COVID19 X-Ray Image Classification using Voting Ensemble CNNs Transfer Learning

Phuwadol Viroonluecha
Department of Information and Communication Technologies
Technical University of Cartagena
30202 Cartagena, Spain
phuwadol.viroonluecha@upct.es

Thanwarat Borisut
Department of Statistics
Ramkhamhaeng University
10240 Bangkok, Thailand
thanwarat.borisut1@gmail.com

Jose Santa
Department of Electronics and Computer Technology
Technical University of Cartagena
30202 Cartagena, Spain
jose.santa@upct.es

***Abstract*— COVID-19 is a novel pandemic and infected COVID-19 people are overgrowing, involving an outbreak that is changing lifestyles around the world. A global issue when trying to contain the propagation of the illness is how to efficiently detect infected people and isolate them. Medical image classification is one of the medical screening tools being used nowadays. Apart from manual inspection, several automatic methods can be applied to exploit artificial intelligence, such as Convolutional Neural Networks (CNNs). Transfer learning can also be applied to make predictive models more effective worldwide, due to the expected small amount of COVID-19 chest X-ray images available. In this paper, we propose the Voting Ensemble CNNs Transfer Learning to recognize the COVID-19 footprint and classify it in a chest X-ray image. The dataset used for training and evaluation was collected from several sources: COVID-19 image data collection and NIH Chest X-ray Dataset of 14 Common Thorax Disease Categories. Our voting ensemble comprises CNNs architectures: ResNet18, ResNet34, and AlexNet. The results illustrate that our model performs with an accuracy of 0.9, recall of 0.825, precision of 0.971, and F1 score of 0.892.**

Keywords*— *coronavirus disease 2019, CXR image, majority voting, deep learning, convolutional neural networks.

I. Introduction

The COVID-19 virus outbreak has created chaos around the world, impacting the economy, societies, political measures, and people's livelihoods. Infected people are rapidly growing globally, as shown in Fig. 1. The number of cases is 17 million people counted at the end of July 2020. Currently, the mortality rate is about 3.90% [1]. However, all of us acknowledge that this viral epidemic is more severe, rapid, and broader than it has been assessed and has murdered too many people around the world [2]. As a result, all frontline public health professionals, including public and private sector workers worldwide, have worked hard to heal this enormous impact. People must harmonize to keep social distancing, especially staying at home to stop the transmission and infection during this challenging time [3]. In parallel, innovations and technologies are identified as useful tools for making prevention and treatment more effective against COVID-19. These include the automatic inspection of chest X-ray images by using image recognition.

Image classification can become one of the first essential mechanisms that hospitals use to screen patients and identify people who are being infected with COVID-19 [4]. In addition to measuring the temperature, controlling, and tracking the data of people at risk of infection, separating COVID-19 from pneumonia using medical or X-rays images is an additional method of detecting illness. In the laboratory, the genetic material of the virus is detected by Reverse Transcription PCR (RT PCR), but it may be insufficient because the number of reagents is limited to the rapidly increasing number of patients. Image classification could support and confirm the results from the laboratory, save time, and perform the lab's results more accurately [5].

Deep Learning is a subset of machine learning which mimics the function of the human brain to do specific tasks. Deep learning could be applied for medical image classification, such as lung cancer and pneumonia [6]. Convolutional Neural Networks (CNNs) is a well-known model for image classification and segmentation [7]. CNNs are a class of deep learning that automatically learns the filters that humans have to do in traditional algorithms. This paper proposes an ensemble approach using CNNs and transfers learning to classify COVID-19 infection in X-ray images derived from medical image libraries. We modified three CNNs architectures, which are ResNet18, ResNet34, and AlexNet. These three models are applied in our voting ensemble.

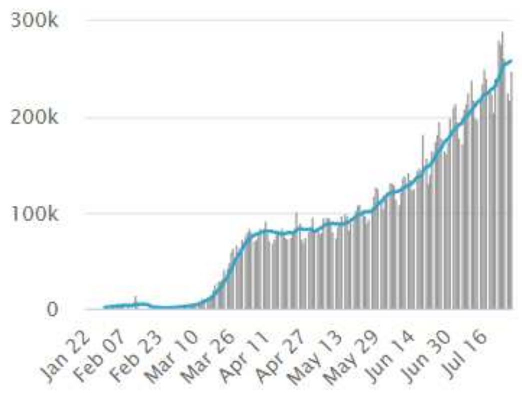


Fig. 1 Global daily new cases of COVID-19 with 7-day moving average (blue line) from January to July 2020 [1].

This paper's structure is the following: after describing the reason for COVID-19 image classification, Section II reviews related works. Section III includes the proposed methodology with the CNN-based voting ensemble transfer learning. In Section IV, we evaluate the proposal discussing the results assessing the suitability of the automatic COVID-19 image classification.

This work was supported by the European Commission, through the PHARAON project (ref. 857188), and the Spanish Government, through the projects ONOFRE-2 (ref. TEC2017-84423-C3-2-P) and Go2Edge (ref. RED2018-102585-T).

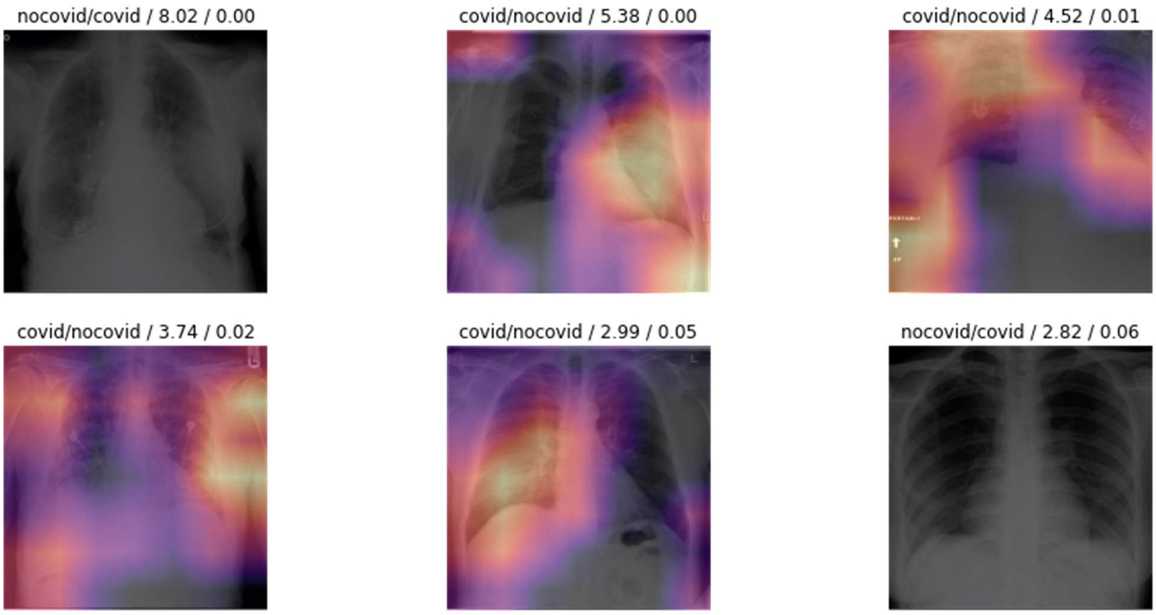


Fig. 2 Top losses images with heatmap.

II. Related Works

The emergence of the new coronavirus encourages the attention of medical image classification research to itself. Various researchers proposed methods to define the disease in X-Ray images such as A. Abbas et al. proposed Decay Transfer and Writing (DeTraC) [8], E. E. Hemdan et al. presented COVIDX-Net, consisting of seven CNNs models [9]. R. M. Pereira et al. propose multi-class and hierarchical classification to detect COVID-19 from normal and pneumonia chest X-ray images [10]. The pre-training model is applied to COVID-CAPS presented by P. Afshar et al. The results of all proposed models acquired high accuracy [11].

From the literature review, we found that all of them used CNNs as a core algorithm. Moreover, ensemble and pre-trained models are interesting and could improve medical image classification accuracy. As a result, we applied CNNs with two approaches: pre-trained and ensemble in our method, described in the next section.

III. Methodology

There are five steps in our research. First, data collection is describing in the dataset section. Data preparation is how we processed images with augmentation, including improvement of augmentation using heatmap from the top loss images. There are two main steps in modeling: separate model training and consolidate three models to the voting ensemble. In the end, the evaluation is presenting with chosen metrics.

A. Dataset

The dataset is taken from the chest x-ray images collected from two different sources: COVID-19 image data collection [12] and NIH Chest X-ray Dataset of 14 Common Thorax Disease Categories [13]. The dataset obtains two classes of images: COVID-19 class, including 200 images, and non-COVID-19 class, with another 200 images. While positive COVID-19 images are in the first class, the second class includes healthy chest images and chest images with regular pneumonia in the same ratio.

B. Data Preparation

The dataset is divided into three parts: 60% for the training set, 20% as the validation set, and the remaining 20% as the testing set. After that, we trained the model with a few epochs and analyzed the incorrectly identified images (images with the highest loss value) using a heatmap, as shown in Fig. 2. This analysis was applied to determine image augmentation parameters. Image augmentation is a regularisation technique to support training models for computer vision. It performs small random transformations but not changing the core of the image. In this experiment, the parameters used in this process are do flip horizontal, do not flip vertical, maximum rotation = 25, and maximum zoom = 1.17. Fig. 3 shows a sample of normal and COVID-19 images extracted from the dataset.

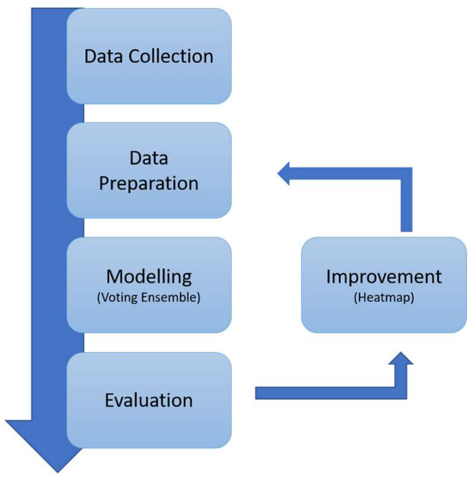


Fig. Methodology of this research

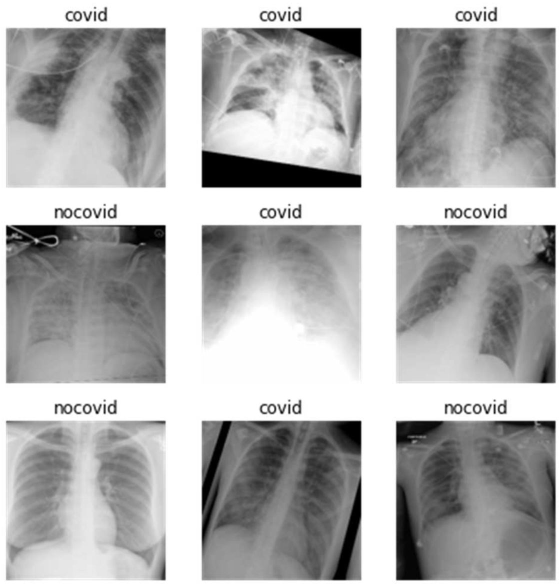


Fig. 3 Example images after image augmentation.

C. Data Modeling

We want to obtain the optimal algorithms in our study. Therefore, we compare 14 pre-trained CNNs models available in the FastAI library:

- ResNet18, ResNet34, ResNet50, ResNet101, ResNet152
- SqueezeNet1_0, SqueezeNet1_1
- DenseNet121, DenseNet169, DenseNet201, DenseNet161
- VGG16_bn, VGG19_bn
- AlexNet

The performance of the 14 models with train loss, valid loss, accuracy, and execution time is shown in the results included in Table I. The selection criteria followed considers the models with the highest accuracy of 3 models, ResNet18, ResNet34, and AlexNet. If there are models with the same accuracy, it will be determined by the valid loss.

Then, we selected the top three models by accuracy and validation loss scores. As presented in Table I, the top three models are AlexNet, ResNet18, and ResNet34. These three algorithms will be the voters in the majority voting system for COVID-19 classification.

- Alexnet is designed by A. Krizhevsky [14]. It is an eight-layer CNNs in which the first five are convolutional layers.
- ResNet is designed by K. He, X. Zhang, S. Ren, and J. Sun [15]. It is a convolutional neural network up to 152 layers deep. We picked ResNet18 and ResNet 34, which have 18 and 34 layers deep, respectively.

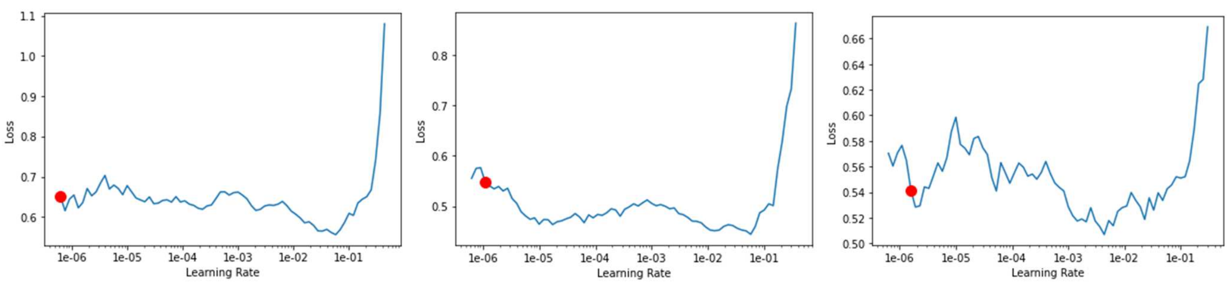
Before training models, we defined the optimal learning rate for each model, as shown in Fig. 4. After the model selection, every model has been trained for another 30 epochs with a predefined learning rate.

TABLE I. PERFORMANCE OF 14 REFERENCE MODELS

Model	Train loss	Valid loss	Accuracy	Time
resnet18	0.69976	0.75995	0.7625	0:22
resnet34	0.60526	0.8179	0.75	0:22
resnet50	0.91096	2.03975	0.6375	0:24
resnet101	1.0916	5.28986	0.6125	0:25
resnet152	1.24716	3.70274	0.5625	0:26
squeezenet1_0	0.79798	0.9064	0.75	0:22
squeezenet1_1	0.91698	3.36416	0.6125	0:22
densenet121	0.89914	2.03398	0.6375	0:25
densenet169	0.63103	1.15615	0.725	0:25
densenet201	0.91675	5.21541	0.5625	0:26
densenet161	0.71805	2.3503	0.5875	0:37
vgg16_bn	0.90201	2.13799	0.55	0:28
vgg19_bn	0.75693	0.88414	0.6	0:24
alexnet	0.90383	0.72282	0.8125	0:22

D. Voting Ensemble

Finally, the majority voting method was created with ResNet18, ResNet34, and AlexNet. The voting system is a simple system designed to improve the prediction performance of our models. Each model provides only one predicted result, outputting a 0 value for the positive COVID-19 cases and a 1 value for the negative COVID-19 cases. This voting schema avoids false positives when the predictions are not in agreement. For instance, if the three models do not have consensus output, the highest frequency value from models will represent the final output [16]. Fig. 5 shows the example of the final result with negative COVID-19 (top) and the final result with positive COVID-19 (bottom).



(a) Resnet18 (b) Resnet34 (c) AlexNet

Fig. 4 Learning rates of Resnet18, Resnet34 and Alexnet

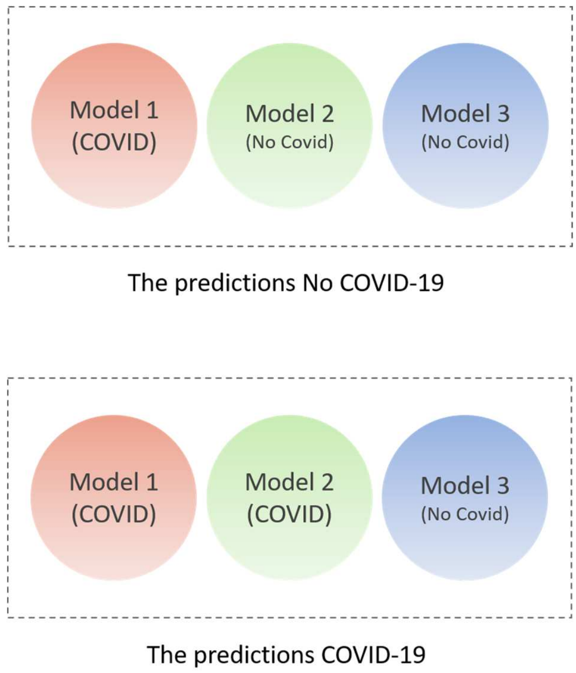


Fig. 5 Example outputs from voting ensemble when the predictions are not in agreement

E. *Evaluation Metrics*

The conventional method to evaluate the classification approach is to calculate metrics from the confusion matrix. The confusion matrix is a table with two columns and two rows that demonstrates the number of true positives – correct prediction of positive COVID-19, true negatives – correct prediction of negative COVID-19, false positives – incorrect prediction of positive COVID-19, and false negatives – incorrect prediction of negative COVID-19. Table II shows the confusion matrix.

TABLE II. CONFUSION MATRIX

	Predicted Positive	**Predicted Negative**
Actual Positive	True Positive (TP)	False Negative (FN)
Actual Negative	False Positive (FP)	True Negative (TN)

We chose the accuracy, precision, recall, and F1 score to evaluate our model performance. The accuracy shows how many correct predictions our model outputs. The formula of accuracy is shown as:

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \quad (1)$$

Precision is a measurement that gives a ratio of correct positive predictions as compared to the positive predictions (true and false). The formula is shown as:

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

The recall is known as the sensitivity ratio of correct positive predictions as compared to the addition of true positives and false negatives. The formula is shown as:

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

F1 score is a combination of recall and precision. The formula is shown as:

$$F1\ score = 2 * \frac{Recall * Precision}{Recall + Precision} \quad (4)$$

All metrics range from 0 to 1, being 1 the best possible score.

TABLE III. PERFORMANCE OF PROPOSED MODEL WITH ITS VOTERS

Metrics	**Voting ensemble**	**ResNet18**	**ResNet34**	**AlexNet**
Accuracy	**0.9**	0.825	0.8375	0.8625
Recall	0.825	0.725	**0.85**	0.75
Precision	**0.970588**	0.90625	0.829268	0.967742
F1-Score	**0.891892**	0.805556	0.839506	0.84507

TABLE IV. COMPARISON OF CONFUSION MATRICES FROM THE FOUR MODELS

	PREDICTED							
	Voting Ensemble		**ResNet18**		**ResNet34**		**AlexNet**	
	COVID	**NOCOVID**	**COVID**	**NOCOVID**	**COVID**	**NOCOVID**	**COVID**	**NOCOVID**
ACTUAL COVID	33	7	29	11	34	6	30	10
ACTUAL NOCOVID	1	39	3	37	7	33	1	39

IV. RESULTS

We applied the specific learning rates indicated in Section III. The learning rates for training ResNet18, ResNet34, and AlexNet models are e^{-7}, e^{-6}, and e^{-6}, respectively.

Table III shows that the voting ensemble model based on three CNNs voters performed well, with true negatives results equal to the ones obtained with AlexNet. Both the voting ensemble and the AlexNet model classified for non-COVID-19 class correctly at 97.5%, with only one incorrect prediction. As seen further in Table IV, ResNet34 overperformed our proposed model when attending to recall, presenting only one prediction different. The ranking for positive COVID-19 class prediction (TP cases) is ResNet34, our voting ensemble, AlexNet, and ResNet18. Furthermore, the ranking for negative COVID-19 class prediction (TN cases) is, first, our proposed model and AlexNet, and then ResNet18 and ResNet34.

As a result, the voting ensemble performed better than the other models in three metrics. The voting model obtained 90% accuracy, 82.5% recall, 97.06% precision and 89.19% F1-score. AlexNet, ResNet34, and ResNet18 performed at 2nd, 3rd, and 4th positions by accuracy, respectively.

V. CONCLUSION

The unexpected pandemic caused by COVID-19 disease is a global problem that all countries around the world are fighting. Numerous tools are developed, and some are still under development, to fight against the virus to prevent new cases and support treatments. Machine learning is identified as a useful supporting tool in the area of automatic medical image classification. It is a supporting tool for medical professionals to detect COVID-19 using chest X-ray images.

In this paper, we propose to use CNNs to identify COVID-19 cases from X-ray images. Our voting ensemble CNNs transfer learning method is able to efficiently classify COVID-19 from medical images. The voting system consists of ResNet18, ResNet34, AlexNet as its voters. Our proposed model performed better than its individual voting members with an accuracy of 90%, recall of 82.5%, the precision of 97.06%, and F1-score 89.19%. The F1-score of our proposed model performed over a multi-class approach of [11] and performed equivalently to a hierarchical classification of [11] and healthy image classification of [9]. This means that it is an efficient supporting tool to support common PCR tests when X-ray imaging is possible, and a number of patients must be checked. Accuracy detected is, in fact, in the same range of PCR tests currently being used. Moreover, this system is able to detect patients not suffering from COVID-19, but at risk of developing the illness critically, due to other respiratory problems.

As a part of our future work, we plan to extend our dataset to better adapt our prediction models and improve accuracy and reliability. Also, we intend to pre-train the model with other chest X-ray images to improve performance because pre-trained models in FastAI is based on ImageNet but not X-Ray images. Finally, we plan to develop and experiment the system with a more complex and weighted (discounted) voting system to get better prediction results.

ACKNOWLEDGMENT

We would like to express my very great appreciation to Associate Professor Dr. Esteban Egea López for his valuable and constructive suggestions for this research work. His willingness to give his time so generously has been very much appreciated.

REFERENCES

[1] Worldometer, "Coronavirus Update (Live) from COVID-19 Virus Pandemic." Accessed: Jul. 30, 2020. [Online]. Available: https://www.worldometers.info/coronavirus

[2] WHO, "Coronavirus Disease (COVID-19) Dashboard," Accessed: Jul. 30, 2020. [Online]. Available: https://covid19.who.int/

[3] WHO, "Argentina: There is no economy without health," Accessed: Jul. 30, 2020. [Online]. Available: https://www.who.int/news-room/feature-stories/detail/argentina-there-is-no-economy-without-health

[4] M. Anthimopoulos, S. Christodoulidis, L. Ebner, A. Christe, and S. Mougiakakou, "Lung pattern classification for interstitial lung diseases using a deep convolutional neural network," *IEEE Transactions on Medical Imaging*, vol. 35, no.5, pp. 1207–1216, May 2016, doi: 10.1109/TMI.2016.2535865.

[5] LabCorp, "ACCELERATED EMERGENCY USE AUTHORIZATION (EUA) SUMMARY COVID-19 RT-PCR TEST (LABORATORY CORPORATION OF AMERICA)," Accessed: Jul. 24, 2020. [Online]. Available: https://www.fda.gov/media/136151/download

[6] J. Kuruvilla and K. Gunavathi, "Lung cancer classification using neural networks for CT images," *Computer Methods and Programs in Biomedicine*, vol.113, no.1, pp. 202–209, January 2014, doi:10.1016/j.cmpb.2013.10.011.

[7] L. Wang and A. Wong, "COVID-Net: A Tailored Deep Convolutional Neural Network Design for Detection of COVID-19 Cases from Chest X-Ray Images," 2020. [Online]. Available: arXiv:2003.13815.

[8] A. Abbas, M. M. Abdelsamea, M. M. Gaber, "Classification of COVID-19 in chest X-ray images using DeTraC deep convolutional neural network," 2020. [Online]. Available: arXiv:2003.09871.

[9] E. E. Hemdan, M. A. Shouman, M. E.l Karar, "COVIDX-Net: A Framework of Deep Learning Classifiers to Diagnose COVID-19 in X-Ray Images," 2020. [Online]. Available: arXiv:2003.11055.

[10] R. M. Pereira, D. Bertolini, L. O. Teixeira, C. N. Silla, and Y. M.G. Costa, "COVID-19 identification in chest X-ray images on flat and hierarchical classification scenarios," *Computer Methods and Programs in Biomedicine*, vol.194, pp. 1-18, 2020.

[11] P. Afshar, S. Heidarian, F. Naderkhani, A. Oikonomou, K. N. Plataniotis, A. Mohammadi, "COVID-CAPS: A Capsule Network-based Framework for Identification of COVID-19 cases from X-ray Images," 2020. [Online]. Available: arXiv:2004.02696.

[12] *COVID-19 Image Data Collection: Prospective Predictions Are the Future*, J. P. Cohen, P. Morrison, L. Dao, K. Roth, T. Q. Duong, and M. Ghassemi, 2020, [Online]. Available: https://github.com/ieee8023/covid-chestxray-dataset

[13] *NIH Chest X-ray Dataset of 14 Common Thorax Disease Categories*, National Institutes of Health, Sep. 27, 2017, [Online]. Available: https://nihcc.app.box.com/v/ChestXray-NIHCC

[14] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," In *Advances in neural information processing systems*, 2012, pp. 1097-1105.

[15] K. He, X. Zhang, S. Ren, and J. Sun, "Identity mappings in deep residual networks," In *European conference on computer vision*, 2016, pp. 630-645.

[16] A. M Rossetto, W. Zhou, "Ensemble Convolution Neural Network with a Simple Voting Method for Lung Tumor Detection," In *Proceedings of the 8th ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics*, 2017 pp. 729–734, doi:10.1145/3107411.310