Automated Detection of COVID-19 From CXR Image Using Voting Ensemble CNNs Transfer Learning

Phuwadol Viroonluecha, Thanwarat Borisut and Jose Santa

***Abstract*— Controlling the spread and investigating the COVID-19 coronavirus outbreak is a big challenge, attending to the exponentially growing number of cases. The high number of suspected cases causes medical personnel to work hard with time restrictions to perform tests. Although patients are mainly screened with Polymerase Chain reaction (PCR) tests, the chest X-ray is an effective method for detecting infections in patients, including COVID-19. In this paper, we propose to automate this virus detection by image processing using chest X-rays, using Convolutional Neural Networks (CNNs). We introduce the Voting Ensemble CNNs Transfer Learning technique, which is applied with pre-trained weights from another dataset to solve the training data insufficiency issue that can lead to low classification accuracy and strengthen a prediction result. The training and evaluation datasets were collected from several sources, including a COVID-19 image data collection and NIH Chest X-ray Dataset of 14 Common Thorax Disease Categories. Several data preparation techniques such as data augmentation and heatmap were used to improve the prediction during the experiment. After comparing 14 CNNs algorithms' performance, we chose five of the finest algorithms for the voting ensemble to build 3-voter, 4-voter and 5-voter ensembles. The best performer is the 3-voter ensemble which its voters are ResNet18, ResNet34 and AlexNet. The results show that our model outputs an accuracy of 0.9, recall of 0.825, precision of 0.971, and F1 score of 0.892.**

***Index Terms*—coronavirus disease 2019, chest X-ray 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

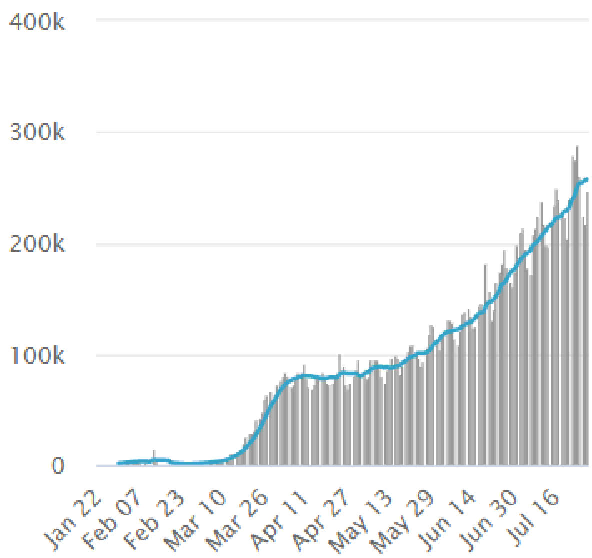


Fig. 1: Global daily new cases of COVID-19 with 7-day moving average (blue line) from January to July 2020 [1].

Phuwadol Viroonluecha, Department of Information and Communications Technology, Technical University of Cartagena, 30202 Cartagena, Spain, email: phuwadol.viroonluecha@upct.es.

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

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

Manuscript received February 8, 2021; accepted March 19, 2021; revised April 24, 2021; published online April 26, 2021

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.

The contribution of this paper is an automatic method to identify an image of the new type of respiratory disease COVID-19 based on the different numbers of voting ensemble members with pre-trained various CNNs algorithms. Furthermore, the heatmap analysis is applied for finding optimal parameters for image augmentation. We modified five CNNs architectures, which are ResNet18, ResNet34, AlexNet, SqueezeNet1.0, and DenseNet169. These five models are applied in our voting ensembles.

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.

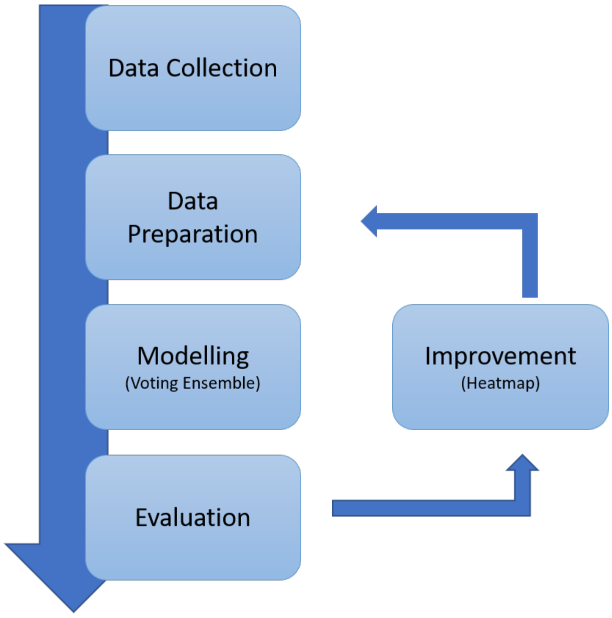


Fig. 2: Methodology of this research.

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 as shown in Fig. 2. The experiment was conducted in Google Colab with GPU runtime. 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 five models to the voting ensembles. 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 Grad-CAM heatmaps, as shown in Fig. 3. These heatmaps are generated using the final layer of Convolutional Neural Networks. 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, we analyzed heatmaps then attempted several image augmentation parameters. Finally, the optimal parameters used in this process are do flip horizontal, do not flip vertical, maximum rotation = 25, and maximum zoom = 1.17. Fig. 4 shows a sample of normal and COVID-19 images extracted from the dataset.

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

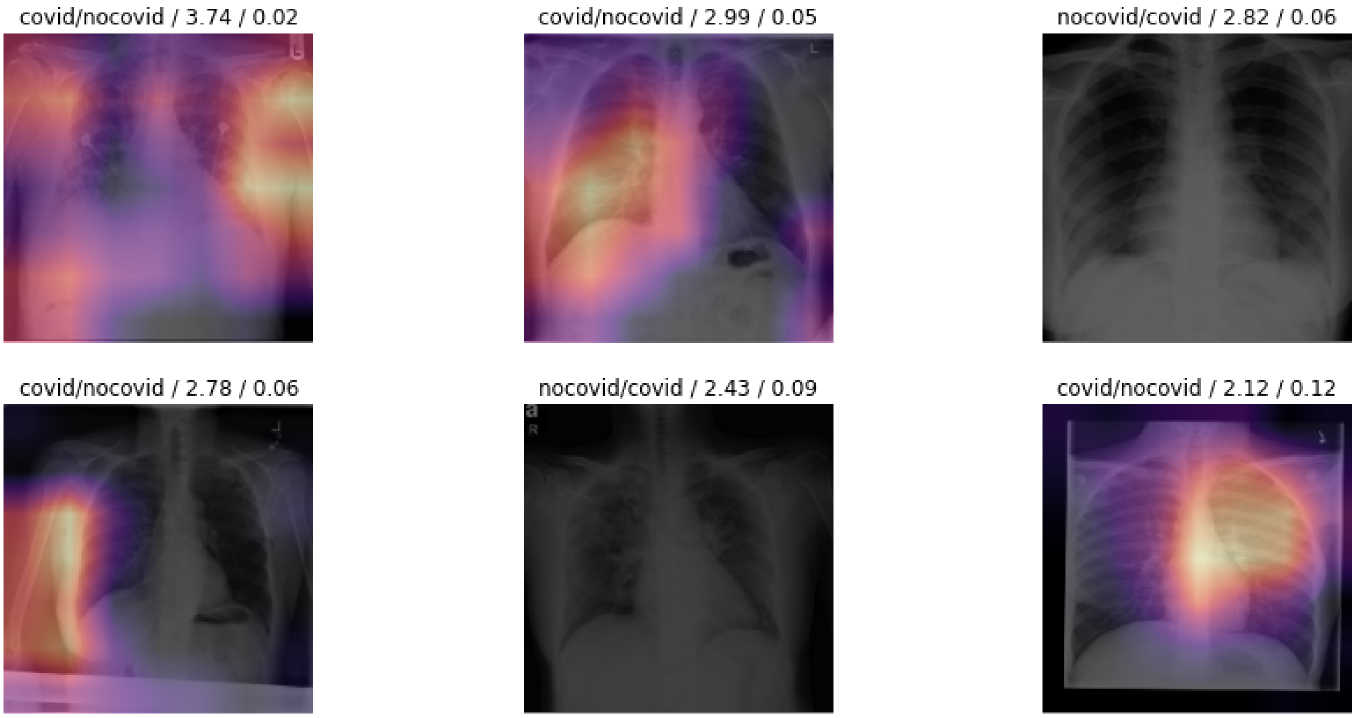


Fig. 3: Top losses images with heatmap

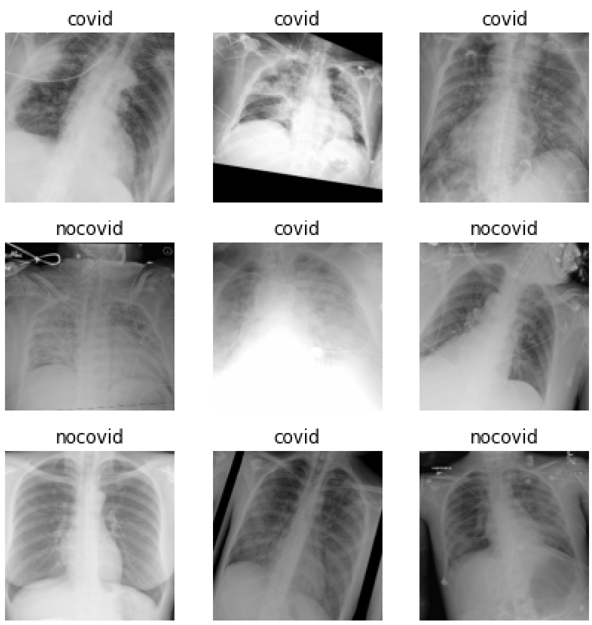


Fig. 4: Example images after image augmentation.

- 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 5 models. If there are models with the same accuracy, it will be determined by the valid loss.

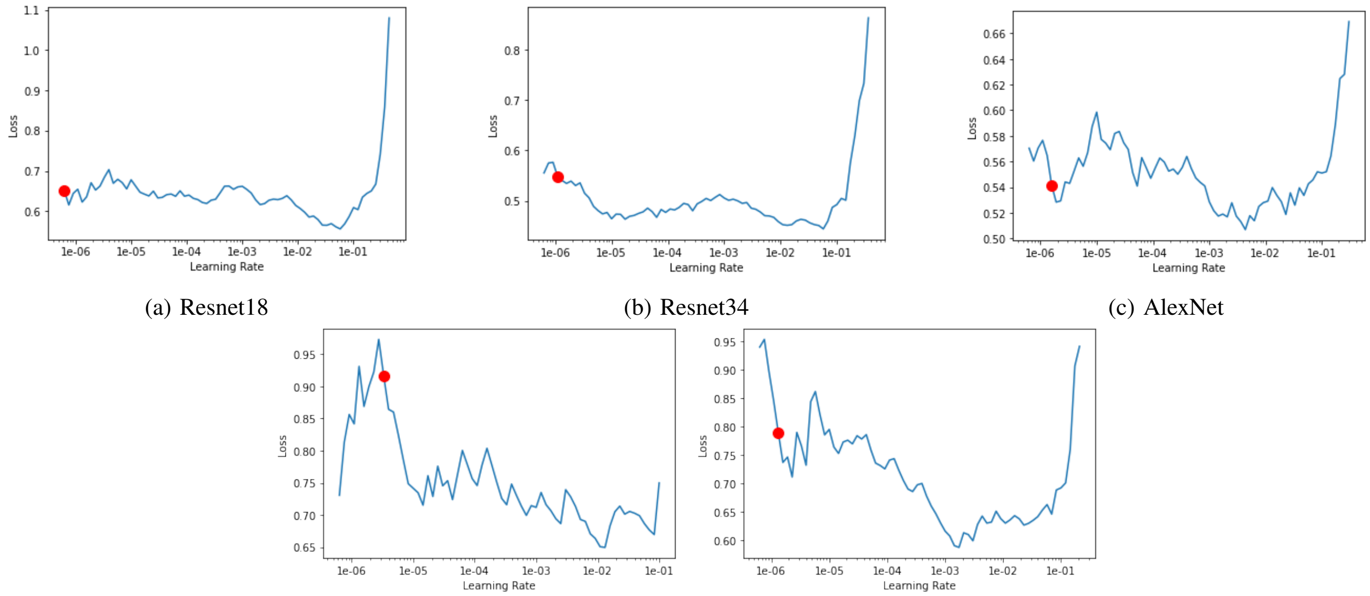
Then, we selected the top five models by accuracy and validation loss scores. As presented in Table I, the top five models are AlexNet, ResNet18, ResNet34, SqueezeNet1.0, and DenseNet169. These five algorithms will be the voters in the majority voting systems for COVID-19 classification which consist of 3-voter, 4-voter, and 5-voter ensembles.

- 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.
- SqueezeNet is a convolutional neural network that is 18 layers deep. It was designed by researchers from DeepScale and UC Berkeley, and Stanford University [16]. Its small size benefits computer memory and transferring in a computer network.
- G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger proposed densely connected convolutional networks, a.k.a. DenseNet. Each layer of DenseNet obtains inputs from all preceding layers and passes on its feature-maps to all subsequent layers [17]. So the network can be thinner and compact.

Before training models, we defined the optimal learning rate for each model, as shown in Fig. 5. After the model selection, every model has been trained for another 30 epochs with a predefined learning rate.

D. Voting Ensemble

Finally, the majority voting method was created with AlexNet, ResNet18, ResNet34, SqueezeNet1.0, and



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

(d) SqueezeNet1.0 (e) DenseNet169

Fig. 5: Learning rates of ResNet18, ResNet34, AlexNet, SqueezeNet1.0 and DenseNet169

TABLE I: Performance of 14 Reference Models

Model	Train loss	Valid loss	Accuracy	Time
resnet18	0.6998	0.7600	0.7625	0:22
resnet34	0.6053	0.8179	0.7500	0:22
resnet50	0.9110	2.0400	0.6375	0:24
resnet101	1.0916	5.2899	0.6125	0:25
resnet152	1.2472	3.7027	0.5625	0:26
squeezenet1.0	0.7980	0.9064	0.7500	0:22
squeezenet1.1	0.9170	3.3641	0.6125	0:22
densenet121	0.8991	2.0340	0.6375	0:25
densenet169	0.6310	1.1562	0.7250	0:25
densenet201	0.9168	5.2154	0.5625	0:26
densenet161	0.7181	2.3503	0.5875	0:37
vgg16_bn	0.9020	2.1380	0.5500	0:28
vgg19_bn	0.7569	0.8841	0.6000	0:24
alexnet	0.9038	0.7228	0.8125	0:22

TABLE II: Confusion Matrix

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

DenseNet169. We compared three voting ensembles, which are 3-voter, 4-voter, and 5-voter systems. The 3-voter consists AlexNet, ResNet18, ResNet34 as the voters. We added SqueezeNet1.0 to the 4-voter ensemble together with the previous three algorithms. The final one consists of DenseNet169, besides algorithms in the 4-voter system. 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 [18]. Fig. 6 shows the example of the final result with negative COVID-19 (top) and the final result with positive COVID-19 (bottom).

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.

We chose the accuracy, precision, recall, and F1 score to evaluate our model performance. The accuracy shows how many correct predictions our model outputs. Precision is a measurement that gives a ratio of correct positive predictions as compared to the positive predictions (true and false). The recall is known as the sensitivity ratio of correct positive predictions as compared to the addition of true positives and false negatives. F1 score is a combination of recall and

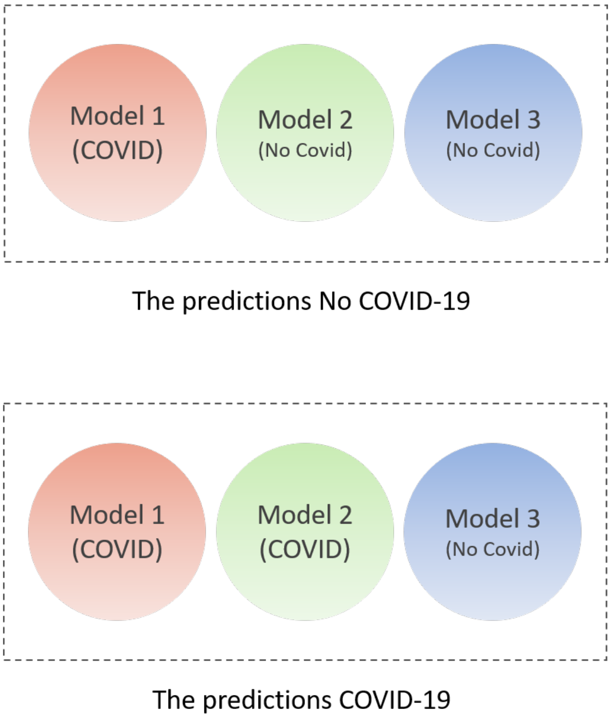


Fig. 6: Example outputs from voting ensemble when the predictions are not in agreement.

TABLE III: Performance of Proposed Models with Its Voters

Metrics	Accuracy	Recall	Precision	F1-Score
3-Voter ensemble	0.9000	0.8250	0.9710	0.8919
4-Voter ensemble	0.8500	0.7500	0.9375	0.8333
5-Voter ensemble	0.8875	0.8000	0.9697	0.8767
ResNet18	0.8250	0.7250	0.9063	0.8056
ResNet34	0.8375	0.8500	0.8293	0.8395
AlexNet	0.8625	0.7500	0.9677	0.8451
SqueezeNet1.0	0.8125	0.7500	0.8571	0.8000
Densenet169	0.8000	0.7250	0.8529	0.7838

precision. Note that all metrics range from 0 to 1, being 1 the best possible score.

IV. Results

We applied the specific learning rates indicated in Section III. The learning rates for training ResNet18, ResNet34, AlexNet, SqueezeNet1.0 and Densenet169 models are 6×10^{-7}, 1×10^{-6}, 1×10^{-6}, 5×10^{-6} and 1×10^{-6} respectively.

As a result, in Table III, the 3-voter ensemble performed better than the other models in three metrics. The 3-voter model obtained 90% accuracy, 82.5% recall, 97.06% precision and 89.19% F1-score. The 5-voter, AlexNet, and 4-voter performed at 2nd, 3rd, and 4th positions by accuracy, respectively. Another highlight in this table is ResNet34 overperformed our proposed model when attending to recall.

TABLE IV: Comparison of Confusion Matrices

PREDICTED		ACTUAL COVID	ACTUAL NOCOVID
3-Voter ensemble	**COVID**	33	1
	NOCOVID	7	39
4-Voter ensemble	**COVID**	30	2
	NOCOVID	10	38
5-Voter ensemble	**COVID**	32	1
	NOCOVID	8	39
ResNet18	**COVID**	29	3
	NOCOVID	11	37
ResNet34	**COVID**	34	7
	NOCOVID	6	33
AlexNet	**COVID**	30	1
	NOCOVID	10	39
Squeezenet1_0	**COVID**	30	5
	NOCOVID	10	35
Densenet169	**COVID**	29	5
	NOCOVID	11	35

Table IV shows that the voting ensemble model based on three CNNs voters performed well, with true negatives results equal to the ones obtained with AlexNet. The voting ensemble and the AlexNet model were classified for non-COVID-19 class correctly at 97.5%, with only one incorrect prediction. The ranking for positive COVID-19 class prediction (TP cases) is ResNet34, 3-voter ensemble, and 5-voter ensemble. Furthermore, the ranking for negative COVID-19 class prediction (TN cases) is, first, our proposed 3-voter model, 5-voter model and AlexNet, and then 4-voter model and ResNet18 with 39, 38 and 37 TNs respectively.

Precision represents the proportion of the models' predictions of the COVID-19 class where the COVID-19 class is actually present, which is relatively high precision in this experiment. Recall of the non-COVID-19 class is more than recall of COVID-19, meaning that this model has better predictions of the non-COVID-19 than the COVID-19 class. In other words, we may see that our model is more inclined to non-COVID-19. So the model might have a greater chance of predicting the non-COVID-19 class. Overall, the F1 score of 89.19% confirmed that the 4-voter ensemble model correctly identifies real threats and is not disturbed by false alarms.

V. Conclusion

Pre-screening by computational image analysis can play an essential role in fighting with COVID-19 spreading. We believe that deep learning can be used as a computer-aided tool for diagnosing this virus disease. This paper proposes voting ensemble CNNs transfer learning method to effectively classify COVID-19 from medical images. We proposed three voting systems: 3-voter, 4-voter, and 5-voter ensemble which consist of ResNet18, ResNet34, AlexNet, SqueezeNet1.1 and DenseNet169 as its voters. Our proposed models performed better than its separately voting members, with over 90 percent of accuracy and precision from the 3-voter model. The F1-score of our proposed 3-voter model, with 89.19%, outperforms a multi-class approach [11], which has equivalent results to a hierarchical classification [11] and a healthy image classification [9]. Our results show the solution as an effective support tool when X-ray imaging is available, reducing the dependency on standard PCR or blood tests. This method can also detect patients who do not suffer from COVID-19 but are at risk of developing a critical condition due to other respiratory problems.

As part of our future work, we plan to expand our data set 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 the pre-trained models in FastAI are based on ImageNet but not X-ray images. Our final aim is to develop and experiment with a more complex and weighted (discounted) voting system in order to get better predictive results.

ACKNOWLEDGMENT

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).

We would also like to appreciate the support of Prof. Dr. Esteban Egea López for his willingness and valuable and constructive suggestions for this research work.

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] 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,” [Online]. Available: arXiv:1602.07360, 2016.

[17] G. Huang, Z. Liu, K. Q. Weinberger and L. van der Maaten, ”Densely connected convolutional networks”, [Online]. Available: arXiv:1608.06993, 2016.

[18] 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

Phuwadol Viroonluecha is a Ph.D. student and early-stage researcher in GIRTEL, Department of Information and Communications Technology at Technical University of Cartagena, Spain. Currently, he works on the application of machine learning for blood glucose control in type I diabetic patients. His research interests are framed in the application of data science, AI and IoT technologies in e-health and industry.

Thanwarat Borisut is a bachelor’s student in Department of Biotechnology at Ramkhamhaeng University, Thailand. Her research interests are biomedical engineering, biotechnology, developmental biology, machine learning, information technology, and image processing.

Dr. José Santa is a Senior Research Fellow (Ramón y Cajal) at Department of Electronics, Computer Technology and Projects, at Technical University of Cartagena, Spain. He works in mobile communications, intelligent transportation systems, next-generation networks, cyber-physical systems and Internet of Things (IoT), with special emphasis on real prototypes and evaluation.