and COVID-Net in normal and

non-COVID19 class. However, it shows 3% lower PPV compared to COVID-Net in COVID-19. Finally, all the above performance evaluation demonstrates that our proposed model brings an important contribution not only in terms of computational complexity and number of parameters but also in its competitive performance compared to the state of the art architectures.

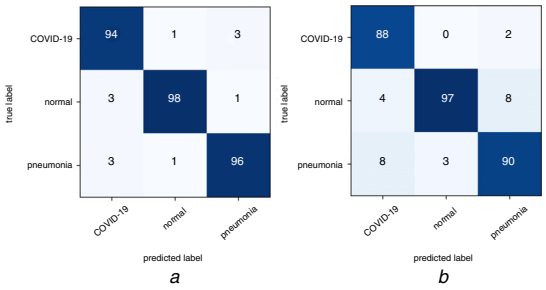

Fig. 3 *Confusion matrix*
a EDANet
b EDANet–DUA module

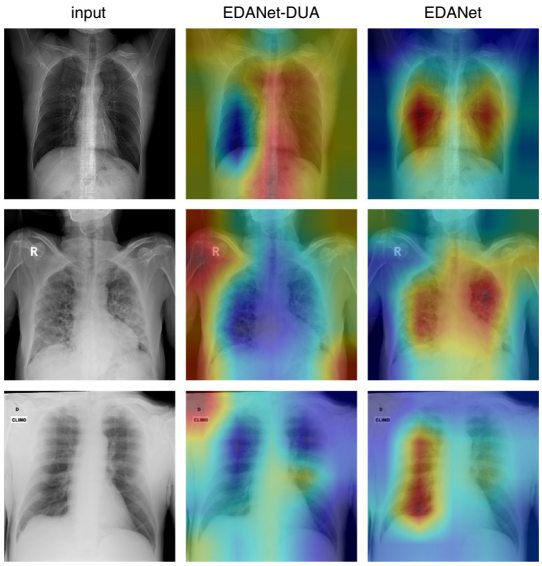

Fig. 4 *Visualisation results of the activation maps. For every row, we show an input image, the corresponding activation maps from the outputs of EDANet-DUA and the EDANet*

We assessed the effects of the DUA module by illustrating the confusion matrix of EDANet and EDANet-DUA modules shown in Fig. 3. The confusion matrix in Fig. 3*a* shows that the proposed model with the DUA module (EDANet) can correctly classify most of the images in all classes. However, it misclassifies 3% COVID-19 to pneumonia and only 1% COVID-19 to the normal class. On the other hand, the confusion matrix in Fig. 3(b) shows the proposed model without DUA module (EDANet-DUA module) can not correctly classify images in all classes. It misclassifies 8% pneumonia to COVID-19 that means it can not distinguish the features of COVID-19 and pneumonia class, separately. It can easily confuse between the COVID-19 and pneumonia because of the features similarities of these classes. Finally, it is proven that by adding the DUA module the proposed EDANet can differentiate the class-wise feature representation and yields the best performance without increasing the model complexity.

The performance analysis over different ablation studies is probably not good enough to judge the advantages and the behaviour of the proposed model. However, the proposed model shows performance improvement in the results with our proposed DUA module, it is interesting to investigate how it works. Thus, we visualise the activation maps with and without the DUA module shown in Fig. 4. In Fig. 4, the EDANet-DUA column shows the activation maps without the DUA modules where the model can classify all these images correctly to the COVID-19 class but activated in different regions of the input CXR images. More preciously, the COVID-19 infection symptom can be conformed through some opacity (white spots) on the CXR image. The activation maps by the EDANet (with DUA module) can remarkably overlay with opacity regions, which could signify the presence of COVID-19. On the other hand, the activation maps by EDANet-DUA can not focus on the opacity area, it considered the other body parts features as a COVID-19 indication, which is not correct. Finally, we can conclude it that by adding the DUA module, the model can able to differentiate the relevant and non-relevant features and learn the proper features related to the individual class.

Conclusion: In this Letter, a lightweight and efficient web-based model, named EDANet, has been proposed for identifying COVID-19 infection from the CXR images. EDANet consists of MBConv layers and DUA modules. The proposed architecture, tested on COVIDx data set, yields precise classification results with accuracy, sensitivity and PPV of 96, 94 and 95%, respectively. Compared to the existing COVID-19 detection models, it stands out by being significantly less complex and faster with only 4.2 million parameters. Future work would consider including more CXR images with additional text features (e.g. COVID-19 symptom), and implement the model for clinical trials to detect COVID-19 from the CXR images.

Submitted: *03 July 2020*
doi: 10.1049/el.2020.1962
One or more of the Figures in this Letter are available in colour online.

Md. Mostafa Kamal Sarker and Petia Radeva (*Department of Mathematics and Computer Science, University of Barcelona, 08007 Barcelona, Spain*)

✉ E-mail: mdmostafakamalsarker@ub.edu

Yasmine Makhlouf (*School of Medicine, Dentistry and Biomedical Sciences, Queen's University Belfast, Belfast BT9, UK*)

Sylvie Chambon (*Institut de Recherche en Informatique de Toulouse, ENSEEIHT-INP, Toulouse 31000, France*)

Syeda Furruka Banu and Domenec Puig (*Department of Computer Engineering and Mathematics, Universitat Rovira i Virgili, Tarragona 43007, Spain*)

References

1 Wang, W., Xu, Y., Gao, R., *et al.*: 'Detection of sars-cov-2 in different types of clinical specimens', *JAMA*, 2020, **323**, (18), pp. 1843–1844
2 Ng, M.Y., Lee, E.Y., Yang, J., *et al.*: 'Imaging profile of the covid-19 infection: radiologic findings and literature review', *Radiol., Cardiothorac. Imag.*, 2020, **2**, (1), p. e200034
3 Anthimopoulos, M., Christodoulidis, S., Ebner, L., *et al.*: 'Lung pattern classification for interstitial lung diseases using a deep convolutional neural network', *IEEE Trans. Medic. Imag.*, 2016, **35**, (5), pp. 1207–1216
4 Jaiswal, A.K., Tiwari, P., Kumar, S., *et al.*: 'Identifying pneumonia in chest x-rays: a deep learning approach', *Measurement*, 2019, **145**, pp. 511–518
5 Hemdan, E.E.D., Shouman, M.A., and Karar, M.E.: 'Covidx-net: a framework of deep learning classifiers to diagnose covid-19 in x-ray images'. arXiv preprint arXiv:200311055, 2020
6 Wang, L., and Wong, A.: 'Covid-net: a tailored deep convolutional neural network design for detection of covid-19 cases from chest radiography images'. arXiv preprint arXiv:200309871, 2020
7 Tan, M., and Le, Q.V.: 'Efficientnet: rethinking model scaling for convolutional neural networks'. arXiv preprint arXiv:190511946, 2019
8 Fu, J., Liu, J., Tian, H., *et al.*: 'Dual attention network for scene segmentation'. Proc. IEEE Conf. on Computer Vision and Pattern Recognition, Long Beach, CA, USA, June 2019, pp. 3146–3154
9 Cui, Y., Jia, M., Lin, T.-Y., *et al.*: 'Class-balanced loss based on effective number of samples'. Proc. IEEE Conf. on Computer Vision and Pattern Recognition, Long Beach, CA, USA, June 2019, pp. 9268–9277
10 Goyal, P., and Kaiming, H.: 'Focal loss for dense object detection', *IEEE Trans. Pattern Anal. Mach. Intell.*, 2018, **39**, pp. 2999–3007
11 Paszke, A., Gross, S., Chintala, S., *et al.*: 'Pytorch', Automatic differentiation in PyTorch. 31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA, 2017, pp. 1–4
12 Gulcehre, C., Sotelo, J., and Bengio, Y.: 'A robust adaptive stochastic gradient method for deep learning'. 2017 Int. Joint Conf. on Neural Networks (IJCNN), Anchorage, AK, USA, May 2017, pp. 125–132