A Two-Headed Deep Learning Framework for Predicting Severity of COVID-19 Disease

Coşku ÖKSÜZ [1, *], Oğuzhan URHAN [2], Mehmet Kemal GÜLLÜ [3]

[1] *University of Kastamonu, Bozkurt Vocational School, Dept. of Electronics and Automation, Turkey*
[2] *University of Kocaeli, Dept. of Electronics and Communication Engineering, Turkey*
[3] *University of Bakircay, Dept. of Electrical and Electronics Engineering, Turkey*

* Corresponding Author: University of Kastamonu, Bozkurt Vocational School, Turkey
 Tel.: +90 3662805803. E-Mail: coksuz@kastamonu.edu.tr

A B S T R A C T

Introduction-Objectives: The high contagiousness of the SARS-COV-2 virus has resulted in many people being infected worldwide. In many countries, the capacity of intensive care units has been insufficient and has become unable to accept new patients. Imaging-based non-invasive methods developed as an alternative to the RT-PCR technique to control the spread of the virus during the pandemic process generally focus on the presence or absence of the disease. However, these methods do not provide information about how severe the disease is and how it progresses. Therefore, in this study, a deep learning-based estimation framework with low computational load is proposed to predict severity scores using chest radiographs.

Materials-Methods: The pre-trained ImageNet models are used as feature extraction networks to extract generic features. A two-headed estimation subnetwork each with the same number of layers is created to learn task-specific features. Eventually, an end-to-end trainable lightweight deep model is created by connecting the estimation subnetwork to the feature extraction network.

Results: The proposed model is evaluated on a publicly available Cohen's *covid-chestxray-data set*. The best cross-validation performance in terms of RMSE, MAE, and R^2 in the prediction of lung involvement and opacity is obtained as 1.39/0.98, 1.1/0.81, 0.65/0.66, respectively.

Conclusions: Although the model has been trained with limited data, promising results are achieved with an end-to-end framework for estimating the severity of the COVID-19 disease.

1. Introduction

The COVID-19 outbreak that started in Wuhan, China in December 2019 continues its impact worldwide. As of April 3, 2021, the number of deaths worldwide is approximately 2.8M, while the number of cases is 129.9M [1]. Those infected with the disease show mild symptoms, spreading the virus at an early stage, making it difficult to control the spread [2]. Although vaccines against the virus have been developed, the mutations that the virus has undergone cause it to be transmitted faster than before. For this reason, early detection and isolation of the patients with the disease is of great importance for the correct treatment, planning, and control of the spread.

After proving that medical images carry various information about COVID-19, many researchers aimed to develop rapid screening tools that could be an alternative to RT-PCR (Reverse Transcriptase-Polymerase Chain Reaction) technique [3]. In some of these developed methods, hand-crafted features are extracted from medical images, and classical machine learning methods are trained [4]–[6]. Deep learning-based methods have been proposed in another significant number of studies [4], [7]–[9]. It is possible to categorize the proposed deep learning-based methods into three categories: X-Ray, CT, and methods developed using multimodal images. On the other hand, because all these studies focusing on the presence and absence of the disease, it is not possible to obtain information on how severe the disease is. The information to be obtained about the severity of the disease will provide important clues about the course of the disease. Thus, an idea can be obtained to reduce or increase the care for the patient.

Some studies have been conducted to determine the severity of COVID-19 disease. In [3], RT-PCR test and X-Ray imaging are used on all patients suspected of being infected with the SARS-COV-2 virus. It is concluded that the RALE score computed is a numerical value that can express the severity of COVID-19 pneumonia and is an important precursor of emergency care. In [10], a scoring system is developed to determine the severity of COVID-19. Risk factors that cause severe illness are determined by regression on the patient population infected with SARS-COV-2. The scoring system developed in this study reveals that patients in the high-risk group are 20.24 times more likely to develop severe illness than the low-risk group. In [11], a Random Forest model is trained using the quantitative and radiomics features as well as laboratory indexes to assess COVID-19 disease severity. Promising results are obtained in the study.

Developing effective, reliable, and low-cost diagnostic tools is a necessity in the fight against COVID-19. Therefore, it is aimed to develop a reliable method with the low computational cost for predicting severity scores for COVID-19 disease in this study.

The main contributions of this work are as follows:

1- With the proposed model, lung involvement and opacity scores, which explain the severity of COVID-19 pneumonia, are simultaneously estimated by an end-to-end trained deep learning model.

2- The proposed model is lightweight with 2.19M parameters that have been trained on the images obtained from the X-Ray imaging technique, which is available in almost all health institutions.
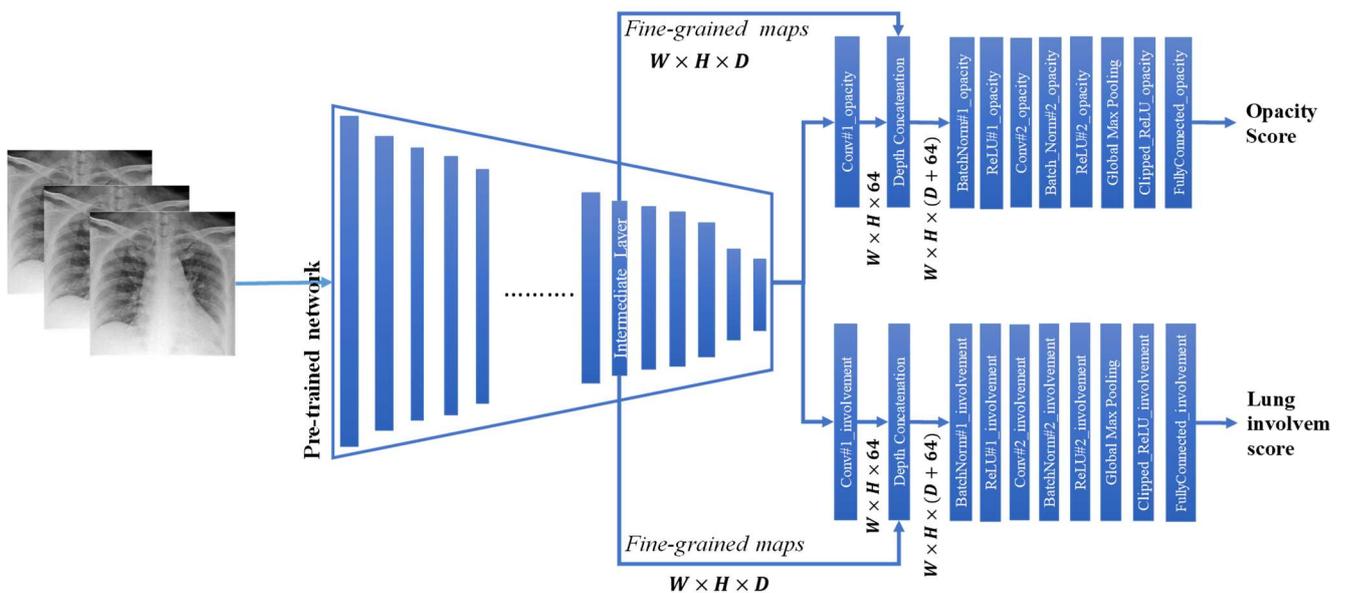
Fig. 1. The proposed severity score estimation framework.

The study is organized as follows. The proposed method is presented in Section 2. Section 3 is devoted to experimental results. Section 4 is devoted to the discussion. Finally, the conclusions are presented in Section 5.

2. Proposed Method

The proposed deep learning-based severity score estimation framework is given in Fig. 1. As seen in Fig. 1, the framework consists of two major parts, i.e, Feature Extraction Network (FEN) and Severity Score Estimation Subnetwork (SSEN). Accordingly, activation maps generated at the top of the FEN are fed to the SSEN. Finally, the severity scores i.e. lung involvement and opacity are estimated by SSEN. The details of each network are given in below.

2.1. Feature extraction network

One of the main problems in the field of medical image analysis is the insufficiency of training data. The deep transfer learning strategy [12] is a technique that can overcome this. In this strategy, features learned by a network trained over a larger data set are transferred to the new task. In the study, pre-trained ImageNet [13] models at different depths are used as the feature extractors of the proposed framework to see which model performs well. Accordingly, the models used in the study are ResNet-18 [14], SqueezeNet [15], GoogLeNet [16], ShuffleNet [17], MobileNetv2 [18] and EfficientNet-B0 [19], respectively. All the weights of each model to be used as feature extractors are frozen to avoid overfitting.

2.2. Severity score estimation subnetwork

The severity score estimation subnetwork consists of additional layers working in parallel to learn task-specific features regarding opacity and lung involvement. First, at the beginning of each parallel branch, activation maps are processed by a convolutional layer with 3×3 sized 64 filters. Then, a depth concatenation layer is used to consolidate the obtained maps with fine-grained maps incoming from an intermediate layer. In Table 1, the last and intermediate layer names of each model are given.

Table 1. Pre-trained models are given with the layer names.

Model	Last Layer	Tensor	Intermediate Layer	Tensor
ResNet-18	res5b_relu	7×7×512	res4b_relu+maxpool	7×7×256
SqueezeNet	fire9-concat	13×13×512	fire6-concat	13×13×384
GoogLeNet	inception_5b-output	7×7×1024	inception_4b-output+maxpool	7×7×528
ShuffleNet	node_199	7×7×544	node_163	7×7×544
MobileNetv2	out_relu	7×7×1280	block_15_add	7×7×160
EfficientNet-B0	efficientnet-b0|Model|Head|MulLayer	7×7×1280	efficientnet-b0|model|blocks_14|Add	7×7×192

The obtained feature maps are further processed along with the batch normalization, ReLU, convolutional layer, batch normalization, and ReLU stages, respectively. A global max-pooling layer is used to extract more prominent features at each branch. Accordingly, a 1×1×64 dimensional vector is obtained by picking one feature from each of the 64 channels. Then, a clipped ReLU is used to threshold opacity and lung involvement scores that exceed expert-defined ceiling values. Finally, a fully connected layer is used to estimate scores at each branch.

3.   Experiments and Results

3.1. Data set

In the study, the publicly available data set given in [20] is used. The data set consists of 94 X-ray images in which each image corresponds to the COVID-19 case. The lung involvement extent (or geographic extent) and opacity extent are predicted by three-blinded experts following a scoring system [21]. The extent for each lung i.e. '0', '1', '2', '3', '4' indicate 'no involvement', '<25% involvement', '25%-50% involvement', '50%-75% involvement', and '>75% involvement', respectively [22]. Accordingly, the total extent of the lung involvement is aligned between 0-8 to be included left and right lung [22]. Similarly, the extent of opacity for each lung is aligned between 0-3, where '0', '1', '2' and '3' indicate 'no opacity', 'ground-glass opacity', 'consolidation' and 'white-out', respectively. As a result, the extent of total opacity is aligned between 0-6 [22]. Therefore, the ceiling threshold of the clipped ReLU layer in each branch of the proposed framework given in Fig. 1 is set as in (1). This setting facilitates the convergence of the network.

$$\text{Clipped\_Relu\_involvement} = \begin{cases} 0, & x < 0 \\ x, & 0 < x \leq 8 \\ 8, & x > 8 \end{cases} \quad \text{Clipped\_Relu\_opacity} = \begin{cases} 0, & x < 0 \\ x, & 0 < x \leq 6 \\ 6, & x > 6 \end{cases} \tag{1}$$

3.2. Hyperparameters, Data Augmentation, and Loss Function

Each model has been trained for 200 epochs with SGDM (Stochastic Gradient Descent with Momentum) optimizer. The other hyperparameters i.e. *Learn rate*, *L_2 regularization and Momentum* are set as 10^{-3}, 10^{-4}, and 0.9, respectively. The data is randomly augmented at each training epoch using the affine transformations i.e. rotation, scaling, reflection and translation. As the different version of the original training data is seen by the network at each training epoch, data is augmented 200-fold (~15,200 samples) in total. The half mean square error function given in (2) is used to calculate the loss, where N, h_φ, and y indicate the number of observations, network responses and targets, respectively.

$$\mathcal{L}_{opacity, \ involvement} = \frac{1}{2} \underbrace{\frac{1}{N} \sum_{i=1}^{N} \left(h_\varphi(x^i) - y^i \right)^2}_{MSE} \tag{2}$$

Accordingly, the loss values obtained at each head of the network i.e. the lung involvement loss ($\mathcal{L}_{involvement}$) and the opacity loss ($\mathcal{L}_{opacity}$) are summed up to obtain total loss ($\mathcal{L}_{total}$) as given in (3). Finally, the gradients are updated using the total loss.

$$\mathcal{L}_{total} = \mathcal{L}_{opacity} + \mathcal{L}_{involvement} \tag{3}$$

3.3. Performance metrics

The performance metrics utilized in the study are root-mean-square-error (*RMSE*), determination coefficient (R^2), and mean absolute error (MAE). The *RMSE*, R^2, and *MAE* metrics are computed as given in (4).

$$RMSE = \sqrt{MSE}, \quad R^2 = 1 - \frac{\sum_{i=1}^{N}\left(h_\varphi(x^i) - y^i\right)^2}{\sum_{i=1}^{N}\left(y^i - \bar{y}\right)^2}, \quad MAE = \frac{1}{N}\sum_{i=1}^{N}\left|h_\varphi(x^i) - y^i\right| \tag{4}$$

3.4. Performance comparison

In this section, the performance of each model is evaluated using the RMSE metric. The evaluation process is demonstrated in Fig. 2. The data set is first randomly divided into training and test sets (hold-out) using the rates of 80% and 20%. Then, the training set is further partitioned into five non-overlapping test sets (folds) for five-fold cross-validation. Each model is validated using the five-fold cross-validation. Accordingly, the RMSE scores obtained at each test fold for opacity and lung involvement are averaged to reveal overall performance. Finally, the best-performing model is evaluated on the hold-out set.
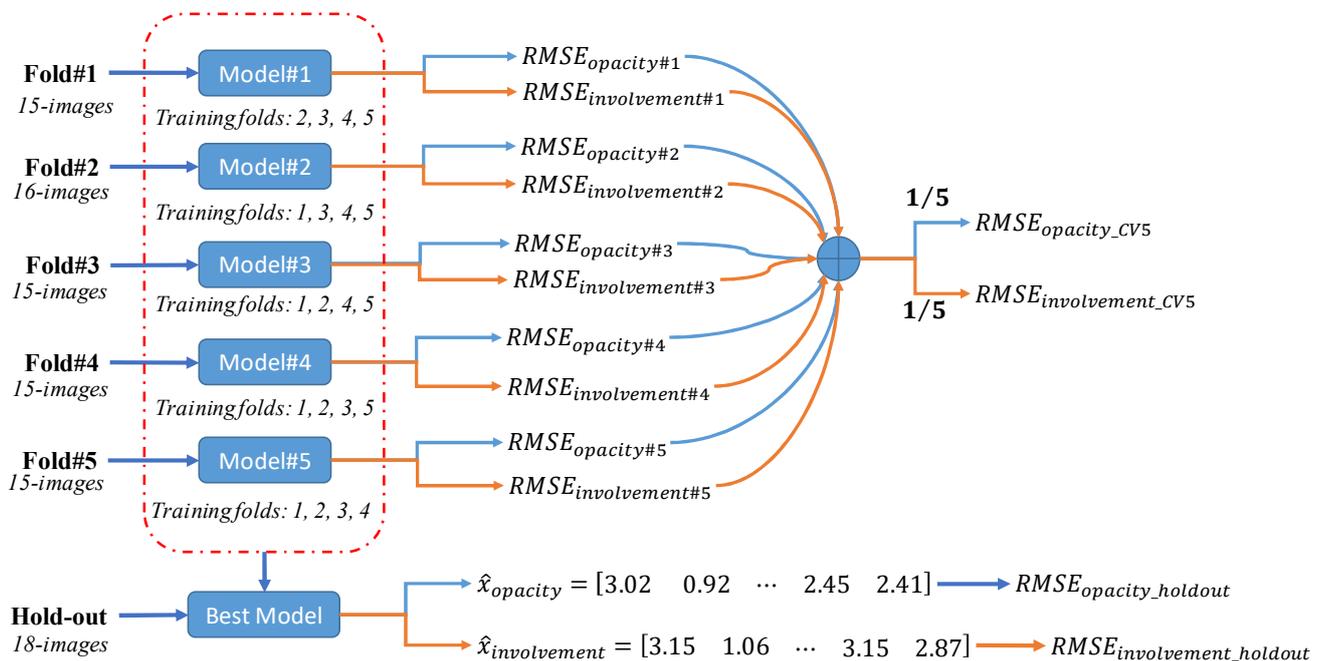
Fig. 2. Evaluation process of the proposed framework.

In Table 2, the RMSE scores achieved by each method on each fold are given in detail. The five-fold cross-validation and the hold-out scores achieved by each method are also given. As seen in Table 2, since the deviation of the results obtained at each fold is relatively low, more stable performance is obtained with ShuffleNet and ResNet-18 based models compared to other models. On the other hand, it is seen from Table 2 that all models provide better performance in the hold-out set due to the limited sample size.

Table 2. Performance comparison.

Feature extractor	RMSE						
	Fold-1 15-Images	Fold-2 16-Images	Fold-3 15-Images	Fold-4 15-Images	Fold-5 15-Images	*Avg.*	Hold-out 18-Images
SqueezeNet	1.55/1.77	1.61/1.36	2.32/2.08	1.94/1.11	2.19/0.89	1.93/1.45	1.70/1.08
ShuffleNet	1.12/1.01	1.18/0.78	1.51/1.17	1.52/1.03	1.58/0.90	**1.39/0.98**	**1.28/0.82**
MobileNetv2	1.41/1.32	1.60/1.00	2.27/1.68	1.92/1.07	1.97/1.62	1.84/1.34	1.32/1.20
ResNet-18	1.39/1.04	1.39/0.94	1.79/1.30	1.27/0.64	1.62/1.24	1.50/1.04	1.34/0.85
GoogLeNet	1.19/1.03	1.89/1.25	2.72/1.15	1.77/1.01	1.62/1.08	1.84/1.11	1.53/0.97
EfficientNet-B0	1.59/1.27	1.48/0.70	1.92/1.29	1.94/1.15	1.58/1.06	1.71/1.10	1.30/0.85

In Fig. 3, the average five-fold cross-validation score obtained with each method is given as a bar graph to easily compare the results based on MAE, R^2, and RMSE. As seen in Fig. 3, the minimum MAE and RMSE for both lung involvement and opacity prediction are obtained with the ShuffleNet-based model. The maximum R^2 score for lung involvement prediction is obtained by the ShuffleNet-based and ResNet-18-based models. The maximum R^2 score for opacity prediction is obtained by the ShuffleNet-based model. On the other hand, the SqueezeNet-based model underperforms all other methods based on MAE, R^2, and RMSE.
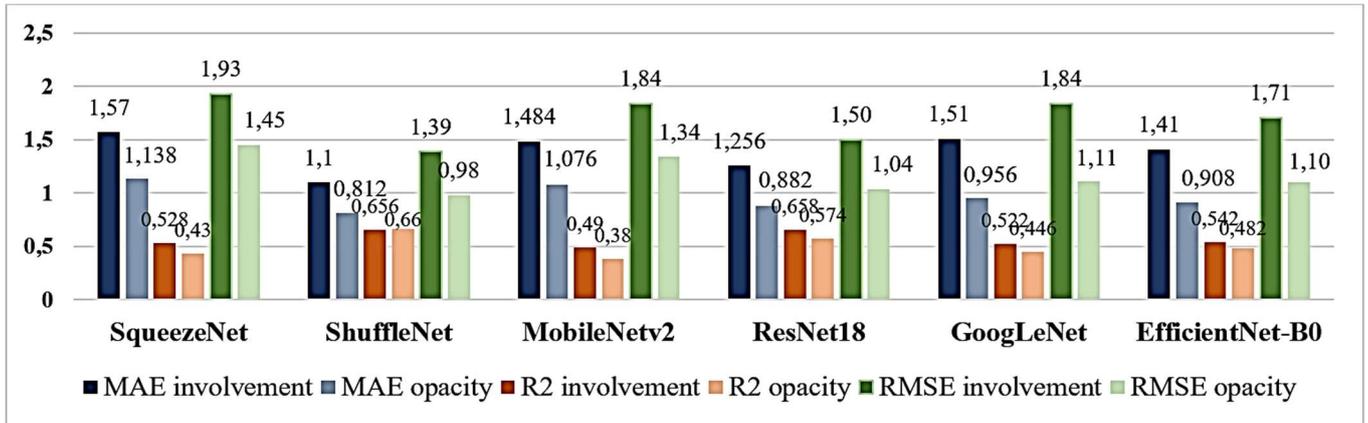
Fig. 3. Comparison of the methods based on the metrics i.e., MAE, RMSE, and R^2.

Since the best performance is achieved with the ShuffleNet-based model, predictions of this model on hold-out set are plotted versus actuals. In Fig. 4(a), the estimates (oval shapes in black) made by the lung involvement head are plotted versus the actual scores (oval shapes in blue). In Fig. 4(b), the estimates (oval shapes in black) made by the opacity score head are plotted versus the actual scores (oval shapes in blue). The vertical red lines in Fig. 4(a) and Fig. 4(b) show the errors between the actuals and the network estimates. In Fig. 4(c)-4(d), the scatter plots show the predictions versus actuals corresponding to the lung involvement head and the opacity score head, respectively. The dashed red line drawn in Fig. 4(c) and Fig. 4(d) shows the perfect prediction line. As seen in Fig. 4 (c) and 4 (d), the predicted points lie along with the perfect prediction line. Also, the mean absolute distance of points to the perfect prediction line is 1.11 and 0.69 in Fig. 4(c) and Fig. 4(d), respectively.

(a)   **(b)**

(c)   **(d)**
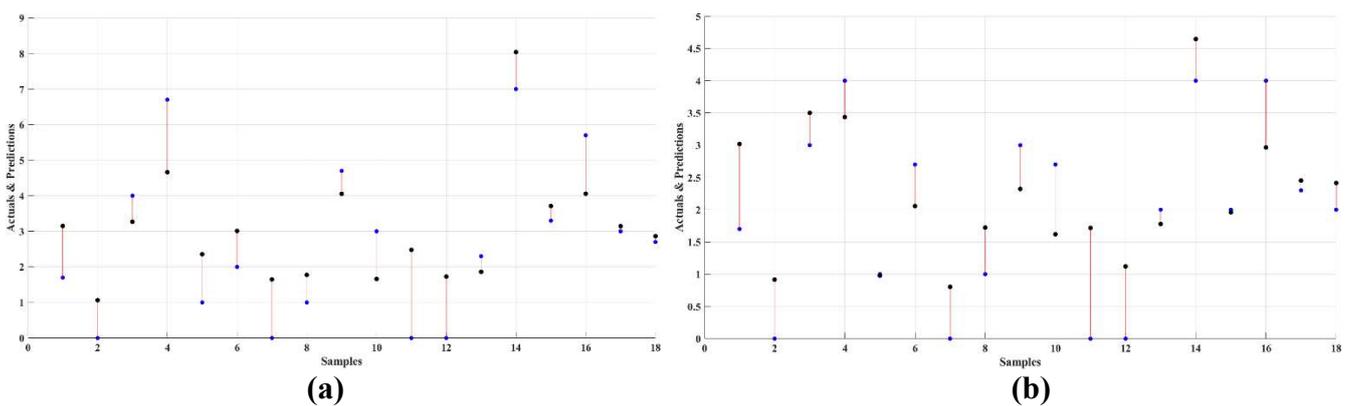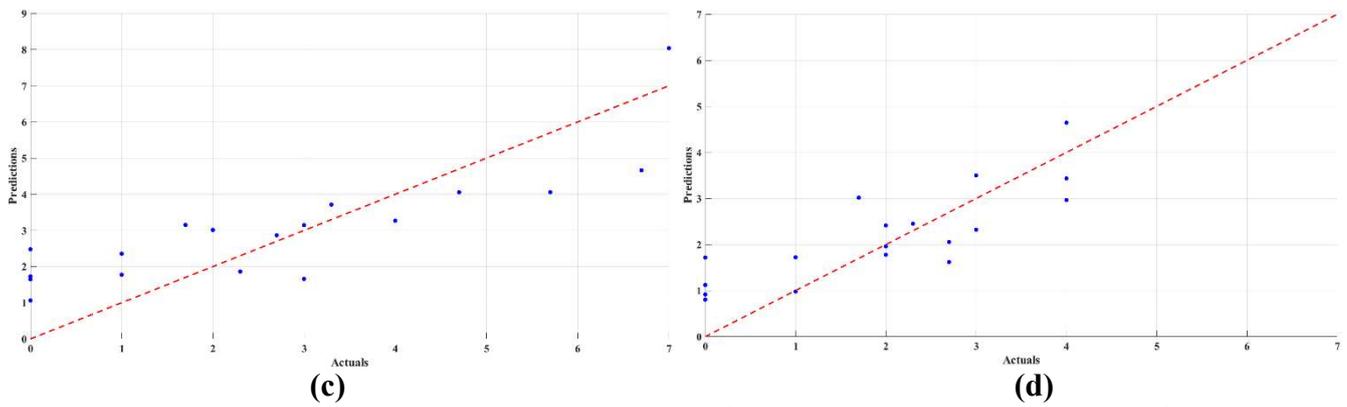
Fig. 4. Predictions versus actuals are plotted with error lines in (a) for the lung involvement head and in (b) for the opacity score head. Predictions versus actuals are plotted without error lines in (c) for the lung involvement score head and in (d) for the opacity score head.

3.5. Visualization

In this section, Gradient-weighted Regression Activation Maps (Grad-RAM) are given for images selected from the hold-out set. Since the network is double-headed, two types of RAM are provided in Fig. 5. Accordingly, the images given in the leftmost column of Fig. 5 are original. While the RAM images given in the middle column are obtained from the lung involvement score head, images given in the rightmost column are obtained using the opacity score head. Predicted scores are also given under each image, with the corresponding actual scores (Gth) assigned by experts.

Gth: 2.7, predicted: 2.86   **Gth: 2.0, predicted: 2.41**

Gth: 3.0, predicted: 3.14   **Gth: 2.3, predicted: 2.45**

Gth: 4.7, predicted: 4.05   **Gth: 3.0, predicted: 2.32**

Gth: 3.3, Predicted: 3.71   **Gth: 2.0, Predicted: 1.96**
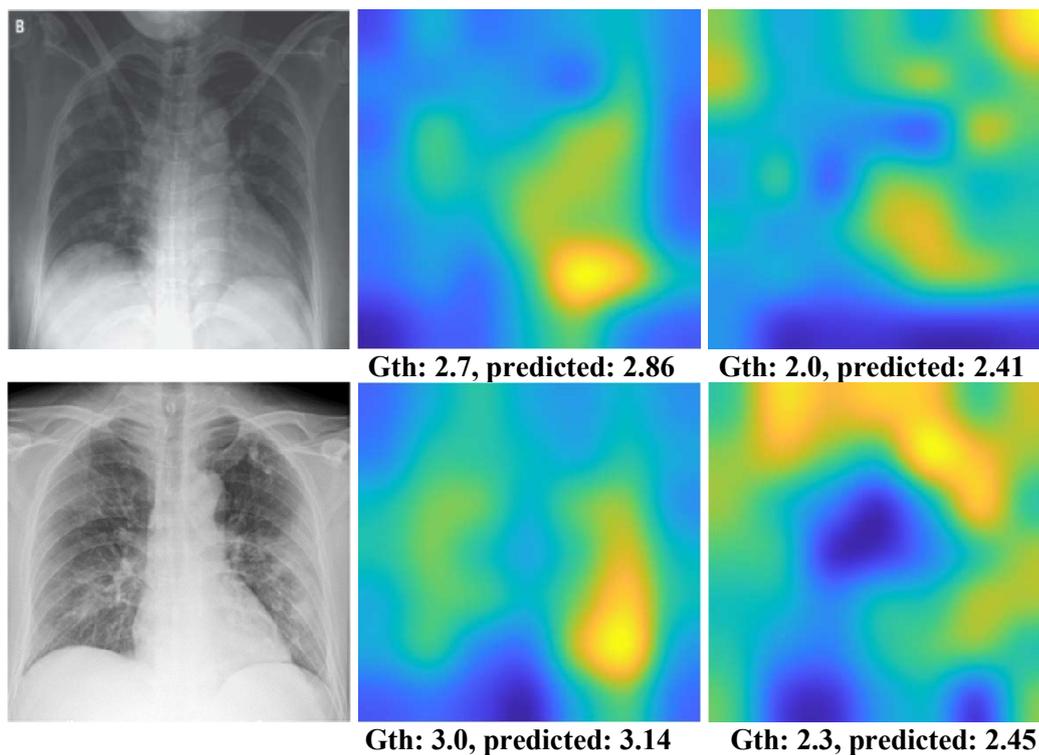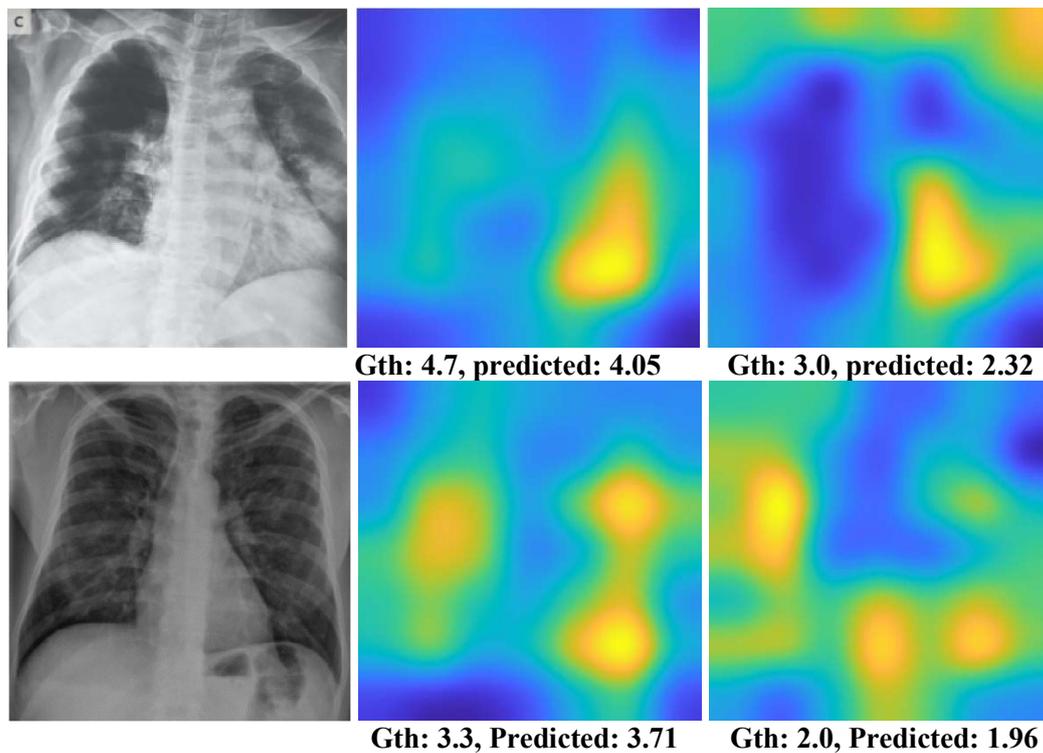
Fig. 5. Grad-RAM images given in the middle and rightmost columns are obtained using the lung involvement head of the model and the opacity score head, respectively.

In Fig. 5, RAM images from each head generally focus on the regions inside the lungs, particularly the opaque regions. However, unrelated areas outside the lungs are also returned with warm colors, especially from the opacity score head. The most likely reason for the model to look at these regions is the insufficient training examples.

4. Discussion

The only difference between the models formed to predict the severity scores is the feature extraction backbone. To find an optimal feature extraction backbone that provides superior performance for any given task is extremely important. The feature maps generated by the pre-trained networks consist of universal and complex information. When this information is consolidated with the task-related features learned by additional layers, performance is improved. In this work, the better combination in terms of this is accomplished by the ShuffleNet-based model. This model yields better generalization ability on a hold-out set. In Table 3, the proposed method is compared with the reference studies using the same data set. Cohen et al. [22] used a pre-trained DenseNet [23] model to predict severity scores. The model is trained with non-COVID-19 datasets collected from different data sources before creating a final model. Promising results are obtained. Amer et al. [24] proposed a deep learning-based model to predict severity scores. Before the severity score estimation, lungs are segmented in the work. Only the lung involvement score is estimated from the segmented lungs. As seen in Table 3, superior results are obtained in terms of RMSE, MAE, and R^2 compared to the reference studies with the proposed method.

Table 3. Comparing the results with the reference studies.

Work	Validation method	Test set	Parameters	RMSE	MAE	R^2
Cohen et al. **[22]**	Hold-out	50-images	-	1.43/0.92	1.14/0.78	0.60/0.58
Amer et al. **[24]**	Hold-out	50-images	-	-	-	0.67/-
Proposed method	Hold-out	18-images	2.19M	**1.28/0.82**	1.11/**0.69**	**0.69/0.66**
	CV-5	76-images		1.39/0.98	**1.10**/0.81	0.65/0.66

5. Conclusion

In this study, the COVID-19 pneumonia severity scores, namely lung involvement, and opacity are estimated by a two-headed ShuffleNet-based network. Since the proposed method has a total of 2.19M parameters, it offers a low complexity solution to the problem. The proposed model can be used to triage patients as well as monitor the course of COVID-19 patients in the intensive care unit. One of the limitations of this study is that the model has yet been trained with a limited number of samples. As the data set is updated with more expert-labeled data, performance can be improved by fine-tuning the model. Besides, performance can be further improved by using only lungs in the severity prediction following a segmentation. These issues will be taken into consideration in our next study.

References

[1]   'WHO Coronavirus (COVID-19) Dashboard'. https://covid19.who.int (accessed Apr. 03, 2021).

[2]   K. R. Peck, 'Early diagnosis and rapid isolation: response to COVID-19 outbreak in Korea', *Clin. Microbiol. Infect.*, vol. 26, no. 7, pp. 805–807, Jul. 2020, doi: 10.1016/j.cmi.2020.04.025.

[3]   D. Cozzi *et al.*, 'Chest X-ray in new Coronavirus Disease 2019 (COVID-19) infection: findings and correlation with clinical outcome', *Radiol. Med. (Torino)*, vol. 125, no. 8, pp. 730–737, Aug. 2020, doi: 10.1007/s11547-020-01232-9.

[4]   A. M. Ismael and A. Şengür, 'Deep learning approaches for COVID-19 detection based on chest X-ray images', *Expert Syst. Appl.*, vol. 164, p. 114054, Feb. 2021, doi: 10.1016/j.eswa.2020.114054.

[5]   K. Shankar and E. Perumal, 'A novel hand-crafted with deep learning features based fusion model for COVID-19 diagnosis and classification using chest X-ray images', *Complex Intell. Syst.*, Nov. 2020, doi: 10.1007/s40747-020-00216-6.

[6]   Ş. Öztürk, U. Özkaya, and M. Barstuğan, 'Classification of Coronavirus (COVID-19) from X-ray and CT images using shrunken features', *Int. J. Imaging Syst. Technol.*, vol. 31, no. 1, pp. 5–15, 2021, doi: https://doi.org/10.1002/ima.22469.

[7]   C. Öksüz, O. Urhan, and M. K. Güllü, 'Ensemble-CVDNet: A Deep Learning based End-to-End Classification Framework for COVID-19 Detection using Ensembles of Networks', *ArXiv201209132 Eess*, Dec. 2020, Accessed: Dec. 20, 2020. [Online]. Available: http://arxiv.org/abs/2012.09132.

[8]   L. Wang, Z. Q. Lin, and A. Wong, 'COVID-Net: a tailored deep convolutional neural network design for detection of COVID-19 cases from chest X-ray images', *Sci. Rep.*, vol. 10, no. 1, Art. no. 1, Nov. 2020, doi: 10.1038/s41598-020-76550-z.

[9]   M. K. Hasan, M. T. Jawad, K. N. I. Hasan, S. B. Partha, and M. M. A. Masba, 'COVID-19 identification from volumetric chest CT scans using a progressively resized 3D-CNN incorporating segmentation, augmentation, and class-rebalancing', *ArXiv210206169 Cs Eess*, Feb. 2021, Accessed: Apr. 04, 2021. [Online]. Available: http://arxiv.org/abs/2102.06169.

[10]   C. Zhang *et al.*, 'A Novel Scoring System for Prediction of Disease Severity in COVID-19', *Front. Cell. Infect. Microbiol.*, vol. 10, p. 318, Jun. 2020, doi: 10.3389/fcimb.2020.00318.

[11]   Z. Tang *et al.*, 'Severity assessment of COVID-19 using CT image features and laboratory indices', *Phys. Med. Biol.*, vol. 66, no. 3, p. 035015, Jan. 2021, doi: 10.1088/1361-6560/abbf9e.

[12]   C. Tan, F. Sun, T. Kong, W. Zhang, C. Yang, and C. Liu, 'A Survey on Deep Transfer Learning', in *Artificial Neural Networks and Machine Learning – ICANN 2018*, Cham, 2018, pp. 270–279, doi: 10.1007/978-3-030-01424-7_27.

[13]   O. Russakovsky *et al.*, 'ImageNet Large Scale Visual Recognition Challenge', *ArXiv14090575 Cs*, Jan. 2015, Accessed: Nov. 17, 2020. [Online]. Available: http://arxiv.org/abs/1409.0575.

[14]   K. He, X. Zhang, S. Ren, and J. Sun, 'Deep Residual Learning for Image Recognition', 2016, pp. 770–778, Accessed: Mar. 26, 2021. [Online]. Available: https://openaccess.thecvf.com/content_cvpr_2016/html/He_Deep_Residual_Learning_CVPR_2016_paper.html.

[15] F. N. Iandola, S. Han, M. W. Moskewicz, K. Ashraf, W. J. Dally, and K. Keutzer, 'SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <0.5MB model size', *ArXiv160207360 Cs*, Nov. 2016, Accessed: Nov. 17, 2020. [Online]. Available: http://arxiv.org/abs/1602.07360.

[16] C. Szegedy *et al.*, 'Going Deeper With Convolutions', 2015, pp. 1–9, Accessed: Mar. 26, 2021. [Online]. Available: https://www.cv-foundation.org/openaccess/content_cvpr_2015/html/Szegedy_Going_Deeper_With_2015_CVPR_paper.html.

[17] X. Zhang, X. Zhou, M. Lin, and J. Sun, 'ShuffleNet: An Extremely Efficient Convolutional Neural Network for Mobile Devices', *ArXiv170701083 Cs*, Dec. 2017, Accessed: Nov. 17, 2020. [Online]. Available: http://arxiv.org/abs/1707.01083.

[18] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, 'MobileNetV2: Inverted Residuals and Linear Bottlenecks', *ArXiv180104381 Cs*, Mar. 2019, Accessed: Nov. 18, 2020. [Online]. Available: http://arxiv.org/abs/1801.04381.

[19] M. Tan and Q. V. Le, 'EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks', *ArXiv190511946 Cs Stat*, Sep. 2020, Accessed: Nov. 17, 2020. [Online]. Available: http://arxiv.org/abs/1905.11946.

[20] J. P. Cohen, P. Morrison, L. Dao, K. Roth, T. Q. Duong, and M. Ghassemi, 'COVID-19 Image Data Collection: Prospective Predictions Are the Future', *ArXiv200611988 Cs Eess Q-Bio*, Dec. 2020, Accessed: Mar. 26, 2021. [Online]. Available: http://arxiv.org/abs/2006.11988.

[21] H. Y. F. Wong *et al.*, 'Frequency and Distribution of Chest Radiographic Findings in Patients Positive for COVID-19', *Radiology*, vol. 296, no. 2, pp. E72–E78, Aug. 2020, doi: 10.1148/radiol.2020201160.

[22] J. P. Cohen *et al.*, 'Predicting COVID-19 Pneumonia Severity on Chest X-ray With Deep Learning', *Cureus*, Jul. 2020, doi: 10.7759/cureus.9448.

[23] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, 'Densely Connected Convolutional Networks', *ArXiv160806993 Cs*, Jan. 2018, Accessed: Apr. 05, 2021. [Online]. Available: http://arxiv.org/abs/1608.06993.

[24] R. Amer, M. Frid-Adar, O. Gozes, J. Nassar, and H. Greenspan, 'COVID-19 in CXR: from Detection and Severity Scoring to Patient Disease Monitoring', *IEEE J. Biomed. Health Inform.*, pp. 1–1, 2021, doi: 10.1109/JBHI.2021.3069169.