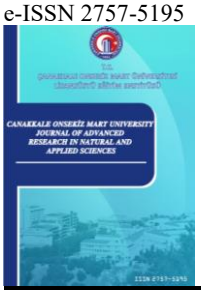
e-ISSN 2757-5195



Çanakkale Onsekiz Mart University
Journal of Advanced Research in Natural and Applied Sciences
Open Access

doi.org/10.28979/jarnas.952700 2021, Vol. 7, Issue 4, Pages: 486-503 dergipark.org.tr/tr/pub/jarnas

COVID-19 Detection from Chest X-Ray Images and Hybrid Model Recommendation with Convolutional Neural Networks

Hacer Karacan[1], Furkan Eryılmaz[2,*]

[1]Department of Computer Engineering, Faculty of Engineering, Gazi University, Ankara, Turkey
[2]Department of Computer Science, Graduate School of Informatics, Gazi University, Ankara, Turkey

Article History

Received: 21.06.2021
Accepted: 22.09.2021
Published: 15.12.2021

Research Article

Abstract − The coronavirus pandemic, which emerged at the end of 2019, continues to be effective. Although various vaccines have been developed, the diagnosis of the disease and the 14-day isolation method still remain valid. *Purpose:* To classify diseases such as COVID-19, tuberculosis, other pneumonias, and lung opacity. For this goal, it was aimed to automate disease diagnosis, which was done manually by expert radiologists with chest X-Ray images, with a hybrid model created with the four most commonly used convolutional neural networks. In addition, it is aimed to reach a more accurate result by combining four CNN models instead of one model during the epidemic when diagnosis is of critical importance. *Method:* In the binary classification, the images were categorised as COVID-19-positive and negative, and multi-class classification was conducted according to labels including Lung Opacity, COVID-19, Normal, and Viral Pneumonia. The recommended model consisted of MobileNetV2, DenseNet121, InceptionResNetV2, and Xception networks. Transfer learning was used for initial weights. These models, were combined with ensemble learning to obtain better classification performance. *Results*: The best binary classification result was obtained from the MobileNetv2 with an accuracy rate of 98.84%. The hybrid model improved slightly, with a value 0.16. In the multi-class classification, the DenseNet121 accomplished the highest classification rate, with 93.17%. The hybrid model improvement in the multi-class classification was 0.81. According to these results, the proposed method will alleviate the burden of health personnel during the epidemic. It will also aid in the diagnosis of COVID-19 in areas where facilities are limited.

Keywords *– CNN, Chest X-Ray, COVID-19, ensemble learning, model averaging*

1. Introduction

In December 2019, a new virus was informed to the World Health Organization (WHO) by Chinese specialists, whose origin was unknown, but with symptoms similar to pulmonary pneumonia (Allam, 2020). The WHO is called this new pneumonia as severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), has been accepted as an epidemic disease, and according to statistics, it has infected 209 million people and caused the deaths of 4.38 million people (Qu, Cao, & Chen 2021; WHO Coronavirus (COVID-19) Dashboard, 2021). The virus causing SARS was detected around Guandong, China in 2003, the Middle East Respiratory Syndrome (MERS) was exposed in the Middle East in 2012, and the SARS-CoV-2 virus was detected in the Hubei region of the Chinese city of Wuhan (Allam, 2020; Qu et al., 2021).

The COVID-19 epidemic has deeply affected the entire world, and no effective drug has been found against the virus to date. The existing medical infrastructures of countries have collapsed against the virus and most patients have not been able to reach treatment. Specialists have recommended the use of masks, social distancing, quarantine, and filiation studies after disease detection to communities and governments as a

[1] hkaracan@gazi.edu.tr
[2] eryilmaz.furkan@gmail.com
*Sorumlu Yazar / Corresponding Author

precaution against the virus. In order to keep the pandemic under control, rapid and accurate detection of the disease has been of critical importance.

As stated in WHO data dated the end of August 2021, the total number of reports since the beginning is 81 million in America, 64 million in Europe, 41 million in Asia, and 5.5 million in Africa. The highest number of cases comprise 37 million in the United States, 32.4 million in India, 21 million in Brazil, 6.7 million in the Russian Federation, and 6.4. million in France. The number of incidents is directly proportional to the number of tests, and when the access of African countries to the test is considered, these numbers show that the situation is more serious.

Reverse transcription polymerase chain reaction (RT-PCR), diffractive phase interferometry, saliva, antibody tests, lung biopsy, and radiological imaging methods [computed tomography (CT), lung X-Ray] are used in the detection of coronavirus (Huang et al., 2020; Ozturk et al., 2020; Wang, Lin, & Wong, 2020; Bouchareb et al., 2021). The PCR test is the most commonly used method and has been recognized as the gold standard. However, this test method has limitations, such as producing erroneous results (Huang et al., 2020), supply shortages, a long diagnostic time, and the requirement of manual processing (Wang et al., 2020). The sensitivity rate of this diagnostic method is between 60% and 70%, and it can give negative results in cases with COVID-19 symptoms (Ozturk et al., 2020).

As stated in October 2020 dated announcement of the WHO, medical imaging methods have been accepted as a valid method for the identification of clinical symptoms of patients who have been infected and recuperated (Alam et al., 2021). In the examinations performed by radiologists on CT and chest radiograph images of COVID-19 positive patients, findings specific to the SARS-CoV-2 virus were detected (Wang et al., 2020). These coronavirus findings are expressed as the two-sided distribution of irregular shades and the opacity of opal glass (Wang et al., 2020). These infected lung areas were found in lung X-Ray images in 9% of patients with negative RT-PCR tests. The results are dependent on the antigen concentration in the test kit (Karthik, Meneja & Hariharan, 2020). The output time of the test results varies from one hour to several days (Bouchareb et al., 2021)

During the epidemic, most countries did not have access to RT-PCR test kits or had to fight against the virus with very few test kits. Countries like Turkey, where the number of RT-PCR test kits is insufficient, have tried to COVID-19 diagnosis with CT (Ozturk et al., 2020). CT devices have long imaging times and are not available in most underdeveloped countries (Alam, Ahsan, Based, Haider & Kowalski, 2021). The disadvantages of this method include that it cannot be applied to children and pregnant women due to high radiation, high cost (Wang et al., 2021), insufficient equipment and personnel, and CT laboratories increase the risk of infection (Oh, Park, & Ye, 2020). The CT imaging method is useful in monitoring the condition of COVID-19 patients rather than diagnosis (Horry, Chakraborty, Paul, Pradhan & Shukla, 2020).

Today, most hospitals and medical clinics have X-Ray equipment (Narin, Kaya, & Pamuk, 2021). The chest X-Ray imaging method offers advantages such as working with lower exposure to radiation, rapid output, and low cost when compared to the CT method (Wang et al., 2021). X-Ray imaging take approximately 15–20 min (Karthik et al., 2020). Chest X-Ray images are preferred in the process of diagnosing COVID-19 due to its modularity and accessibility. Repeated features observed on the X-Ray images of COVID-19 cases include irregularity of nitrates or opacities, similar to other pneumonia features (Horry et al., 2020).

The analysis of the disease can only be determined by specialist radiologists by interpreting the chest radiograph images. The similarity of the findings on the images of COVID-19 and other pneumonias makes the diagnosis process more difficult (Wang et al., 2021). It takes an average of 5–6 min to make a diagnosis if examination of the image is performed by a specialist radiologist (Rubin et al., 2020).

Despite the negative aspects described above in the diagnosis of coronavirus, lung radiograph images are accepted as a valid and effective method. These problems in the evaluation of the image can be eliminated with machine learning (ML) approaches and an agile solution can be presented. Artificial intelligence

techniques need large and balanced data sets in order to give high performance results. (Abbas, Abdelsamea, & Gaber, 2021). However, the number of data in open sources for COVID-19 is limited (Abbas, Abdelsamea, & Gaber, 2021). In cases where there are insufficient data, the approach to be followed in order to obtain successful results is convolutional neural networks (CNNs) (Abbas, Abdelsamea, & Gaber, 2021).

Successful results in detecting corona virus from lung X-ray images have been obtained using CNNs and have made classification with higher accuracy than support vector machines(SVM), artificial neural networks (ANNs), and k-nearest neighbour (KNN) algorithms (Monshi, Pool, Chung, and Monshi, 2021).

Recently, it has become very popular to ensemble more than one classifier in the field of ML and DL and use them together for classification problems (Bhardwaj & Kaur, 2021). Ensembled architectures have produced effective results in real world problems and computer vision problems (Bhardwaj & Kaur, 2021).

Within the scope of the this study, MobileNetV2, DenseNet121, Xception, and InceptionResNetV2, which are the most used CNNs in the literature, were applied to lung radiograph images. Instead of starting from scratch with the transfer learning approach, the initial weights were transferred to these models from the ImageNet dataset. Combining the advantages of the above-mentioned four networks with the ensemble learning approach, a hybrid solution capable of binary and multiple classification is proposed. One of the largest known datasets in open sources was used. The accuracy rate was chosen as the evaluation parameter. It is hoped that the hybrid solution, which included four different models, will demonstrate robust performance against different types of datasets that it may encounter in the future.

The main contributions can be outlined as follows:

- Developing a test method within existing medical equipment (X-Ray Devices) that can be found in each health centre
- Supporting existing testing capabilities for more accurate results
- Providing testing capabilities to countries that cannot access PCR tests
- Proposing a more effective system by combining four models instead of one at epidemic when accurate diagnosis is critical
- Building an effective model for larger datasets that may be encountered in the future

This study is organised as follows. The reminder of this chapter presents a literature review on coronavirus classification from radiological images. In section 2, the theoretical background about the models used in the research, CNN, transfer and ensemble learning are discussed. In addition, the dataset and information about the application in the research are included in this section. In section 3, the outputs of the study, results, and discussion are included. In Section 4, a brief summary of the study, the outputs, and future studies are given.

1.1. Literature Review

In this section, there is a literature review of ML approaches and applications in the field of COVID-19 detection. With the increase in data, the AI community has started to support to the control COVID-19 epidemic with traditional ML methods, new approaches, and hybrid model applications. The most common and biggest problem seen in the studies included in the literature review was the lack of a sufficient size data set. It was aimed to overcome the related constraint by various methods.

Wang et al. (2020) classified normal, viral, or bacterial, COVID-19 labels on chest radiograph images in their study presenting COVID-NET. The network, whose initial weights were transferred from the ImageNet data set, made classification with 92.6% accuracy. Ozturk et al. (2020) studied the diagnosis of COVID-19 from lung radiograph images in their work presenting DarkNet-19-based DarkCovidNet. They reached a result with accuracy rates for binary and multiple classification 98.08% and 87.02% respectively. Hemdan et al. (2020) proposed COVIDX-Net in their study by using VGG19, InceptionV3, DenseNet201, ResNetV2,

InceptionResNetV2, MobileNetV2, and Xception models. As a result of the study, the most successful results were obtained from the VGG19 and DenseNet201 networks.

Rahimzadeh & Attar (2020) studied Xception and ResNet50V2 models on a dataset containing very few images of a COVID-19 class. To overcome this problem, the dataset was divided into subsets according to the least number of class labels and the models were trained in 8 consecutive steps. By combining the output layer of the two models, a performance evaluation was made with 5 cross-validation methods and an accuracy rate of 92.4% was obtained. Arkadani, Kanafi, Acharya, Khadem & Mohammadi (2020) surveyed CT images, and examined 10 commonly used CNN models (AlexNet, GoogleNet, VGG16-19, SqueezeNet, MobileNetV2, ResNet18-50-101, and Xception). When the model performances were evaluated, the highest results were obtained from ResNet101 with 99.26% accuracy and Xception with an accuracy rate of 99.38%.

Feki, Ammar, Kessentini & Muhammad (2021) proposed a federated and centralised system solution based on VGG16 and ResNet50, which will cover clinics that cannot share their data due to data privacy. The initial weights are obtained from the training in the central environment and transferred to the client nodes. Client nodes start with these weights in their training and transfer possible updates to the centre. The process is managed iteratively. In federated architecture, VGG16 performed with 93.57 % accuracy, ResNet50 95.4%, and in centralised architecture VGG16 93.75% and ResNet50 95.3% accuracy. The CVDNet model, developed by Ouchicha, Ammor & Meknassi (2020) and based on ResNet networks, aimed to extract local and general characteristics of the data. The architecture, which consists of two parallel blocks, applies two different filters to the image. The solution presented was evaluated with 5-fold cross-validation, and it worked with a success rate of 97.20% in binary classification and 96.69% in multiple classification.

Bozkurt (2021) applied CNN networks to X-ray images from simple to complex and compared performance. All networks were multi-classified and labelled data according to COVID-19, Normal, and Other pneumonia labels. Several networks were used in the study. The best result was achieved from the DenseNet121 network with an accuracy rate of 97.26%. Bhardwaj & Kaur (2021) tried to diagnose COVID-19 and other respiratory diseases by applying an ensemble learning approach to X-Ray images. Ensemble deep learning consists of InceptionV3, DenseNet121, Xception, and InceptionResNetV2 models. The proposed method made binary and multi-classification with 92.36% and 98.33% accuracy. In another study for the detection of COVID-19 from lung radiograph images, Altan & Karsu (2021) proposed a hybrid study combining chaotic salp swarm algorithm, 2D Curvelet Transformation, and EfficientNet-B0 model. The feature map was created with the coefficients obtained with the 2D Curvelet transformation, the coefficient optimization in the feature map was made with CSSA and the classification was performed with EfficentNET-B0. The hybrid architecture has classified data with a high accuracy rate of 99.59%.

Monshi, Pool, Chung, & Monshi (2021) focused on data augmentation and hyper parameters to get better performance for CNN models. The best data augmentation as a result of trials was determined as 20° rotation, 1.2 zooming, 0.2 warping, 0.3 lighting, and normalizing. The most efficient hyper parameters were 30 epochs, 32 bath size, and cross-entropy loss function. The accuracy rate of the VGG-19 network was improved by 11.93% and the ResNet50 network by 4.97%. The proposed CovidXrayNet network, based on EfficientNet-50, achieved triple classification with an accuracy of 95.82% as a result of the evaluations. Alam, Ahsan, Based, Haider, & Kowalski (2021) focused on feature extraction processes in their study. After collected features from X-ray images with histogram and CNN methods, they combined the features and gave them as input to the model. The VGGNet model was used and the outputs were compared with ANNs, KNNs, and CNNs. It performed better than these networks and classified with 99.49% accuracy.

Afifi, Hafsa, Ali, Alhumam & Alsanman (2021), analysed the entire lung x-ray image and local anomalies simultaneously using the ResNet18, InceptionV4 and DenseNet161 architectures. Each model is ensembled by averaging the weights which obtained from global and local points. The most successful result was obtained from the ensembled version of the DenseNet161 network. The ensemble architecture classified the data according to control, Pneumonia, and COVID-19 tags with 91.2% accuracy. Saha et al. (2020), ensembled the

MobileNet, InceptionV3, Xception, DenseNet121 and DenseNet201 models in order to be able to classify with higher performance. In the ensembling process, two different strategies were used, namely the averaging of the feature maps and the max vote method for the predictions in the output layer. The classification made in 4 labels as COVID-19, Normal, Viral and Bacterial pneumonia resulted in an accuracy rate of 89.21%. Focusing on overfitting, high variance, and noise-induced generalization errors of small datasets, Tang et al. (2021) proposed a method they named EDL-COVID. This method is based on taking samples of the COVID-Net architecture according to different sensitivity values. Then, these models with different weights are combined by taking the average of the weights. The sensitivity value was accepted as a performance parameter and it was obtained at a rate of 96%. The accuracy rate, on the other hand, is 95% of the binary classification accuracy of the COVID-19 tag. Ghernea & Neagoe (2021) combined two concurrent convolutional neural networks based on VGG-Net with ensemble learning. The main difference of the models in the same architecture is that they are trained with asymmetrical data sets. The best asymmetric value for the binary classification was determined as 5, and the ensemble learning model performed with 95.02% accuracy.

2.1. Technical Background

2.1.1. Convolutional Neural Network

CNN architectures used in computer vision and image processing are a deep learning approach applied to data exhibiting similar topology, such as sequential pixels in images (Goodfellow, Bengio, & Courville, 2017). The serious processing complexity encountered by other neural network models in image processing is overcome with CNN models and reduced to more manageable states (Ankile, Heggland, & Krange, 2020). CNN architectures consist of different and ordered layers. First layer is convolution, middle layer is pooling, and last layer is fully-connected layer (Aloysius & Geetha, 2017).

- Convolutional Layer: In this layer, local features are extracted by applying a filter or a matrix called kernel to the image (Arora, Garg, & Gupta, 2020). The output of the convolution layer is called the feature map (Ouchicha et al., 2020). Most of the computational load of the entire network occurs at this layer (Arora et al., 2020). Optimizations to be made on hyper parameters such as filter size, stride, and padding used in the extraction of the feature map provide a significant improvement in the complexity of the network (Aloysius & Geetha, 2017).
  The outputs of the convolution layer are linear, and it needs to be converted to nonlinear structure for improve the speed of the model and to get the desired outputs from the input functions (Gulcu & Kus, 2020; Hassantabar, Ahmadi, & Sharifi, 2020; Ouchicha et al., 2020).
  - Rectified Linear Unit (ReLU): Inputs are converted to values between 0 and ∞. ReLU is widely preferred due to its better performance against vanishing gradient problem, low cost and high performance in the training phase (Ouchicha et al., 2020).
- Pooling Layer: The number of parameters of the model is reduced by applying pooling layer to the outputs of the convolution layer (Aloysius & Geetha, 2017). In this way, improvements in computational complexity are achieved and overfitting is prevented (Aloysius & Geetha, 2017).
- Fully-Connected Layer: Using this layer, a link is established between the features obtained from the image and the classification label (Ouchicha et al., 2020). It is not possible to apply the convolution after flatten layer (Arora et al., 2020).

Within the scope of the related research, MobileNetV2, DenseNet121, InceptionResNetV2, and Xception models were used in binary and multiclass classification.

2.1.1.1. MobileNetV2

MobileNet networks classify with fewer computation and parameters than other networks. These networks, which exhibit similar classification performance to other networks, also have low memory requirements. Compared to other CNN models, the computational cost is 8–9 times less. In the convolution layer of these

networks, two different sub-convolutional layers are used, as point(1×1) and deep(3×3) (Sandler, Howard, Zhu, Zhmoginov, & Chen, 2018).

2.1.1.2. DenseNet121

In DenseNet networks, each layer takes the feature map of the former layer as input and adds the output it produces to this feature map and transfers it to the next layer (Huang, Liu, Maaten, & Weinberger, 2017). This advantageous situation complicates the control of the main feature map and the management of the network (Wang, & Zhang, 2020). In order to keep the feature map manageable by the network, the subsampling method is used in the merge operations made in the layers. In addition, transition blocks are placed between the layers, which include stack normalization, convolution, and pooling operations (Wang, & Zhang, 2020).

With the DenseNet architecture, the convolutional model becomes resistant to the vanishing gradient problem, exhibits a strong feature distribution, and works with less total number of parameters than other networks (Huang et al., 2017).

2.1.1.3. Xception

Xception networks are formed by combining depthwise separable layers with residual connections (Chollet, 2017). While the network achieves computational time improvement with the help of separable convolution layers, it is made more resistant to the vanishing gradient problem with the residual connections (Chollet, 2017; Rsimiyati, Endah, Khadijah & Shiddiq, 2020).

2.1.1.4. InceptionResNetv2

It is an effective hybrid model that combines the InceptionV3 architecture and residual connections. Although more effective results are obtained with increasing depth in Inception networks, the problem of vanishing gradient is encountered. By combining the two network architectures, the computational efficiency advantage of Inception networks is retained, while all of the advantages of the residual networks, such as vanishing gradient problem solving, are utilised (Szegedy,Ioffe, Vanhoucke & Alemi, 2016; Wu, Liu,Yang & Zen, 2020).

2.1.2. Transfer Learning

A large quantity of training and test data is needed in order to construct and effectively implement ML methods (Tan, Sun, Kong, Zgang, & Liu, 2018). However, it is often not possible to collect sufficient data (Tan et al., 2018). Transfer learning is the process of increasing the performance of applications in these fields of study by transferring the knowledge obtained from fields that consist of healthy data and that show a balanced data distribution to different but related fields and target algorithms.

With transfer learning, there is an increase in the learning performance of the target study area and a significant decrease in the data requirement and time required for training. Transfer learning approaches are divided into four groups as example, feature, parameter and relation-based (Zuhang et al., 2021).

2.1.3. Ensemble Deep Learning

Ensemble deep learning is the process of training ANNs simultaneously using one or more data sets to solve a common problem (Goodfellow et al., 2017). The final model that emerges as a result of ensemble deep learning combines the advantages of its sub-models and shows much better generalization performance (Ganaie, Hu, Tanveer, & Suganthan, 2021). High-performance results are obtained with these integrated network applications in large-scale challenges such as natural language processing, computer vision and speech recognition (Yang et al., 2021).

2.1.4. Dataset

The dataset studied in this research were obtained from the Coronavirus Radiology database. The dataset (COVID-19 Radiography Database, 2021) was generated by the researchers of Qatar and Dhaka Universities and Pakistani and Malaysian medical doctors who supported this team. The dataset consists of the following sources, BIMCV Medical Image Bank of the Valencia Region (BIMCV, 2021), GitHub (Ml-Workgroup, 2021), SIRM Società Italiana di Radiologia Medica y Interventisca (Redazione, 2021), European Society of Radiology (Eurorad.org, 2021), GitHub (Ieee8023, 2021), FigShare COVID-19 Chest X-Ray Image Repository (COVID-19 Chest X-Ray Image Repository, 2021), GitHub (Armiro, 2021), Radiology Society of North America (RSNA Pneumonia Detection Challenge, 2021), and Kaggle Chest X-Ray Images(Pneumonia) (Chest X-Ray Images (Pneumonia), 2018). The dataset consists of healthy, viral pneumonia, lung opacity, and COVID-19 patient radiograph images (Figure 1).

Table 1
X-Ray Image dataset structure

Subset	COVID-19	Lung Opacity	Normal	Viral Pneumonia
Train	2.908	4.799	8.152	1.076
Test	711	1.213	2.040	269
Total	3.616	6.012	10.192	1.345

Within the scope of the study, the data set was divided into training, validation, and test subsets, for both classification methods. The training and test set are separated by 80%–20%. In order to prevent overfitting of the model and to monitor its behaviour, 20% of the training set was reserved as a validation set. The distribution of dataset according to the subsets is placed in Table 1.

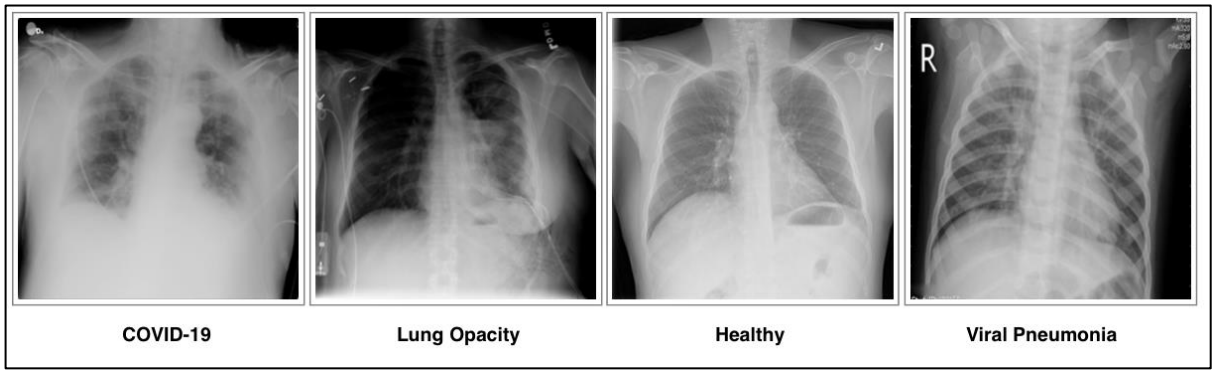


Figure 1. X-Ray images with classification labels.

Over-fitting is inevitable when the dataset size is small. There is a wide spectrum of methods to prevent this problem. Monshi, Pool, Chung, & Monshi (2021) focused on data augmentation and hyper parameters for better performance. In this study, the outputs of Monshi, Pool, Chung, & Monshi (2021)'s study were used. Data augmentation methods applied to X-Ray images; 20° rotation, 1.2 zooming, 0.2 warping, 0.3 lighting, 0.2 width and height shifting.

2.2. Proposed Method

The insufficient amount of the dataset, the presence of outlier data and the high randomness of deep learning models bring along the problems of variance and generalization error (Chen, Dobrian & Lee, 2019).

Although the effect of this problem is reduced by regularization and data augmentation, these disadvantages of a single model can be effectively overcome with ensemble learning (Tang et al.., 2021).

This study consisted of 3 phases. In the first phase, binary classification was made on the chest X-Ray images. In binary classification, each image is classified according to COVID-19 and other class labels. In the second phase, multiple classifications were made and the models were expected to classify the data according to the Normal, COVID-19, Lung Opacity, and Viral Pneumonia labels. In the third phase, these models, which have advantages over each other, were ensembled and it was aimed to reach a common solution that could perform binary and multiple classification with much higher performance.

The individually trained networks were combined with the unweighted model averaging ensemble method for binary and multiple classification. The approach applied within the scope of the study was based on creating the final model by averaging of the output layers of the sub-models. The probability that each model calculates for the classification labels was averaged. Final classification was based on the class label with the highest average probability. Initial weights were transferred from the ImageNet dataset by using the sample-based transfer learning methodology instead of starting from scratch. Transfer learning methodology is shown in Figure 2.

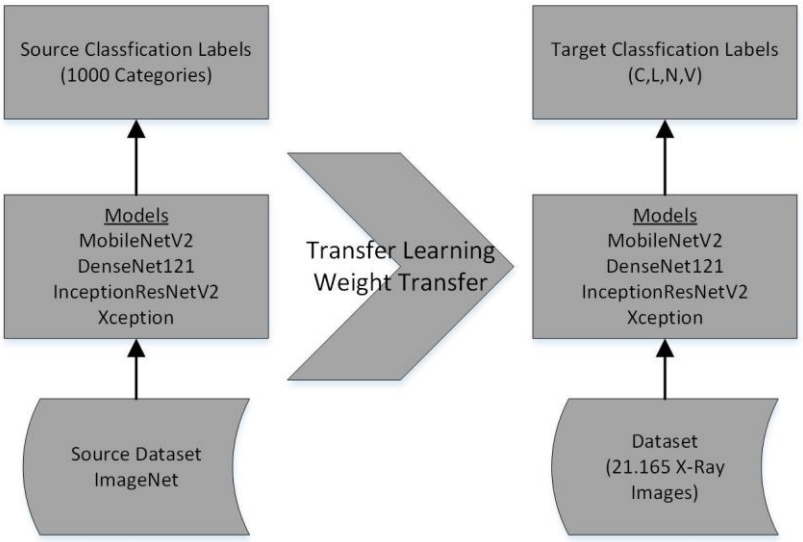


Figure 2. Transfer learning approach in multi classification.

2.2.1. Application

MobileNetV2, DenseNet121, InceptionResNetV2 and Xception CNNs were used as base classification models. In the first stage, the networks were individually trained. In these trainings, the transfer learning approach was followed and the process was started with the initial weights obtained from the ImageNet data set.

New layers were added to increase the performance in the output layer of the models and to show compatibility with the classification labels. There was a 4 × 4 filter size average pooling layer, flatten layer, hidden layer consisting of 256 neurons, dropout, hidden layer (128 neurons), dropout, hidden layer (64 neurons), and lastly, an output layer containing neurons belonging to labels. Rectified linear unit activation (ReLU) was used in the previous layers and the softmax was used in the last layer. These new layers are given in Figure 3. Experimented with a dropout value of 0.2, 0.5 and 0.8. The best result was obtained at 0.5.

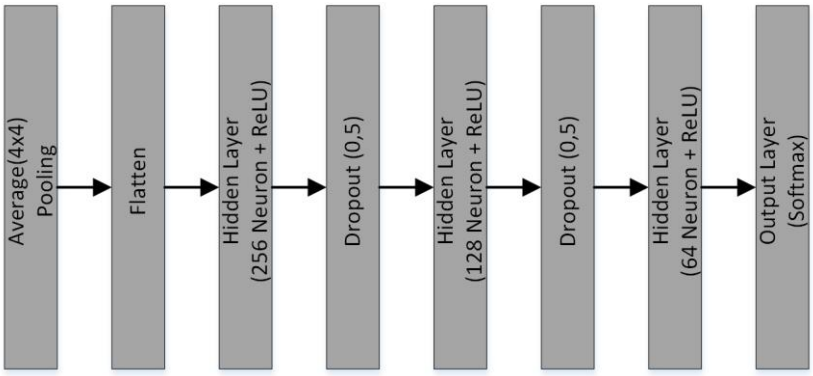


Figure 3. New output layers.

In the models, binary cross-entropy and categorical cross-entropy loss function were used in binary and multiple classification, respectively (Figure 4). The adadelta optimization algorithm was used in both classification models.

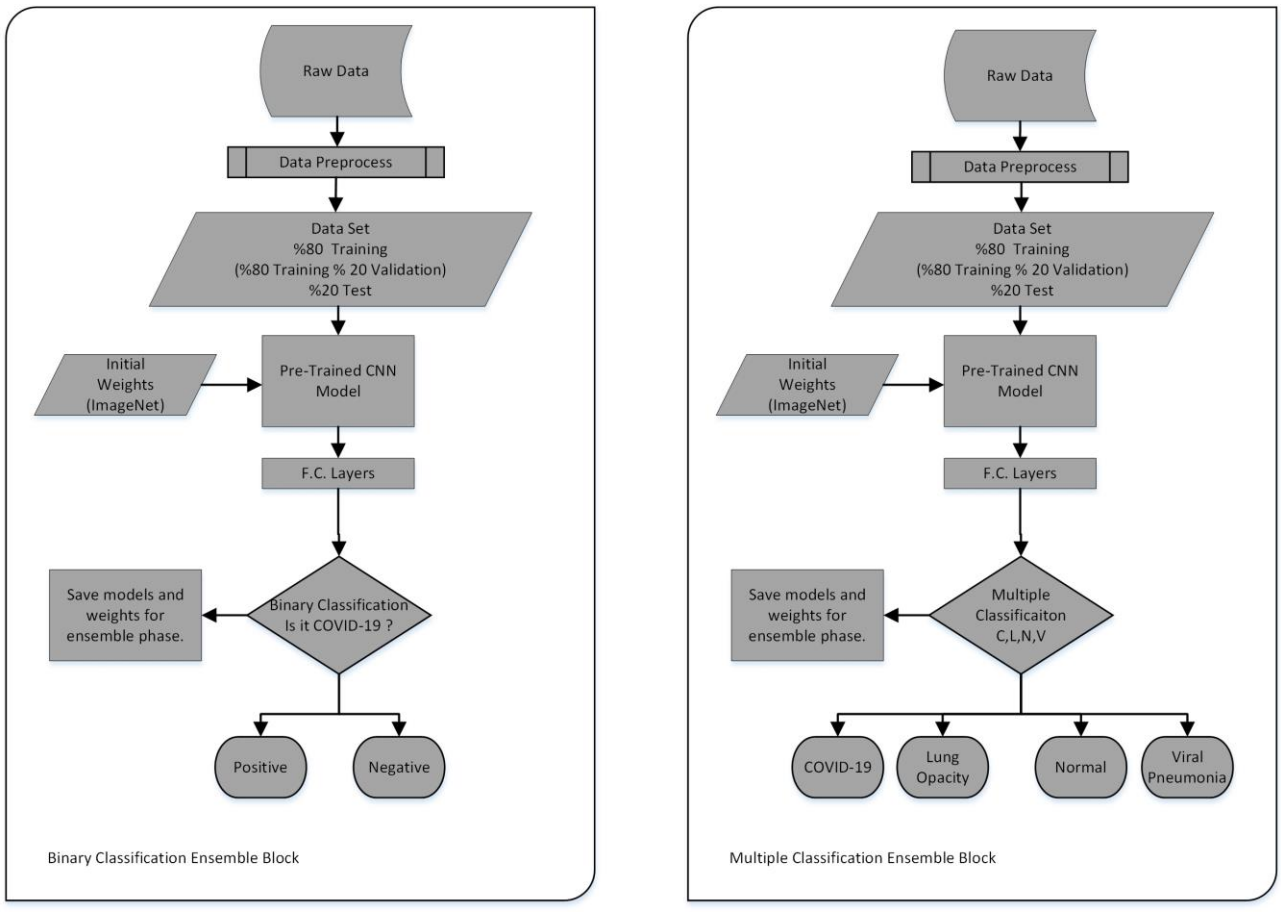


Figure 4. Binary and multiple classification blocks of CNN Models.

While applying the models, different methods were followed due to the limited resources available, such as time slots of the graphics processing unit (GPU) and central processing unit (CPU). During the training, the epoch was determined as 100 and the batch size was 32. In addition, the Early Stopping mechanism of the Keras library was used. This mechanism monitors the validation loss in the model. If the model is saturated and there is no change in the validation loss, the weights that give the best results are saved and the model is

terminated. The validation loss monitoring value was taken as 15. The accuracy rate was used as a performance metric. Other performance metrics such as Precision, Recall, Specificity, AUC, MSE were also examined.

The trained convolutional models and the weights of these models were saved and transferred to ensemble learning, which was the last step. The final model was generated with the unweighted model averaging ensemble method. This method is based on averaging the of the output layers of the saved sub-models. In this step, the average of the probabilistic values of the class labels is calculated and the classification is made according to the highest value. In the combined model, both the Adam and Adadelta optimization methods were used and better results were obtained with Adadelta. The trainable and non-trainable parameters of CNN models and hybrid architecture are demonstrated in Table 2. Despite the high complexity, each model was trained individually before combining them and only its weights were used in the hybrid model.

Table 2
Model parameter list.

Model	Input Shape	Trainable	Non Trainable	Total Parameter
DenseNet121	224x224x3	8.043.970	83.648	8.127.618
MobileNetV2	224x224x3	3.576.130	34.112	3.610.242
Xception	299x299x3	22.945.642	54.528	23.000.170
InceptionResNetV2	299x299x3	54.710.946	60.544	54.771.490
Ensembled Model	256x256x3	89.276.688	232.832	89.509.520

DenseNet121 block is shown as a representation in Figures 5 and 6. The core models of ensemble learning used in the general architecture were different, and these models were: DenseNet121 is MobileNetV2, Xception and InceptionResNetV2.

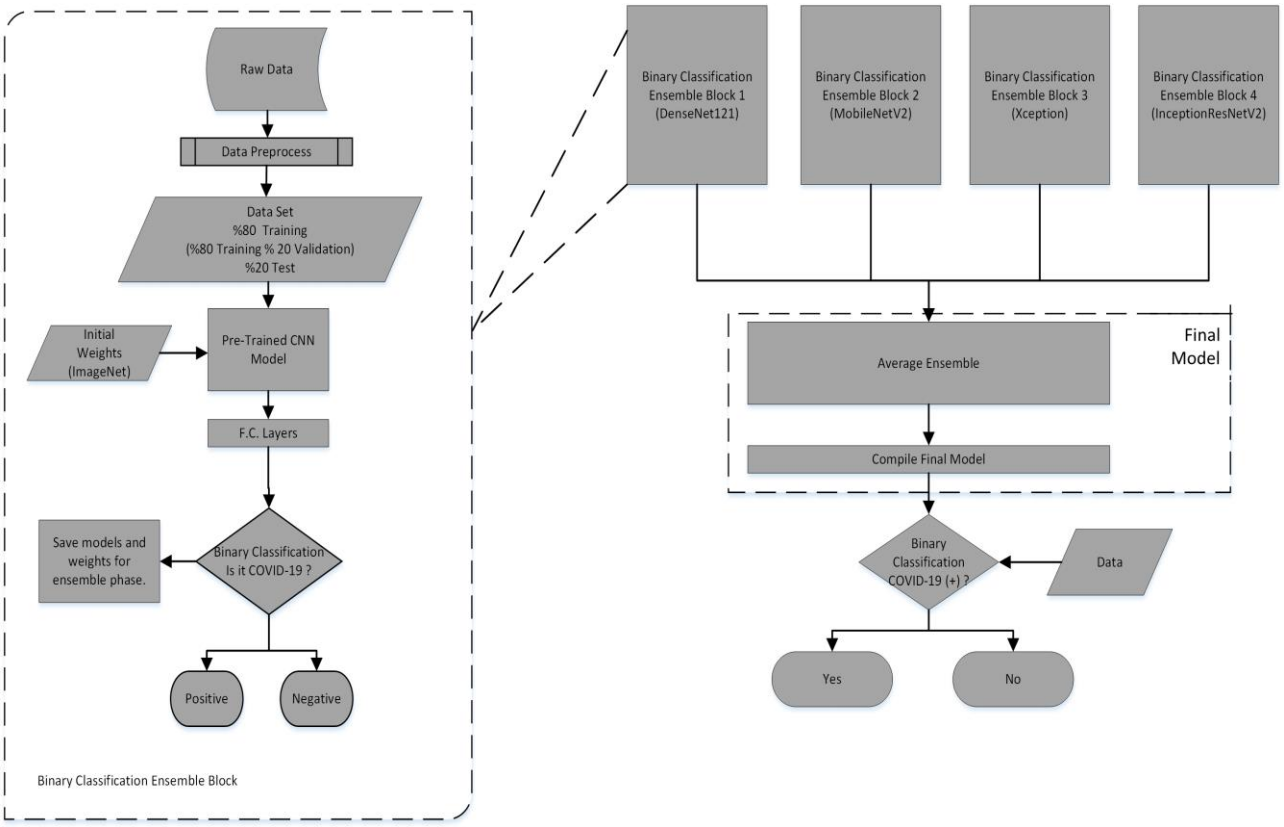


Figure 5. The proposed ensemble learning architecture for binary classification.

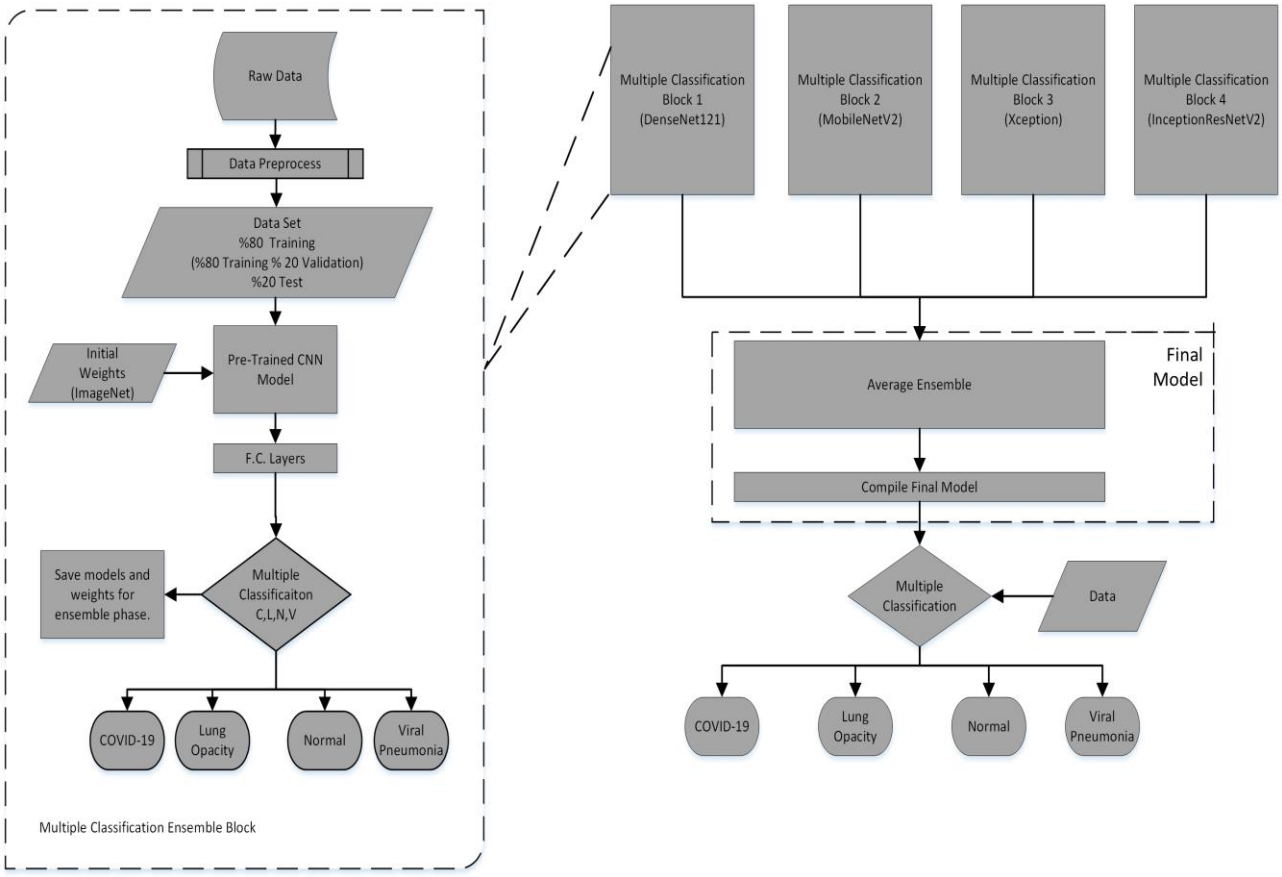


Figure 6. The proposed ensemble learning architecture for multi classification.

3. Results and Discussion

In the study on the data set, two different classification approaches were followed and the detection of SARS-CoV-2 pneumonia was attempted by analysing the chest radiograph images with CNNs. In the first approach, a binary classification was made with the COVID-19 and other class labels, and in the second approach, a multi-classification was made by considering the Lung Opacity, COVID-19, Normal, and Viral Pneumonia labels. In the proposed solution, the models were individually trained for binary (Figure 8), and multiple classification (Figure 7), and these models and weights were saved for ensemble deep learning. Various additions were made to the output layers in order to provide higher performance results for the individual models. In the final stage, the trained sub-models were combined with unweighted model averaging ensemble method. After the final model was created, the results in Tables 3 and 4 were obtained using the test set.

The performance of multi-classification was lower than that of binary classification. In order to overcome this problem, higher performance results can be obtained with detailed studies with activation functions and hyper parameters in future studies.

In the future, the current study can be supported by other clinical data and a much more holistic solution can be developed. CNN-based or other ML approaches, in which parameters such as blood values, saliva test, other lung data, PCR test are combined with X-ray images, can be studied on specific solutions to the problem. Or, focus can be placed on scenarios where there are not enough medical personnel but basic imaging devices, such as x-ray, are available. CNN networks that can work on mobile devices can be examined, and support personnel, such as inexperienced doctors, soldiers, or police, can be assigned during a crisis, as a result of mobile devices that can classify COVID-19.

Table 3
Binary classification results.

Model	Accuracy	Precision	Recall	Specificity	Sensitivity	AUC	MSE
DenseNet121	0.9862	0.9541	0.9648	0.9997	0.9887	0.9933	0.0105
MobileNetV2	0.9884	0.9728	0.9578	1.0	0.9901	0.9944	0.0095
Xception	0.9799	0.9471	0.9324	0.9985	0.9789	0.9864	0.0165
InceptionResNetV2	0.9591	0.8788	0.8776	0.9900	0.9409	0.9588	0.0372
Ensembled Model	0.9900	0.9737	0.9666	1.0	0.9980	0.9974	0.0108

Table 4
Multiple classification results.

Model	Accuracy	Precision	Recall	Specificity	Sensitivity	AUC	MSE
DenseNet121	0.9317	0.9336	0.9308	0.9980	0.9906	0.9883	0.0275
MobileNetV2	0.9216	0.9249	0.9171	0.9983	0.9929	0.9878	0.0302
Xception	0.9152	0.9128	0.9978	0.9975	0.9858	0.9832	0.0338
InceptionResNetV2	0.8909	0.8959	0.8883	0.9910	0.9679	0.9690	0.0453
Ensembled Model	0.9398	0.9437	0.9312	0.9987	0.9979	0.9918	0.0249

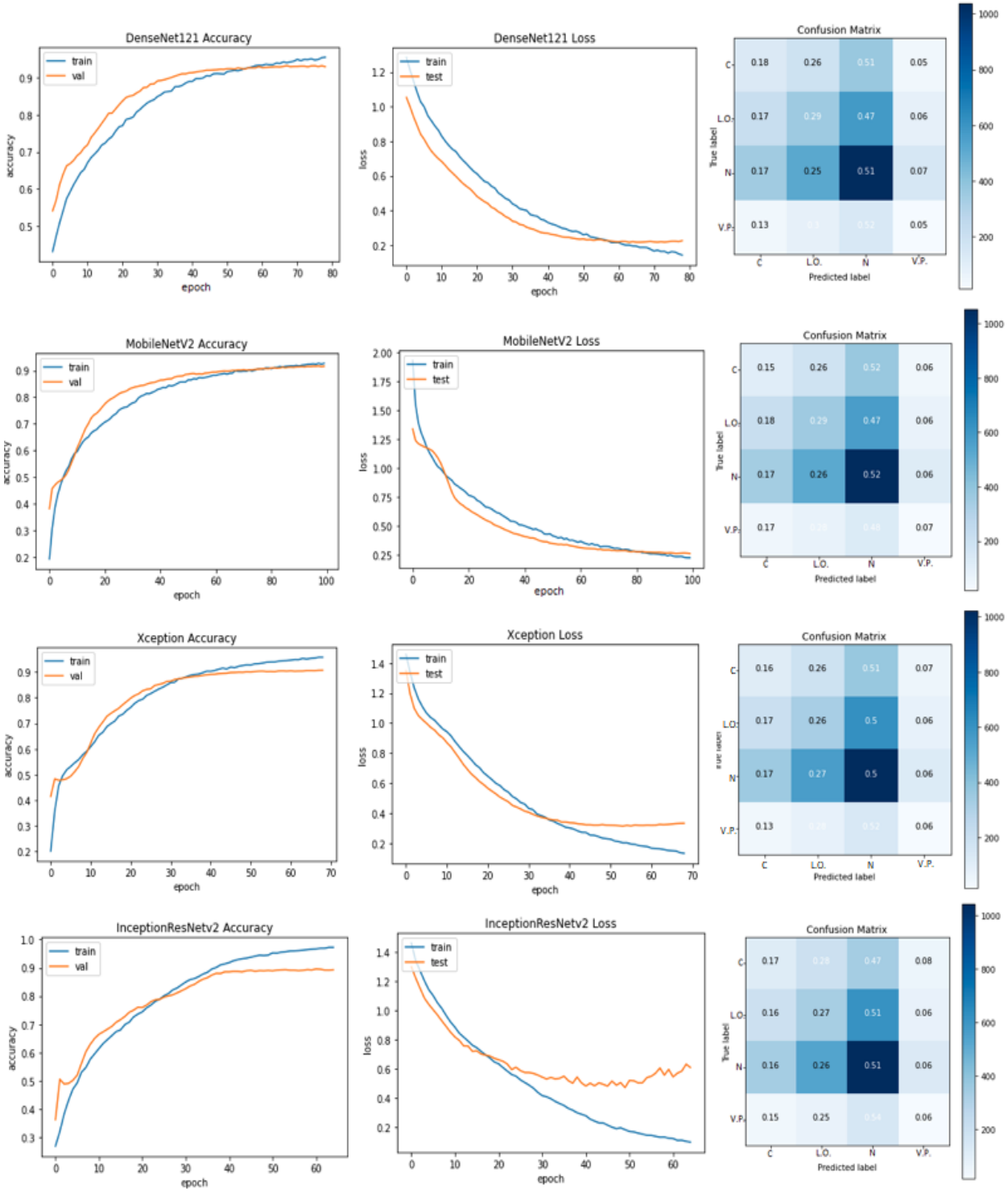


Figure 7. Model performances used in ensemble deep learning for multi classification.

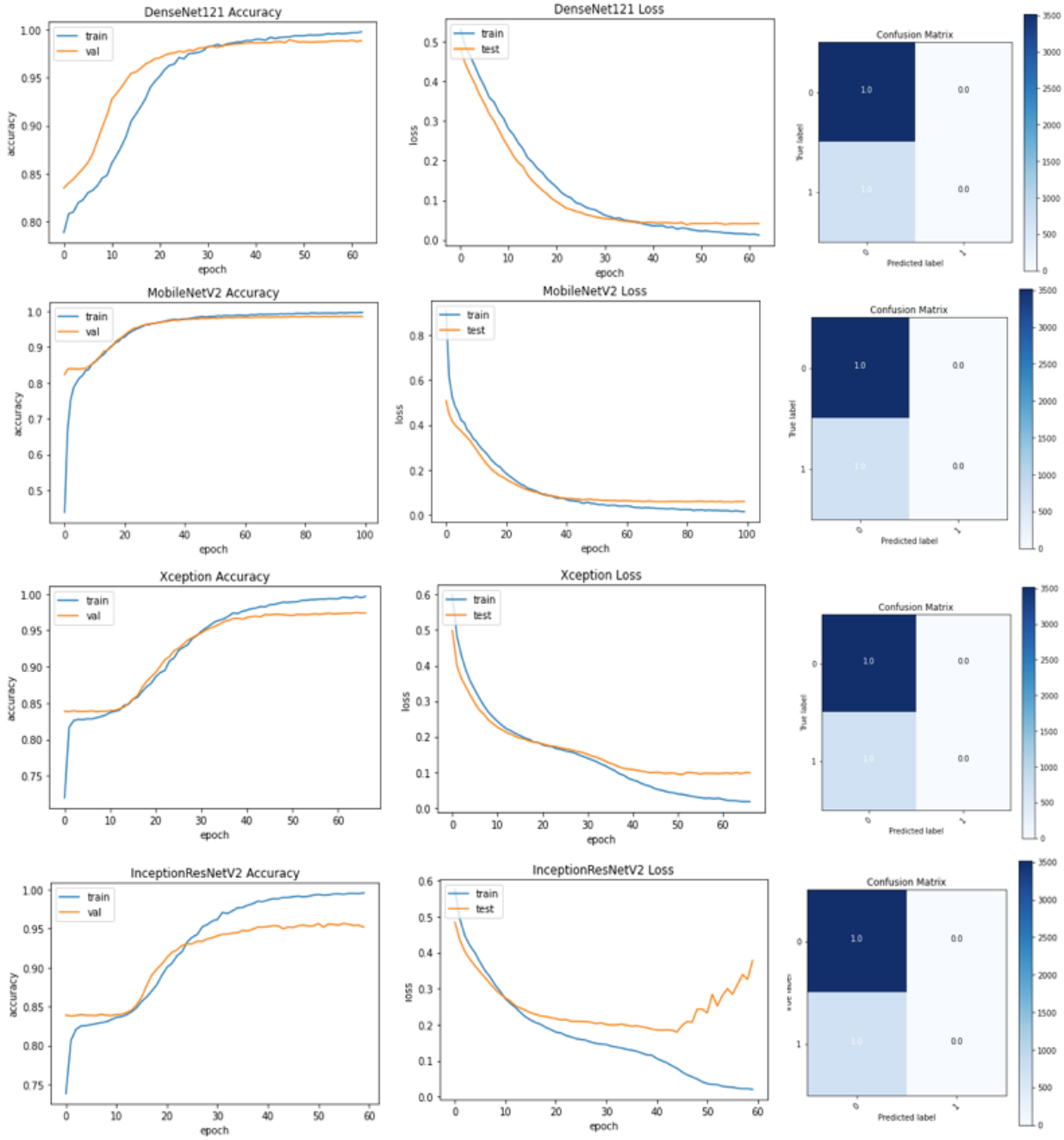


Figure 8. Model performances used in ensemble deep learning for binary classification.

According to the results obtained, it was observed that the ensemble learning approach gave better results in both classification methods. The final model created for binary classification provided a small increase compared to the sub-models that formed it. However, it is believed that the proposed solution will preserve all the advantages of the sub-models and will perform binary classification with high accuracy in future data sets. In multiple learning, on the other hand, a much better generalization was achieved compared to sub-models, and the final model accuracy rate was higher than these models.

Table 5
Proposed methods with related works

Reference	Model	Chest X-Ray Dataset	Classification Accuracy (Binary: B / Multiple: M)
Wang et al.(2020)	COVID-NET	13.800 Images (8.066 Normal, 5.538 Pneumonia, 183 COVID-19)	B: 92.6% M: 83.5%
Ozturk et al.(2020)	DarkCovidNet	1.127 Images (500 Normal, 500 Pneumonia, 127 COVID-19)	B: 98.08% M: 87.02%
Afifi et al.(2021)	Ensemble Deep Learning (DenseNet161 with Different Weights)	11.197 Images (7.217 Normal and Other Lung pulmonary diseases, 5.451 Pneumonia, 1.056 COVID-19)	M: 91.2%
Hemdan et al.(2020)	COVIDX-Net	50 Images (25 COVID-19, 25 Normal	B: 90%
Saha et al.(2020)	Ensemble Deep Learning (MobileNet, InceptiobV3, Xception, DenseNet121, DenseNet201)	1.622 Images (408 Normal, 408 COVID-19, 408 Bacterial Pneumonia, 408 Viral Pneumonia)	M: 89.21%
Rahimzadeh et al.(2020)	Hibrid CNN	15.085 Images (180 COVID-19, 6.054 Pneumonia, 8.851 Normal)	B: 99.50% M: 91.4%
Tang et al. (2021)	Ensemble Deep Learning (Different snapshots of COVID-NET)	15.477 Images (6.053 Pneumonia, 8.851 Normal, 573 COVID-19)	B: 96 %
Ghenea et al. (2021)	Ensemble Deep Learning (VGG-NET based CNN1, CNN2)	8.088 Images (4.044 COVID-19, 4.044 NON-COVID-19)	B: 95.02%
Ouchicha et al.(2020)	CVDNet	2.905 Images (1.341 Normal, 219 COVID-19, 1.345 Viral Pneumonia)	B: 97.20% M: 96.69%
Bhardwaj & Kaur (2021)	Ensemble Deep Learning (InceptionV3, DenseNet121, Xception, InceptionResNetV2)	10.000 Images (2.022 Other Pneumonia, 2.161 COVID-19, 5.563 Normal)	B: 98.33% M: 92.36%
#	Proposed Method	21.165 Images (3.616 COVID-19, 6.012 Lung Opacity, 10.192 Normal, 1.345 Viral Pneumonia)	B: 99.00% M: 93.98%

4. Conclusion

Within the scope of the related research, a solution that can distinguish between coronavirus and other respiratory diseases using chest radiograph images was presented. Limited data are available in open sources due to the new outbreak of the epidemic and limitations such as legal regulations regarding the sharing of patient X-ray images. Despite these limitations, the models in the study were developed with a much larger dataset than the datasets used in the studies in Table 5. The main purpose of the study was determined as finding a solution to produce effective results in data sets of various sizes that may be encountered in the future. For this purpose, a hybrid model based on the DenseNet121, MobileNetV2, Xception, and InceptionResNetV2 models, which show high performance in computer vision and image processing problems, was built. Instead of starting from scratch in the training stages of the preferred models, the initial weights were transferred from the ImageNet dataset.

The advantages of the four CNNs were combined with unweighted model averaging ensemble learning. As a result, better generalization was obtained. In the final model for binary classification, a slight increase was obtained compared to the sub-models, and the result model was 99.00% accuracy, and a higher increase in multi-classification was achieved and the classification was made with an accuracy rate of 93.98%. These results showed that the hybrid architecture built on proven and high-performance CNN models can also show promising performance in future datasets.

Author Contributions

Hacer Karacan: Designed the architecture, supervised and coordinated the study.

Furkan Eryılmaz: Designed the architecture, performed the analysis and wrote the paper

Conflicts of Interest

The authors declare no conflict of interest.

References

Abbas, A., Abdelsamea, M. M., & Gaber, M. M. (2020). Classification of COVID-19 in chest X-ray images using DeTraC deep convolutional neural network. doi:10.1101/2020.03.30.20047456

Afifi, A., Hafsa, N. E., Ali, M. A., Alhumam, A., & Alsalman, S. (2021). An ensemble of global and local-attention Based convolutional neural networks FOR COVID-19 diagnosis on chest x-ray images. Symmetry, 13(1), 113. doi:10.3390/sym13010113

Alam, N.-A.-A., Ahsan, M., Based, M. A., Haider, J., & Kowalski, M. (2021). COVID-19 Detection from Chest X-ray Images Using Feature Fusion and Deep Learning. Sensors, 21(4), 1480. doi:10.3390/s21041480

Allam, Z. (2020). The First 50 days of COVID-19: A Detailed Chronological Timeline and Extensive Review of Literature Documenting the Pandemic. Surveying the Covid-19 Pandemic and Its Implications, 1–7. doi: 10.1016/b978-0-12-824313-8.00001-2

Aloysius, N., & Geetha, M. (2017). A review on deep convolutional neural networks. 2017 International Conference on Communication and Signal Processing (ICCSP), 0588-0592. doi: 10.1109/ICCSP.2017.8286426

Altan, A., & Karasu, S. (2020). Recognition of COVID-19 disease from X-ray images by hybrid model consisting of 2D curvelet transform, chaotic salp swarm algorithm and deep learning technique. Chaos, Solitons & Fractals, 140, 110071. doi: 10.1016/j.chaos.2020.110071

Ankile, L.H., Heggland, M.F., Krange, K. (2020). Deep Convolutional Neural Networks: A survey of the foundations, selected improvements, and some current applications. arXiv:2011.12960

Ardakani, A. A., Kanafi, A. R., Acharya, U. R., Khadem, N., & Mohammadi, A. (2020). Application of deep learning technique to manage COVID-19 in routine clinical practice using CT images: Results of 10 convolutional neural networks. Computers in Biology and Medicine, 121, 103795. doi:10.1016/j.compbiomed.2020.103795

Armiro. (n.d.). Armiro/COVID-CXNet. Retrieved from https://github.com/armiro/COVID-CXNet

Arora, D., Garg, M., & Gupta, M. (2020). Diving deep in Deep Convolutional Neural Network. 2020 2nd International Conference on Advances in Computing, Communication Control and Networking (ICACCCN). doi:10.1109/icacccn51052.2020.9362907

Benmalek, E., Elmhamdi, J., & Jilbab, A. (2021). Comparing CT scan and chest X-ray imaging for COVID-19 diagnosis. Biomedical Engineering Advances, 1, 100003. doi:10.1016/j.bea.2021.100003

Bhardwaj, P., & Kaur, A. (2021). A novel and efficient deep learning approach for COVID -19 detection using X-ray imaging modality. International Journal of Imaging Systems and Technology. doi:10.1002/ima.22627

BIMCV. (n.d.). Retrieved from https://bimcv.cipf.es/bimcv-projects/bimcv-covid19/#1590858128006-9e640421-6711

Bouchareb, Y., Khaniabadi, P. M., Kindi, F. A., Dhuhli, H. A., Shiri, I., Zaidi, H., & Rahmim, A. (2021, July 21). Artificial intelligence-driven assessment of radiological images for COVID-19. Retrieved from https://www.sciencedirect.com/science/article/pii/S0010482521004595

Bozkurt, F. (2021). Derin Öğrenme Tekniklerini Kullanarak Akciğer X-Ray Görüntülerinden COVID-19 Tespiti. European Journal of Science and Technology, 149-156. doi:10.31590/ejosat.898385

Chen, S., Dobriban, E., & Lee, J.H. (2019). Invariance reduces Variance: Understanding Data Augmentation in Deep Learning and Beyond. ArXiv: abs/1907.10905

Chollet, F. (2017). Xception: Deep Learning with Depthwise Separable Convolutions. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). doi:10.1109/cvpr.2017.195

Eurorad.org. (n.d.). Retrieved from https://www.eurorad.org/advanced-search?filter[0]=section:40

Feki, I., Ammar, S., Kessentini, Y., & Muhammad, K. (2021). Federated learning for COVID-19 screening from Chest X-ray images. Applied Soft Computing, 106, 107330. doi:10.1016/j.asoc.2021.107330

Ganaie, m. A., Hu, M., Tanveer, M., Suganthan, P. N., (2021). Ensemble deep learning: A review. https://arxiv.org/abs/2104.02395

Ghenea, G. L., & Neagoe, V. E. (2021, July). Concurrent Convolutional Neural Networks with Decision Fusion to Diagnose COVID-19 using Chest X-ray Imagery. In 2021 13th International Conference on Electronics, Computers and Artificial Intelligence (ECAI) (pp. 1-4). IEEE. doi: 10.1109/ECAI52376.2021.9515174

Goodfellow, I., Bengio, Y., & Courville, A. (2017). Deep learning. Cambridge, MA: The MIT Pr

Gulcu, A., & Kus, Z. (2020). Hyper-Parameter Selection in Convolutional Neural Networks Using Microcanonical Optimization Algorithm. IEEE Access, 8, 52528-52540. doi:10.1109/access.2020.2981141

Haghanifar, A., Majdabadi, M. M., & Ko, S. (2021, May 20). COVID-19 Chest X-Ray Image Repository. Retrieved from https://figshare.com/articles/dataset/COVID-19_Chest_X-Ray_Image_Repository/12580328

Hassantabar, S., Ahmadi, M., & Sharifi, A. (2020). Diagnosis and detection of infected tissue of COVID-19 patients based on lung x-ray image using convolutional neural network approaches. Chaos, Solitons & Fractals, 140, 110170. doi:10.1016/j.chaos.2020.110170

Hemdan, E. E., Shouman, A. M., Karar, M. E. (2020). COVIDX-Net: A Framework of Deep Learning Classifiers to Diagnose COVID-19 in X-Ray Images. arXiv:2003.11055

Horry, M. J., Chakraborty, S., Paul, M., Ulhaq, A., Pradhan, B., Saha, M., & Shukla, N. (2020). COVID-19 Detection Through Transfer Learning Using Multimodal Imaging Data. IEEE Access, 8, 149808-149824. doi:10.1109/access.2020.3016780

Huang, G., Liu, Z., Maaten, L. V., & Weinberger, K. Q. (2017). Densely Connected Convolutional Networks. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). doi:10.1109/cvpr.2017.243

Huang, P., Liu, T., Huang, L., Liu, H., Lei, M., Xu, W., . . . Liu, B. (2020). Use of Chest CT in Combination with Negative RT-PCR Assay for the 2019 Novel Coronavirus but High Clinical Suspicion. Radiology, 295(1), 22-23. doi:10.1148/radiol.2020200330

Ieee8023. (n.d.). Ieee8023/covid-chestxray-dataset. Retrieved from https://github.com/ieee8023/covid-chestxray-dataset

Karthik, R., Menaka, R., & M, H. (2021). Learning distinctive filters for COVID-19 detection from chest X-ray using shuffled residual CNN. Applied soft computing, 99, 106744. doi: 10.1016/j.asoc.2020.106744

Konig, J., Jenkins, M. D., Barrie, P., Mannion, M., & Morison, G. (2019). A Convolutional Neural Network for Pavement Surface Crack Segmentation Using Residual Connections and Attention Gating. 2019 IEEE International Conference on Image Processing (ICIP). doi:10.1109/icip.2019.8803060

Ml-Workgroup. (n.d.). Ml-workgroup/covid-19-image-repository. Retrieved from https://github.com/ml-workgroup/covid-19-image-repository/tree/master/png

Monshi, M. M. A., Poon, J., Chung, V., & Monshi, F. M. (2021). CovidXrayNet: Optimizing data augmentation and CNN hyperparameters for improved COVID-19 detection from CXR. Computers in Biology and Medicine, 133, 104375. doi: 10.1016/j.compbiomed.2021.104375

Mooney, P. (2018, March 24). Chest X-Ray Images (Pneumonia). Retrieved from https://www.kaggle.com/paultimothymooney/chest-xray-pneumonia

Narin, A., Kaya, C., & Pamuk, Z. (2021). Automatic detection of coronavirus disease (COVID-19) using X-ray images and deep convolutional neural networks. Pattern Analysis and Applications. doi:10.1007/s10044-021-00984-y

Oh, Y., Park, S., & Ye, J. C. (2020). Deep Learning COVID-19 Features on CXR Using Limited Training Data Sets. IEEE Transactions on Medical Imaging, 39(8), 2688-2700. doi:10.1109/tmi.2020.2993291

Ouchicha, C., Ammor, O., & Meknassi, M. (2020). CVDNet: A novel deep learning architecture for detection of coronavirus (Covid-19) from chest x-ray images. Chaos, Solitons & Fractals, 140, 110245. doi:10.1016/j.chaos.2020.110245

Ozturk, T., Talo, M., Yildirim, E. A., Baloglu, U. B., Yildirim, O., & Acharya, U. R. (2020). Automated detection of COVID-19 cases using deep neural networks with X-ray images. Computers in Biology and Medicine, 121, 103792. doi:10.1016/j.compbiomed.2020.103792

Qu, J., Cao, B., & Chen, R. (2021). Respiratory virus and COVID-19. Covid-19, 1-6. doi:10.1016/b978-0-12-

824003-8.00001-2

Rahimzadeh, M., & Attar, A. (2020). A modified deep convolutional neural network for detecting COVID-19 and pneumonia from chest X-ray images based on the concatenation of Xception and ResNet50V2. Informatics in Medicine Unlocked, 19, 100360. doi:10.1016/j.imu.2020.100360

Rahman, T. (2021, March 06). COVID-19 Radiography Database. Retrieved from https://www.kaggle.com/tawsifurrahman/covid19-radiography-databas

Redazione. (2021, May 20). COVID-19 DATABASE. Retrieved from https://www.sirm.org/category/senza-categoria/covid-19/

Rismiyati, Endah, S. N., Khadijah, & Shiddiq, I. N. (2020). Xception Architecture Transfer Learning for Garbage Classification. 2020 4th International Conference on Informatics and Computational Sciences (ICICoS). doi:10.1109/icicos51170.2020.9299017

RSNA Pneumonia Detection Challenge. (n.d.). Retrieved from https://www.kaggle.com/c/rsna-pneumonia-detection-challenge/data

Rubin, G. D., Ryerson, C. J., Haramati, L. B., Sverzellati, N., Kanne, J. P., Raoof, S., . . . Leung, A. N. (2020). The Role of Chest Imaging in Patient Management during the COVID-19 Pandemic: A Multinational Consensus Statement from the Fleischner Society. Radiology, 296(1), 172-180. doi:10.1148/radiol.2020201365

Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., & Chen, L. (2018). MobileNetV2: Inverted Residuals and Linear Bottlenecks. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. doi:10.1109/cvpr.2018.0047

Saha, O., Tasnim, J., Raihan, M. T., Mahmud, T., Ahmmed, I., & Fattah, S. A. (2020). A multi-model BASED Ensembling approach to Detect COVID-19 from chest x-ray images. 2020 IEEE REGION 10 CONFERENCE (TENCON). doi: 10.1109/tencon50793.2020.9293802

Szegedy, C., Ioffe, S., Vanhoucke, V., & Alemi, A. (2017). Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning. Proceedings of the AAAI Conference on Artificial Intelligence, 31(1). Retrieved from https://ojs.aaai.org/index.php/AAAI/article/view/11231

Tan, C., Sun, F., Kong, T., Zhang, W., Yang, C., & Liu, C. (2018). A Survey on Deep Transfer Learning. ICANN. Arxiv: abs/1808.01974

Tang, S., Wang, C., Nie, J., Kumar, N., Zhang, Y., Xiong, Z., & Barnawi, A. (2021). EDL-COVID: Ensemble Deep Learning for COVID-19 Case Detection From Chest X-Ray Images. IEEE Transactions on Industrial Informatics, 17, 6539-6549. doi: 10.1109/TII.2021.3057683

Wang, D., Hu, B., Hu, C., Zhu, F., Liu, X., Zhang, J., . . . Peng, Z. (2020). Clinical Characteristics of 138 Hospitalized Patients With 2019 Novel Coronavirus–Infected Pneumonia in Wuhan, China. Jama, 323(11), 1061. doi:10.1001/jama.2020.1585

Wang, L., Lin, Z. Q., & Wong, A. (2020). COVID-Net: A tailored deep convolutional neural network design for detection of COVID-19 cases from chest X-ray images. Scientific Reports, 10(1). doi:10.1038/s41598-020-76550-z

Wang, S., & Zhang, Y. (2020). DenseNet-201-Based Deep Neural Network with Composite Learning Factor and Precomputation for Multiple Sclerosis Classification. ACM Transactions on Multimedia Computing, Communications, and Applications, 16(2s), 1-19. doi:10.1145/3341095

Wang, Z., Xiao, Y., Li, Y., Zhang, J., Lu, F., Hou, M., & Liu, X. (2021). Automatically discriminating and localizing COVID-19 from community-acquired pneumonia on chest X-rays. Pattern Recognition, 110, 107613. doi:10.1016/j.patcog.2020.107613

WHO Coronavirus (COVID-19) Dashboard. (n.d.). Retrieved from https://covid19.who.int/

Wu, X., Liu, R., Yang, H., & Chen, Z. (2020). An Xception Based Convolutional Neural Network for Scene Image Classification with Transfer Learning. 2020 2nd International Conference on Information Technology and Computer Application (ITCA). doi:10.1109/itca52113.2020.00063

Yang, Y., Lv, H., Chen, N., Wu, Y., Zheng, J., & Zheng, Z. (2021). Local minima found in the subparameter space can be effective for ensembles of deep convolutional neural networks. Pattern Recognition, 109, 107582. doi:10.1016/j.patcog.2020.107582

Zhuang, F., Qi, Z., Duan, K., Xi, D., Zhu, Y., Zhu, H., Xiong, H., & He, Q. (2021). A Comprehensive Survey on Transfer Learning. Proceedings of the IEEE, 109, 43-76. Arxiv: abs/1911.02685