sciendo

Polish Journal of Medical Physics and Engineering
The Journal of Polish Society of Medical Physics
ISSN 1898-0309, doi: 10.2478/pjmpe-2022-0014

September 2022
Vol 28, Issue 3

Scientific Paper

Automatic diagnosis of severity of COVID-19 patients using an ensemble of transfer learning models with convolutional neural networks in CT images

Ahmad SHALBAF [1,ABCDEF], Parisa GIFANI[2,ABCDEF], Ghazal MEHRI-KAKAVAND[3,ABC], Mohamad PURSAMIMI[3,ABCDF], Mahdi GHORBANI[1,CDF,*], Amirhossein Abbaskhani DAVANLOO[4,BC], Majid VAFAEEZADEH[5,ABCE]

[1]*Biomedical Engineering and Medical Physics Department, School of Medicine, Shahid Beheshti University of Medical Sciences, Tehran, Iran*

[2]*Department of Biomedical Engineering, Science and Research Branch, Islamic Azad University, Tehran, Iran*

[3]*Department of Medical Physics, School of Medicine, Semnan University of Medical Sciences, Semnan, Iran*

[4]*Department of Radiology, Faculty of Medicine, Mashhad University of Medical Sciences, Mashhad, Iran*

[5]*School of Electrical Engineering, Iran University of Science and Technology, Tehran, Iran*

[*]*Corresponding author: Mahdi Ghorbani; mhdghorbani@gmail.com*

(received 23 March 2022; revised 9 June and 2 July 2022; accepted 4 July 2022)

Abstract

Introduction: Quantification of lung involvement in COVID-19 using chest Computed tomography (CT) scan can help physicians to evaluate the progression of the disease or treatment response. This paper presents an automatic deep transfer learning ensemble based on pre-trained convolutional neural networks (CNNs) to determine the severity of COVID -19 as normal, mild, moderate, and severe based on the images of the lungs CT.

Material and methods: In this study, two different deep transfer learning strategies were used. In the first procedure, features were extracted from fifteen pre-trained CNNs architectures and then fed into a support vector machine (SVM) classifier. In the second procedure, the pre-trained CNNs were fine-tuned using the chest CT images, and then features were extracted for the purpose of classification by the softmax layer. Finally, an ensemble method was developed based on majority voting of the deep learning outputs to increase the performance of the recognition on each of the two strategies. A dataset of CT scans was collected and then labeled as normal (314), mild (262), moderate (72), and severe (35) for COVID-19 by the consensus of two highly qualified radiologists.

Results: The ensemble of five deep transfer learning outputs named EfficientNetB3, EfficientNetB4, InceptionV3, NasNetMobile, and ResNext50 in the second strategy has better results than the first strategy and also the individual deep transfer learning models in diagnosing the severity of COVID-19 with 85% accuracy.

Conclusions: Our proposed study is well suited for quantifying lung involvement of COVID-19 and can help physicians to monitor the progression of the disease.

Keywords: severity of COVID-19; convolutional neural network; ensemble.

Introduction

The new coronavirus 2019 (COVID-19) is an infection that has involved more than 200 million people worldwide from October 2021.[1] There are different techniques for the diagnosis of this disease which are mainly based on polymerase chain reaction (PCR). However, with the PCR technique, there is a need for a long time to obtain the diagnosis results. In addition, the severity of the disease is not detectable by this method. For these reasons, the alternative testing method, Computed tomography (CT) from the chest, is used to provide faster, more efficient, and more accessible testing.[2] Chest images CT also contain important information for diagnosing the severity and magnitude of lung involvement qualitatively and quantitatively in patients.

The assessment of severity is critical for the effective administration of appropriate therapeutic agents based on the patient's condition.[3-4] The assessment of the three severity classes (namely, mild, moderate, and severe) was well described in the world health organization (WHO) report on the clinical control of COVID-19 disease to determine an adequate treatment.[5] Currently, severity is evaluated based on the symptoms, clinical parameters, and chemical analyses. On the

© 2022 Authors This is an open access article licensed under the Creative Commons Attribution-NonCommercial-NoDerivs License (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Authors' contribution: A – Research concept and design, B – Collection and/or assembly of data, C – Data analysis and interpretation, D – Writing the article, E – Critical revision of the article, F – Final approval of the article.

other hand, most of the specific symptoms of the disease are seen differently at different stages of the disease[6] and indicate the indirect status of the lung, which is the most affected organ. According to the guidelines, chest examinations by CT in the coronavirus disease indicate ground-glass opacities (GGOs), which are developed in size and develop with crazy-paving patterns and subsequent consolidation.[7-9] Therefore, processing the CT images from the patient's lung is an appropriate tool to evaluate the severity of the involvement of the lung in order to monitor and follow the specified treatment. Currently, these images are used to quantify disease burden and follow-up patients every 3-5 days to evaluate the disease progression or response to treatment[10] and can help physicians to perform a better control of the infections. Recent studies have also shown that lung CT quantification is used to evaluate the severity and progression of the COVID-19 disease.[11-13] Diagnosis of COVID-19 and also evaluation of the severity of this disease can be performed based on CT images by a highly expert radiologist by visually scoring. However, visual development is tedious and subjective and the validity of the diagnosis depends on the experience of the radiologists. Quantitative analysis of CT images using advanced machine learning techniques could provide an automatic, reliable, and objective assessment of the disease burden, facilitating and expediting imaging interpretation during the COVID-19 pandemic.

Various machine learning methods are widely used for image processing in medicine. Among these methods, deep learning ones, especially convolutional neural networks (CNNs), have been introduced in recent years as the newest and most advanced methods.[14] However, there is a need for large amounts of training data for these models. To solve this problem, transfer learning methods have been introduced in which a neural architecture with pre-trained weights is considered in a large dataset, and then is applied to the target task with a limited amount of training data.[15] In other words, a powerful deep neural network is learned on a set of large datasets for comprehensive extraction of features and then is applied to a target task that includes a small-sized dataset. Recently, several studies have applied deep learning methods for the automatic detection of COVID-19 diseases.[16-21] Wang et al. developed a 3D deep CNN[22], Bansal et al. proposed a genetic algorithm based deep learning model[23], Almalki et al. used COVID Inception-ResNet model[24], and Irfan et al. proposed hybrid deep neural networks to detect COVID-19.[25] However, the mentioned studies only addressed the automatic detection of this disease by the CT images and did not consider the severity of the disease.

Few studies have used CT images to determine the severity of patients based on lung involvement. In one study, texture features and a random forest classifier are used for the classification of severity of COVID-19 patients.[26] In another study, the changes in chest radiographs CT of 21 COVID-19 patients were determined on the basis of initial diagnosis until the patient recovered.[27] From that study, most patients who recovered from COVID-19, had the most lung changes on CT scans approximately 10 days after symptom onset. In another study, the U-Net deep learning method was applied to segment opacities on lung CT images and then classified the CT images of 126 patients diagnosed with COVID-19 disease into mild, moderate, severe, and critical classes.[28] In another study, an image segmentation system was developed based on deep learning methods for automatic quantification of the infected regions on the CT chest scans of COVID-19 patients.[29] Moreover, lesions were segmented in terms of opacities, crazy-paving patterns, and consolidations in 101 expert-annotated cases from a database of different centers in Italy. Then, the severity was determined based on intensity and texture features for three classes, including mild, moderate, and severe.[30] Finally, percent lesion coverage and mean lesion density were used to evaluate the three aforementioned severity classes in 44 patients.[31] However, these methods require image segmentation of the percentage of lesions in the lung images of the COVID-19 patients, which is a problematic task.

The innovation of this study is to determine the severity of COVID-19 patients into four classes: normal, mild, moderate, and severe, on a different set of CT images with two different automatic ensemble methods of deep transfer learning outputs based on different powerful pre-trained CNNs architectures without the need for image segmentation. We want to optimize this ensemble of deep learning classifiers to improve model generalization and prediction accuracy by leveraging the capabilities of each classifier. It should be noted that this study for the diagnosis of severity of COVID-19 patients is different from other studies with only detecting COVID from non-COVID patients and is used to track the changes in patients' conditions and make better decisions to cure and on the time period required for hospitalization.

Materials and methods

Database

Images of the lung CT were collected from people with suspected COVID-19 in 3D format. A total of 683 patients (408 men and 275 women) were included in this study (**Table 1**). The chest CT was acquired using a Hispeed CT dual slice scanner (Healthcare, USA) in spiral mode at Qaboos Teb Golestan Medical Imaging Center (Gonbad-e-Kavous, Iran). To minimize the artifacts due to motion, the imaging was performed during a single breath-hold. During the acquisition of CT, the tube voltage and exposure were 120 kVp and 60 mAs, respectively. Images from CT were stored in standard DICOM format with a size of 512 × 512 pixels as axial images (slice thickness, 10 mm; step size, 10 mm). The average number of 2D images for each patient in a 3D scan CT is 21. It should be noted that, in COVID-19 patients, thinner thicknesses from 10 mm slices are not advisable in order to reduce the time of scanning, minimize patients' exposure (radiation dose), and acquire better image

quality. In these patients, it is very important to minimize imaging time, since the risk of recurrent coughs in these patients is high, resulting in motion artifacts and thus lowering image quality. Moreover, the sheer volume of patients referring to CT centers makes it even more necessary to reduce imaging time.

Chest CT images were used to assess the severity of involvement of lungs in COVID-19 patients. The main CT observations were based on the international standard nomenclature which was defined in the Fleischner Society glossary and peer-reviewed studies on viral pneumonia using terms including GGOs, crazy-paving pattern (GGO with superimposed inter-and intralobular septal thickening), and consolidation.[7-9] The lung findings on the chest scan CT were small unilateral or bilateral GGOs and became larger and developed as crazy-paving patterns and subsequent consolidation. Based on the regions involved in the lung, a scoring system was utilized for quantitative evaluation of the severity of involvement of the lung. Each of the five lung lobes (left lower, upper lobe, right lower, middle, and upper lobe) was visually labeled by scores in the range of 0 to 5. In this case, a 0 score indicates no involvement; a 1 score means less than 5% involvement; a 2 score indicates 5%-25% involvement; a 3 score 26%-49% involvement; and a 4 score means 50%-75% involvement; and 5 greater than 75% involvement. The total score in the CT images was calculated as the sum of the individual lobar scores and ranged from 0 (which means no involvement) to 25 (which means maximum involvement). The infection score COVID-19 was determined based on the condition of the lungs by two experienced radiologists who had more than ten years of experience in thoracic radiology and thousands of patients examined. The final scores were specified by consensus. Overall, there was a higher probability of the affection of lower lobes, and their CT scores were higher. Subpleural lesions were more commonly seen than those lesions in the central lung. In most CT scans, bilateral lung involvement was seen. Based on these considerations, the CT images are classified into four classes, including normal lung tissue and COVID-19 related findings in three different disease levels, depending on the percentage of pulmonary involvement and the CT total score: mild with a total score of 1-10, moderate with a total score of 11-15, and severe with a total score of 16-25. Thus, our data set based on the consensus of two highly qualified radiologists included 314 normal cases, 262 mild stage cases of COVID-19, i.e., lung parenchymal involvement of less than 25%, 72 moderate stage cases, i.e., with lung parenchymal involvement between 25 and 50%, and 35 severe stage cases of COVID-19, i.e., with lung parenchymal involvement of more than 50%. Some slices of CT images for 4 individuals in our dataset in different classes, including normal, mild, moderate, and severe stages are shown in **Figure 1**. The data of our research is available for other researchers. The download link is as follows: "https://github.com/medical-dataset/CT-Images" which contains the path to Google drive to download the dataset.

Table 1. Our dataset in each class

Categories	Number of patients	Males/females	Number of patients in training	Number of patients in validation	Number of patients in test
Normal	314	186/128	252	31	31
Mild stage of Covid-19	262	158/104	210	26	26
Moderate stage of Covid-19	72	43/29	58	7	7
Severe stage of Covid-19	35	22/13	27	4	4
Total	683	408/275	547	68	68

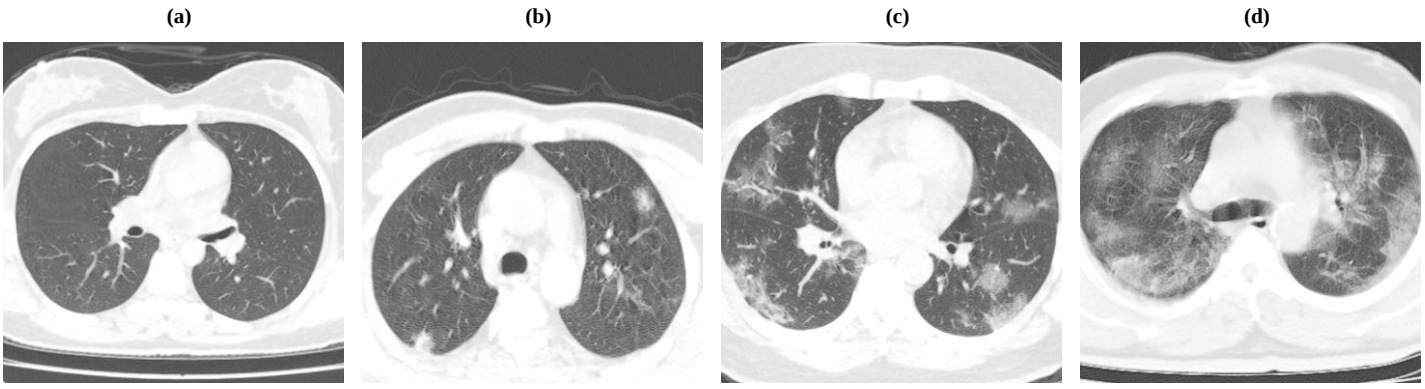


Figure 1. CT images in different classes, including normal state (a), mild (b), moderate (b), and severe stage (d)

Preprocessing

As we mentioned earlier, the images of the lungs CT of COVID-19 patients were collected in 3D format. In 3D format, 21 2D images are required for each patient with a slice thickness of 10 mm. Then, in a preprocessing phase, each of these 21 2D images is evaluated to find the best images from the collection of images that completely contain the part of the lung volume (**Figure 2**). For this purpose, the following operations are performed for each subject. In the first step of preprocessing, for each original 2D image, which is 512×512 in size, the central region was extracted with 333×333 pixels. The 2D images were then converted to binary images using the method of Otsu[32] and then complemented. In the following, the actual numbers of pixels set to 1 on the binary 2D image in the region (area) were extracted as scalar numbers. These pixels correspond to the part of lung volume in each image. Usually, in the 3D CT scan, the part of lung volume in these 21 2D images gradually increases and decreases again. By extracting the maximum curve of the pixels set to 1, we can obtain the 2D images corresponding to the maximum lung volume (**Figure 2**). To increase the information, we also used 2D images adjacent to this 2D image. For this purpose, in addition to the 2D image that has the largest lung volume, we automatically extracted the previous five 2D images and the next five from the collection of 2D images of each subject (**Figure 2**). Therefore, from 21 2D images in 3D format, 11 2D images were selected from each person.

Pre-trained convolutional neural network

CNNs are the most popular deep learning methods that are widely used in medical applications.[33] In CNNs, by applying multiple building blocks such as convolutional layers, pooling layers, nonlinear layers, batch normalization, and fully connected layers, spatial hierarchies of features can be learned automatically through backpropagation. Finally, fully connected layers prepare the extracted features for classification by the softmax layer. Pooling layers are types of convolutional layers and are utilized to increase the translation invariance and reduce the map size of features. In the case of nonlinear problems, nonlinear layers (mainly the ReLU function) are utilized to increase the strength of the network. The number of parameters used in the CNN models is very large, which leads to more complex calculations and a need for too much training data. Therefore, transfer learning methods were used to compensate for the small size of the used datasets.[15] In these methods, a powerful deep neural network is learned on large datasets to extract comprehensive features and then applied or fine-tuned to the target task while the small-sized dataset is used. The pre-trained CNNs model is formed by training different CNN architectures on very large sets of images with different categories. The models are trained on the ImageNet dataset, which contains a number of 14 million images from 1000 categories of animals, objects, and other samples.[34] ResNet-50, SeResnet50, InceptionV3, DenseNet121, ResNext50, Xception, EfficientNets (B0-B5), NasNetLarge, NasNetMobile, and Inception_resnet_v2 are other types of pre-trained CNNs which are commonly used[35] and they were applied in this study. The advantages of these networks for research purposes include less training time, weaker hardware requirements, and lower computation time.

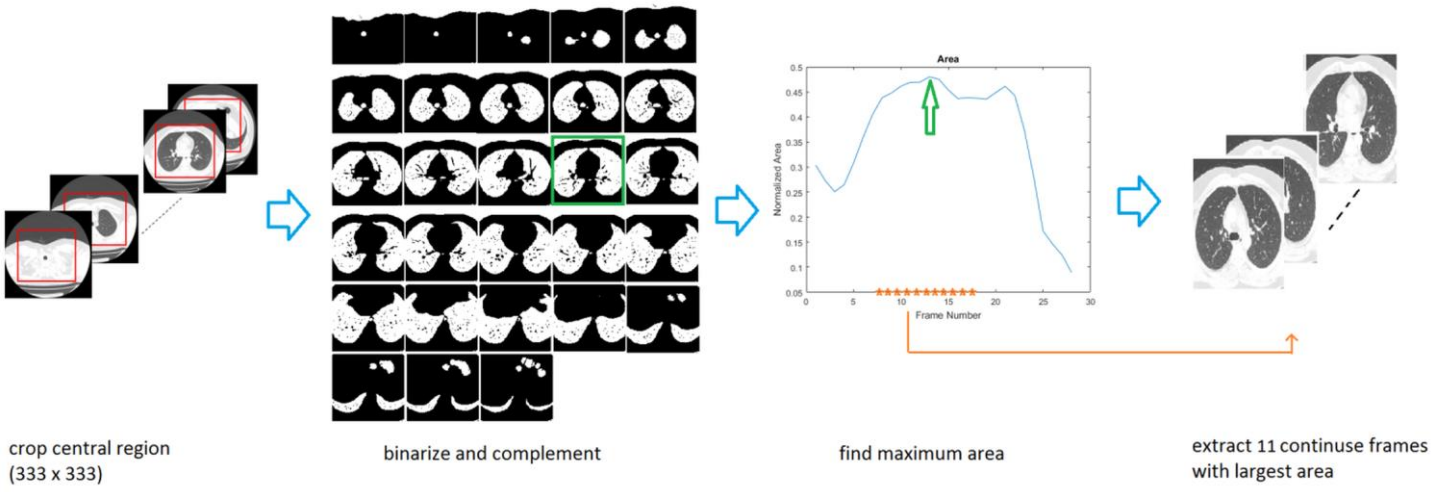


Figure 2. Preprocessing data for extracting the best CT images from the collection of CT images from each patient

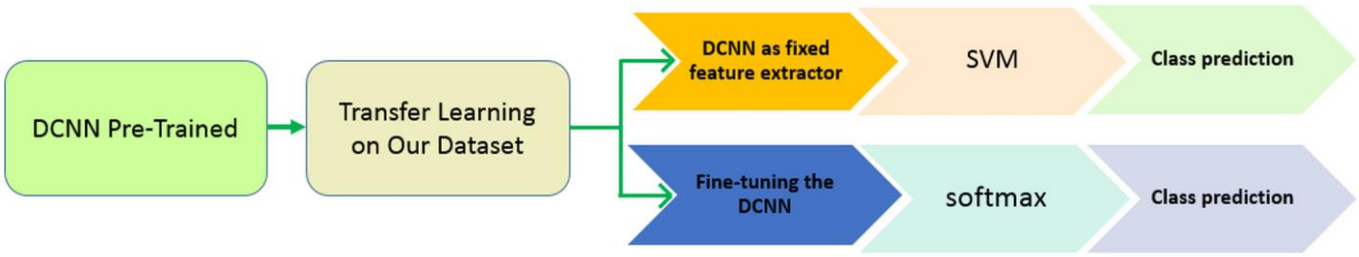


Figure 3. Two different procedures to find the best method for detecting the severity of COVID-19 from CT images by transfer learning methods with pre-trained CNN

Proposed deep learning methods

As shown in **Figure 3**, two different procedures were performed to find the best method for detecting the severity of COVID-19 disease by the use of CT images. In the first procedure, only features from pre-trained CNN architectures were extracted, and they were fed to an SVM classifier to classify the image data. In the second procedure, the pre-trained CNNs were fine-tuned on the chest CT images, and then the extracted features were classified separately by the softmax layer. When fine-tuning the CNN models, data augmentation was performed during training to avoid overfitting. In this study, we applied a random 10% shift to the vertical and horizontal axes to expand the original dimensions. In addition, a small random zoom combined with a random rotation of 20 degrees was applied to the training image group. The size of the dataset was also increased by the horizontal flipping of the images. For the purpose of fine-tuning all networks, we removed all fully connected layers and only the convolutional portion of each model's architecture was used. We also added a layer above the last convolutional layer to summarize the global average. After the flattening layer, we applied softmax non-linearity in the final classification layer. For network fine-tuning, all models were fine-tuned with 50 epochs using optimization of stochastic gradient descent (SGD). The initial learning rate was 0.0001, the Nesterov momentum was 0.9, and the batch size was 32. In all cases, the categorical cross-entropy was used as the loss function. It was noted that the input size of each pre-trained CNN network was different. Therefore, in the first step of data preparation, all 2D images were resized to 512 × 512 pixels to account for the different sizes of model inputs. These models were trained with the same initialization and learning rate rules. In table 2, the analyzed pre-trained CNNs used in this study are compared from computational complexity and computing time. Floating-point operations (FLOPs) are often used to describe how many operations are required to run a single instance of a given model and can serve as an accurate proxy for any kind of efficiency, including run-time. According to **Table 2**, the best models in terms of FLOPs are EfficientNetB0, NASNetMobile and EfficientNetB1, respectively. Finally, the most computational burden are corresponds to NASNetLarge, Inception-ResNet and EfficientNetB5, respectively.

Evaluation metrics

As the second strategy, 15 independent pre-trained CNN models were fine-tuned, and then prediction was performed for the test dataset. Due to the limited number of cases in the dataset, the 10-fold cross-validation technique was used for partitioning the data. As the experiment step, the dataset was partitioned into training, validation, and test sets (including 80%-10%-10% percentages of sample cases in the sets). The training data was used to introduce an appropriate model which was evaluated using the validation data. The performance of the developed model was evaluated using the test data. A common classification metric was utilized to evaluate the proposed method and its accuracy in each class separately and also in the four mentioned classes. Accuracy is used as a statistical measure of how well a deep learning classification method correctly identifies a score (normal, mild, moderate, or severe) compared to the reference score method by the consensus of the two radiologists. Accuracy is defined as the proportion of correct predictions among the total number of cases examined.

Table 2. Comparison of the analyzed pre-trained CNNs used from computational complexity, computing time and input shape

Architecture Name	Parameters	FLOP	# trainable layers	Input shape
EfficientNetB0	4,049,564	0.39B	237	224x224
EfficientNetB1	6,575,232	0.70B	338	240 x 240
EfficientNetB2	7,768,562	1.0B	338	260 x 260
EfficientNetB3	10,783,528	1.8B	383	300 x 300
EfficientNetB4	17,673,816	4.2B	473	380 x 380
EfficientNetB5	28,513,520	9.9B	575	456 x 456
NasnetLarge	84,916,818	24B	1038	331 x 331
NasnetMobile	4,269,716	0.56 B	768	224x224
Inception-V3	21,802,784	5.7B	310	299x299
Inception-ResNet	54,336,736	13B	779	299x299
ResNet 50	23,546,057	4.1B	189	224x224
Xception	20,861,480	8.4B	131	299x299
ResNeXt	23,048,137	4 B	215	224x224
SEResNet	26,092,144	4 B	286	224x224
DenseNet 121	7,037,504	3 B	426	224x224

Results

The general pipeline of assessing the severity of lung involvement in the diagnosis of COVID-19 disease, which was developed based on the proposed deep learning structure, is shown in **Figure 4**. After data acquisition of 3D images of the lungs CT, the next necessary step is data preparation, which includes data preprocessing and data partitioning. Then, the proposed methods are applied with two strategies presented in the previous section to optimize the performance for the diagnosis of the severity of COVID-19. Finally, an ensemble method was used by majority voting for the best deep transfer learning outputs with different CNNs. Finally, the developed system is evaluated using the classification metric. The details of each stage are described above. Fifteen independent pre-trained CNNs were fine-tuned. These models were applied to the COVID-19 CT image dataset to transfer the information into our task with limited training data. In the first procedure, only features from the pre-trained CNN architectures were extracted and fed to the SVM classifier. In the second procedure, the pre-trained CNNs were fine-tuned for the chest CT images, and then features were extracted and classified with a softmax layer. Deep learning was performed using Python and Keras software with a graphic processing unit (GeForce GTX 1080 Ti, NVIDIA).

Accuracy and loss function in the training sets for fine-tuning different pre-trained CNN models in the second strategy are presented in **Figures 5** and **6**. As it can be seen from these figures, after the 50th epoch, the model converges in the training process. Therefore, after 50 steps of training, diagnostic accuracy was calculated based on the test set.



Figure 4. A block diagram of the proposed method for detecting the severity of COVID-19 from CT images

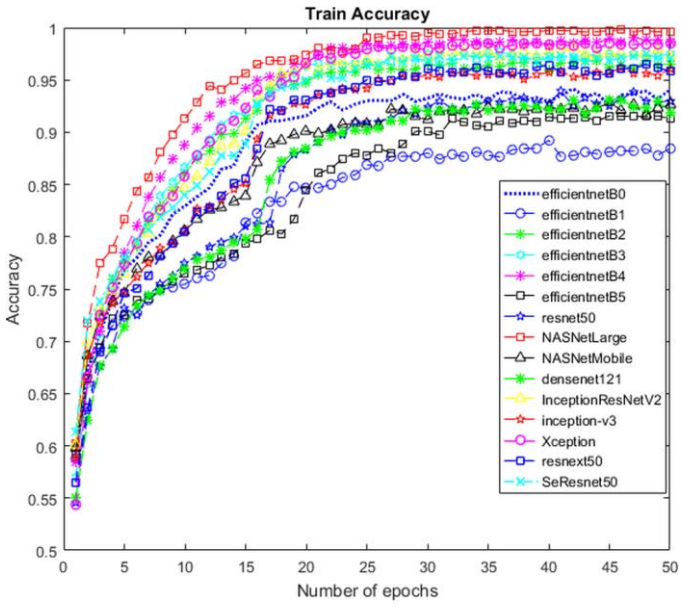


Figure 5. The training accuracy for different pre-trained CNN models in the second strategy

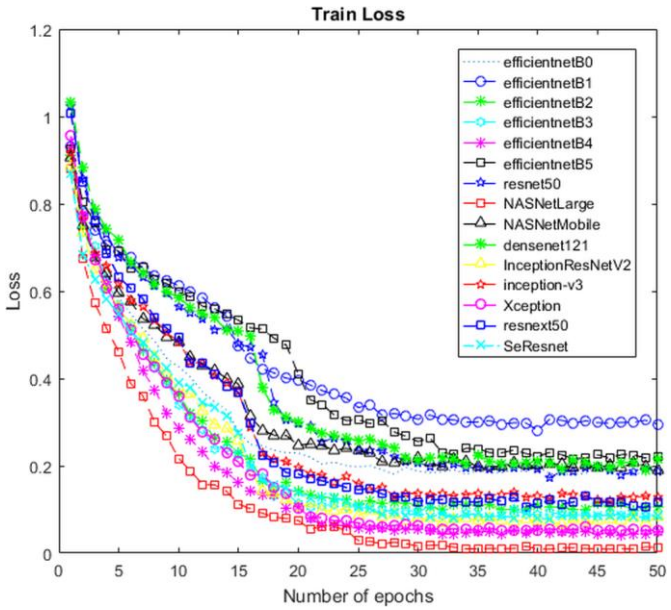


Figure 6. The cross-entropy loss function for different pre-trained CNN models in the second strategy

The results obtained from the two deep learning strategies are listed in **Table 3** using the accuracy metric. For the first and second strategies, EfficientNetB4 has higher results than the other models with an accuracy of 76% and 80%, respectively. Following this, EfficientNetsB3, InceptionV3, NasnetMobile, and ResNext50 have better results with an accuracy of 75% and 79% in the first and second strategies, respectively. Among the two strategies, the results show that the transfer learning models in the second strategy with fine-tuning achieve better results than the first strategy. To improve the overall accuracy in the classification of the severity of COVID-19 disease from CT images, an ensemble model was created using majority voting among all possible basic network architectures to assign the final prediction to the test dataset. So, all possible three, five, seven, nine, eleven, thirteen and fifteen network architectures are evaluated to select the best ones. The experimental results in **Table 3** indicate that an ensemble of five pre-trained CNN models named EfficientNetsB4, EfficientNetsB3, InceptionV3, NasnetMobile, and ResNext50 improves the overall performance in each of the two strategies compared to each individual basic model and also all other possible ensemble models. The accuracy of this ensemble of the above five transfer learning models in the second strategy has higher results with a value of % 85. Finally, the confusion matrix for the ensemble model was calculated and presented in **Tables 4** and **5** for the first and second strategies, respectively.

Table 3. The results of classification accuracy on the test dataset obtained by the used deep transfer learning systems based on the different architecture of pre-trained CNNs and also proposed ensemble method in two strategies

	Accuracy of First strategy with SVM					Second strategy with fine-tuning and softmax				
	Normal	**Mild**	**Moderate**	**Sever**	**Four classes**	**Normal**	**Mild**	**Moderate**	**Sever**	**Four classes**
EfficientNetB0	77	73	28	50	69	77	76	57	50	73
EfficientNetB1	83	57	71	50	71	83	69	71	50	75
EfficientNetB2	90	61	57	50	74	90	69	57	50	76
EfficientNetB3	83	73	57	50	75	83	76	71	75	79
EfficientNetB4	87	69	71	50	76	90	73	71	75	80
EfficientNetB5	83	65	71	25	72	87	69	71	50	76
Inception_resnet_v2	87	61	57	25	71	87	61	57	50	72
InceptionV3	83	77	57	25	75	83	80	71	50	79
NASNetLarge	87	61	42	0	68	87	61	42	25	69
NASNetMobile	83	69	57	75	75	87	73	71	75	79
ResNet50	74	84	42	50	74	77	84	57	50	76
Xception	67	84	57	50	72	74	84	57	50	74
DenseNet121	87	57	57	25	69	90	61	57	50	73
SeResnet50	87	61	28	75	71	87	61	57	75	73
ResNext50	93	61	57	50	75	93	69	71	50	79
Proposed Ensemble model	87	80	71	50	81	90	85	71	75	85

Table 4. Confusion matrix obtained for the proposed ensemble model in the first strategy for different stages of COVID-19 disease (Normal, Mild stage, Moderate stage, and Sever stage) on the test data set.

		Estimated labels by the proposed method			
		Normal	**Mild stage**	**Moderate stage**	**Sever stage**
Reference (diagnosis of radiologist)	Normal	27	2	2	0
	Mild stage	2	21	2	1
	Moderate stage	0	1	5	1
	Sever stage	0	0	2	2

Table 5. Confusion matrix obtained for the proposed ensemble model in the second strategy for different stages of COVID-19 disease (Normal, Mild stage, Moderate stage, and Sever stage) on the test data set.

		Estimated labels by the proposed method			
		Normal	**Mild stage**	**Moderate stage**	**Sever stage**
Reference (diagnosis of radiologist)	Normal	28	2	1	0
	Mild stage	1	22	2	1
	Moderate stage	0	1	5	1
	Sever stage	0	0	1	3

Discussion

In this research, two automatic deep transfer learning systems were developed based on powerful pre-trained CNNs for the classification of the severity of lung involvement in COVID-19 disease into mild, moderate, and severe using CT lung images. For this purpose, 15 state-of-the-art pre-trained CNN architectures were used and trained on the imageNet dataset that showed excellent performance. The pre-trained CNN architectures include: InceptionV3, ResNet-50, SeResnet50, EfficientNets (B0-B5), NasNetLarge, ResNext50, NasNetMobile, Xception, DenseNet121, and Inception_resnet_v2. In the first procedure, we only extracted features from pre-trained CNNs and fed them to the SVM classifier. In the second procedure, the pre-trained CNNs were fine-tuned on the chest CT images and then extracted features that are classified by the softmax layer. Finally, an ensemble method was developed based on majority voting among the deep transfer learning outputs to improve the performance of recognition for each of the two strategies. The obtained experimental results imply that the ensemble of five deep transfer learning outputs named EfficientNetsB4, EfficientNetsB3, InceptionV3, NasnetMobile, and ResNext50 in the second strategy with the fine-tuning of pre-trained CNN achieves better results than the first strategy and also the individual deep transfer learning models with 85% accuracy. Thus, this automated model can help physicians to make a better decision in monitoring and then reviewing the treatment with an accurate diagnosis of the severity of COVID-19 lung involvement.

An ensemble of base classifiers is a common method for improving model generalization and prediction accuracy and normally improves a single classifier by reduction of variance in the final prediction. An ensemble of base classifiers can achieve better results than any single classifier by leveraging the capabilities of each classifier. In our research, an ensemble of deep transfer learning classifiers is created using the different architectures of pre-trained CNN networks. A total of 5 different deep transfer learning architectures were utilized to develop an ensemble of classifiers. The final classification was performed based on majority voting (**Table 3**). The accuracy of classification for the ensemble model in the second strategy is improved compared to all individual transfer learning classifiers. The results in **Table 3** show that the ensemble model reduces the number of error predictions and increases the number of correct predictions compared to the best single transfer learning model.

Our main problem with the proposed method is to separate CT images from the moderate and severe stages of COVID-19. Our classifier separated normal and mild classes very well, but moderate and severe stages are confused. The reason is the similarity of the found CT patterns in these two stages. So, the patterns of these images are very similar, and it was too difficult for the classifier to separate them accurately. Another reason could be the smaller number of datasets in moderate and severe stages compared to normal and mild stages, so, moderate and severe stages are falsely identified as normal and mild stages (**Table 3**). Thus, due to the similarity of the patterns and the low number of datasets in moderate and severe stages, our classifier is not able to separate these two stages very well.

The results obtained by the second strategy are better than those obtained by the first strategy. There is one main problem in using the first strategy. The large image data used for training the CNN models mostly belong to the general domain, namely, cats, dogs, chairs, etc. However, the images of CT in the present study of COVID-19 disease had different visual appearances. Therefore, the visual representations which are learned by the use of these large images are not capable of adequately representing the CT images from the lungs. This leads to a bias in network extraction features on the dataset and lower generalization of the target data. Therefore, in the second strategy, the pre-trained CNN models were fine-tuned and modified to match the targeted task on the used dataset. Transfer learning methods were applied to compensate for the lack of large datasets and achieve better results. In these methods, a powerful deep neural network is learned on large datasets to extract comprehensive features, and then fine-tuned to the target task with a dataset with a small size. In other words, we transfer the information from a large database to the problem of this study with an insufficient database using a transfer learning approach.

Table 6 compares the results of our work with other studies[29,36] in classifying the severity of COVID-19 patients using CT images. As can be seen, the performance obtained in this study is promising. The advantage of the present study over other studies in this field is the comparison of deep transfer learning models with high performance pre-trained CNNs and an ensemble of a deep transfer learning system without the need for image segmentation. The major limitation of this study is the size of the dataset used for training the deep neural networks, especially in the moderate and severe stages of COVID-19. By performing regularization terms and simplification of the deep models, it is possible to overcome this problem. Furthermore, the study population was only contained from only one site.

Table 6. Comparison of classification results of our work with other studies in the classification of severity of COVID-19 patients from CT images

Study	Methods or features	Results
Shan et al., 2020[28]	quantification of infection regions by Deep learning	Accuracy = 85.1 %
Tang et al., 2020[35]	Infection volume of the whole lung and the volume of GGO and RF probabilities	Accuracy = 81.5 %
Our work	ensemble of transfer learning models with convolutional neural networks	Accuracy = 85 %

In future work, we want to further expand the experimental space by collecting a large number of samples. Also, data input from various hospitals could be used to develop and incrementally train the proposed model to predict the stage of COVID-19 patients more precisely in clinical settings. Finally, employing the developed methodology on other CT images can be very useful.

Conclusion

We proposed a new automatic ensemble of deep transfer learning systems based on powerful pre-trained CNNs to determine the severity of lung involvement in COVID-19 disease based on the images of CT into normal, mild, moderate and severe. The obtained results imply that the ensemble method, which was developed based on majority voting among the five deep transfer learning outputs (EfficientNetsB4, EfficientNetsB3, InceptionV3, NasnetMobile, and ResNext50) in the second strategy, has higher results than the first strategy and also single deep transfer learning models with an accuracy metric of 85%. So, this automated model can help physicians to monitor the progression or response of the disease to treatment with an accurate diagnosis of the severity of COVID-19 lung involvement.

Acknowledgement

The article is supported by Research Department of School of Medicine, Shahid Beheshti University of Medical Sciences (Grant No.: 28599).

Ethics declarations

Conflict of interest

The authors declare that they have no conflict of interest.

Ethical approval

All procedures performed in studies involving human participants met the ethical standards of the institutional and/or national research committee and the 1964 Declaration Helsinki and its subsequent amendments or comparable ethical standards. The study was approved by the ethics committee of Shahid Beheshti University of Medical Sciences.

Informed consent

Informed consent was obtained from all individual participants included in the study.

References

1. Global COVID-19 Tracker & Interactive Charts. Available at: https://coronavirus.1point3acres.com/en/world
2. Bernheim A, Mei X, Huang M, Yang Y, Fayad ZA, Zhang N, et al. Chest CT findings in coronavirus disease-19 (COVID-19): relationship to duration of infection. Radiology. 2020;295(3): 200463. https://doi.org/10.1148/radiol.2020200463
3. Zumla A, Chan JF, Azhar EI, Hui DS, Yuen KY. Coronaviruses-drug discovery and therapeutic options. Nat Rev Drug Discov. 2016;15(5):327-347. https://doi.org/10.1038/nrd.2015.37
4. Cote A, Ternacle J, Pibarot P. Early prediction of the risk of severe coronavirus disease 2019: A key step in therapeutic decision making. eBioMedicine. 2020;59:102948. https://doi.org/10.1016/j.ebiom.2020.102948
5. World Health Organization. (2020). Clinical management of COVID-19: interim guidance, 27 May 2020. World Health Organization. Available at: https://apps.who.int/iris/handle/10665/332196
6. Hachim A, Kavian N, Cohen CA, Chin AW, Chu DK, Mok CK, et al. ORF8 and ORF3b antibodies are accurate serological markers of early and late SARS-CoV-2 infection. Nat Immunol. 2020;21(10):1293-1301. https://doi.org/10.1038/s41590-020-0773-7
7. MacMahon H, Naidich DP, Goo JM, Soo Lee K, Leung ANC, Mayo JR, et al. Guidelines for management of incidental pulmonary nodules detected on CT images: From the Fleischner Society 2017. Radiology. 2017;284(1):228-243. https://doi.org/10.1148/radiol.2017161659
8. Koo HJ, Lim S, Choe J, Choi SH, Sung H, Do KH. Radiographic and CT features of viral pneumonia. Radiographics. 2018;38(3):719-739. https://doi.org/10.1148/rg.2018170048
9. Franquet T. Imaging of pulmonary viral pneumonia. Radiology. 2011;260(1):18-39. https://doi.org/10.1148/radiol.11092149
10. Feng Z, Yu Q, Yao S, Luo L, Zhou W, Mao X, et al. Early prediction of disease progression in COVID-19 pneumonia patients with chest CT and clinical characteristics. Nat Commun. 2020;11(1):4968. https://doi.org/10.1038/s41467-020-18786-x
11. Francone M, Iafrate F, Masci GM, Coco S, Cilia F, Manganaro L, et al. Chest CT score in COVID-19 patients: correlation with disease severity and short-term prognosis. Eur Radiol. 2020;30(12):6808-6817. https://doi.org/10.1007/s00330-020-07033-y
12. Matos J, Paparo F, Mussetto I, Bacigalupo L, Veneziano A, Bernardi SP, et al. Evaluation of novel coronavirus disease (COVID-19) using quantitative lung CT and clinical data: prediction of short-term outcome. Eur Radiol Exp. 2020;4(1):39. https://doi.org/10.1186/s41747-020-00167-0

13. Chassagnon G, Vakalopoulou M, Battistella E, Christodoulidis S, Hoang-Thi TN, Dangeard S, et al. AI-driven quantification, staging and outcome prediction of COVID-19 pneumonia. Med Image Anal. 2021;67:101860. https://doi.org/10.1016/j.media.2020.101860
14. Litjens G, Kooi T, Bejnordi BE, Setio AAA, Ciompi F, Ghafoorian M, et al. A survey on deep learning in medical image analysis. Med Image Anal. 2017;42:60-88. https://doi.org/10.1016/j.media.2017.07.005
15. Tan C, Sun F, Kong T, Zhang W, Yang C, Liu C. A survey on deep transfer learning. arXiv:1808.01974v1 [cs.LG]. 2018. https://doi.org/10.48550/arXiv.1808.01974
16. Asgharnezhad H, Shamsi A, Alizadehsani R, et al. Objective evaluation of deep uncertainty predictions for COVID-19 detection. Sci Rep. 2022;12:815. https://doi.org/10.1038/s41598-022-05052-x
17. Alizadehsani R, Sharifrazi D, N Hoseini Izadi. et al. Uncertainty-Aware Semi-Supervised Method Using Large Unlabeled and Limited Labeled COVID-19 Data. arXiv:2102.06388 [eess.IV], 2021. https://doi.org/10.48550/arXiv.2102.06388
18. Ayoobi N, Sharifrazi D, Alizadehsani R, et al. Time series forecasting of new cases and new deaths rate for COVID-19 using deep learning methods. Results in Physics. 2021;27:104495. https://doi.org/10.1016/j.rinp.2021.104495
19. Khozeimeh F, Sharifrazi D, Izadi NH, et al. Combining a convolutional neural network with autoencoders to predict the survival chance of COVID-19 patients. Sci Rep. 2021;11:15343. https://doi.org/10.1038/s41598-021-93543-8
20. Sharifrazi D, Alizadehsani R, Roshanzamir N, et al. Fusion of convolution neural network, support vector machine and Sobel filter for accurate detection of COVID-19 patients using X-ray images. Biomed Signal Process Control. 2021;68102622. https://doi.org/10.1016/j.bspc.2021.102622
21. Shoeibi A, Khodatars M, Alizadehsani R. Automated Detection and Forecasting of COVID-19 using Deep Learning Techniques: A Review. arXiv:2007.10785 [cs.LG]. 2021. https://doi.org/10.48550/arXiv.2007.10785
22. Wang X, Deng X, Fu Q, Zhou Q, Feng J, Ma H, et al. A weakly-supervised framework for COVID-19 classification and lesion localization from chest CT. IEEE Transact Med Imaging. 2020;39(8):2615-2625. https://doi.org/10.1109/TMI.2020.2995965
23. Bansal S, Singh M, Dubey RK, et al. Multi-objective Genetic Algorithm Based Deep Learning Model for Automated COVID-19 Detection Using Medical Image Data. Journal of Medical and Biological Engineering. 2021;41:678-689. https://doi.org/10.1007/s40846-021-00653-9
24. Almalki YE, Qayyum A, Irfan M, et al. A Novel Method for COVID-19 Diagnosis Using Artificial Intelligence in Chest X-ray Images. Healthcare (Basel). 2021;9(5):522. https://doi.org/10.3390%2Fhealthcare9050522
25. Irfan M, Iftikhar MA, Yasin S, et al. Role of Hybrid Deep Neural Networks (HDNNs), Computed Tomography, and Chest X-rays for the Detection of COVID-19. Int J Environ Res Public Health. 2021;18(6):3056. https://doi.org/10.3390%2Fijerph18063056
26. Amini N, Shalbaf A. Automatic classification of severity of COVID-19 patients using texture feature and random forest based on computed tomography images. Int J Imaging Syst Technol. 2022;32(1):102-110. https://doi.org/10.1002/ima.22679
27. Pan F, Ye T, Sun P, Gui S, Liang B, Li L, et al. Time course of lung changes on chest CT during recovery from 2019 novel coronavirus (COVID-19) pneumonia. Radiology 2020:295(3):715-721. https://doi.org/10.1148/radiol.2020200370
28. Huang L, Han R, Ai T, Yu P, Kang H, Tao Q, Xia L, et al. Serial quantitative chest CT assessment of COVID-19: deep-learning approach. Radiol Cardiothorac Imaging. 2020;2(2):e200075. https://doi.org/10.1148/ryct.2020200075
29. Shan F, Gao Y, Wang J, Shi W, Shi N, Han M, et al. Lung infection quantification of COVID-19 in CT images with deep learning. arXiv:2003.04655. 2020. https://doi.org/10.48550/arXiv.2003.04655
30. Ghosh B, Kumar N, Singh N, Sadhu AK, Ghosh N, Mitra P, et al. A quantitative lung computed tomography image feature for multi-center severity assessment of COVID-19. medRxiv 2020.07.13.20152231. https://doi.org/10.1101/2020.07.13.20152231
31. Shen C, Yu N, Cai S, Zhou J, Sheng J, Liu K, et al. Quantitative computed tomography analysis for stratifying the severity of coronavirus disease 2019. J Pharmaceut Analysis 2020;10(2):123-129. https://doi.org/10.1016/j.jpha.2020.03.004
32. Otsu N. A threshold selection method from gray-level histograms. IEEE Trans Sys Man Cyber. 1979;9(1):62-66. https://doi.org/10.1109/TSMC.1979.4310076
33. Tajbakhsh N, Shin JY, Gurudu SR, Hurst RT, Kendall CB, Gotway MB et al. Convolutional neural networks for medical image analysis: full training or fine tuning?. IEEE Transact Med Imaging 2016;35(5):1299-1312. https://doi.org/10.1109/TMI.2016.2535302
34. Deng J, Dong W, Socher R, Li L, Li K, Fei-Fei L. ImageNet: A large-scale hierarchical image database. 2009 IEEE Conference on Computer Vision and Pattern Recognition. 2009:248-255. https://doi.org/10.1109/CVPR.2009.5206848
35. Khan A, Sohail A, Zahoora U, Qureshi AS, et al. A survey of the recent architectures of deep convolutional neural networks. Artif Intell Rev. 2020;53:5455-5516. https://doi.org/10.1007/s10462-020-09825-6
36. Tang Z, Zhao W, Xie X, Zhong Z, Shi F, Ma T, et al. Severity assessment of COVID-19 using CT image features and laboratory indices. Phys Med Biol. 2021;66(3):035015. https://doi.org/10.1088/1361-6560/abbf9e