OKU Fen Bilimleri Enstitüsü Dergisi
6(Ek Sayı): 1-17, 2023

OKU Journal of The Institute of Science and Technology, 6(Suppl): 1-17, 2023

Osmaniye Korkut Ata Üniversitesi Fen Bilimleri Enstitüsü Dergisi

Osmaniye Korkut Ata University Journal of The Institute of Science and Technology

COVID-19 Teşhisi İçin Derin Topluluk Öğrenmeye Dayalı Modellerde Temel Füzyon Tekniklerinin Etkisi

Yaşar DAŞDEMİR[1*], Hafize ARDUÇ[2]

[1,2]Erzurum Teknik Üniversitesi Mühendislik ve Mimarlık Fakültesi Bilgisayar Mühendisliği Bölümü, Erzurum

[1]https://orcid.org/0000-0002-9141-0229
[2]https://orcid.org/0000-0002-2231-3580
*Sorumlu yazar: yasar.dasdemir@erzurum.edu.tr

Araştırma Makalesi

Makale Tarihçesi:
Geliş tarihi: 17.12.2022
Kabul tarihi:16.04.2023
Online Yayinlanma: 20.12.2023

Anahtar Kelimeler:
COVID-19
Topluluk öğrenimi
Transfer öğrenimi
Füzyon
Maksimum oylama

ÖZ

Küresel salgın hastalık (pandemi) olarak deklare edilen koronavirüs hastalığı (COVID-19), yeni bir viral solunum yolu hastalığıdır. Hastalık insandan insana damlacık veya temas yoluyla bulaşmaktadır. Hastalığın yayılmasını önlemek için hızlı tanı oranları ile hastalığı erken tespit etmek çok önemlidir. Ancak uzun süren patolojik laboratuvar testleri ve test sonuçlarındaki düşük tanı oranı araştırmacıları farklı teknikleri uygulamaya yöneltmiştir. Radyolojik görüntüleme ile birlikte derin öğrenme tekniklerinin uygulanması bu hastalığın doğru tespitinde oldukça önemli bir yere sahiptir. Bu çalışmada, COVID-19 X-ray veri seti kullanılarak temel füzyon fonksiyonlarının topluluk öğrenme algoritmaları üzerindeki sınıflandırma performansına etkisi araştırılmıştır. Farklı derin öğrenme modellerini birleştirmek için iki farklı topluluk modeli oluşturuldu; Topluluk-1 (Ens-1) ve Topluluk-2 (Ens-2). Bu topluluk modellerinde Maksimum, Mod, Toplam, Ortalama ve Çarpım gibi temel füzyon fonksiyonları test edilmiştir. Elde edilen değerler incelendiğinde Maksimum ve Çarpım temel füzyon fonksiyonlarının sınıflandırma performansı üzerinde olumlu bir etkiye sahip olduğu görülmektedir. Çoklu sınıflandırmada, hem Ens-1 hem de Ens-2 için Max işlevi sırasıyla %85 ve %86 doğruluk oranıyla öne çıkıyor. Product fonksiyonu, ikili sınıflandırmada %99 ile en yüksek performansı elde etmiştir. Sonuçlar, füzyon yöntemlerinin ikili sınıflandırmada daha iyi sınıflandırma performansı elde edebileceğini göstermektedir.

The Effect of Basic Fusion Techniques in Deep Ensemble Learning-Based Models For COVID-19 Diagnosis

Research Article

Article History:
Received: 17.12.2022
Accepted: 16.04.2023
Published online: 20.12.2023

Keywords:
COVID-19
Ensemble learning
Transfer learning
Fusion
Max voting

ABSTRACT

The coronavirus (COVID-19), declared a global epidemic (pandemic), is a new viral respiratory disease. The disease is transmitted from person to person through droplets or contact. It is very important to detect the disease early with rapid diagnosis rates to prevent the spread of the disease. However, long-term pathological laboratory tests and low diagnosis rates in test results led researchers to apply techniques. Radiological imaging has begun to be used to monitor COVID-19 disease as well as being useful in detecting various lung diseases. The application of deep learning techniques together with radiological imaging has a very important place in the correct detection of this disease. This study investigated the effect of basic fusion functions on classification performance on ensemble learning algorithms using the COVID-19 X-ray

dataset. Two ensemble models were created to combine deep learning models; Ensemble-1 (Ens-1) ve Ensemble-2 (Ens-2). The basic fusion rules of Max, Mode, Sum, Average, and Product were tested in these ensemble models. When the obtained values are examined, it is seen that the Max and Product basic fusion functions have a positive effect on the classification performance. In multi-classification, the Max function for Ens-1 and Ens-2 becomes prominent with an accuracy rate of 85% and 86%, respectively. The Product function achieved the highest performance with 99% in binary classification. The results show that the fusion methods can achieve better classification performance in binary classification.

To Cite: Daşdemir Y., Arduç H. The Effect of Basic Fusion Techniques in Deep Ensemble Learning-Based Models For COVID-19 Diagnosis. Osmaniye Korkut Ata Üniversitesi Fen Bilimleri Enstitüsü Dergisi 2023; 6(Ek Sayı): 1-17.

1. Introduction

The World Health Organization (WHO) has named the deadly virus originating in China "COVID-19". The institution used the following abbreviations while giving this name; "co" means corona, "vi" means a virus, and "d" comes from the initials of the English word "disease". The number 19 in COVID-19 indicates December 31, the date the disease was first identified. In retrospect, coronaviruses have caused 3 epidemics in the last 20 years: MERS, SARS, and COVID-19 (Li et al., 2020).

The WHO used the term COVID-19 to describe the disease caused by the virus, making the virus the official name SARS-COV-2 (Severe Acute Respiratory Syndrome-Coronavirus-2). On January 31, 2020, COVID-19 was declared a global health emergency by the WHO. Due to the virus, a pandemic, that is, a global epidemic, was declared on March 11, 2020. Clinical symptoms of the disease include shortness of breath, cough, fatigue, fever, sore throat-head-muscle pain (Singhal, 2020).

The most common testing technique used for diagnosing COVID-19 is RT-PCR, a real-time reverse testing technique. However, RT-PCR infrastructure is insufficient in most regions with the COVID-19 outbreak. RT-PCR testing is considered time-consuming and tedious with a complex manual procedure. The standard screening tools for early detection and diagnosis of thoracic and lung diseases, including COVID-19, are radiological imaging techniques such as chest x-ray (CXR) and computed tomography (CT) (Ozturk et al., 2020). The advantages of radiology techniques are that they can be detected in the early stages of the disease even if the scan is negative in some cases, and the disadvantages are that X-ray analysis requires a radiologist and manual reading is time-consuming. Therefore, it is necessary to develop an automatic analysis system that will detect abnormalities in scans to save time for healthcare professionals (Ouchicha et al., 2020). Thus, on March 16, 2020, America's white house announced that it is encouraging experts and researchers to use artificial intelligence (AI) techniques to combat the new COVID-19 pandemic (Alimadadi et al., 2020).

Two important scientific communities have emerged to combat COVID-19. The first is a collection of Artificial Intelligence (AI) in the form of automatic COVID-19 detection from a dataset of computed

tomography (CT) scans and X-ray images. The second are mathematicians and epidemiologists who develop complex patterns of virus spread and transmission (Shuja et al., 2020).

This study uses commonly used deep learning approaches, fusion techniques, and X-Ray images to create an effective model. Among these pre-trained deep learning approaches, ResNet-50, ResNet-152, VGG-19, and DenseNet201 methods were selected (Narin et al., 2021). Three different scenarios were applied with these methods. First, patients with COVID-19 were classified among three different conditions (Normal, COVID-19, and Pneumonia). Secondly, cases with COVID-19 were classified among 2 different conditions (COVID-19, Normal). As the third, cases with COVID-19 were classified among 2 different conditions (COVID-19, Pneumonia). Two new models were created using these models. Our first model ENSEMBLE-1 consists of ResNet152, VGG-19, DenseNet201 models, and our second model ENSEMBLE-2 consists of ResNet50, VGG-19, DenseNet201 models. The results of these models are combined with the fusion methods at the decision level, and the effects of these methods on the classification performance are examined.

The remainder of this paper is organized as follows: Section 2 describes related work on COVID-19 and CNN models. Section 3 introduces the details of the methodology, and the results are discussed in Section 4. Finally, the paper ends with a general conclusion.

2. Related Work

When chest X-ray and CT studies in the literature are examined, it is seen that handmade and recently CNN-based methods have been used. Convolutional neural network (CNN) has shown promising results in the diagnosis and classification of various diseases. However, CNN needs a huge amount of labeled datasets for initial training. When it comes to medical images, it is difficult to obtain a large number of labeled images. In such cases, pre-trained CNNs can use large numbers of images (such as ImageNet). Pre-trained CNN has been successful in predicting COVID-19. Scientists were motivated by the need for rapid interpretation of X-ray images. Therefore, they proposed deep learning models (especially convolutional neural networks) to detect COVID-19-infected cases from chest X-ray imaging.

Autoregressive Integrated Moving Average (ARIMA), Nonlinear Autoregression Neural Network (NARNN), and Long Short-Term Memory (LSTM) approaches have been used to predict COVID-19 (Kırbas et al., 2020). Studies also present a new model called DarkCovidNet, which is based on the DarkNet model using X-Ray images (Ozturk et al., 2020). DarkCovidNet has been tested on binary classification (COVID-19andNormal) and multi-class classification (NormalandCOVID-19andPneumonia) for COVID-19 detection. DarkCovidNet achieved an accuracy of 98.08% for binary classes and 87.02% for multi-class situations. Abraham et al. proposed a method for high-speed recognition of COVID-19 from chest X-ray images, using features from multiple CNNs. The method used a feature selection approach based on correlation and the BayesNet classifier (Abraham and Nair, 2020). Ouchicha et al. used a neural network-focused model that uses local and global features

of chest X-ray images using two parallel layers with various kernel sizes. With this model, called CVDNet, they achieved an average accuracy of 97.20% (Ouchicha et al., 2020).

In addition, manual techniques were used together with CNN models. Detection and diagnosis were performed using MLP on fractal features and CNN on CXR. After the images were first converted to 1-dimensional vectors, the covariance matrix was calculated and fractal features were extracted with the help of eigenvalues and eigenvectors. The CNN architecture achieved a higher accuracy of 93.2% (Hassantabar et al., 2020). Studies have used pre-trained models ResNet-18, ResNet-50, ResNet-101, VGG-16, and VGG-19 for feature extraction from chest X-ray images. Using a Support Vector Machine (SVM) classifier implemented with different kernel parameters, they achieved 94.7% accuracy on ResNet and SVM with linear kernel (Ismael and Şengür, 2021).

It is challenging to find labeled images of new diseases such as COVID-19, as training CNN requires large amounts of data. Transfer learning methods were also used to solve this problem (Xu et al., 2020). Using X-Ray images, they compared the performance of GoogleNet, ResNet101, Xception, and MobileNetv2 with the transfer learning method to solve the medical image imbalance problem. Similarly, with X-Ray images, Şahinbaş and Çatak used VGG16, VGG19, ResNet, DenseNet, and InceptionV3 models to diagnose COVID-19 (Sahinbas and Catak, 2020).

Tuncer et al. tried a COVID-19 classification method using a residual sample local binary model called ResExLBP. However, a limited number of positive and negative labeled data were used in the study (Tuncer et al., 2020). Nilanjan et al. developed a new method to detect Pneumonia by calculating features from chest X-ray using a modified VGGNet19 (Visual Geometry Group Network) (Dey et al., 2021). This method achieved 97.94% classification accuracy. Ardakani et al. applied the deep learning technique to manage COVID-19 in routine clinical practice using CT images. They used 10-CNNs to distinguish COVID-19 disease from healthy individuals and presented the characteristics of these networks (Ardakani et al., 2020).

In the study where the performances of 5 different machine learning and deep learning algorithms were compared (Özbay and Altunbey Özbay, 2020), 98.1% accuracy was obtained with the Convolutional Neural Network algorithm. In another experimental study (Bozkurt, 2021), 98.17% success was achieved with the DenseNet121 model. In another study (Korkmaz and Atila B, 2018) that detects lung infection caused by Covid-19 using deep learning techniques, three different classes were used, and the classification success was found to be 97%.

The fusion method has been applied before and has been found to work well in many areas. The main contribution of this study is testing this method on current data such as Covid-19. The basic fusion functions have achieved superior classification performance, especially with the max and product functions.

3. Material and Method

3.1. X-ray Image DataSet

The data set was created by combining two data sets (Ozturk et al., 2020). The COVID-19 X-ray image database using images from various open-access sources was used in the study (Cohen et al., 2020). The authors compiled radiology images of COVID-19 cases from various sources (Radiology Society (RSNA), Radiopaedia, etc.) for research purposes, and most of the studies on COVID-19 use images from this source.

This dataset is constantly updated with data shared by researchers from different regions. In addition, the ChestX-ray8 dataset provided by (X. Wang et al., 2017) was used for images of Pneumonia and a normal chest. A total of 1125 pictures were collected from these two sources. The dataset comprises 500 normal, 500 pneumonia cases, and 125 COVID-19 cases. The summary of the dataset used in the study is given in Table 1.

Table 1. Dataset summary

	Disease	**#No. of Images**
1	Normal (Healthy)	500
2	Pneumonia	500
3	COVID-19	125
4	**Total**	**1125**

Some chest X-ray image samples from the prepared data set are given in Figure 1.

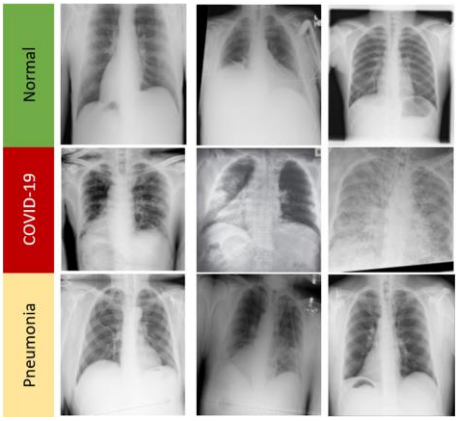


Figure 1. Chest X-ray images samples from the prepared dataset
(Top: Normal, Medium: COVID-19, Bottom: Pneumonia)

3.2. Deep Learning Methods

Deep learning, a sub-branch of Machine Learning, is a study area that covers artificial neural networks and similar machine learning algorithms that contain one or more hidden layers. Convolutional neural networks (CNN), on the other hand, are a sub-branch of deep learning. CNN is generally used to analyze visual information. A CNN has three basic types of layers: Convolutional Layer, Pooling Layer, and Fully-connected layer. Multiple convolution+pooling can be done in succession. Then there are several fully connected layers. Feature extraction occurs in both convolutional and pooling

layers. Deep learning techniques continue to perform impressively in medical image processing, as in many other fields. Deep learning algorithms are widely used in many areas, such as classification, segmentation, and lesion detection of medical data.

Transfer Learning is the storage of the information obtained while solving a problem and using that information when faced with another problem. Because Transfer Learning uses previous knowledge, it can achieve higher-performing and faster learning models with less training data. One of the biggest challenges researchers face in analyzing medical data is the limited number of data sets. Deep learning models often need many data. Labeling this data by experts is both time-consuming and costly. The Transfer Learning method enables data training with fewer data and requires less computational cost.

In this study, CNN-based ResNet-34, ResNet-50, ResNet-101, ResNet-152, VGG-19 and DenseNet201 models were created for COVID-19 classification. In addition, the transfer learning method using ImageNet data was applied to eliminate the problems of insufficient data and training time.

The Residual Neural Network (ResNet) is an artificial neural network (ANN) that uses residual neural networks. It is residual-based and uses shortcuts or jumpers to jump over some layers. It is called ResNet-50 because it has 50 layers. ResNet adds links between layers to solve a problem (Figure 2). In this way, it prevents the disruptions that occur in the deepening and complex network. In addition, training is tried to be done faster with bottleneck blocks. The general flow of parameters used for ResNet50 training is given in Table 2. In this study, Collaboratory (or Colab for short), a product of Google Research, was preferred as a training environment.

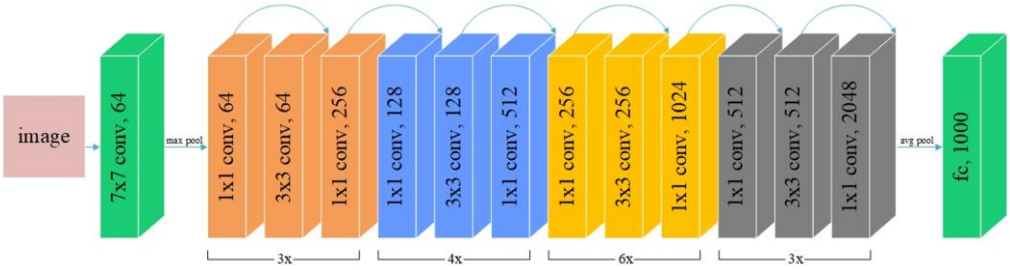


Figure 2. ResNet-50 block diagram

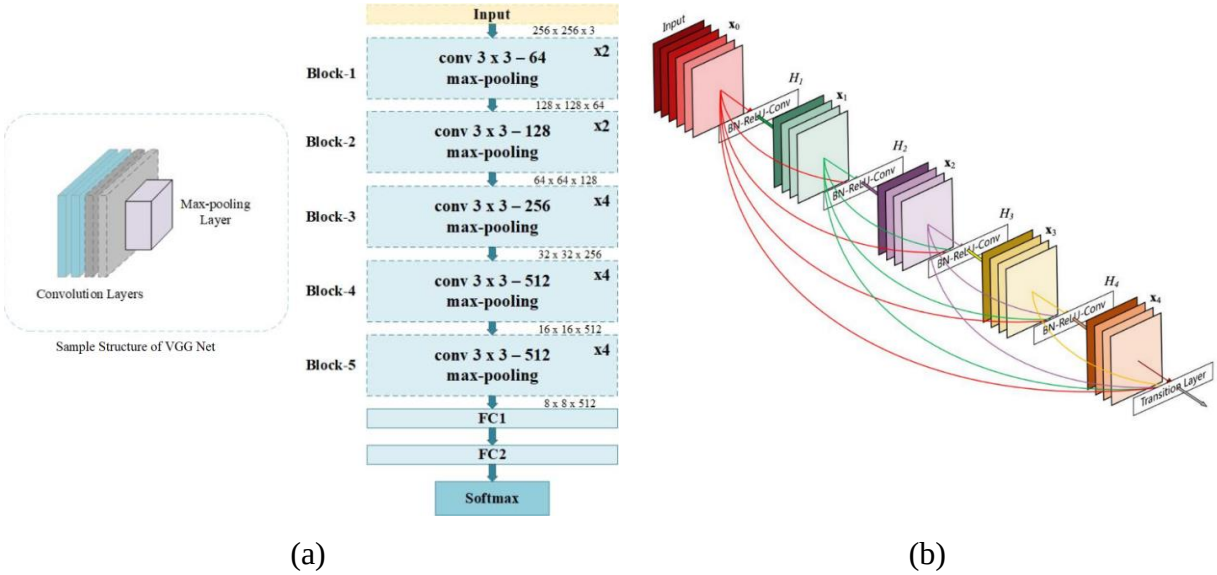
Deep convolutional networks use a deeper number of layers to improve classification and recognition accuracy, solving complex tasks for image classification. In neural networks, however, this causes the accuracy to saturate in the deep layers and then degrade. Residual training deals with this problem.

Table 2. ResNet50 training parameters flow
(Total params: 25,616,451; Total trainable params: 2,161,539; Total non-trainable params: 23,454,912)

Layer type	Output Shape	Param #
Conv2d/BatchNorm2d/ReLU	[64, 128, 128]	9408/128/0
MaxPool2/Conv2d/ BatchNorm2d/Conv2d/BatchNorm2d	[64, 64, 64]	4096/128/36864/128
Conv2d/BatchNorm2d/ReLU/Conv2d/BatchNorm2d	[256, 64, 64]	16384/512/0/16384/512
Conv2d/BatchNorm2d/Conv2d/BatchNorm2d	[64, 64, 64]	16384/128/36864/128
. . .	. . .	. .
Conv2d/BatchNorm2d/ReLU/AdaptiveAvgPool2d /AdaptiveMaxPool2d	[2048, 8, 8]	1048576/4096/0/0
Flatten/BatchNorm1d/Dropout	[4096]	0/8192/0
Linear/ReLU/BatchNorm1d/Dropout	[512]	2097664/0/1024/0
Linear	[3]	1539

The architecture of the VGG (Visual Geometry Group) network (VGG19) includes 19 trainable weighted layers (Simonyan and Zisserman, 2014). The block structure of the VGG network consists of a combination of convolution and maximum pooling operations. VGG blocks start with two convolutional layers with 64 and 128 filters, respectively. Then, the third layer contains 256 filters (Figure 3-a). VGG19 has been initialized with ImageNet weights and is preferred as a transfer learning model.

Densely Connected Convolutional Networks (DenseNet) is a deep CNN proposed by Gao Huang et al (Huang et al., 2017). In DenseNet, each layer is fed-forward to all other layers, so DenseNet has n(n+1)/2 connections in total (Figure 3-b). The feature maps of the previous layers are used as input to the active layer, and the feature map of the active layer is used as the input to the next layer.



(a) (b)

Figure 3. A visualization of the models (a) Fine-tuned architecture of VGG19 for COVID-19 classification (b) DenseNet 201 architecture (Huang et al., 2017).

3.3. Ensemble Learning and Fusion Rules

Ensemble Learning is a machine learning prototype created by weak learners coming together to solve a similar problem. This learning method enables to create a model with more than one learner instead of training the model with a single learner. The output of each individual or single learner is taken as 'votes'. The final output (decision) is based on the majority response (Figure 4). Ensemble methods attempt to generate a set of hypotheses and combine them for use. Ensemble-based classification systems require a scoring mechanism to combine the prediction results of the elements of the ensemble and produce the final output. For this purpose, different fusion approaches such as max, majority vote, sum, mean, and product are used.

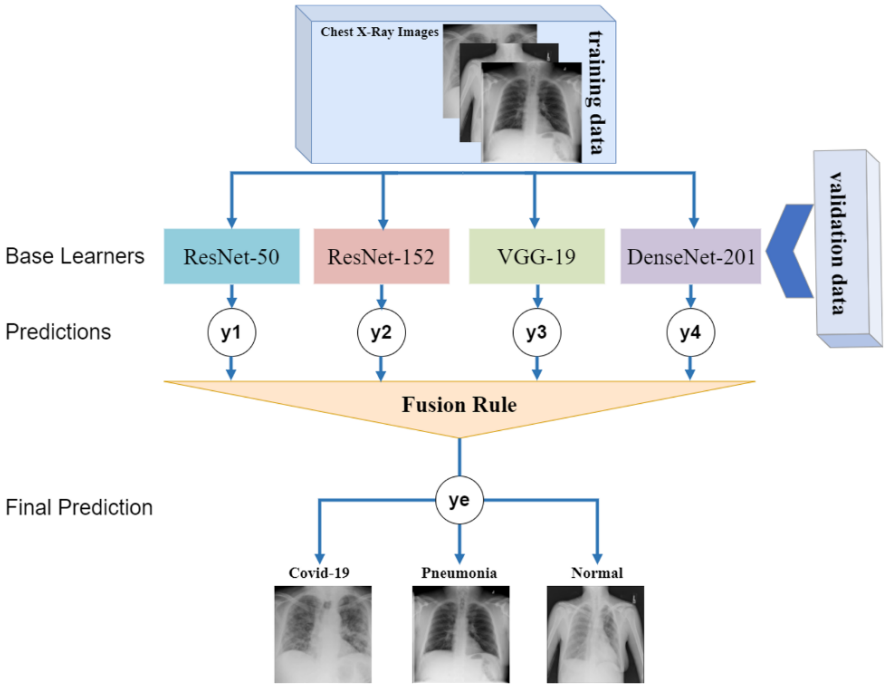


Figure 4. The flow diagram of the study.

The proposed framework is presented in Figure 4. Here, 4 CNN-based classifiers are seen. These classifiers are considered base learners. Each learner y_i will predict a result $y_{ensemble}$ from the class set {c_1:COVID-19,c_2:Pneumonia,c_3:Normal}. The ensemble strategy is formulated as follows.

$$y_{e(nsemble)} = c_{\arg\max_j \sum_{i=1}^{T} y_i^j} \quad (1)$$

wherein, multi-dimensional vector (y_{i1},y_{i2},…y_{iT}) represents the predicted output of y_i on the training data sample and $y_{i,j}$ represents the output of y_i on class c_j. T is the number of learners with a value of 4. Given the input feature vector X, the i.output of the classifier is shown as Eq. 2 for the three classes.

$$y_i(X) = \left[y_{i,1}(X), y_{i,2}(X), y_{i,3}(X)\right]^T \quad (2)$$

The fused output of the four classifiers is constructed as in Eq. 3.

$$y(X) = \text{Fusion}(y_1(X), y_2(X), y_3(X), y_4(X)) \quad (3)$$

The output of all basic learners can be represented as a decision profile in the form of a 4x3 matrix (Eq.4).

$$\mathrm{DP}(X) = \begin{bmatrix} y_{1,1}(X) & y_{1,2}(X) & c_{1,3}(X) \\ y_{2,1}(X) & y_{2,2}(X) & y_{2,3}(X) \\ y_{3,1}(X) & y_{3,2}(X) & c_{3,3}(X) \\ y_{4,1}(X) & y_{4,2}(X) & y_{4,3}(X) \end{bmatrix} \tag{4}$$

Additionally, the fusion result y(X) is a 3-dimensional vector represented by the measure layer form (d represents the decision for each class) shown in Eq. 5.

$$y(X) = [d_1(X), d_2(X), d_3(X)]^T \tag{5}$$

The fusion rules given in Table 3 are applied to each column of DP (X) and give the fusion output as a result of the operation. The ensemble scoring by fusion rules is given by Eq.1.

Table 3. Fusion rules (i: classifier number, j: class label)

Fusion rules	**Formula**
Max	$d_j(X) = \max_{i=1}^{l} y_{i,j}(X), j = 1,2,\dots,m$
Majority Vote	$d_j(X) = \underset{i=1}{l} \operatorname{argmax} y_{i,j}(X), j = 1,2,\dots,m$
Sum	$d_j(X) = \sum_{i=1}^{l} y_{i,j}(X), j = 1,2,\dots,m$
Mean	$d_j(X) = \frac{1}{l}\sum_{i=1}^{l} y_{i,j}(X), j = 1,2,\dots,m$
Product	$d_j(X) = \prod_{i=1}^{l} y_{i,j}(X), j = 1,2,\dots,m$

4. Results and Discussions

Using Machine Learning techniques to classify the COVID-19 disease has become quite common. Especially in binary classification, classification studies between healthy and COVID-19 are common. It is also very important to distinguish other types of lung disease from COVID-19 positive. Therefore, binary-classification (COVID-19 vs Normal, COVID-19 vs Pneumonia) and multi-classification (Normal, COVID-19, Pneumonia) were preferred in this study. An ensemble usually consists of basic learners. In our approach, these key learners were chosen as ResNet50, ResNet152, VGG-19, and DenseNet201.

In this study, a bagging approach is considered to combine different deep learning models with basic fusion methods to increase classification accuracy. Two ensemble models were created to combine different deep learning models; Ensemble-1 (ENS-1) ve Ensemble-2 (ENS-2). Ensemble-1 model consists of ResNet152, VGG-19, DenseNet201 models, and Ensemble-2 model consists of ResNet50, VGG-19, DenseNet201 models. Three different scenarios were applied to these models. First, the models were trained for three classes (COVID-19, Pneumonia, and Normal) to classify X-Ray images

into three categories, as seen in Table 4. Second, the models are trained to classify two classes (COVID-19 and Normal), as seen in Table 5. Third, the models are trained to classify two classes (COVID-19 and Pneumonia), as seen in Table 6. A dataset of 1125 images (125 COVID-19, 500 Pneumonia, and 500 Normal) was used to develop the model. The dataset was separated by 5-fold cross-validation with 80% training and 20% validation (Figure 5). Both the binary and multi-classification performance of the models were evaluated for each fold, and finally, the classification performances of the basic fusion methods were compared.

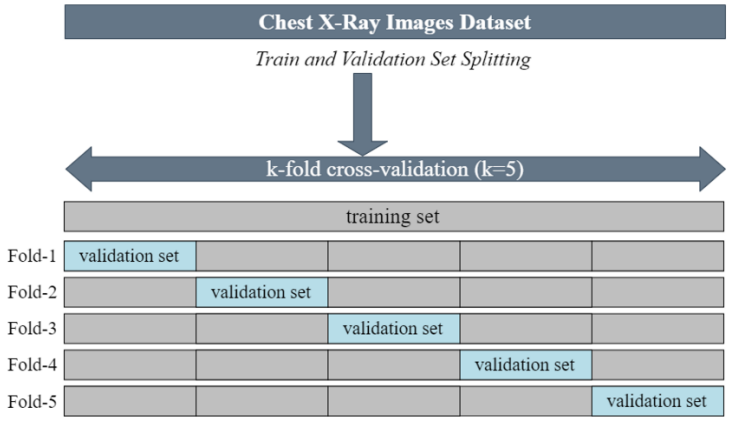


Figure 5. 5-fold cross-validation used to split the dataset into training and validation

The biggest strength of our ensemble deep learning system is that two selected models make their predictions according to fusion methods. Our model also reduces the misclassification error by predicting the new sample with multiple ensemble models.

Among the fusion methods, the Max function achieved the highest classification performance in three classification scenarios for both Ens-1 and Ens-2 (Table 4).

Table 4. Ensemble-1 and Ensemble-2 Models Accuracy for three classifications (COVID-19, Normal, Pneumonia)

	ENS-1					ENS-2				
	MAX	**MODE**	**SUM**	**MEAN**	**PRODUCT**	**MAX**	**MODE**	**SUM**	**MEAN**	**PRODUCT**
Fold-1	0,87	0,86	0,88	0,88	0,88	0,87	0,85	0,88	0,88	0,87
Fold-2	0,79	0,78	0,79	0,79	0,78	0,79	0,76	0,78	0,78	0,79
Fold-3	0,83	0,80	0,80	0,80	0,82	0,85	0,81	0,83	0,83	0,83
Fold-4	0,84	0,82	0,80	0,80	0,80	0,86	0,84	0,85	0,85	0,85
Fold-5	0,90	0,93	0,92	0,92	0,92	0,91	0,93	0,90	0,90	0,88
Average	**0,85**	**0,84**	**0,84**	**0,84**	**0,84**	**0,86**	**0,84**	**0,85**	**0,85**	**0,84**

Figure 6 shows the accuracy rates of the fusion functions for each fold for Ens-1 (left) and Ens-2 (right). While all fusion methods continued close to the same mean in Ens-1, the Max function showed the highest performance. Similarly, while the max function showed the highest performance in Ens-2, the lowest performance was obtained with Mode and Product.

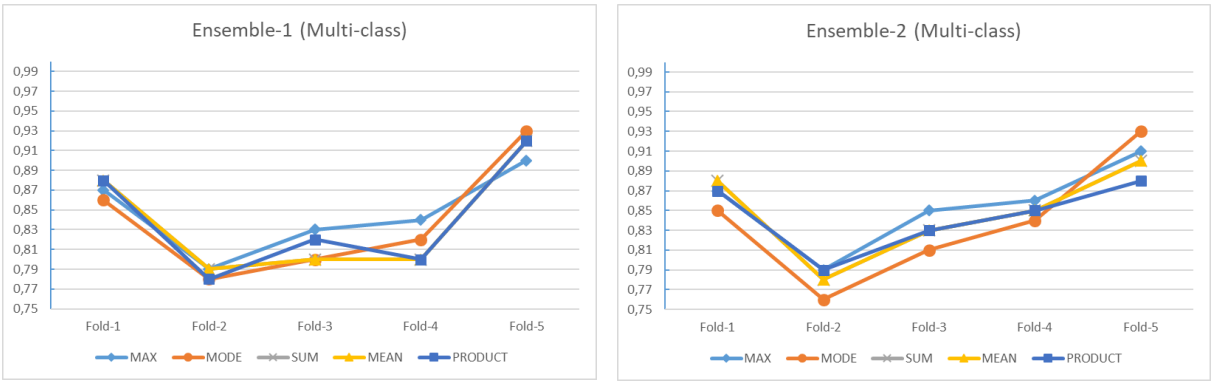


Figure 6. Multi-classification results of basic fusion functions

Binary classification performances are higher for each fusion function than for 3-classification (Table 5). The classification performance for COVID-19 vs Normal is 99% for Ens-1 and Ens-2. Among the fusion methods, the Sum function obtained the lowest value, while the other functions achieved the level of 99% (Table 5).

Table 5. ENSEMBLE-1 and ENSEMBLE-2 Models Accuracy for binary classification (COVID-19 vs Normal)

	ENS-1					**ENS-2**				
	MAX	**MODE**	**SUM**	**MEAN**	**PRODUCT**	**MAX**	**MODE**	**SUM**	**MEAN**	**PRODUCT**
Fold-1	1,00	0,99	0,98	1,00	1,00	1,00	1,00	1,00	1,00	0,99
Fold-2	0,98	0,98	0,98	0,98	0,98	0,99	0,99	0,99	0,99	0,99
Fold-3	0,98	0,99	0,98	0,99	0,99	0,99	0,99	0,99	0,99	1,00
Fold-4	0,99	0,99	0,99	1,00	1,00	0,99	0,99	0,99	0,99	1,00
Fold-5	0,98	0,99	0,98	0,99	0,99	0,98	0,98	0,98	0,98	0,98
Average	**0,99**	**0,99**	**0,98**	**0,99**	**0,99**	**0,99**	**0,99**	**0,99**	**0,99**	**0,99**

The obvious difference in healthy and COVID-19 x-rays (Figure 1) is also reflected in binary classification performances. While the Sum function is at the lowest level for almost all folds in Ens-1, the other functions have exceeded 99% in both Ens-1 and Ens-2 (Figure 7).

Figure 7. Binary classification (COVID-19 and Normal) results of basic fusion functions

Since COVID-19 and Pneumonia are two diseases with similar symptoms, the classification performance here is relatively lower than the COVID-19 and Normal classification. While all fusion methods showed the same performance for Ens-1, 98% success was achieved with the Product function for Ens-2 (Table 6).

Table 6. ENSEMBLE-1 and ENSEMBLE-2 Models Accuracy for binary classification (COVID-19 vs Pneumonia)

	ENS-1					ENS-2				
	MAX	**MODE**	**SUM**	**MEAN**	**PRODUCT**	**MAX**	**MODE**	**SUM**	**MEAN**	**PRODUCT**
Fold-1	0,97	0,97	0,98	0,97	0,97	0,96	0,95	0,96	0,96	0,97
Fold-2	0,93	0,94	0,92	0,93	0,93	0,94	0,95	0,95	0,95	0,95
Fold-3	1,00	0,99	0,99	1,00	1,00	0,99	0,99	1,00	1,00	0,99
Fold-4	0,98	0,98	0,97	0,99	0,99	0,99	0,98	0,99	0,99	0,99
Fold-5	0,96	0,96	0,97	0,96	0,96	0,95	0,97	0,97	0,97	0,98
Average	**0,97**	**0,97**	**0,97**	**0,97**	**0,97**	**0,97**	**0,97**	**0,97**	**0,97**	**0,98**

In the COVID-19 vs Pneumonia binary classification, the Product function for Ens-2 achieved the highest performance with 98%. In Ens-1, 92% and 93% values were obtained for Mean and Sum in only one-fold, respectively. The gap between performances in folds in Ens-1 is higher than in Ens-2 (Figure 8).

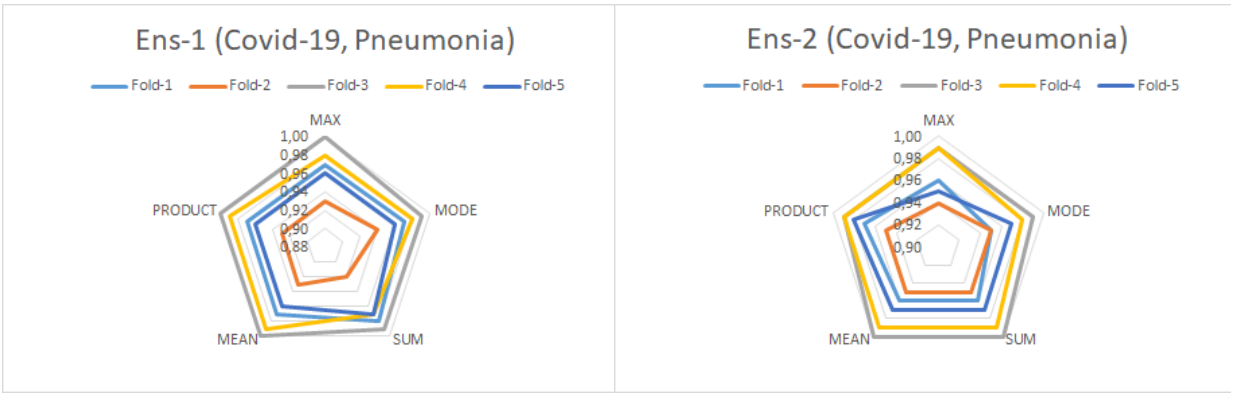


Figure 8. Binary classification (COVID-19 and Pneumonia) results of basic fusion functions

Due to the low rate of COVID-19 X-ray images in the dataset, it was evaluated according to precision, sensitivity, and F1 score as well as accuracy parameters (Table 7, Table 8, Table 9). The values in these tables were obtained according to the Max Voting function. When the values are examined, it is seen that many basic fusion functions have a positive effect on classification performance. For the three-classification, it is seen that the maximum 83% success rate in Table 7 reaches 85% for Max, and 84% for Mode, Sum, Mean, and Product with basic fusion functions. These values reached 86% with the Max function applied in Ens-2. As can be seen from Table 7, the model with the lowest 3-classification performance was DenseNet-201 with 82%.

Table 7. Precision, Recall, F1-Score, and Accuracy of 4 different models per fold for 3-classification (Normal, COVID-19, Pneumonia)

Models	Folds	Fold-1	Fold-2	Fold-3	Fold-4	Fold-5	Average
ResNet-50	Precision	1,00	0,94	1,00	1,00	1,00	0,99
	Recall	0,92	0,68	0,96	0,92	0,96	0,89
	F1-Score	0,96	0,79	0,98	0,96	0,98	0,93
	Accuracy	0,85	0,79	0,80	0,84	0,89	**0,83**
ResNet-152	Precision	1,00	0,94	1,00	0,96	1,00	0,98
	Recall	0,84	0,64	1,00	0,88	0,92	0,86
	F1-Score	0,91	0,76	1,00	0,92	0,96	0,91
	Accuracy	0,87	0,78	0,84	0,77	0,89	**0,83**
VGG-19	Precision	0,95	1,00	1,00	1,00	0,96	0,98
	Recall	0,76	0,64	0,96	0,84	0,88	0,82
	F1-Score	0,84	0,78	0,98	0,91	0,92	0,89
	Accuracy	0,83	0,80	0,76	0,77	0,86	**0,80**
DenseNet-201	Precision	1,00	1,00	1,00	1,00	1,00	1,00
	Recall	0,72	0,64	1,00	0,84	0,84	0,80
	F1-Score	0,84	0,78	1,00	0,91	0,91	0,89
	Accuracy	0,83	0,76	0,82	0,82	0,89	**0,82**

Table 8. Performance parameters of 4 different models on each fold for binary classification (COVID-19 vs Normal)

Models	Folds	Fold-1	Fold-2	Fold-3	Fold-4	Fold-5	Average
ResNet-50	Precision	1,00	1,00	1,00	1,00	1,00	1,00
	Recall	0,96	1,00	0,88	0,96	0,92	0,95
	F1-Score	0,98	1,00	0,94	0,98	0,96	0,97
	Accuracy	0,99	1,00	0,98	0,99	0,98	**0,99**
ResNet-152	Precision	1,00	1,00	1,00	1,00	1,00	1,00
	Recall	1,00	0,88	0,92	0,96	0,92	0,94
	F1-Score	1,00	0,94	0,96	0,98	0,96	0,97
	Accuracy	1,00	0,98	0,95	0,99	0,98	**0,98**
VGG-19	Precision	1,00	1,00	1,00	0,96	0,95	0,98
	Recall	0,96	0,92	1,00	1,00	0,88	0,95
	F1-Score	0,98	0,96	1,00	0,98	0,94	0,97
	Accuracy	0,99	0,98	1,00	0,99	0,98	**0,99**
DenseNet-201	Precision	1,00	1,00	1,00	1,00	1,00	1,00
	Recall	1,00	0,72	0,92	0,96	0,96	0,91
	F1-Score	1,00	0,84	0,96	0,98	0,98	0,95
	Accuracy	1,00	0,95	0,95	0,99	0,99	**0,98**

Binary classification performances of the models remained stable at 98% and 99% levels. The ResNet-152 and DenseNet-201 models showed the lowest performance in both cases (COVID-19 and Normal and COVID-19 and Pneumonia). ResNet-50 and VGG-19 showed the highest performance in both cases (Table 8, Table 9).

Table 9. All performance parameters of 4 different models on each fold for binary classification (COVID-19 vs Pneumonia)

Models	Folds	Fold-1	Fold-2	Fold-3	Fold-4	Fold-5	Average
ResNet-50	Precision	0,92	1,00	0,85	1,00	1,00	1,00
	Recall	0,96	0,52	0,88	0,88	0,96	0,95
	F1-Score	0,94	0,68	0,86	0,94	0,98	0,97
	Accuracy	0,97	0,90	0,94	0,98	0,99	**0,99**
ResNet-152	Precision	0,96	1,00	1,00	1,00	1,00	1,00
	Recall	0,88	0,64	0,96	0,92	0,92	0,94
	F1-Score	0,92	0,78	0,98	0,96	0,96	0,97
	Accuracy	0,97	0,93	0,99	0,98	0,98	**0,98**
VGG-19	Precision	0,94	0,93	0,95	0,95	0,90	0,98
	Recall	0,68	0,52	0,84	0,84	0,76	0,95
	F1-Score	0,79	0,67	0,89	0,89	0,83	0,97
	Accuracy	0,93	0,90	0,96	0,96	0,94	**0,99**
DenseNet-201	Precision	1,00	0,74	1,00	1,00	1,00	1,00
	Recall	0,80	0,56	0,96	0,96	0,76	0,91
	F1-Score	0,89	0,64	0,98	0,98	0,86	0,95
	Accuracy	0,96	0,87	0,99	0,99	0,95	**0,98**

As can be seen in Table 10, it is seen that the best results are generally obtained in binary classification studies. Considering that we obtain the accuracy rates by combining only two models, it is seen that this proposed system will find a remarkable use in other big datasets, as well as the use of different fusion functions.

Table 10. Comparison of accuracy rates of COVID-19 classification methods using chest X-ray images

Previous study	Data type	Methods/classifier	# classes	Accuracy (%)
(S, Wang et al,, 2020)	Chest CT	M-Inception	2	82,90
(Loey et al,, 2020)	Chest CT	AlexNet, VGG16, VGG19, GoogleNet, and ResNet-50	2	82,91
(Sahinbas and Catak, 2020)	Chest X-ray	VGG16, VGG19, ResNet, DenseNet ve InceptionV3	2	80,00
(Xu et al,, 2020)	Chest X-ray	ResNet+Location Attention	3	86,70
(Ouchicha et al,, 2020)	Chest X-ray	CVDNet	3	97,20
(Ozturk et al,, 2020)	Chest X-ray	DarkCovidNet	2	98,08
			3	87,02
This Study	Chest X-ray	ENS-1	3	**85,00**
		ENS-2	3	**86,00**
		ENS-1 (COVID-19andPneumonia)	2	**97,00**
		ENS-2 (COVID-19andPneumonia)	2	**98,00**
		ENS-1 (COVID-19andNormal)	2	**99,00**
		ENS-2 (COVID-19andNormal)	2	**99,00**

4. Conclusion

Early diagnosis is vital in preventing the transmission of disease to other people. This study proposes a transfer learning-based ensemble deep learning system to automatically detect and classify COVID-19 patients from 1125 X-ray images obtained from patients with COVID-19, Pneumonia, and Normal. Examining the effect of fusion functions on classification performance in these ensemble systems is important. Fusion functions have been studied before and have been shown to work well in many areas. Therefore, it is important to apply these functions to COVID-19 datasets. In this study, only five of the basic fusion functions were used. Our system used deep CNN-based ResNet-50, ResNet-152, VGG-19, and DenseNet201 models for multi and binary-class problems. In the COVID-19, Normal, and Pneumonia classification, the Ensemble-1 model achieved 85% performance with the Max function. The Ensemble-2 model, on the other hand, achieved 86% performance with the Max function. In binary classification, almost all functions showed a 99% performance. The general performance accuracy achieved by this system is 86% for the three classes and 99% for the binary class. This system can be used to detect COVID-19 disease without the need for radiologists. In addition, the proposed system can be used to classify other diseases related to the chest, such as Pneumonia. In future studies, methods such as Dempster-Shafer can be used instead of basic fusion functions.

Statement of Conflict of Interest

The authors declare that he has no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Author's Contributions

The contribution of the authors is equal.

References

Abraham B., Nair MS. Computer-aided detection of COVID-19 from X-ray images using multi-CNN and Bayesnet classifier. Biocybernetics and Biomedical Engineering 2020; 40(6): 1436-1445.

Alimadadi A., Aryal S., Manandhar I., Munroe PB., Joe B., Cheng X. Artificial intelligence and machine learning to fight COVID-19. Physiol Genomics 2020; 52(4): 200–202.

Ardakani AA., Kanafi AR., Acharya UR., Khadem N., Mohammadi A. 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 2020; 121(103795): 1-12.

Bozkurt F. Derin öğrenme tekniklerini kullanarak akciğer x-ray görüntülerinden COVID-19 tespiti. European Journal of Science and Technology 2021; 24: 149-156.

Cohen JP., Morrison P., Dao L. COVID-19 image data collection. Computers and Education 2020; 164(11597): 1-11.

Dey N., Zhang YD., Rajinikanth V., Pugalenthi R., Raja NSM. Customized VGG19 architecture for pneumonia detection in chest x-rays. Pattern Recognition Letters 2021; 143(1): 67–74.

Hassantabar S., Ahmadi M., Sharif A. Diagnosis and detection of infected tissue of COVID-19 patients based on lung x-ray image using convolutional neural network approaches. Chaos, Solitons and Fractals 2020; 140(110170): 1-11.

Huang G., Liu Z., van der Maaten L., Weinberger KQ. Densenet: densely connected convolutional networks. IEEE Computer Society Conference on Computer Vision and Pattern Recognition 2017; 30(1): 82–84.

Ismael AM., Şengür A. Deep learning approaches for COVID-19 detection based on chest X-ray images. Expert Systems with Applications 2021; 164(114054): 1-11.

Kırbas I., Sözen A., Tuncer AD., Kazancıoglu FS. Comparative analysis and forecasting of COVID-19 cases in various European countries with ARIMA, NARNN and LSTM approach. American Physiological Society Bethesda 2020; 138(110015): 1-11.

Korkmaz A., Atila BÜ. Derin öğrenme teknikleriyle akciğer röntgeninden Covid-19 tespiti. Artificial Intelligence Studies, 2018; 1: 1–13.

Li X., Geng M., Peng Y., Meng L., Lu S. Molecular immune pathogenesis and diagnosis of COVID-19. Journal of Pharmaceutical Analysis 2020; 10(1): 102-108.

Loey M., Manogaran G., Khalifa NEM. A deep transfer learning model with classical data augmentation and CGAN to detect COVID-19 from chest CT radiography digital images. Neural Computing and Applications 2020; 1(26): 1-13.

Narin A., Kaya C., Pamuk Z. Automatic detection of coronavirus disease (COVID-19) using X-ray images and deep convolutional neural networks. Pattern Anal Applic 2021; 24(1): 1207–1220.

Ouchicha C., Ammor O., Meknassi M. CVDNet: A novel deep learning architecture for detection of coronavirus (Covid-19) from chest x-ray images. Chaos, Solitons and Fractals 2020; 140(110245): 1-11.

Ozturk T., Talo M., Yildirim EA., Baloglu UB., Yildirim O., Acharya UR. Automated detection of COVID-19 cases using deep neural networks with x-ray images. Computers in Biology and Medicine 2020; 121(103792): 1-11.

Özbay E., Altunbey Özbay F. Derin öğrenme ve sınıflandırma yaklaşımları ile BT görüntülerinden Covid-19 tespiti. DÜMF Mühendislik Dergisi 2021; 12(2): 211-219.

Sahinbas K., Catak FO. Transfer learning-based convolutional neural network for COVID-19 detection with X-ray images (pp 451-466). Academic Press; 2021.

Shuja J., Alanazi E., Alasmary W., Alashaikh A. COVID-19 open source data sets: a comprehensive survey. Appl Intell 2021; 51: 1296-1325.

Simonyan K., Zisserman A. Very deep convolutional networks for large-scale image recognition. American Physiological Society Bethesda 2014; 1(6): 1-14.

Singhal TA. Review of coronavirus disease-2019 (COVID-19). The Indian Journal of Pediatrics 2020; 87(1): 281–286.

Tuncer T., Dogan S., Ozyurt F. An automated residual exemplar local binary pattern and iterative ReliefF based COVID-19 detection method using chest X-ray image. Chemometrics and Intelligent Laboratory Systems 2020; 203(104054): 1-11.

Wang S., Kang B., Ma J., Zeng X., Xiao M., Guo J., Cai M., Yang J., Li Y., Meng X., Xu Bo. A deep learning algorithm using CT images to screen for Coronavirus disease (COVID-19). European Radiology 2021; 31(8): 6096-6104.

Wang X., Peng Y., Lu L., Lu Z., Bagheri M., Summers RM. ChestX-ray8: Hospital-scale Chest X-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition 2017; 2097–2106.

Xu X., Jiang X., Ma C., Du P., Li X., Lv S., Yu L., Ni Q., Chen Y., Su J., Lang G., Li Y., Zhao H., Liu J., Xu K., Ruan L., Sheng J., Qiu Y., Wu W., Li L. A Deep Learning system to screen novel coronavirus disease 2019. Pneumonia. Engineering 2020; 6(10): 1122–1129.

Wang X., Peng Y., Lu L., Lu Z., Bagheri M., Summers RM. ChestX-ray8: Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition 2017; 2097–2106.