

Improving RepVGG model with variational data imputation in COVID-19 classification

Kien Trang^{1,2}, An Hoang Nguyen^{1,2}, Long TonThat^{1,2}, Bao Quoc Vuong^{1,2,3}

¹School of Electrical Engineering, International University, Ho Chi Minh City, Vietnam

²Vietnam National University, Ho Chi Minh City, Vietnam

³Univ Brest, CNRS, Lab-STICC, CS 93837, 6 avenue Le Gorgeu, CEDEX 3, Brest, France

Article Info

Article history:

Received Sep 29, 2021

Revised Apr 5, 2022

Accepted May 5, 2022

Keywords:

Chest X-Ray

COVID-19

Deep learning

RepVGG

Variational autoencoder

ABSTRACT

Millions of fatal cases have been reported worldwide as a result of the Coronavirus disease 2019 (COVID-19) outbreak. In order to stop the spreading of disease, early diagnosis and quarantine of infected people are one of the most essential steps. Therefore, due to the similar symptoms of SARS-CoV-2 virus and other pneumonia, identifying COVID-19 still exists some challenges. Reverse transcription-polymerase chain reaction (RT-PCR) is known as a standard method for the COVID-19 diagnosis process. Due to the shortage of RT-PCR toolkit in global, Chest X-Ray (CXR) image is introduced as an initial step to support patient's classification. Applying deep learning in medical imaging becomes an advanced research trend in many applications. In this research, RepVGG pre-trained model is considered to be used as the main backbone of the network. Besides, variational autoencoder (VAE) is firstly trained to perform lung segmentation. Afterwards, the encoder part in VAE is preserved as an additional feature extractor to combine with RepVGG performing classification. A COVID-19 radiography database consisting of 3 classes COVID-19, Normal and Viral Pneumonia is conducted. The obtained average accuracy of the proposed model is 95.4% and other evaluation metrics also show better results compared with the original RepVGG model.

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



Corresponding Author:

An Hoang Nguyen

School of Electrical Engineering, International University

Quarter 6, Linh Trung Ward, Thu Duc City, Ho Chi Minh City, Vietnam

Email: nhan@hcmiu.edu.vn

1. INTRODUCTION

After more than a year, Coronavirus disease 2019 (COVID-19) has been declared a worldwide pandemic and a public health emergency by the World Health Organization (WHO). The rapid transmission of this disease in humans makes COVID-19 become one of the major threats to humanity. This is also one of the main factors leading to complex situations. According to the report of WHO, the confirmed cases have reached over 180 billion and there are 4 billion people who have fallen into fatal cases [1]. Reverse transcription-polymerase chain reaction (RT-PCR) is the main tool to diagnose COVID-19. This method is to conduct the detection of the viral nucleic acid in SARS-CoV-2. Nevertheless, the error during the process of sampling caused by samples with low viral load is one of the disadvantages of the RT-PCR test. Antigen testing is known to give rapid results, but the sensitivity to detect viruses in patient samples is low. Besides, this normally requires a huge number of testing kits and health human resources, followed by the increase of infected patients. To overcome the high demand for the usage of the toolkit, medical imaging methods such as X-ray and computerized tomography (CT) are considered as an alternative solution for detecting

disease [2]. Therefore, in compared to other imaging techniques, chest X-ray (CXR) is known as an imaging approach to aid radiologists in rapid identification and reduction of expenses.

Recently, deep learning has emerged as a cutting-edge method for computer vision and pattern recognition achieving outstanding results in image-based classification. Therefore, the implementation of deep learning in medical imaging such as X-ray is considered as a new trend in automated disease identification and several of these types of research have been already conducted in the field of COVID-19 detection. In recent work, COVID-19 detection based on CXR images is proposed by Wang *et al.* [3] using a deep convolutional neural network (DCNN). Compared with other pre-trained networks, their proposed COVID-Net gives an outstanding result of over 93% in terms of average accuracy. Besides, COVID-ResNet is proposed by [4] based on the traditional ResNet-50 to process the CXR images in COVIDx dataset. The better generalization coming from 3 different resolutions of images training leads to that the accuracy is higher than the COVID-Net - reaching 96.23%. To increase the number of images for training, the usage of X-ray projected generative adversarial network (XPGAN) is mentioned by Quan *et al.* [5]. This is able to synthesize more CXR images based on the current images which can give an improvement in data augmentation and classification accuracy. Applying convolutional neural network (CNN) based on spectral analysis is introduced by Singh and Singh [6]. Integration of multiresolution analysis (MRA) by wavelet decomposition is conducted to produce frequency sub-bands before feeding into CNN for classification. Grad-CAM is also applied to visualize gradient information in the forms of a heatmap for diagnosis and the final obtained result achieves over 95%. Oh *et al.* [7] employed a CNN working with small patches to deal with COVID-19 identification in the limited dataset. Lung area is extracted and then its background is removed by FC-DenseNet. The processed images are separated into many small patches to feed in CNN for training and testing. Despite the amelioration in sensitivity, the classification accuracy is 91.9% which is less than the COVID-Net. Transfer learning from pre-trained models is also applied in the work [8]–[10]. By fine-tuning, the network is able to learn the new features based on the CXR images and this process also gives a better result compared with training from scratch in terms of accuracy and training time.

In this paper, the combination of RepVGG pre-trained model and variational data imputation is proposed. Data imputation of variational autoencoder (VAE) is conducted with U-net for segmentation training at initial. Then, the encoder part of U-net and VAE is preserved and treated as an additional feature extractor for combining with RepVGG. By this connection, the extra features are added to the pre-trained model which can give benefits to the classification performance. The structure of the paper is organized as follow. The proposed approach is described in section 2, followed by the results and discussions are given in section 3. Finally, the conclusion of the paper is given in section 4.

2. PROPOSED APPROACH

The proposed model to improve RepVGG in the case of COVID-19 disease identification based on CXR image is illustrated in Figure 1. In the beginning, the CXR images are prepared and re-scaled to feed into the two following networks. Initially, the U-Net type network is trained for lung segmentation with the support of variational data imputation. The decoder part of the network is then removed while the output of encoder part is flattened as a feature vector. Simultaneously, the final layer of RepVGG model is adjusted to concatenate the features from the encoder part. Thanks to this connection, the classification layer is employed based on two types of features extracted from two networks. While training the COVID-19 disease classification, all the layers in U-net and a small part of layers in RepVGG are frozen. This can give an improvement for learning as a result of the pre-trained model.

2.1. Dataset

The training classification in this work is conducted by COVID-19 radiography dataset [11] and [12]. The dataset is collected by international research teams in Qatar, Pakistan, and Malaysia from different sources and publications. All the images have a resolution of 299×299 with the PNG format. In this research, we employ three classes in the dataset including COVID-19, Normal and Viral Pneumonia which is described in Table 1.

Table 1. Dataset description

	Number of images
COVID-19	3,616
Normal	10,200
Viral Pneumonia	1,345

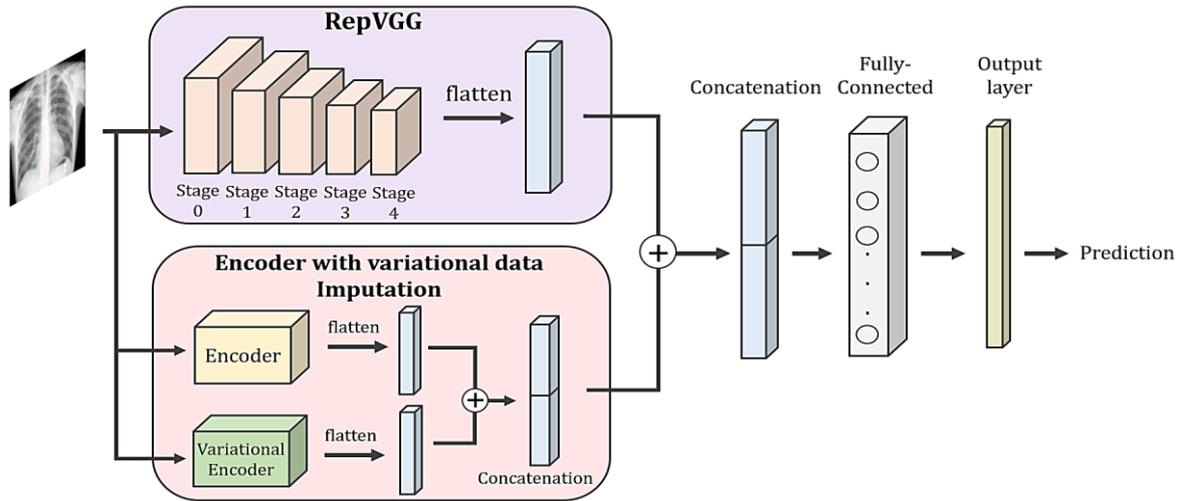


Figure 1. The proposed model architecture

2.2. U-net with variational data imputation

U-net is a symmetrical end-to-end network which is introduced by Ronneberger *et al.* [13] to perform biomedical image segmentation at initial. Encoder and decoder are two main parts of the U-net architecture. Compared with fully convolutional network (FCN), U-shape architecture contributes concatenation of the features in the shallow layers with those in the deeper layers. Besides, the robust idea of this structure is the usage of skip connection between the encoder and decoder layers to increase the performance in the process of reconstruction. Encoder is also treated as a contraction path, which employs feature extraction task to seek the context information of the input image. Downsampling is then performed with the usage of deep convolutional layers, which results in shrinking the dimension of input while increasing the depth. In contrast, decoder is considered as an expansion path, which assembles the precise localization of output pixels based on the latent space. Upsampling is applied to perform reconstruction by using transposed convolution. Afterwards, skip connection from encoder is merged to concatenate learned feature information from encoder at the same level. This helps decoder to perform back learning the features from the corresponding stage of encoder getting better precise location in upsampling operation. Then, a pixel-wise Softmax layer is used to get the final output by calculating the probability of each pixel.

On the other hand, VAE is known as one of the common unsupervised learning approaches and is firstly proposed by Kingma and Welling [14]. The standard structure of VAE is depicted in Figure 2. The traditional autoencoder is using input as ground truth and performs learning to minimize the distance between encoder and decoder. This process can extract the potential features to describe the information of original data and store these features in latent representation. However, the minimization process is conducted irrespective of whether the distribution of hidden features is appropriate leading to the problems of overfitting and unacceptable results. Thus, VAE is introduced as an improved version of autoencoder to overcome this problem which is able to seek more accurate features. In comparison with traditional autoencoder, instead of performing the reconstruction of input data by encoding-decoding process, the distribution of probability over latent space is employed in VAE in order to match with the distribution from decoder. Thus, the mean and variance are sampled in the final layer of encoder before passing to the decoder part. Based on the sample from distribution, the decoder computes a latent vector and then, reconstruction of the original data is performed. Nevertheless, the backpropagation computation is not able to compute through a random sampling distribution. To overcome this problem, a reparameterization trick is mentioned which is able to perform the resampling operation to variable $\epsilon \sim N(0,1)$. Shifting by the mean of latent distribution μ and scaling by the standard deviation of latent distribution σ are then conducted and contribute to the latent variable z . For the general cases, the optimization purpose in VAE is defined by the evidence lower bound (ELBO) loss which is described in (1).

$$ELBO = \mathcal{L}_{rec} + \mathcal{K}\mathcal{L}_{loss} \quad (1)$$

The ELBO loss is the summation of two types of loss. The first term \mathcal{L}_{rec} indicates the reconstruction loss for determining the loss between the original data and the reconstructed data which is

measured by the expected negative log-likelihood as given in (2). Another term is the Kulback-Leibler divergence \mathcal{KL}_{loss} , also called relative entropy which is used for computing the difference of one probability distribution $Q(z|\mathcal{X})$ and the reference probability distribution $P(z)$, as shown in (3). In the case of multivariate normal distributions, the Kulback-Leibler divergence can be written in form of (4) where k is the dimensionality of latent space.

$$\mathcal{L}_{rec} = -\mathbb{E}[\log P(\mathcal{X}|z)] \tag{2}$$

$$\mathcal{KL}_{loss} = \mathbb{E}[\log Q(z|\mathcal{X}) - \log P(z)] \tag{3}$$

$$\mathcal{KL}_{loss} = -\frac{1}{2} \sum_{n=1}^k (1 + \log(\sigma_n^2) - \mu_n^2 - \sigma_n^2) \tag{4}$$

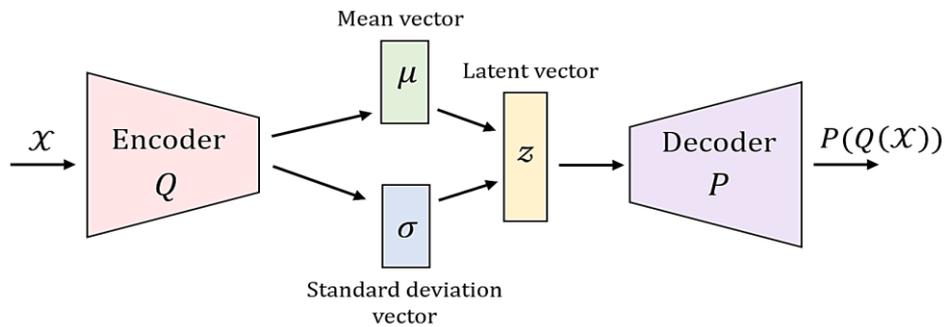


Figure 2. The structure of VAE

VAE, in our research, is used as data imputation to contribute more features for the segmentation process. This runs against the traditional applications of VAE that some researchers focused on data generation [15]–[17]. Besides, denoising based on VAE is proposed in [18] while Anh *et al.* [19] introduce VAE as a feature extraction to combine with random forest for fraud detection.

The combination of U-net and VAE in [20] is conducted in this work. The block diagram of this model is illustrated in Figure 3. Firstly, the training for lung area segmentation is conducted. Encoder in VAE is stacked to the U-net through latent variable z . This connects directly with the latent space of U-Net which is used to represent a low dimensionality of the input image. The combination of features is conducted and fed into the decoder part for segmentation instead of the reconstruction as the origin. Therefore, the first term in (1) would be modified to compute the loss of segmentation, while the second term remains objective as a regularization.

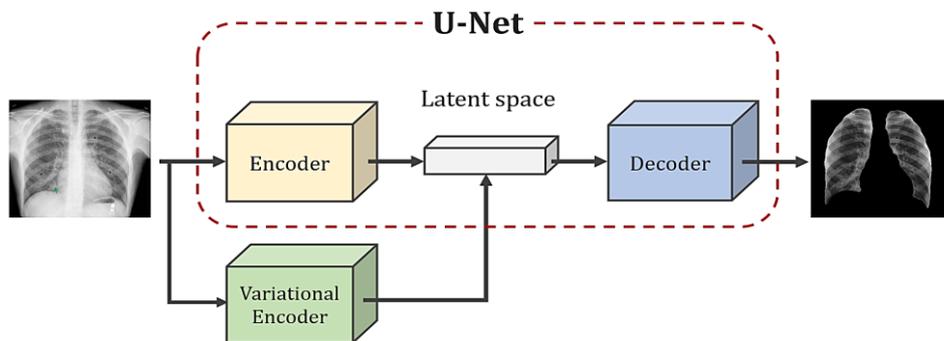


Figure 3. The block diagram of U-net with variational data imputation model

2.3. RepVGG

Currently, deep learning is one of the advanced techniques which can deal with an enormous dataset applying many fields such as image processing, and pattern recognition. DCNN takes advantage of the

connection of more layers compared with the traditional neural network, which results in applying for complicated tasks. Many pre-trained networks have been introduced to deal with different datasets such as AlexNet [21], VGG [22], ResNet [23], GoogleNet [24], and DenseNet [25]. As a result, transfer learning approaches are widely employed in many applications which could transfer the previous learning features to the new tasks with impressive outcomes in terms of training time and accuracy. In the beginning, researchers tend to propose single-branch model such as AlexNet and VGG which are easily implemented and give impressive results in many competitions. This leads to the fact that many new models are attached with more layers to carry out more complex datasets [26]. However, making a deeper network also exists some disadvantages in terms of training because of gradient vanishing [27] which results in unable updating for the initial layers in backpropagation. To address this phenomenon, skip-connection is applied as a solution described in ResNet which encourages the development of later multi-branch models such as GoogleNet and DenseNet. This type of models not only avoids the dependence in one branch preventing from gradient vanishing but also allows to transmit the information in previous layers to the latter ones by concatenated connection. Despite the multi-branch architecture brings an improvement in accuracy, the complexity is increased giving rise to longer time for training and inference. This may cause an increase in the usage of memory and be difficult to implement on some devices.

Inspired by the previous models, RepVGG [28] is born to resolve these disadvantages. Multi-branch architecture is applied in RepVGG, however, the separation of training and inference model makes the increase in performance while retaining the advantages of the multi-branch model. RepVGG model consists of 5 main stages which contain similar blocks in terms of structure. Convolutional layer (Conv), ReLU function and batch normalization (BN) are the main components in RepVGG. The first blocks in each stage carry a 3×3 Conv with stride-2 and a 1×1 Conv with stride-2 for down-sampling. From the second blocks onwards, there are 3 branches: 3×3 Conv with stride-1 added BN, 1×1 Conv added BN and identity using BN. The summation of branches is conducted before feeding into ReLU activation function. To conquer the disadvantages of multi-branch architecture, RepVGG mentions a process called reparameterization which is used to convert from multi-branch to single-branch model before inference. Initially, the convolutional layer is fused with batch normalization. The function of convolutional layer $Conv$ is described in (5) where \mathcal{W} and \mathcal{B} are weight and bias. Then, normalization is performed by subtracting the mean value μ and dividing by the standard deviation of the batch. In terms of batch normalization \mathcal{BN} , γ and β are the scaling and shifting factors, which are mentioned in (6). Substituting (5) into (6), the result is obtained in (7). It is noticed that the first term and second term are similar to $Conv$ function in terms of characteristics. Therefore, let \mathcal{W}_{fused} and \mathcal{B}_{fused} be the first and second terms, the fusion function can be rewritten in (8). Besides, 1×1 Conv is able to be replaced by 3×3 Conv by zero-padding and adjusting the center value of kernel. Subsequently, to convert the identity branch to feed with 3×3 Conv, the identity matrix is employed as a convolutional kernel. Thanks to the mentioned process, the fusion of three different branches is presented as one 3×3 Conv block. Therefore, the architecture is converted from multi-branch to single-branch for inference process as depicted in Figure 4.

$$Conv(x) = \mathcal{W}(x) + \mathcal{B} \quad (5)$$

$$\mathcal{BN}(x) = \gamma \cdot \frac{x - \mu}{\sigma} + \beta \quad (6)$$

$$\mathcal{BN}(Conv(x)) = \gamma \cdot \frac{\mathcal{W}(x) + \mathcal{B} - \mu}{\sigma} + \beta = \frac{\gamma \cdot \mathcal{W}(x)}{\sigma} + \left(\frac{\gamma \cdot (\mathcal{B} - \mu)}{\sigma} + \beta \right) \quad (7)$$

$$\mathcal{BN}(x) = \mathcal{W}_{fused} + \mathcal{B}_{fused} \quad (8)$$

In this work, the final layer in the last stage of RepVGG is replaced by a flattened layer for joining with the encoders as shown in Figure 1. This allows performing the concatenation of the feature vectors in RepVGG and encoders. Due to the encoding-decoding process focused on segmentation, the latent feature may concentrate on the lung area. Thanks to this connection, the additional features are obtained to enlarge more useful information before classification. Finally, the fully-connected and output layers are subsequently added and adjust to fit with the number of classes. Since the pre-trained model dealt with a large scale of images, in this case, transfer learning is applied in the training process to preserve the previous training weights by freezing the low-level layers. Thus, the first three stages and the encoders block are frozen while the remains process re-training.

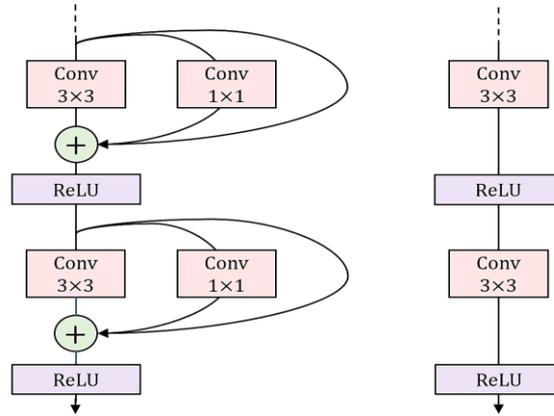


Figure 4. The conversion from multi-branch to single-branch architecture

3. RESULTS AND DISCUSSION

In this experiment, we employ the joint two networks which are U-net with variational data imputation and RepVGG. Since the U-net need to be trained for lung segmentation before removing the decoder part to merge with RepVGG model. This work is applied by [20] which is conducted by the Pulmonary CXR Abnormalities dataset [29]. The training and validation ratio is 75/25 and the augmentation for training data is also employed for enriching the number of training images. The training process is done by GPU NVIDIA Titan X – 12 GB with Adam optimizer and the number of epochs is up to 200. The mentioned results reach around 88% in terms of accuracy higher than the baseline case – 86%. After training for segmentation, the decoder part is removed while the decoder parts of U-net and VAE are preserved to connect with RepVGG later.

For training the classification, the joined model is implemented on GPU NVIDIA Tesla P100 PCIe with 16 GB VRAM and CPU Xeon (R) with 26 GB RAM. The dataset presented in [11] and [12] is applied for this stage which is divided into training and test sets following the 80/20 ratio. Since the joined network has two inputs requiring different image sizes, the re-scaling of image resolution is done with 640×512 pixels for the encoder parts and 224×224 pixels for RepVGG. Besides, histogram equalization is also applied to enhance the contrast of input images. In order to perform a comparison, we conduct two experiments: classification only using RepVGG and using the proposed joining network. The training progress of two networks is shown in Figure 5. Overall, the training progress of only using RepVGG in Figure 5(a) has less fluctuation over 30 epochs than our proposed model in Figure 5(b) has in the first 8 epochs. After 30-epoch of training, the proposed network gives a higher accuracy in terms of training and validation compared with the only using RepVGG case. This would be explained that due to the modification in some last layers for joining with encoder parts, the classifier needs to be trained to get familiar with the merging features. The gaps between training and validation of the two cases are narrow which indicates that there is no big problem of underfitting or overfitting in both cases.

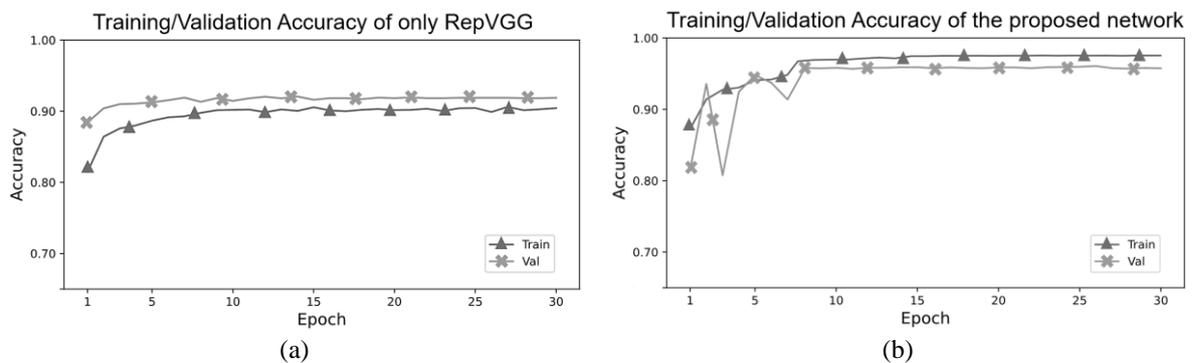


Figure 5. The training progress of (a) only using RepVGG and (b) the proposed network

In addition, to perform a specific analysis in each class, the confusion matrices are depicted in Figure 6. Over the experiments, the mean accuracy is measured by obtaining the number of correct predictions over the total images for each class. Mean accuracy is one of the vital factors to evaluate the performance which indicates the ability to deal with the new data. Generally, compared to the original RepVGG in Figure 6(a), all of the classification cases of the proposed network in Figure 6(b) give better results in terms of accuracy. There is a critical improvement in the case of COVID-19 from 83% to 90% which plays an essential role because our main purpose concentrates on seeking the COVID-19 patients. Besides, compared to the original RepVGG model, our proposed network also provides better outcomes of 97% and 91% in the case of Normal and Viral Pneumonia, respectively.

The summary results of the experiments are described in Table 2. The metrics used for evaluating two models are accuracy, precision, recall, F1 score and time processing per image. The overall accuracy of the joined network is higher than the original network-95.4% and 91.8%, respectively. Precision and recall are known as powerful metrics to measure imbalanced data, especially in the case of collecting Covid class due to the scarcity of samples. High precision implies the high confidence of detection in the corresponding class, while high recall demonstrates the low rate of missing detection of the true class. Our proposed model also has improved results reaching 96.1% for precision and 97.5% for recall, while F1 score is 96.7%. Despite attaching more layers, the processing time in one image of the joined network only lasts over 1/3 times of the original RepVGG case.

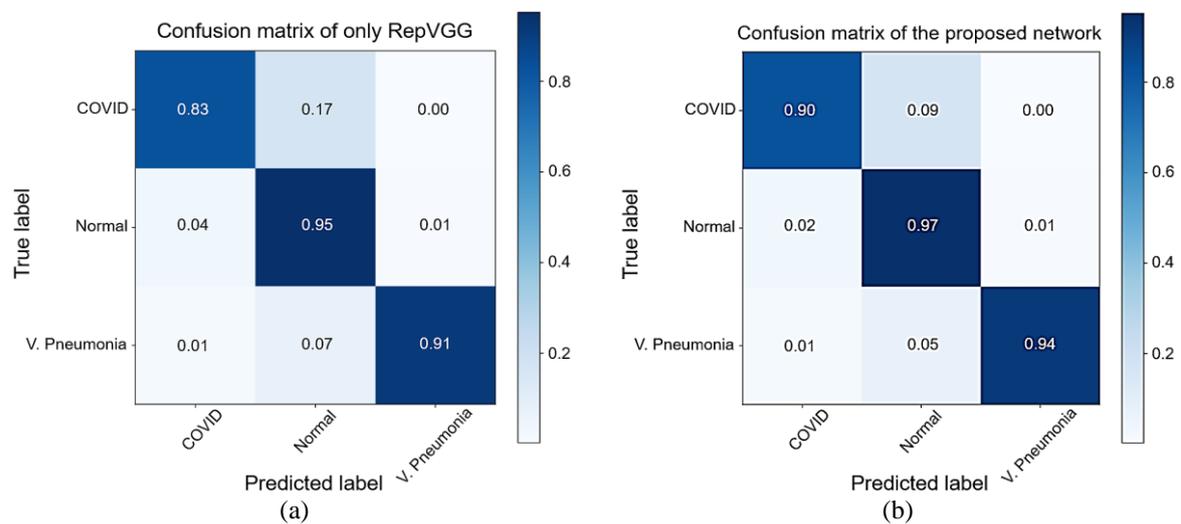


Figure 6. The training progress of (a) only using RepVGG and (b) the proposed network

Table 2. Performance comparison

	Only RepVGG	Proposed Network
Accuracy	91.8%	95.4%
Precision	93.1%	96.1%
Recall	95.2%	97.5%
F1 score	94.2%	96.7%
Processing time per image	0.01(s)	0.014(s)

4. CONCLUSION

In this paper, an improvement of RepVGG model has been presented by integrating with variational data imputation to deal with the classification of COVID-19 based on CXR images. Although the original RepVGG pre-trained reaches a fairly high accuracy, improvement in image-based disease classification still needs to be performed continuously. Inspired by the lung segmentation, the proposed model makes good use of latent features to support the RepVGG for increase the classification performance by concatenating RepVGG and encoder with data imputation. As a result, our proposed model implies that the average accuracy reaches 95.4% which is higher than the initial RepVGG. Besides, other evaluation metrics also give improved results over the original model. This also indicates that the use of deep learning for COVID-19

disease classification could become a reference method for medical therapy to aid in the prevention of corona-virus spread.

REFERENCES

- [1] "WHO coronavirus (COVID-19) dashboard." World Health Organization. Accessed on: Jul. 28, 2021. [Online]. Available: <https://covid19.who.int/>
- [2] M. Hasan Jahid, M. Alom Shahin, and M. Ali Shikhar, "Deep learning based detection and segmentation of COVID-19 pneumonia on chest X-ray image," in *2021 International Conference on Information and Communication Technology for Sustainable Development, ICT4SD 2021-Proceedings*, Feb. 2021, pp. 210–214., doi: 10.1109/ICT4SD50815.2021.9396878.
- [3] 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," *Scientific Reports*, vol. 10, no. 1, Dec. 2020, doi: 10.1038/s41598-020-76550-z.
- [4] M. Farooq and A. Hafeez, "COVID-ResNet: a deep learning framework for screening of COVID19 from radiographs," Mar. 2020, *arXiv:2003.14395*.
- [5] T. M. Quan *et al.*, "XPGAN: X-ray projected generative adversarial network for improving COVID-19 image classification," in *2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI)*, Apr. 2021, pp. 1509–1513., doi: 10.1109/ISBI48211.2021.9434159.
- [6] K. K. Singh and A. Singh, "Diagnosis of COVID-19 from chest X-ray images using wavelets-based depthwise convolution network," *Big Data Mining and Analytics*, vol. 4, no. 2, pp. 84–93, Jun. 2021, doi: 10.26599/BDMA.2020.9020012.
- [7] Y. Oh, S. Park, and J. C. Ye, "Deep learning COVID-19 features on CXR using limited training data sets," *IEEE Transactions on Medical Imaging*, vol. 39, no. 8, pp. 2688–2700, Aug. 2020, doi: 10.1109/TMI.2020.2993291.
- [8] S. Degadwala, D. Vyas, and H. Dave, "Classification of COVID-19 cases using fine-tune convolution neural network (FT-CNN)," in *2021 International Conference on Artificial Intelligence and Smart Systems (ICAIS)*, Mar. 2021, pp. 609–613., doi: 10.1109/ICAIS50930.2021.9395864.
- [9] S. Asif, Y. Wenhui, H. Jin, and S. Jinhai, "Classification of COVID-19 from chest X-ray images using deep convolutional neural network," in *2020 IEEE 6th International Conference on Computer and Communications (ICCC)*, Dec. 2020, pp. 426–433., doi: 10.1109/ICCC51575.2020.9344870.
- [10] E. T. Hastuti, A. Bustamam, P. Anki, R. Amalia, and A. Salma, "Performance of true transfer learning using CNN DenseNet121 for COVID-19 detection from chest X-ray images," in *2021 IEEE International Conference on Health, Instrumentation and Measurement, and Natural Sciences (InHeNce)*, Jul. 2021, pp. 1–5., doi: 10.1109/InHeNce52833.2021.9537261.
- [11] M. E. H. Chowdhury *et al.*, "Can AI help in screening viral and COVID-19 pneumonia?," *IEEE Access*, vol. 8, pp. 132665–132676, 2020, doi: 10.1109/ACCESS.2020.3010287.
- [12] T. Rahman *et al.*, "Exploring the effect of image enhancement techniques on COVID-19 detection using chest X-ray images," *Computers in Biology and Medicine*, vol. 132, May 2021, doi: 10.1016/j.compbiomed.2021.104319.
- [13] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 9351, Springer International Publishing, 2015, pp. 234–241., doi: 10.1007/978-3-319-24574-4_28.
- [14] D. P. Kingma and M. Welling, "Auto-encoding variational bayes," Dec. 2013, *arXiv:1312.6114*.
- [15] D. Liu and G. Liu, "A transformer-based variational autoencoder for sentence generation," in *2019 International Joint Conference on Neural Networks (IJCNN)*, Jul. 2019, pp. 1–7., doi: 10.1109/IJCNN.2019.8852155.
- [16] S. Semeniuta, A. Severyn, and E. Barth, "A hybrid convolutional variational autoencoder for text generation," in *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, 2017, pp. 627–637., doi: 10.18653/v1/D17-1066.
- [17] A. Sagar, "Generate high resolution images with generative variational autoencoder," Aug. 2020, *arXiv:2008.10399*.
- [18] V. Zilvan, A. Ramdan, E. Suryawati, R. B. S. Kusumo, Di. Krisnandi, and H. F. Pardede, "Denosing convolutional variational autoencoders-based feature learning for automatic detection of plant diseases," *3rd International Conference on Informatics and Computational Sciences: Accelerating Informatics and Computational Research for Smarter Society in The Era of Industry 4.0, Proceedings*, Oct. 2019., doi: 10.1109/ICICoS48119.2019.8982494.
- [19] N. T. N. Anh, T. Q. Khanh, N. Q. Dat, E. Amouroux, and V. K. Solanki, "Fraud detection via deep neural variational autoencoder oblique random forest," in *2020 IEEE-HYDCON*, Sep. 2020, pp. 1–6., doi: 10.1109/HYDCON48903.2020.9242753.
- [20] R. Selvan *et al.*, "Lung segmentation from chest x-rays using variational data imputation," May 2020, *arXiv:2005.10052*.
- [21] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Communications of the ACM*, vol. 60, no. 6, pp. 84–90, May 2017, doi: 10.1145/3065386.
- [22] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*, Sep. 2014
- [23] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2016, pp. 770–778., doi: 10.1109/CVPR.2016.90.
- [24] C. Szegedy *et al.*, "Going deeper with convolutions," in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2015, pp. 1–9., doi: 10.1109/CVPR.2015.7298594.
- [25] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jul. 2017, pp. 2261–2269., doi: 10.1109/CVPR.2017.243.
- [26] J. Kolbusz, P. Rozycki, and B. M. Wilamowski, "The study of architecture MLP with linear neurons in order to eliminate the 'vanishing gradient' problem," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 10245, Springer International Publishing, 2017, pp. 97–106., doi: 10.1007/978-3-319-59063-9_9.
- [27] K. J. M. Tarnate, M. Devaraj, and J. C. De Goma, "Overcoming the vanishing gradient problem of recurrent neural networks in the ISO 9001 quality management audit reports classification," *International Journal of Scientific and Technology Research*, vol. 9, no. 3, pp. 6683–6686, 2020
- [28] X. Ding, X. Zhang, N. Ma, J. Han, G. Ding, and J. Sun, "RepVGG: making VGG-style ConvNets great again," in *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2021, pp. 13728–13737., doi: 10.1109/CVPR46437.2021.01352.
- [29] S. Jaeger, S. Candemir, S. Antani, Y.-X. J. Wang, P.-X. Lu, and G. Thoma, "Two public chest X-ray datasets for computer-aided screening of pulmonary diseases," *Quantitative imaging in medicine and surgery*, vol. 4, no. 6, pp. 475–477, 2014

BIOGRAPHIES OF AUTHORS

Kien Trang    holds a Master of Engineering in Electronic Engineering degree from International University, Vietnam National University (VNU) Ho Chi Minh city in 2020. He is currently a researcher in the School of Electrical Engineering at International University – VNU. His current research interests focus on image processing, computer vision and deep learning. He can be contacted at email: tkien@hcmiu.edu.vn.



An Hoang Nguyen    received his B.S. and Master of Electronic Engineering degree from the International University, Vietnam National University (VNU), Ho Chi Minh city, in 2018 and 2020, respectively. Currently, he is a Researcher in the School of Electrical Engineering at International University, VNU. His research interests are signal processing, computer visions, and machine learning applications. He can be contacted at email: nhan@hcmiu.edu.vn.



Long TonThat    obtained M.Sc. and Ph.D. degrees in 2008 and 2012 respectively at The University of Manchester (UK). Since 2014, he joined International University (IU), Vietnam National University Ho Chi Minh City (VNU-HCMC) as a Lecturer in Department of Automation and Control Engineering. His research interests lie mainly in control theory, nonlinear observer design, biological systems, and computational intelligence. He was also involved in some papers in biomedical signal processing. He can be contacted at email: tlong@hcmiu.edu.vn.



Bao Quoc Vuong    received the B.Eng. degree and M.Eng. degree in Electrical Engineering from School of Electrical Engineering, International University, Vietnam National University-Ho Chi Minh City (IU- VNUHCMC), in 2014 and 2017, respectively. He is currently working toward the Ph.D. degree with the Information Science and Technology, Communication and Knowledge Laboratory (Lab-STICC), Department of Electrical Engineering, University of Western Brittany, Brest, France. His main interests are in the areas of signal processing, wireless communication, information theory, Full-duplex transmission. He can be contacted at email: bao.vuong@univ-brest.fr.