2020 IEEE REGION 10 CONFERENCE (TENCON)
Osaka, Japan, November 16-19, 2020

# A Multi-Model Based Ensembling Approach to Detect COVID-19 from Chest X-Ray Images

Oishy Saha[1], Jarin Tasnim[1], Md. Tanvir Raihan[1], Tanvir Mahmud[1],
Istak Ahmmed[2] and Shaikh Anowarul Fattah, Senior Member, IEEE[1]
[1]Department of Electrical and Electronic Engineering (EEE)
Bangladesh University of Engineering and Technology (BUET), Dhaka - 1205, Bangladesh.
[2]PrimeSilicon Technology
10241 Santa Clara Avenue, Cupertino, CA-95014, USA

***Abstract*—Since the onset of COVID-19, radiographic image analysis coupled with artificial intelligence (AI) has become popular due to insufficient RT-PCR test kits. In this paper, an automated AI-assisted COVID-19 diagnosis scheme is proposed utilizing the ensembling approach of multiple convolutional neural networks (CNNs). Two different strategies have been carried out for ensembling: A feature level fusion-based ensembling method and a decision level ensembling method. Several traditional CNN architectures are tested and finally in the ensembling operation, MobileNet, InceptionV3, DenseNet201, DenseNet121 and Xception are used. To handle the computational complexity of multiple networks, transfer learning strategy is incorporated through ImageNet pre-trained weight initialization. For feature-level ensembling scheme, global averages of the convolutional feature maps generated from multiple networks are aggregated and undergo through fully connected layers for combined optimization. Additionally, for decision level ensembling scheme, final prediction generated from multiple networks are converged into a single prediction by utilizing the maximum voting criterion. Both strategies perform better than any individual network. Outstanding performances have been achieved through extensive experimentation on a public database with 96% accuracy on 3-class (COVID-19/normal/pneumonia) diagnosis and 89.21% on 4-class (COVID-19/normal/viral pneumonia/bacterial pneumonia) diagnosis.**

***Index Terms*—COVID-19, Deep Learning, Ensembling, Transfer Learning**

## I. Introduction

Since its origin, COVID-19 has become a great threat to mankind taking about 0.7 million lives globally. Socio-economic life has also been affected due to its high contagion. Therefore, early detection of COVID-19 can contribute to minimize further transmission and alleviate the pressure of the medical sector. In recent times, deep learning is applied extensively in image processing based medical studies. COVID-19 research is of no exception. Khan et al. [1] came up with the model Coronet which is based on the Xception architecture and initialized the model with the weights of ImageNet to perform binary and multi-class classification. Apostopolous et al. [2] also explored the transfer learning method and compared five CNN models namely VGG19, MobileNetV2, Inception, Xception and Inception ResNet V2 pre-trained by ImageNet data for COVID detection. In these two studies, any particular CNN Model was used at a time and the model weights were initialized with ImageNet weights. However, from the study of Apostopolous et al. [2] the comparison of the performances of various models can be obtained. Ozturk et al. [3] designed a model named DarkCovidNet for both binary classes (COVID vs no-Findings) and multi-class classification (COVID vs no-Findings vs pneumonia). As they used a custom network, they couldn't apply any transfer learning strategies utilizing ImageNet weights. Mahmud et al. [4] designed a model named CovXNet which utilizes depthwise convolution with varying dilation rates. Though a large number of chest X-ray images of normal and viral/bacterial were used to train the model, no initiatives were taken to extract classwise diversified features. N. Narayan Das et al. [5] approached a simple transfer learning technique to detect COVID-19 from Xception architecture, pre-trained by large datasets. Ucar et al. [6] proposed COVIDiagnosis-Net, a tuned SqueezeNet architecture, pre-trained by ImageNet weights with Bayesian optimization additive. Despite high accuracy, their dataset consisted of only 76 COVID X-ray images that went through augmentation before training. Das et al. [7] proposed a model based on InceptionV3 architecture. In their study, they truncated the model and took three inception modules and one grid size-reduction block from the beginning followed by a max-pooling and global average pooling layer and SVM was used as a classifier. Vaid et al. [8] developed a deep learning-based model, modification of VGG-19, previously trained on the ImageNet dataset. Togacar et al. [9] used models like MobileNetV2 and SqueezeNet and SVM classifiers. However, they had a small dataset and did not perform challenging 4-class classification. Various methods have individual advantages and disadvantages. One possible way could be to consider some efficient models together to achieve better performance for three-class and four-class classification.

The main objective of this study is to explore the advantages and opportunities of different traditional well-known architectures as well as to converge multiple networks into an integrated one for achieving optimum performance. Two basic ensembling strategies have been carried out for the integration process, which are feature-based ensembling scheme and decision-based ensembling scheme. The feature based scheme considers the feature space of multiple networks for group optimization while the decision based scheme converges the decision of multiple networks using maximum voting criterion. These ensembling strategies provide additional opportunities to utilize the architectural diversity of multiple networks that provides considerable improvement of performances consistently compared to other individual

network.

## II. Methodology

Two major problems in dealing with the chest X-ray image based COVID-19 detection scheme are: (1) lack of chest X-ray images with confirmed COVID-19 cases and (2) high similarity in the overall appearance of the chest X-ray images confirmed as COVID-19, viral pneumonia and bacterial pneumonia. In order to overcome these problems, in most of the existing deep learning based COVID-19 detection schemes, generally a suitable conventional architecture or its variants are used. In this research, numerous deep convolutional neural network (CNN) architectures have been tested to detect COVID-19 X-ray images. Moreover, various modifications on those deep CNN architectures have also been analyzed. However, it is very difficult to justify the variants of an existing architecture against the variation in classification performance and a certain type of modification may not found suitable for in all cases. It is well known that the state of the deep CNN architectures are very well optimized with respect to computational complexity as well as the classification performance. Hence in this research work, an ensembling approach is introduced for COVID-19 X-ray image detection by utilizing multiple state of the art deep CNNs, MobileNet, InceptionV3, DenseNet201, DenseNet121 and Xception. It is to be noted that the number of CNNs can easily be varied but in this paper, above five CNN models are chosen to demonstrate the performance.

For the purpose of ensembling the CNN models, two schemes are designed:

- Ensembling of Multiple CNN Models by Feature Concatenation
- Ensembling of Multiple CNN Models by Decision Fusion

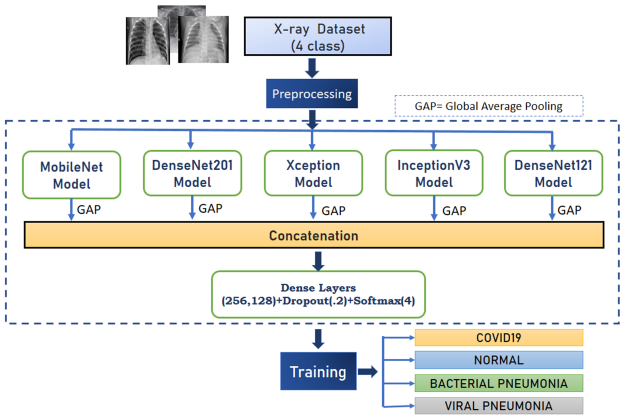


Fig. 1. Ensembling of models using Feature Concatenation

In Fig. 1, with the help of a simple block diagram, the proposed multiple CNN models ensembling scheme using feature concatenation is presented. In order to handle the computational complexity of multiple networks, a transfer learning strategy is incorporated where the ImageNet pre-trained weights are utilized for the purpose of initialization. It can be observed that the five deep CNN models are trained on chest X-ray images of various classes. Four major classes

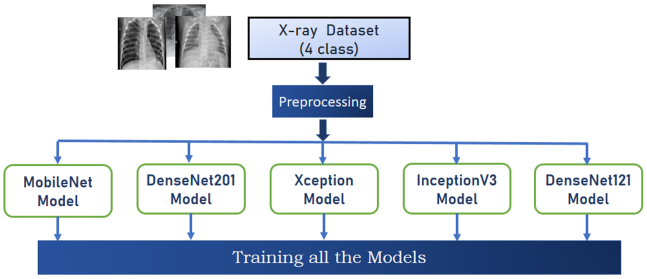


Fig. 2. Training approach for ensembling of models using Decision Fusion

of chest X-ray images are considered here, such as COVID-19, normal, viral pneumonia, and bacterial pneumonia. The objective here is to extract four class separable features. The extracted features from each CNN architecture, after the Global Average Pooling layer, are concatenated. This concatenation of extracted features from different models is expected to improve the separability among various classes. Two dense layers, a dropout and a *softmax* activated dense layer are added after concatenation to ensure further training after the concatenation. It is to be noted that the convolutional feature maps are aggregated and passed through a shared stack of fully connected layers for combined optimization. For 3 class (COVID-19/ normal/ pneumonia) classification, same strategy is followed.

In Fig. 2, the training approach for the ensembling of multiple CNN models by decision fusion is shown. In this case, the deep CNN models are first trained individually using the 4 class X-ray images in the training phase.

During the testing phase of the chest X-ray images, the implementation of the decision fusion approach comes into account. Using each of the trained CNN models, predictions are made for each of the test images. By decision fusion, we mean taking the "mode" of the predictions made by the five models. The predictions produced by various models are converged into single prediction by utilizing the maximum voting criterion. The testing approach for the decision fusion strategy is shown in Fig. 3.

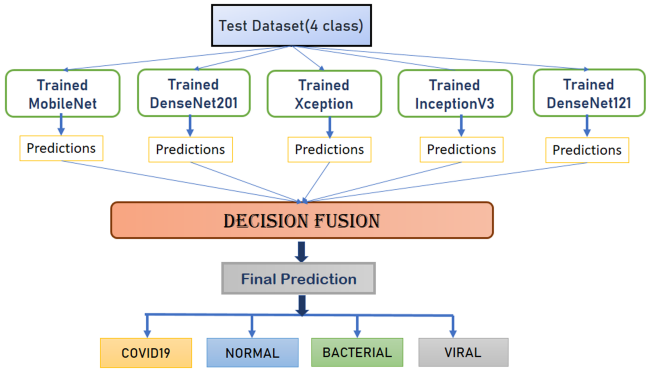


Fig. 3. Testing approach for ensembling of models using Decision Fusion

For 3 class (COVID-19/ normal/ pneumonia) classification, the same training and testing approaches are followed as described above. In this approach, all the predictions made

by different models are taken into account than depending only on the prediction made by an individual model.

### A. Pre-processing

The balanced chest X-ray (CXR) datasets used for feature concatenation and decision fusion are kept the same. The chest X-ray images are passed through minimal preprocessing for easier implementation in the testing phase. The images are only reshaped to a uniform size before further processing with the ensembled deep neural network.

### B. Training approach for ensembling by feature concatenation strategy

As shown in the structure of this approach in Fig. 1, five networks i.e MobileNet, DenseNet201, Xception, InceptionV3, DenseNet121 are trained on balanced chest X-ray dataset. All the networks are initialized with 'ImageNet' weights and these weights are updated in the training process. This transfer learning approach provides better initialization and faster convergence. After concatenation, the newly added classification head contains a dropout(.2) layer followed by two dense layers with 128 and 256 neurons, respectively. The dense layers are associated with *relu* activation. For the final classification, *softmax* activation is used. The final network is trained for 50 epochs. Also, adaptive momentum (Adam) optimization with learning rate $0.0001$ and minimum learning rate $0.00001$ are used for training.

### C. Training approach for ensembling by decision-fusion strategy

As shown in Fig. 2, each of the five networks i.e MobileNet, DenseNet201, Xception, InceptionV3, DenseNet121 is trained individually on a balanced chest x-ray dataset. Similar transfer learning approach is followed as described above. Each network follows the same classification head containing a dropout(.2) layer followed by two dense layers with 128 and 256 neurons, respectively. Afterwards, similar steps like the case of feature concatenation are followed (*relu* activation, *softmax* activation, adam optimization).

### D. Dataset

The dataset used in this study comprises COVID-19, normal, viral pneumonia and bacterial pneumonia chest X-ray images. The images are gathered from two different sources. COVID-19 images are collected from GitHub [10] repository developed by Joseph Paul Cohen. 408 COVID-19 chest X-ray images are sorted out from GitHub. The normal and pneumonia images are collected from Kaggle [11]. The Kaggle dataset is composed of 1583 normal and 4273 pneumonia images of pediatric patients from Guangzhou Women and Children's Medical Center, Guangzhou. For creating a balanced chest X-ray dataset, a 408 number of normal, viral pneumonia and bacterial pneumonia classes are collected.

For 4 class classification, the dataset contains 408 images of each of the 4 classes of which 326 images of each class are used for training and 82 images are used for testing.

For 3 class classification, the dataset contains 328 images of COVID-19, 328 images of normal and 328 images of pneumonia class. 246 images of each class are used for training and 82 images are used for testing.

TABLE I
PERFORMANCE OF PROPOSED FEATURE CONCATENATION SCHEME FOR 4 CLASS PROBLEM

| Folds | Precision(%) | Sensitivity(%) | Specificity(%) | F-1 Score(%) | Accuracy(%) |
|---|---|---|---|---|---|
| **Fold 1** | 90.27 | 89.94 | 96.64 | 89.70 | 89.94 |
| **Fold 2** | 89.79 | 86.89 | 91.46 | 85.67 | 86.89 |
| **Fold 3** | 96.14 | 96.04 | 98.68 | 96.06 | 96.04 |
| **Fold 4** | 89.40 | 88.27 | 96.09 | 87.62 | 88.27 |
| **Fold 5** | 85.29 | 85.19 | 95.06 | 84.56 | 85.19 |
| **Average** | 90.18 | 89.27 | 95.59 | 88.72 | 89.26 |

TABLE II
PERFORMANCE OF PROPOSED FEATURE CONCATENATION SCHEME FOR 3 CLASS PROBLEM

| Folds | Precision(%) | Sensitivity(%) | Specificity(%) | F-1 Score(%) | Accuracy(%) |
|---|---|---|---|---|---|
| **Fold 1** | 97.24 | 97.15 | 98.57 | 97.14 | 97.15 |
| **Fold 2** | 94.78 | 94.72 | 97.34 | 94.73 | 94.72 |
| **Fold 3** | 95.33 | 95.12 | 97.56 | 95.09 | 95.12 |
| **Fold 4** | 96.67 | 96.30 | 98.15 | 96.31 | 96.30 |
| **Fold 5** | 95.98 | 95.88 | 97.94 | 95.90 | 95.89 |
| **Average** | 96.00 | 95.83 | 97.91 | 95.83 | 95.84 |

## III. RESULTS AND DISCUSSIONS

In this paper, basically two separate schemes are proposed: ensembling of CNN models using feature concatenation and ensembling of CNN models using decision fusion. In this section, the classification performance of these proposed schemes are investigated individually. For the purpose of analysis, five widely used performance measures are considered: Precision, Sensitivity, Specificity, F1-Score and Accuracy. Classification results for the dataset mentioned before for four class and three class scenarios are computed for each proposed scheme.

Using the feature concatenation approach, the accuracy obtained by the proposed approach is 89.26% and 95.84% for 4-class and 3-class classification, respectively. The combined confusion matrix for 4-class and 3-class classification using the Feature Concatenation scheme is shown in Fig. 4.

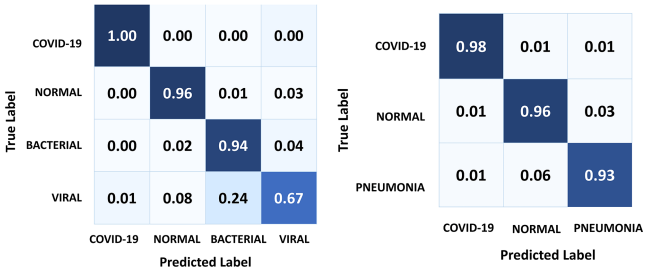


Fig. 4. Combined confusion matrix for four class and three class classification using Feature Concatenation approach

Classification performance obtained by the first scheme, ensembling of CNN models using feature concatenation, is presented in Table I and Table II for three-class and four-class problems, respectively. Results obtained in each fold of the five fold cross validation scheme during the testing phase is presented in the tables along with the average results. A very satisfactory overall performance is achieved in case of all performance measures. As expected the the classification performance is found better in case of three class problem. It

TABLE III
PERFORMANCE OF PROPOSED DECISION FUSION SCHEME FOR 4 CLASS PROBLEM

| Folds | Precision(%) | Sensitivity(%) | Specificity(%) | F-1 Score(%) | Accuracy(%) |
|---|---|---|---|---|---|
| **Fold 1** | 91.92 | 91.77 | 91.73 | 97.25 | 91.77 |
| **Fold 2** | 91.67 | 90.55 | 90.12 | 96.85 | 90.55 |
| **Fold 3** | 90.13 | 89.02 | 96.34 | 89.07 | 89.02 |
| **Fold 4** | 88.22 | 88.27 | 96.10 | 88.07 | 88.27 |
| **Fold 5** | 86.82 | 86.42 | 95.47 | 86.25 | 86.42 |
| **Average** | 89.75 | 89.21 | 93.95 | 91.50 | 89.21 |

TABLE IV
PERFORMANCE OF PROPOSED DECISION FUSION SCHEME FOR 3 CLASS PROBLEM

| Folds | Precision(%) | Sensitivity(%) | Specificity(%) | F-1 Score(%) | Accuracy(%) |
|---|---|---|---|---|---|
| **Fold 1** | 98.38 | 98.38 | 99.19 | 98.37 | 98.37 |
| **Fold 2** | 92.96 | 92.68 | 96.34 | 92.70 | 92.68 |
| **Fold 3** | 96.35 | 96.35 | 98.17 | 96.33 | 96.34 |
| **Fold 4** | 96.94 | 96.71 | 98.35 | 96.68 | 96.71 |
| **Fold 5** | 95.96 | 95.88 | 97.94 | 95.88 | 95.88 |
| **Average** | 96.11 | 96.00 | 98.00 | 96.00 | 96.00 |

is to be mentioned that the task of differentiating two types of pneumonia is relatively difficult. In case of three class problem better performance is achieved as the two pneumonia classes are kept under one class.

In a similar fashion, the classification performance obtained by the second scheme, ensembling of CNN models using decision fusion, is presented in Table III and Table IV for three-class and four-class problems, respectively. In case of 4-class classification, decision fusion approach produces accuracy of 89.21% and in 3-class classification it offers 96% accuracy. It is evident that a very satisfactory performance is obtained in case of all performance measures. The performance is found very similar in both the schemes.

The combined confusion matrix for 4-class and 3-class classification using Decision Fusion approach is shown in Fig. 5.

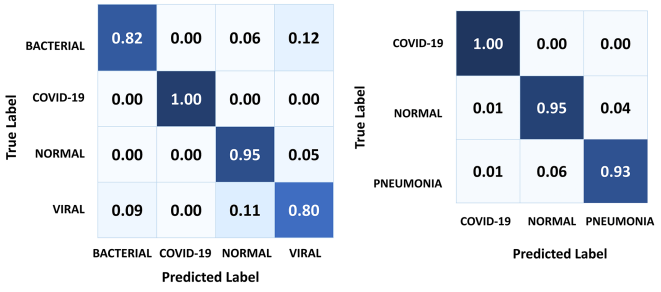


Fig. 5. Combined confusion matrix for four class and three class classification using Decision Fusion approach

The classification accuracy obtained by the proposed methods is compared with that obtained by some existing state-of-the-art approaches for COVID-19 detection from X-ray images in Table V. Ioannis et al. [2] compared five CNN models namely VGG19, MobileNetV2, Inception, Xception and Inception ResNet V2 with a conventional transfer learning scheme from the ImageNet database. MobileNetV2 and VGG19 generated the best result claimed by the authors. Dipayan et al. [7] also came up with a model based on

TABLE V
PERFORMANCE COMPARISON OF THE PROPOSED SCHEMES WITH SOME EXISTING METHODS

| Study | Total Number of X-ray images used | Architecture | Accuracy(%) |
|---|---|---|---|
| Ioannis et al. [2] | 224 COVID-19+500 normal+500 pneumonia | MobileNetV2 | 92.85 |
| | | VGG-19 | 93.48 |
| | 224 COVID-19+1204 Non-COVID-19 | MobilenetV2 | 97.40 |
| | | VGG-19 | 98.75 |
| Dipayan et al. [7] | 162 COVID-19 vs. Non COVID(500 normal+500 pneumonia) | Truncated Inception Net | 99.96 |
| | 162 COVID-19 vs. Non COVID(500 normal+500 pneumonia+400 TB) | | 99.92 |
| Tanvir et al. [4] | 305 COVID-19 vs. 305 normal | | 97.4 |
| | 305 COVID-19 vs. 305 viral pneu. | | 87.3 |
| | 305 COVID-19 vs. 305 bacterial pneu. | | 94.7 |
| | 305 COVID-19 vs. 305 viral pneu. vs.305 bacterial pneu. | CovXNet | 89.6 |
| | 305 COVID-19 vs. normal vs. 305 viral pneu. vs.305 bacterial pneu. | | 90.3 |
| Ozturk et al. [3] | 125 COVID-19 vs. 500 No-findings | DarkCovidNet | 98.08 |
| | 125 COVID-19 vs. 500 No-findings vs 500 pneumonia | | 87.02 |
| Wang and Wong [12] | 53 COVID-19 vs. 5526 Non-COVID | CovidNet | 92.4 |
| Khan et al. [1] | 284 COVID-19 vs.310 normal vs. 330 bacterial pneu. vs. 327 viral pneu.) | | 89.6 |
| | 284 COVID-19 vs.310 normal vs. combined(330 bacterial+327 viral pneumonia) | CoroNet | 94.59 |
| | 284 COVID-19 vs.310 normal | | 99 |
| **Proposed Ensembling Method** | 408 COVID-19 vs.408 normal vs. 408 bacterial pneu. vs 408 viral pneu. | **Feature Concatenation** | 89.26 |
| | | **Decision Fusion** | 89.21 |
| | 408 COVID-19 vs. 408 normal vs 408 pneumonia | **Feature Concatenation** | 95.84 |
| | | **Decision Fusion** | 96.00 |

InceptionV3 architecture. They also initialized the weights from the ImageNet data-set. Ozturk et al. [3] proposed a deep neural network DarkCovidNet and didn't apply any transfer learning strategies. Khan et al. [1] came up with the model CoroNet which is based on Xception architecture and initailized the model with the weights of ImageNet to do multi-class and binary classification.

In all these state-of-the-art approaches, an imbalanced dataset containing a small number of COVID-19 X-ray images is used. They have used only one CNN model and trained it using the traditional transfer learning approach. Any traditional transfer learning-based models is less likely to extract diversified features for challenging classification process. The proposed ensembling approach outperforms in this regard as by feature concatenation we get to use the diversified features obtained from different models. Also in decision fusion, all the predictions made by different models are taken into account .

## IV. CONCLUSION

In this research, both feature concatenation and decision fusion approaches are explored. In feature concatenation scheme, features obtained from the base models are concatenated and then fed to the classification head and in the decision fusion approach, the majority voting class from the predictions of the base models is taken as the final class. Both approaches have significant performance in the classification process compared to other methods available in the literature. Although both approaches perform similar in the 4 class classification but feature concatenation gets an edge in the performance for 3 class classification. Hence it is proved that the presented method can increase the performance of simple transfer learning techniques and can be suggested as

a preliminary tool to detect COVID-19 and to widen the afterward treatment for the positive patients. In the proposed method, the poor performance of one base model deteriorates the overall performance as in the case of fold 5 of the 4 class classification of decision fusion approach. So, similar performing other base models will be explored and different ensembling techniques will be pursued and large data-set will be searched constantly for further improvements.

## REFERENCES


[1] A. I. Khan, J. L. Shah, and M. M. Bhat, "Coronet: A deep neural network for detection and diagnosis of covid-19 from chest x-ray images," *Computer Methods and Programs in Biomedicine*, vol. 196, p. 105581, 2020. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0169260720314140

[2] I. D. Apostolopoulos and T. A. Mpesiana, "Covid-19: automatic detection from x-ray images utilizing transfer learning with convolutional neural networks," *Physical and Engineering Sciences in Medicine*, p. 1, 2020.

[3] T. Ozturk, M. Talo, E. A. Yildirim, U. B. Baloglu, O. Yildirim, and U. R. Acharya, "Automated detection of covid-19 cases using deep neural networks with x-ray images," *Computers in Biology and Medicine*, p. 103792, 2020.

[4] T. Mahmud, M. A. Rahman, and S. A. Fattah, "Covxnet: A multi-dilation convolutional neural network for automatic covid-19 and other pneumonia detection from chest x-ray images with transferable multi-receptive feature optimization," *Computers in Biology and Medicine*, vol. 122, p. 103869, 2020. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0010482520302250

[5] N. Narayan Das, N. Kumar, M. Kaur, V. Kumar, and D. Singh, "Automated deep transfer learning-based approach for detection of covid-19 infection in chest x-rays," *IRBM*, 2020. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1959031820301172

[6] F. Ucar and D. Korkmaz, "Covidiagnosis-net: Deep bayes-squeezenet based diagnosis of the coronavirus disease 2019 (covid-19) from x-ray images," *Medical Hypotheses*, vol. 140, p. 109761, 2020. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0306987720307702

[7] D. Das, K. Santosh, and U. Pal, "Truncated inception net: Covid-19 outbreak screening using chest x-rays," *Physical and Engineering Sciences in Medicine*, pp. 1–11, 2020.

[8] S. Vaid, R. Kalantar, and M. Bhandari, "Deep learning covid-19 detection bias: accuracy through artificial intelligence," *International Orthopaedics*, p. 1, 2020.

[9] M. Toğaçar, B. Ergen, and Z. Cömert, "Covid-19 detection using deep learning models to exploit social mimic optimization and structured chest x-ray images using fuzzy color and stacking approaches," *Computers in Biology and Medicine*, vol. 121, p. 103805, 2020. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0010482520301736

[10] J. P. Cohen, P. Morrison, and L. Dao, "Covid-19 image data collection," *arXiv 2003.11597*, 2020. [Online]. Available: https://github.com/ieee8023/covid-chestxray-dataset

[11] (2018) Chest x-ray images (pneumonia). [Online]. Available: https://www.kaggle.com/paultimothymooney/chest-xray-pneumonia

[12] 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," *arXiv preprint arXiv:2003.09871*, 2020.