AUT Journal of Electrical Engineering

AUT J. Elec. Eng., 53(2) (2021) 223-232
DOI: 10.22060/eej.2021.19467.5397

# A Hybrid Deep Transfer Learning-based Approach for COVID-19 Classification in Chest X-ray Images

K. Rezaee[1*], A. Badiei[2], H. Ghayoumi Zadeh[3], S. Meshgini[4]

[1] Department of Biomedical Engineering, Meybod University, Meybod, Iran
[2]Department of Biomedical Engineering, Faculty of Electrical and Computer Engineering, Tabriz University, Tabriz, Iran
[3]Department of Electrical Engineering, Faculty of Engineering, Vali-e-Asr University of Rafsanjan, Rafsanjan, Iran
[4]Department of Biomedical Engineering, Faculty of Electrical and Computer Engineering, Tabriz University, Tabriz, Iran

**ABSTRACT:**The COVID-19 pandemic is a severe public health hazard. Hence, proper and early diagnosis is necessary to control the infection progression. We can diagnose this disease by employing a chest X-ray (CXR) screening, which is ordinarily cheaper and less harmful than a Computed Tomography scan (CT scan) and is continuously accessible in small or rustic hospitals. Since the COVID-19 dataset is inadequate and cannot be strictly distinguished from CXR, Deep Transfer Learning (DTL) models can be used to diagnose coronavirus even with access to a small number of images. In this paper, we presented an approach to diagnosis COVID-19 using CXR images based on the concatenated features vector of the three DTL structures and soft-voting feature selection procedure, including Receiver of Curve (ROC), Entropy, and signal-to-noise ratio (SNR) techniques. Our hybrid model reduces the feature vector size and classifies it in optimize manner to improve the decision-making process. A collection of 2,863 CXR images comprising normal, bacterial, viral, and COVID-19 cases were prepared in JPEG format from the Medical Imaging Center of Vasei Hospital, Sabzevar, Iran. The proposed approach obtained an Accuracy of 99.34%, Sensitivity of 99.48%, Specificity of 99.27% while having a far fewer number of trainable parameters in contrast to its counterparts. Compared to the latest similar methods, the diagnosis accuracy has increased from 1.5 to 2.2%. The comparative experiment reveals the advantage of the suggested COVID-19 classification pattern based on DTL over other competing schemes.

**Review History:**

Received: Jan. 0, 2021
Revised: Feb. 11, 2021
Accepted: Feb. 21, 2021
Available Online: Sep. 01, 202

**Keywords:**

COVID-19
Chest X-ray
Deep transfer learning
Convolutional neural network
Feature selection.

## 1- INTRODUCTION

COVID-19 is a worldwide disease and epidemic infectious sickness [1,2] that has been widespread in almost every country since its first origin. Nearly 108.1 M approved cases and the deaths of more than 2.36 M people, in mid-February 2021 was confirmed by WHO [3]. COVID-19 is a pandemic originated by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) [4]. The exponential growth of COVID-19 not only endangers life, but also further affects financial business and disturbs world trip. The virus originated from pneumonitis [5], an infection caused to inflame alveoli [5]. One method of diagnosing lung swelling is to engage in CXR radiography. The artificial intelligence (AI) and ML procedures assist physicians in recognizing pneumonia fast and precisely. Furthermore, CT scans are usually more expensive and more deleterious than X-rays and are not regularly accessible in small or rustic hospitals.

The radiologists favor chest X-Ray images, since most hospitals acquire X-ray devices [6]. Moreover, considering its low cost, CXR is a preferable procedure to diagnose COVID-19 in some undeveloped countries. Although sometimes the X-ray radiography devices cannot identify muscle impairment, soft-tissue lesions, or the defects of other parts of the body, a CT scan can overcome these challenges. Accordingly, ML-based approaches can efficiently aid in analyzing images to minimize the mistake in distinguishing infectious diseases.

Lately, processing procedures have been broadly studied to improve the processing of medical images [7]. Conventional techniques implement ML algorithms on CXR images to detect COVID-19 patients. However, they do not focus on detecting false-negatives that cause further spread of COVID-19 [8]. Usually, DL methods are famous for screening tuberculosis in CXR. DTL improves such medical methods to produce more reliable outcomes, extend diagnosing scope, and perform real-time employed medical and diagnostic schemes [9-11]. These techniques can modify and train the weight of networks on a big database, as well as fine-tuning the weight of pre-trained networks on a dataset with inadequate images. Up to now, many DTL models have been proposed to classify various patterns, (e. g., AlexNet [12], VGGNet [13], GoogleNet [14], ResNet [15], Xception [16], DenseNet [17] and Inception-V3 [18]).

Various novel investigations have considered medical images of COVID-19 utilizing several types of DL-based methods. Wang et al., [19] introduced an open-source DL pattern and fulfilled an extensive standard dataset called

*Corresponding author's email: kh.rezaee@meybod.ac.ir

COVIDx with 13,975 X-ray radiographs of the patient's chest to detect COVID-19. Their research had an efficiency of 93.3% over the COVIDx dataset. Messina and Apostolopoulos [20] implemented transfer learning (TL), preparing CNNs on a small-size medical image dataset. Their dataset contains 1427 X-ray images, including 224 COVID-19 confirmed images. Their evaluation metrics including accuracy, sensitivity, and specificity have been calculated at 96.78%, 98.6%, and 96.46%, respectively. Narin et al., [9] applied three CNN models, particularly InceptionResNet V2, InceptionV3, and ResNet50, for automatic determination of COVID-19 to interpret CXR images. The outcomes showed the highest accuracy of 98% relating to five-fold cross-validation. Abbas et al., [21] introduced the De TraC CNN in-depth model, which proposes a method based on the transfer of knowledge from object detection. They used the mentioned transfer of knowledge to recognize COVID 19 CXR images, and their plan illustrated an accuracy of 95.55% (precision of 93.36% and specificity of 91.87%). In [22], a COVIDX-Net comprises several CNN in-depth structures proposed that analyzed 50 chest radiographs with 25 cases of COVID-19. Their outcomes revealed an F-score of 91% and an accuracy of 90%. Zhang et al., [23] introduced the construction of a DL-based anomaly detection pattern that is utilized in the CXR image dataset, and the sensitivity and specificity were 96% and 70.65%.

As stated in prior investigations, CT scans of the chest perform a low false-positive rate compared to other imaging machines, such as X-rays. Therefore, some studies have only analyzed CT images to diagnose COVID-19. In [25], a DL method has been offered for automatically analyzing infected pulmonary areas. They assessed their approach to 300 people with coronavirus, leading accuracy of 91%. Extended systems cannot ascertain the severity of other cases of pneumonia. Gozes et al., [26] formed a DL-based approach to recognize and quantify COVID-19 severity from chest CT scan images. They evaluated more than 110 infected individuals, and the classification accuracy obtained by this process was 94.80%. Wang et al., [27] employed Inception Net to identify COVID-19-related abnormalities in CT scan images of the lungs. They examined the InspectionNet architecture on 1065 CT scan images and distinguished 325 infected individuals with 85.20% accuracy. Xu et al. [28] applied 3D CNN architecture to detect infections affected by coronaviruses from influenza-A (H1N1) viral pneumonia on CT scans. The accuracy of the CNN structure has been reported as 86.70%. Chen et al., [29] handled UNet++ construction to identify coronavirus pneumonia. Following training their architecture with 106 patients, they achieved a classification accuracy of 98.85%. The time demanded analysis was decreased by up to 65%. Das et al. [30], suggested a deep learning-based Convolutional Neural Network (CNN) architecture, which utilizes the Truncated Inception Net to screen COVID-19 positive CXRs from other non-COVID and/or healthy cases. They strived to extract chest CXR images' features to detect coronavirus infection by improving the COVID-19 diagnose neural network (COVNet).

Zheng et al. [31], presented a supervised DL-based approach to detect COVID-19 patients by employing a 3D CT scan. They formed a UNET-trained structure to segment 3D images of the lungs. The segmented areas were implemented to the DL plan to recognize infected regions. The accuracy achieved from their model was 95.9%. Minaee et al. [32], exhibited a DL structure as per 5000 images called COVID-Xray-5k in which it adopted SqueezeNet, ResNet18, ResNet50, Densenet-121, and illustrated an average sensitivity of 97.5% and a specificity of 90%. Concerning many new DLs, models have been formed for the detection of the COVID-19. The study [33] intended to consider the accurate tuning of pre-trained CNNs for COVID-19 classification using CXR images. Recently, methods based on combining features extracted from DTL-based networks have been proposed, each of which has reported satisfactory results from CXR image classification for Covid-19 diagnosis [34-36].

We need the application of patterns that are automatically inferred from the CXR images and various infections similar to COVID-19. It can be understood from this respect that different models can be appropriated beside the designs, such as different classifiers as an integrated representation in recognizing abnormal pneumonia. One of the proper methods that have lately matured as the essential model in classification is deep learning-based methods. In this investigation, hybrid deep learning procedures are suggested for detecting COVID-19 based on chest X-ray radiographs, unlike the several intended approaches in the literature. Accordingly, our study's major contribution is presenting a hybrid utilization of a pre-train deep learning framework to extract the proper features from CXR images to classify various types of pneumonia and COVID-19. Therefore, we introduce a COVID-19 detection algorithm that does not need an identical decision layer process (i.e. used in advanced architectures). We also have solved challenges such as vanishing gradient and over-fitting problems, using a new methodology based on pre-train structures, hybrid feature selection, and optimized classifier. This methodology utilizes pre-trained CNN structures (i. e., AlexNet [12], Vgg-f [37], and CaffeNet [38]) to extract features from CXR images without further preprocessing. This is followed by adopting the properties acquired for an optimized SVM-based classification. We applied a CXR database from Vasei Hospital in Sabzevar to evaluate the achievement of the proposed model.

The remainder of the research is prepared as follows: Section 2 presents the material and method. This section explains the chest X-ray datasets employed in experiments and the evaluation of the scheme. Section 3 reflects the experimental results and outcomes of the proposed model. Subsequently, Section 4 consummate the paper and donates aspects of future studies.

## 2- MATERIALS AND METHODS

We stated that the main contributions of COVID-19 diagnosis have been throughout the analysis of CXR images,

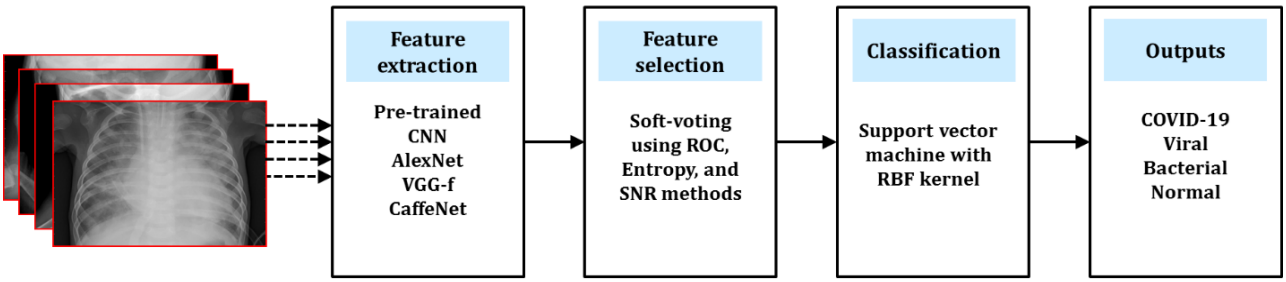


**Fig. 1. A schematic of the proposed method for identifying COVID-19 and similar infectious diseases.**

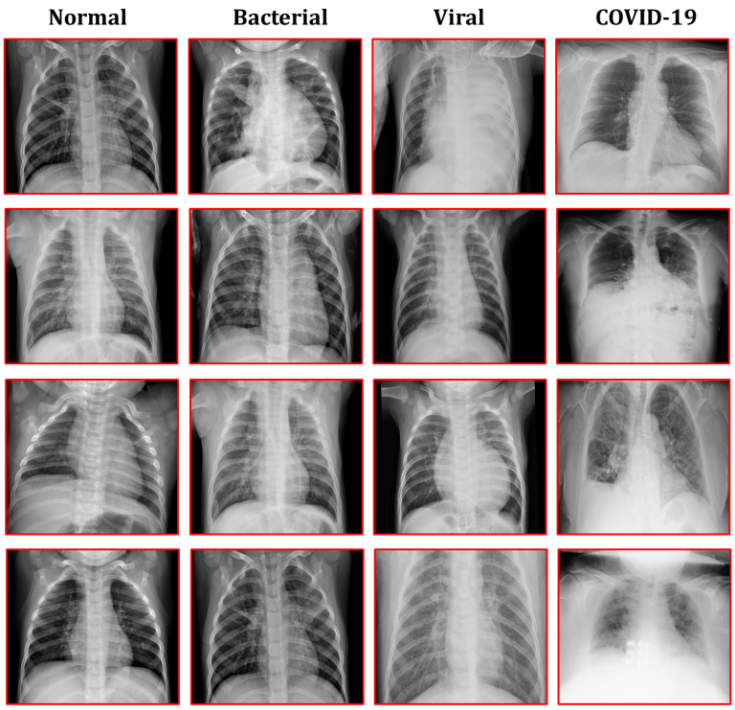


**Fig. 2. A sample of CXR images collected from four different classes.**

concatenated feature extraction via three pre-train structures (i.e. non-handcrafted features), feature selection, and classification plan. The suggested approach tries to detect COVID-19 utilizing CXR images. According to Fig. 1, this model employs CXR images without further pre-processing or segmentation as input, marking the significant distinction between the offered procedure and similar approaches. The input radiographs are represented relating CNNs, followed by feature selection. The classification step comprises image classification as validated COVID-19 cases applying the SVM with Gaussian Radial Basis Function (RBF) kernel.

## 2-1- CXR Images

Our implemented pre-train learning models were prepared in MATLAB 2019b environment. Several deep learning techniques and classifiers were employed to analyze the CXR image for healthy, pneumonia patient, and COVID-19. We used a single CPU to process the information.

Pattern recognition and deep learning toolboxes were adopted to carry out the architectures and classifier implementation. A sum of 2,863 chest X-ray images in JPEG format was gathered, which contained two major categories: healthy and pneumonia. Two physicians have clarified the type of induced infectious diseases from CXR images in the Medical Imaging Center of Vasei Hospital, Sabzevar, Iran. All chest X-ray images in the dataset were considered to remove low-quality scans. They were also classified by a physician and a third-party specialist to prevent any classification errors.

In this manner, radiographs were administered as a part of clinical analysis. Only 676 CXR images (331 normal; 183 COVID-19; 83 bacterial, and 79 viral images) were chosen from the received dataset to show the effectiveness, which could generate acceptable results with an inadequate number of images depicting about 23.61% of the data available. In Fig. 2, the CXR images are represented in four separate classes.

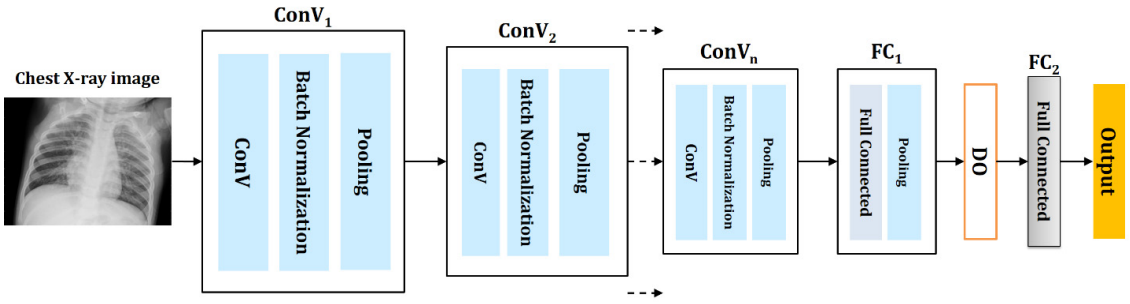


**Fig. 3. An illustration of the pre-train structures.**

## 2-2- Pre-train Models

Fig. 3 depicts the overall structure of the DTL. Pre-trained CNNs in a great natural database use 1000-category images, referred to as ImageNet [12,39]. Among them, this database comprises medical images as well. While improving the suggested approach, three models were preferred and concatenated: Vgg-F, AlexNet, and CaffeNet. They are similar in structure, with the main variation being the number of neurons in the fully-connected (FC) layers and the size of the filters used in convolutional layers (CLs). The principal variation between this structure and the AlexNet pattern is the order of the normalization and pooling layers. AlexNet was designed to train and classify the ImageNet database for the ILSVRC-2010 architectural purpose. The AlexNet architecture incorporates eight layers requiring to be training, five CLs with 5×5 and 7×7 filters, following by three FC layers as well as max-pooling layers (MPLs). The CaffeNet structure is one of the most popular CNNs in DL. It comprises five circular layers, each of which is joined by a single layer and three FC layers. The VGG-F structure is an 8-layer deep convolutional neural network (DCNN). This structure is analogous to the one utilized by Krizhevsky et al. [12]. In VGG-f, the radiography image size should be applied as 224×224×3 sizes. Quick processing is assured by the four-pixel stride in the first convolutional layer. The structure has been trained on ILSVRC data applying gradient descent (GD) with momentum. The VGG-f additionally is similar to AlexNet and presented with two other items: Vgg-s and Vgg-m.

The main variation between the three models is the size of convolutional filters and the number of layers. The filter size influences enhancement of the computational complexity of the network. By reducing the size of the filter, a less number of processors will be recquired. Notwithstanding, larger-size filters utilize more neighborhood data. The main variation between Vgg-f and AlexNet prevails that Vgg-f has a lower number of filters in the 1st, 3rd, and 4th CLs. Approximately 4096 features are extracted from the last FC layer of any structure.

## 2-3- Feature Selection

The effectiveness of the suggested soft-voting for subset selection has already been investigated [40]. Feature selection is conducted by soft-voting three schemes employing the signal-to-noise ratio, Entropy, and Receiver operating characteristic (ROC). A ROC plot is a graphical plotting of the fraction of true positives (TP) against the fraction of false positives (FP) for a paired classification scheme that the discrimination threshold is different. The area under a ROC curve, AUC, is a global estimation of the discrimination representation in a pattern. The AUC can be computed according as (1) [41]:

$$AUC = \int_0^1 ROC(t)dt \tag{1}$$

If the distribution functions of X in the two statistical samples are $F_1(x)$ and $F_2(x)$, the function T can be brought up a complement of $F_i(x)$, and as an outcome, we express ROC as (2) [41]:

$$ROC(t) = T_1(T_2^{-1}(t)),\ t \in (0,1) \tag{2}$$

A large amount of AUC shows a high overlap between the samples' labels in chest X-rays, and thus, the features with a larger AUC are included in the feature selection vector.

The Entropy measure is represented as (3), also known as the Kullback–Liebler (K-L) divergence, and assumes that the considered test classes are distributed normally. This equation, also called the entropy score, is provided as follows [42]:

$$Entropy = \frac{1}{2}\left[ (\frac{\sigma_1^2}{\sigma_2^2} + \frac{\sigma_2^2}{\sigma_1^2} - 2) + (\frac{1}{\sigma_2^2} + \frac{1}{\sigma_2^2})(\mu_1 - \mu_2)^2 \right] \tag{3}$$

Where, $\mu_1$, $\mu_2$, as well as $\sigma_1$ and $\sigma_2$, are the mean of the samples and their standard deviation. Selected features are ordered according to the highest entropy score.

The third approach is the SNR method, which is utilized to measure the distinction between feature labels that are shown as (4) [43]:

$$SNR(f_i, c) = \frac{\mu_1 - \mu_2}{\sigma_1 + \sigma_2} \tag{4}$$

The variables c and $f_i$ are supposed to be the vectors of the chest X-ray labels and the feature vector of the $i_{th}$. Fig. 4 presents the trend of feature selection based on the soft-voting method.

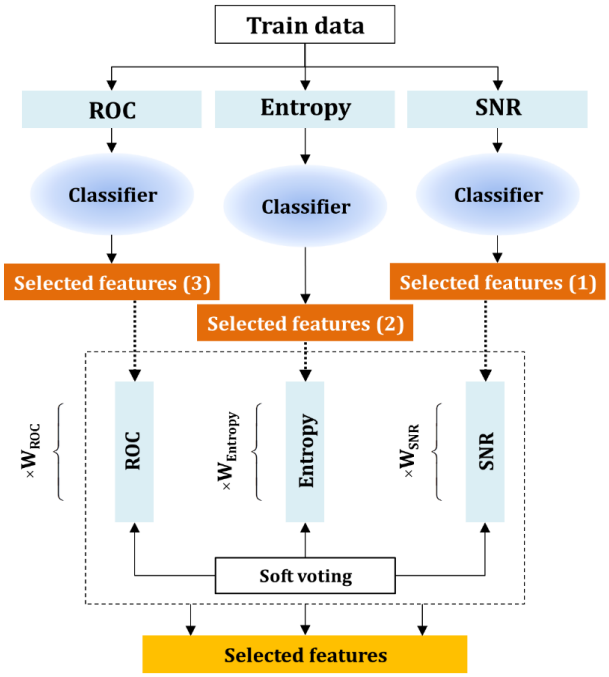


**Fig. 4. Soft-voting feature selection**

### 2-4- Classification

The subset selection of concatenated features is applied to SVM classifier through different cross-validation (CV). We choose the Gaussian RBF kernel and apply the grid search function to determine the optimal parameters C and gamma. Therefore, the over-fitting problem will be solved, and the classification accuracy will increase. The SVM with the final aim of finding a hyper-plane as a decision hyper-plan in such a way that maximizes the margin between the two or four classes in the COVID-19 classification design. The parameters of the Gaussian RBF kernel is expressed as (5) [44]:

$$K(x,x_i) = \exp(-\frac{\gamma\|x-x_i\|^2}{2\sigma^2}) \quad (5)$$

The chief reason for applying the Gaussian RBF kernel function is that this kernel has fewer numerical complexities, possesses less hyper-parameter than other kernels, and a low number of regulatory variables (C and gamma) compared to other kernels [45].

## 3- EXPERIMENTAL RESULTS

### 3-1- Setting

The concatenated features are constructed by implementing a deep learning model including CNN, DNN, and pre-trained models (i.e. Vgg-f, CaffeNet, and Alexnet). We employed other CXR images for training, validation, and testing steps. We apply the deep learning package and MATLAB software as a library. The models were configured with an initial learning rate (μ) of 0.001 and epochs of 50-2000. Furthermore, the mini-batch size was set to 32 and early-stopping to 5 epochs if the accuracy was not improved. This optimization method fuses Root Mean Square Propagation (RMSprop) and Stochastic Gradient Descent (SGD) with momentum. Additionally, to eschew the over-fitting problem in the deep learning network, this problem was appropriated using the dropout process along the early-stopping procedure to determine the best training iteration.

### 3-2- Assessments

We utilize the confusion matrix that is a table ordinarily handled to represent the performance of the classification method. The accuracy, precision, recall, F-measure, and AUC were determined from this table. To validate the outcomes, a k-fold CV was selected (k=5.) This value expressed a broader set of images to be tested, which considered 20% of images related to tests, 20% for validation, and 60% for training. In the SVM classifier, the RBF kernel was used with standard parameters.

We conducted two experiments in our study: (1) binary classification and (2) multi-class classification. In the first experiment, the overall CXR images are divided into two categories: COVID-19 and Healthy (i.e. normal CXR image). The multi-class category incorporates Healthy, Bacterial, Viral, and COVID-19 classes. The highest, lowest, and mean accuracies of binary and multi classifications are illustrated in Tables 1 and 2 based on various pre-train designs in two scenarios (i.e. low and high sizes of the extracted vector). The experiments were replicated five times with a 5-fold CV for concatenated of Vgg-f, CaffeNet, and Alexnet architectures. The selected vectors with a soft ensembling feature selection

**Table 1. Experimental outcomes of binary and multi-class classifications in scenario 1 (i. e., the high size of extracted features).**

| Data Divi. | Pre-train model | Binary classification (Scenario 1) | | | Multi classification (Scenario 1) | | |
|---|---|---|---|---|---|---|---|
| | | Best | Mean | Worst | Best | Mean | Worst |
| 5-fold (1) | AlexNet | 0.99 ± (0.008) | 0.98 ± (0.018) | 0.97 ± (0.028) | 0.94 ± (0.043) | 0.92 ± (0.057) | 0.91 ± (0.076) |
| | CaffeNet | 0.99 ± (0.007) | 0.98 ± (0.014) | 0.97 ± (0.021) | 0.94 ± (0.047) | 0.93 ± (0.063) | 0.92 ± (0.068) |
| | Vgg-f | 0.99 ± (0.010) | 0.98 ± (0.010) | 0.98 ± (0.018) | 0.94 ± (0.039) | 0.93 ± (0.055) | 0.92 ± (0.073) |
| | Concatenate | **0.99 ± (0.002)** ↑ | **0.99 ± (0.005)** ↑ | **0.98 ± (0.010)** ↑ | **0.95 ± (0.028)** ↑ | **0.93 ± (0.041)** ↑ | **0.92 ± (0.054)** ↑ |
| 5-fold (2) | AlexNet | 0.99 ± (0.009) | 0.98 ± (0.016) | 0.98 ± (0.019) | 0.95 ± (0.032) | 0.93 ± (0.062) | 0.92 ± (0.058) |
| | CaffeNet | 0.99 ± (0.008) | 0.98 ± (0.016) | 0.97 ± (0.023) | 0.93 ± (0.046) | 0.92 ± (0.047) | 0.92 ± (0.054) |
| | Vgg-f | 0.99 ± (0.008) | 0.98 ± (0.018) | **0.98 ± (0.017)** | 0.94 ± (0.038) | 0.93 ± (0.051) | 0.91 ± (0.063) |
| | Concatenate | **0.99 ± (0.001)** ↑ | **0.99 ± (0.006)** ↑ | 0.98 ± (0.018) ↓ | **0.95 ± (0.030)** ↑ | **0.93 ± (0.042)** ↑ | **0.92 ± (0.047)** ↑ |
| 5-fold (3) | AlexNet | 0.99 ± (0.009) | 0.98 ± (0.018) | 0.98 ± (0.019) | 0.94 ± (0.052) | 0.91 ± (0.068) | 0.90 ± (0.073) |
| | CaffeNet | 0.99 ± (0.006) | 0.98 ± (0.015) | 0.97 ± (0.027) | 0.95 ± (0.039) | 0.92 ± (0.056) | 0.91 ± (0.065) |
| | Vgg-f | 0.99 ± (0.007) | 0.98 ± (0.016) | 0.97 ± (0.024) | 0.95 ± (0.038) | 0.92 ± (0.051) | **0.91 ± (0.068)** |
| | Concatenate | **0.99 ± (0.002)** ↑ | **0.99 ± (0.006)** ↑ | **0.98 ± (0.013)** ↑ | **0.96 ± (0.024)** ↑ | **0.92 ± (0.038)** ↑ | 0.91 ± (0.070) ↓ |
| 5-fold (4) | AlexNet | 0.99 ± (0.007) | 0.98 ± (0.019) | 0.97 ± (0.024) | 0.94 ± (0.040) | **0.93 ± (0.041)** | 0.92 ± (0.064) |
| | CaffeNet | 0.99 ± (0.007) | 0.98 ± (0.017) | 0.98 ± (0.019) | 0.94 ± (0.051) | 0.93 ± (0.067) | 0.92 ± (0.058) |
| | Vgg-f | 0.99 ± (0.008) | 0.98 ± (0.016) | 0.97 ± (0.019) | 0.93 ± (0.056) | 0.92 ± (0.038) | 0.92 ± (0.066) |
| | Concatenate | **0.99 ± (0.001)** ↑ | **0.99 ± (0.006)** ↑ | **0.98 ± (0.012)** ↑ | **0.94 ± (0.032)** ↑ | 0.92 ± (0.041) ↓ | **0.92 ± (0.051)** ↑ |
| 5-fold (5) | AlexNet | 0.99 ± (0.008) | 0.98 ± (0.016) | 0.97 ± (0.025) | 0.94 ± (0.053) | 0.92 ± (0.048) | 0.91 ± (0.054) |
| | CaffeNet | 0.99 ± (0.010) | 0.98 ± (0.014) | 0.97 ± (0.025) | 0.94 ± (0.048) | **0.93 ± (0.041)** | 0.90 ± (0.081) |
| | Vgg-f | 0.99 ± (0.008) | 0.98 ± (0.011) | 0.97 ± (0.019) | 0.95 ± (0.037) | 0.93 ± (0.053) | **0.92 ± (0.066)** |
| | Concatenate | **0.99 ± (0.001)** ↑ | **0.99 ± (0.005)** ↑ | **0.98 ± (0.015)** ↑ | **0.95 ± (0.030)** ↑ | 0.93 ± (0.048) ↓ | 0.92 ± (0.072) ↓ |

**Table 2. The experimental outcomes of binary and multi-class classifications in scenario 2 (i. e., the low size of extracted features).**

| Data Divi. | Pre-train model | Binary classification (Scenario 2) | | | Multi classification (Scenario 2) | | |
|---|---|---|---|---|---|---|---|
| | | Best | Mean | Worst | Best | Mean | Worst |
| 5-fold (1) | AlexNet | 0.99 ± (0.007) | 0.98 ± (0.016) | 0.97 ± (0.022) | 0.95 ± (0.041) | 0.93 ± (0.051) | 0.90 ± (0.082) |
| | CaffeNet | 0.99 ± (0.009) | 0.98 ± (0.018) | 0.97 ± (0.022) | 0.95 ± (0.037) | 0.94 ± (0.054) | 0.91 ± (0.076) |
| | Vgg-f | 0.99 ± (0.007) | 0.98 ± (0.011) | 0.98 ± (0.021) | 0.95 ± (0.036) | 0.93 ± (0.052) | **0.91 ± (0.071)** |
| | Concatenate | **0.99 ± (0.004)** ↑ | **0.99 ± (0.003)** ↑ | **0.98 ± (0.013)** ↑ | **0.95 ± (0.023)** ↑ | **0.94 ± (0.047)** ↑ | 0.91 ± (0.078) ↓ |
| 5-fold (2) | AlexNet | 0.99 ± (0.008) | 0.98 ± (0.011) | 0.98 ± (0.021) | 0.95 ± (0.042) | 0.94 ± (0.052) | **0.91 ± (0.073)** |
| | CaffeNet | 0.99 ± (0.007) | 0.98 ± (0.018) | 0.97 ± (0.024) | 0.95 ± (0.039) | 0.93 ± (0.057) | 0.91 ± (0.078) |
| | Vgg-f | 0.99 ± (0.009) | 0.98 ± (0.014) | 0.98 ± (0.026) | 0.95 ± (0.031) | 0.93 ± (0.061) | 0.91 ± (0.082) |
| | Concatenate | **0.99 ± (0.005)** ↑ | **0.99 ± (0.004)** ↑ | **0.98 ± (0.017)** ↑ | **0.96 ± (0.026)** ↑ | **0.94 ± (0.044)** ↑ | 0.91 ± (0.081) ↓ |
| 5-fold (3) | AlexNet | 0.99 ± (0.008) | 0.98 ± (0.017) | 0.98 ± (0.023) | 0.95 ± (0.046) | 0.93 ± (0.061) | 0.91 ± (0.074) |
| | CaffeNet | 0.99 ± (0.009) | 0.98 ± (0.012) | 0.97 ± (0.020) | 0.95 ± (0.041) | 0.93 ± (0.060) | 0.91 ± (0.081) |
| | Vgg-f | 0.99 ± (0.006) | 0.98 ± (0.014) | 0.97 ± (0.019) | 0.96 ± (0.033) | 0.94 ± (0.056) | 0.90 ± (0.085) |
| | Concatenate | **0.99 ± (0.003)** ↑ | **0.99 ± (0.002)** ↑ | **0.98 ± (0.011)** ↑ | **0.96 ± (0.022)** ↑ | **0.94 ± (0.043)** ↑ | **0.91 ± (0.067)** ↑ |
| 5-fold (4) | AlexNet | 0.99 ± (0.008) | 0.98 ± (0.017) | 0.97 ± (0.028) | 0.95 ± (0.047) | **0.94 ± (0.047)** | 0.91 ± (0.073) |
| | CaffeNet | 0.99 ± (0.008) | 0.98 ± (0.016) | 0.98 ± (0.019) | 0.94 ± (0.056) | 0.93 ± (0.063) | 0.92 ± (0.063) |
| | Vgg-f | 0.99 ± (0.009) | 0.98 ± (0.015) | 0.97 ± (0.026) | 0.94 ± (0.051) | 0.93 ± (0.061) | 0.91 ± (0.074) |
| | Concatenate | **0.99 ± (0.006)** ↑ | **0.99 ± (0.002)** ↑ | **0.98 ± (0.014)** ↑ | **0.95 ± (0.029)** ↑ | 0.94 ± (0.054) ↓ | **0.92 ± (0.057)** ↑ |
| 5-fold (5) | AlexNet | 0.99 ± (0.009) | 0.98 ± (0.016) | 0.97 ± (0.028) | 0.94 ± (0.051) | 0.93 ± (0.053) | 0.90 ± (0.085) |
| | CaffeNet | 0.99 ± (0.009) | 0.98 ± (0.014) | 0.97 ± (0.021) | 0.95 ± (0.041) | 0.93 ± (0.050) | 0.91 ± (0.076) |
| | Vgg-f | 0.99 ± (0.007) | 0.98 ± (0.011) | 0.97 ± (0.023) | 0.94 ± (0.057) | 0.93 ± (0.049) | 0.91 ± (0.070) |
| | Concatenate | **0.99 ± (0.005)** ↑ | **0.99 ± (0.003)** ↑ | **0.98 ± (0.017)** ↑ | **0.95 ± (0.038)** ↑ | **0.93 ± (0.042)** ↑ | **0.91 ± (0.066)** ↑ |

of less than 500 had a high hit rate of above 99% in the binary classification and 0.95 in multi-classification. By increasing the size of the vector, these rates are maintained throughout this percentage. Moreover, when only 5% of the features are applied, the suggested method can achieve the highest accuracy rate. By increasing the size of the vector, these rates are maintained throughout this percentage. Moreover, when only 5% of the features are applied, the suggested method can achieve the highest accuracy rate. Therefore, we emphasize the need for feature selection, which expedites the classification level and reduces the computational complexity in both classification forms.

Our further approach is the proper design with 5% of features as it conceded the highest robustness and accuracy. As displayed in Fig. 5, converges of the algorithm has the minimum value based on differences in the number of features. In this figure, the two diagrams describe binary and multi classifications. Additionally, the loss value is computed

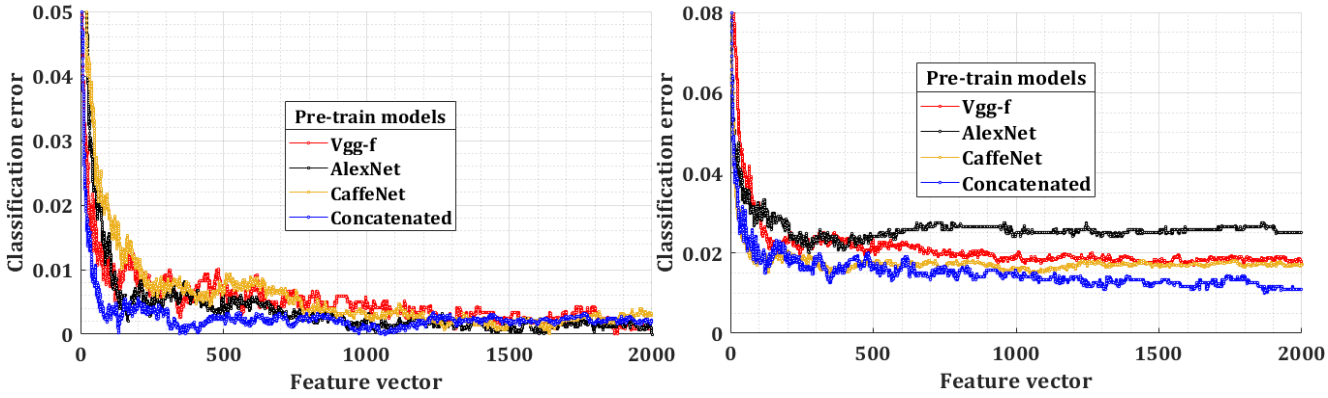


**Fig. 5. Accuracy of binary (left) and multi (Right) classifications obtained by the proposed approach, varying the size of the feature vector for Vgg-f, CaffeNet, AlexNet, and Concatenated models.**

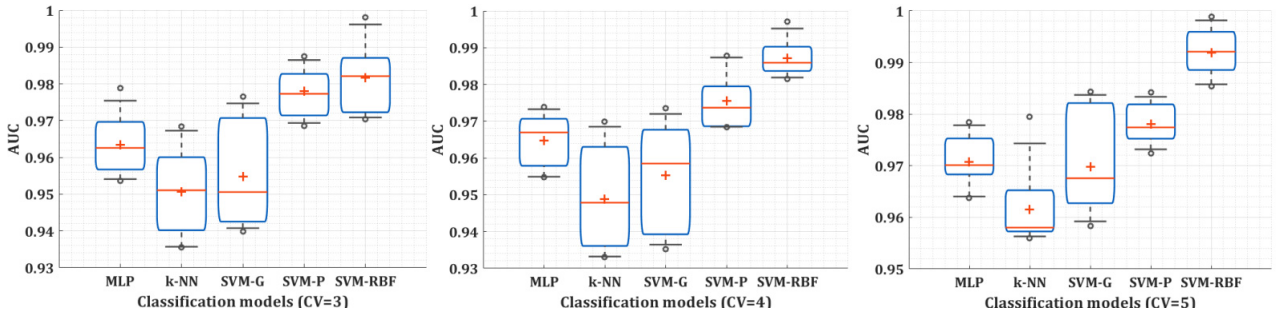


**Fig. 6. Results are compared based on the box-plots shown for different classifiers in which different values of K in K-fold CV are analyzed.**

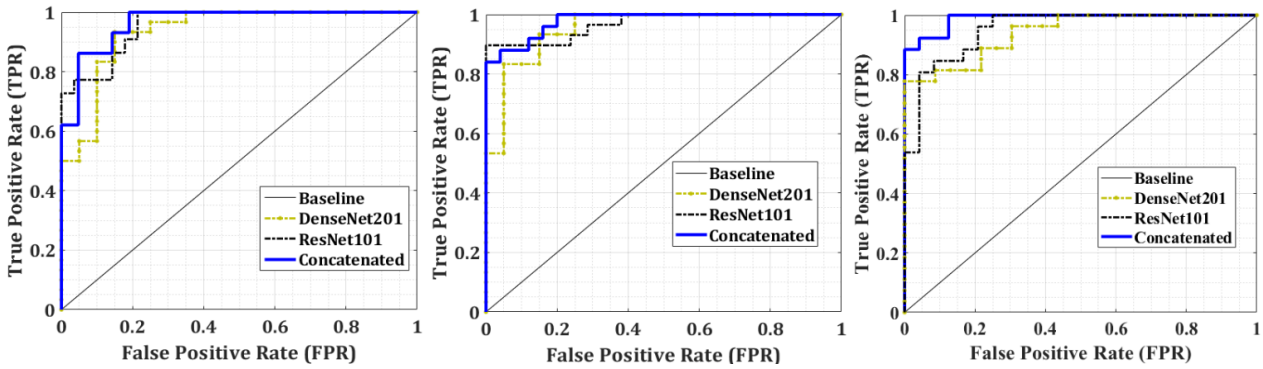


**Fig. 7. Comparing AUC and ROC curves between DenseNet201, ResNet101, and Concatenated pre-train model (K or CV range of 3 to 5).**

based on the extracted feature from Vgg-f, AlexNet, CaffeNet, and Concatenate models in all three pre-train purposes.

The experiments are fulfilled for the Multi-Layer Perceptron (MLP), k-nearest neighbor (k-NN), SVM-Gaussian, SVM-Polynomial, and SVM-RBF kernels classifiers (the K selected in the range of 3 to 5). In Fig. 6, the outcomes are examined and compared as a box-plot based on various classifiers. As can be observed, the SVM-RBF kernel yields the highest accuracy with the lowest number of features. Tables 3 and 4 compare the accuracy, precision, recall, and f-measure evaluation metrics of the introduced method with other state-of-the-art approaches for validation and test CXRs. As displayed in Tables 3 and 4, the suggested scheme outperforms other methods exhibited in the literature. We believe that the outcomes could still be improved by a fine-tuning function such as C and γ parameters in the SVM-RBF kernel classifier. We calculated the AUC for similar pre-train designs, as illustrated in Fig. 7. These techniques involve DenseNet201, ResNet101, and the concatenate model for AlexNet, CaffeNet, and Vgg-f networks. In these plots, the AUC value of the concatenate strategy is higher.

**Table 3. Comparison of the accuracy, precision, recall, and f-measure evaluation metrics of the proposed strategy with other state-of-the-art schemes for binary classifications**

| Model | Precision | Recall | F-measure | Specificity | Accuracy |
|---|---|---|---|---|---|
| Inception | 0.953 | 0.952 | 0.955 | 0.957 | 0.954 |
| DenseNet201 | 0.964 | 0.965 | 0.966 | 0.967 | 0.962 |
| ResNet101 | 0.922 | 0.954 | 0.950 | 0.923 | 0.943 |
| Vgg16 | 0.987 | 0.973 | 0.982 | 0.983 | 0.988 |
| **Proposed** | **0.995** | **0.992** | **0.990** | **0.991** | **0.996** |

**Table 4. Comparison of the accuracy, precision, recall, and f-measure evaluation metrics of the proposed strategy with other state-of-the-art schemes for multi-class classifications**

| Model | Precision | Recall | F-measure | Specificity | Accuracy |
|---|---|---|---|---|---|
| Inception | 0.915 | 0.916 | 0.917 | 0.911 | 0.913 |
| DenseNet201 | 0.916 | 0.911 | 0.912 | 0.914 | 0.910 |
| ResNet101 | 0.909 | 0.906 | 0.901 | 0.905 | 0.904 |
| Vgg16 | 0.912 | 0.913 | 0.916 | 0.919 | 0.915 |
| **Proposed** | **0.934** | **0.938** | **0.931** | **0.933** | **0.935** |

**Table 5. Performance comparison of recent works based on deep learning models**

| Author(s) | Method | Dataset | Accuracy | Samples |
|---|---|---|---|---|
| Wang et al. [19] | DL model | Collected | 0.933 | 13,975 |
| Jaiswal et al. [5] | DenseNet201 | Kaggle | 0.952 | 1262 |
| Chen et al. [29] | UNet++ | Collected | 0.988 | **106** |
| Wang et al. [27] | InspectionNet | COVIDx | 0.852 | 1065 |
| Mpesiana et al. [20] | DTL model | Public | 0.967 | 1427 |
| Xu et al. [28] | 3D CNN | Kaggle | 0.867 | 618 |
| Narin et al. [9] | InceptionResNet | Kaggle | 0.980 | - |
| Abbas et al. [21] | TraC CNN | Collected | 0.955 | - |
| Zhang et al. [23] | DL-based | xVIRAL | 0.960 | 5000 |
| Shan et al. [25] | DL-based | Collected | 0.916 | 300 |
| Gozes et al. [26] | DL-based | Collected | 0.948 | 110 |
| Zheng et al. [31] | UNET-trained | Collected | 0.959 | 499 |
| Minaee et al. [32] | DL framework | 5k | 0.954 | 3100 |
| **Proposed** | Concatenated | Collected | **0.993** | 676 |

In Table 5, our suggested usage has been compared with existing procedures. Significant correspondence with the conclusions of Wang et al. [26], Jaiswal et al. [5], Mpesiana et al. [20], and Xu et al. [28] can be witnessed in Table 5. Wang et al. [26] extended transfer learning procedures employing the Inception model. They conducted experiments on 453 CT scans of confirmed COVID-19 cases, reporting an accuracy of 89.5%. Another deep learning-based scheme was introduced by Xu et al. [27], who recognized COVID-19 pneumonia from influenza-A viral pneumonia.

Our suggested recommendation approach was able to detect the COVID-19 from bacterial pneumonia (99.31% accuracy) and COVID-19 from viral pneumonia (98.83% accuracy). We employed three pre-trained CNN networks to design a combined transfer learning model and concluded that fine-tuning pre-trained CNN networks can be successfully used in limited-sized CXR images even without data augmentation to discriminate COVID-19 disease. However, augmenting the statistical community, investigating patients, determining the COVID-19 severity, and exploring their influence can contribute to more generalizable outcomes. The constraints of our identification procedure based on CXR images involve the reliance on the COVID-19, while this disease is made by a variety of determinants. Given all these challenges, our suggested approach is nevertheless able to be considered a robust diagnosis based on CXR images as a novel strategy for COVID-19 disease.

## 4- CONCLUSION

This paper represents a hybrid scheme for diagnosing COVID-19 in CXR images using pre-train CNNs. Based on the suggested method's outcomes, the robustness of pre-trained CNNs for feature extraction could be measured corresponding to other classical state-of-the-art approaches. Utilizing the feature selection, we realized that more features were required to analyze images, including similar viruses, while fewer features were required for images that only revealed COVID-19. As discussed, CXR lets local hospitals and small clinics carry out analysis in a short time. The suggested procedure's major superiority is that it surpasses other state-of-the-art approaches without needing a segmentation process. Moreover, we demonstrated the robustness of the recommended model applying the collected CXR dataset. It is offered that future investigations prospect the manner of fine-tuning in the CNN and optimized SVM classifiers to develop the generality of COVID-19 information. Further to the improvement of CNNs performance, a new and large CXR could be utilized to provide validation in daily life, promoting this disease's diagnosis for physicians and patients.

## CONFLICT OF INTEREST

We have no conflict of interest to declare.

## ACKNOWLEDGMENT

We gratefully acknowledge the generous support of Meybod and Tabriz Universities. The authors wish to thank physicians, radiologists and Medical Imaging Center of Vasei Hospital, without their helps, this study could not have been accomplished

## REFERENCES

[1] F. Wu, S. Zhao, B. Yu, Y. M. Chen, W. Wang, Z. G. Song, Y. Hu, Z. W. Tao, J. H. Tian, Y. Y. Pei, M. L. Yuan, A new coronavirus associated with human respiratory disease in China, *Nature*, 579 (7798) (2020) 265-269.

[2] C. Huang, W. Yeming L. Xingwang R. Lili, Z. Jianping, H. Yi, Z. Li, and et al., Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China, *The lancet*, 39 (10223) (2020) 497-506.

[3] World Health Organization, WHO coronavirus disease (COVID- 19) dashboard, (2020). https://covid19.who.int/. Accessed 10 august 2020.

[4] S. Stoecklin, et al., First cases of coronavirus disease 2019 (COVID-19) in France: surveillance, investigations and control measures, January 2020, *Eurosurveillance*, 25 (6) (2020) 2000094.

[5] A. Jaiswal, N. Gianchandani, D. Singh, V. Kumar, and M. Kaur, Classification of the covid-19 infected patients using densenet201 based deep transfer learning, *Journal of Biomolecular Structure and Dynamics*, (2020) 1–8.

[6] S. Rakshit, I. Saha, M. Wlasnowolski, U. Maulik and D. Plewczynski, Deep Learning for Detection and Localization of Thoracic Diseases Using Chest X-Ray Imagery, *International Conference on Artificial Intelligence and Soft Computing*, (2019).

[7] M. Kaur and D. Singh, Fusion of medical images using deep belief networks, *Cluster Computing*, 23 (2020) 1439–1453.

[8] N. N. Das, N. Kumar, M. Kaur, V. Kumar, and D. Singh, Automated deep transfer learning-based approach for detection of covid-19 infection in chest x-rays, *IRBM*, 2020.

[9] A. Narin, C. Kaya and Z. Pamuk, Automatic detection of coronavirus disease (COVID-19) using X-ray images and deep convolutional neural networks, arXiv:2003.10849, (2020), [online] Available: http://arxiv.org/abs/2003.10849.

[10] A. S. Lundervold and A. Lundervold, An overview of deep learning in medical imaging focusing on MRI, *Zeitschrift für Medizinische Physik*, 29 (2) (2019) 102-127, 2019.

[11] A. Maier, C. Syben, T. Lasser and C. Riess, A gentle introduction to deep learning in medical image processing, *Zeitschrift für Medizinische Physik*, 29 (2) (2019) 86-101.

[12] A. Krizhevsky, I. Sutskever and G. Hinton, Imagenet classification with deep convolutional neural networks. *Communications of the ACM*, 60(6) (2017) 84-90.

[13] S. Liu and W. Deng, Very deep convolutional neural network based image classification using small training sample size, *Pattern Recognition (ACPR) 2015 3rd IAPR Asian Conference on*, (2015) 730-734.

[14] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, et al., Going deeper with convolutions, *arXiv:1409.4842*, (2014).

[15] K. He, X. Zhang, S. Ren and J. Sun, Deep residual learning for image recognition, *In Proceedings of the IEEE conference on computer vision and pattern recognition* (2016) 770-778.

[16] F. Chollet, Xception: Deep learning with depthwise separable convolutions, *In Proceedings of the IEEE conference on computer vision and pattern recognition* (2017) 1251-1258.

[17] G. Huang, Z. Liu, L. Van Der Maaten, K. Q. Weinberger, Densely connected convolutional networks. *In Proceedings of the IEEE conference on computer vision and pattern recognition,* (2017) 4700-4708.

[18] M. Loey, F. Smarandache, N. E. Khalifa, Within the Lack of Chest COVID-19 X-ray Dataset: A Novel Detection Model Based on GAN and Deep Transfer Learning," *Symmetry*, 12 (4) (2020) 651.

[19] L. Wang and A. Wong, COVID-net: A tailored deep convolutional neural network design for detection of COVID-19 cases from chest X-ray images, *arXiv:2003.09871*, (2020), [online] Available: http://arxiv.org/abs/2003.09871.

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

[21] A. Abbas, M. M. Abdelsamea and M. M. Gaber, Classification of covid-19 in chest x-ray images using detrac deep convolutional neural network, *In: arXiv*

*preprint arXiv:2003.13815*, (2020).

[22] E. El-Din Hemdan, M. A. Shouman and M. Esmail Karar, COVIDX-net: A framework of deep learning classifiers to diagnose COVID-19 in X-ray images, arXiv:2003.11055, (2020), [online] Available: http://arxiv.org/abs/2003.11055.

[23] J. Zhang et al., Covid-19 screening on chest x-ray images using deep learning based anomaly detection, *In: arXiv preprint arXiv:2003.12338*, (2020).

[24] T. Ai, Z. Yang, H. Hou, C. Zhan, C. Chen, W. Lv, et al., Correlation of chest CT and RT-PCR testing in coronavirus disease 2019 (COVID-19) in China: A report of 1014 cases, *Radiology*, 26 (2020).

[25] F. Shan, Y. Gao, J. Wang, W. Shi, N. Shi, M. Han, et al., Lung infection quantification of COVID-19 in CT images with deep learning, *arXiv:2003.04655*, (2020), [online] Available: http://arxiv.org/abs/2003.04655.

[26] O. Gozes, M. Frid-Adar, H. Greenspan, P. D. Browning, H. Zhang, W. Ji, et al., Rapid AI development cycle for the coronavirus (COVID-19) pandemic: Initial results for automated detection &; Patient monitoring using deep learning CT image analysis, *arXiv:2003.05037*, (2020), [online] Available: http://arxiv.org/abs/2003.05037.

[27] S. Wang, B. Kang, J. Ma, X. Zeng, M. Xiao, J. Guo, et al., A deep learning algorithm using CT images to screen for corona virus disease (COVID-19), *in MedRxiv*, (2020), [online] Available: https://doi.org/10.1101/2020.02.14.20023028.

[28] X. Xu, X. Jiang, C. Ma, P. Du, X. Li, S. Lv, et al., Deep learning system to screen coronavirus disease 2019 pneumonia, *arXiv:2002.09334*, (2020), [online] Available: http://arxiv.org/abs/2002.09334.

[29] J. Chen et al., Deep learning-based model for detecting 2019 novel coronavirus pneumonia on high-resolution computed tomography: A prospective study, *MedRxiv*, (2020).

[30] D. Das, K. C. Santosh, U. Pal, Truncated inception net: COVID-19 outbreak screening using chest X-rays, *Physical and engineering sciences in medicine* 43 (3) (2020) 915-925.

[31] C. Zheng, X. Deng, Q. Fu, Q. Zhou, J. Feng, H. Ma, et al., Deep Learning-based Detection for COVID-19 from Chest CT using Weak Label, *Medrxiv*, (2020).

[32] S. Minaee, R. Kafieh, M. Sonka, S. Yazdani, G.J. Soufi, Deep-covid: Predicting covid-19 from chest x-ray images using deep transfer learning, *arXiv preprint arXiv:2004*.09363 (2020).

[33] T. D. Pham, Classification of COVID-19 chest X-rays with deep learning: new models or fine tuning?, *Health Information Science and Systems*, *9*(1) (2021) 1-11.

[34] M. Loey, G. Manogaran, M. H. Taha, N. E. Khalifa. A hybrid deep transfer learning model with machine learning methods for face mask detection in the era of the COVID-19 pandemic. *Measurement*. 167 (2021), 108288.

[35] M. F. Aslan, M. F. Muhammed, K. Sabanci, A. Durdu, CNN-based transfer learning–BiLSTM network: A novel approach for COVID-19 infection detection, *Applied Soft Computing* 98 (2021): 106912.

[36] K. Rezaee, A. Badiei, S. Meshgini, A hybrid deep transfer learning based approach for COVID-19 classification in chest X-ray images. *In 2020 27th National and 5th International Iranian Conference on Biomedical Engineering (ICBME)* (2020) 234-241.

[37] K. Chatfield, K. Simonyan, A. Vedaldi and A. Zisserman, Return of the devil in the details: Delving deep into convolutional nets, *arXiv preprint arXiv:1405.3531* (2014).

[38] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, et al., Caffe: Convolutional architecture for fast feature embedding, *In Proceedings of the 22nd ACM international conference on Multimedia*, (2014) 675-678.

[39] O. Russakovsky et al., ImageNet large scale visual recognition challenge, *International journal of computer vision* 115 (3) (2015) 211-252.

[40] K. Rezaee, F. Ghaderi, H. Taheri Gorji, J. Hadadnia, Hand gesture and movement recognition based on electromyogram signals using novel soft ensembling feature selection and optimized classifier, *Journal of Biomedical Engineering*, 14 (3) (2020) 221-230 (In persian). https://doi.org/10.22041/ijbme.2020.123570.1580.

[41] X. W. Chen, M. Wasikowski, Fast: a roc-based feature selection metric for small samples and imbalanced data classification problems. *In Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*, (2008) 124-132.

[42] S. Cateni, V. Colla , M.Vannucci, A hybrid feature selection method for classification purposes. *In 2014 European Modelling Symposium* (2014), 39-44.

[43] T. Nguyen, A. Khosravi, D. Creighton, S. Nahavandi, Hidden Markov models for cancer classification using gene expression profiles, *Information Sciences*, 316 (2015) 293-307.

[44] J. P. Vert, K. Tsuda, B. Schölkopf, A primer on kernel methods. *Kernel methods in computational biology*, 47(2004) 35-70.

[45] K. Rezaee, J. Haddadnia, Intelligent and Online Evaluation of Diabetes using Wireless Sensor Networks and Support Vector Machines Algorithm, *Iranian Journal of Diabetes and Obesity*, 6 (2) (2014) 56-66.

**HOW TO CITE THIS ARTICLE**

*K.Rezaee, A.Badiei, Hossein Gh.Zadeh, S.Meshgini4, A Hybrid Deep Transfer Learning-based Approach for COVID-19 Classification in Chest X-ray Images, AUT J. Elec. Eng., 53(2) (2021) 223-232.*

**DOI:** 10.22060/eej.2021.19467.5397