Turkish Journal of Electrical Engineering & Computer Sciences

http://journals.tubitak.gov.tr/elektrik/

Research Article

Turk J Elec Eng & Comp Sci
(2021) 29: 2663 – 2679
© TÜBİTAK
doi:10.3906/elk-2104-184

A deep transfer learning based model for automatic detection of COVID-19 from chest X-rays

Prateek CHHIKARA*, **Prakhar GUPTA**, **Prabhjot SINGH**, **Tarunpreet BHATIA**
Department of Computer Science & Engineering, Thapar Institute of Engineering and Technology, Punjab, India

Received: 26.04.2021 • **Accepted/Published Online:** 06.08.2021 • **Final Version:** 04.10.2021

Abstract: Deep learning in medical imaging has revolutionized the way we interpret medical data, as high computational devices' capabilities are far more than their creators. With the pandemic causing havoc for the second straight year, the findings in our paper will allow researchers worldwide to use and create state-of-the-art models to detect affected persons before it reaches the R number. The paper proposes an automated diagnostic tool using the deep learning models on chest x-rays as an input to reach a point where we surpass this pandemic (COVID-19 disease). A deep transfer learning-based model for automatic detection of COVID-19 from chest x-rays using the Inception-V3 model is proposed, in which we added flattening, node dropping, normalization, and dense layer. The proposed architecture is compared with existing state-of-the-art ImageNet models. The model's efficacy is tested on three different COVID-19 radiography datasets with three classes: COVID, normal, and viral pneumonia. The proposed model has reached an accuracy of 97.7%, 84.95%, and 97.03% on the mentioned datasets, respectively. The proposed work introduces the deep neural networks applied to medical images to analyze image enhancement techniques and emphasize the field's clinical aspects.

Key words: COVID-19, chest X-ray, computer vision, digital ımage processing, transfer learning

1. Introduction

The year 2020 has been harsh on the whole world and will be remembered throughout history as a year like no other; the world has witnessed a novel coronavirus (SARS-CoV-2 or COVID-19). The pandemic's impact is not limited to millions of lives lost. It has vast implications on mental health, a state of constant fear, economic depression and social disruption, and much more. The COVID-19 virus infiltrates the lungs causing respiratory difficulties. In difficult situations, a patient may have pneumonia, which leads to the accumulation of fluid and pus in the lungs' air sacs. In most cases, the COVID-19 virus shows only moderate signs or the patient may be asymptomatic [1]. Still, it can be life-threatening in advanced cases, particularly for those already suffering from specific health issues. Treatment of COVID-19 does not make the virus go away; instead, it only helps relieve symptoms. Doctors typically test blood and sputum specimens to diagnose COVID-19 by adopting a reverse transcription polymerase chain reaction (RT-PCR) test, which examines traces of the virus's biogenetic material in the patient's blood. Being a susceptible test, the remains of the virus does not show up well. The outcomes of the various current research suggest that chest computer tomography (CT) performs better in screening for

*Correspondence: pchhikara_be16@thapar.edu

This work is licensed under a Creative Commons Attribution 4.0 International License.

COVID-19 than lab testing[1]. According to[2], in comparison to RT-PCR, chest CT imaging is a more reliable, efficient, and fast process in COVID-19 diagnosis. For the diagnosis of pneumonia, doctors routinely use chest CT images. A detailed 3D image of the lung is created using CT scans. Doctors examine the chest CT images, viewing for fluid or pus in the lungs or other particular COVID-19 infection symptoms. Chest CTs are relatively faster and comparatively simple to implement, and are painless. They are also more sensitive to COVID-19 infection [2].

Ever since technology leveraged in medical science, medical devices and diagnosis have reached a different level. The state-of-the-art IoT devices have made complex tasks much more straightforward and have helped in real-time monitoring. Technology is improving all the time. Today, with the help of deep learning, researchers can generate a diverse and wide range of extensive feature sets, which would be difficult even for a specialist [3]. Machine learning (ML) and deep learning (DL) have earned many applications over the past couple of decades and have proved to be a boon in solving compound medical use-cases. ML algorithms require structured data to understand, for example, to differentiate between pneumonia-infected lungs and healthy lungs; it requires a feature set to predict the output. On the contrary, in DL, the classifier classifies both pneumonia and normal images through the automatically extracted features from the network. There is no requirement of feeding a handcrafted filter. With the power of generating features independently, DL has become a powerful tool where we do not have to create features manually. Over the years, DL algorithms have gone through many improvements and are prominently used to detect abnormalities in medical images. We have DL algorithms that can use previously learned knowledge to solve the new one. The symptoms of COVID-19 are similar to severe pneumonia. Since we already have success in detecting pneumonia using chest x-rays, e.g., CheXNet, a 121-layer convolutional neural network (CNN) that outputs the confidence score of a diagnosis based on an input image of a chest x-ray. The model is trained on more than 100,000 images and produces results better than practicing radiologists. With slight modification, it can detect 14 pathological diseases in chest x-rays using the most prominent openly accessible chest x-ray dataset [4]. DL's applicability in AI includes millions of data entries and complicated model training to perform better results. Finding ways to increase neural network efficiency is a critical challenge. This paper discusses how DL is used to detect COVID-19 disease using chest x-ray images as input and how our model overcomes the problems described in previous research. With automation at the level of experts, we expect that this technology can contribute to the testing of COVID-19 patients and put a pause on this pandemic. The major contributions of the paper are as follows:

1. We designed a pre-processing pipeline for enhancing the essential features of the input chest x-ray images. For the enhancement, a Laplacian filter is used to detect the edges in the image. To remove the noise in the chest x-ray images, we have used denoising with the help of an average filter. For the contrast enhancement, we have used an advanced histogram equalization method. Also, for the brightness part, we have used the gamma correction.
2. CNN architecture has been proposed, which uses InceptionV3 as a backbone. The output of the InceptionV3 is flattened, and then we add three dense layers having the weights learned from our image dataset. The final output layer contains three nodes signifying the number of classes of chest x-rays (COVID, normal, or viral pneumonia).

[1]ScienceDaily. CT Provides Best Diagnosis For COVID-19 (2020). Website https://www.sciencedaily.com/releases/2020/02/200226151951.htm [accessed 25 Sep 2020]

[2]Rsna.org. CT Provides Best Diagnosis For COVID-19 (2020). Website https://www.rsna.org/en/news/2020/February/COVID-19-CT-Diagnosis-Study [accessed September 25, 2020]

3. The efficacy of the proposed model is compared with existing state-of-the-art ImageNet models on three different chest x-ray datasets. Finally, the evaluation is performed using various classification metrics like precision, recall, f1-score, etc.

The paper is organized as follows. The previous research work is described in Section 2. Section 3 mentions the used dataset and basic techniques. The proposed approach of this paper is described in Section 4. In Section 5, the obtained results are discussed and compared with existing state-of-the-art models. Finally, Section 6 concludes the paper with future directions.

2. Related work

This section illustrated some latest related research work in image enhancement using pre-processing steps and DL approaches to detect COVID-19 from the chest x-ray images.

2.1. Image pre-processing

Many researchers have used a wide range of digital image processing (DIP) techniques before feeding the images to DL-based methods to detect COVID-19. Tartaglione *et al.* [5] applied histogram equalization to get uniform image dynamics in the dataset. Lung segmentation was applied to the pre-processed images, which retained only areas of interest in an image and discarded the rest of the portion. Later, image intensity was normalized between 0 and 1. Asnaoui *et al.* [6] used image datasets of CXR and CT scans and applied CLAHE to the images as a pre-processing step before applying the DL models. They used transfer learning on several DL models and did a comparative study of these models. Abbas *et al.* [7] applied different image augmentation techniques such as translation, rotation, and flipping to overcome the limitation of a small dataset. Histogram modification techniques were applied to enhance image contrast. Heidari *et al.* [8] used image processing algorithms to find the diaphragm in a chest x-ray image and remove it for further analysis. They have used morphological filters (like open, close, and dilate) followed by two noise removal filters and histogram equalization before feeding images to CNN. Siddhartha *et al.* [9] applied white balancing and CLAHE to enhance the visibility of chest x-ray images. White balancing was used to overcome the low lighting conditions in medical images by fixing color fidelity in digital images. Finally, they applied a depth-wise separable CNN (DSCNN) to make their COVIDLite model. Siracusano *et al.* [10] used a fast and adaptive bidimensional decomposition and CLAHE for enhancing the necessary features from the chest x-rays. They used it in the medical domain for image retrieval and fusion. It is applied to rectify the luminance contributions but preserve other details and information.

2.2. DL approaches to detect COVID-19

Researchers have started focusing on DL algorithms in detecting COVID-19 from chest x-rays. Das *et al.* [11] stated that using a DL-based network in place of CT scan systems is a cheaper alternative. They used a truncated Inception network as a screening tool to segregate positive images from negative ones. Ozturk *et al.* [12] created a binary classifier using the DarkNet architecture for detecting positive and negative COVID-19 chest x-ray images and a multi-class classifier to detect COVID-19, pneumonia, and normal images. Khan *et al.* [13] used the Xception model as a pre-trained network on the ImageNet dataset. The testing was performed on publically available datasets. Apostolopoulos *et al.* [14] used MobileNet to train a model from scratch and extracted the features for the classification task. Ucar *et al.* [15] used the SqueezeNet model tuned on the COVID-19 diagnosis using the Bayesian optimization technique. Augmentation is also performed on the image

dataset for better performance. Loey *et al.* [16] used generative adversarial networks (GANs) and transfer learning for constructing the model. GAN helped generate 30 times more images from the original dataset and helped overcome the overfitting problem by making the proposed model more robust. Luz *et al.* [17] extended the EfficientNet model using chest x-ray images to perform COVID-19 detection. They reduced the number of parameters by a factor of 30 compared to the baseline research model and used 5 and 28 times less parameters than ResNet50 and VGG16, respectively. Shan *et al.* [18] developed a DL segmentation system that automatically distinguishes infection regions of interest and their volumetric ratios in the lungs. Model's performance is evaluated by analyzing the manually detected infection regions with the automatically segmented ones on around 300 chest CT scans images. Ghoshal *et al.* [26] reviewed how Bayesian CNN based on drop-weights can predict the uncertainty in DL use cases. The expected uncertainty is used to increase the human-machine unit's diagnostic performance using an openly accessible COVID-19 chest x-ray dataset and discussed the uncertainty in prediction and prediction accuracy that are highly correlated. Apostolopoulos *et al.* [21] used two datasets in the experiment. The results imply that DL with x-ray images as input extracts significant features from the images. Moreover, Table 1 illustrates the use of AI-based ML and DL methods for COVID-19 detection using chest x-ray images as an input. The proposed work handles the limitations mentioned in Table 1. Following are the significant limitations observed in the related work and are handled by the proposed approach:

Table 1. An overview of AI-based ML and DL methods for detection of COVID-19 using chest x-rays.

Authors	**Methods Used**	**Evaluation Metrics**	**Research Limitations**
Narin *et al.* [19]	Deep ConvNet Models such as ResNet50, InceptionV3 and Inception-ResNetV2	Accuracy, ROC, and Confusion Matrices	They have used a dataset with fewer images. The performance could be better if they selected a large dataset. The used dataset contains 50 COVID-19 chest X-ray and 50 for non-COVID-19.
Huang *et al.* [20]	CNN Models	Opacification %	The correlation between the COVID-19 virus and the opacities of CT lung images is not incorporated.
Apostolopoulos *et al.* [21]	CNN models using Transfer Learning	Accuracy, Sensitivity and Specificity	The proposed model should be able to distinguish between the COVID-19 class and the pneumonia class.

Table 1. (Continued).

Authors	Methods Used	Evaluation Metrics	Research Limitations
Loey *et al.* [16]	Combined GAN with AlexNet, GoogLeNet, and ResNet18	Precision, recall, f1-score, and accuracy	A smaller dataset is used which contains a total of 306 images are present in the four classes: COVID-19, normal, pneumonia bacterial, and pneumonia virus.
Maia *et al.* [22]	Different Convolutional Support Vector Machines	Accuracy, recall, Matthew's correlation, and f1-score	A dataset with fewer images was selected, having a total of 437 X-ray images (217 COVID-19, 108 other diseases, and 112 healthy).
Bai *et al.* [23]	Multi-layered perceptron with LSTM	Accuracy, precision, recall, and AUC	They combined both LSTM and multi-layered perceptron, which will increase the number of parameters and thus the inference time will be high.
Ozturk *et al.* [12]	DarkCovidNet Architecture	Binary accuracy, multi-class accuracy, precision, and recall	Used a very small sample set of COVID-19 X-ray images containing 127 images.
Sun *et al.* [24]	Support Vector Machine	AUC and recall	A detailed comparison is not made between the proposed classifier and the existing ones.
Ibrahim *et al.* [25]	AlexNet	Accuracy, sensitivity, and specificity	A detailed comparison is not made between the proposed classifier and the existing ones.

1. Most researchers have used datasets with fewer images. Hence, due to fewer training samples, the model is not generalized, and there is a possibility that the model has overfitted on the training samples.
2. Many related works have not used the pre-processing step to enhance the image features, which is crucial in the computer vision pipeline.
3. The comparison is made based on a single dataset, which is not a best practice to prove the model's

efficacy.

4. A separate comparison has not been made with other state-of-the-art computer vision models.

3. Material and methods

This section provides a concise explanation of the used dataset and the techniques during this research.

3.1. Dataset

The following three datasets have been used (where 'C' is COVID-19, 'N' is normal and 'P' is viral pneumonia class).

1. **Dataset 1:**[3] It consists of three classes. The total number of images are 576, 1583, and 4273, respectively, for 'C', 'N', and 'P' classes. The dataset is divided into two parts training and testing in the ratio of 4:1. This gives 116 images in the test set of the 'C' class. To make the test set exactly equal, we use 116 images of each of the three classes. Image augmentation is applied for the training set such that the images in the training set are almost equal. The augmentation of eight and three times is performed on 'C' and 'N' classes.

2. **Dataset 2:**[4] The dataset contains a total of 6871 chest x-ray images. The dataset is divided into train and test sets, with the test set being 20% (1375 images). The test set includes 449, 463, and 463 images of 'C', 'N', and 'P' class.

3. **Dataset 3:**[5] This dataset contains only two classes, 'C' and 'N', with 7593 and 6893 images. The dataset is divided into train and test in the ratio 4:1. The test set contains 1519 and 1379 images for 'C' and 'N' classes.

For each image in the datasets, the images have been resized to 224×224, keeping RGB channels intact. The distribution of the dataset is shown in Table 2.

Table 2. Dataset distribution into train-test.

Dataset	Classes	COVID-19	Normal	Pneumonia
Dataset1	Train	3680	3798	3418
	Test	116	116	116
Dataset2	Train	2245	2313	2313
	Test	449	463	463
Dataset3	Train	6074	5514	-
	Test	1519	1379	-

[3]Patel P. Chest X-ray (Covid-19 and Pneumonia) 2020. Website https://www.kaggle.com/prashant268/chest-xray-covid19-pneumonia [accessed 7 Nov 2020]

[4]Asraf A. COVID19 Pneumonia Normal Chest Xray PA Dataset 2020. Website https://www.kaggle.com/amanullahasraf/covid19-pneumonia-normal-chest-xray-pa-dataset [accessed 10 May 2021]

[5]Maftouni M. Large COVID-19 CT scan slice dataset 2021. Website https://www.kaggle.com/maedemaftouni/large-covid19-ct-slice-dataset [accessed 7 May 2021]

3.2. Transfer learning

In the bioinformatics domain, the development is limited because of the non-availability of large-scale structured datasets, mainly because of the expense of data acquisition. Transfer learning solves this problem of limited availability of training data. In transfer learning, learning is the result of the transfer of knowledge from a similar model that is trained to perform similar tasks [27]. It is a methodology where weights and features from a trained model are used to train and fine-tune the new model. In our problem, some low-level features, such as intensity, edges, corners, and shapes, can be learned through already trained models, and hence for our task, transfer learning is among the best available options.

3.3. Data augmentation

In bioinformatics, data acquisition is costly, and there is limited availability of large-scale well-annotated datasets. Limited data hinders the performance of DL models such as CNN. Image augmentation is one of the valuable techniques that solve the problem of limited availability of data. It increases the datasets' size by artificially creating images through different processing techniques, such as random rotation, flips, shifts, image resizes, and image processing algorithms, such as histogram equalization, applying random noise to the image throughout. Image augmentation can solve class imbalance, as there may not be sufficient data for some classes that can be generated using image augmentation[6].

3.4. Feature extraction using CNN

Traditionally data scientists create feature sets manually from domain knowledge. However, this is tedious, time-consuming, suffers from human bias, and almost impossible for datasets consisting of images. Using DL algorithms, extracting as much information as possible from the available datasets is crucial for creating an effective solution. A CNN has two essential parts, (a) feature extractor and (b) classifier. The last sigmoid/logistic dense layer concerning the target variable is removed to use CNN as a feature extractor. This feature extractor can be used for any new business problem data without retraining the new data model. In many cases, the last layer needs to be changed, and then feed forwarding the network adjusts the weights, and it will directly map final weights to calculate features [28]. CNN-based feature extractor provides the flexibility based on our requirement of several features and features at any dense intermediate layer with the desired dimensions.

4. Proposed work

This research aims to develop an automated diagnostic tool using a deep neural network that classifies x-ray images as viral pneumonia, COVID-19, or normal. Initially, DIP algorithms are used for the preprocessing of the images. Moreover, the dataset is split into training and testing sets. The preprocessed train data is used as an input to the model during the training stage. Furthermore, the model's performance is evaluated on the test set.

4.1. Dataset preprocessing using DIP

Medical images are often degraded due to noise by several sources of intrusion that affect the acquisition of medical images. Image enhancement methods are mathematical tools that improve image quality. The result

[6]ODSC - Open Data Science. Image Augmentation for Convolutional Neural Networks (2019). Website https://medium.com/@ODSC/image-augmentation-for-convolutional-neural-networks-18319e1291c [accessed 26 Sep 2020]

is a different image that exhibits specific features that are better corresponding to the original image. Image enhancement methods are widely used in fields where images' subjective quality is necessary during diagnosis, such as radiology. Low-contrast structures demand enhancement in all kinds of medical images. In examining x-rays, capturing high contrast in the raw image directly from the imaging device is expensive. The low contrast in an image is viewed as a consequence of poor pixel intensities spread over the display device's dynamic range. Few improvements in the medical images' visual quality may assist in better interpretation than a medical specialist. Therefore, DIP performs an essential part in enhancing the image quality before feeding it to a deep neural network [29].

Better image enhancement is not achieved using a single process. Preferably a combination of several techniques is used to achieve the final result. The following DIP tools are used in combinations to enhance our dataset's quality, as shown in Figure 1. First, we apply the Laplace filter to the input image (1a). The Laplace filter result (1b) is not an enhanced image; we subtract the Laplace result from the original image to get the final sharpened image (1c). We then apply the Sobel filter on (1a) to get (1d). Further to (1d), we apply an average filter for smoothing. Now we create a mask by multiplying (1c) and (1e). This mask is then added to the original image to get a sharpened image. For accurate reproduction of colors in the image, gamma correction is done (1g). At last, to improve the image contrast, we use CLAHE (1h).

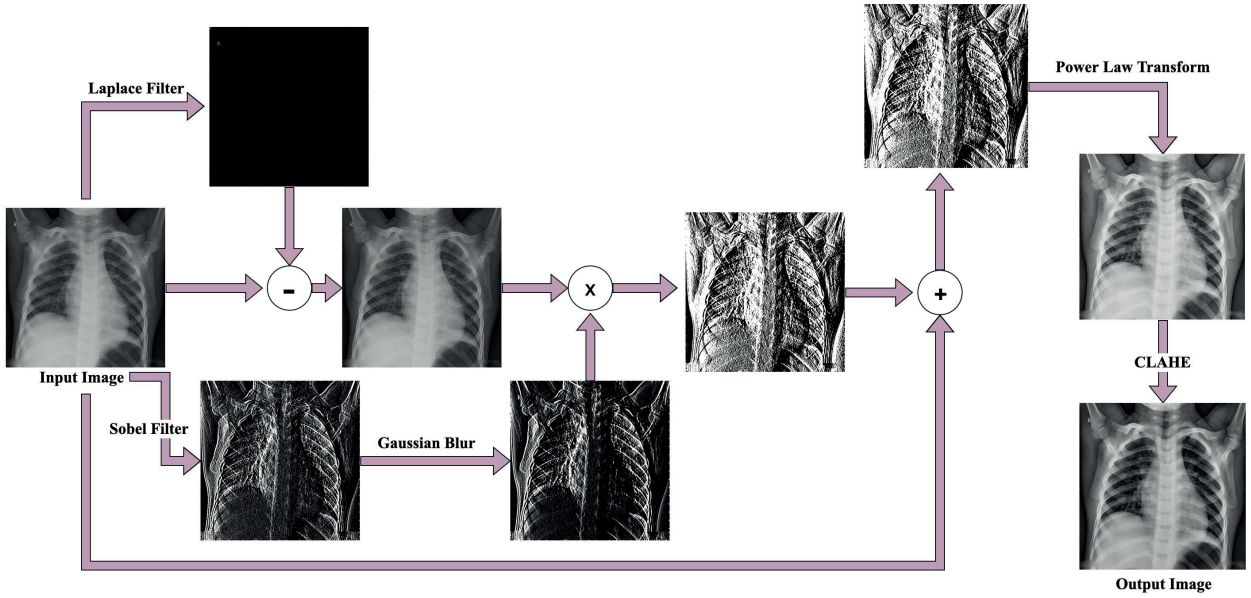

Figure 1. Dataset pre-processing.

4.1.1. Laplacian Filter

It is a 2-D isotropic finite ımpulse response (FIR) edge detector filter that highlights rapid intensity change regions. It enhances the image by calculating the 2-D spatial derivative of an image[7].

$$L(x,y) = \frac{\partial^2 I}{\partial x^2} + \frac{\partial^2 I}{\partial y^2} \tag{1}$$

[7]Fisher R, Perkins S, Walker A, Wolfart E. Laplacian/Laplacian of Gaussian (2003). Website http://homepages.inf.ed.ac.uk/rbf/HIPR2/log.htm [accessed 4 Dec 2020]

where L(x,y) is the Laplacian of an Image with pixel Intensity values I(x,y).

4.1.2. Sobel filter

It is a discrete differentiation operator that enhances the edges by emphasizing the high spatial frequency regions corresponding to edges. The Sobel kernel is expressed as a product of an averaging and a differentiation kernel; it also smooths the image while calculating gradients. Hence, it is less sensitive to noise than the Robert Cross filter.

$$G_x = \begin{bmatrix} +1 & 0 & --1 \\ +2 & 0 & --2 \\ +1 & 0 & --1 \end{bmatrix} * A \tag{2}$$

can be written as:

$$G_x = \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix} * (\begin{bmatrix} +1 & 0 & --1 \end{bmatrix} * A) \tag{3}$$

Where G_x is the derivative approximation of a point 'x' along x-axis, 'A' is the source image and '*' is the convolution operator.

4.1.3. Average filter

Average or mean filter is the simplest low pass filter. It is a smoothing filter that reduces the image's noise by lowering the amount of intensity variations between two neighboring pixels.

4.1.4. Gamma transformation

Many medical imaging devices have non-linear luminance that affects the performance of image processing algorithms. Gamma correction is applied to correct the images that are changed by devices that respond to Power law.

$$I_{out} = c.I_{in}^{\gamma} \tag{4}$$

where I_{out} is the output image, I_{in} is the input image, c and γ are positive constants.

4.1.5. Contrast-limited adaptive histogram equalization (CLAHE)

It enhances the contrast of an image by dividing the image into small parts to exploit local spatial coherence. It is the advancement to adaptive histogram equalization (AHE), which, unlike AHE, does not overamplify noise. In CLAHE, different regions of an image are combined once they are enhanced, and induced edges are handled by interpolation [30].

4.2. Model training

The proposed model is based on the well-known InceptionV3 model, which is used as a backbone to obtain compelling features from the input images. It allows us to leverage the benefits of the InceptionV3 model. The concern of the availability of limited data is mitigated by using transfer learning. The InceptionV3 network

consists of 311 layers with 94 layers of convolutional, batch normalization, ReLU activation layers, and a few pooling layers. The pre-trained layers of InceptionV3 are used as it is without any modification. The output of the pre-trained layers is flattened with a dropout of 0.4. After that, we have stacked three fully connected dense layers with nodes 128 (dropout = 0.5), 64 (dropout = 0.5), and 64, respectively, after flattening. These three fully connected dense layers contain 'ReLU' as an activation function. Considering our problem, the last dense layer with 1000 nodes of the pre-trained model is removed. In our case, the number of classes to be predicted is two/three (for binary and three class classification). Therefore, we add a final output fully connected layer having two/three nodes using 'sigmoid' as an activation function. The weights are randomly initialized in the network. To boost the proposed model's training efficiency, we have used ındependent component (IC) layer before each weight layer [31]. IC layer consists of two layers, batch normalization, and dropout which linearly decrease the similarity among pairs of neurons concerning the dropout layer parameter 'p'. As shown empirically, the IC layer beats the baseline strategies with a more steady training process and faster convergence speed. The problem of overfitting is handled by introducing the dropout and batch normalization layer. The plot for the added layers in the InceptionV3 architecture is shown in Figure 2. Figure 3 shows the detailed version of 11 layers that were added, including respective input-output dimensions.

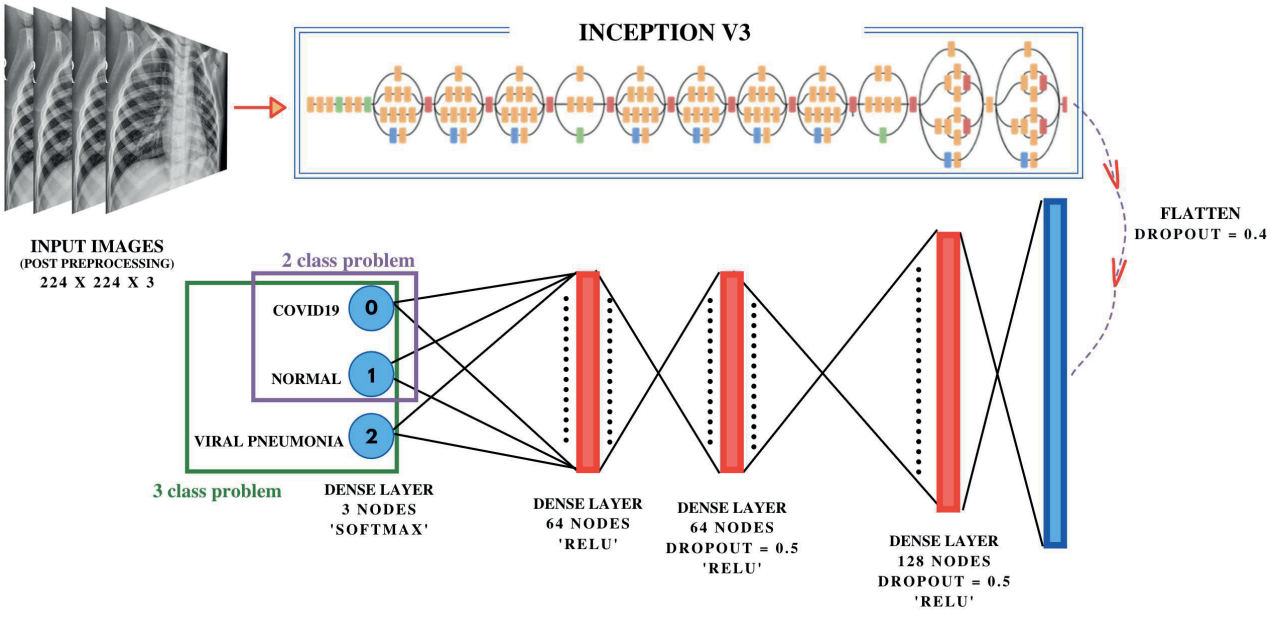

Figure 2. Architecture of the proposed model.

The final architecture is 322 layered with total parameters equal to 28,382,755. Training is accomplished by taking 'categorical crossentropy' as the loss measure, which is minimized using the 'adam' optimizer. For the training, the batch size is selected as 64. The learning rate is reduced when there is no significant reduction in the loss value. All activation layers (except the final dense layer) use ReLU as the activation function. 'Softmax' is used as the activation function in the last output layer. If a significant increase in accuracy is not seen until five epochs, the learning rate is reduced by 10%. The model is trained for 50 epochs, and the model, which gives maximum accuracy, is saved. The model is trained by keeping the layers of the InceptionV3 frozen. Freezing means the weights for these layers are not updated whenever the model is retrained on new data for a new

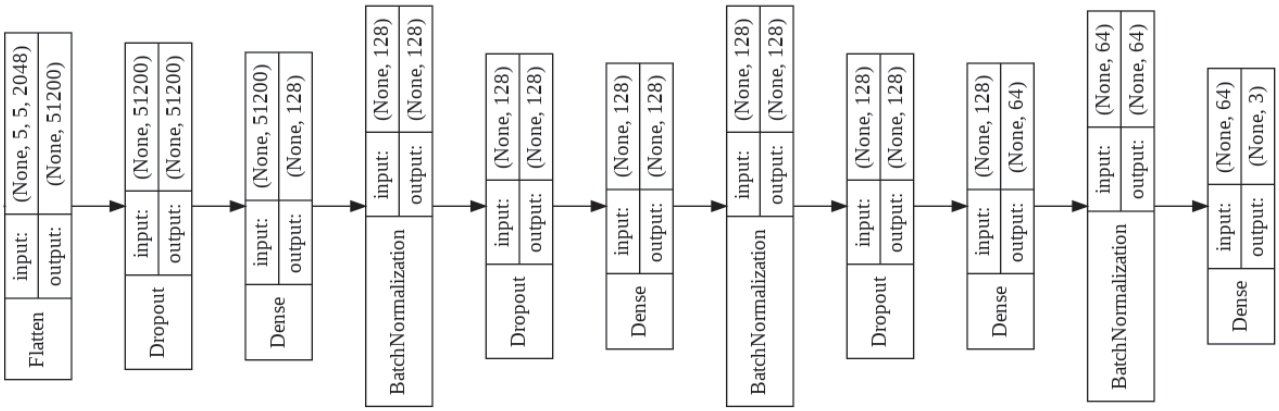

Figure 3. Plot of the added layers on the InceptionV3 network.

task. Then, the input images are fed to this model, but backpropagation is stopped at the first added layer by freezing other layers. This is done to prevent backpropagation from destroying these essential features from the base model. The learning rate is kept small during training, so each fully connected layer can learn features from the previously learned layers in the network. After the fully connected layer has started to learn from our dataset, we unfreeze the complete network in the second step, and training is continued with a small learning rate. This prevents drastic changes in already learned features and modifying weights of last layers containing features specific to the task at hand. Training is done until the desired accuracy is obtained.

5. Experimentation and results

Google collaboratory is used to achieve significant results, which offers free 1xTesla K80 GPU has 2496 CUDA cores and 12GB GDDR5 VRAM. GPU helped in parallel processing, and thus training time is reduced by a high margin. All the models are trained similarly for 50 epochs. The training set is divided into training and validation set in the ratio 85:15. The hyper-parameters are tuned using the validation set. For evaluating the model's potential, multiclass classification metrics are used as a benchmark, such as precision, recall, and f1-score, based on the confusion matrix. Classification accuracy is generally used because it tends to summarize model performance. On the other hand, f1-score provides a way to combine precisions and recall into a single measure that captures both properties. In the proposed work, we created two models P_P and P_R. The P_P is trained and tested using the pre-processed images, whereas the P_R model is trained and tested on the raw images. The performance evaluation of the proposed models with other state-of-the-art models is done separately on each of the datasets.

5.1. Performance evaluation of models on Dataset 1

On Dataset 1, P_P shows the highest overall test accuracy of 97.70%, followed by VGG19 having a test accuracy of 97.41%. VGG16 model also shows the test accuracy of almost 97%. The MobileNet and DenseNet121 model shows the least test accuracy. If we consider individual class accuracy, for the COVID-19 class, P_P and VGG19 offer the perfect 100% accuracy. Also, the accuracy of the other two classes is almost similar for both P_P and VGG19. For Dataset 1, both P_P and VGG19 perform the best. The detailed comparison of all proposed models with existing models using Dataset 1 is shown in Table 3. The confusion matrix of the proposed model on the test Dataset 1 is shown in Figure 4.

Table 3. Evaluation of models on dataset 1 with respect to classification metrics (where, P is precision, R is recall and f1 is f1-score).

Model	Class	n (truths)	n (classified)	Acc (%)	P	R	f1	Acc_overall (%)
Inception V3	0	116	110	98.28	1.00	0.95	0.97	95.54
	1	116	114	95.40	0.94	0.92	0.93	
	2	116	124	95.40	0.90	0.97	0.93	
VGG16	0	116	114	99.43	1.00	0.98	0.99	96.83
	1	116	112	97.13	0.97	0.94	0.96	
	2	116	122	97.13	0.93	0.98	0.96	
ResNet50	0	116	117	98.56	0.97	0.96	0.98	95.11
	1	116	119	96.26	0.93	0.98	0.94	
	2	116	112	95.40	0.95	0.91	0.93	
VGG19	0	116	116	100	1.00	1.00	1.00	97.41
	1	116	119	97.41	0.85	0.97	0.96	
	2	116	113	97.41	0.97	0.95	0.96	
MobileNet	0	116	110	98.28	1.00	0.95	0.97	94.25
	1	116	115	95.11	0.93	0.92	0.93	
	2	116	123	95.11	0.9	0.96	0.93	
InceptionResNetV2	0	116	112	98.85	1.00	0.97	0.98	95.68
	1	116	111	96.84	0.97	0.93	0.95	
	2	116	125	95.69	0.90	0.97	0.94	
DenseNet121	0	116	112	98.85	1.00	0.97	0.98	93.96
	1	116	108	94.25	0.94	0.88	0.91	
	2	116	128	94.83	0.88	0.97	0.93	
Proposed Model (P_R) (without image pre-processing)	0	116	114	99.43	1.00	0.98	0.99	95.11
	1	116	108	95.4	0.96	0.90	0.93	
	2	116	126	95.4	0.90	0.97	0.93	
Proposed Model (P_P) (with image pre-processing)	0	116	116	100	1.00	1.00	1.00	97.70
	1	116	112	97.70	0.98	0.95	0.96	
	2	116	120	97.70	0.95	0.98	0.97	

5.2. Performance evaluation of models on Dataset 2

On Dataset 2, P_P, P_R, VGG19, and VGG16, all four models show an overall accuracy above 84%. And out of these four models, P_P outperforms the rest models with an overall accuracy of 84.95%. The most petite performing models on Dataset 2 are MobileNet and DenseNet121, with an overall accuracy of 82.18% and 83.78%, respectively. If the accuracy of COVID-19 is concerned, P_P, P_R, VGG19, VGG16, and DenseNet121, all the five models show almost the same accuracy of above 96%. Table 4 illustrates a detailed comparison of all proposed models with existing models using Dataset 2. The confusion matrix of the proposed model on the test Dataset 2 is shown in Figure 4.

5.3. Performance evaluation of models on Dataset 3

Dataset 3 contains only two classes: COVID-19 and Normal. On the test set, the P_P model outperforms all the models in terms of test accuracy of 97.03%, followed by P_R and VGG16. DenseNet121 performs the worst with an overall test accuracy of 92.34%. The detailed comparison is performed in Table 5, and the confusion

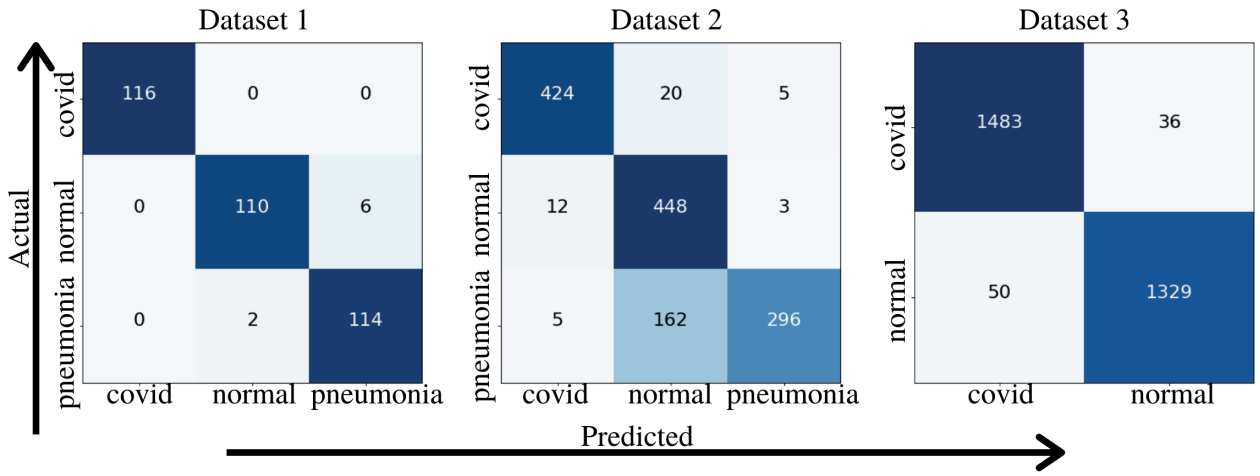

Figure 4. Confusion matrix of the proposed model on the three used datasets.

matrix of the P_P model is shown in Figure 4.

5.4. Receiver operating characteristic (ROC) curve analysis of the proposed model on the three datasets.

The ROC curve is used to get a complete insight into the model performance. The curve plots true positive and false-positive rates for a model by changing the values of different threshold values. The curve's steepness is important since our goal is to maximize the true positive rate and minimize the false positive rate. ROC curves are used in binary classification problems for the analysis of the output of a classifier. ROC curve is extended to multi-label classification by converting the output in binary form; then, one ROC curve is drawn for one label. There are two evaluation metrics in the ROC curve, (a) Micro averaging: weighs individual element of the class indicator matrix as a binary prediction, (b) Macro averaging: gives equivalent weight to each label. The area under the curve corresponds to model performance; the area with the highest magnitude determines the test is better. The ROC curve for the P_P model is shown in Figure 5. The average AUC for P_P (Figure 5) is 98%, 89%, and 97% for the dataset 1, 2 and 3, respectively. The results show that the proposed model outperforms others in terms of performance.

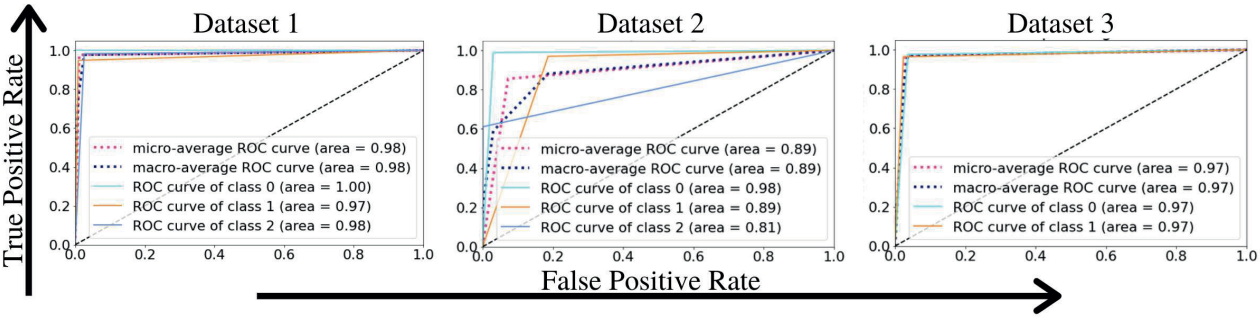

Figure 5. ROC curve of the proposed model on the three used datasets.

Table 4. Evaluation of models on dataset 2 with respect to classification metrics (where, P is precision, R is recall and f1 is f1-score).

Model	Class	n (truths)	n (classified)	Acc (%)	P	R	f1	Acc_overall (%)
Inception V3	0	449	445	91.27	0.87	0.86	0.87	79.42
	1	463	603	82.40	0.88	0.89	0.77	
	2	463	327	85.16	0.90	0.63	0.74	
VGG16	0	449	444	96.87	0.96	0.95	0.95	84.22
	1	463	648	84.95	0.71	0.98	0.81	
	2	463	283	86.62	0.99	0.61	0.75	
ResNet50	0	449	694	74.76	0.57	0.89	0.70	64.58
	1	463	300	72.44	0.64	0.41	0.50	
	2	463	381	81.96	0.78	0.64	0.71	
VGG19	0	449	485	96.22	0.91	0.98	0.94	84.29
	1	463	608	85.67	0.72	0.94	0.82	
	2	463	282	86.69	1.00	0.61	0.75	
MobileNet	0	449	469	95.64	0.91	0.96	0.93	82.18
	1	463	641	83.13	0.68	0.94	0.79	
	2	463	265	85.60	1.00	0.57	0.73	
InceptionResNetV2	0	449	426	93.38	0.92	0.87	0.90	78.33
	1	463	712	79.71	0.63	0.97	0.76	
	2	463	237	83.56	1.00	0.51	0.68	
DenseNet121	0	449	438	96.44	0.96	0.93	0.94	83.78
	1	463	649	84.44	0.69	0.97	0.81	
	2	463	288	86.69	0.99	0.61	0.76	
Proposed Model (P_R) (without image pre-processing)	0	449	438	96.87	0.96	0.94	0.95	84.36
	1	463	643	85.02	0.70	0.97	0.81	
	2	463	294	86.84	0.98	0.62	0.76	
Proposed Model (P_P) (with image pre-processing)	0	449	441	96.95	0.96	0.94	0.95	84.95
	1	463	630	85.67	0.71	0.97	0.82	
	2	463	304	87.27	0.97	0.64	0.77	

6. Conclusion

An attempt to find the best automated diagnostic tool for classifying COVID-19, viral pneumonia from the chest x-rays has been proposed in this paper. Combining image processing to enhance relevant features from chest x-rays with deep neural network architecture helped us achieve 97.70%, 84.95%, and 97.03% across the three COVID-19 test datasets. The proposed work also compares our model with models proposed in previous research, which comments on the credibility, authenticity, and importance of our work. Based on the results achieved, we consider our attempt successful and hope to get a vast, diverse dataset to generalize our model parameters and make our model clinically feasible. Simple, efficient, accurate tool using state of the art technology was the core emphasis of this research that paves the way for unbiased and precise medical diagnosis. Humankind has been struck with one of the worst pandemics; we are grateful that we could contribute in our attempt to fight it. To make the proposed model clinically feasible, further training of the model is required, including training on a large dataset with diverse x-ray images that are not yet publicly available. It's good to have a large dataset because the larger the data set, the more we can extract insights that we trust from

Table 5. Evaluation of models on dataset 3 with respect to classification metrics (where, P is precision, R is recall and f1 is f1-score).

Model	Class	n (truths)	n (classified)	Acc (%)	P	R	f1	Acc_overall (%)
Inception V3	0	1519	1529	95.17	0.95	0.96	0.95	95.17
	1	1379	1369	95.17	0.95	0.95	0.95	
VGG16	0	1519	1530	95.96	0.96	0.97	0.96	95.96
	1	1379	1368	95.96	0.96	0.95	0.96	
ResNet50	0	1519	1522	94.65	0.95	0.95	0.95	94.65
	1	1379	1376	94.65	0.94	0.94	0.94	
VGG19	0	1519	1536	96.51	0.96	0.97	0.97	95.51
	1	1379	1362	96.51	0.97	0.96	0.96	
MobileNet	0	1519	1547	93.17	0.93	0.94	0.94	93.17
	1	1379	1351	93.17	0.94	0.92	0.93	
InceptionResNetV2	0	1519	1508	95.27	0.96	0.95	0.95	95.27
	1	1379	1390	95.27	0.95	0.95	0.95	
DenseNet121	0	1519	1553	92.34	0.92	0.94	0.93	92.34
	1	1379	1345	92.34	0.93	0.91	0.92	
Proposed Model (P_R) (without image pre-processing)	0	1519	1532	96.58	0.97	0.97	0.97	96.58
	1	1379	1366	96.58	0.97	0.96	0.96	
Proposed Model (P_P) (with image pre-processing)	0	1519	1533	97.03	0.97	0.98	0.97	97.03
	1	1379	1356	97.03	0.97	0.96	0.97	

that dataset and, hence, the more confidence in the model. The large dataset will improve accuracy and will generalize the model well.

References

[1] Barnawi A, Chhikara P, Tekchandani R, Kumar N, Alzahrani B. Artificial intelligence-enabled Internet of Things-based system for COVID-19 screening using aerial thermal imaging. Future Generation Computer Systems 2021; 124: 119-132. doi: 10.1016/j.future.2021.05.019

[2] Fang Y, Zhang H, Xie J, Lin M, Ying L et al. Sensitivity of chest CT for COVID-19: comparison to RT-PCR. Radiology 2020; 296 (2): E115–E117. doi: 10.1148/radiol.2020200432

[3] Singh D, Kumar V, Kaur M. Classification of COVID-19 patients from chest CT images using multi-objective differential evolution–based convolutional neural networks. European Journal of Clinical Microbiology & Infectious Diseases 2020; 39 (7): 1379–1389. doi: 10.1007/s10096-020-03901-z

[4] Rajpurkar P, Irvin J, Zhu K, Yang B, Mehta H et al. Chexnet: Radiologist-level pneumonia detection on chest x-rays with deep learning 2017. https://arxiv.org/abs/1711.05225

[5] Tartaglione E, Barbano C, Berzovini C, Calandri M, Grangetto M. Unveiling covid-19 from chest x-ray with deep learning: a hurdles race with small data. International Journal of Environmental Research and Public Health 2020; 17 (18): 6933. doi: 10.3390/ijerph17186933

[6] Asnaoui K, Chawki Y. Using X-ray images and deep learning for automated detection of coronavirus disease. Journal of Biomolecular Structure and Dynamics 2020; 1–12. doi: 10.1080/07391102.2020.1767212

[7] Abbas A, Abdelsamea MM, Gaber MM. Classification of COVID-19 in chest X-ray images using DeTraC deep convolutional neural network. Applied Intelligence 2021; 51: 854–864. doi: 10.1007/s10489-020-01829-7

[8] Heidari M, Mirniaharikandehei S, Khuzani A, Danala G, Qui Y et al. Improving the performance of CNN to predict the likelihood of COVID-19 using chest X-ray images with preprocessing algorithms. International Journal of Medical Informatics 2020; 144: doi: 10.1016/j.ijmedinf.2020.104284

[9] Siddhartha M, Santra A. COVIDLite: A depth-wise separable deep neural network with white balance and CLAHE for detection of COVID-19 2021. https://arxiv.org/abs/2006.13873

[10] Siracusano G, Corte A, Gaeta M, Cicero G, Chiappini M et al. Pipeline for Advanced Contrast Enhancement (PACE) of chest X-ray in evaluating COVID-19 patients by combining bidimensional empirical mode decomposition and CLAHE. Sustainability 2020; 12 (20): 8573. doi: 10.3390/su12208573

[11] Das D, Santosh K, Pal U. Truncated inception net: COVID-19 outbreak screening using chest X-rays. Physical and engineering sciences in medicine 2020; 43: 915–925. doi: 10.1007/s13246-020-00888-x

[12] Ozturk T, Talo M, Yildirim E, Baloglu U, Yildirim O, et al. Automated detection of COVID-19 cases using deep neural networks with X-ray images. Computers in Biology and Medicine 2020; 121: doi: 10.1016/j.compbiomed.2020.103792

[13] Khan A, Shah J, Bhat M. Coronet: A deep neural network for detection and diagnosis of COVID-19 from chest x-ray images. Computer Methods and Programs in Biomedicine 2020. 196: doi: 10.1016/j.cmpb.2020.105581

[14] Apostolopoulos I.D., Aznaouridis S.I., Tzani M.A. Extracting Possibly Representative COVID-19 Biomarkers from X-ray Images with Deep Learning Approach and Image Data Related to Pulmonary Diseases. Journal of Medical and Biological Engineering 2020; 40:462–469. doi: 10.1007/s40846-020-00529-4

[15] Ucar F, Korkmaz D. COVIDiagnosis-Net: Deep Bayes-SqueezeNet based Diagnostic of the Coronavirus Disease 2019 (COVID-19) from X-Ray Images. Medical Hypotheses 2020; 140: doi: 10.1016/j.mehy.2020.109761

[16] Loey M, Smarandache F, Khalifa N. Within the Lack of Chest COVID-19 X-ray Dataset: A Novel Detection Model Based on GAN and Deep Transfer Learning. Symmetry 2020; 12 (4): 651. doi: 10.3390/sym12040651

[17] Luz E, Silva P, Silva R, Silva L, Guimaraes J et al. Towards an Effective and Efficient Deep Learning Model for COVID-19 Patterns Detection in X-ray Images. Research on Biomedical Engineering 2021. doi: 10.1007/s42600-021-00151-6

[18] Shan F, Gao Y, Wang J, Shi W, Shi N et al. Lung infection quantification of covid-19 in ct images with deep learning 2020. https://arxiv.org/pdf/2003.04655.pdf

[19] Narin A, Kaya C, Pamuk Z. Automatic detection of coronavirus disease (covid-19) using x-ray images and deep convolutional neural networks. Pattern Analysis and Applications 2021. doi: 10.1007/s10044-021-00984-y

[20] Huang L, Han R, Ai T, Yu P, Kang H et al. Serial quantitative chest CT assessment of COVID-19: a deep learning approach, Radiology: Cardiothoracic Imaging 2020; 2 (2). doi: 10.1148/ryct.2020200075

[21] Apostolopoulos ID, Mpesiana TA. Covid-19: automatic detection from X-ray images utilizing transfer learning with convolutional neural networks. Physical and Engineering Sciences in Medicine 2020; 43: 635–640. doi: 10.1007/s13246-020-00865-4

[22] Maia M, Pimentel J, Pereira I, Gondim J, Barreto M et al. Convolutional Support Vector Models: Prediction of Coronavirus Disease Using Chest X-rays. Information; 11 (12): 548. doi: 10.3390/info11120548

[23] Bai X, Fang C, Zhou Y, Bai S, Liu Z et al. Predicting COVID-19 malignant progression with AI techniques. SSRN 2020; 1–29. doi: 10.2139/ssrn.3557984

[24] Sun L, Song F, Shi N, Liu F, Li S et al. Combination of four clinical indicators predicts the severe/critical symptom of patients infected COVID-19. Journal of Clinical Virology 2020; 128. doi: 10.1016/j.jcv.2020.104431

[25] Ibrahim A, Ozsoz M, Serte S, Al-Turjman F, Yakoi P. Pneumonia Classification Using Deep Learning from Chest X-ray Images During COVID-19. Cognitive Computation 2021. 1–13. doi: 10.1007/s12559-020-09787-5

[26] Ghoshal B, Tucker A. Estimating uncertainty and interpretability in deep learning for coronavirus (COVID-19) detection 2020. https://arxiv.org/abs/2003.10769

[27] Chhikara P, Singh P, Gupta P, Bhatia T. Deep Convolutional Neural Network with Transfer Learning for Detecting Pneumonia on Chest X-Rays. In: Advances in Bioinformatics, Multimedia, and Electronics Circuits and Signals 2020; 155–168. doi: 10.1007/978-981-15-0339-9_13

[28] Liu H. Feature extraction and image recognition with convolutional neural networks. Journal of Physics: Conference Series 2018; 1087 (6). doi: 10.1088/1742-6596/1087/6/062032

[29] Meyer-Baese A, Schmid V. In: Chapter 4 - The Wavelet Transform in Medical Imaging. Pattern Recognition and Signal Analysis in Medical Imaging (Second Edition) 2014. 113–134. doi: 10.1016/B978-0-12-409545-8.00004-2

[30] Kurt B, Nabiyev V, Turhan K. Medical images enhancement by using anisotropic filter and clahe. In: 2012 International Symposium on Innovations in Intelligent Systems and Applications 2012. pp. 1–4. doi: 10.1109/INISTA.2012.6246971

[31] Chen G, Chen P, Shi Y, Hsieh C, Liao B et al. Rethinking the Usage of Batch Normalization and Dropout in the Training of Deep Neural Networks 2019.https://arxiv.org/abs/1905.05928