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 71 (2022) 103216

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

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

CNN-based bi-directional and directional long-short term memory network for determination of face mask

Murat Koklu [a,*], Ilkay Cinar [a], Yavuz Selim Taspinar [b]

[a] Department of Computer Engineering, Selcuk University, Konya, Turkey
[b] Doganhisar Vocational School, Selcuk University, Konya, Turkey

ARTICLE INFO

Keywords:
AlexNet
BiLSTM
Convolutional neural network
LSTM
Transfer learning
VGG16

ABSTRACT

Context: The COVID-19 virus, exactly like in numerous other diseases, can be contaminated from person to person by inhalation. In order to prevent the spread of this virus, which led to a pandemic around the world, a series of rules have been set by governments that people must follow. The obligation to use face masks, especially in public spaces, is one of these rules.
Objective: The aim of this study is to determine whether people are wearing the face mask correctly by using deep learning methods.
Methods: A dataset consisting of 2000 images was created. In the dataset, images of a person from three different angles were collected in four classes, which are "masked", "non-masked", "masked but nose open", and "masked but under the chin". Using this data, new models are proposed by transferring the learning through AlexNet and VGG16, which are the Convolutional Neural network architectures. Classification layers of these models were removed and, Long-Short Term Memory and Bi-directional Long-Short Term Memory architectures were added instead.
Result and conclusions: Although there are four different classes to determine whether the face masks are used correctly, in the six models proposed, high success rates have been achieved. Among all models, the TrVGG16 + BiLSTM model has achieved the highest classification accuracy with 95.67%.
Significance: The study has proven that it can take advantage of the proposed models in conjunction with transfer learning to ensure the proper and effective use of the face mask, considering the benefit of society.

1. Introduction

COVID-19 is a human-to-human respiratory disease caused by a competing agent through aerosol and droplet transmission. It has vital importance to bring under the control of this disease's spread. It is stated that controlling or preventing the spread of COVID-19 can be possible by observing the social distance and mask rules [1–3]. The World Health Organization remarks that people's wearing face masks as a basic non-pharmaceutical intervention (NPI) measure, when they have respiratory symptoms or interact with the ones with symptoms, can be used as an effective means of preventing respiratory infectious diseases [4–6]. It is also stated that even if face masks have a limited protective effect in terms of protecting people from large respiratory droplets, it may have a great impact on reducing the number of infected people and COVID-19-related deaths in society [7–9].

Many countries have enacted laws to ensure that face masks are worn in common areas and public spaces. Despite that, some individuals still refuse to wear face masks, while others do not wear them properly. It is known that the protection of even the most effective mask will disappear if it is not worn correctly [2,6,10]. Especially in public spaces, it is quite difficult to manually detect those who wear the face mask incorrectly or do not wear it at all, only by human labor. On the other hand, thanks to deep learning, it can be much easier to detect people who wear the face mask incorrectly or not at all by automatically processing visual data.

Recently, thanks to deep learning, the performance of computer vision applications has been upgraded. In the field of deep learning, convolutional neural networks (CNN) in particular have achieved remarkable success in image classification [11–13]. CNN has gain popularity in various fields such as image-based content retrieval, tracking of various objects, license plate recognition for vehicles, and autonomous vehicle applications [14,15].

To be used in all areas where it is mandatory to wear a face mask,

* Corresponding author at: Department of Computer Engineering, Faculty of Technology, Selcuk University, 42031 Konya, Turkey.
E-mail addresses: mkoklu@selcuk.edu.tr (M. Koklu), ilkay.cinar@selcuk.edu.tr (I. Cinar), ytaspinar@selcuk.edu.tr (Y.S. Taspinar).

https://doi.org/10.1016/j.bspc.2021.103216
Received 1 May 2021; Received in revised form 30 August 2021; Accepted 30 September 2021
Available online 9 October 2021

especially in public areas, with this article, a study has been carried out to detect those who wear the face mask incorrectly/undesirably and who do not wear a face mask.

The contributions of this article are as follows;

A dataset was created in order to be used in the study and to contribute to the literature.
In the dataset, there are a total of 2000 images in four classes, as "masked", "unmasked", "masked but nose open, and "masked but under the chin".
Transfer learning has been applied to the dataset of the study through CNN-based VGG16 and AlexNet architectures in order to determine how people use the face masks. In addition, classification processes have been carried out by using these architectures together with Long short-term memory (LSTM) and Bidirectional long short-term memory (BiLSTM) architectures.
The comparative results for the methods used after the classification process were shared.

2. The remaining parts of this article are structured as follows;

In the second chapter, studies on face mask detection by deep learning and machine learning is included. In the third chapter, information about the dataset and the proposed methodology are presented. In the fourth chapter, experimental results are given and the last chapter includes the results of the study.

3. Related works

Many studies in the literature have revealed the benefits of mask usage to prevent the spread of the COVID-19 virus and other respiratory viruses. There are many facial recognition technics designed based on deep learning and machine learning to detect mask-wearing status. Studies based on face mask detection are given below.

In their studies, Dey et al. [16] proposed MobileNetMask as a deep learning-based face mask detection model. The testing and training procedures were carried out by using two different datasets with two classes, with more than 5200 images. As a result of the test, which uses 770 verification samples, a classification accuracy was obtained approximately 93%.

Using the YOLOv3 architecture, Bhuiyan et al. [17] YOLOv3 developed a study to determine whether the individuals are wearing a mask or not. A classification accuracy of 96% was achieved at the end of the study.

Loey et al. [18] proposed a hybrid model to detect the status of mask-wearing through deep learning and machine learning methods. The feature extraction was carried out via ResNet50 which is one of the deep learning architectures. Decision Tree (DT), Support Vector Machine (SVM) and Ensemble methods were used to perform the classification process. During the training and testing stages, the real face masks, fake face masks and the datasets, which was created by combining them, were used. As a result of the classification, accuracy values varying between 92% and 98% were obtained.

Islam et al. [19] used CNN to identify whether the people are wearing a mask or not. In the study, it was detected whether the person is wearing a mask or not by instant monitoring with the camera and the official was informed. As a result of the study, the developed model provided 98% classification accuracy.

Mohan et al. [20] proposed a model for detecting the mask-wearing status. The results obtained from the proposed model were compared with the results obtained from the SqueezeNet architecture. A classification accuracy of 98.53% was achieved with SqueezeNet while the proposed model showed 99.83% accuracy in detecting the face mask.

Rahman et al. [1] used the deep learning algorithm to detect the mask on the person's face. The study was carried out through a dataset that consists of 1539 images and two classes, as masked and unmasked. As result of the study, 98.7% classification accuracy was obtained.

In their studies, Razavi et al. [21] developed an automated system, using the R-CNN Inception ResNet V2 architecture, to detect the physical distances of construction workers and whether they are wearing face masks. Pixel values taken from images with Euclidean distance have been converted into real distance. They set a limit value of 6 m to detect physical distance violations. At the end of the study, it was stated that physical distance violations of construction workers were successfully detected. Also, the classification accuracy achieved in the detection of workers' face masks was reported as 99.8%.

Using the ResNet deep learning architecture, Basha et al. [22] worked on detecting whether the person is wearing a mask or not. As a result of the study, an accuracy of 97%, the highest classification accuracy in the detection of face masks, was achieved.

With the aim of detecting face masks, Oumina et al. [4] carried out classification processes with SVM and K-NN machine learning methods by extracting features via VGG19, Xception and MobileNetV2, which are among the deep learning architectures. The highest classification accuracy, 97.1%, was achieved with MobilNetV2-SVM model.

Loey et al. [23], in their studies, carried out masked face detection with YoloV2 and performed classification processes by extracting image features with ResNet-50, which is one of the deep learning architectures. With the proposed model, they obtained an average precision value of 81%.

Yadav [24] proposed a real-time computer vision system for face mask and social distance detection. MobileNet V2 architecture was utilized to analyse the video stream. The proposed model was designed to run on Raspberry pi 4 and as a result, classification accuracy was achieved in the range of 85% to 95%.

Militante and Dionisio [13] developed a CNN model to determine the physical distance between people and to detect whether they are wearing a face mask. In the model, Tensorflow and Keras modules were used together with VGG-16 architecture. In the study, which obtained a classification accuracy of more than 97% in face mask detection, the physical distance of the people was also successfully determined.

Pagare et al. [25] developed an application to detect whether the person is wearing a face mask or not, and their suitability for social distance. Within the scope of the study, MobileNet architecture was used in face mask detection while YOLO object detection algorithm was used in social distance detection. It was stated that the study became successful in both detecting face masks and determining social distance compliance.

Sanjaya and Rakhmawan [26] used CNN-based MobileNetV2 for face mask detection. In the study, the highest classification accuracy, 96.85%, was obtained on a dataset with two classes, as masked and unmasked.

In order to automate the recognition of people who do not wear masks, Chowdary et al. [27], proposed a transfer learning model. InceptionV3 architecture is used for the proposed model. In the study using a simulated dataset, 100% classification accuracy was achieved.

Sandesara et al. [28] proposed a CNN-based model for face mask detection. In the proposed model, 95% classification accuracy was achieved in determining whether people are wearing a mask or not.

Chavda et al. [29] presented a deep learning-based model to detect people who do not wear face masks properly. In their study, where NasNet-Mobile, DenseNet-121 and MobileNetV2 architectures were used, the highest classification accuracy has been achieved via DenseNet-121 architecture, with 99.49%.

In the study conducted by Said [12], a face mask detection system, which is using YOLO object detection algorithm and CNN, has been proposed. It was stated that a classification accuracy of 97% was achieved in the tests performed on the dataset used with the proposed system.

Aiming at detecting multi-scale face masks in real-time, Addagarla et al. [15] proposed two different models. YOLOv3, NasNetMobile and ResNet-SSD300 algorithms were used in the proposed models. As a

result, the recall rates were reported as 98% and 99% for both models.

Militante and Dionisio [30] used deep learning techniques for facial recognition besides predicting whether the person is wearing a mask or not. Moreover, they have developed a Raspberry Pi-based system that will alarm in case a person without a mask are detected. It was stated that a classification accuracy of 96% was achieved in face mask detection.

In the study by Vijitkunsawat and Chantngarm [31], MobileNet deep learning architecture and K-NN and SVM machine learning algorithms were used to determine whether people are wearing face masks. As a result of the study, it was reported that the highest classification accuracy, 94.2%, was achieved via CNN-based MobileNet.

A comparative summary of the studies found in the literature based on face mask detection is given in Table 1.

When the studies in the literature are examined, it can be seen that a major part of the datasets used in face mask detection consists of collected or simulated images. In addition, it is seen that two-class datasets, "masked" and "non-masked", have been used.

4. Material and methods

Deep learning-based artificial intelligence applications, recently, are obtaining significant success in computer vision. CNN, which is a deep learning method, is highly preferred in different discipline applications at present since it can easily distinguish small details that the human eye cannot notice in image recognition applications. The fact that they do not require much preprocessing and recognize visual patterns directly from pixel images is the most important feature of CNNs [32–35].

CNN architectures AlexNet and VGG16 were used within the scope of the study. In addition to these architectures, models were created using LSTM and BiLSTM. Training and testing processes of these models were performed through the MATLAB application.

Table 1
Comparative summary of the studies found in the literature based on face mask detection.

Model	The Number of classes	The Number of images	Classification Accuracy (%)	References
MobileNetMask	2	3835 1376	93 100	[16]
YOLOv3	2	600	96	[17]
ResNet50 + SVM	2	15,000 1570 13,000	between 92 and 98	[18]
CNN	2	1376	98	[19]
CNN	2	135,849	99.83	[20]
CNN	2	1539	98.7	[1]
R-CNN Inception ResNet V2	2	1853	99.8	[21]
ResNet	2	95,000	97	[22]
MobilNetV2 + SVM	2	1376	97.1	[4]
YoloV2 + ResNet-50	2	1415	81 (Precision)	[23]
MobileNet V2	2	3165	between 85 and 95	[24]
VGG-16	2	20,000	97	[13]
MobileNetV2	2	3846	96.85	[26]
InceptionV3	2	1570	100	[27]
Yolo + CNN	2	–	95	[28]
DenseNet-121	2	7855	99.49	[29]
CNN	2	95,000 24,771 500,000 65,617	97	[12]
YoloV3	3	680	98	[15]
NASNetMobile	2	1400	99	
VGG-16	2	25,000	96	[30]
MobileNet	2	3216	94.2	[31]

4.1. AlexNet

AlexNet, which is a version of the traditional LeNet, was presented in the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) in 2012 by Krizhevsky et al. [36], and showed the best performance in this challenge. It utilizes large-scale ImageNet [37] training dataset and GPUs that provide approximately 10 times acceleration in terms of computing power [38].

AlexNet has five convolutional layers, three fully connected layers, and a Softmax output layer. Following each convolutional layer, there is the Straightened Linear Unit (ReLU) activation function [39]. Each convolution layer has maximum pooling to reduce network size. After the convolutional layer at the end of the network, there are two fully connected layers with 4096 outputs. Finally, there is one more fully connected layer to classify input data. This last layer classifies 1000 objects with using the Softmax function [40].

4.2. Vgg16

VGG was presented by Simonyan and Zisserman [41] in the ILSVRC 2014 challenge and came in second. Small filters (3×3) existing in each convolutional layer of VGG show improved performance. This is due to the fact that multiple small filters in order can mimic the effects of larger ones. The simplicity of using small-size filters throughout the network results in very high performance of generalization. Therefore, today, multiple versions of VGG are still widely available since it has simplicity and high generalization performance.

VGG16, which is one of the most popular versions [38,42], contains thirteen convolutional layers, three fully connected layers and a Softmax output layer. After each convolutional layer, there is the ReLU activation function. Each convolution layer has maximum pooling to reduce mesh size. After the convolutional layer at the end of the network, there are two fully connected layers with 4096 outputs. Finally, there is one more fully connected layer to classify the input data. This last layer classifies 1000 objects with using the Softmax function [43–46].

4.3. Long Short-Term memory (LSTM)

LSTM, which is a kind of artificial neural network, has a different structure compared to the traditional neural networks. It is a special type of Recurrent Neural Networks (RNN) [47] that can learn long-term relationships between data. LSTMs were developed to prevent the long-term dependency problem experienced in repetitive neural networks. RNNs are in the form of a chain of repeating modules of a neural network. This module, which is repeated in standard RNNs, has a simple structure just like the tanh activation layer. The structure of this duplicate module in LSTMs is different. LSTMs remember information for a long time and perform this process through learning. LSTM architecture uses hidden units, called memory cells, in cases where long-term dependencies are required. It stores inputs that need to be remembered in a long time interval in these memory units. It decides whether this information is important, via the doors it hosts in its architecture [48–50]. Fig. 1 shows the example LSTM architecture.

RNN and LSTM, having the same network structures, has only one node with different content. Different from input and output connections, an LSTM block contains three different gates that are input gate, output gate and forget gate. Data from the gates pass through a specific activation function (tanh or sigmoid). Subsequent values go through certain operations (multiplication, addition, etc.) with the input content and exit the node as output [49,50]. In Fig. 2, LSTM cell structure is illustrated. In Fig. 2, i represents the input gate, C cell state, O output gate, f forget gate.

As a consequence, an LSTM structure is the case that an RNN cell works together with memory. With this memory, information from the previous time is received and transmitted to the next time. It is decided by training which information will or will not be evaluated. LSTM is

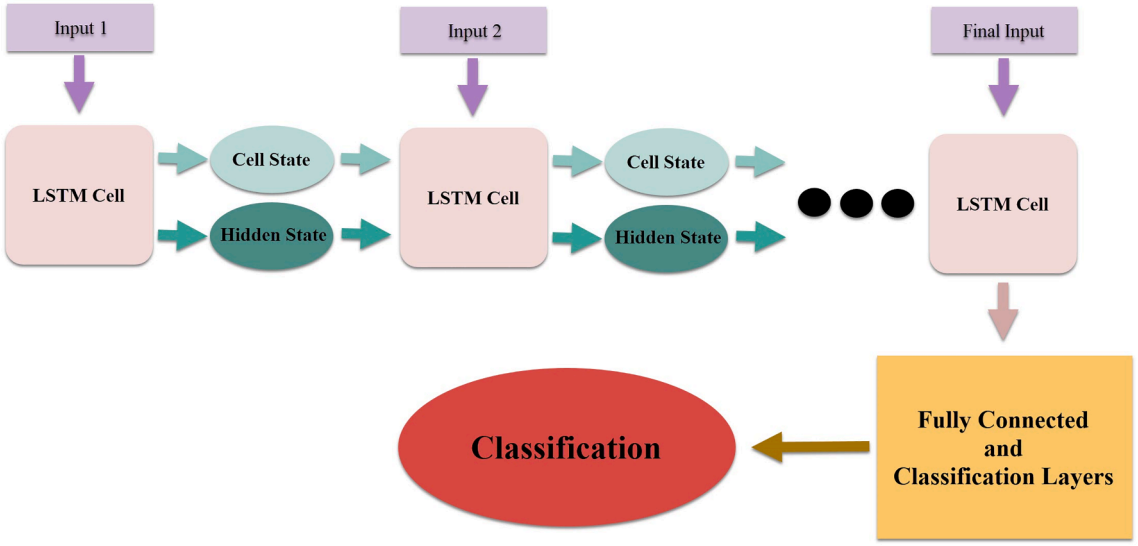


Fig. 1. LSTM architecture.

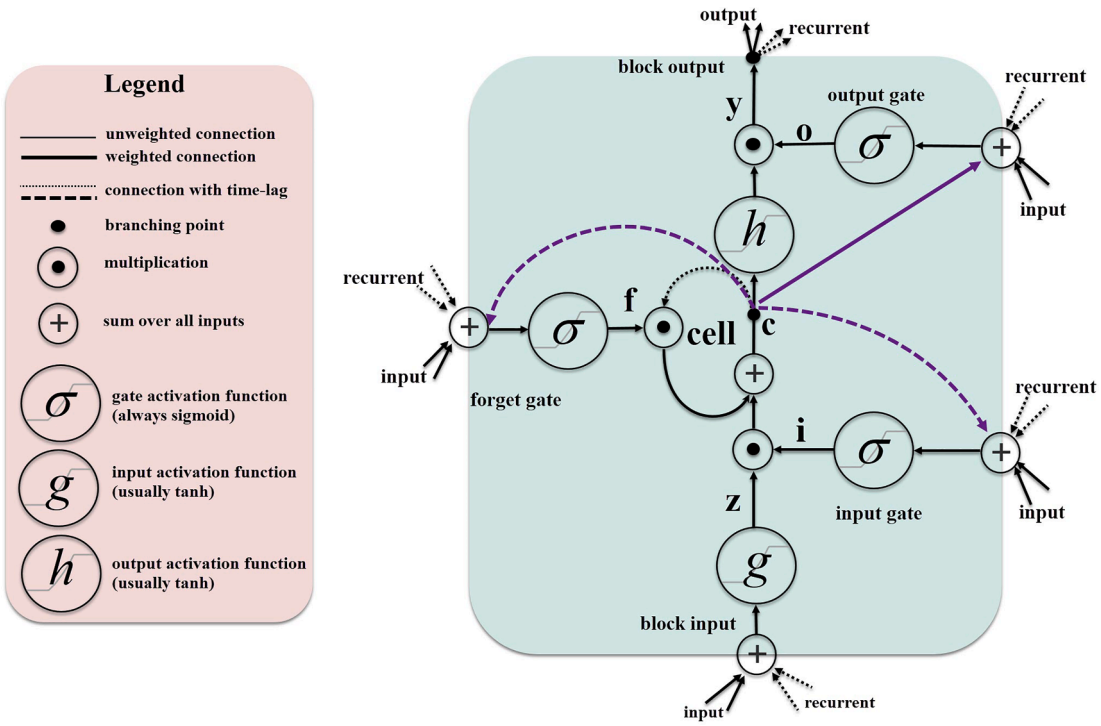


Fig. 2. LSTM cell structure.

frequently used in various data fields and artificial intelligence applications, such as analysis of time series data, speech recognition, natural language processing, financial analysis, language translation, text compression, and etc. [51,52].

4.4. Bi-directional long Short-Term memory (BiLSTM)

BiLSTM, is the bi-directional RNN modified with the LSTM architecture. Bi-directional RNN consists of two independent RNN layers that process input signals back and forth, respectively [53]. In Fig. 3, the bi-directional RNN structure is illustrated. x represents the inputs, y represents the outputs, t represents the state time, h represents the hidden state. Furthermore, t-1 refers to previous, and $t + 1$ refers to next.

RNNs, which have a simple neural network structure, usually show poor performance caused by the disappearing gradient problem [53]. In order to overcome this problem, LSTM, which is a special RNN cell, is proposed [54]. The hidden state in a one-way forward LSTM captures only previous features and does not consider the future. Combining the Bi-directional RNN with the LSTM ensures that previous and future features are used effectively. Dissimilar to the LSTM network, network has two parallel layers in both propagation directions. Previous features are extracted by a forward LSTM layer while future features are captured by a backward LSTM layer [55,56]. In Fig. 4, BiLSTM sample architecture is given in 3 consecutive steps.

The main idea in BILSTMs is to provide two separate hidden layers, forward and backward, to capture past and future information respectively. Following that, the two hidden states are combined to create the final output [57].

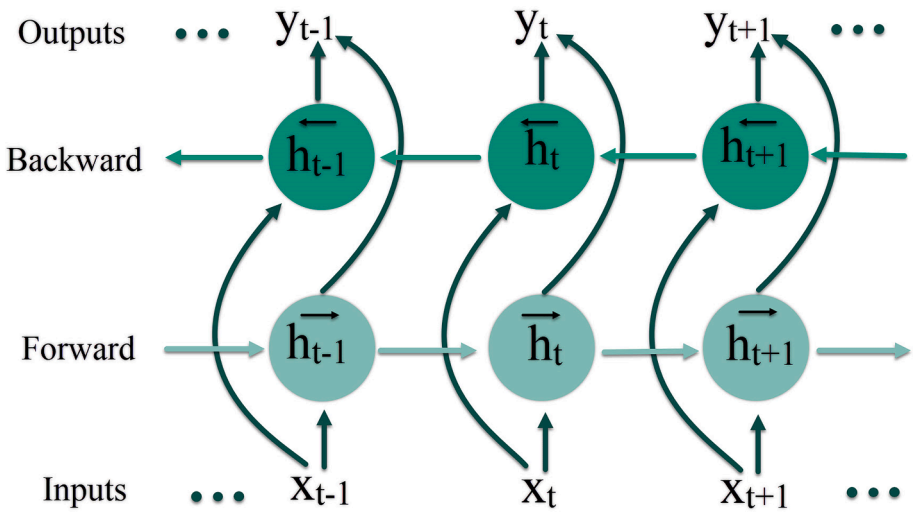


Fig. 3. Bi-directional RNN structure.

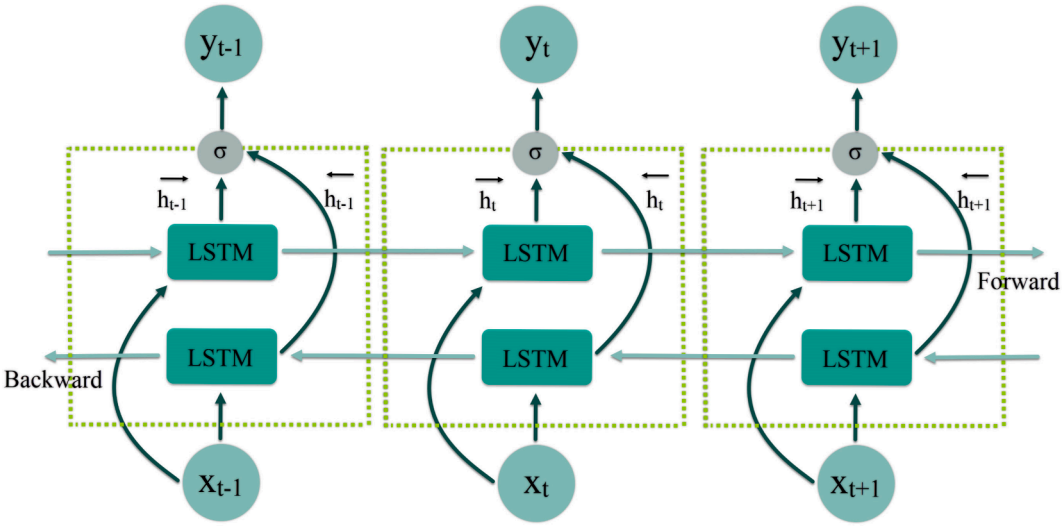


Fig. 4. The example of BiLSTM architecture in 3 consecutive steps.

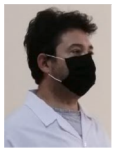
4.5. Dataset description

The dataset used in the study consists of four different classes, and there are 2000 images in total, 500 for each class. Images of people from 3 different angles were collected in the dataset. The necessary permissions were obtained from the people whose images were collected and filed. The detailed information on the dataset is given in Table 2.

4.6. The proposed models

This study proposes six different models using the architectures of AlexNet and VGG16 that are the convolutional neural networks. In Fig. 5, the block diagram of the proposed models is given. Each proposed model will be detailed under the following headings.

Table 2
Dataset features.



4.6.1. TrAlexNet

In this proposed model, the images in the face mask dataset, were extracted by using modified AlexNet architecture at first, and then the classification process was carried out. The last three layers of the AlexNet model have been removed and modified in order to properly classify the face mask dataset. The classification process was carried out with the four outputs taken from the fully connected fc_optimized layer. Fc_optimized layer is the layer where the output size in the fully connected layer is set to the output size in the study due to the use of the transfer learning method in the study. In Fig. 6, TrAlexNet architecture is given.

The parameters of the remaining layers of the AlexNet model have been preserved, except for the last three layers. The layers and their parameters used in the model are given in Table 3.

4.6.2. TrAlexNet + LSTM

In this proposed model, the images in the face mask dataset, were extracted by using modified AlexNet architecture. The features taken from the last fully connected layer of AlexNet, fc8, were transferred via the flatten layer, to the LSTM layer which was added to the network later. The classification process was carried out with the four outputs

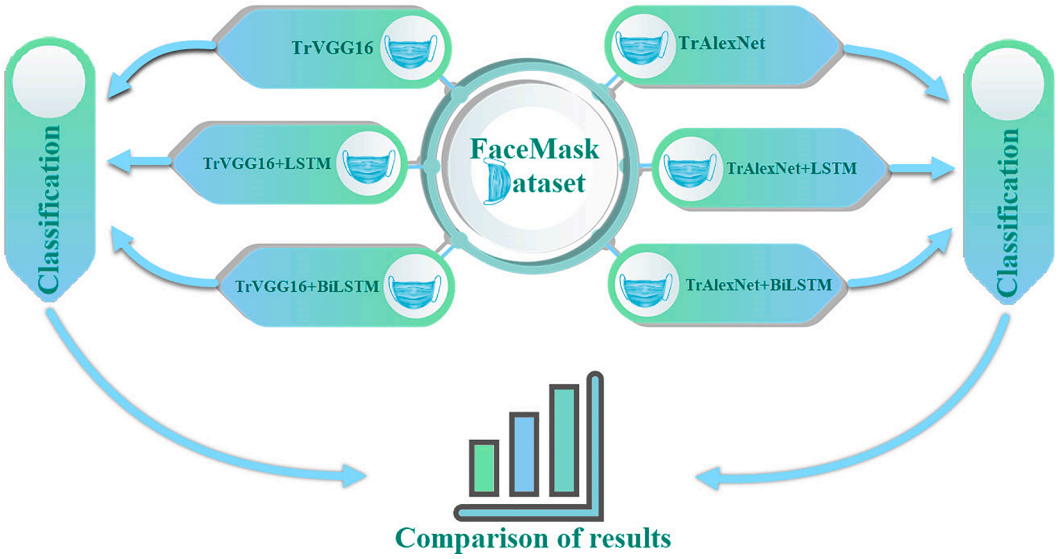


Fig. 5. Block diagram of the proposed models.

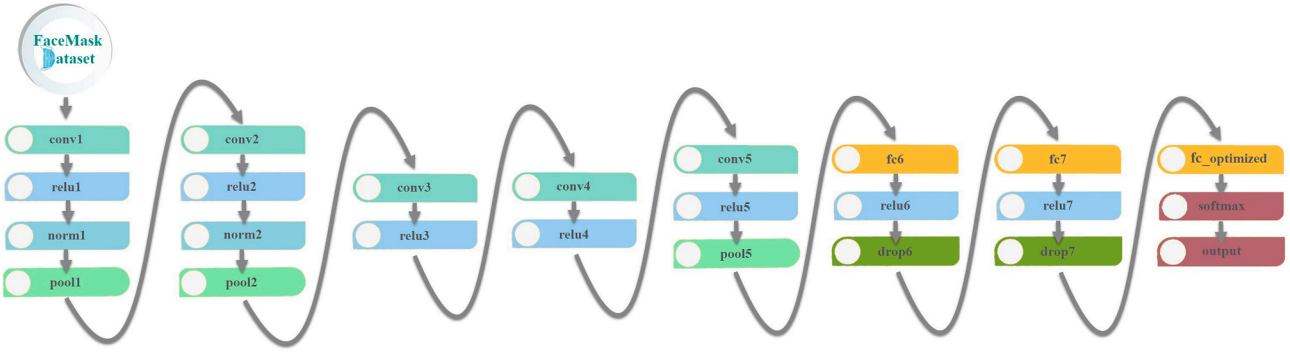


Fig. 6. TrAlexNet architecture.

Table 3
Layers and parameters of the proposed TrAlexNet.

Layer Name	Layer Type	Filter Size	Stride	Padding	Output Channel	Activation Function
conv1	convolution 2d	11,11	4,4	0	96	relu
pool1	max pooling 2d	3,3	2,2	0	96	–
conv2	grouped convolution 2d	5,5	1,1	2	256	relu
pool2	max pooling 2d	3,3	2,2	0	256	–
conv3	convolution 2d	3,3	1,1	1	384	relu
conv4	grouped convolution 2d	3,3	1,1	1	384	relu
conv5	grouped convolution 2d	3,3	1,1	1	256	relu
pool5	max pooling 2d	3,3	2,2	0	256	–
fc6	fully connected	–	–	–	4096	relu
fc7	fully connected	–	–	–	4096	relu
fc_optimized	fully connected	–	–	–	4	softmax

taken from the fully connected fc_optimized layer. In Fig. 7, TrAlexNet + LSTM architecture is given.

In this model, the parameters of the AlexNet model's remaining layers have been preserved, except for the layers added after the last connected layer, fc8. The layers and their parameters used in the model are given in Table 4.

4.6.3. *TrAlexNet + BiLSTM*

In this proposed model, the images in the face mask dataset were extracted by using a modified AlexNet architecture. The features taken from the fc8 layer, the last fully connected layer of AlexNet, were transferred to the BiLSTM layer via the flatten layer added to the network later. The classification process was carried out with the four outputs taken from the fully connected fc_optimized layer. In Fig. 8, TrAlexNet + BiLSTM architecture is given.

In this model, the parameters of the remaining layers of the AlexNet model have been preserved, except for the layers added after the last connected layer, fc8. The layers and their parameters used in the model are given in Table 5.

4.6.4. *TrVGG16*

In the proposed model, firstly, the images in the face mask dataset were extracted by using the modified VGG16 architecture, and then classification process was performed. The last three layers of the VGG16

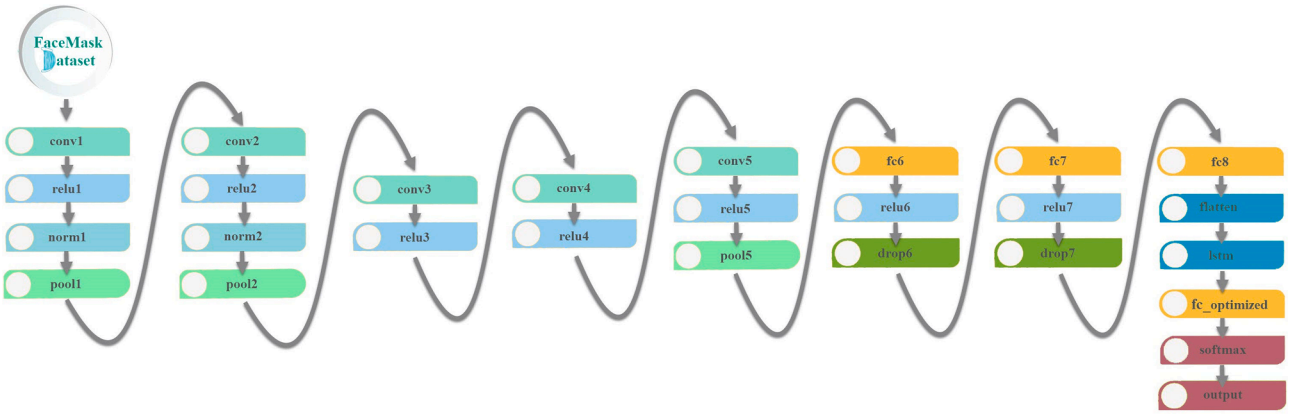


Fig. 7. TrAlexNet + LSTM architecture.

Table 4
Layers and parameters of the proposed TrAlexNet + LSTM.

Layer Name	Layer Type	Filter Size	Stride	Padding	Output Channel	Activation Function
conv1	convolution 2d	11,11	4,4	0	96	relu
pool1	max pooling 2d	3,3	2,2	0	96	–
conv2	grouped convolution 2d	5,5	1,1	2	256	relu
pool2	max pooling 2d	3,3	2,2	0	256	–
conv3	convolution 2d	3,3	1,1	1	384	relu
conv4	grouped convolution 2d	3,3	1,1	1	384	relu
conv5	grouped convolution 2d	3,3	1,1	1	256	relu
pool5	max pooling 2d	3,3	2,2	0	256	–
fc6	fully connected	–	–	–	4096	relu
fc7	fully connected	–	–	–	4096	relu
fc8	fully connected	–	–	–	1000	relu
flatten	flatten	–	–	–	–	–
lstm	lstm	–	–	–	–	–
fc_optimized	fully connected	–	–	–	4	softmax

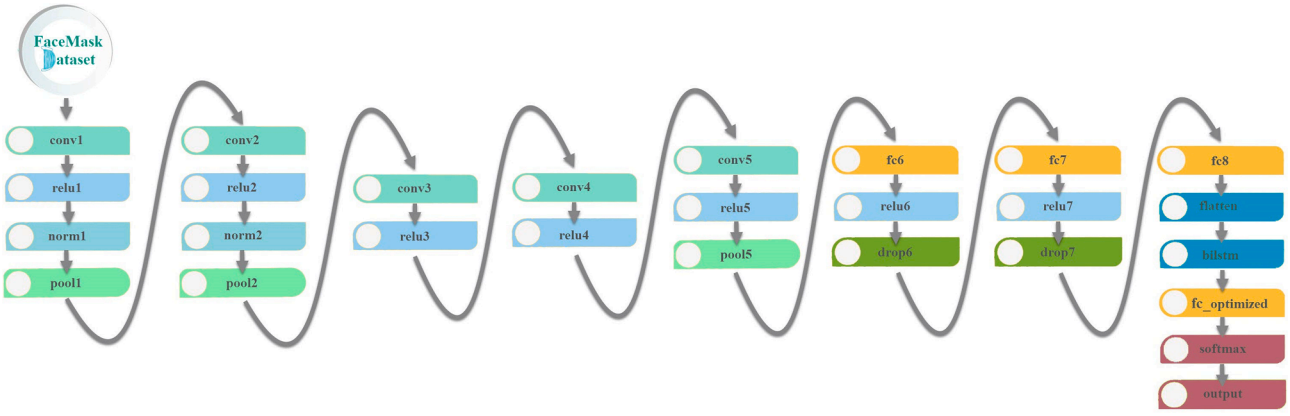


Fig. 8. TrAlexNet + BiLSTM architecture.

model were removed and modified in order to classify the face mask dataset appropriately. The classification process was carried out with the four outputs taken from the fully connected fc_optimized layer. Fc_optimized layer is the layer where the output size in the fully connected layer is set to the output size in the study due to the use of the transfer learning method in the study. TrVGG16 architecture is given in Fig. 9.

The parameters of the VGG16 model's remaining layers have been preserved, VGG16 model, except for the last three layers. The layers and their parameters used in the model are given in Table 6.

4.6.5. TrVGG16 + LSTM

In this proposed model, the images in the face mask dataset were extracted by using the modified VGG16 architecture. The features taken from the fc8 layer, the last fully connected layer of AlexNet, were transferred to the LSTM layer, which was added to the network later, through the flatten layer. The classification process was carried out with the four outputs taken from the fully connected fc_optimized layer. In Fig. 10, the architecture of TrVGG16 + LSTM is given.

In this model, the parameters of the remaining layers of the VGG16 model have been preserved, except for the layers added after the last connected layer, fc8. The layers and their parameters used in the model are given in Table 7.

Table 5
Layers and parameters of the proposed TrAlexNet + BiLSTM.

Layer Name	Layer Type	Filter Size	Stride	Padding	Output Channel	Activation Function
conv1	convolution 2d	11,11	4,4	0	96	relu
pool1	max pooling 2d	3,3	2,2	0	96	–
conv2	grouped convolution 2d	5,5	1,1	2	256	relu
pool2	max pooling 2d	3,3	2,2	0	256	–
conv3	convolution 2d	3,3	1,1	1	384	relu
conv4	grouped convolution 2d	3,3	1,1	1	384	relu
conv5	grouped convolution 2d	3,3	1,1	1	256	relu
pool5	max pooling 2d	3,3	2,2	0	256	–
fc6	fully connected	–	–	–	4096	relu
fc7	fully connected	–	–	–	4096	relu
fc8	fully connected	–	–	–	1000	relu
flatten	flatten	–	–	–	–	–
bilstm	bilstm	–	–	–	–	–
fc_optimized	fully connected	–	–	–	4	softmax

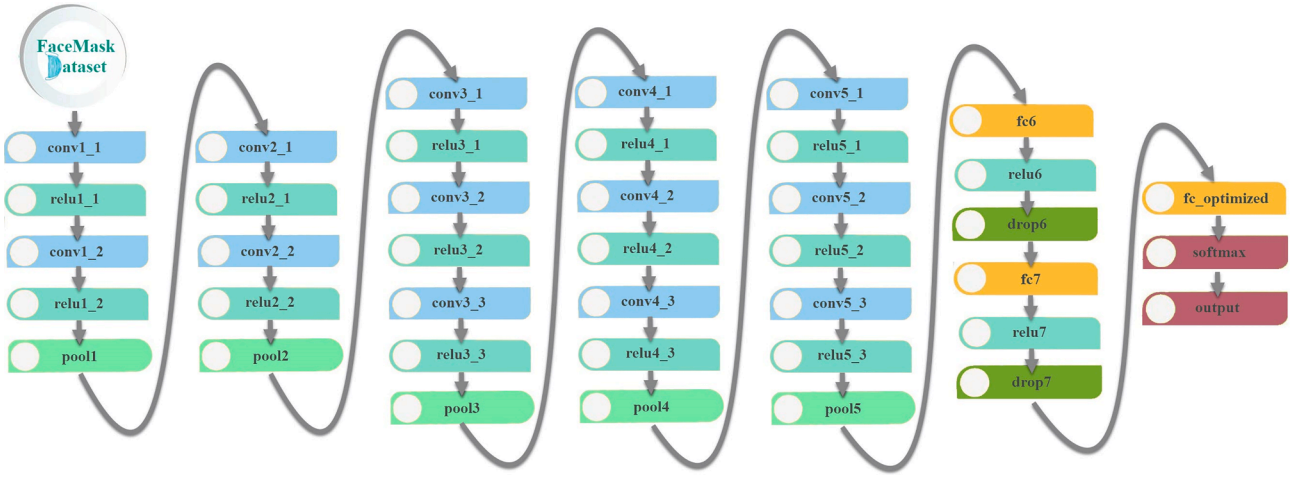


Fig. 9. TrVGG16 architecture.

Table 6
Layers and parameters of the proposed TrVGG16.

Layer Name	Layer Type	Filter Size	Stride	Padding	Output Channel	Activation Function
conv1_1	convolution 2d	3,3	1,1	1	64	relu
conv1_2	convolution 2d	3,3	1,1	1	64	relu
pool1	max pooling 2d	2,2	2,2	0	64	–
conv2_1	convolution 2d	3,3	1,1	1	128	relu
conv2_2	convolution 2d	3,3	1,1	1	128	relu
pool2	max pooling 2d	2,2	2,2	0	256	–
conv3_1	convolution 2d	3,3	1,1	1	256	relu
conv3_2	convolution 2d	3,3	1,1	1	256	relu
conv3_3	convolution 2d	3,3	1,1	1	256	relu
pool3	max pooling 2d	2,2	2,2	0	256	–
conv4_1	convolution 2d	3,3	1,1	1	512	relu
conv4_2	convolution 2d	3,3	1,1	1	512	relu
conv4_3	convolution 2d	3,3	1,1	1	512	relu
pool4	max pooling 2d	2,2	2,2	0	512	–
conv5_1	convolution 2d	3,3	1,1	1	512	relu
conv5_2	convolution 2d	3,3	1,1	1	512	relu
conv5_3	convolution 2d	3,3	1,1	1	512	relu
pool5	max pooling 2d	2,2	2,2	0	512	–
fc6	fully connected	–	–	–	4096	relu
fc7	fully connected	–	–	–	4096	relu
fc_optimized	fully connected	–	–	–	4	softmax

4.6.6. TrVGG16 + BiLSTM

In this proposed model, the features of the images in the face mask dataset are extracted by using the modified VGG16 architecture. Features obtained from the fc8 layer, the last fully connected layer of VGG16, were transferred to the BiLSTM layer via the flatten layer, which was added to the network later. The classification process is carried out with the four outputs taken from the fully connected fc_optimized layer. In Fig. 11, the architecture of TrVGG16 + BiLSTM is given.

In this model, the parameters of the remaining layers of the VGG16 model have been preserved, except for the layers added after the last

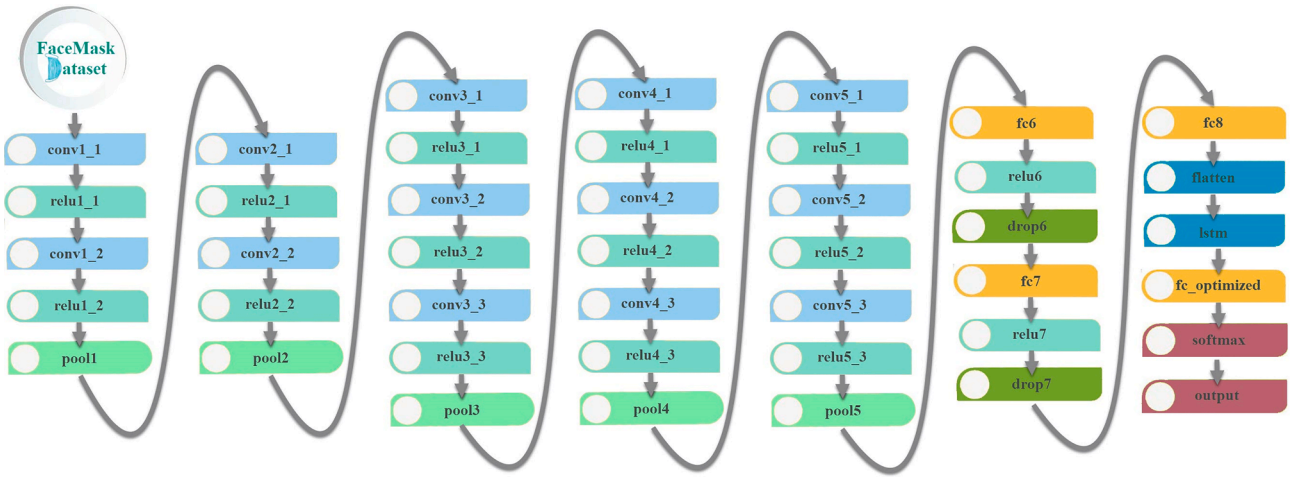


Fig. 10. TrVGG16 + LSTM architecture.

Table 7
Layers and parameters of the proposed TrVGG16 + LSTM.

Layer Name	Layer Type	Filter Size	Stride	Padding	Output Channel	Activation Function
conv1_1	convolution 2d	3,3	1,1	1	64	relu
conv1_2	convolution 2d	3,3	1,1	1	64	relu
pool1	max pooling 2d	2,2	2,2	0	64	–
conv2_1	convolution 2d	3,3	1,1	1	128	relu
conv2_2	convolution 2d	3,3	1,1	1	128	relu
pool2	max pooling 2d	2,2	2,2	0	256	–
conv3_1	convolution 2d	3,3	1,1	1	256	relu
conv3_2	convolution 2d	3,3	1,1	1	256	relu
conv3_3	convolution 2d	3,3	1,1	1	256	relu
pool3	max pooling 2d	2,2	2,2	0	256	–
conv4_1	convolution 2d	3,3	1,1	1	512	relu
conv4_2	convolution 2d	3,3	1,1	1	512	relu
conv4_3	convolution 2d	3,3	1,1	1	512	relu
pool4	max pooling 2d	2,2	2,2	0	512	–
conv5_1	convolution 2d	3,3	1,1	1	512	relu
conv5_2	convolution 2d	3,3	1,1	1	512	relu
conv5_3	convolution 2d	3,3	1,1	1	512	relu
pool5	max pooling 2d	2,2	2,2	0	512	–
fc6	fully connected	–	–	–	4096	relu
fc7	fully connected	–	–	–	4096	relu
fc8	fully connected	–	–	–	1000	relu
flatten	flatten	–	–	–	–	–
lstm	lstm	–	–	–	–	–
fc_optimized	fully connected	–	–	–	4	softmax

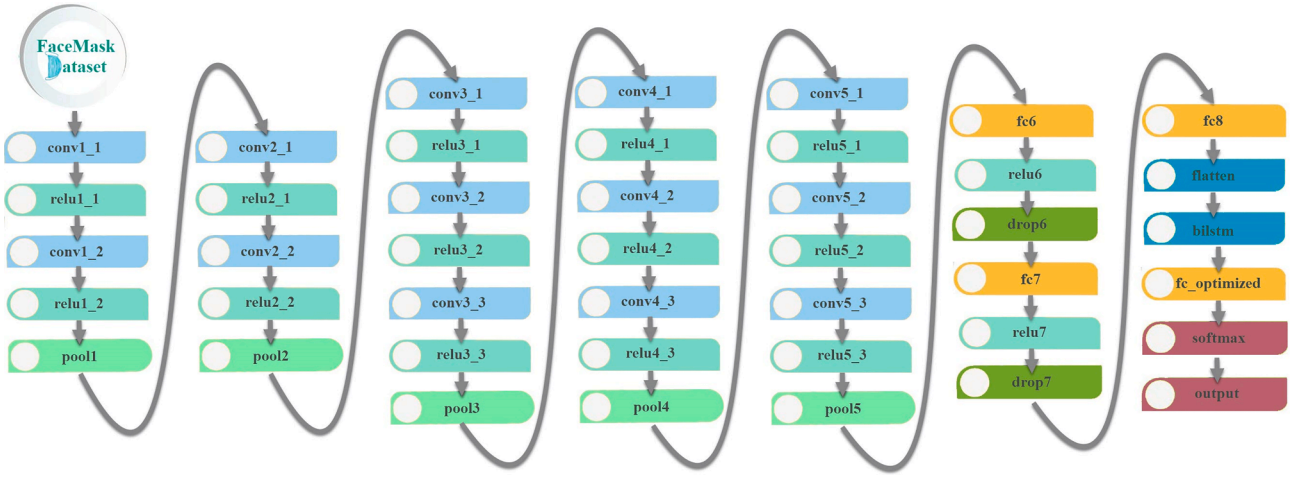


Fig. 11. TrVGG16 + BiLSTM architecture.

connected layer, fc8. The layers and their parameters used in the model are given in Table 8.

4.6.7. Experimental setup

All models proposed in the study were trained by using 8 GB memory, 2.20 GHz Intel I7-8750H processor and NVIDIA GTX 1050 Ti graphics processing unit (GPU). Table 9 shows the training parameters of network, common to all proposed models. In addition to that, Fc_optimized layer parameters are given in Table 10, and LSTM and BiLSTM layer parameters are given in Table 11.

5. Experimental results

Using the MATLAB application, suggested models were created and the results were calculated. Fig. 12 shows training and verification charts of all created models.

According to Fig. 12, validation curves in graphs for TrVGG16 models increased faster in the 0–100 iteration range than validation curves for TrAlexNet models. Furthermore, in parallel with this result, the loss curve also decreased more rapidly. This may indicate that higher success will be achieved in subsequent iterations than in the model. The validation curves of TrVGG16 models are more stable in the 800–1000 iteration range than the validation curves of TrAlexNet models. This can also be interpreted as an indication that the model is well trained. The Train and Validation curves are similar in all models. This indicates that the models created do not fall into underfitting and overfitting states.

The training set was determined as 80% and the validation set determined as 20%, for all models used in the study. In addition, the number of epochs was determined as 8 and the number of iterations as 1016. The classification accuracy of the trained models has been calculated according to Eq. (1). Classification accuracies of the created models and training times of the network are given in Table 12.

$$Accuracy = \frac{Numberofcorrectlydetected}{Totalnumberofsamples} \quad (1)$$

Examining Table 12, it is seen that the results obtained from all models are successful in determining how the person uses the mask. Besides, it can be stated that adding LSTM and BiLSTM layers to TrAlexNet and TrVGG16 models improves the classification accuracy. The highest classification accuracy of 95.67% has been achieved through the TrVGG16 + BiLSTM model. The impact of the network structure of architectures used in the study can be seen when considering the training times. It is possible to point out that the long time for the training performed with the VGG16 architecture depends on the number of convolutional layers in the network and the filter size of this layer. It has been observed that the use of LSTM and BiLSTM layers as the classification layer of CNN is more effective in solving long-term dependency problems. It is thought that this is due to the repetitive training of the LSTM and BiLSTM layers compared to the fully connected layer. Using the TrVGG16 + BiLSTM model, which has the highest classification success, average classification success as well as classification accuracy for each class was obtained. The results are shown in Table 13.

Table 9
Training parameters for all proposed models.

Solver	Initial Learn Rate	Validation Frequency	Max Epochs	Mini Batch Size
sgdm	0.0001	5	8	11

Table 10
Parameters of fc_optimized layer for all proposed models.

Weight Learn Rate Factor	Weight L2 Factor	Bias Learn Rate Factor	Bias L2 Factor	Weight Initializer	Bias Initializer
10	1	10	0	glorot	zeros

Table 11
Parameters of LSTM and BiLSTM layer for all proposed models.

Number Hidden Units	Output Mode	State Activation Function	Gate Activation Function
100	last	tanh	sigmoid

6. Conclusion and discussion

This article aims to use transfer learning to determine whether people are wearing the face mask correctly, based on the fact that face

Table 8
Layers and parameters of the proposed TrVGG16 + BiLSTM.

Layer Name	Layer Type	Filter Size	Stride	Padding	Output Channel	Activation Function
conv1_1	convolution 2d	3,3	1,1	1	64	relu
conv1_2	convolution 2d	3,3	1,1	1	64	relu
pool1	max pooling 2d	2,2	2,2	0	64	–
conv2_1	convolution 2d	3,3	1,1	1	128	relu
conv2_2	convolution 2d	3,3	1,1	1	128	relu
pool2	max pooling 2d	2,2	2,2	0	256	–
conv3_1	convolution 2d	3,3	1,1	1	256	relu
conv3_2	convolution 2d	3,3	1,1	1	256	relu
conv3_3	convolution 2d	3,3	1,1	1	256	relu
pool3	max pooling 2d	2,2	2,2	0	256	–
conv4_1	convolution 2d	3,3	1,1	1	512	relu
conv4_2	convolution 2d	3,3	1,1	1	512	relu
conv4_3	convolution 2d	3,3	1,1	1	512	relu
pool4	max pooling 2d	2,2	2,2	0	512	–
conv5_1	convolution 2d	3,3	1,1	1	512	relu
conv5_2	convolution 2d	3,3	1,1	1	512	relu
conv5_3	convolution 2d	3,3	1,1	1	512	relu
pool5	max pooling 2d	2,2	2,2	0	512	–
fc6	fully connected	–	–	–	4096	relu
fc7	fully connected	–	–	–	4096	relu
fc8	fully connected	–	–	–	1000	relu
flatten	flatten	–	–	–	–	–
bilstm	bilstm	–	–	–	–	–
fc_optimized	fully connected	–	–	–	4	softmax

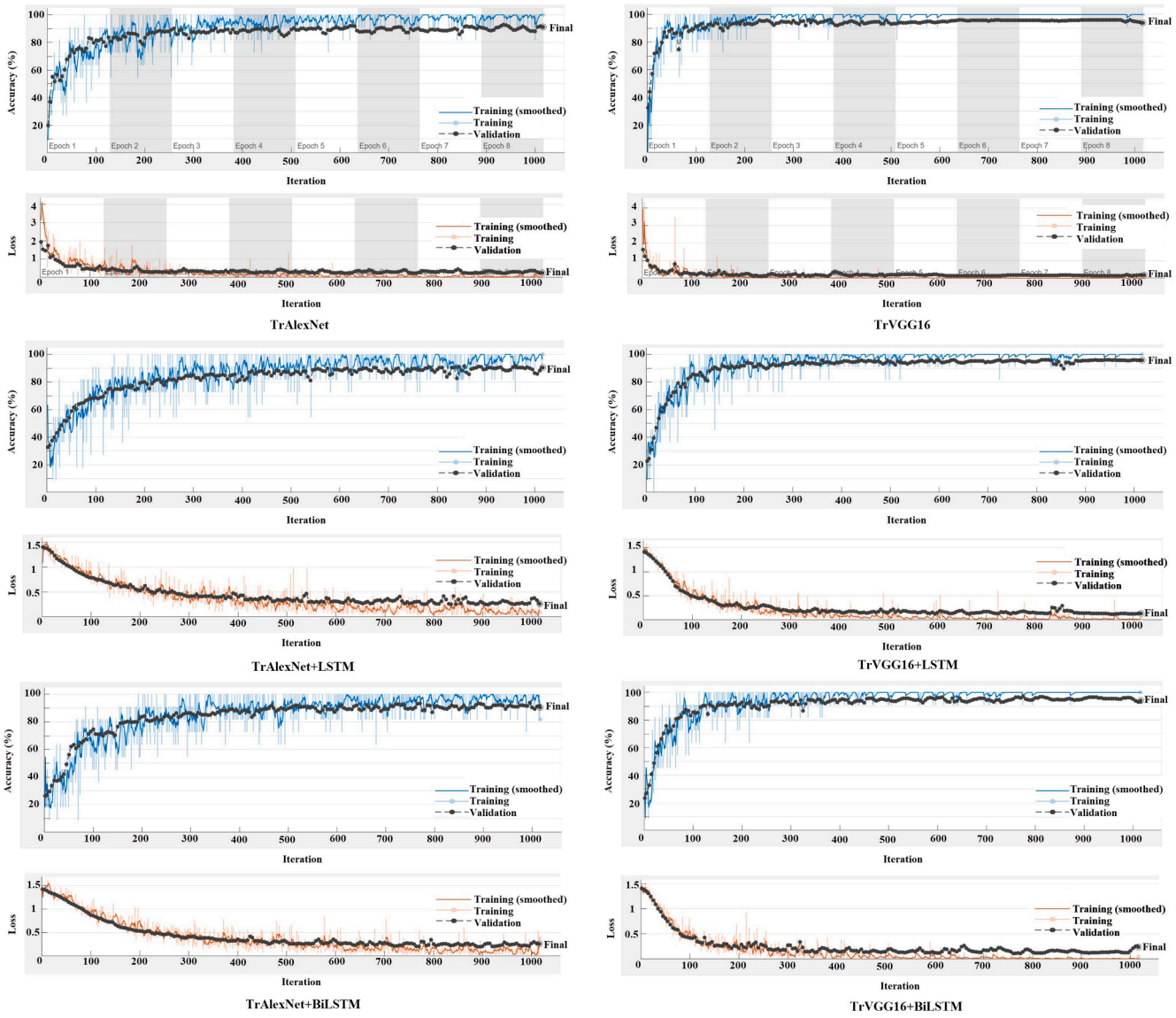


Fig. 12. Training and validation graphs of model.

Table 12
Classification accuracies of models and training times of the network.

Model	Accuracy (%)	Training time
TrAlexNet	90.33	8 min 38 sec
TrAlexNet + LSTM	90.55	7 min 25 sec
TrAlexNet + BiLSTM	91.17	7 min 43 sec
TrVGG16	94.00	74 min 55 sec
TrVGG16 + LSTM	94.17	74 min 23 sec
TrVGG16 + BiLSTM	95.67	75 min 23 sec

Table 13
Classification accuracy achieved by the TrVGG16 + BiLSTM model for all classes.

Class	Accuracy (%)
Masked	95.65
No_Mask	98.55
Masked but nose open	94.3
Masked but under the chin	94.2

mask usage can be an effective tool in preventing respiratory tract infectious diseases. Within the scope of the study, AlexNet and VGG16 CNN architectures have been used since the transmitted CNNs require less data compared to pre-trained CNNs.

The data used in the study consists of real images that have not been simulated, or collected over the web. The literature usually contains datasets consisting of images containing masked and unmasked situations. These datasets are datasets that are not created specifically, but are combined with different combinations within studies using publicly shared datasets. In addition, it is seen that the images used in many studies are also reproduced by data augmentation. The dataset used in this study was created specifically. It contains four different classes: Masked, No_masked, Masked but under the chin, Masked but nose open. From the face mask dataset, 1600 images were used for training and 400 images were used for testing, and necessary adjustments have been made to recognize 4 classes. Although there are 4 classes in the dataset created, it can be said that high success rates are achieved in contrast to other studies in the literature. However, due to the fact that the characteristics of the datasets found in the literature are different, the final state of the datasets obtained after data augmentation or various combinations of datasets has not been made public, our model could not be compared with other studies.

Among the proposed models, the highest classification accuracy of

95.67% was achieved with the TrVGG16 + BiLSTM model. Classification accuracy was also obtained for each of the 4 classes with this model. Because the Masked but nose open class is similar to the Masked class in the images contained in the dataset, classification accuracy has been negatively affected. The classification accuracy of the Masked but nose open class was achieved by 94.30% and the Masked class by 95.65%. Classification accuracy has been negatively affected because the Masked but under the chin class is similar to the No_mask class. The classification success of the Model masked but under the chin class was achieved by 94.2% and the No_mask class by 98.55%. The study has proven that it can be benefited from the proposed models besides transfer learning, to ensure the correct and effective use of the face mask, taking into account the public interest.

Although the proposed models have achieved success in face mask recognition, in future studies, different models, based on deep learning will be proposed, aiming at maximum success. Additionally, it is thought that the success will further increase by increasing the number of tagged data in the dataset. Moreover, it is aimed to detect violations in the use of face masks via real-time video recordings.

CRediT authorship contribution statement

Murat Koklu: Conceptualization, Data curation, Formal analysis, Supervision, Writing - original draft, Writing - review & editing. **Ilkay Cinar:** Investigation, Methodology, Project administration, Resources, Software, Supervision, Writing - original draft, Writing - review & editing. **Yavuz Selim Taspinar:** Supervision, Validation, Visualization, Writing - original draft, 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 would like to thank Selcuk University Scientific Research Coordinator for their support. This study was produced from Ilkay CINAR's unpublished Ph.D. thesis.

References

[1] M.M. Rahman, et al., An Automated System to Limit COVID-19 Using Facial Mask Detection in Smart City Network, in: 2020 IEEE International IOT, Electronics and Mechatronics Conference (IEMTRONICS). 2020. IEEE. DOI: 10.1109/IEMTRONICS51293.2020.9216386.

[2] R. Liu, Z. Ren, Application of Yolo on Mask Detection Task. arXiv preprint arXiv: .05402, 2021.DOI: arXiv:2102.05402.

[3] T. Li, Y. Liu, M. Li, X. Qian, S.Y. Dai, K. Thavorn, Mask or no mask for COVID-19: A public health and market study, PloS one 15 (8) (2020) e0237691, https://doi.org/10.1371/journal.pone.023769110.1371/journal.pone.0237691.g00110.1371/journal.pone.0237691.g00210.1371/journal.pone.0237691.g00310.1371/journal.pone.0237691.t00110.1371/journal.pone.0237691.t00210.1371/journal.pone.0237691.s00110.1371/journal.pone.0237691.r00110.1371/journal.pone.0237691.r00210.1371/journal.pone.0237691.r00310.1371/journal.pone.0237691.r004.

[4] A. Oumina, N. El Makhfi, M. Hamdi, Control The COVID-19 Pandemic: Face Mask Detection Using Transfer Learning, in: 2020 IEEE 2nd International Conference on Electronics, Control, Optimization and Computer Science (ICECOCS). 2020. IEEE. DOI: 10.1109/ICECOCS50124.2020.9314511.

[5] J. Wang, L. Pan, S. Tang, J.S. Ji, X. Shi, Mask use during COVID-19: A risk adjusted strategy, Environ. Pollut. 266 (2020) 115099, https://doi.org/10.1016/j.envpol.2020.115099.

[6] M.H. Haischer, R. Beilfuss, M.R. Hart, L. Opielinski, D. Wrucke, G. Zirgaitis, T. D. Uhrich, S.K. Hunter, Y. Kotozaki, Who is wearing a mask? Gender-, age-, and location-related differences during the COVID-19 pandemic, PloS one 15 (10) (2020) e0240785, https://doi.org/10.1371/journal.pone.024078510.1371/journal.pone.0240785.g00110.1371/journal.pone.0240785.g00210.1371/journal.pone.0240785.g00310.1371/journal.pone.0240785.g00410.1371/journal.pone.0240785.t00110.1371/journal.pone.0240785.s00110.1371/journal.pone.0240785.s00210.1371/journal.pone.0240785.s003.

[7] V. Offeddu, et al., Effectiveness of masks and respirators against respiratory infections in healthcare workers: a systematic review and meta-analysis, Clin. Infect. Dis. 65(11) (2017) 1934–1942. DOI: 10.1093/cid/cix681.

[8] C.J. Worby, H.-H. Chang, Face mask use in the general population and optimal resource allocation during the COVID-19 pandemic, Nat. Commun. 11 (1) (2020) 1–9, https://doi.org/10.1038/s41467-020-17922-x.

[9] D.N. Fisman, A.L. Greer, A.R. Tuite, Bidirectional impact of imperfect mask use on reproduction number of COVID-19: A next generation matrix approach, Infect. Dis. Model. 5 (2020) 405–408, https://doi.org/10.1016/j.idm.2020.06.004.

[10] R. Tirupathi, et al., Comprehensive review of mask utility and challenges during the COVID-19 pandemic, Le Infezioni in Medicina 28 (suppl 1) (2020) 57–63.

[11] I. Goodfellow, et al., Deep learning, Vol. 1, MIT press Cambridge, 2016. DOI: 10.4258/hir.2016.22.4.351.

[12] Y. Said, Pynq-YOLO-Net: An Embedded Quantized Convolutional Neural Network for Face Mask Detection in COVID-19 Pandemic Era, Int. J. Adv. Comput. Sci. Applications 11 (9) (2020) 100–106.

[13] S.V. Militante, N.V. Dionisio, Deep Learning Implementation of Facemask and Physical Distancing Detection with Alarm Systems, in: 2020 Third International Conference on Vocational Education and Electrical Engineering (ICVEE). 2020. IEEE. DOI: 10.1109/ICVEE50212.2020.9243183.

[14] L.i. Liu, W. Ouyang, X. Wang, P. Fieguth, J. Chen, X. Liu, M. Pietikäinen, Deep learning for generic object detection: A survey, Int. J. Comput. Vis. 128 (2) (2020) 261–318, https://doi.org/10.1007/s11263-019-01247-4.

[15] S.K. Addagarla, G.K. Chakravarthi, P. Anitha, Real Time Multi-Scale Facial Mask Detection and Classification Using Deep Transfer Learning Techniques, Int. J. Adv. Trends Comput. Sci. Eng. 9(4) (2020). DOI: 10.30534/ijatcse/2020/33942020.

[16] S.K. Dey, A. Howlader, C. Deb, MobileNet Mask: A Multi-phase Face Mask Detection Model to Prevent Person-To-Person Transmission of SARS-CoV-2, in: Proceedings of International Conference on Trends in Computational and Cognitive Engineering. 2021. Springer. DOI: 10.1007/978-981-33-4673-4_49.

[17] M.R. Bhuiyan, S.A. Khushbu, M.S. Islam, A Deep Learning Based Assistive System to Classify COVID-19 Face Mask for Human Safety with YOLOv3, in: 2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT). 2020. IEEE. DOI: 10.1109/ICCCNT49239.2020.9225384.

[18] M. Loey, G. Manogaran, M.H.N. Taha, N.E.M. 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, https://doi.org/10.1016/j.measurement.2020.108288.

[19] M.S. Islam, et al., A Novel Approach to Detect Face Mask using CNN, in: 2020 3rd International Conference on Intelligent Sustainable Systems (ICISS). 2020. IEEE. DOI: 10.1109/ICISS49785.2020.9315927.

[20] P. Mohan, A.J. Paul, A. Chirania, A Tiny CNN Architecture for Medical Face Mask Detection for Resource-Constrained Endpoints. arXiv preprint arXiv:.14858, 2020. DOI: arXiv:2011.14858.

[21] M. Razavi, et al., An Automatic System to Monitor the Physical Distance and Face Mask Wearing of Construction Workers in COVID-19 Pandemic. arXiv preprint arXiv:.01373, 2021.DOI: arXiv:2101.01373.

[22] C. Basha, B. Pravallika, E.B. Shankar, An Efficient Face Mask Detector with PyTorch and Deep Learning, EAI Endorsed Trans. Pervasive Health Technol. 7 (25) (2021) e4, https://doi.org/10.4108/eai.8-1-2021.167843.

[23] M. Loey, G. Manogaran, M.H.N. Taha, N.E.M. Khalifa, Fighting against COVID-19: A novel deep learning model based on YOLO-v2 with ResNet-50 for medical face mask detection, Sustain. Cities Soc. 65 (2021) 102600, https://doi.org/10.1016/j.scs.2020.102600.

[24] S. Yadav, Deep Learning based Safe Social Distancing and Face Mask Detection in Public Areas for COVID-19 Safety Guidelines Adherence, Int. J. Res. Appl. Sci. Eng. Technol. 8 (7) (2020) 1368–1375.

[25] R. Pagare, Face Mask Detection and Social Distancing Monitoring, Int. J. Res. Appl. Sci. Eng. Technol. 9 (1) (2021) 374–379.

[26] S.A. Sanjaya, S.A. Rakhmawan, Face Mask Detection Using MobileNetV2 in The Era of COVID-19 Pandemic, in: 2020 International Conference on Data Analytics for Business and Industry: Way Towards a Sustainable Economy (ICDABI). 2020. IEEE.DOI: 10.1109/ICDABI51230.2020.9325631.

[27] G.J. Chowdary, et al., Face mask detection using transfer learning of inceptionv3, in: International Conference on Big Data Analytics. 2020. Springer. DOI: 10.1007/978-3-030-66665-1_6.

[28] A.G. Sandesara, D.D. Joshi, S.D. Joshi, Facial Mask Detection Using Stacked CNN Model, Int. J. Sci. Res. Comput. Sci. Eng. Inform. Technol. (2020), https://doi.org/10.32628/CSEIT206553.

[29] A. Chavda, et al., Multi-Stage CNN Architecture for Face Mask Detection. arXiv preprint arXiv:.07627, 2020.DOI: arXiv:2009.07627.

[30] S.V. Militante, N.V. Dionisio, Real-Time Facemask Recognition with Alarm System using Deep Learning, in: 2020 11th IEEE Control and System Graduate Research Colloquium (ICSGRC). 2020. IEEE. DOI: 10.1109/ICSGRC49013.2020.9232610.

[31] W. Vijitkunsawat, P. Chantngarm, Study of the Performance of Machine Learning Algorithms for Face Mask Detection, in: 2020-5th International Conference on Information Technology (InCIT). 2020. IEEE. DOI: 10.1109/InCIT50588.2020.9310963.

[32] M.F. Aslan, M.F. Unlersen, K. Sabanci, A. Durdu, CNN-based transfer learning–BiLSTM network: A novel approach for COVID-19 infection detection, Appl. Soft Comput. 98 (2021) 106912, https://doi.org/10.1016/j.asoc.2020.106912.

[33] K. O'Shea, R. Nash, An introduction to convolutional neural networks. arXiv preprint arXiv:.08458, 2015.DOI: arXiv:1511.08458.

[34] M.M. Qanbar, S. Tasdemir, Detection of Malaria Diseases with Residual Attention Network, Int. J. Intell. Syst. Applications Eng. 7 (4) (2019) 238–244, https://doi.org/10.18201/ijisae.2019457677.
[35] G. Çinarer, B.G. Emiroğlu, A.H. Yurttakal, Prediction of Glioma Grades Using Deep Learning with Wavelet Radiomic Features, Appl. Sci 10 (18) (2020) 6296, https://doi.org/10.3390/app10186296.
[36] A. Krizhevsky, I. Sutskever, G.E. Hinton, Imagenet classification with deep convolutional neural networks, Adv. Neural Inform. Process. Syst. 25 (2012) 1097–1105.
[37] J. Deng, et al. Imagenet: A large-scale hierarchical image database. in 2009 IEEE conference on computer vision and pattern recognition. 2009. Ieee. DOI: 10.1109/CVPR.2009.5206848.
[38] K. Nguyen, C. Fookes, A. Ross, S. Sridharan, Iris recognition with off-the-shelf CNN features: A deep learning perspective, IEEE Access 6 (2018) 18848–18855, https://doi.org/10.1109/ACCESS.2017.2784352.
[39] K. Eckle, J. Schmidt-Hieber, A comparison of deep networks with ReLU activation function and linear spline-type methods, Neural Netw. 110 (2019) 232–242, https://doi.org/10.1016/j.neunet.2018.11.005.
[40] A. Abd Almisreb, N. Jamil, N.M. Din, Utilizing AlexNet deep transfer learning for ear recognition, in: 2018 Fourth International Conference on Information Retrieval and Knowledge Management (CAMP). 2018. IEEE.DOI: 10.1109/INFRKM.2018.8464769.
[41] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:.00784, 2014.DOI: arXiv:1409.1556.
[42] M. Bicakci, O. Ayyildiz, Z. Aydin, A. Basturk, S. Karacavus, B. Yilmaz, Metabolic Imaging Based Sub-Classification of Lung Cancer, IEEE Access 8 (2020) 218470–218476, https://doi.org/10.1109/Access.628763910.1109/ACCESS.2020.3040155.
[43] D.I. Swasono, H. Tjandrasa, C. Fathicah, Classification of tobacco leaf pests using VGG16 transfer learning, in: 2019 12th International Conference on Information & Communication Technology and System (ICTS). 2019. IEEE.DOI: 10.1109/ICTS.2019.8850946.
[44] H. Qassim, A. Verma, D. Feinzimer, Compressed residual-VGG16 CNN model for big data places image recognition, in: 2018 IEEE 8th Annual Computing and Communication Workshop and Conference (CCWC). 2018. IEEE. DOI: 10.1109/CCWC.2018.8301729.
[45] S.-H. Wang, Q. Zhou, M. Yang, Y.-D. Zhang, ADVIAN: Alzheimer's Disease VGG-Inspired Attention Network Based on Convolutional Block Attention Module and Multiple Way Data Augmentation, Front. Aging Neurosci. 13 (2021), https://doi.org/10.3389/fnagi.2021.687456.
[46] S.-H. Wang, et al., AVNC: Attention-based VGG-style network for COVID-19 diagnosis by CBAM. IEEE Sens. J., 2021. DOI: 10.1109/JSEN.2021.3062442.
[47] R. Pascanu, T. Mikolov, Y. Bengio, On the difficulty of training recurrent neural networks, in: International conference on machine learning. 2013. PMLR.
[48] V.Y. Senyurek, M.H. Imtiaz, P. Belsare, S. Tiffany, E. Sazonov, A CNN-LSTM neural network for recognition of puffing in smoking episodes using wearable sensors, Biomed. Eng. Lett. 10 (2) (2020) 195–203, https://doi.org/10.1007/s13534-020-00147-8.
[49] Huang, Z., W. Xu, K. Yu, Bidirectional LSTM-CRF models for sequence tagging. arXiv preprint arXiv:.01991, 2015. DOI: arXiv:1508.01991.
[50] X. Ma, E. Hovy, End-to-end sequence labeling via bi-directional lstm-cnns-crf. arXiv preprint arXiv:.01354, 2016. DOI: arXiv:1603.01354.
[51] G. Jain, M. Sharma, B. Agarwal, Optimizing semantic LSTM for spam detection, Int. J. Inform. Technol. 11 (2) (2019) 239–250, https://doi.org/10.1007/s41870-018-0157-5.
[52] B. Yang, S. Sun, J. Li, X. Lin, Y. Tian, Traffic flow prediction using LSTM with feature enhancement, Neurocomputing 332 (2019) 320–327, https://doi.org/10.1016/j.neucom.2018.12.016.
[53] Y. Bengio, P. Simard, P. Frasconi, Learning long-term dependencies with gradient descent is difficult, IEEE Trans. Neural Netw. Learn. Syst. 5 (2) (1994) 157–166, https://doi.org/10.1109/TNN.7210.1109/72.279181.
[54] S. Hochreiter, J. Schmidhuber, LSTM can solve hard long time lag problems, Adv. Neural Inform. Process. Syst. (1997) 473–479.
[55] A. Graves, S. Fernández, J. Schmidhuber. Bidirectional LSTM networks for improved phoneme classification and recognition, in: International conference on artificial neural networks. 2005. Springer. DOI: 10.1007/11550907_163.
[56] M. Jia, et al., Analysis and research on stock price of LSTM and bidirectional LSTM neural network, in: 3rd International Conference on Computer Engineering, Information Science & Application Technology (ICCIA 2019). 2019. Atlantis Press. DOI: 10.2991/iccia-19.2019.72.
[57] A. Graves, J. Schmidhuber, Framewise phoneme classification with bidirectional LSTM and other neural network architectures, Neural Netw. 18 (5–6) (2005) 602–610, https://doi.org/10.1016/j.neunet.2005.06.042.