Multimedia Tools and Applications (2023) 82:12653–12677
https://doi.org/10.1007/s11042-022-13739-6

TRACK 2: MEDICAL APPLICATIONS OF MULTIMEDIA

Multi-modal fusion of deep transfer learning based COVID-19 diagnosis and classification using chest x-ray images

A. Siva Krishna Reddy[1] · K. N. Brahmaji Rao[2] · Narasimha Reddy Soora[3] · Kotte Shailaja[4] · N. C. Santosh Kumar[5] · Abel Sridharan[6] · J. Uthayakumar[7]

Received: 25 January 2022 / Revised: 19 March 2022 / Accepted: 25 August 2022 /
Published online: 16 September 2022

Abstract
COVID-19 pandemic has a significant impact on the global health and daily lives of people living over the globe. Several initial tests are based on the detecting of the genetic material of the coronavirus, and they have a minimum detection rate with a time-consuming process. To overcome this issue, radiological images are recommended where chest X-rays (CXRs) are employed in the diagnostic process. This article introduces a new Multi-modal fusion of deep transfer learning (MMF-DTL) technique to classify COVID-19. The proposed MMF-DTL model involves three main processes, namely pre-processing, feature extraction, and classification. The MMF-DTL model uses three DL models namely VGG16, Inception v3, and ResNet 50 for feature extraction. Since a single modality would not be adequate to attain an effective detection rate, the integration of three approaches by the use of decision-based multimodal fusion increases the detection rate. So, a fusion of three DL models takes place to further improve the detection rate. Finally, a softmax classifier is employed for test images to a set of six different. A wide range of experimental result analyses is carried out on the Chest-X-Ray dataset. The proposed fusion model is found to be an effective tool for COVID-19 diagnosis using radiological images with the average $sens_y$ of 92.96%, $spec_y$ of 98.54%, $prec_n$ of 93.60%, $accu_y$ of 98.80%, F_{score} of 93.26% and kappa of 91.86%.

Keywords COVID-19 · Chest X-rays · Deep learning · Fusion model · Radiological images

1 Introduction

In late 2019, a virus related to severe acute respiratory syndrome (SARS) coronavirus has been identified in Wuhan city, China. The virus is shortly known as severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), causing an infection called COVID-19. It is highly

✉ J. Uthayakumar
uthayresearchscholar@gmail.com

Extended author information available on the last page of the article

relevant to SARS and has been initially identified by the symptoms of fever with respiratory problems. This virus is highly infectious and can transmit to other humans easily [21]. At the end of the year 2019, the first case of COVID-19 has been identified in Wuhan city and it has been rapidly spread to all other parts of the world from Europe to the United States. The increased mortality rate by COVID-19 has been threatened in the whole world [15]. Using the spike exterior structure of COVID-19, it can easily bind with the receptors [17]. The presence of COVID-19 in infected persons cannot be identified from the primary clinical symptoms [20]. The diagnostic procedure of COVID-19 becomes highly essential due to the faster spreading rate [22].

Nowadays, healthcare professionals utilize the Reverse Transcription Polymerase Chain Reaction (RT-PCR) test to detect the COVID-19. To perform this test, the respiratory specimens are collected and the major problem lies in the receipt place of the specimens. Besides, this diagnosis procedure can also be prone to error by the expert mistakes [11], and also it consumes more time. During the testing period and report generation, the patients have to be secluded in unsuitable circumstances for hours or days till the test reports are received. Besides, these kinds of tests have resulted in a poor detecting rate in the range of 30% to 50%. Therefore, it is needed to repeat the test several times for confirming the reports [6]. In this situation, the crucial problem to tackle with the problem of COVID-19 is the deployment of rapid treatment to the patients. Due to the quick spreading of the pandemic, the patients are drastically admitted to the hospitals. It leads to the requirement of quicker diagnosis models, which need to be resolved. Another option is the use of radiological images like Chest X-ray (CXR) or Computed Tomography (CT) images to diagnose CoV-2 infections.

Presently, Machine Learning (ML) approaches are widely employed to diagnose diseases in an automated way. Deep learning (DL) is a familiar research area in artificial intelligence (AI), which allows the design of dedicated models for attaining better outcomes with no need for manual feature extraction. It is commonly employed for several problems like a breast cancer diagnosis, diabetic retinopathy, brain tumor identification, pneumonia detection from CXR images, etc. The COVID-19 pandemic is rapidly rising the need for knowledge. It has enhanced the consciousness in designing an automatic AI-based detection model. The modest, precise, and rapid AI methods find useful to resolve the issue and offer support to patients at appropriate times.

Due to the advancements in the healthcare sector, the predictive and diagnosis models are increased rapidly. Although the operative feature extraction methods are required for achieving effective performance, DL models are commonly employed in the healthcare sector for automated deep feature extraction [10]. Earlier researchers proved that the COVID-19 abnormalities can be recognized in the CXRs and CT images. Though the radiological image-based diagnosis process seems faster and attains an increased detection rate in the earlier stages of a pandemic, it necessitates medical experts for interpreting the images. Essentially, the AI-based diagnosis process recommends the experts get a fast and precise clarification over the X-ray images while detecting COVID-19 [1]. Keeping this in mind, several works have been carried out in the AI and DL-based diagnosis process of COVID-19 using radiological images [2–4].

The convolutional neural network (CNN) is found to be an effective DL model for medical image classification. Since the outbreak of COVID-19, much research has been conducted on processing the data related to DL algorithms, especially CNN. On the other hand, transfer learning is the cornerstone of computer vision. Various categorization tasks related to images can accomplish effective performance with datasets of a limited size with transfer learning than using any other method. With the urgent need for solutions to cope with the COVID-19

pandemic and based on the recent efforts among researchers to design AI-based solutions, this study introduces a new multi-modal fusion of deep transfer learning (MMF-DTL) model for COVID-19 diagnosis on CXRs. The presented MMF-DTL model involves a pre-processing step to improve the image quality by contrast enhancement, artefacts removal, and image sharpening. Then, the MMF-DTL undergoes the fusion of three DL models namely VGG-16, ResNet 50, and Inception v3. Finally, softmax classified is used for the classification of applied images into a set of six different classes, namely acute respiratory distress syndrome (ARDS), COVID-19, No Finding, Pneumocystis, SARS, and Streptococcus. The proposed MMF-DTL model is developed as a user interface to simplify the diagnosis process and assist physicians, healthcare professionals, researchers, and so on. A detailed experimental validation takes place on the CXR dataset. The overall paper contribution is summarized as follows

- Preprocessing includes contrast enhancement, artefact removal, and image sharpening
- Present a new MMF-DTL model for COVID-19 classification by the fusion of three DL approaches such as VGG-16, ResNet 50, and Inception v3.
- Classifies the CXR images into six classes, such as ARDS, COVID-19, No Finding, Pneumocystis, SARS, and Streptococcus
- Develop a simple user interface for the MMF-DTL model to simplify the diagnosis process and control the present pandemic

2 Literature survey

DL models have gained significant attention among researchers in the computer vision community. DL comprises multiple layers and it makes it different from the artificial neural networks. Xu et al. [1] devised a DL-enabled COVID-19 detection model to categorize the pulmonary CT images. The author has used a 3D convolution neural network (CNN) to segment candidate infection regions and the classification process takes place with respective confidence scores utilizing local attention mechanism with CNN. At last, the infection type, and the total confidence score of the CT cases were determined using Noisy-or Bayesian function. This model has achieved a maximum detection accuracy of 86.7%. Ali Narin et al. [2] had undergone an examination of three CNN models for the identification of COVID-19 and normal cases in CXR radiographs. The authors have designed an end-to-end structure with automated feature extraction and selection approaches. The experimental results indicated that the pre-trained ResNet-50 model outperformed the other two models with the maximum accuracy of 98% whereas the InceptionV3 and Inception-ResNetV2 models have ended up with the accuracy values of 97% and 87% respectively.

Ayyar et al. [4] presented a hierarchical classifier model for COVID-19 using CXR images. It comprises several binary classification models and it is exhibited that the DL models integrated into the global attention mechanism outperformed the baseline COVID-Net model. Ioannis et al. [3] have performed experimental validation of the existing CNN models for the classification of COVID-19. Particularly, the authors have used transfer learning for the identification of different abnormalities in image datasets and offered significant results. It has reached the accuracy values of 96.78% and 94.72% on the applied first and second datasets respectively. Ozturk et al. [13] presented a machine learning (ML) model for the identification of viral epidemics using X-ray images. At the earlier stage, the images are extracted by the use of 4 feature extraction techniques. Then, the SMOTE algorithm is applied

for the elimination of unbalanced data problems. Afterward, the feature reduction process takes place using the stacked autoencoder (SA) and principal component analysis (PCA). Lastly, the support vector machine (SVM) model is utilized for the classification task. This model has classified a set of six classes with higher accuracy of 94.23%. Nguyen et al. [12] aimed to investigate the severity of the problem by the use of DL models for COVID-19 detection and classification. Rasheed et al. [16] presented a new ML-based COVID-19 detection model using CNN and logistic regression (LR) models. The authors in [14] performed a study on the detection of COVID-19 using ML models. Sharmila [9] presented an effective CNN and deep convolutional generative adversarial network (DCGAN) which categorizes CXR images into different classes. Though several DL-based COVID-19 diagnosis models were presented, still there is a need to improve the detection rate and thereby increase the survival rate of infected persons.

3 The proposed MMF-DTL model

In this study, a novel MMF-DTL technique has been developed to detect and classify COVID-19. The presented MMF-DTL model incorporates several processes such as preprocessing, feature extraction, fusion process, and classification. The working principle of the MMF-DTL technique is demonstrated in Fig. 1.

3.1 Preprocessing

The MMF-DTL model undergoes preprocessing in three stages, namely contrast enhancement, artefacts removal, and image sharpening.

3.1.1 Contrast enhancement

CLAHE has demonstrated the most applicable usefulness in the intensity of medical images. The main aim of this model is to enable the observer to examine the disease easily in an image [23]. It is used to investigate the intensities of histograms in a contextual area that is placed at every pixel and fix the demonstrated pixel intensities as a pixel intensity rank in the concerned histogram. The histogram can be defined as an extended form of the normal histogram where the contrast level improvement is intended by a model at every intensity level that is restricted to a user-selectable extent. CLAHE is fundamentally a conventional technique of contrast improvement that maximizes the local information of an image effectively. CLAHE is mainly applied to enhance the contrast of an image when related to the actual CXR image.

The procedures involved in CLAHE are:

- **Step 1:** Classify the image into continuous and non-overlapping tiles.
- **Step 2:** Fix the histogram of every tile of a threshold and re-share the fixed pixels over the entire histogram in a uniform manner.
- **Step 3:** Use Histogram Equalization on every tile.
- **Step 4:** Use bilinear interpolation for mapping among the adjacent tiles. Hence, the mapped pixel is said to be the outcome of interpolation from the intensity mapping of 4 corresponding tiles.

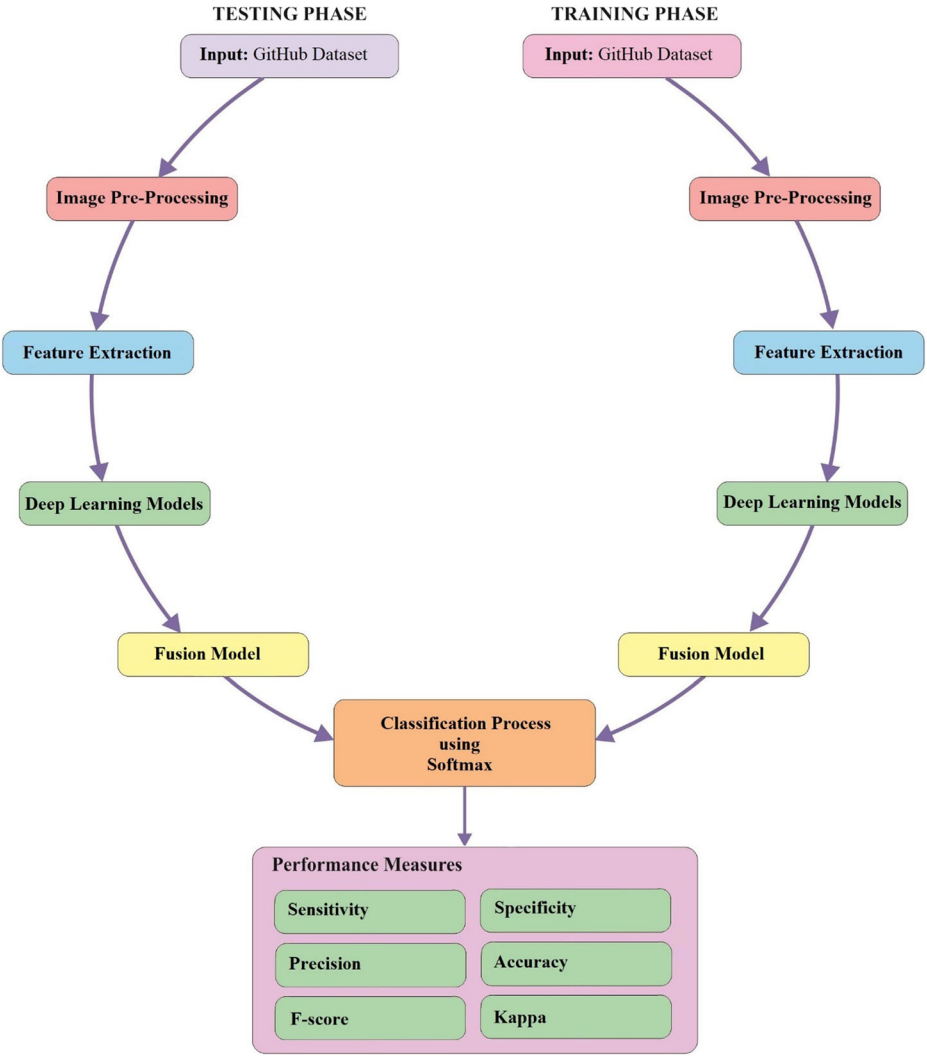


Fig. 1 Overall process of MMF-DTL model

3.1.2 Artefacts removal

There are massive diverse artifacts present in the medical images, where some of them affect the diagnostic process, whereas others might confuse with pathology. The elimination of an artifact is a mandatory operation in the medical image examination [5]. The processes involved in the artifact removal technique are listed below.

- **Step 1.** A grayscale medical image has been acquired as the input.
- **Step 2.** A threshold value of an image is estimated under the application of the standard deviation (SD) model
- **Step 3.** The image has been binarized with the help of a threshold value where the pixel values cross a threshold are fixed to 1 and pixels lower than thresholds are fixed to 0.

- **Step 4.** The binary form of the image is modeled and regions of linked elements were evaluated with the application of equivalent categories.
- **Step 5.** The linked element with the maximum area as well as a connected component that has partial area can be identified.
- **Step 6.** The ratio of the higher area to that of the second higher area is determined.
- **Step 7.** When the ratio is maximum, then the component with a greater area is maintained and the remaining values are eliminated. Else, when a ratio is low, then the component with higher and second maximum area is provisioned and residuals are removed.
- **Step 8.** A convex hull is computed for 1 pixel of an image and the region inside the convex hull has been allocated as 1.
- **Step 9.** Then, the attained image matrix can be improved to the actual image matrix to get the image of a problem-free medical image.

3.1.3 Image sharpening

The image sharpening is used to improve the contrast among bright as well as dark sites for obtaining the features. Once the noise is removed from an image, sharpening the images are carried out under the application of a filter 2D model. The performance of filter 2D is used to convolve a kernel along with an image. A maximum filter might be utilized on the image that estimates the average of pixel values that are present in a window. A 3 × 3 averaging filter kernel (K) is presented as given in the following:

$$K = \frac{1}{9}\begin{bmatrix} 0 & -1 & 0 \\ -1 & 5 & -1 \\ 0 & -1 & 0 \end{bmatrix} \tag{1}$$

Sharpening the images using the predefined kernel operates as follows: for each pixel, a 3 × 3 window should be located in the pixel, each pixel that arrives in a window is calculated and a sum is divided by 9. These functions are finally computed with the average of pixel values in a window. This process is carried out on all pixels in the image for producing the resultant image.

3.2 Feature extraction

The preprocessed input images undergo the feature extraction process by the use of VGG-16, ResNet 50, Inception v3, and fusion models. The processes involved in these models are discussed in the following subsections.

3.2.1 VGGNet-16

VGG-Net was presented by [19], the standard performance of VGG-Net eliminates the complexities involved in the system. The convolution of AlexNet and ZFNet was developed by [24] in the primary convolutional layer are 11 × 11 with stride 4 and 7 × 7 with stride 2, correspondingly. Also, VGG-Net shows that 2 successive 3 × 3 convolutions are similar to 5 × 5 receptive fields whereas 3 are identical to 7 × 7. This model is comprised of 2 major benefits in applying 3 × 3 convolutions rather than using a single 7 × 7 convolution: firstly,

as an alternative of one, 3 ReLU layers have been applied to make the decision function more distinguishable; then, minimizes the parameters. The main goal of using a 1 × 1 convolution layer is to enhance the non-linearity of the decision functions. Even though the 1 × 1 convolutional task is linear, ReLU improves the nonlinearity. In this study, a tremendous VGG16 network has been selected to fine-tune the 6 classes of coronavirus diagnosis. The architecture of VGG16, as well as the parameter settings, are provided at every level. Assume that there is a training dataset with m instances, $\{(x(l), y(l), \ldots, (x(m), y(m))\}$. In the entire sample, the network overall cost performance can be represented as follows:

$$J(W,b) = \left[\frac{1}{m}\sum_{i=1}^{m}\left(\frac{1}{2}\left\|K_{w,b}\left(x^{(i)}\right) - y^{(i)}\right\|^2\right)\right] + \frac{\lambda}{2}\sum_{l=1}^{n_{l-1}}\sum_{i=1}^{s_l}\sum_{j=1}^{s_{l+1}}\left(W_{ji}^{(l)}\right)^2 \tag{2}$$

Where $K_{w,b}(x^{(i)})$ represents the NN model, $W_{ji}^{(l)}$ is a connection weight among the jth element of layer 1 and i^{th} element of layer l + 1, b refers to the bias term of the hidden layer neuron. Besides, a popular batch gradient descent optimization approach is developed.

3.2.2 ResNet 50

ResNet employs the residual block for resolving degradation as well as gradient vanishing issues which often occur in the CNN model. The residual block lengthens the network intensity and enhances network performance. ResNet networks are capable of producing the best results in ImageNet [7] classification competitions. The residual blocks of the ResNet model execute the residual under the inclusion of input present in a residual block as well as the simulation outcome of the residual blocks. The residual function can be represented as follows:

$$y = F(x, W) + x \tag{3}$$

where x means residual block input; W represents residual block weights; y denotes the outcome of residual blocks. It comprises different residual blocks where the convolution kernel size of a convolution layer is diverse. The conventional structure of ResNet contains RetNet18, RestNet50, and RestNet101. The building block of ResNet50 is shown in Fig. 2.

ResNet50 is referred to as a 50-layer Residual Network. This model has produced a highly convenient outcome in the ImageNet and MS-COCO competitions. This structure is constrained with 49 convolutional layers and 1 fully connected (FC) layer along with the input map of size 224 × 224. The convolutional part includes 5 blocks with tiny filter sizes of 1 × 1 and 3 × 3. The initial block is 1 convolution layer and max-pooling layer, which has a stride of 2. In the case of alternate blocks, it is filled with massive residual blocks with a stride of 2 for the final 3 blocks. The ResNet50 structure has concluded with an FC layer and softmax activation functions. The common suggestion of the transfer learning models is to employ a pre-trained dataset that has to be suitable for a novel operation. Hence, the key objective is to begin the learning task from patterns that are already learned to overcome diverse issues and it leads to the model normalization. Transfer learning has been utilized for tiny datasets and applied as a feature extractor. The major aim is to maintain the convolutional portions and interchange the FC layers with the presented classifier.

3.2.3 Inception v3

The Inception model of CNN is said to be the series of neural networks (NN) which should be rejected from the records of CNN. Many NN performs only the network lengthening operation

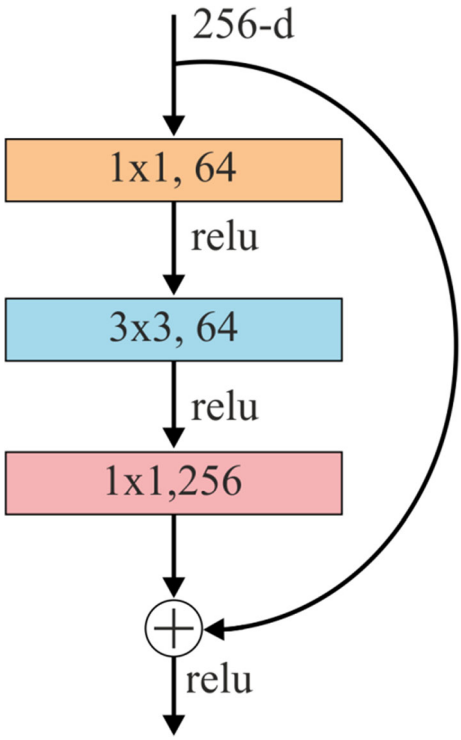


Fig. 2 Building block of ResNet50

by improving the convolutional layer to attain the optimal function of Inception NN. This Inception NN has modified the predefined principle. The Inception module evolved from the Inception NN applies diverse filter sizes and higher pooling to limit the data dimension. There is a major benefit of reaching effective features with meaningfully minimized processes and limited variables. Inceptionv3 employs the asymmetric model to degrade large-scale convolution kernels into small-scale convolution kernels, which reduces the 3 × 3 convolution kernels into 2 convolution kernels like 1 × 3, 3 × 1, and also limit the network variables at the time of provisioning network performance. Under the application of asymmetric decomposition, the depths of the NN could be deepened and the non-linearity of the system has been maximized [8]. Conventionally, the CNN pooling task is applied for lessening the mesh size of feature maps. To eliminate the problems, the activation dimension of network activation has to be extended before using maximum pooling or average pooling.

3.2.4 Fusion model

In this paper, the MMF model is utilized for improving the classification process of COVID19 and makes it effective in the detection of six different classes using CXR images. Since a single modality would not be adequate to attain an effective detection rate, the integration of three models by the use of decision-based multimodal fusion increases the detection rate. In MMF architecture, the improvements in unimodal CNNs are beneficial and significant progress has been made. The number of features in the fusion model is 256,512 features.

Table 1 General Parameters Settings

Description	Values
Optimizer	Adam
Learning Rate	0.01
Beta_1	0.9
Beta_2	0.999
Batch Size	4098
Epochs	35
Models	VGG16/ResNet50/Inceptionv3/Fusion

3.3 Softmax classification

The extracted feature vectors are provided as input into the softmax classifier, which is generally employed for multiclass classification problems [25]. It undergoes mapping of an input vector i to a N dimensional space to K output classes, as defined in Eq. (4):

$$y_j = \frac{\exp\left(\theta_j^T i\right)}{\sum_{k=1}^{K} \exp\left(\theta_k^T i\right)} \tag{4}$$

Where $\theta_k = [\theta_{k1}, \theta_{k2}, .., \theta_{kN}]^T$ indicates the weights, which have been tuned by an optimization algorithm. The output value of the layer is identical to the class count generated, i.e. six classes.

4 Results and discussion

The implementation of the MMF-DTL model takes place using Intel i5, 8th generation PC with 16GB RAM, MSI L370 Apro, Nividia 1050 Ti4 GB. We used Python 3.6.5 tool along with pandas, sklearn, Keras, Matplotlib, TensorFlow, opencv, Pillow, seaborn and pycm. The general set of parameters involved in the simulation process has been provided in Table 1.

Besides, the model specific parameters of VGG-16, ResNet-50, and Inception v3 model are offered in Table 2.

4.1 Dataset description

The effectiveness of the proposed MMF-DTL model on COVID-19 diagnosis takes place on the Chest-X-Ray dataset (https://github.com/ieee8023/covid-chestxray-dataset). It contains a total of 305 images under six classes such as ARDS, COVID-19, No Finding, Pneumocystis, SARS, and Streptococcus. The details of the classes along with the corresponding image count are tabulated in Table 3. A few of the sample test images from six classes are also displayed in Fig. 3.

Table 2 Model Specific Parameters

Descriptions	VGG-16	ResNet-50	Inception V3
Input Shape	224,224	224,224	299,299
Number of Features	25,088	131,072	256,512
Total Params	14,714,688	23,587,712	21,802,784
Trainable Params	14,714,688	23,534,592	21,768,352
Non-Trainable Params	0	53,120	34,432

Table 3 Dataset Description

Classes	Label	Number of Images
ARDS	0	15
COVID-19	1	220
No Finding	2	27
Pneumocystis	3	15
SARS	4	11
Streptococcus	5	17

4.2 Results analysis

Figure 4 shows the sample visualization of the images produced from the different stages of preprocessing. Figure 4a shows the actual input images and the respective contrast improved images by the CLAHE model are shown in Fig. 4b. Then, Fig. 4c shows the images with artifacts removed and finally, the sharpened image is illustrated in Fig. 4d. The class activation

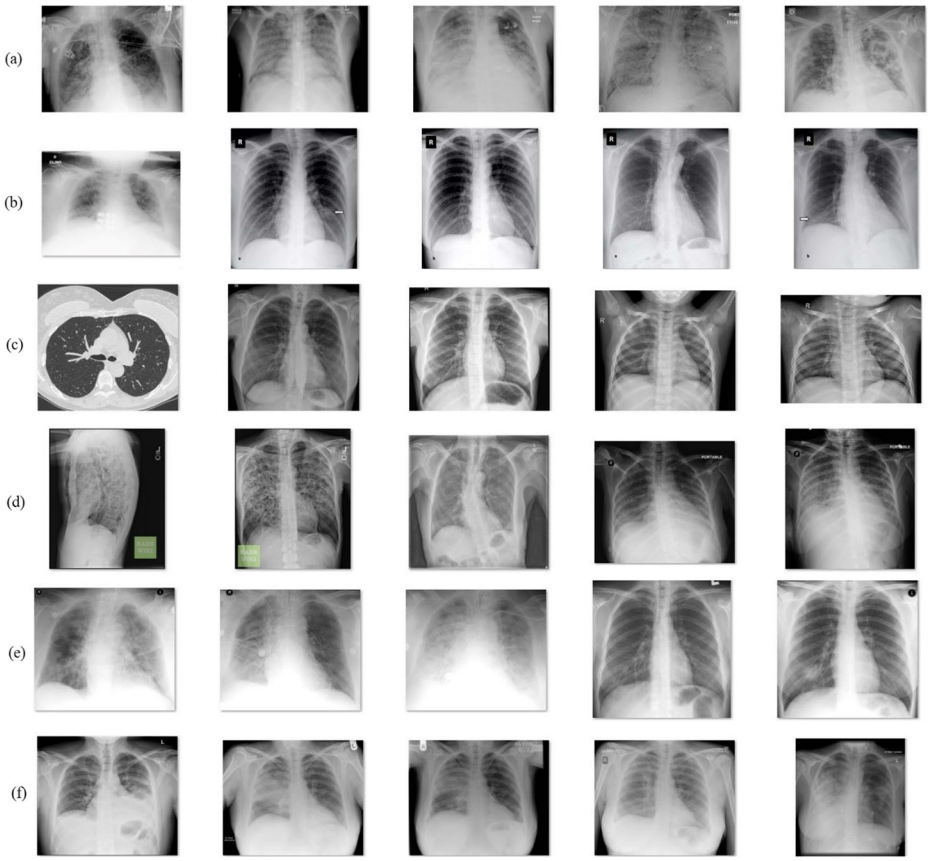


Fig. 3 Sample Test Images (**a**) ARDS (**b**) COVID19 (**c**) No Findings) Pneumocystise) SARS (**f**) Streptococcus

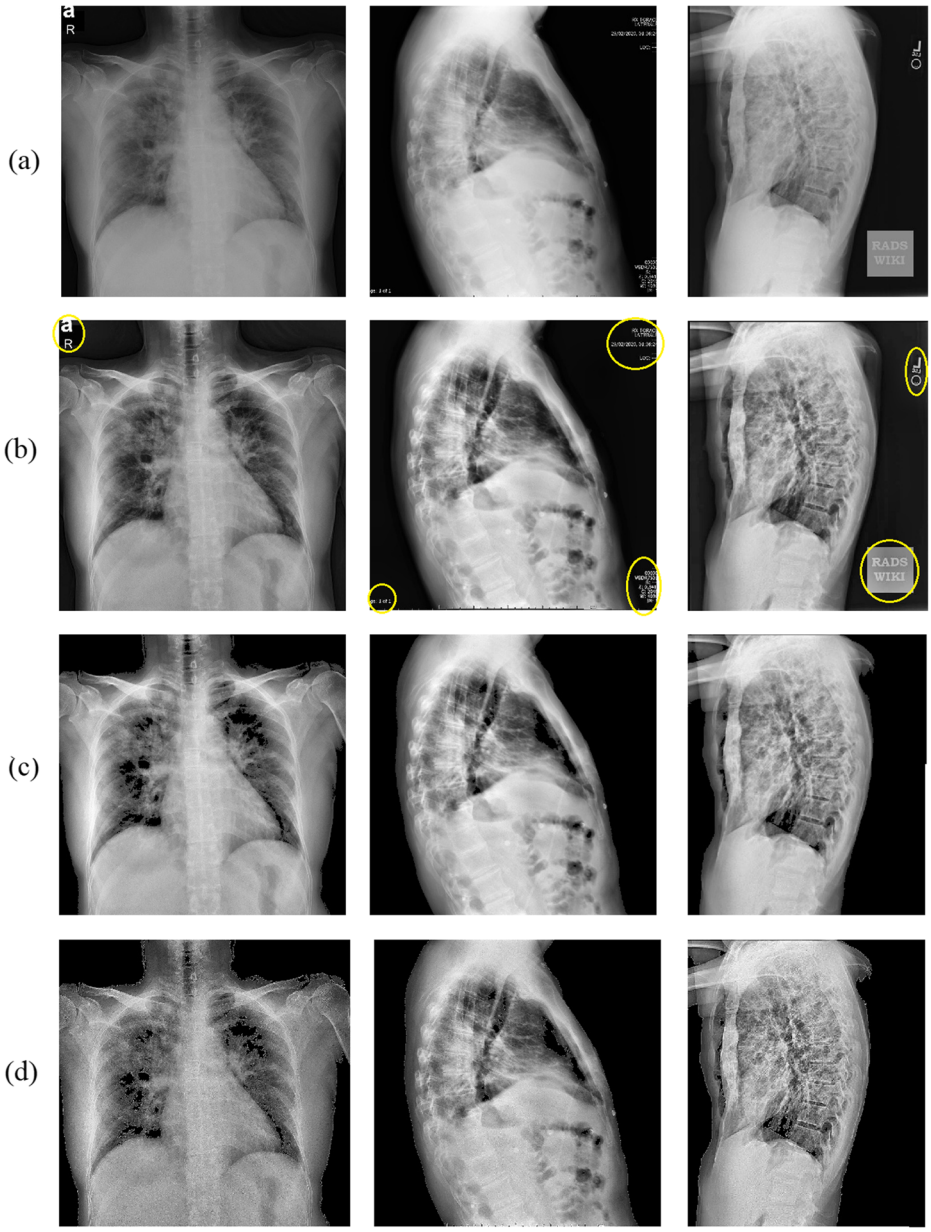


Fig. 4 **a** Original Images **b** Contrast Enhanced Images **c** Artefacts Removed Images **d** Sharpened Images

map of the test images for different classes is shown in Appendix 1 Fig. 13. The class activation map helps to detect the infected portions of the lungs for the respective class label (https://towardsdatascience.com/detection-of-covid-19-presence-from-chest-x-ray-scans-using-cnn-class-activation-maps-c1ab0d7c294b). Besides, the subprocesses involved in the

MMF-DTL model in the classification of six classes have been visualized in Appendix 2 Figs. 14, 15, 16, 17, 18 and 19.

Figure 5 shows the accuracy graph generated by four DL models at the time of training and validation. It is noted that the training and validation accuracies get improved with a rise in epoch count. The figures portrayed that the fusion model has offered a higher accuracy rate over the other three methods. Likewise, Fig. 6 depicts the loss graph generated by four DL models during the training and validation processes. The figure showed that the loss rate gets declined with an increase in the number of epochs. Besides, it is found that the fusion technique has shown proficient outcomes by the attainment of a minimal loss rate.

Figure 7 shows the receiver operating characteristic (ROC) curves offered by the different DL models and fusion models. The ROC analysis demonstrated that the DL models have showcased effective classification process with improved values of ROC. Particularly, it is evident that the fusion model has accomplished maximum ROC values on the classification process.

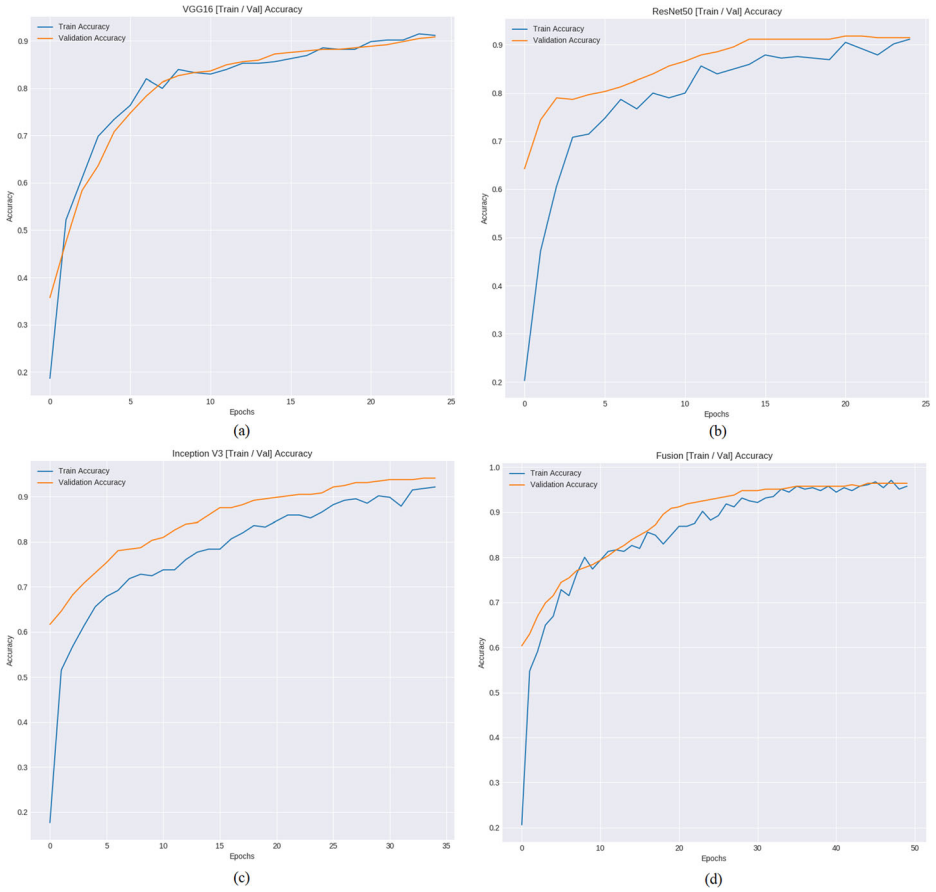


Fig. 5 Accuracy Graph (**a**) VGG16 (**b**) ResNet50 (**c**) Inception V3 (**d**) Fusion Method

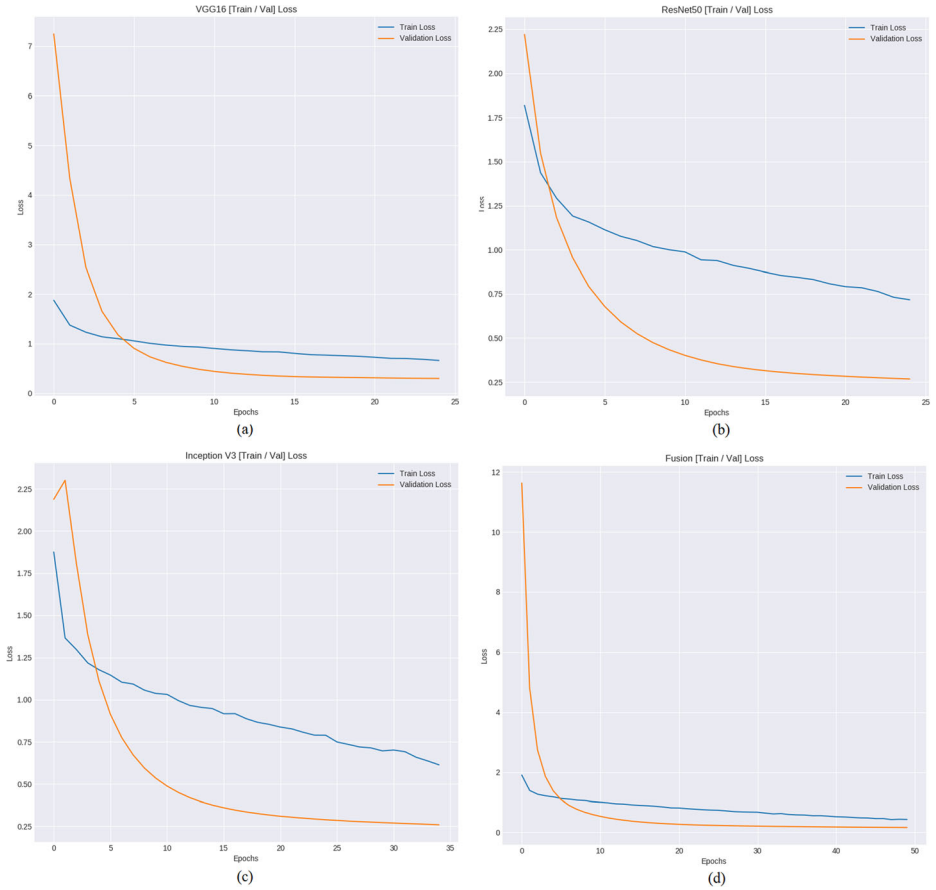


Fig. 6 Loss Graph (**a**) VGG16 (**b**) ResNet-50 (**c**) Inception V3 (**d**) Fusion Model

Figure 8 illustrates four confusion matrices generated by different models. From Fig. 8a, it is shown that the VGG-16 model has correctly classified 10 instances to ARDS, 201 instances to COVID-19, 27 instances to No Finding, 15 instances to Pneumocystis, 7 instances to SARS, and 17 instances to Streptococcus classes respectively. From Fig. 8b, it is depicted that the ResNet model has properly classified 11 instances to ARDS, 202 instances to COVID-19, 27 instances to No Finding, 15 instances to Pneumocystis, 7 instances to SARS, and 17 instances to Streptococcus classes respectively. Similarly, Fig. 8c indicated that the Inception v3 model has classified a set of 11, 206, 27, 15, 11, and 17 images under ARDS, COVID-19, No Finding, Pneumocystis, SARS, and Streptococcus classes respectively. Finally, Fig. 8d indicated that the Fusion model has resulted in effective performance and classified 9 images to ARDS, 215 images to COVID-19, 27 images to No Finding, 15 images to Pneumocystis, 11 images to SARS, and 17 images to Streptococcus classes respectively.

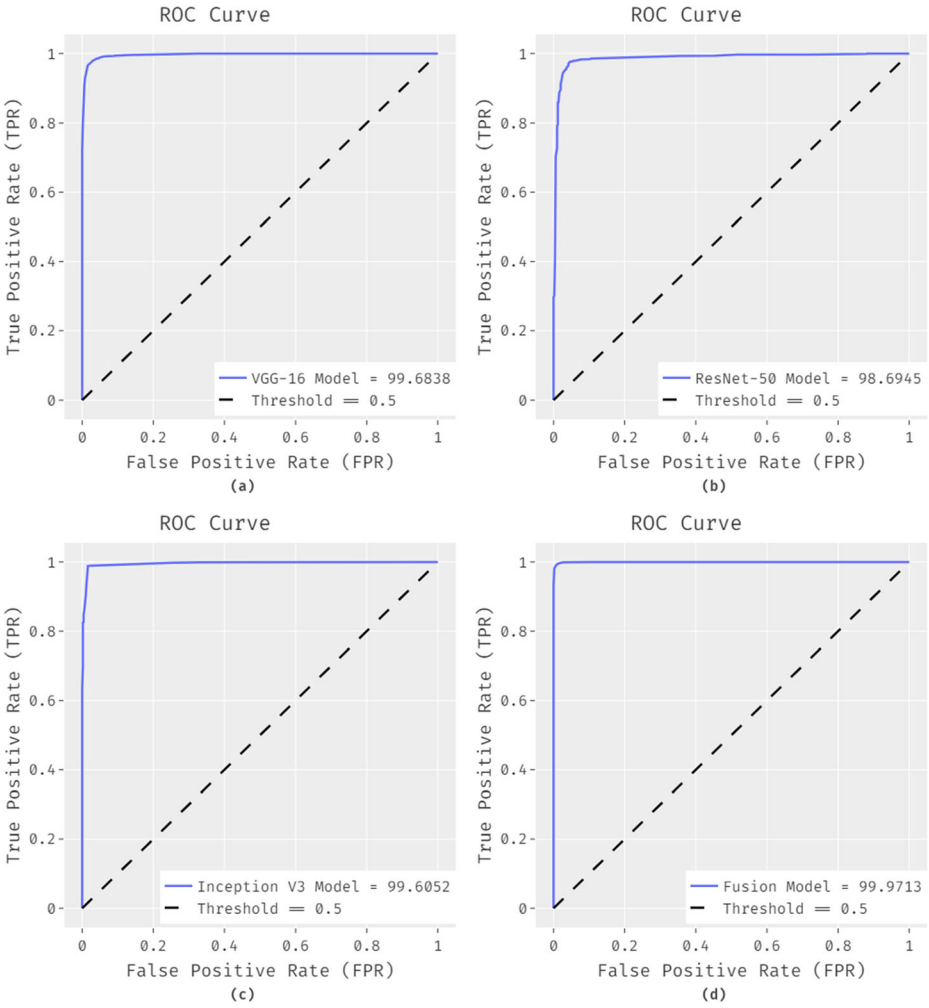


Fig. 7 ROC Curves (**a**) VGG16 (**b**) ResNet-50 (**c**) Inception V3 (**d**) Fusion Model

Table 4 exhibits the manipulation of the values presented in the confusion matrix with respect to True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN).

Table 5 offers a comprehensive comparative classification result analysis of the presented technique and other DL models. Figure 9 displays the average classifier outcomes of the proposed models on the test images. The experimental outcome pointed out that the VGG-16 model has classified the CXR images with average $sens_y$ of 86.94%, $spec_y$ of 97.94%, $prec_n$ of 82.46%, $accu_y$ of 96.79%, F_{score} of 83.12% and kappa value of 80.04%. Besides, the ResNet-50 model has attained a slightly higher classifier outcome over the VGG-16 model with the average $sens_y$of 88.13%, $spec_y$ of 97.95%, $prec_n$ of 84.07%, $accu_y$ of 97.04%, F_{score} of 84.60%

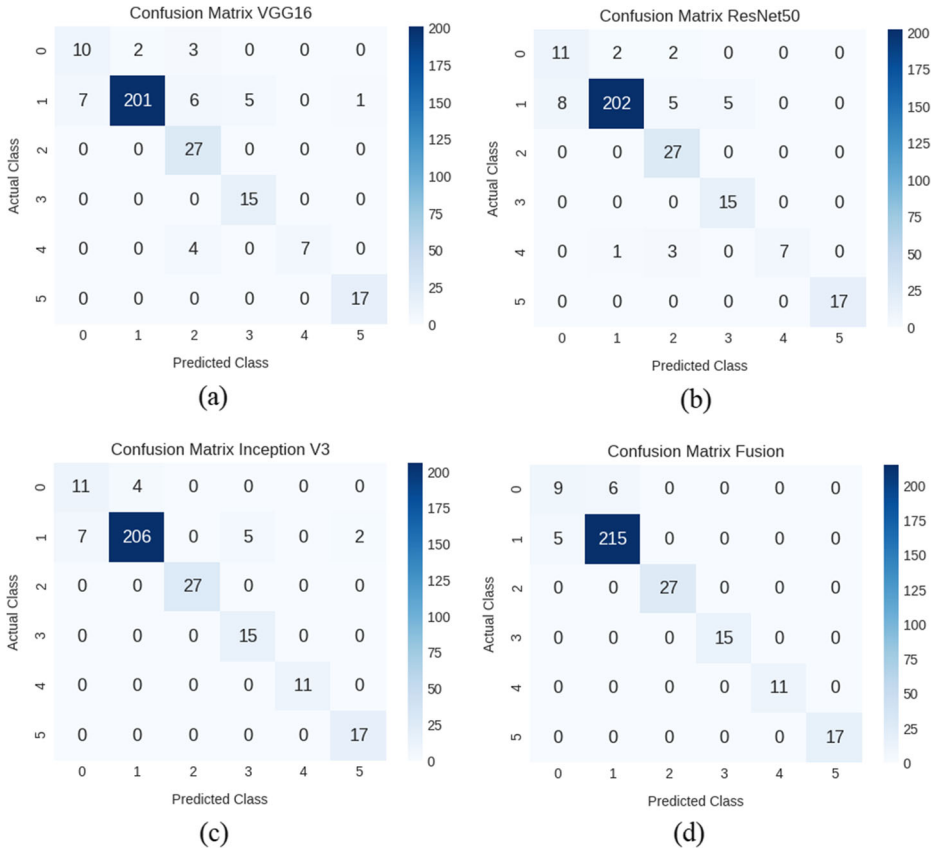


Fig. 8 Confusion Matrix (**a**) VGG-16 (**b**) ResNet-50 (**c**) Inception-v3 (**d**) Fusion Model

Table 4 Manipulation from Confusion Matrix

Classes	**VGG-16 Model**						**ResNet-50 Model**					
	0	**1**	**2**	**3**	**4**	**5**	**0**	**1**	**2**	**3**	**4**	**5**
TP	10	201	27	15	7	17	11	202	27	15	7	17
TN	267	76	250	262	270	260	268	77	252	264	272	262
FP	7	2	13	5	0	1	8	3	10	5	0	0
FN	5	19	0	0	4	0	4	18	0	0	4	0
Classes	**Inception V3 Model**						**Fusion Model**					
	0	**1**	**2**	**3**	**4**	**5**	**0**	**1**	**2**	**3**	**4**	**5**
TP	11	206	27	15	11	17	9	215	27	15	11	17
TN	276	81	260	272	276	270	285	79	267	279	283	277
FP	7	4	0	5	0	2	5	6	0	0	0	0
FN	4	14	0	0	0	0	6	5	0	0	0	0

Table 5 Performance Study of Proposed Method with different DL models

Model	Class	$Sens_y$	$Spec_y$	$Prec_n$	$Accu_y$	F_{score}	Kappa
VGG-16	ARDS	66.66	97.44	58.82	95.85	62.50	60.31
	COVID-19	91.36	97.44	99.01	92.95	95.03	82.96
	No Finding	100	95.05	67.50	95.51	80.59	78.16
	Pneumocystis	100	98.12	75.00	98.22	85.71	84.79
	SARS	63.63	100	100	98.57	77.77	77.07
	Streptococcus	100	99.61	94.44	99.64	97.14	96.95
	Average	**86.94**	**97.94**	**82.46**	**96.79**	**83.12**	**80.04**
ResNet-50	ARDS	73.33	97.10	57.89	95.87	64.70	62.54
	COVID-19	91.81	96.25	98.53	93.00	95.05	83.10
	No Finding	100	96.18	72.97	96.53	84.37	82.48
	Pneumocystis	100	98.14	75.00	98.23	85.71	84.79
	SARS	63.63	100	100	98.58	77.77	77.08
	Streptococcus	100	100	100	100	100	100
	Average	**88.13**	**97.95**	**84.07**	**97.04**	**84.60**	**81.67**
Inception V3	ARDS	73.33	97.52	61.11	96.30	66.66	64.72
	COVID-19	93.63	95.29	98.09	94.09	95.81	85.83
	No Finding	100	100	100	100	100	100
	Pneumocystis	100	98.19	75.00	98.28	85.71	84.82
	SARS	100	100	100	100	100	100
	Streptococcus	100	99.26	89.47	99.31	94.44	94.07
	Average	**94.49**	**98.38**	**87.28**	**98.00**	**90.44**	**88.24**
Fusion	ARDS	60.00	98.28	64.29	96.39	62.07	60.17
	COVID-19	97.73	92.94	97.28	96.39	97.51	90.99
	No Finding	100	100	100	100	100	100
	Pneumocystis	100	100	100	100	100	100
	SARS	100	100	100	100	100	100
	Streptococcus	100	100	100	100	100	100
	Average	**92.96**	**98.54**	**93.60**	**98.80**	**93.26**	**91.86**

and kappa value of 81.67%. Moreover, the Inception v3 model has attained an even higher average $sens_y$ of 94.49%, $spec_y$ of 98.38%, $prec_n$ of 87.28%, $accu_y$ of 98%, F_{score} of 90.44% and kappa of 88.24%. Furthermore, the Fusion model has attained maximum classifier results with the average $sens_y$ of 92.96%, $spec_y$ of 98.54%, $prec_n$ of 93.60%, $accu_y$ of 98.80%, F_{score} of 93.26% and kappa of 91.86%.

Table 6 and Figs. 10, 11 and 12 shows the comparative examination of the classification outcome provided by the presented and existing techniques on test CXR images. The table values indicated that the diagnosis technique presented by the DLS-SCD model classifies the images into distinct class labels with $sens_y$ of 86.67%, $prec_n$ of 86.86%, and $accu_y$ of 86.70% respectively. Next to that, a slighter better result has been offered by ADCD-DCNN technique classifies a set of two classes with the $sens_y$ of 84%, $spec_y$ of 90%, $prec_n$of 91%, and $accu_y$ of 87%. In line with, the DLA-CVD technique classifies two classes with the $sens_y$ of 88%, $spec_y$ of 87%, $accu_y$ of 89.50%, and F_{score} of 86%. In addition, the AD-TLCNN (3 Class) method has resulted to $sens_y$ of 98.66%, $spec_y$ of 96.46%, and $accu_y$ of 96.78%.

Besides, the AD-TLCNN (2 Class) method has obtained $sens_y$ of 98.66%, $spec_y$ of 96.46%, and $accu_y$ of 96.72%. Simultaneously, the VGG-16 model has outperformed the earlier models with $sens_y$ of 86.94%, $spec_y$ of 97.94%, $prec_n$ of 82.46%, $accu_y$ of 96.79%, and F_{score} of

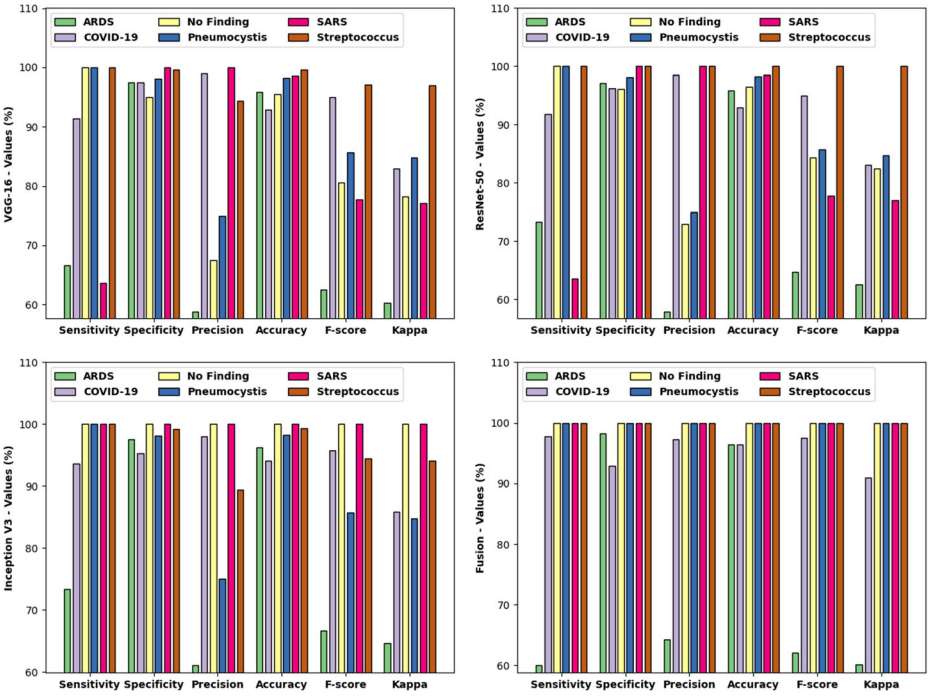


Fig. 9 Average Classifier results analysis of proposed model

83.12%. In line with, the ResNet-50 model has offered somewhat better results over the VGG-16 model with the average $sens_y$ of 88.13%, $spec_y$ of 97.95%, $prec_n$ of 84.07%, $accu_y$ of 97.04%, and F_{score} of 84.60%. At the same time, the Inception v3 model has achieved near-optimal results with $sens_y$ of 94.49%, $spec_y$ of 98.38%, $prec_n$ of 87.28%, $accu_y$ of 98%, and F_{score} of 90.44%. However, the proposed fusion model has shown effective results over all the

Table 6 Comparison Study of proposed with different models [18]

Methods	No. of Classes	$Sens_y$	$Spec_y$	$Prec_n$	$Accu_y$	F_{score}
Fusion Model	6	92.96	98.54	93.60	98.80	93.26
Inception V3 Model	6	94.49	98.38	87.28	98.00	90.44
ResNet-50 Model	6	88.13	97.95	84.07	97.04	84.60
VGG-16 Model	6	86.94	97.94	82.46	96.79	83.12
DLS-SCD	3	86.67	86.75	86.86	86.70	86.70
ADCD-DCNN	2	84.00	90.00	91.00	87.00	86.00
DLA-CVD	2	88.00	87.00	89.97	89.50	77.00
AD-TLCNN (2 Class)	2	98.66	96.46	91.93	96.78	92.09
AD-TLCNN (3 Class)	3	98.66	96.46	91.89	94.72	93.13
FM-HCF-DLF	4	93.61	94.56	92.85	94.08	93.20

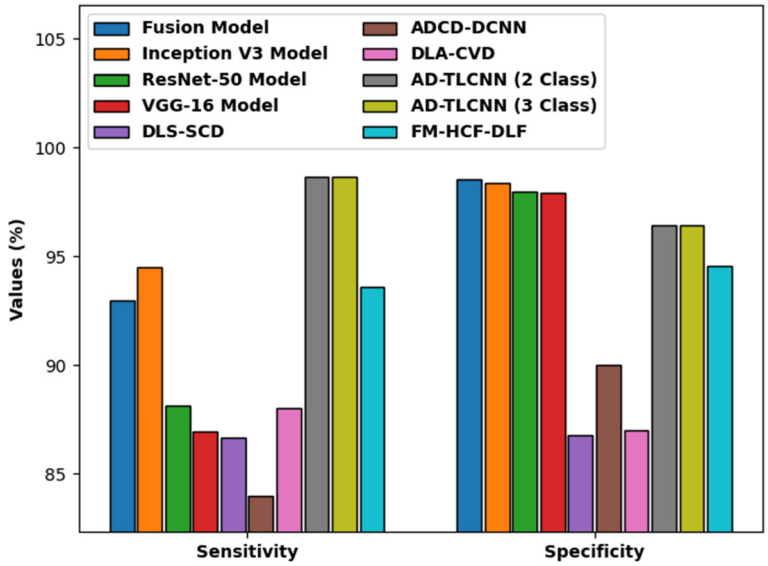


Fig. 10 Comparative results analysis of different models in terms of $sens_y$ and $spec_y$

compared methods by offering maximum $sens_y$ of 92.96%, $spec_y$ of 98.54%, $prec_n$ of 93.60%, $accu_y$ of 98.80%, and F_{score} of 93.26%.

By looking into the above-mentioned results, it is evident that the MMF-DTL technique is found to be an effective tool for COVID-19 diagnosis using radiological images. The

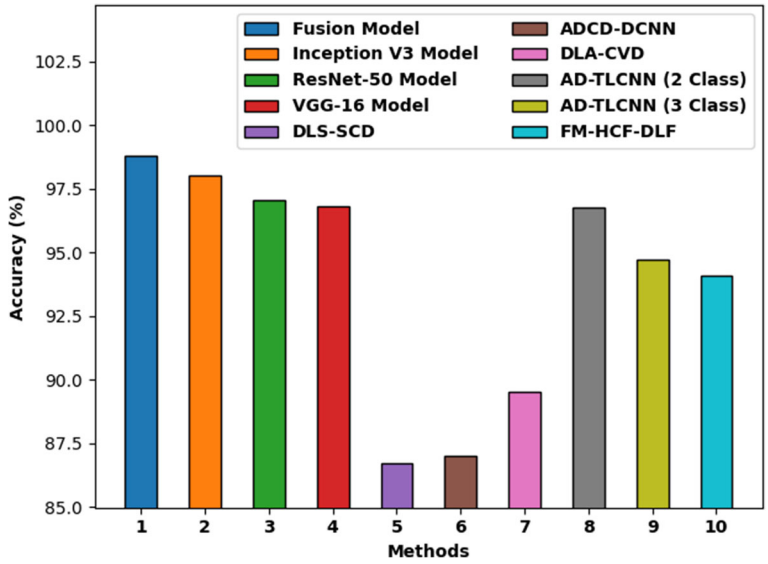


Fig. 11 Comparative results analysis of different models in terms of $accu_y$

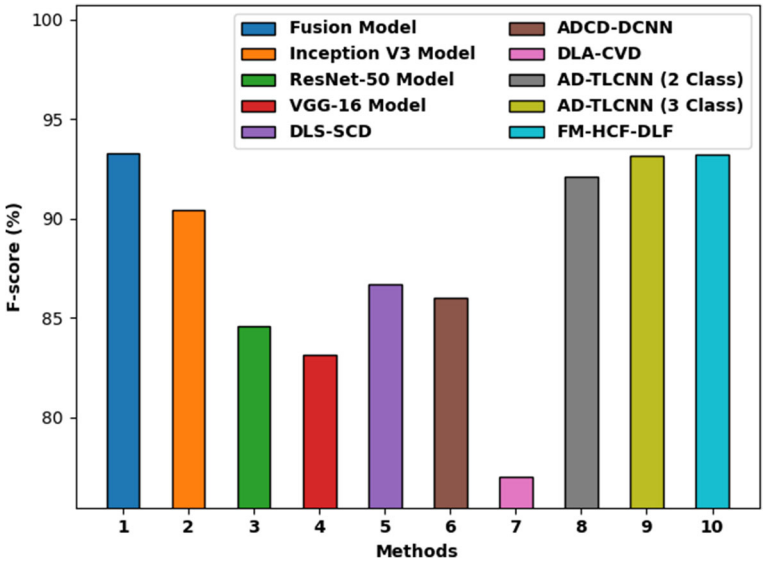


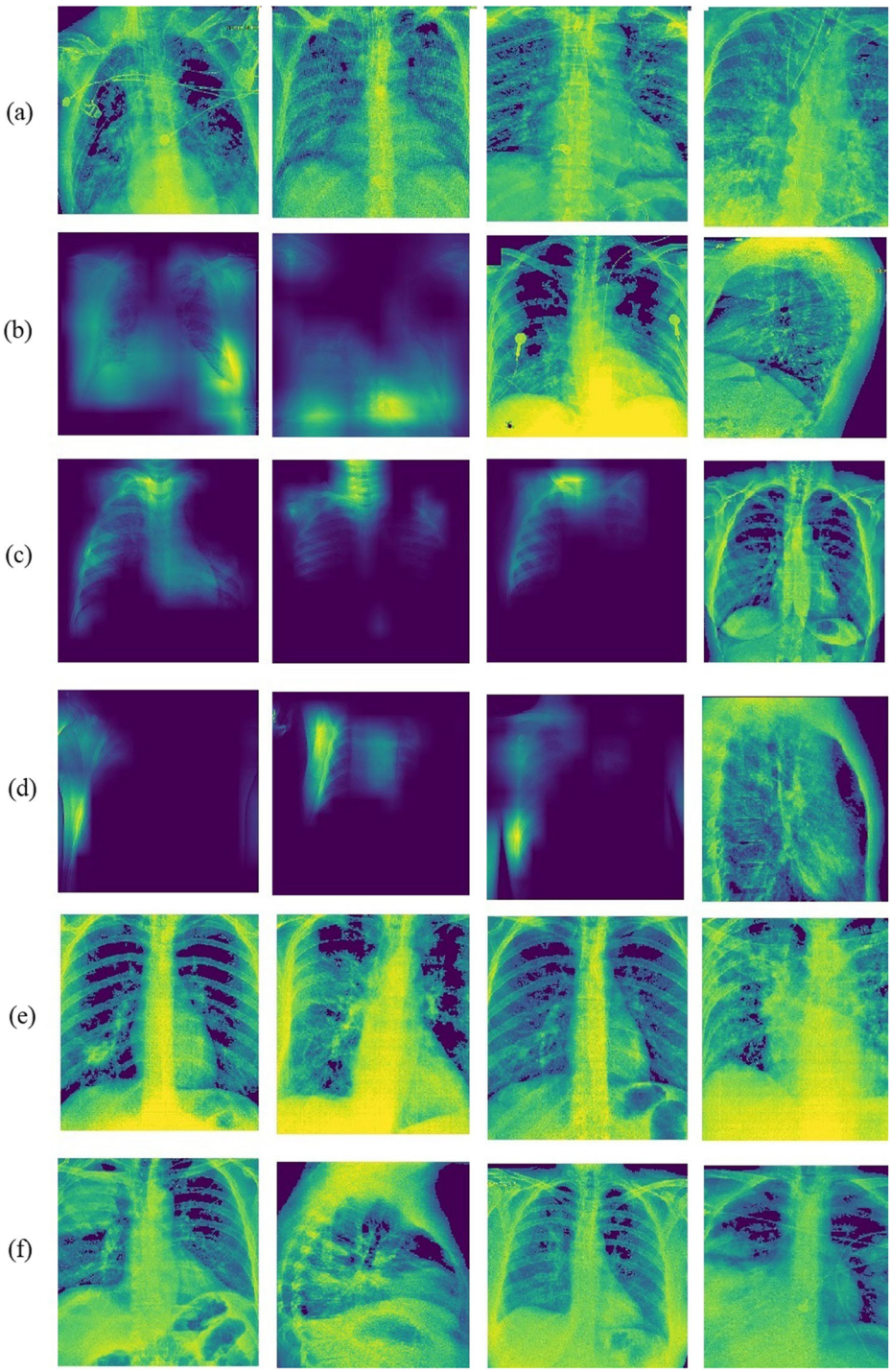
Fig. 12 Comparative results analysis of different models in terms of *F _ score*

simulation outcome concluded that the MMF-DTL model has offered better diagnosis results by attaining average $sens_y$ of 92.96%, $spec_y$ of 98.54%, $prec_n$ of 93.60%, $accu_y$ of 98.80%, and an F-score of 93.26%. The application of the fusion process in the MMF-DTL model helps to achieve a higher detection rate.

5 Conclusion

This study has presented a new MMF-DTL technique to detect COVID-19 using CXRs. A new user interface has also been created to assist healthcare professionals in the classification process. The MMF-DTL model has initially preprocessed in three stages, namely contrast enhancement, artefacts removal, and image sharpening. Then, the feature extraction and classification processes are performed. The validation of the MMF-DTL technique is performed against the Chest X-Ray dataset and the images are classified into a set of six classes. The experimental values demonstrated that the MMF-DTL model has offered better diagnosis results with an average sensitivity of 92.96%, specificity of 98.54%, precision of 93.60%, the accuracy of 98.80%, and an F-score of 93.26%. Therefore, the proposed MMF-DTL model is found to be more effective than other DL models and can be employed for real-time classification processes. In the future, the performance of the MMF-DTL technique can be enhanced by the use of a large-sized dataset to improve the training process, and thereby the detection rate can be significantly improved. Besides, the proposed model can be tested on real-time images in our future work.

Appendix 1



◀ **Fig. 13** Class Activation Map (**a**) ARDS (**b**) COVID-19 (**c**) No Findings (**d**) Pneumocystis (**e**) SARS (**f**) Streptococcus (The figure in the second row indicates that the COVID-19 affects right side of the organ compared to other classes)

Appendix 2

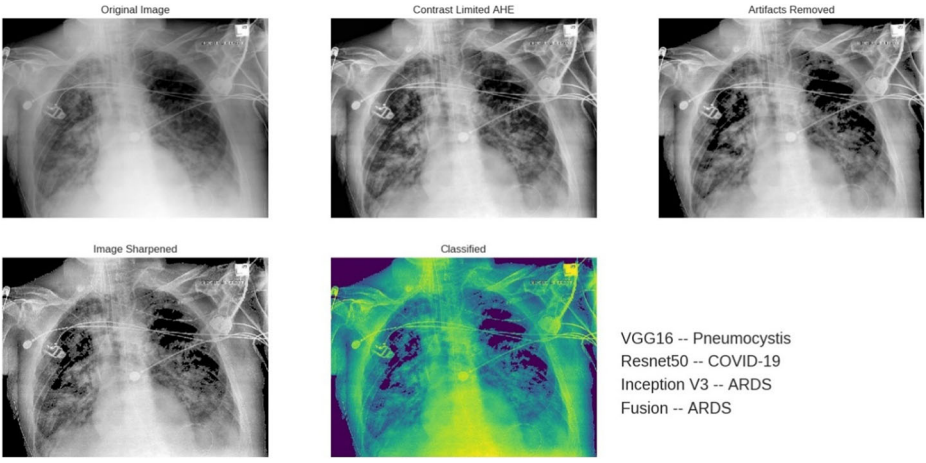


Fig. 14 Visualization Steps of MMF-DTL model: ARDS

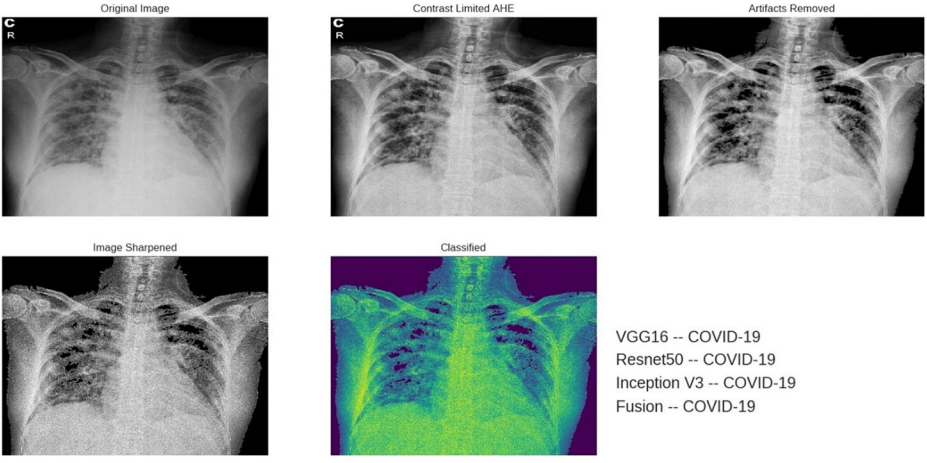


Fig. 15 Visualization Steps of MMF-DTL model: COVID-19

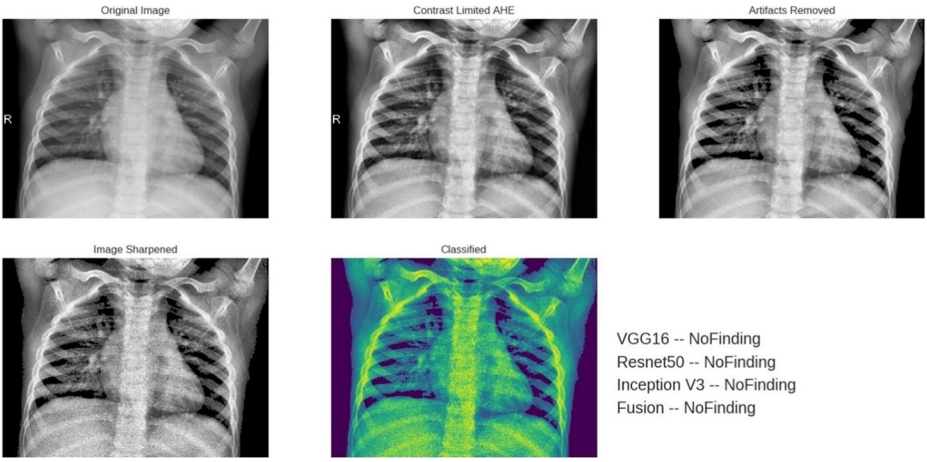


Fig. 16 Visualization Steps of MMF-DTL model: No Finding

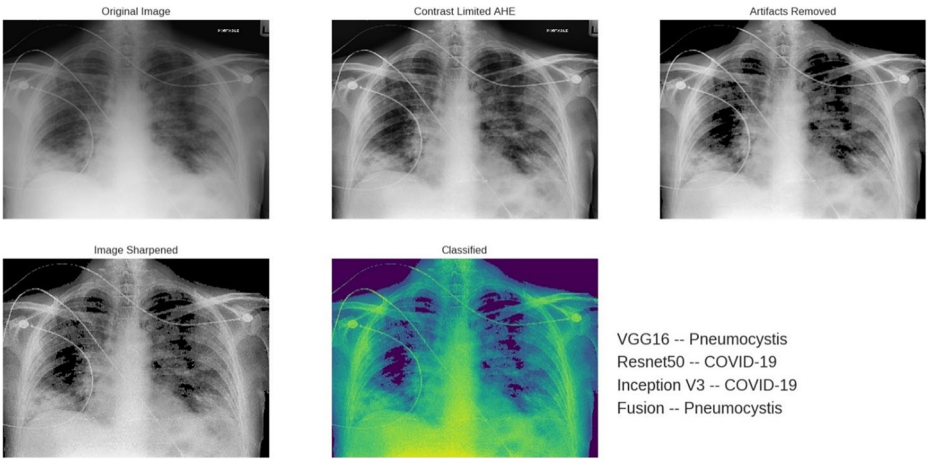


Fig. 17 Visualization Steps of MMF-DTL model: Pneumocystis

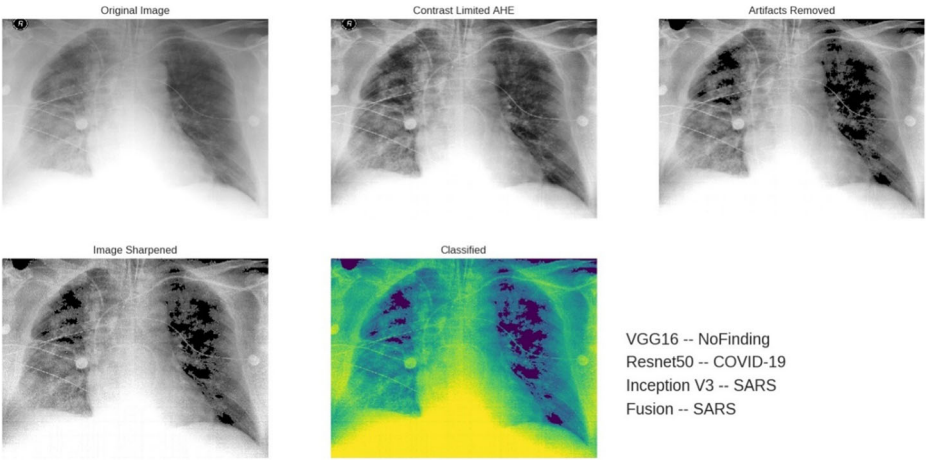


Fig. 18 Visualization Steps of MMF-DTL model: SARS

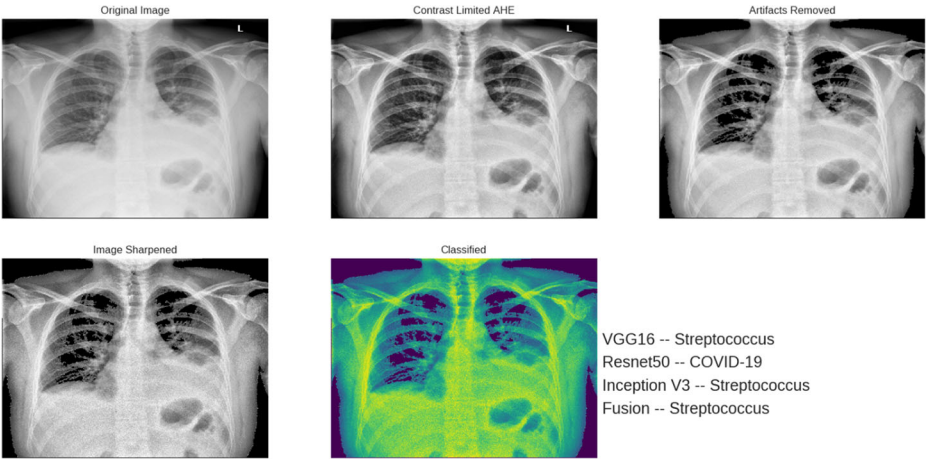


Fig. 19 Visualization Steps of MMF-DTL model: Streptococcus

Declarations

Conflict of interest The authors declare that they have no conflicts of interest to report regarding the present study.

References

1. Al-Turjman F (2020) AI-powered IoT for COVID-19. CRC Press
2. Al-Turjman F, Zahmatkesh H, Mostarda L (2019) Quantifying uncertainty in internet of medical things and big-data services using intelligence and deep learning. IEEE Access 7:115749–115759
3. Apostolopoulos ID, Mpesiana TA (2020) Covid-19: automatic detection from x-ray images utilizing transfer learning with convolutional neural networks. Phys Eng Sci Med 1:635–640

4. Ayyar MP, Benois-Pineau J, Zemmari A (2021) A hierarchical classification system for the detection of Covid-19 from chest X-ray images. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 519-528
5. Basu N, Nag S, Maitra IK, Bandyopadhyay SK (2016) Artefact removal and edge detection from medical image. European J Biomed 3(4):493–502
6. Chu DKW, Pan Y, Cheng SMS, Hui KPY, Krishnan P, Liu Y et al (2020) Molecular Diagnosis of a Novel Coronavirus (2019-nCoV) Causing an Outbreak of Pneumonia. Clin Chem 66:549–555. https://doi.org/10.1093/clinchem/hvaa029
7. He K Zhang X Ren S Sun J (2016) Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 770-778
8. Identification of Plant Leaf Diseases Based on Inception V3 Transfer Learning and Fine-Tuning (n.d.)
9. Kamal KC, Yin Z, Wu M, Wu Z (2021) Evaluation of deep learning-based approaches for COVID-19 classification based on chest X-ray images. *SIViP*:1–8
10. Kermany DS, Goldbaum M, Cai W, Valentim CCS, Liang H, Baxter SL et al (2018) Identifying Medical Diagnoses and Treatable Diseases by Image-Based Deep Learning. Cell 172:1122–1131.e9. https://doi.org/10.1016/j.cell.2018.02.010
11. Liu R, Han H, Liu F, Lv Z, Wu K, Liu Y, Feng Y, Zhu C (2020) Positive rate of RT-PCR detection of SARSCoV-2 infection in 4880 cases from one hospital in Wuhan, China, from Jan to Feb 2020. Clin Chim Acta 505:172–175. https://doi.org/10.1016/j.cca.2020.03.009
12. Nguyen D, Kay F, Tan J, Yan Y, Ng YS, Iyengar P, Peshock R, Jiang S (2021) Deep learning-based COVID-19 pneumonia classification using chest CT images: model generalizability. *arXiv preprint arXiv: 2102.09616*
13. Ozturk S, Ozkaya U, Barstugan M (2020) Classification of coronavirus images using shrunken features. medRxiv 1–13. https://doi.org/10.1101/2020.04.03.20048868
14. Quiroz-Juarez MA, Torres-Gomez A, Hoyo-Ulloa I, Leon-Montiel RDJ, U'Ren AB (2021) Identification of high-risk COVID-19 patients using machine learning. *medRxiv*
15. Rahman MA, Zaman N, Asyhari AT, Al-Turjman F, Bhuiyan MZA, Zolkipli MF (2020) Data-driven dynamic clustering framework for mitigating the adverse economic impact of Covid-19 lockdown practices. Sustain Cities Soc 62:102372
16. Rasheed J, Hameed AA, Djeddi C, Jamil A, Al-Turjman F (2021) A machine learning-based framework for diagnosis of COVID-19 from chest X-ray images. Interdiscip Sci: Computational Life Sciences 13(1):103–117
17. Salman S, Salem ML (2020) Routine childhood immunization may protect against COVID-19, vol 140. Churchill Livingstone. https://doi.org/10.1016/j.mehy.2020.109689
18. Shankar K, Perumal E (2020) A novel hand-crafted with deep learning features based fusion model for COVID-19 diagnosis and classification using chest X-ray images. Complex Intell Syst 7:1277–1293
19. Simonyan K, Zisserman A (2014) Very deep convolutional networks for large-scale image recognition. Computer Science
20. Wang D, Hu B, Hu C, Zhu F, Liu X, Zhang J et al (2020) Clinical Characteristics of 138 Hospitalized Patients with 2019 Novel Coronavirus-Infected Pneumonia in Wuhan, China. JAMA - J Am Med Assoc 323:1061–1069. https://doi.org/10.1001/jama.2020.1585
21. WHO - Coronavirus disease 2019 info web site n.d. https://www.who.int/emergencies/diseases/novel-coronavirus-2019. Accessed 6 April 2020
22. Xie M, Chen Q (2020) Insight into 2019 novel coronavirus — an updated intrim review and lessons from SARS-CoV and MERS-CoV. Int J Infect Dis 94:119–124. https://doi.org/10.1016/j.ijid.2020.03.071
23. Yadav G, Maheshwari S, Agarwal A (2014, September) Contrast limited adaptive histogram equalization based enhancement for real time video system. In 2014 international conference on advances in computing, communications and informatics (ICACCI). IEEE 2392–2397
24. Zeiler MD, Fergus R (2014) Visualizing and understanding convolutional networks. European Conference on Computer Vision. Springer, Cham. Vol.8689, pp.818–833
25. Zhang Y, Zhang E, Chen W (2016) Deep neural network for halftone image classification based on sparse auto-encoder. Eng Appl Artif Intell 50:245–255

Affiliations

A. Siva Krishna Reddy[1] · K. N. Brahmaji Rao[2] · Narasimha Reddy Soora[3] · Kotte Shailaja[4] · N. C. Santosh Kumar[5] · Abel Sridharan[6] · J. Uthayakumar[7]

A. Siva Krishna Reddy
reddyandluru@gmail.com

K. N. Brahmaji Rao
brahmaji77@gmail.com

Narasimha Reddy Soora
snreddy75@gmail.com

Kotte Shailaja
shailaja.wgl@gmail.com

N. C. Santosh Kumar
santosh.naliganti@gmail.com

Abel Sridharan
abel.sridharan@gmail.com

1 School of CS and AI, Department of CS and AI, SR University, Warangal, Telangana, India

2 Raghu Institute of Technology, Vishakhapatnam, Andhra Pradesh, India

3 Kakatiya Institute of Technology and Science, Warangal-15, Telangana, India

4 Department of EIE, Kakatiya Institute of Technology and Science, Warangal-15, Telangana, India

5 Department of Computer Science and Engineering, Kakatiya Institute of Technology and Science, Warangal-15, Telangana, India

6 Senior Manager Department of Computer Science, University of Madras, Madras, India

7 Genesys Academy of Computer Science, Puducherry, India