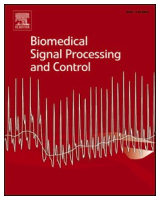
Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Biomedical Signal Processing and Control 69 (2021) 102862

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

journal homepage: www.elsevier.com/locate/bspc



Automated detection of Covid-19 disease using deep fused features from chest radiography images

Emine Uçar [a,*], Ümit Atila [b], Murat Uçar [a], Kemal Akyol [c]

[a] Department of Management Information Systems, Faculty of Business and Management Science, Iskenderun Technical University, Hatay, Turkey
[b] Department of Computer Engineering, Faculty of Engineering, Gazi University, Ankara, Turkey
[c] Department of Computer Engineering, Faculty of Engineering and Architecture, Kastamonu University, Kastamonu, Turkey

ARTICLE INFO

Keywords:
Covid-19
Pneumonia
X-ray
Deep learning
Bi-LSTM
Automatic medical diagnosis

ABSTRACT

The health systems of many countries are desperate in the face of Covid-19, which has become a pandemic worldwide and caused the death of hundreds of thousands of people. In order to keep Covid-19, which has a very high propagation rate, under control, it is necessary to develop faster, low-cost and highly accurate methods, rather than a costly Polymerase Chain Reaction test that can yield results in a few hours. In this study, a deep learning-based approach that can detect Covid-19 quickly and with high accuracy on X-ray images, which are common in every hospital and can be obtained at low cost, was proposed. Deep features were extracted from X-Ray images in RGB, CIE Lab and RGB CIE color spaces using DenseNet121 and EfficientNet B0 pre-trained deep learning architectures and then obtained features were fed into a two-stage classifier approach. Each of the classifiers in the proposed approach performed binary classification. In the first stage, healthy and infected samples were separated, and in the second stage, infected samples were detected as Covid-19 or pneumonia. In the experiments, Bi-LSTM network and well-known ensemble approaches such as Gradient Boosting, Random Forest and Extreme Gradient Boosting were used as the classifier model and it was seen that the Bi-LSTM network had a superior performance than other classifiers with 92.489% accuracy.

1. Introduction

Advances in the field of artificial intelligence increase its importance in the interpretation of medical images in order to support the early detection, correct diagnosis and treatment of diseases [1]. The Covid-19 disease that occurred in Wuhan, China in December 2019 has spread rapidly and has become a pandemic. Investigating the causes and effects of the Covid-19 pandemic, which is a serious threat to human health, has become a focus for scientists and healthcare professionals. The effects of this disease on people are followed anxiously by the world. In addition to the damage caused by Covid-19 disease to the organs in the human body, many researches are conducted on their psychological effects [2–4]. Researchers are constantly making efforts to control this epidemic and to find possible solutions in their fields [5–7]. One of the important steps in the fight against Covid-19 is to ensure that the infected patients can be screened effectively and thus they can be isolated and treated. Real-time reverse transcription polymerase chain reaction is the main screening method currently used for scanning [8,9]. As an alternative to this method, the researchers [10,11] stated that chest radiography images may be useful in Covid-19 detection. Studies have reported that patients with Covid-19 symptoms have mist-darkened spots in their lungs that can separate these patients from Covid-19 non-infected individuals [11,12]. Therefore, systems based on chest radiology are considered an effective material for the detection and classification of Covid-19. The method used in the acquisition of these images is the use of computerized tomography scan (CT-Scan) and X-rays in a hospital with medical equipment. Since most of the hospitals have CT imaging machines, the systems developed based on these images can be useful in order to test many patients quickly in hospitals where there are no test kits or in limited numbers. Moreover, diagnosis and interpretation of Covid-19 disease using chest CT images requires additional time for a field specialist, and an increase in the workload on field specialist due to densities in hospitals may result in unintentional erroneous decisions.

Today, in many countries, researchers from different disciplines are doing research to fight Covid-19, as well as other researchers are conducting many experimental studies to detect Covid-19 from chest radiography images. In recent years, the increase in the speed and capacities

* Corresponding author.
E-mail address: emine.ucar@iste.edu.tr (E. Uçar).

https://doi.org/10.1016/j.bspc.2021.102862
Received 5 December 2020; Received in revised form 12 April 2021; Accepted 7 June 2021
Available online 11 June 2021

of CPUs and GPUs has enabled the development of new high-performance computing models that can directly process raw data without the need for features [13]. Deep neural network architectures with multiple layers and neurons can efficiently perform high-complexity tasks such as voice and image recognition by processing large-size data. The use of deep learning models in the diagnosis and classification of diseases from medical image is quite common [14–19]. Extracting the features needed to perform classification in classical machine learning methods involve complex processes and must be done carefully. Moreover, hand-designed feature extraction directly affects classification performance. For this reason, many researchers [20–28] carried out deep learning-based studies for accurate and reliable detection of Covid-19.

In this study, we propose a deep learning model that can be used in the design of an expert system that can automatically classify Covid-19, pneumonia and no-finding cases. This model extracts deep features on different color spaces such as RGB, CIE Lab and RGB-CIE obtained from chest radiography images using DenseNet and EfficientNet transfer learning approaches and performs classification with Bidirectional Long Short-Term Memory (Bi-LSTM) network. RGB-CIE color space is one of many RGB color spaces distinguished by a specific monochromatic main color group [29]. The most prominent feature of the CIE-Lab color space is the smooth change of the color space in terms of perception. This color space defines all colors that the human eye can perceive, and is more useful in digital image processing than RGB color space for image sharpening and removing artifacts from the image [30]. This study examines the effects of deep features on image classification which reveal the different characteristics of mentioned color spaces, and also evaluates the contribution of combining deep features obtained from different color spaces to classification accuracy.

With this study, it is aimed to have a model that may contribute to the detection of Covid-19 at low cost, low error and high speed, especially in hospitals without diagnostic kits. The contributions of this paper are as follows:

- The proposed approach performs direct detection of Covid-19 disease on chest radiography images.
- The proposed approach realizes 3-class classification with high accuracy without the need for handcrafted features.
- Features that are automatically extracted from X-Ray images in different color spaces with deep transfer learning architectures are fused to increase the classification accuracy.
- The proposed approach can be used as an efficient expert system to support field experts, as it can quickly perform direct classification on X-ray images that are easily obtained in hospitals.

The rest of this work is organized as follows. Section 2 gives a literature on deep-learning based Covid-19 detection. Section 3 describes the dataset and presents proposed model used for Covid-19 detection. Experiments are presented in Section 4. The obtained results are given and discussed in Section 5. Finally, the study is concluded in Section 6.

2. Literature review

Khan et al. [20] proposed a method based on Xception deep learning architecture for classification of normal, pneumonia and Covid-19 disease from Chest X-Ray images. They tested the proposed CoroNet model on two different datasets, and in the first dataset, there were 89.5% average accuracy in the 4-class case including normal, pneumonia-bacterial, pneumonia-viral and Covid-19 images, 94.59% in the 3-class case including normal, pneumonia and Covid-19 images, 99% in the 2-class case including normal and Covid-19 images. In the second dataset, they reported that they achieved 90% success in three-class classification including normal, pneumonia and Covid-19 images. Pathak et al. [21] proposed a model based on ResNet-50 deep learning architecture for the diagnosis of Covid-19. In the study, they used images containing 413 Covid-19 and 439 normal or pneumonia diseases obtained from various sources. They used 60% of the dataset for training, 40% for test and 10% of training data for validation. They reported that the success rate in binary classification was approximately 93.02%. Ozturk et al. [22] suggested a deep learning network called DarkCovidNet for diagnosis of Covid-19 disease from Chest X-Ray images. Their model uses the modified layers and filters of DarkNet-19 architecture based on YOLO which is a real-time object recognition system. Their proposed model achieved average accuracy of 98.08% in 2-class classification including Covid-19 and normal cases, and 87.02% in three-class classification including Covid-19, normal and pneumonia cases. Panwar et al. [23] proposed a new model called nCOVnet based on VGG16 deep learning architecture for the diagnosis of Covid-19 disease. They applied data augmentation techniques to Chest X-ray images of normal and Covid-19 patients that are collected from various public datasets and achieved 88% overall accuracy. Apostolopoulos et al. [24] applied CNN-based transfer learning approaches to classify 3 cases including normal, pneumonia and Covid-19 and obtained overall accuracy of 94.72% using MobileNet v2. Rahimzadeh et al. [25] presented a new dataset containing 15,589 images of 95 Covid-19 patients and 48,260 images of normal patients. They reported that the classification success of their model proposed as a modified version of the ResNet50V2deep learning architecture was 98.49%. Punn and Agarwal [26] used deep learning architectures such as ResNet, Inception-v3, Inception ResNet-v2, DenseNet169 and NASNetLarge for automatic diagnosis of Covid-19 disease. They reported that NASNetLarge model achieved the best success rate with 0.96 in multi-class classification and 0.98 in binary classification. Narin et al. [27] used ResNet50, InceptionV3 and Inception ResNetV2 deep learning architectures to detect Covid-19 disease from X-Ray images and reported that the ResNet50 model achieved 98% accuracy in binary classification. Wang and Wong [28] proposed a deep learning-based model for the detection of Covid-19 disease called Covid-Net. They reported that the proposed Covid-Net model achieved 93.3% success in classifying Normal, Non-Covid-19 and Covid-19 conditions.

3. Materials and methods

The main focus of this study is to reveal a stable and efficient model that successfully detects Covid-19 disease from X-Ray images. To this aim, radiography images belonging to Covid-19, pneumonia and no-finding classes were used. This model extracts features from chest radiography images RGB, CIE Lab and RGB CIE color spaces using DenseNet121 and EfficientNetB0 pre-trained deep learning models and performs two-stage classification on the deep fused features. In the first stage, the images are handled in two classes as no-finding and others (Covid-19 and pneumonia). The images labeled as patients in the first stage are passed to the second stage for detection of Covid-19 or pneumonia. Fig. 1 shows the block diagram of the proposed model.

As stated in the introduction section, the main goal of the study is to create a low-cost, fast-running diagnostic model with a low error rate. In this context, DenseNet121 (8,062,504) and EfficientNetB0 (5,330,571) models, which have much lower number of parameters compared to other state-of-the-art deep learning architectures, were selected in the feature extraction part of the proposed model. While most of the state-of-the-art models experience losses in the features obtained from the image as they progress through the layers, in DenseNet121 architecture, each layer connects to the next layers and thus layers can access the properties of the previous layers. EfficientNet architecture, on the other hand, can reduce the size of the model by performing compound scaling, thus obtaining more efficient results. These two architectural studies have come to the fore as the most appropriate options for the goal of realizing the desired fast and efficient disease diagnosis.

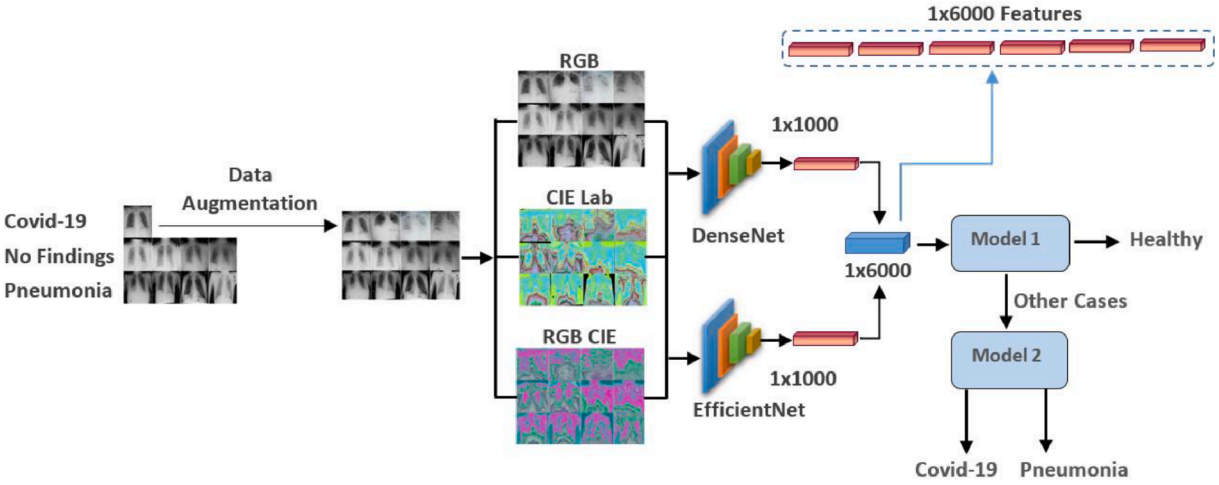


Fig. 1. Block diagram of the proposed model.

3.1. Chest radiography images

X-ray images used in the study belong to Covid-19, pneumonia and no-finding cases, andwere collected from different resources [31,32]. In the datasets, there are 1125 images in RGB color space in different sizes, and while 125 of these images belong to the Covid-19 class, there are 500 images belonging to each of the other classes. In order to achieve balanced data distribution, which is an important issue in machine learning, Covid-19 images were increased from 125 to 500 by applying data augmentation. Image rotation at 30-60-90 angles was chosen as the data augmentation method. Fig. 2 presents sample images of Covid-19 and other classes.

3.2. Pre-trained models for feature extraction

3.2.1. DenseNet121

When neural networks are trained, there is a decrease in feature maps due to convolution and subsampling processes. At the same time, there are losses in the image feature in the transition between layers. DenseNet121 architecture was proposed by Huang et al. [33] for more effective use of features extracted from images. In this architecture, each layer is connected to the other layers in a feed-forward manner. In this way, any layer can access the property information of all previous layers. In addition, DenseNet's other advantages stand out as lightening the lost angle problem and generously reducing the number of parameters.

3.2.2. EfficientNet B0

The EfficientNet model, which reaches 84.4% accuracy with 66 M parameter calculation load in the ImageNet classification problem, can be considered as a group of convolutional neural network models. EfficientNet includes 8 models between B0-B7 and as this model number grows, the number of parameters calculated does not increase much, while the accuracy increases remarkably. The purpose of deep learning is to obtain more efficient models with least number of parameters.

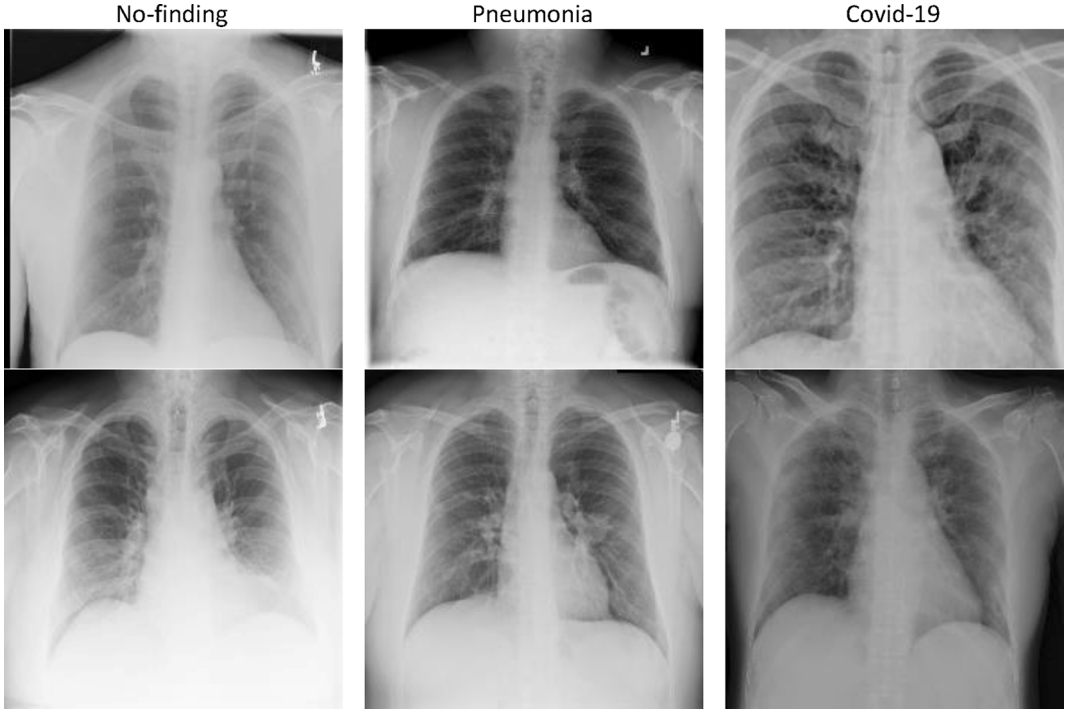


Fig. 2. Sample images of no-finding, pneumonia and Covid-19 classes.

Unlike other state-of-the-art deep learning models, the EfficientNet shrinks the model by compound scaling which performs scaling in terms of depth, width and resolution to produce more efficient results. In compound scaling method, firstly a grid search is performed to find the relationship between different scaling dimensions of the baseline network under a fixed resource constraint. Therefore, suitable scaling factors are determined for depth, width and resolution dimensions. These factors are then applied to scale the baseline network to the desired target network [34].

3.3. Bi-LSTM

LSTM networks have been developed due to the gradient vanishing problem in traditional RNNs [35]. A memory cell is used for learning and recall tasks in LSTM. This memory cell stores information about learning. Gate mechanisms are created with nonlinear activation functions in the memory cell. These mechanisms regulate the transmission of stored information to the next layer or forgetting it. Unlike the traditional LSTM structure, the bidirectional LSTM has two hidden layers (forward and backward) connected to the outputs. While the output layer in the LSTM network obtains the information from the input data at time t and the previous hidden state (t-1, t-2,…, t-N), bidirectional LSTM network also uses subsequent hidden state (t + 1, t + 2,…, t + N). Processing data in two-way hidden layers opposite each other provides additional information to the network including future data, which enables faster and better learning.

3.4. Performance metrics

There are 3 classes in the dataset used in this study, and the proposed approach realizes the classification with two independent models, each of which performs binary classification. In the first stage of the proposed model, healthy samples are treated as negative, and the diseased ones as positive. In the second stage, Covid-19 samples are considered positive and Pneumonia samples are considered negative. Considering confusion matrix, true positive (TP) is the number of correctly classified samples in the positive category and true negative (TN) is the number of correctly classified samples in the negative category. False negative (FN) is the number of misclassified samples in the positive category and false positive (FP) is the number of misclassified samples in the negative category.

The performances of the classifiers used in both stages of the proposed approach in this study were measured using different metrics such as Sensitivity (Sen), Specificity (Spe), Accuracy (Acc), Precision (Pre) and F1-score. The formulas for these metrics are given in Eqs. (1)–(10) respectively.

For a class k,

$$Sen(k) = \frac{TP(k)}{TP(k) + FN(k)} \tag{1}$$

$$Spe(k) = \frac{TN(k)}{TN(k) + FP(k)} \tag{2}$$

$$Acc(k) = \frac{TP(k) + TN(k)}{TP(k) + FN(k) + TN(k) + FP(k)} \tag{3}$$

$$Pre(k) = \frac{TP(k)}{TP(k) + FP(k)} \tag{4}$$

$$F_1 - score(k) = \frac{2}{\frac{1}{Sen(k)} + \frac{1}{Pre(k)}} \tag{5}$$

Then,

$$AverageSen = \frac{1}{classes}\sum_{k=1}^{classes} Sen(k) \tag{6}$$

$$AverageSpe = \frac{1}{classes}\sum_{k=1}^{classes} Spe(k) \tag{7}$$

$$AverageAcc = \frac{1}{classes}\sum_{k=1}^{classes} Acc(k) \tag{8}$$

$$AveragePre = \frac{1}{classes}\sum_{k=1}^{classes} Pre(k) \tag{9}$$

$$AverageF_1 - score = \frac{2}{\frac{1}{AverageSen} + \frac{1}{AveragePre}} \tag{10}$$

In addition to these metrics, the performance of the classifiers can also be evaluated with the Area Under the Receiver Operating Characteristic (AUC-ROC) metric. The ROC curve shows the relationship between True Positive Rate (TPR) and False Positive Rate (FPR). The TPR is the same with the sensitivity metric and gives the ratio of correctly predicted positives to all positives. FPR is the ratio of negatives that are incorrectly predicted as positive to all negative data and can be expressed as 1-specifity.

If the area under the ROC curve (Area Under Curve-AUC) is close to 1, this indicates that the model has a high success in separating the classes, if it is close to 0, this indicates that the model has a low success in separating the classes. If AUC equals to 0.5 this indicates that the model does not have the ability to discriminate between classes and selects a class randomly at each time [36].

4. Experiments

4.1. Experimental setup

All models used in this study were compiled with GPU support. All experimental studies were conducted in Google cloud environment using 64-bitUbuntu 18.04.3 LTS operating system with NVIDIA T80 GPU having 12 GB memory and Intel (R) Xeon (R) 2.00 GHz CPU and 12 GB RAM. All codes were realized with Keras 2.3.1, Scikit-image and Scikit-learn 0.22.2 libraries in Python.

4.2. Forming deep fused features

The X-ray images used in this study are in RGB color space. In addition to this color space, X-ray images in the dataset were converted into other color spaces such as LAB and CIE in order to examine the effect of using different color spaces on classification performance. DenseNet121 and Efficient B0 pre-trained deep learning models were used to extract deep features from X-ray images. All images were resized to 224x224 dimensions and normalized by the min–max method before inputting to these models. The feature size that each pre-trained model extracts on each image in the fully connected layer was 1x1000. In this context, DenseNet121 and Efficient B0 architectures extracted 3000 features on 3 different color spaces. Finally, these features were fused to obtain 6000 features for each image. These features were normalized in the range between 0 and 1 with min–max normalization. Thus, 1500x6000 dimensional feature vector was extracted.

4.3. Model training

Fig. 1 demonstrates the proposed approach that discriminates Covid-19 from other cases through two stages. In the first stage, a classifier model (Model 1) distinguishes healthy ones from infected (Covid-19 and pneumonia). In the second stage, Covid-19 cases are distinguished from pneumonia by another classifier model (Model 2). These two models work independently of each other. In the proposed approach, while all images including Covid-19, pneumonia and healthy cases are given as input to Model 1, Model 2 only takes Covid-19 and pneumonia images as input.

In this study, the k-fold cross validation technique where k was set to 5 was used to reveal and compare the performances of the algorithms more accurately and thus determine the best algorithm. In this context, as seen in Fig. 3, in each fold, 80% of the X-ray images were used for training and the rest of the images were used for testing. In the first stage, the classifiers in Model 1 were fed with a total of 1200 images including 400 healthy and 800 diseased for training. Besides, a total of 300 images, 100 from each class, were used to test the performance of Model 1. In the second stage, the Model 2 was fed by 800 diseased images for training, and then the performance of this model was tested on a total of 200 test images, including 100 images for each class of Covid-19 and pneumonia. Thus, Model 1 and Model 2 in the proposed approach were built.

In order to evaluate the final classification performance of the proposed approach on the basis of 3 classes, the test dataset containing 100 images from each class was given to a two-stage model. Accordingly, an input image was labeled as healthy or diseased by Model 1. If Model 1 labeled an image as diseased, Model 2 classified this image as Covid-19 or pneumonia. Thus, a 3-class classification was performed on all test images.

In the study, 6 feature vectors of 1x1000 were obtained from images in each color space using DenseNet and EfficientNet pre-trained models. For all classifiers except Bi-LSTM, 6 feature vectors were combined and a 1x6000 feature vector was used as input, while in the Bi-LSTM network, 1×1000 feature vectors from each color space were given as input to the model in 6 steps sequentially.

Along with selecting the Bi-LSTM as a classifier in the proposed approach, experiments were also conducted with other machine learning algorithms to compare their performance with Bi-LSTM. In this context, trainings were carried out in both stages of the proposed approach by recent popular ensemble learning algorithms such as Random Forest (RF), Gradient Boosting (GB) and Extreme Gradient Boosting (XGboost). These algorithms were run with default parameter values defined in Scikit-learn library.

Bi-LSTM network was built in Keras environment. For both stages of the proposed approach, the number of input layer units of Bi-LSTM, was set as 1000 and the number of output units as 2. The Bi-LSTM model was designed with 3 hidden layers and number of hidden layer units was set to 16. The hidden layer weights of the Bi-LSTM model were randomly initialized to have uniform distribution in the range between -1 and 1. Adam optimization method was chosen for Bi-LSTM network and the parameters of this method were set as $B1 = 0.9$ and $B2 = 0.999$. The learning rate was set as 5e-04 and decay as 0 and mean squared error (MSE) was used as the loss function of the network. The training of the Bi-LSTM network was completed in 100 epochs and the weights in the epoch, which obtained best validation accuracy, were stored and used in the final model to prevent overfitting. During the training, batch size was chosen as 200 since the number of samples per class in the training set is divisible with this value, and this results in more efficient use of memory. A dropout layer has been added to the outputs of all LSTM hidden layers to prevent overfitting. In addition, specific to the LSTM network, the recurrent dropout method which makes forward and backward dropout in each LSTM layer was also applied. For both dropout technique, parameter values were applied by being varied between 0.0 and 0.6 with step value of 0.1.

In accordance with 5-fold cross-validation technique, each test data in the dataset was used as validation data in the Bi-LSTM network. In both stages of the proposed approach, as a result of trial and error method, the Bi-LSTM network provided the best performance when dropout and recurrent dropout values were selected as 0.2 and 0.2, respectively. Accuracy / Loss curves obtained from training and validation sets during the training process with these values were presented in Fig. 4.

5. Results

In this study, a two-stage classification approach was presented on deep features obtained from Chest X-ray images using DenseNet121 and EfficientNet B0 pre-trained architectures. The focus of the study was to predict Covid-19 disease with an acceptable accuracy. For the detection of Covid-19 disease, Bi-LSTM network was used in both stages of the proposed approach. In addition, the performance of the Bi-LSTM network used in the proposed approach was compared with ensemble learning algorithms such as Random Forest (RF), Gradient Boosting (GB) and Extreme Gradient Boosting (XGboost).

The approach proposed in this study performs 3-class classification with two separate models that each make binary classifications consecutively. Binary classification accuracies of the classifiers used in Model 1 and Model 2 in the proposed approach for RGB, CIE Lab and RGB CIE color spaces and as well as for the combined color space were summarized in Tables 1 and 2. It was seen that both Model 1 and Model 2 used in the proposed approach achieved the highest accuracies using Bi-LSTM in all color spaces. Moreover, when the successes of the classifiers in Model 1 and Model 2 for the combined color space were

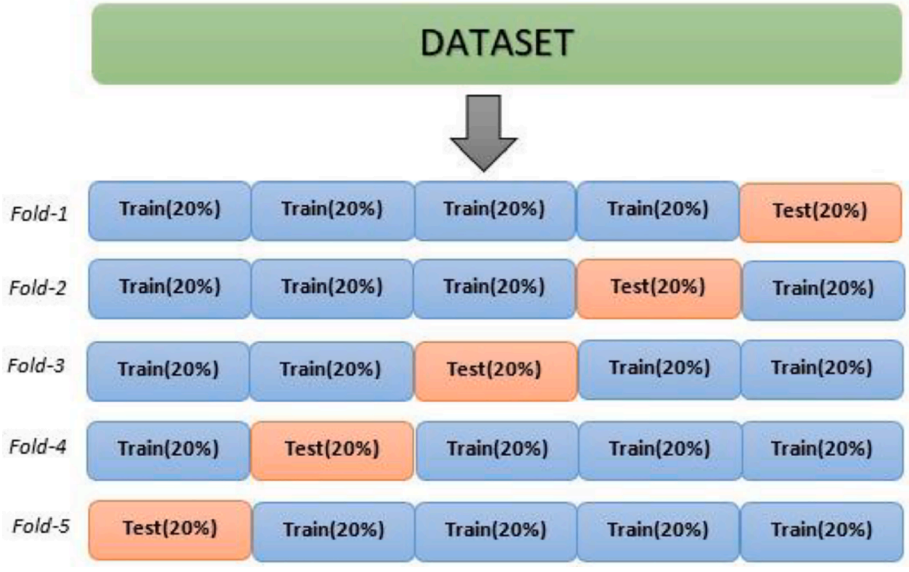


Fig. 3. Training and validation scheme applied in 5-fold cross-validation.

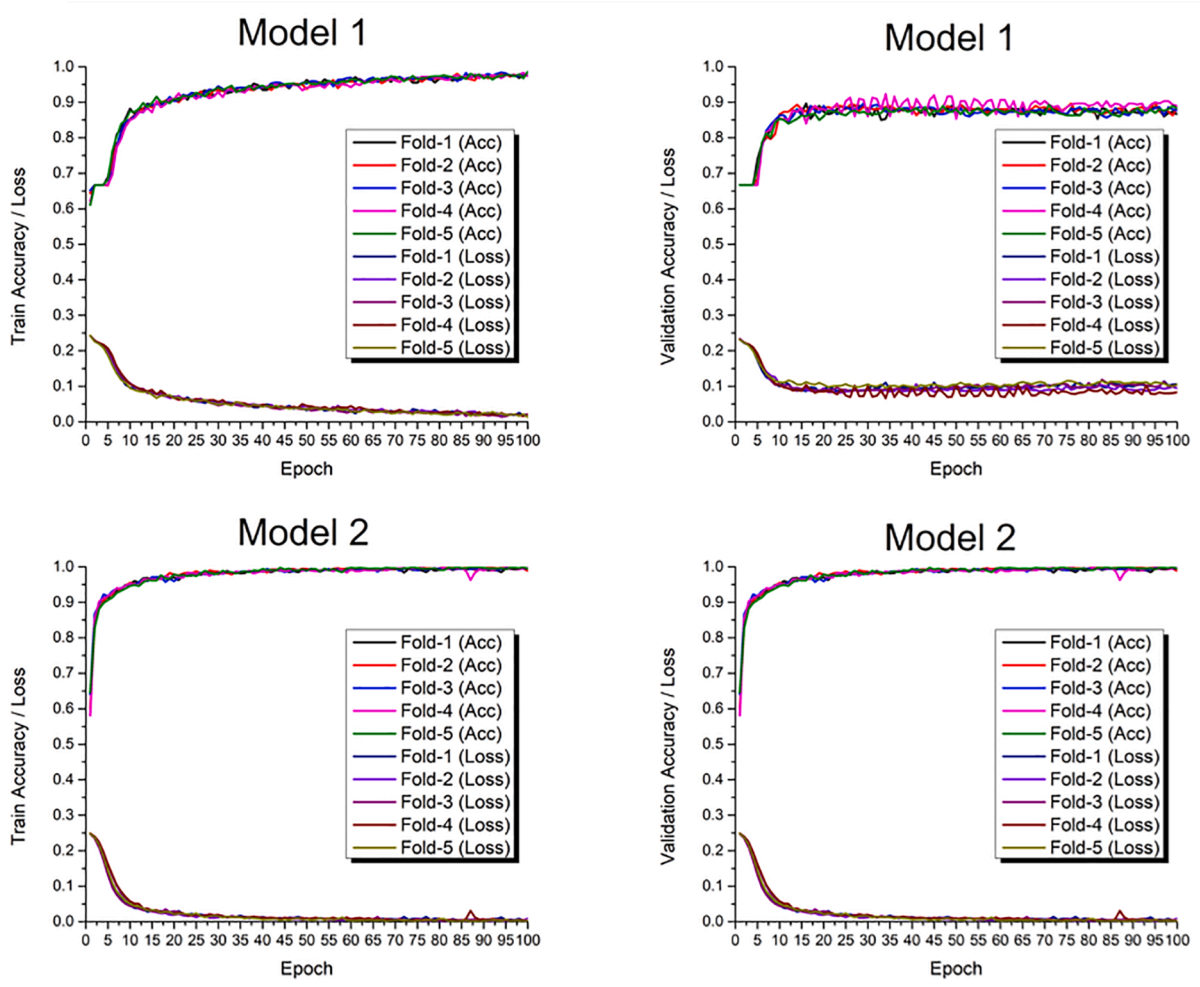


Fig. 4. Accuracy/loss curves for training and validation sets.

Table 1
Binary classification results of Model 1 (No-finding vs Others).

	Average Acc (%)			
	GB	RF	XGB	Bi-LSTM
RGB	86.800	83.067	86.867	89.400
RGB CIE	87.000	84.933	86.867	89.400
CIE Lab	86.200	84.733	86.200	88.600
All color spaces (Concatenated)	**87.600**	**85.533**	**87.133**	**90.200**

Table 2
Binary classification results of Model 2 (Covid-19 vs Pneumonia).

	Average Acc (%)			
	GB	RF	XGB	Bi-LSTM
RGB	96.000	93.800	95.700	97.700
RGB CIE	95.100	93.200	95.300	97.600
CIE Lab	86.200	84.733	86.200	97.900
All color spaces (Concatenated)	**96.100**	**94.300**	**96.800**	**98.000**

examined, it was seen that Bi-LSTM offers the highest accuracy compared to others.

In addition, Table 3 summarizes the performance of the classifiers tested in the 3-class classification with the average Acc, Sen, Spe, F1-score and Pre values obtained based on the 5-fold cross-validation method. As can be seen in this table, the best results were obtained with 92.489% average accuracy when Bi-LSTM was used in both stages

Table 3
Performance comparison of classifiers based on 3-class classification.

Classifier	Average Sen	Average Spe	Average Pre	Average F1-score	Average Acc
GB	85.267	92.633	85.659	85.406	90.178
RF	81.800	90.900	83.102	82.084	87.867
XGB	85.133	92.567	85.542	85.285	90.089
Bi-LSTM	88.933	94.367	88.886	88.799	92.489

of the proposed two-step approach. Among other classifiers in experimental studies, GB and XGboost algorithms showed the closest performances to Bi-LSTM. The values obtained by these two algorithms in the context of all evaluation metrics were lower than the Bi-LSTM method in the range between 2% and 3%.

In this context, since the highest performances in the proposed approach were obtained with Bi-LSTM network, results given below in this section were based on Bi-LSTM. 3-class classification performance obtained for each fold was evaluated with the accuracy, sensitivity, specificity, precision and F1-score metrics, and the average classification performances of 5-fold were presented in Table 4. Accordingly, the proposed approach presented an average of 88.933% sensitivity, 94.367% specificity, 89.028% precision, 88.760% F1-score and 92.489% accuracy in the 3-class classification. In addition, obtained confusion matrices for each fold and overlapped confusion matrix were shown in Fig. 5. As can be seen in the confusion matrices, the proposed approach classified Covid-19 patients more successfully in each fold

Table 4
Performance results of proposed approach on each fold based on Bi-LSTM.

	Average Sen	Average Spe	Average Pre	Average F1-score	Average Acc
Fold-1	88.667	94.333	89.120	88.614	92.444
Fold-2	88.333	94.167	88.679	88.380	92.222
Fold-3	89.000	94.500	89.102	88.947	92.667
Fold-4	90.000	95.000	90.378	90.109	93.333
Fold-5	87.667	93.833	87.861	87.750	91.778
Average	88.933	94.367	89.028	88.760	92.489

compared to other classes. For example, when overlapped confusion matrix was examined, the proposed approach misclassified 15 out of 500 Covid-19 cases, while it misclassified 73 out of 500 pneumonia and 78 out of 500 no-finding samples. Therefore, the proposed model can detect Covid-19 case more successfully compared to other cases. In the first stage of the two-stage model proposed here, Sensitivity and Precision values decrease due to misclassifications of the samples in pneumonia class as no-finding or vice versa.

When the overlapped confusion matrices (Fig. 6a–6b) of the Bi-LSTM were examined, it was seen that the number of misclassified samples in the no-finding and pneumonia classes was higher than Covid-19. Especially when the overlapped confusion matrices of Model 1 and Model 2 were compared, it is understood that Model 1 labeled 73 of 500 healthy images as diseased and 74 of 1000 diseased images as healthy. Moreover, it was seen that the performance of Model 2 in classifying Covid-19 and pneumonia images was higher than Model 1′s performance in classifying healthy and diseased images. As can be observed here, the performance of the proposed approach was lower in distinguishing between pneumonia and no-finding compared to Covid-19. It should also be noted that there were very small differences between the values in the overlapped confusion matrix obtained as a result of the 3-class classification and the values in the overlapped confusion matrix obtained in the 2-class classification of Model 1 and Model 2. This was because, in the proposed approach, Model 2 was dependent on classification according to the labeling result from Model 1. In other words, healthy images that Model 1 has mistakenly classified as patients must be classified as Covid-19 or Pneumonia by Model 2 and the image that Model 1 classified as healthy was not handled by Model 2.

Some of the samples that were misclassified by the proposed model were examined by a pulmonologist and the possible reasons that were thought to cause the images to be misclassified were explained. Fig. 7 shows eight sample images that were misclassified by the proposed two-stage model in the study. The images given in Fig. 7a and 7b were Covid-19 but misclassified as Pneumonia. In Fig. 7a, the faint infiltrative area in the left paracardiac region led to misclassification. In Fig. 7b, existence of indistinct infiltrations in bilateral basals led to misclassification. The images given in Fig. 7c and 7d belong to no-finding case but misclassified as Pneumonia. In Fig. 7c, scattered dense views in the right lower mediobasal region led to incorrect classification. The X-Ray image shown in Fig. 7d was taken without adequate inspiration and also has scattered dense views in the right lower mediobasal region. The

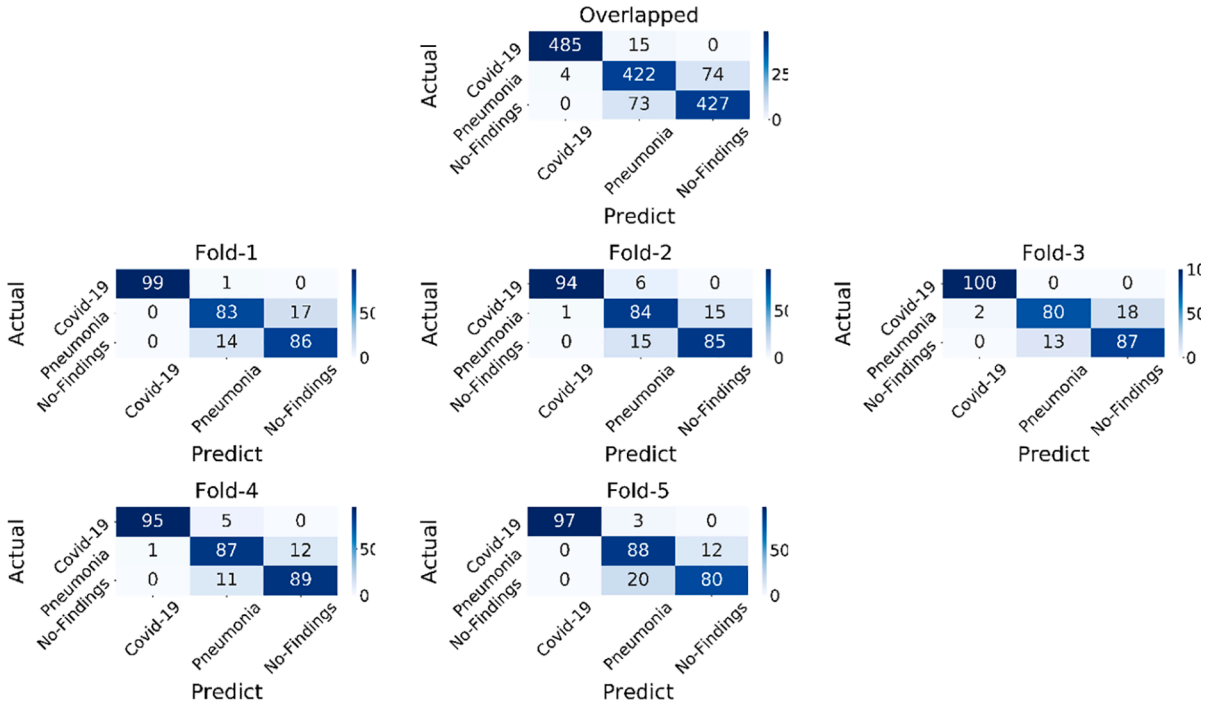


Fig. 5. Confusion matrices obtained for each fold and overlapped confusion matrix.

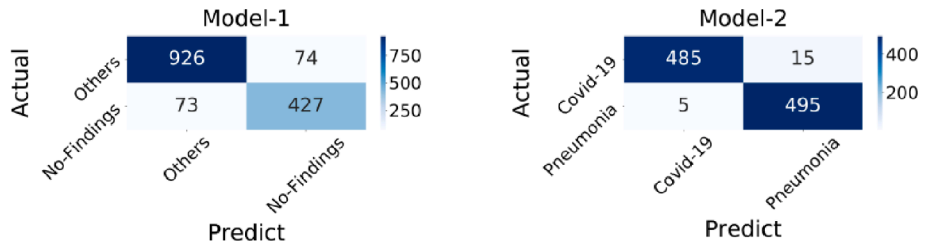


Fig. 6. Overlapped confusion matrices obtained for binary classifications: (a) Model 1 with Bi-LSTM, (b) Model 2 with Bi-LSTM.

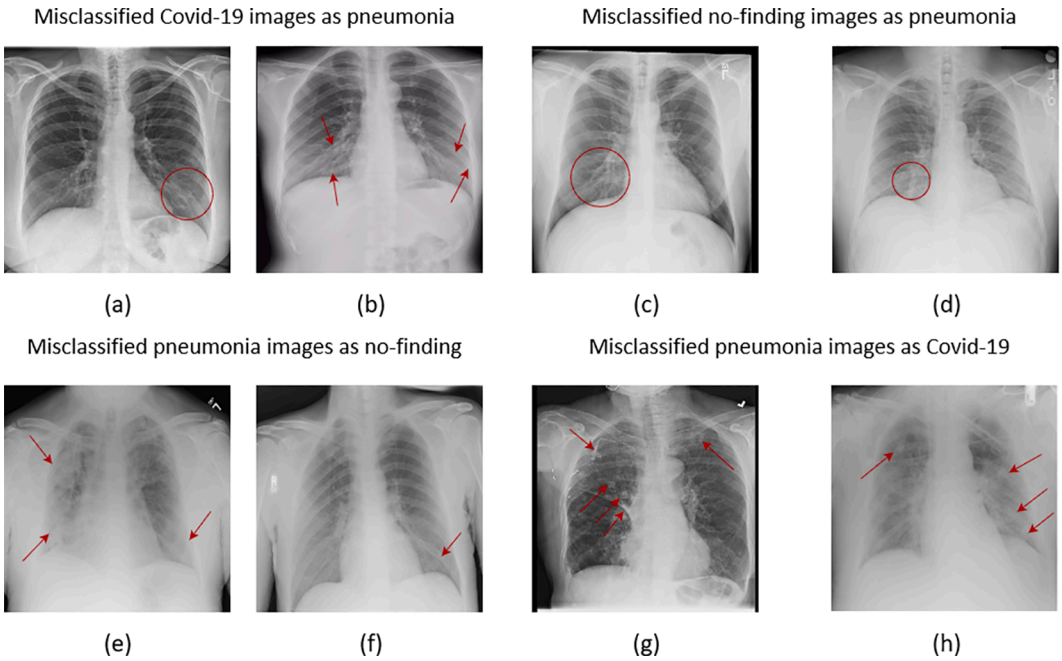


Fig. 7. Misclassified sample images.

Pneumonia image given in Fig. 7e was misclassified as no-finding due to difficulties in discriminating the upper and lower right regions as widespread density, consolidation or ground glass. Also this image has ground glass infiltration in the lower left region. In Fig. 7f, uncertain linear density at the bottom left caused misclassification. The images given in Fig. 7g and 7h belong to pneumonia class but misclassified as Covid-19. In Fig. 7g, several cases such as cavity at the left apex, two nodular views adjacent to the right hilus, appearance of branule at the right apex led to misclassification. Finally in Fig. 7h, diffuse reticular infiltrations in both lungs in the image, nodule in the upper right and nodular infiltrations on the left region led to misclassification.

Fig. 8 shows the class-based ROC curves and AUC values of the classifiers. As can be seen from the ROC curves in Fig. 8, the two-stage model we propose detects Covid-19 positive patients, which is the main goal of this study, with higher success compared to pneumonia and no-finding classes. As shown in the graphics, Bi-LSTM obtained the best AUC values for each class with 0.983 for Covid-19, 0.878 for pneumonia and 0.890 for no-finding.

6. Discussions

Table 5 analyzes the performances of the deep learning models

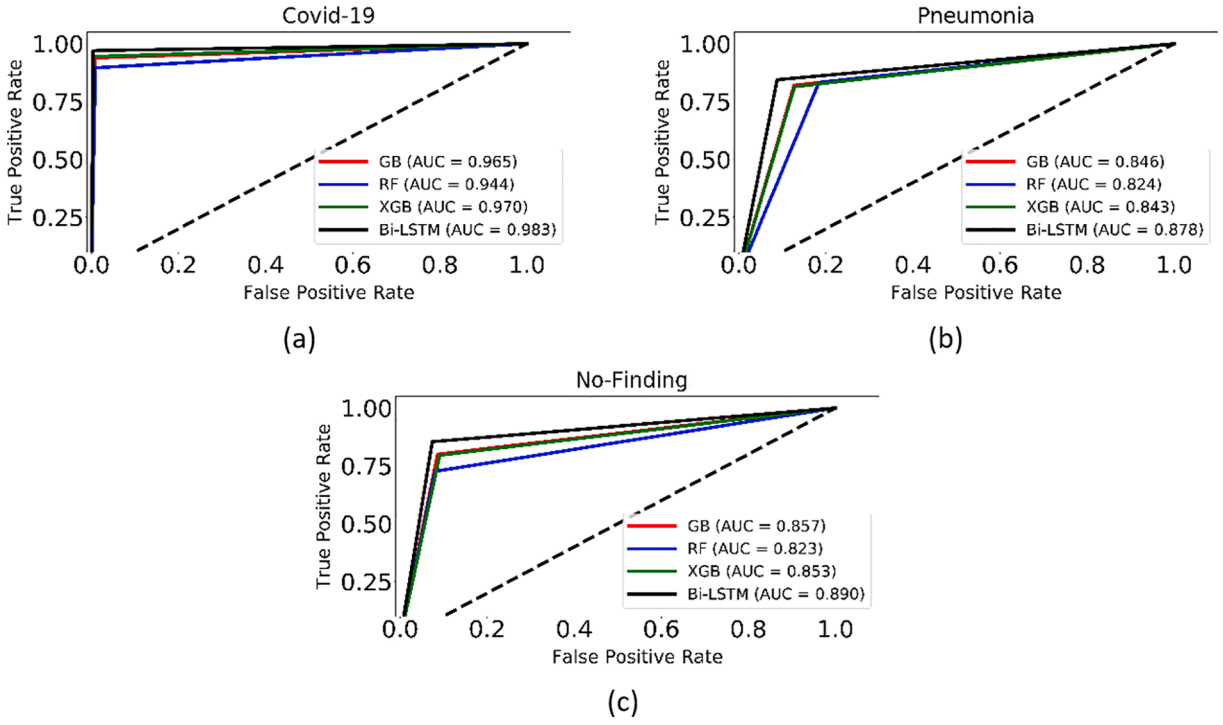


Fig. 8. ROC Curves for each class: (a) Covid-19, (b) pneumonia, (c) no-finding.

created for the diagnosis of Covid-19 and the datasets used in the literature. As can be seen in the table, the studies were generally carried out with two or three classes. While Covid-19 and normal classes were used in the studies with two classes, Covid-19, pneumonia and normal classes were used in three-class studies. Only in the study of Khan et al., the pneumonia case was divided into viral and bacterial, and the study was carried out with four classes [20]. It stands out that different numbers of images belonging to the relevant classes are used in the studies. Therefore, direct comparison of the performance of the models proposed in these studies with the performance of our model would be misleading. While the images used in the modified ResNet50 model of Rahimzadeh et al. were CT images, the images used in other studies were X-Ray images. In addition, the number of images used in their study was much higher than the number of images used in other studies. Consequently, the success of their model was higher than other studies with an overall accuracy of 98.49% in two-class classification [25]. Among the two-class studies using X-ray images, the DarkCovidNet model of Öztürk et al achieved the highest accuracy rate with 98.08% [22]. On the other hand, among the three-class studies, CoroNet model proposed by Khan et al. achieved the highest accuracy with 95% [20].

Since the dataset we used in this study was the same as the dataset in the study of Ozturk et al. [22], a direct comparison was performed only with this study. Ozturk et al. proposed a deep learning model named DarkNet for the diagnosis of Covid-19. The authors reported that the low number of Covid-19 images posed a disadvantage for their studies. In our study to overcome this disadvantage, we increased the number of images belonging to the Covid-19 class from 125 images to 500 by applying augmentation techniques and brought to the same number with the other two classes. As a result, while the DarkNet model presented by Ozturk et al. achieved an average accuracy of 91.35% which is calculated using their confusion matrix in 3-class classification, our proposed approach achieved an accuracy of 92.49% with the Bi-LSTM classifier, providing a slightly better performance.

Table 5
Covid-19 diagnosis performance comparison with other deep learning methods.

Author	Method	Dataset	Accuracy
Punn and Agarwal [26]	NASNetLarge	108 Covid-19 515 pneumonia 453 normal	2-class (Covid-19 and normal): 97% 3-class (Covid-19, pneumonia and normal): 94%
Pathak et al [21]	CNN	413 Covid-19 439 normal/ pneumonia	2-class (Covid-19 and others): 93.01%
Khan et al. [20]	CoroNet	290 Covid-19 310 normal 330 pneumonia-bacterial 327 pneumonia-viral	3-class (Covid-19, pneumonia and normal): 95% 4-class (Covid-19, pneumonia-bacterial, Pneumonia-viral and normal): 89.6%
Apostolopoulos & Mpesiana [24]	VGG19	224 Covid-19, 714 pneumonia 504 normal	2-class (Covid-19 and others): 96.78% 3-class (Covid-19, pneumonia and normal): 94.72%
Panwar et al. [23]	nCOVnet	142 Covid-19 142 normal	2-class (Covid-19 and normal): 97.62%
Rahimzadeh et al. [25]	Modified ResNet50	15,589 Covid-19 48,260 normal	2-class (Covid-19 and normal): 98.49 (overall accuracy)
Narin et al. [27]	ResNet-50	50 Covid-19 50 normal	2-class (Covid-19 and normal): 98%
Öztürk et al. [22]	DarkCovidNet	125 Covid-19 500 Pneumonia 500 normal	2-class (Covid-19 and normal): 98.08% 3-class(Covid-19, pneumonia and normal): 87.02% (overall accuracy)
Proposed model	**Two-stages model (Bi-LSTM)**	**500 Covid-19 (augmented) 500 Pneumonia 500 no-finding**	**3-class(Covid-19, pneumonia and normal): 92.489%**

7. Conclusions

Developing a method that can diagnose Covid-19 disease, which has turned into a pandemic that threatens the whole world, quickly and accurately, at low cost, is very important to prevent the collapse of health systems. Today, the most common method used for the detection of Covid-19 is a Polymerase Chain Reaction (PCR) test. Although this test is concluded within a few hours, it can still be considered slow, given the rate of spread of the Covid-19 virus. In addition, the fact that PCR is a high cost method limits the number of people to be tested in countries with low welfare and crowded populations. This gives the pandemic an opportunity to get out of control by causing it to spread faster in such countries. In this study, in order to avoid such limitations, a deep learning-based approach that can detect Covid-19 disease on X-ray-based images in a very short time without the need for any feature extraction method is proposed.

In order to have a balanced dataset and thus to have increased diagnostic success, the number of images belonging to the Covid-19 class was increased to have equal number of samples with the other two classes using various augmentation techniques. Secondly, for increasing the diagnostic success of the model, the features extracted with DenseNet121 and EfficientNet B0 pre-trained models in different color spaces, such as RGB, CIE Lab and RGB-CIE were fused. In both stages of the two-step approach proposed in the study, Bi-LSTM provided the best performance with 92.489% accuracy compared to other ensemble methods. Although the model proposed in this study had difficulty in separating the images of pneumonia and no-finding cases, it stood out with its high success in the detection of Covid-19 disease, which is the starting point of this study.

Since the proposed approach can be directly applied on X-ray images that can be easily obtained in almost every hospital, it avoids the disadvantages of antibody and PCR tests mentioned above in terms of both cost and time. Thus, the proposed approach can help the hospital workflow in determining which patients need PCR test, and thus reduce hospital workload, rather than making a definitive Covid-19 diagnosis. We achieved acceptable diagnosis performance with the limited number of Covid-19 image, and it is planned to build a more stable and successful model by working with more Covid-19 images in the future study. Besides, the hybrid machine learning-based feature extraction approaches proposed in recent years can be used to extract important features from X-Ray images, thus increasing the classification success by overcoming the difficulties in distinguishing between pneumonia and no-finding classes [37,38]. High-performance models to be developed in this way can contribute to slow down the spread of Covid-19, especially in hospitals lacking high-capacity devices.

CRediT authorship contribution statement

Emine Uçar: Conceptualization, Methodology, Writing - review & editing. **Ümit Atila:** Supervision, Methodology, Writing - original draft. **Murat Uçar:** Software, Visualization, Investigation. **Kemal Akyol:** Data curation, Investigation, Writing - review & editing.

Declaration of Competing Interest

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

Acknowledgements

We specially thank to Assoc. Prof. Dr. Sertaç Arslan for his invaluable evaluations on misclassified image samples.

References

[1] F. Jiang, Y. Jiang, H. Zhi, Y. Dong, H. Li, S. Ma, Y. Wang, Q. Dong, H. Shen, Y. Wang, Artificial intelligence in healthcare: past, present and future, Stroke Vasc. Neurol. 2 (2017) 230–243, https://doi.org/10.1136/svn-2017-000101.

[2] Y.E. Aksoy, V. Koçak, Psychological effects of nurses and midwives due to COVID-19 outbreak: the case of Turkey, Arch. Psychiatr. Nurs. (2020), https://doi.org/10.1016/J.APNU.2020.07.011.

[3] X. Cai, X. Hu, I.O. Ekumi, J. Wang, Y. An, Z. Li, B. Yuan, Psychological distress and its correlates among COVID-19 survivors during early convalescence across age groups, Am. J. Geriatr. Psychiatry (2020), https://doi.org/10.1016/J.JAGP.2020.07.003.

[4] Z. Zhu, Q. Liu, X. Jiang, U. Manandhar, Z. Luo, X. Zheng, Y. Li, J. Xie, B. Zhang, The psychological status of people affected by the COVID-19 outbreak in China, J. Psychiatr. Res. 129 (2020) 1–7, https://doi.org/10.1016/J.JPSYCHIRES.2020.05.026.

[5] E. Rutayisire, G. Nkundimana, H.K. Mitonga, A. Boye, S. Nikwigize, What works and what does not work in response to COVID-19 prevention and control in Africa, Int. J. Infect. Dis. 97 (2020) 267–269, https://doi.org/10.1016/J.IJID.2020.06.024.

[6] J. Lin, W. Huang, M. Wen, D. Li, S. Ma, J. Hua, H. Hu, S. Yin, Y. Qian, P. Chen, Q. Zhang, N. Yuan, S. Sun, Containing the spread of coronavirus disease 2019 (COVID-19): meteorological factors and control strategies, Sci. Total Environ. 744 (2020) 140935, https://doi.org/10.1016/J.SCITOTENV.2020.140935.

[7] Z. Baloch, Z. Ma, Y. Ji, M. Ghanbari, Q. Pan, W. Aljabr, Unique challenges to control the spread of COVID-19 in the Middle East, J. Infect. Public Health (2020), https://doi.org/10.1016/J.JIPH.2020.06.034.

[8] W. Wang, Y. Xu, R. Gao, R. Lu, K. Han, G. Wu, W. Tan, Detection of SARS-CoV-2 in different types of clinical specimens, JAMA 323 (2020) 1843–1844, https://doi.org/10.1001/jama.2020.3786.

[9] V.M. Corman, O. Landt, M. Kaiser, R. Molenkamp, A. Meijer, D.K.W. Chu, T. Bleicker, S. Brünink, J. Schneider, M.L. Schmidt, others, Detection of 2019 novel coronavirus (2019-nCoV) by real-time RT-PCR, Eurosurveillance 25 (2020) 2000045.

[10] A. Bernheim, X. Mei, M. Huang, Y. Yang, Z.A. Fayad, N. Zhang, K. Diao, B. Lin, X. Zhu, K. Li, S. Li, H. Shan, A. Jacobi, M. Chung, Chest CT findings in Coronavirus Disease-19 (COVID-19): relationship to duration of infection, Radiology 295 (2020) 200463, https://doi.org/10.1148/radiol.2020200463.

[11] X. Xie, Z. Zhong, W. Zhao, C. Zheng, F. Wang, J. Liu, Chest CT for typical 2019-nCoV pneumonia: relationship to negative RT-PCR testing. Radiology (2020) 200343–200343.

[12] Y. Fang, H. Zhang, J. Xie, M. Lin, L. Ying, P. Pang, W. Ji, Sensitivity of chest CT for COVID-19: comparison to RT-PCR, Radiology 296 (2020) E115–E117, https://doi.org/10.1148/radiol.2020200432.

[13] Y. LeCun, Y. Bengio, G. Hinton, Deep learning, Nature 521 (2015) 436–444, https://doi.org/10.1038/nature14539.

[14] F. Ciompi, B. de Hoop, S.J. van Riel, K. Chung, E.T. Scholten, M. Oudkerk, P.A. de Jong, M. Prokop, B. van Ginneken, Automatic classification of pulmonary peri-fissural nodules in computed tomography using an ensemble of 2D views and a convolutional neural network out-of-the-box, Med. Image Anal. 26 (2015) 195–202, https://doi.org/10.1016/J.MEDIA.2015.08.001.

[15] M. Gao, U. Bagci, L. Lu, A. Wu, M. Buty, H.-C. Shin, H. Roth, G.Z. Papadakis, A. Depeursinge, R.M. Summers, Z. Xu, D.J. Mollura, Holistic classification of CT attenuation patterns for interstitial lung diseases via deep convolutional neural networks, Comput. Methods Biomech. Biomed. Eng. Imaging Vis. 6 (2018) 1–6, https://doi.org/10.1080/21681163.2015.1124249.

[16] J. Kleesiek, G. Urban, A. Hubert, D. Schwarz, K. Maier-Hein, M. Bendszus, A. Biller, Deep MRI brain extraction: a 3D convolutional neural network for skull stripping, Neuroimage 129 (2016) 460–469, https://doi.org/10.1016/J.NEUROIMAGE.2016.01.024.

[17] P. Moeskops, M.A. Viergever, A.M. Mendrik, L.S. de Vries, M.J.N.L. Benders, I. Isgum, Automatic segmentation of MR brain images with a convolutional neural network, IEEE Trans. Med. Imaging 35 (2016) 1252–1261, https://doi.org/10.1109/TMI.2016.2548501.

[18] S.M. Plis, D.R. Hjelm, R. Salakhutdinov, E.A. Allen, H.J. Bockholt, J.D. Long, H. J. Johnson, J.S. Paulsen, J.A. Turner, V.D. Calhoun, Deep learning for neuroimaging: a validation study, Front. Neurosci. 8 (2014) 229, https://doi.org/10.3389/fnins.2014.00229.

[19] H.-I. Suk, S.-W. Lee, D. Shen, Latent feature representation with stacked auto-encoder for AD/MCI diagnosis, Brain Struct. Funct. 220 (2015) 841–859, https://doi.org/10.1007/s00429-013-0687-3.

[20] A.I. Khan, J.L. Shah, M.M. Bhat, CoroNet: a deep neural network for detection and diagnosis of COVID-19 from chest x-ray images, Comput. Methods Programs Biomed. 196 (2020) 105581, https://doi.org/10.1016/J.CMPB.2020.105581.

[21] Y. Pathak, P.K. Shukla, A. Tiwari, S. Stalin, S. Singh, P.K. Shukla, Deep transfer learning based classification model for COVID-19 disease, IRBM (2020), https://doi.org/10.1016/J.IRBM.2020.05.003.

[22] T. Ozturk, M. Talo, E.A. Yildirim, U.B. Baloglu, O. Yildirim, U. Rajendra Acharya, Automated detection of COVID-19 cases using deep neural networks with X-ray images, Comput. Biol. Med. 121 (2020) 103792, https://doi.org/10.1016/J.COMPBIOMED.2020.103792.

[23] H. Panwar, P.K. Gupta, M.K. Siddiqui, R. Morales-Menendez, V. Singh, Application of deep learning for fast detection of COVID-19 in X-Rays using nCOVnet, Chaos Solitons Fractals 138 (2020) 109944, https://doi.org/10.1016/J.CHAOS.2020.109944.

[24] I.D. Apostolopoulos, T.A. Mpesiana, Covid-19: automatic detection from x-ray images utilizing transfer learning with convolutional neural networks, Phys. Eng. Sci. Med. (2020) 1.

[25] M. Rahimzadeh, A. Attar, S.M. Sakhaei, A fully automated deep learning-based network for detecting COVID-19 from a new and large lung CT scan dataset. MedRxiv. (2020) 2020.06.08.20121541. https://doi.org/10.1101/2020.06.08.20121541.

[26] N.S. Punn, S. Agarwal, Automated diagnosis of COVID-19 with limited posteroanterior chest X-ray images using fine-tuned deep neural networks (2020). http://arxiv.org/abs/2004.11676 (accessed August 30, 2020).

[27] A. Narin, C. Kaya, Z. Pamuk, Automatic detection of coronavirus disease (COVID-19) using X-ray images and deep convolutional neural networks (2020). http://arxiv.org/abs/2003.10849 (accessed August 30, 2020).

[28] L. Wang, A. Wong, COVID-Net: a tailored deep convolutional neural network design for detection of COVID-19 cases from chest X-ray images (2020). http://arxiv.org/abs/2003.09871 (accessed August 30, 2020).

[29] T. Smith, J. Guild, The C.I.E. colorimetric standards and their use, Trans. Opt. Soc. 33 (1931) 73–134, https://doi.org/10.1088/1475-4878/33/3/301.

[30] W. Mokrzycki, T. Maciej, Perceptual difference in L* a* b* color space as the base for object colour identfication, Image Process Commun. Challenges (2009) 403–412.

[31] J.P. Cohen, P. Morrison, L. Dao, K. Roth, T.Q. Duong, M. Ghassemi, COVID-19 image data collection: prospective predictions are the future, ArXiv 2006.11988. (2020). https://github.com/ieee8023/covid-chestxray-dataset.

[32] X. Wang, Y. Peng, L. Lu, Z. Lu, M. Bagheri, R.M. Summers, ChestX-ray8: hospital-scale chest X-Ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases, in: Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2017.

[33] G. Huang, Z. Liu, L. van der Maaten, K.Q. Weinberger, Densely Connected Convolutional Networks, in: IEEE Conf. Comput. Vis. Pattern Recognit., 2017.

[34] M. Tan, Q. V. Le, EfficientNet: rethinking model scaling for convolutional neural networks, in: 36th Int. Conf. Mach. Learn. ICML 2019. 2019-June (2019) 10691–10700.

[35] S. Hochreiter, J. Schmidhuber, Long short-term memory, Neural Comput. 9 (1997) 1735–1780, https://doi.org/10.1162/neco.1997.9.8.1735.

[36] T. Fawcett, An introduction to ROC analysis, Pattern Recognit. Lett. 27 (2006) 861–874, https://doi.org/10.1016/J.PATREC.2005.10.010.

[37] C. Ieracitano, A. Paviglianiti, M. Campolo, A. Hussain, E. Pasero, F.C. Morabito, A novel automatic classification system based on hybrid unsupervised and supervised machine learning for electrospun nanofibers, IEEE/CAA J. Autom. Sin. 8 (2021) 64–76, https://doi.org/10.1109/JAS.2020.1003387.

[38] C. Pati, A.K. Panda, A.K. Tripathy, S.K. Pradhan, S. Patnaik, A novel hybrid machine learning approach for change detection in remote sensing images, Eng. Sci. Technol. Int. J. 23 (2020) 973–981, https://doi.org/10.1016/j.jestch.2020.01.002.