Radiologist-Level Two Novel and Robust Automated Computer-Aided Prediction Models for Early Detection of COVID-19 Infection from Chest X-ray Images

Munish Khanna[1] · Astitwa Agarwal[1] · Law Kumar Singh[1] · Shankar Thawkar[1] · Ashish Khanna[2] · Deepak Gupta[2]

Abstract

COVID-19 is an ongoing pandemic that is widely spreading daily and reaches a significant community spread. X-ray images, computed tomography (CT) images and test kits (RT-PCR) are three easily available options for predicting this infection. Compared to the screening of COVID-19 infection from X-ray and CT images, the test kits(RT-PCR) available to diagnose COVID-19 face problems such as high analytical time, high false negative outcomes, poor sensitivity and specificity. Radiological signatures that X-rays can detect have been found in COVID-19 positive patients. Radiologists may examine these signatures, but it's a time-consuming and error-prone process (riddled with intra-observer variability). Thus, the chest X-ray analysis process needs to be automated, for which AI-driven tools have proven to be the best choice to increase accuracy and speed up analysis time, especially in the case of medical image analysis. We shortlisted four datasets and 20 CNN-based models to test and validate the best ones using 16 detailed experiments with fivefold cross-validation. The two proposed models, ensemble deep transfer learning CNN model and hybrid LSTMCNN, perform the best. The accuracy of ensemble CNN was up to 99.78% (96.51% average-wise), F1-score up to 0.9977 (0.9682 average-wise) and AUC up to 0.9978 (0.9583 average-wise). The accuracy of LSTMCNN was up to 98.66% (96.46% average-wise), F1-score up to 0.9974 (0.9668 average-wise) and AUC up to 0.9856 (0.9645 average-wise). These two best pre-trained transfer learning-based detection models can contribute clinically by offering the patients prediction correctly and rapidly.

Keywords COVID-19 detection · Deep learning · Chest X-ray images · Ensemble models · Convolutional neural network

✉ Munish Khanna
  munishkhanna.official@rocketmail.com

  Astitwa Agarwal
  astitwa15108824@gmail.com

  Law Kumar Singh
  lawkumarcs@gmail.com

  Shankar Thawkar
  Shankar.thawkar@gmail.com

  Ashish Khanna
  ashishkhanna@mait.ac.in

  Deepak Gupta
  deepakgupta@mait.ac.in

[1] Hindustan College of Science and Technology, Mathura 281122, India

[2] Maharaja Agrasen Institute of Technology, Delhi 110034, India

1 Introduction

The first case of humans infected by Phinolophus bat originated in Zoonotic virus nomenclature in Wuhan City of China as severe acute respiratory syndrome Coronavirus-2 (SARS-CoV-2 virus) December 2019, the world observed. The city would soon become a coronavirus disease epicentre (COVID-19). The epidemic spreads around the globe later, resulting in millions of human deaths and economic losses of billions of dollars. Many people have lost their jobs, along with millions of fatalities, and the GDP of economically sound countries such as the USA, Italy, Spain, India and many others is experiencing negative growth. Most countries, financially, medically and socially, were not prepared for this type of situation where the disease, due to its transmittable nature, enters the stage of population spread. This results in many nations' complete lockdown, contributing

Springer

to an adverse psychological effect on people's health and minds.

The pandemic is widely spreading its wings across the world and is proliferating leading to the spread at the community level. The number of patients increased from 2,810,325(confirmed case) and 193,825(deaths) in April 2020 to 16,114,449(confirmed case) and 646,641(deaths) till July 27, 2020. Till August 17, 2020, 21.59 million confirmed cases were reported and 777.6 thousand death tolls [1], and upto September 17, 2020, there were 29.87million confirmed cases and 940.72 thousand death tolls (Refer Fig. 1). On 05.04.2021, there were approximately 131 million patients worldwide, and the death count reached 2.85 million. On 03.04.2021, India observed 93,249 new cases (7-day average was 73,412) and recorded 2, 40,766 new COVID-19 cases and 3,736 deaths on May 22 (Refer Figs. 2 and 3). The country has so far reported a total of 2, 65, 28,846 cases and 2, 99,296 deaths (till 22.05.2021) (Refer Figs. 2 and 3). India leads the world in the daily average number of new infections reported, accounting for one in every two infections reported worldwide each day. As of date, the top five countries infected by this disease are the USA, India, Brazil, France and the Turkey.

COVID-19 symptoms include acute respiratory disease, fever, dry cough, sore throat, chest pain or pressure, speech loss, myalgia, weakness, taste loss, weakness and nasal
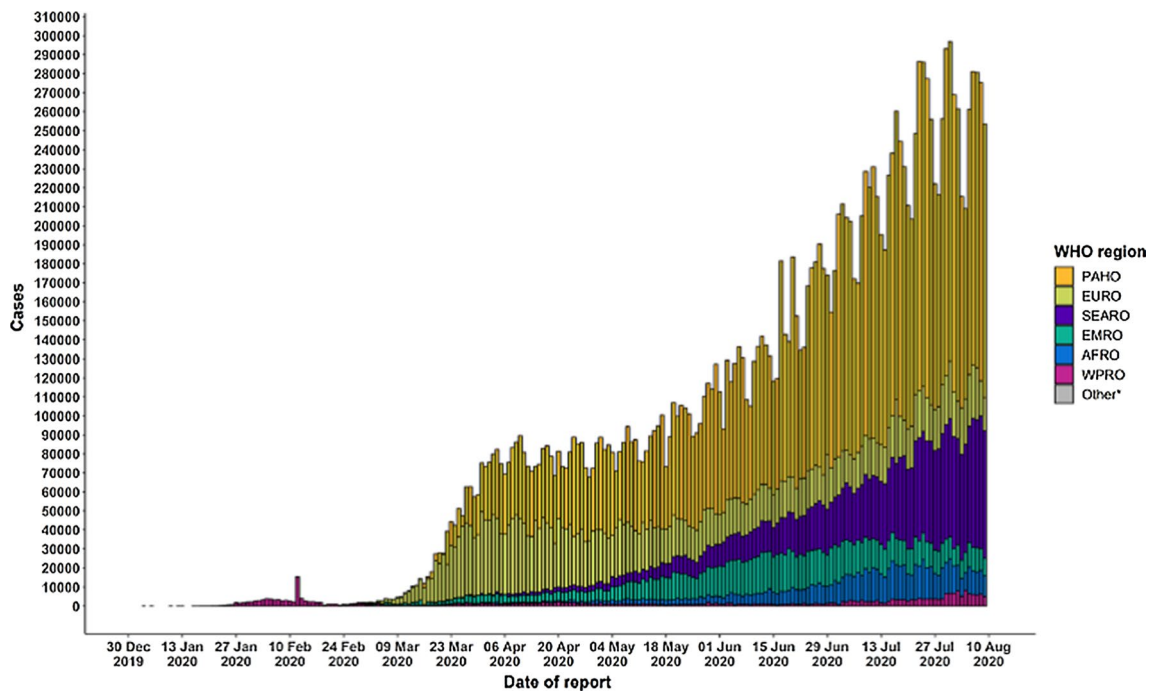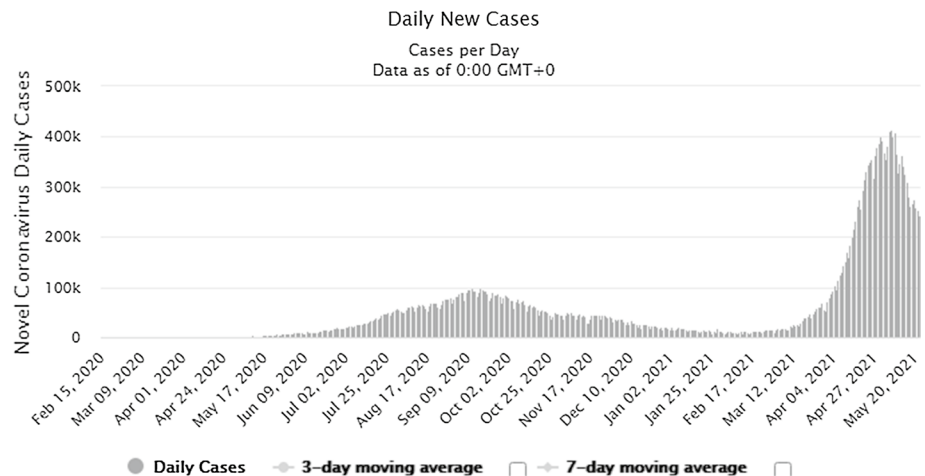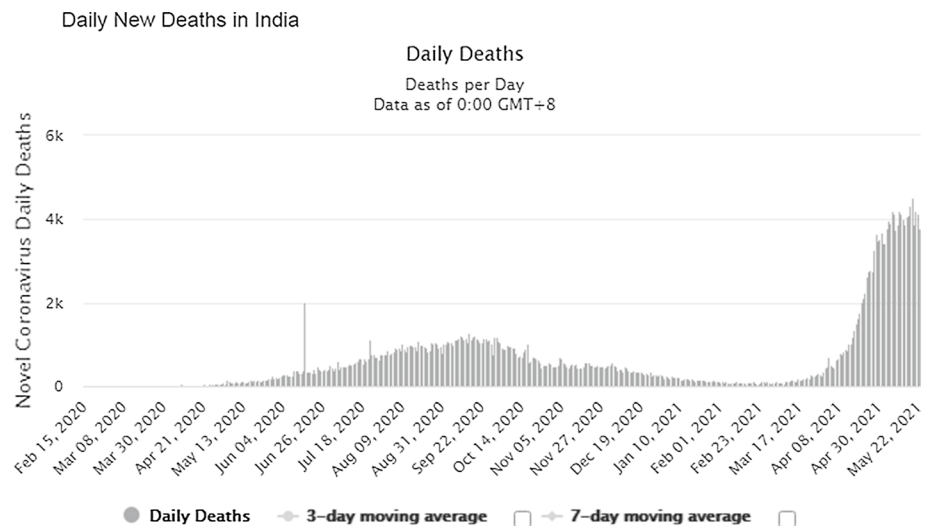

Fig. 1 Number of confirmed COVID-19 cases, by date of report and WHO region, 30 December2019 through 10 August2020 [1]

Fig. 2 Number of COVID-19 cases reported in INDIA during last 18 months. [www. https://www.worldometers.info/]

Fig. 3 Number of COVID-19 deaths reported in INDIA during last 18 months [www. https://www.worldometers.info/]

blockage, critical care respiratory condition breathing and death. The virus is mainly transmitted through respiratory droplets. The virus can also be found on surfaces and in the environment. Patients aged more than 50 years and having diabetes, cancer, cardiovascular diseases and chronic respiratory diseases are advised to follow precautions and preventions strictly. Many studies indicate 80% of mild-to-moderate patients recovered without hospitalization. Some specific patterns such as bilateral involvement, peripheral distribution and basal predominant ground-glass opacity, multifocal patchy focal consolidation, vascular thickness and crazy-paving pattern with peripheral distribution are seen in chest CT images of positive patients [2].

There is no known exact cure; only prevention such as washing hands, covering the mouth and nose and social isolation are available. Early diagnosis will allow for quarantining patients, rapid intubation in specialized hospitals and monitoring of the diseases' spread. There is a great need for the system to identify an infection to predict future patients quickly. Hence, we can minimize the death rate with the help of early diagnosis and immediate treatment. The medical tool specification (PPE package, ventilators, beds and masks) will increase if the disease spreads further. The most appropriate molecular assay is RT-PCR, which detects the SARS-CoV-2 RNA using nasal or throat swabs, but kits of these have some significant disadvantages. These disadvantages include a potential lack of methodology, resource-intensive, low production, less sensitivity, as compared to CT images, and their diagnosis time is 6–9 h [3–6]. PCR tests involve several stages where errors may occur.

While the thoracic radiological examination was accepted as the key to diagnosing potential COVID-19 patients, the approach faces its difficulties. Studies have shown that non-COVID-19 patients with community-acquired infections caused by agents such as Streptococcus pneumoniae,

Mycoplasma pneumonia and Chlamydia pneumoniae are present [7]. The CT symptoms are identical to those found in patients with COVID-19. Symptoms such as fever, cough and tiredness are not peculiar to COVID-19 pneumonia and are seen in pneumonia cases infected with the virus. In comparison, clinical practice has demonstrated no abnormality in CT images of certain COVID-19 patients, raising the complexity of diagnosing new coronavirus pneumonia infections in patients [7]. RT-PCR testing could be inaccessible in all confirmed cases due to the rapid dissemination of COVID-19 pneumonia [7]. A detailed diagnosis of COVID-19 pneumonia in patients with clinical symptoms and CT signs by an easy-to-implement approach will be helpful to implement selective and efficient isolation. Another type of test is also called an antibody test, which is a slow process and takes about 9–28 days to validate the patient as 'positive.' There is a risk that the infected person can multiply the infection these days if the quarantine is not successful.

The radiologist informs the patient of early (and timely) diagnosis, treatment and distinguishing of COVID-19 from other similar viruses (MERS). In this type of pandemic situation, a substantial increase in chest CT's is necessary. We require highly experienced radiologists, but there is a shortage of them [8, 9]. This will help improve their performances by providing them with a computer-aided alternative. Moreover, radiologists have some issues, too; the COVID-19 diagnosis is based on the radiologist's assessment and evaluation of CT and X-Ray images. The work can be time-consuming, and it is often riddled with intra-observer variability. To overcome these limitations, we need an advanced Artificial Intelligence system that is reliable and reproducible. With these systems, we may reduce the dependence on radiologists. Discussing from the other perspective, World Health Organization (WHO) has strongly recommended exhaustive testing of potential suspects and

citizens living in hotspot areas, but the guidelines are not sincerely followed due to the scarcity of the resources/personnel/testing-kit. So it is an immediate priority to develop and test Artificial Intelligence (AI)-based fast and robust alternative approach that can help limit the rate of spread because prior studies, like lung cancer detection [10], heart disease prediction [62], liver lesions in CT scans [63], have justified the performance of AI-based models in clinical practice. These AI-assisted models have high accuracy, which will aid clinical in best decision making. Prior studies [11, 12] conclude that manual examination in the case of X-ray images provides precise recognition of the disease in 60–70%, and in the case of the CT images, manual analysis performance lies between 50 and 75%. The experienced radiologists decide based on prior clinical experience; however, such verdict is easily influenced by subjective factors and individual proficiency. More automated and intelligent methods are needed to search globally, especially in regions with limited resources. Deep learning has been a cornerstone in AI healthcare and medicine for over a decade. Then there should be a computer-based (deep learning approach) alternative for diagnosis of the disease. It supports users in the COVID-19 screening. The last issue in using the manual methods for evaluating chest X-rays is the low sensitivity of ground-glass opacity (GGO) nodules. Hence, early detection of the disease using X-ray is not possible. Deep learning can differentiate between nodules that an untrained human will not detect.

The above discussed factors inspire researchers to discover an artificial intelligence-based alternative health system solution that can diagnose the disease more cheaply, produce results in less time and be highly accurate. X-ray images are limited by low patient exposure and result in the diagnosis of soft tissues or disease areas in the thorax. Computer-aided diagnosis (CAD) systems eliminate the limitations of chest X-rays and help radiologists detect potential conditions automatically in low-contrast X-ray images. Computing technology makes use of more advanced components and recent image processing algorithms. Deep learning has recently become the most advanced approach in AI. They can learn by labeling previously labeled examples, like learning facial expression recognition in computers. A new subfield of computer vision involves using neural networks to analyze medical images. CT results cannot be reproduced accurately when lungs are affected in a large amount. AI-based artificial intelligence can provide a complete analysis.

With the rapid advancement of computers and their applications in multidisciplinary fields, the researcher's fraternity in the medical field, such as organ segmentation and image enhancement and repair, is actively introduced, actively promoting medical diagnosis [13, 14]. By being additional supporting instruments for clinicians, machine learning has proved its applicability in the medical field [15, 16].

Deep learning methods, such as CNN, have been shown to predict clinical results from the classification of skin and breast cancer classification, glaucoma detection, lung segmentation, lung-related abnormalities like cancer and nodule detection, chest X-ray pneumonia detection, classification of brain disease, detection of arrhythmias, viral infections and biomedical studies [11, 17, 18]. The CNN models have also shown their ability to achieve human-like accuracies, or more than that in concise span, in image classification issues. Expertise in diagnosis can be found by digitizing and standardizing image knowledge to help medical professionals make a perfect decision [19].

The above discussion motivates us to develop a Computer-Assisted Diagnosis (CAD) framework based on deep learning to predict COVID-19 infection. The device must deliver an appropriate screening method that can distinguish the CT and X-ray images of potential COVID-19 patients from related viral pneumonia, SARS and MERS diseases. The CAD concept was introduced in 1966 [17] and has been fully applied since 1980 [18]. CAD systems have demonstrated their ability to support radiologists with high trust in making their decisions (such as triage, quantification and follow-up of promising cases). They also have advantages over radiologists, such as reproducibility and subtle improvements that are not apparent by visual examination.

The paper's organization begins with simple COVID-19 coverage and the motivation to construct a diagnostic method based on artificial intelligence. The literature on COVID-19 chest X-ray and CT images using a deep learning method are reviewed in Sect. 2. Information on shortlisted datasets and information on models selected for experimentation is provided in Sect. 3, along with discussing appropriate settings and procedural phases. Section 4 reports the method of investigation, along with results in tabular and graphical presentations. Section 5 provides a discussion of the findings and a comparison of recent works. We presented a detailed debate on the functional applicability and outlook of our presented work in Sect. 6. Finally, we conclude in Sect. 7.

The main contribution of this paper is as follows:

1. In this paper, X-ray and CT images are used to overcome the RT-PCR sensitivity problem. It was noted that COVID-19 reveals some signatures which can easily be observed via X-ray images. Deep learning approaches can automatically analyze chest X-rays, which can accelerate the time of the investigation. Applying multiple pre-trained convolutional neural networks proposed by the scientific community and our proposed ones that are selective and reliable to achieve the best pulmonary classification X-ray picture illnesses. The generated results are remarkable and highly encouraging and prove that deep learning is successful, and more specifically,

transfer learning with CNNs to the automatic detection of abnormal X-ray images related to the COVID-19 disease. The study is an end-to-end solution for the diagnosis of COVID-19 which can be converted into a low-priced, quick, robust and reliable intelligent screening tool to reduce the error rate in binary classification between COVID-19 patients and healthy people (Non-COVID Pneumonia and Normal).

2. Extensive experimentation (16 experiments) and evaluation has been conducted using 20 pre-trained transfer learning deep models on 4 datasets to validate the best model performance while addressing the COVID-19 detection. Performance is measured by standard metrics such as accuracy, AUC and F1 score. Through 21 premium studies recently published in SCI indexed COVID-19 studies, we have concluded that we are the first to do such an extensive experiment with so many models.

3. We also have designed three novel CNN models (one from scratch, next one hybrid and last one an ensemble model) based on deep transfer learning system with different pre-trained CNNs architecture to detect the use of X-ray and CT images in COVID-19 patients. The models are quick in nature and are highly acceptable for its accuracy.

4. This study contributes to scientific research by recommending two novel best classifier models with over 98% accuracy in the detection of COVID-19 cases in all 15 experiments for the binary class task. These two best performed deep learning models would help radiologists and doctors to diagnose and decide on the disease as a hand and also as a second opinion.

1.1 Literature Survey

We have recently examined several research studies in world-renowned journals to present what scientists have recently implemented to diagnose X-ray and/or CT image COVID-19 infection using different approaches based on profound learning. Of these, more than 25 studies (Springer's 10, Elsevier's 14, and the Romanian Journal of Information Science and Technology's remaining 1) were shortlisted and recently published in highly indexed journals. Below is a concise summary of the study's key points. We also built Table 1 displayed at the end of this section to present key points of selected analysis.

Butt et al.[20] observed CT image radiographic patterns of potential COVID-19 patients using a CNN-based approach. Authors evaluated two CNN models; one was RestNet23, and the other was designed. ResNet-18 was used to extract image features. The CT images dataset achieved an overall accuracy of 86.7%. Das et al.[21] selected X-ray images to identify COVID-19 infection using a deep-transfer

learning-based approach using a modified Inception model(Xception). Eleven existing machine and deep learning models/algorithms (including one proposed) were compared; a maximum accuracy of 97.40% justifies the superiority of the proposed model. Alakus et al. [22] selected 18 laboratory findings from 600 COVID-19 patients using four different types of deep learning models and two hybrid models. In the case of tenfold cross-validation, maximum accuracy is achieved by LSTM (86.66%), and maximum accuracy is conducted in the CNNLSTM hybrid model (92.30%) train-test split approach. Ardakani et al. [23] compared ten convolutionary neural networks to diagnose COVID-19 from CT images. Over 1000 images private dataset was selected for training and testing. The authors conclude that ResNet-101 and Xception achieved the best performance of their shortlisted models. Singh et al. [24] used modified XceptionNet to identify COVID-19. The dataset contains 1419 images with fewer than 132 images of COVID cases. Compared to the other three competitive models, the proposed modified model (95.80%) achieved the best accuracy. Panwar et al. [25] proposed a new VGG-16 deep learning neural network, nCOVnet, for X-ray detection. The limited COVID-19 images, 142, are selected for training and testing; the proposed model's performance is not compared with any other pre-trained or modified image. Limited data achieved an overall accuracy of 88.10%. Wang et al. [26] applied deep learning to chest X-ray images for two purposes (discrimination and location). For COVID-19 discrimination, the first objective was lung features extraction, and the second objective was to identify infected pulmonary tissues from each detected chest X-rays. Abraham et al. [27] selected X-ray images to investigate COVID-19. The authors chose 10 CNNs and 15 multiCNNs for performance evaluation. The best result was obtained in combining five pre-trained CNN's: SqueezeNet, DarkNet-53, MobileNetV2, Xception and ShuffleNet. Authors also used correlation-based feature selection and Bayesnet classifier to identify infection. Bayesnet 's performance was compared to seven other classifiers. The authors also showed that Bayesnet's accuracy (91.15%) is equivalent to 5 CNN's best combination.

Authors, Toraman et al. [28], suggested a novel small number of layer-based imagenet, Convolutional CapsNet, for X-ray diagnosis. The model uses tenfold cross-validations to generate outputs in two classes, binary classification (COVID-19 and No-Findings) with 97.24% accuracy and multi-class classification (COVID-19, No-Findings and Pneumonia) with 84.22% accuracy. Researchers, Ozutark et al. [29], used deep neural networks to find COVID-19 infection using potential patient X-ray images. The investigators used the presented model for you to look only once at real-time object detection system. DarkNet model used to perform binary classification with 98.08% accuracy and multi-class classification with 87.02% accuracy when using

Table 1 Comparative chart of the prior studies

Paper title	Selected DataSet and its size	Models Applied	Best Model Justified
Butt et al. [20]	618 transverse section CT samples	ResNet 23 and self-crafted Self-Crafted ResNet-18-based CNN model	ResNet 23 and self-crafted
Das et al. [21]	Approximately public 627 images [28]	Inceptionnet V3 Alexnet,Resnet50 VGGNet,CNN,Deep CNN and Xccetion-based self-crafted model	Deep CNN and Xccetion-based self-crafted model
Alakus et al. [22]	18 laboratory findings of the 600 patients	ANN,CNN,LSTM RNN,CNNLSTM CNNRNN	CNNLSTM
Ardakani et al. [23]	Private dataset of 1020 images	AlexNet,VGG-16,VGG-19,SqueezeNet,GoogLeNet MobileNet-V2, ResNet-18,ResNet-50, ResNet-101,Xception	ResNet-101 and Xception
Singh et al. [24]	Public dataset of 1419 images	Modified XceptionNet	Modified XceptionNet
Panwar et al. [25]	Publicly available 337 images	VGG-16 inspired nCOVnet	VGG-16 inspired nCOVnet
Wang et al. [26]	Publicly available 3545 images	ResNet50+FPN inspired model	ResNet50+FPN inspired model
Abraham et al. [27]	531 COVID-19 images; total 1100 images	25pretrained networks (10 Basic and 15 Hybrid)	SqueezeNet+DarkNet-53+MobileNetV2+Xception+ShuffleNet
Toraman et al. [28]	Approximately public 731 images	Convolutional capsule network architecture	Convolutional capsule network architecture
Ozturk et al.[29]	Approximately public 627 images	Darknet inspired model	Darknet inspired model
Xu et al.[19]	Private 618 images	ResNet-18-based classification model	ResNet-18-based classification model
Khan et al. [30]	1200 images of two public datasets	CNN model based on Xception architecture pre-trained on ImageNet dataset	CNN model based on Xception architecture pre-trained on ImageNet dataset
Ucar et al. [31]	2800 images (consisting 45 images of COVID-19) from two public dataset	Deep Bayes-SqueezeNet inspired model	Deep Bayes-SqueezeNet inspired model
Nour et al. [2]	2905 images (consisting 219 images of COVID-19)	CNN-Machine Learning-Bayesian Optimization-based Model	CNN-Machine Learning-Bayesian Optimization-based Model
Brunese et al. [32]	6523 images (consisting 250 images of COVID-19)	VGG Inspired model	VGG Inspired model
Panwar et al. [33]	Private dataset of 526 images and Public dataset of 1300 images	Applying Grad-CAM technique in VGG-19 inspired model	Applying Grad-CAM technique in VGG-19 inspired model
Goel et al. [34]	800 COVID-19 images; total 2600 images	Self-created CNN-based OptCoNet model	Self-created CNN-based OptCoNet model
Jain et al. [35]	490 COVID-19 images; total 6432 images	Inception V3, Xception and ResNetXt	Xception
Abbas et al. [36]	105 COVID-19 images; total 200 images	Self-composed a Deep CNN-based DeTraC model; (Uses AlexNet,VGG-19,GoogleNet, Resnet, SqueezeNet for the transfer learning stage in DeTraC)	VGG-19 in DeTraC
Zebin et al. [37]	202 COVID-19 images; total 802 images	VGG-16,Resnet50 and EfficientNetB0	EfficientNetB0
Punn et al. [38]	108 COVID-19 images; total 1200 images	ResNet,Inception-v3,InceptionResNet-v2,DenseNet169, and NASNetLarge	NASNetLarge

fivefold cross-validation. The count of COVID-19 images selected for training and testing is less, and the study implements only one model. Xu et al. [19] selected CT images for screening Coronavirus disease with the support of a ResNet-based deep learning system. Initially, deep-learning models were used to segment infection regions. Using the local-attention classification model, these images were then categorized into three classes (including COVID-19). Using Noisy-or Bayesian function, the infection type and overall confidence score were calculated as a final step. Khan et al. [30] introduced a new deep-neural network-based image architecture based on Xcpetion, CoroNet, to detect and infect COVID-19 from X-ray images. They selected 284 COVID images (total 1,600 images) for performance evaluation. The system classifies images into 2-class (99% accuracy), 3-class (95% accuracy) and 4-class (89.6% accuracy) classification. In Ucar et al.'s next study [31], authors explored a novel Deep Bayes-Squeeze-based COVIDiagnosis net for chest X-ray glaucoma screening. Dataset consists of 76 COVID-19 images. The authors demonstrated their model's superiority in accuracy achieved with other existing network designs using fine-tuned hyperparameters and augmented data sets. They justify the performance by presenting the accuracy as 100% (single recognition of COVID-19 among other classes) and 98.3% (among Normal, COVID and Pneumonia). In their study, Nour et al. [2] presented a new system for detecting COVID-19 infection from chest X-rays based on deep features and Bayesian optimization. Model was trained from scratch to transfer learning approach. The deep features were extracted from the self-designed five-layer-based CNN model and finally fed into Machine Learning (ML) classifiers. Model hyperparameters were optimized using the Bayesian optimization algorithm. However, the text lacks a detailed description of extracted features. Brunese et al. [32] applied customized VGG16 deep learning model for COVID-19 chest X-ray detection. The approach consists of three phases: first, pneumonia detection, if the answer is yes, the next phase is called, which distinguishes between COVID-19 and pneumonia, and finally the last phase is implemented to locate the areas in the COVID-19 presence X-ray. In the subsequent study by Panwar et al. [33], a fusion of deep learning (VGG-19 inspired model) and Grad-CAM-based color visualization approach was employed on chest X-ray and CT-scan images for COVID-19 infection detection in the form of binary image classification. Grad-CAM-based color visualization approach was used to interpret and explains the detection of radiology images. The system attains an overall accuracy of 95.61%.

Goel et al. [34] proposed an optimized OptCoNet picture classifying model to automatically detect COVID-19 from chest X-ray images with six CNN architecture layers. The Grey Wolf Optimizer algorithm played the role of optimizing hyperparameters required for training the CNN layers. Authors Jain et al. [35] compared three models, Inception V3, Xception and ResNeXt, in the subsequent study recently published to examine their efficacy detecting COVID-19 infections in patients with radiation images. The public repository has been selected for approximately 575 COVID-19 images; the Xception model is the most exact. Researchers suggested a deep CNN(DeTraC), which DCompose, Transfer and Compose for the classification of X-ray chest pictures of COVID-19 in the subsequent empirical study of Abbas et al. [36]. Five different CNN pre-trained ImageNet models were used for DeTraC 's transfer learning phase. Combining VGG-19 and DeTraC, model achieved the highest accuracy by 97.35%. For X-ray pictures detection, Practitioners Zebin et al. [37] used VGG-16,ResNet50 and EfficientNetB0. In the following study by Pun et al. [38], researchers have used five pre-trained CNN models to identify chest X-ray infections in patients from 153 COVID-19 shortlisted images. In the subsequent article by Gianchandani et al. [5], the authors proposed two different ensemble deep transfer learning models to predict subject disease and classify subject patient images into binary and multi-class problems. Researchers selected two different datasets for validation of the proposed models. The proposed ensemble model's performance (VGG16 + DenseNet) was compared with VGG16, ResNet152V2, InceptionResNetV2 and DenseNet201 and found to be best with 96.15% accuracy on the binary dataset. Later, in the case of Kaur et al. [39] authors used the same dataset of [5] for their experimentation. Researchers show the novelty in their work by proposing modified AlexNet architecture for feature extraction and classification. Strength Pareto evolutionary algorithm-II(SPEA-II) was employed to tune the hyperparameters of this modified imagenet. The proposed model's performance was compared with 09 models in which it performs the best with more than 99% accuracy. They tested for classifying images of the patients into four classes. Another approach was developed by researchers in [40] by ensembling the deep transfer learning models, ResNet152V2, DenseNet101 and VGG16. They gather approximately 11,000 images to classify them into four classes (COVID-19, Pneumonia, Tuberclosis and Healthy). The model's performance was compared with 06 existing models, and the generated results reveal that the proposed one was the best with more than 98% score (in all parameters like Accuracy, AUC, F-measure, sensitivity and specificity). Jaiswal et al. [41] presented a pre-trained deep learning architecture (DenseNet201) to classify COVID-19, infected patients, on the basis of their CT scan. They evaluate the performance with the other three competitive models (VGG16, InceptionResNet and DenseNet) on five efficiency measuring parameters. Authors conclude that their presented model with more than 96% score (in all parameters like Precision, Recall, F-Score, Specificity and Accuracy). In just recently published study

researchers, Tiwari et al. [42] suggested a system based on visual geometry group capsule network (VGG-CapsNet) to screen COVID-19 disease based on input images. The composed dataset contains around 2900 images belonging to three classes (normal, pneumonia and COVID-19). Their presented model achieved an accuracy of 97% for segregating COVID-19 samples from non-COVID-19 samples and 92% accuracy for segregating COVID-19 samples from normal and other viral pneumonia samples.

2 Materials and Models

The dataset used and the methodology used is explained in the subsequent sections.

2.1 Dataset Description

The first three datasets selected for the validation of the shortlisted models are open-access publicly available repositories. The fourth one is a self-crafted dataset whose images (X-ray and CT both) are collected from the internet and private hospitals. As per the best of our knowledge (gathered from various shortlisted prior published studies), we are the first to experiment with four datasets. Next, as per the acquired knowledge, very few studies have tested

on so many COVID-19 images; although the count of the COVID-19 images is not very large still it is greater than experiments performed by previous researchers. After that, regarding the first dataset, we want to communicate that as per the best of our knowledge, none of the earlier studies have experimented on this dataset; this dataset belongs to the University of California, where an open coding competition was going on (01.09.2020). Last observation regarding dataset, while going through these studies, the practitioners have discriminated the images into classes (X-ray or CT). They have selected either X-ray images for experimentations or CT images for the same. We have overcome this discrimination by creating our customized dataset consisting of both X-ray and CT images of COVID and non-COVID patients. The details of the images are stated in Table 2.
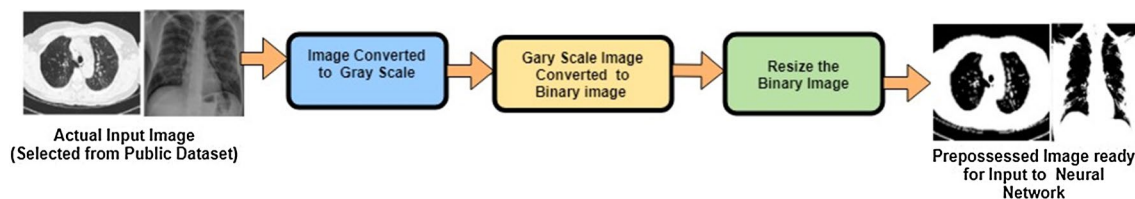
2.2 Preprocessing

Images in the dataset may be of different sizes, so we need to pre-process them using the steps displayed in Figs. 4, 5. Initially, all the images are passed through the preprocessing module, and then

they are converted into a gray scale using OpenCV. After that, we have converted images into the binary conversion of images. We then resized each image into 224×224 pixels for the following processing steps in the deep learning

Table 2 Datasets used in the study

Data Set Serial Number	Total Number of Images	Count and nature (COVID/Non-COVID)	Address (Last Access Date 01/10/2020)	X-ray/CT images
1	746	COVID: 349 Non-COVID: 397	https://github.com/UCSD-AI4H/COVID-CT/tree/master/Images-processed	X-ray images
2	516	COVID: 278 Non-COVID: 238	https://ieee-dataport.org/open-access/COVID19action-radiology-cxr	X-ray images
3	418	COVID: 210 Non-COVID: 218	https://www.kaggle.com/bachrr/COVID-chest-x-ray?select=images	X-ray images
4	770	COVID: 142(CT images) and 183(X-ray) Non-COVID: 207(CT images) and 238(X-ray)	Private Customized dataset. Images collected from different resources available on internet and private hospitals located in author's hometown	X-ray images and CT images

Fig. 4 Pictorial representation of the initial stage

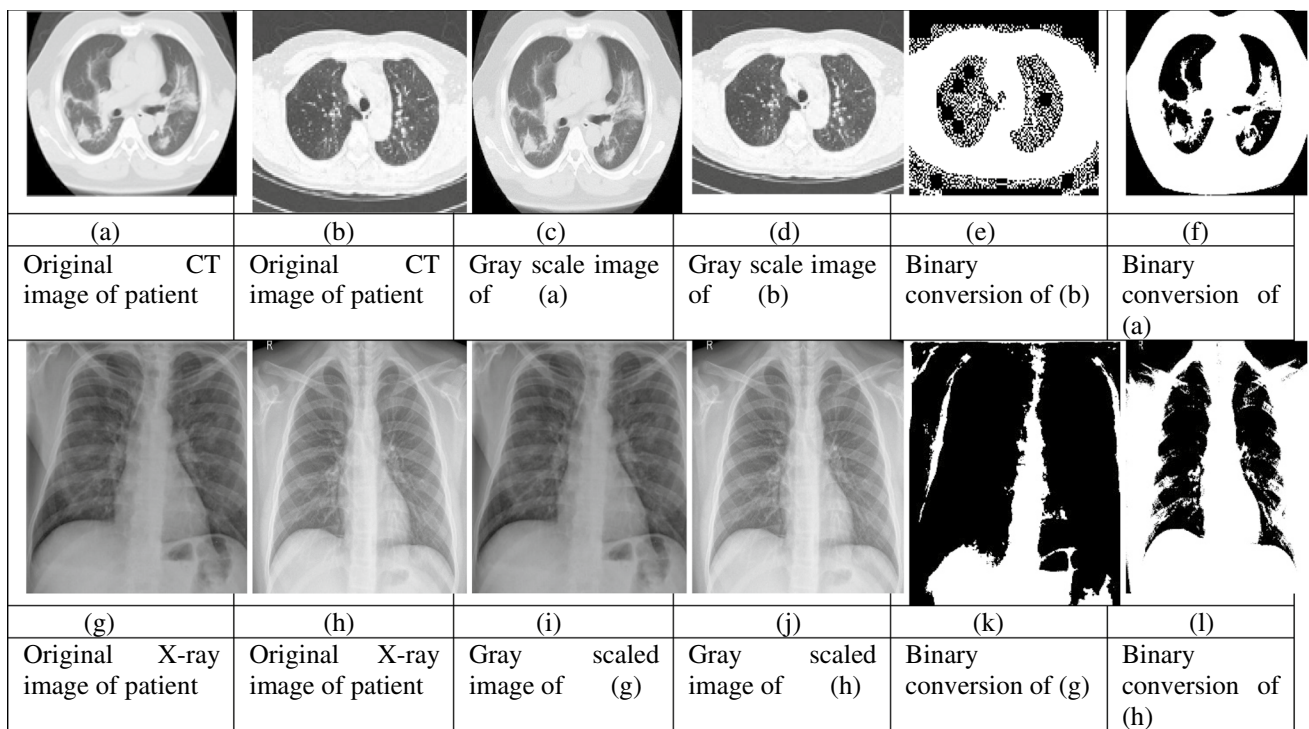| (a)	(b)	(c)	(d)	(e)	(f)
Original CT image of patient	Original CT image of patient	Gray scale image of (a)	Gray scale image of (b)	Binary conversion of (b)	Binary conversion of (a)
(g)	(h)	(i)	(j)	(k)	(l)
Original X-ray image of patient	Original X-ray image of patient	Gray scaled image of (g)	Gray scaled image of (h)	Binary conversion of (g)	Binary conversion of (h)

Fig. 5 Outputs of the various processes implemented during initial stage

framework pipeline. The output of the images after conversion is shown in the above diagram. We have then applied different flips and rotation while executing data augmentation as more images will help train the data out of which are available to us.

2.3 Deep Learning, CNN, Transfer Learning, Fine-Tuning and Ensemble Method

Deep learning (DL) is a branch of machine learning that discusses approaches focused on data representation learning. Deep learning with convolution is an effective way to simplify data processing. Computer models learn to perform classification tasks from pictures, text and sound. DL methods using several neural network layers can precisely extract and acquire specific characteristics from the data. These models are reliable, often sometimes outperforming humans. These models are trained by inputting a wide range of data. In Deep Learning, superior model efficiency can be observed as data quantity increases. With traditional methods, the output is the same after an initial wave. With vast volumes of data, Deep Learning will also be an optimal alternative to current classification methods. Additionally, CNN deep learning has fascinating advantages like low processing time, superior performance, high computing power and reliable outcomes (Prediction). Recent advancements in neural network architecture design and training have allowed

researchers to solve previously unsustainable DL method learning tasks. Profound learning algorithms are contrasted with traditional algorithms. Convolution neural networks are one of the most popular DL approaches in medical imaging. The DL approach is a multi-stacked convolutionary algorithm. The CNN is composed of a convolutionary layer, a nonlinear layer and a softmax layer, a maximum or average pooling layer. Convolutionary layers also use pooling layers to increase invariance and reduce function map computing costs. It subsamples, the convolutional layer to reduce the feature map's size and computes the maximum and average function over the convolutional layer and are called max pooling and average pooling concerning the function they perform. Spacing in the pixels of the image is used with pooling and is called stride.

The number of parameters in the model grows as networks deepen to optimize learning performance. Deep learning models require large volumes of data. Deeper networks lead to more complicated computation and more difficult training data. We don't have many photos to train, check and evaluate data. It appears too small to train a deep CNN, and we need a definition of transfer learning. Transfer learning uses the CNN model in an extensive database to understand the objective mission (e.g., diagnosis of COVID-19 from X-ray CT scans) with limited preparation details. Learning transition is the mechanism by which previously acquired information is passed to an alternate mission. This means

we move information (learning descriptive and widespread representations of features) to our database's insufficient issue. We are concerned about the use of this technique. Big picture data belong primarily to the general domain, such as pet, dog and chair, while our pictures are X-ray and CTs(of COVID-19) with different visual appearances. As a result, the visual representations learned from these large images cannot represent CT images correctly. It makes the network extraction feature partial to the source data and less widespread to the target data. The pre-trained models are also ideally suited to our current data collection of fewer numbers of images. In other words, our task is to change the pre-trained CNN structures. Usually, this approach is much easier than conventional CNN style random weight lifting. Too much data are required to train the enormous CNN model parameters correctly. We use transfer learning to mitigate the scarcity of such datasets and produce improved performance. A vast number of different types of images are trained in various special CNN architectures and are then considered the pre-trained CNN model. However, these models need a lot of data to prevent overfitting.

To solve this dilemma, techniques of transfer learning are applied. Transfer learning (TL) considers the standard neural architecture and prepared weights on broad datasets and then changes the weights of the goal task with limited training data. Transfer learning is a technique used to pretrain models to its best growth so that we can have better efficiency than the present model we are using. It is a machine learning approach in which a model designed for one task is used as the foundation for a model on a different task. Given the large computing and time resources needed to build neural network models on these problems and the massive leaps inability to provide on similar problems, it is a common method in deep learning where pre-trained models are used as the starting point on computer vision and natural language processing tasks. When modeling the second mission, transfer learning is an optimization that makes for faster development or increased results. Transfer learning enhances the learning of a new task by transferring information from a previously learned similar task. When modeling the second mission, transfer learning is an optimization that makes for faster development or increased results. TL is associated with multi-task learning and paradigm drift, and it is not solely a field of research for deep learning. Transfer learning is common in deep learning due to the immense resources needed to train deep learning models or the vast and challenging datasets used to train deep learning models. Deep learning only fits with transfer learning if the model features learned from the first challenge are standard. Shift learning entails initially training a base network on a base dataset and task and then repurposing or moving the learned features to a second target network trained on a target dataset and task. This process is more likely to succeed if the functionality is universal, applicable to both the base and target tasks, rather than unique to the base role. By reducing errors, an ensemble of transfer learning networks can be a stable solution. It generates optimum outcomes from the integrated networks for the fewest errors necessary. The convolutional neural network architecture is designed utilizing pre-trained models after data preprocessing. Researchers also applied the same approach in the recently published study for COVID-19 prediction [5, 41, 43, 44]. Researchers in their work [5] designed ensemble deep transfer learning models to predict this disease from chest X-ray images. Authors focused that TL uses feature extraction capabilities of pre-trained models on large datasets such as ImageNet, capturing class boundaries well. TL not only eliminates the need for a large dataset but also provides quick results. This motivates authors to incorporate these techniques into their proposed deep learning framework for COVID-19 classification. The authors of [43] employed a deep convolutional neural network as the backbone. They suggested three improvements (i) stochastic pooling to replace average classical pooling and maximum pooling methods; (ii) developing conv block by combining conv layer and batch normalization, and (iii) creating a fully connected block by combining dropout layer and fully connected layer. In the following study [44] by Zhang et al., the authors developed a novel method combining DenseNet and optimization of transfer learning setting (OTLS) strategy to solve COVID-19 prediction. Authors suggest that transfer learning is a solution to (i) accelerate developing CNN and (ii) avoid overfitting. Authors explore that with this technique's employability, researchers can develop a better performing deep neural network rapidly. The authors of [41] justify that hyper-tuning of Deep transfer learning models (DTL) can improve the results further; a DTL model with DenseNet201 was suggested in this study.

TL acquires the knowledge from solving another problem and applies that knowledge to a different but related problem. Authors further explored that three elements are paramount to support the transfer are (i) The triumph of pre-trained models supports the user get rid of hyperparameter tuning" (ii)The initial layers in pre- trained models can be thought of as feature descriptors which extract low-level features and (iii) The target model may only need to re-train the last several layers of the pre-trained model. In this technique, we take one model, for example, VGG-16, and train it first; then we freeze all the models' layer; after that, we apply layers from our side on the model so we train the model on that part using the weights of VGG-16. This is a solid and robust technique for increasing the efficiency of the model. In other words, our approach is to train a solid, deep network to extract comprehensive visual properties by preparing broad data sets and then translating the network weights of this deep network into a small dataset's targeting mission. Fine-tuning is used to fine-tune the model we are using after

the transfer learning technique; we used two parameters for fine-tuning early stopping and reducing rl over plateau. This technique helps to make the graph straighter and to reduce ridges in the graph. Early stopping is mainly used to stop the model when the validation accuracy or testing accuracy reaches its best height and when the loss is minimum in validation accuracy and training accuracy. For example, if we have used 50 epochs to train the model, the training is more sure to stop before 50 epochs when minimum loss and maximum accuracy have been reached. A general algorithm for all the models can be visualized as given below.

models must be good in various ways; they must have different prediction errors. Combining projections from several neural networks introduces bias, which offsets the volatility of a single learned neural network construct. The result is forecasted that are less vulnerable to the particulars of the training data, training scheme selection, and the chance of a successful training run. In addition to minimizing prediction uncertainty, the ensemble can provide better predictions than any single model. This method is part of a broader class of techniques known as "ensemble learning," which defines methods that aim to make the best use of predictions from

---

Input: COVID 19 different datasets

μ = learning rate of CNN ; α = iteration per step

ϵ = maximum number of epochs; Π = the number of images covered in one iteration

1. Determining the layers ;   2. Setting the head layer of each model
3. Initializing the CNN;   4. Training the CNN
5. Testing the CNN
   a. For λ =1 to ϵ do
      i. We will select a batch size of 16
      ii. We will compute the loss of forward propagation
      iii. After that using Nadam optimizer we will compute the backward propagation.

Nonlinear methods include neural network simulations. Nonlinear methods include neural network simulations. This means they can discover dynamic nonlinear interactions in results. The disadvantage of this adaptability is that they are vulnerable to initial conditions, both in terms of initial random weights and statistical noise in the testing dataset. Since the learning algorithm is stochastic, each time a neural network model is learned, it can retain a marginally (or dramatically) different version of the mapping function from inputs to outputs, resulting in other performance on the training and holdout datasets. As such, we should consider a neural network to be a system with a low bias and a high variance. When trained on massive datasets to accommodate the high variance, allowing some variance in a final model designed to render predictions can be disappointing.

The high variance of neural networks can be resolved by training several models and integrating their predictions. The aim is to incorporate forecasts from many strong yet distinct models. A good model has talent, which means that its predictions outperform random chance. Notably, the

several models prepared for the same challenge. In general, ensemble learning entails training several networks on the same dataset, then predicting using one of the trained models before integrating the predictions in some manner to produce an outcome or forecast. Indeed, the model ensemble is a common technique in advanced machine learning to ensure the most reliable and best possible prediction. So the main focus for using the ensemble approach is reducing the variance of neural network models.

Ensemble learning is a method in which all models are combined using either soft or hard voting. For this, we use a voting classifier. Using a single model, the model's boundary curve is not as smooth as the ensemble produced. The curve that follows the ensemble better separates the COVID and non-COVID cases than the curve that precedes it. As a result, we used Ensemble learning with a voting classifier to reduce our model's errors. The voting classifier also aids in reducing the likelihood of overfitting, thus improving the models' overall accuracy. Ensemble approaches are among the most appealing techniques. First and foremost,
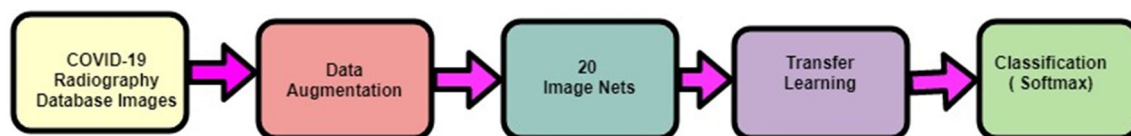

Fig. 6 Block diagram representation of the proposed work

it connects the distinct core model through the Ensemble framework and generates a new optimal predictive model. Bagging is a form of ensemble process that uses bootstrapping and aggregation to create the ensemble system. Bootstrapped selected a random section of the dataset and built a decision tree for each random partition of the dataset. Following the development of each partition of the Decision Tree dataset, an algorithm is used to aggregate the Decision Trees to construct the most valuable predictors. The ensemble method's main aim is to generate the best predictor after combining various models. By reducing errors, an ensemble of transfer learning networks can be a robust solution. It creates optimum results from the combined networks with the fewest errors possible (Refer Fig. 6).

A hyperparameter is a machine learning technique that can improve a classical neural network's overall performance. This optimization method chooses the best parameters to describe the model architecture, also known as hyperparameters. The entire training process, precision, and necessary modifications are used to tune hyperparameters. The model is being revised to find the best combination to meet the challenge. The hyperparameter can be tuned to improve classification techniques. Hidden layer number, learning rate, epochs, batch size, activation functions, and so on are examples of common hyperparameters.

2.4 Motivation

The rapid spread of the novel coronavirus (COVID-19), which has since become an ongoing pandemic, has necessitated fast detection methods to avoid further virus spread and mortality. Globally, three methods are used for screening COVID-19 infectious patients. This includes screening via chest X-ray images, chest CT images, and the RT-PCR examination. Because of the many issues associated with the RT-PCR test, it is not always recommended on its own. Many studies have proposed using both CT/X-ray and RT-PCR parallel.

Meanwhile, the scarcity and overburdening of expert radiologists to diagnose infection from CT/X-ray is a severe problem.

Furthermore, revealing radiological signatures from suspicious patient's photographs is a time-consuming and human-error-prone operation (riddled with intra-observer variability), so we must automate X-ray/CT image analysis. Thus, the development and application of a fast and accurate early-stage classification screening method is critical, and deep learning models may be used, as one of the efficient available alternative, to meet this requirement. The scientific community has made numerous serious attempts to predict suspected infectious patients using deep transfer learning models. However, these models are still plagued by the overfitting problem. In this paper, one novel hybrid model and another created by ensembling deep transfer learning models are developed for classifying suspected artifacts into two categories (COVID-19 infected patients and healthy people). We have worked on four datasets to validate the best performing model(s). The best performing model should perform paramount on all datasets and their combinations and should not be a dataset-specific performer.

2.5 Model Formulation

For this highly intensive experimental study, we have shortlisted different state-of-the-art ImageNet pre-trained CNN networks which have proven their performance previously. We have proposed three models from our side, first one was built from scratch (called as Proposed Model in this study), and the next one is the hybrid model called as LSTMCNN and last one is the ensemble deep transfer learning model called as "**DarkNet-53 + MobileNetV2 + ResNet-101 + NASNetLarge + Xception + GoogLeNet**" in this paper. It should be noted that, to the best of the author's knowledge, this is the first attempt to use the LSTMCNN model for COVID-19 detection from X-ray and CT images, and possibly the first attempt in medical image analysis.

Table 3 List of models implemented in the study

Model	Inspired/Adapted	Model	Inspired/Adapted	Model	Inspired/Adapted
VGG-16	[60]	VGG-19	[23]	MobileNet	[47]
InceptionResNetV2	[48]	InceptionV3	[48]	ResNet-101	[49]
ResNet50V2	[50]	Xception	[51]	SqueezeNet	[52]
DarkNet-53	[53]	SqueezeNet + DarkNet-53 + MobileNetV2 + Xception + ShuffleNet	[27](Adapted)	EfficentNetB7	[54]
DCGAN	[55]	LSTMCNN	Our proposed	U-Net(Glaucoma) (Adapted)	[45]
Proposed Model	(Self-Made Model)	NASNetLarge	[38, 56]	ResNet-101V2	[61]
COVID research paper	[2](Adapted)	DarkNet-53 + MobileNetV2 + ResNet-101 + NASNetLarge + Xception + GoogLeNet	(Self-Made-ensemble deep transfer learning model)		

Moreover three models are adapted from the previous studies where they have shown their incredible performance to solve the problem; first one is named as "**COVID research paper**" which is adapted as such from [2]; next one as "**U-Net(Glaucoma)**" which is exactly adapted from [45] where this model has shown auspicious performance for Glaucoma binary classification on very large dataset; final one is named as "**SqueezeNet + DarkNet-53 + MobileNetV2 + Xception + ShuffleNet**" which is adapted from [27] as this hybrid model yielded impressive performance in COVID-19 classification. The complete list is displayed in Table 3. Next, we discuss the necessary settings, pre-trained backbones, and procedural stages along with some compulsory abstract level discussion on all the models (including abstract level pictorial representation of only some). The self-explanatory block diagram of the proposed work is shown below (Fig. 6).

VGG-16 has been applied in transfer learning fashion. First a model of VGG-16 from scratch is being downloaded and put upon the base model after that own model is created with layers of average pooling 2D with pool size 4, 4, Flatten layer, Dense layer of 64 neurons with ReLU activation, Dropout layer of 0.5 and dense layer of 2 neurons with Softmax activation. All these layers have been fine-tuned using early stopping and reducer lover plateau. Also due to the less quantity of images these images have been passed through data augmentation with batch size 16 and which flips the images, rotation angle set to 15 degrees and fill mode is nearest. The optimizers which have been tested are SGD, Nadam, AdaGrad. Here Nadam comes out to be best which fits the data for each dataset. In early stopping, the patience level is 5 and the learning rate for the optimizer is 0.001. Experiment has been performed up to 50 epochs.

```
Total params: 14,747,650
Trainable params: 32,962
Non-trainable params: 14,714,688.
```

ResNet-101 has been applied in transfer learning fashion here, a model is put on the top of the base model which includes the following layer;Flatten, Dense(64),Batch Normalization, Activation tanh, Dropout 0.25,Dense 64,Batch Normalization, Activation tanh, Dropout 0.25,Dense 32,Batch Normalization, Activation tanh, Dropout 0.5,Dense 2,Activation Softmax. All these layers have been fine tuned and data augmented with batch size 16 and rotation angle 15 degree to increase the quantity of images. In fine tuning we have used early stopping with patience level of 5. Here also we have used optimizer Nadam with learning rate 0.01 to get the accuracy. Experiment has been performed upto 50 epochs.

```
Total params: 49,088,898
Trainable params: 6,430,338
Non-trainable params: 42,658,560.
```

Xception net is used for classification of images in a more précised way with some different layers as its base model and head model keeps the same as we have stated above like from the scratch. Its base model includes the layers of Global Average pooling 2D, Dense 200, Activation ReLU, Dropout 0.4, Dense 170,Activation ReLU, Dense 2 and Activation Softmax. This net is also data augmented with the same rotation angle of 15 degrees and fill mode to the nearest having early stopping with patience value of 6 setting up the best weights quality also using reduce rl over plateau for the softening of the graphs and to increase the accuracy of each dataset on which we have worked on. Optimizer here which we have used is Nadam which remains top in the list. Experiment has been performed upto 50 epochs.

```
Total params: 21,305,792
Trainable params: 21,251,264
Non-trainable params: 54,528.
```

InceptionV3 net is a highly rated net and been applied in the phase of transfer learning having base model of Flatten, Dense 1024,Dropout 0.25,Dense 2 and Activation Softmax. This model also used data augmentation and fine tuning with early stopping having patience level of 5 and batch size of 16; the optimizers tested were nada, Nadam, SGD and AdaGrad. Experiment has been performed upto 50 epochs.

```
Total params: 74,234,658
Trainable params: 52,431,874
Non-trainable params: 21,802,784.
```

InceptionResNetV2 net was proudly said to be used in transfer learning phase with data augmentation and fine tuning with early stopping and reducer lover plateau batch size 16 optimizer Nadam learning rate of 0.001 patience level 5. It's base model includes the following layers, Flatten, Dense(64),Batch Normalization, Activation tanh, Dropout 0.25,Dense 64,Batch Normalization, Activation tanh,Dropout 0.25,Dense 64,Batch Normalization, Activation tanh, Dropout 0.25,Dense 64,Batch Normalization, Activation tanh, Dropout 0.25,Dense 32,Batch Normalization, Activation tanh, Dropout 0.5,Dense 2,Activation Softmax. Experiment has been performed upto 50 epochs.

```
Total params: 56,802,530
Trainable params: 2,465,410
Non-trainable params: 54,337,120
```

In case of EfficientNetB7 for the improvement of accuracy the base model for transfer learning which we have used Batch Normalization, Activation tanh, Dropout 0.25,Dense 32,Batch Normalization, Activation tanh, Dropout 0.5,Dense

2,Activation Softmax. Experiment has been performed upto 50 epochs.

```
Total params: 72,134,041
Trainable params: 8,035,970
Non-trainable params: 64,098,071.
```

VGG-19 is the model which we have worked on with data augmentation, fine tuning and transfer learning including all the concepts of early stopping patience level 5 batch size 16 optimizer Nadam using binary cross entropy having the base model of Average pooling 2D with pool size 4,4 Flatten layer, Dense layer of 64 neurons with ReLU activation, Dropout layer of 0.5 and Dense layer of 2 neurons with Softmax activation. Experiment has been performed upto 50 epochs.

```
Total params: 20,057,346
Trainable params: 32,962
Non-trainable params: 20,024,384.
```

ResNet50V2 has been used with the help of transfer learning having base model content as follows: Flatten, Dense(64),Batch Normalization, Activation tanh, Dropout 0.25,Dense 64,Batch Normalization, Activation tanh, Dropout 0.25,Dense 64,Batch Normalization, Activation tanh, Dropout 0.25, Dense 64,Batch Normalization, Activation tanh, Dropout 0.25,Dense 32,Batch Normalization, Activation tanh, Dropout 0.5,Dense 2,Activation Softmax, The Nadam optimizer is applied with learning rate of 0.001. It also has early stopping with patience level of 5 and having the best weights set this model is the best as compared to any other model having the less data leakage value as compared to other models it also used data augmentation which has a value of 15 degree angle of rotation batch size of 16 and 50 epochs. The epochs in all the cases remains same that is 50.

```
Total params: 29,995,522
Trainable params: 6,430,338
Non-trainable params: 23,565,184.
```

MobileNet is another deep learning net which has been applied in the transfer learning fashion with head model as a mobile net and base model as stated below: Global average pooling 2D, Dense 1024,Activation ReLU,Dense 1024,Activation ReLU,Dense 512,Activation ReLU,Dense 2 and Activation Softmax. The optimizer here which we have used is Nadam with learning rate of 0.001with early stopping patience level 5 batch size 16 50 epochs. Data augmentation is applied with rotation angle of 15 degree. We have also used the best fit here in this case.

```
Total params: 23,301,729
Trainable params: 5,627,522
Non-trainable params: 17,674,207.
```

DarkNet-53 has been trained also in the transfer learning phase with all the optimizers tested namely Adam, AdaGrad and sgd.

```
Total params: 120,466
Trainable params: 119,154
Non-trainable params: 1,312.
```

DCGAN model has been trained on two models which are imposed on discriminator and generator. In this model discriminator has been used for discriminating the images from the right and wrong images the architecture of the discriminator and generator has been added.

NASNetLarge has been implemented in the transfer learning phase and the vital details are shown with the help of screen shot. Batch size is16, Nadam optimizer used to compile the whole model with learning rate of 0.0001.

```
Layer (type)                   Output Shape         Param #    Connected to
================================================================================
input_1 (InputLayer)           [(None, 224, 224, 3) 0

NASNet (Functional)            (None, 7, 7, 1056)   4269716    input_1[0][0]

global_max_pooling2d (GlobalMax (None, 1056)        0          NASNet[0][0]

global_average_pooling2d (Globa (None, 1056)        0          NASNet[0][0]

flatten (Flatten)              (None, 51744)        0          NASNet[0][0]

concatenate_4 (Concatenate)    (None, 53856)        0          global_max_pooling2d[0][0]
                                                               global_average_pooling2d[0][0]
                                                               flatten[0][0]

dropout (Dropout)              (None, 53856)        0          concatenate_4[0][0]

3_ (Dense)                     (None, 2)            107714     dropout[0][0]
================================================================================
Total params: 4,377,430
Trainable params: 4,340,692
Non-trainable params: 36,738
```
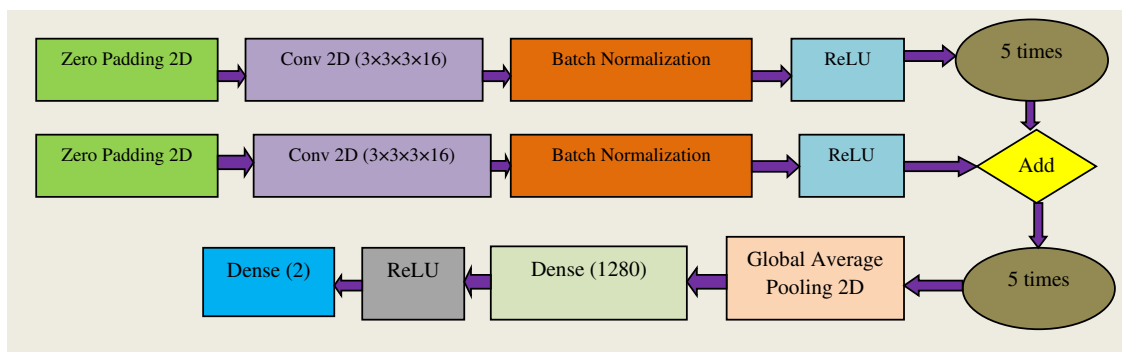
Fig. 7 Abstract form of Proposed Model (Built from Scratch)

Most of the models we use are either pre-installed libraries or on Google; we apply transfer learning to them by creating our small neural network. But making a self-made model (Fig. 7) from scratch is an extraordinary task, so here is the novelty coming out to be. To make our model of our own, we first have made four layers of zero paddings, Convo2D, Batch Normalization with activation ReLU and then added them with one layer of such type; we have put this configuration in such a manner that it fits our each and every dataset on which we have tested our network.

LSTM's (Long short term memory) [60] are rarely used for image classification but it has been discovered that after incorporating certain convolution neural network layers we can use CNN for the classification of images. It can be observed from the diagram of the proposed model, certain layers have been added to the LSTM network to aid in classification of images into COVID-19 and Non-COVID. Such modifications made are highly practical for detection as LSTM runs faster than any CNN network. LSTM has been mostly used to compile the code in regression task but after some modification in transfer learning phase it can be used for image classification as it helps to classify the images; most of the aspects of it remains the same as the other models but the basic difference in this model is that there is large difference between the base model which we have used here.

To apply for classification we have used the following steps with a CNN model: Input, CNN model, LSTM, Dense 2 and Activation Softmax. This is how LSTMCNN (Fig. 8) is used as a tool to classify images; this model comes out with a high expectation of images with good results. We are the frontrunners in using this LSTMCNN model for COVID-19 detection from X-ray and CT images to the best of our knowledge (Fig. 9).

```
Total params: 213,301
Trainable params: 213,301
Non-trainable params: 0
```

It can be observed from Fig. 10, we have proposed ensemble deep transfer learning model (DenseNet-53 + Mobile-NetV2 + ResNet-101 + NASNetLarge + Xception-Net + GoogleNet) where these models have displayed range of performances during classification of similar type of problems. We have combined all the models; after that, we have applied a dense layer of 1024 units with activation function ReLU. Next, we have applied a dense layer with value 2 for this layer, activation function Softmax has been selected; finally we compiled the model. Here we have trained one best model, create its weights and used those weights for another model. From another model, we
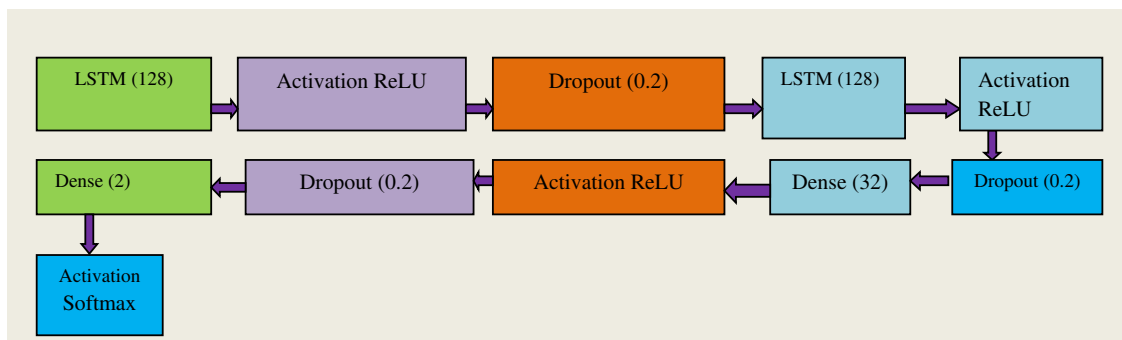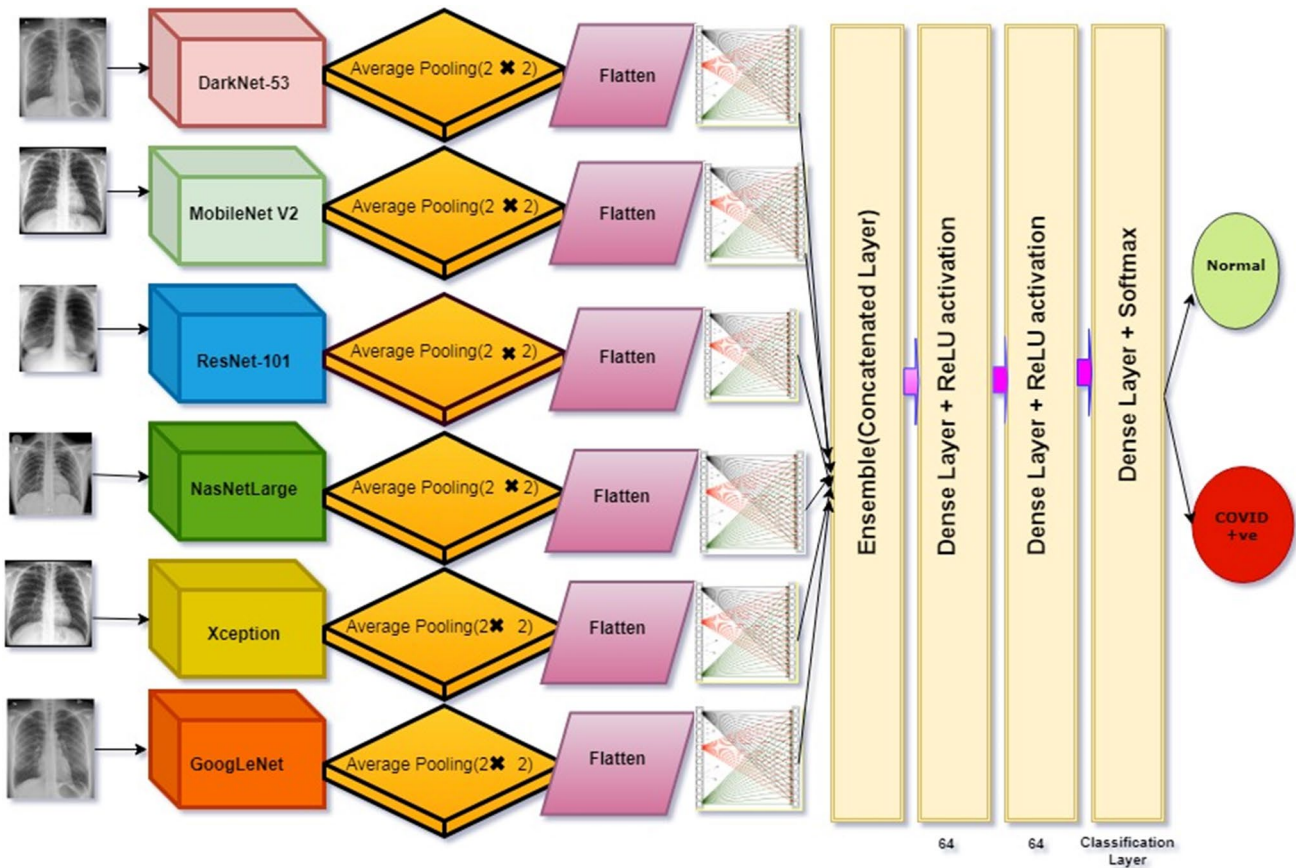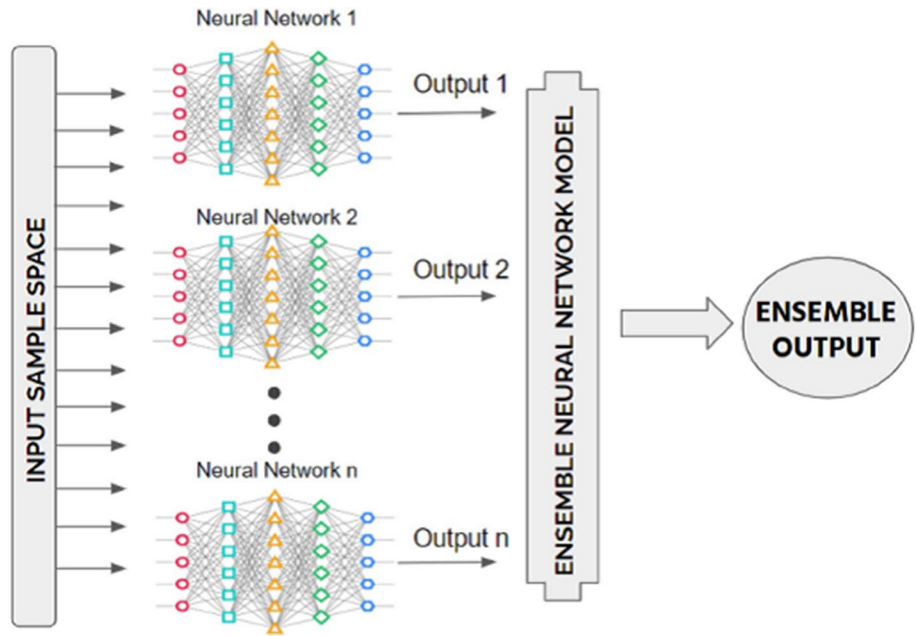

Fig. 8 Abstract form of LSTMCNN Model

Fig. 9 Block diagram of an ensemble learning model by considering *n* number of artificial neural networks[5]

Fig. 10 Detailed representation of proposed ensemble deep transfer learning model

have created weights, and we have used those weights in another model. By doing so, we finally conclude that in between performance may decrease, but after passing those weights of the best model, it helps for sharpening of slopes and ultimately increases in accuracy. For each phase, we have trained the model on batch size 16 with Nadam as an

Table 4 Details of the requisite experimental setup

Experiment Number	Dataset Used to Train the model(s)	Dataset used for Testing	Cross-validation rate (number of folds)	Example Type
1	(1)	(1)	5	Known Examples
2	(2)	(2)	5	
3	(3)	(3)	5	
4	(4)	(4)	5	
5	(1,2)	(1,2)	5	
6	(1,3)	(1,3)	5	
7	(1,4)	(1,4)	5	
8	(2,3)	(2,3)	5	
9	(2,4)	(2,4)	5	
10	(3,4)	(3,4)	5	
11	(1,2,3)	(1,2,3)	5	
12	(1,2,4)	(1,2,4)	5	
13	(2,3,4)	(2,3,4)	5	
14	(1,3,4)	(1,3,4)	5	
15	(1,2,3,4)	(1,2,3,4)	5	
16	(1,2)	(3)	5	Unknown Example
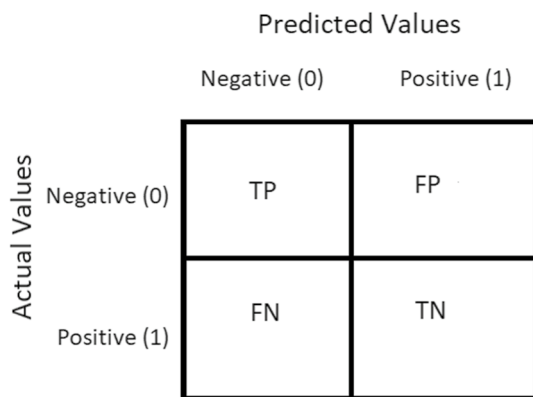
Fig. 11 Representation of the Confusion Matrix used in proposed work

optimizer. It can be analyzed that maybe the accuracy for a single model is less but cumulatively each model's accuracy comes out to be best. Broadly default parameter settings were used for all pre-trained networks used for creating this multiCNN ensemble deep transfer learning model.

3 Experimental Results and Models Evaluation

3.1 Experimental Setup

We have planned 16 experiments to demonstrate each model's robustness, where various combinations of selected datasets are used. Fifteen studies are known examples, while the

other is an unknown one. Established example means that the data for training and testing are the same, while unknown means that training is carried out on a dataset(s) and testing is carried out on different data sets. In Table 4, the specifics of the experimental setup are illustrated. During all 16 studies, we have carried out fivefold cross-validation. The tests were conducted using a 6GBGPU corei7, GTX1060 and coding in Python 3.8 were implemented. Figures 11, 12 display the complete layout of the proposed study.

3.2 Performance Indicators, Evaluation Metrics and Generated Results

In classifying the images into COVID-19 and Non-COVID, this subsection presents the output of the models shortlisted. Confusion Matrix is a performance calculation used to solve classification problems in Deep Learning where output can be two or more classes (Refer Table 5). It is a table composed of four distinct variations of predicted and actual values. The confusion matrices are in the order of 2 X 2. It is a chart that is typically used to portray how the particular classification network is performing on a given test dataset for which the actual values are already known. The matrix of uncertainty consists of four values: True Positive (TP), False Positive (FP), True Negative (TN) and False Negative (FN). In this analysis, various normative performance assessment metrics such as accuracy, precision, recall, F-score, and AUC have been computed. These are based on four words, i.e., TP, FP, TN and FN. TP means prediction is positive and the person is COVID-19 positive, we want that; while FP is the occurrence in which we anticipated true and the real yield was likewise false, a false alarm that is bad. Similarly,
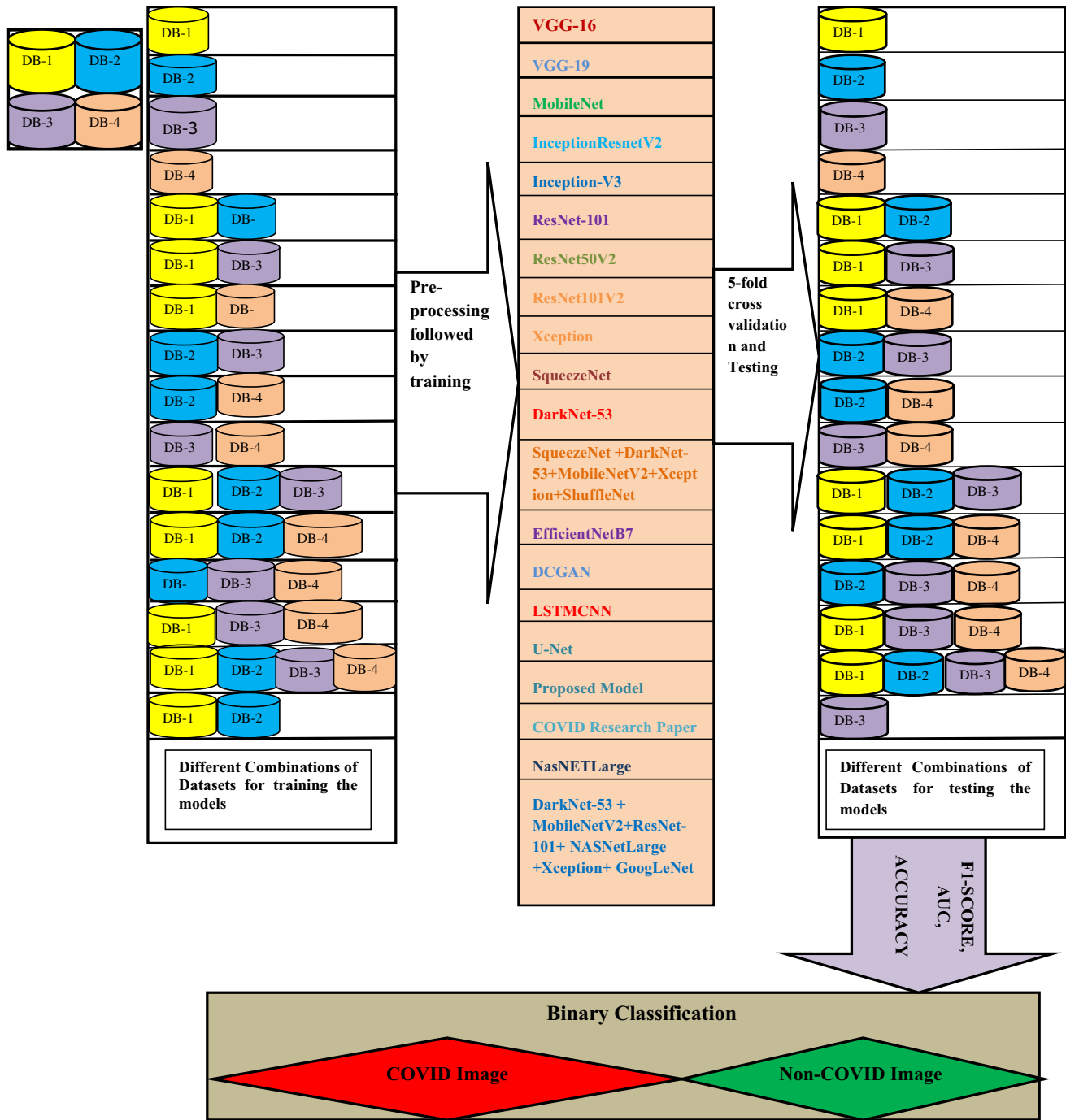
Fig. 12 Layout of the complete process executed in this study for early detection of COVID-19

Table 5 Confusion matrix

	Predicted cases	
	COVID-19	Healthy (Non-COVID)
Actual cases		
Positive	True Positive (TP)	False Negative (FN)
Negative	True Negative (TN)	False Positive (FP)

TN applies to people who do not have an infection, and the result is negative, and FN is a disease patient, but the test is negative which means prediction is negative, and a person is COVID-19 positive, the worst.

Figure 11 shows the confusion matrix for proposed work where Negative (0) represents normal and Positive (1) represents COVID-19.

Sensitivity (a true positive rate) and Recall (the number of true positives separated by the sum of a number of true positives and false negatives) are the preeminent measures of diagnostic precision. The evaluation shows us the frequency at which CNN accurately predicts COVID-19 infection. The sum of sensitivity and false negative incidence is the number of true negatives divided by the total number of checks. The accuracy is the number of true negatives divided by the number of true negatives and false positives. The sensitivity measure calculates just how much of the positive cases were correctly predicted, while the specificity measure determines just how much of the negative cases were correctly expected. The research shows that CNN's medical reporting is reliable for individuals who don't have COVID-19 infection. The recall is the number of accurate outcomes. Precision is how close to the target the sample comes. F1-score tests the accuracy of a source. The F1-score is a metric used to gauge model accuracy. It represents both accuracy and the test which is a recall for the computation of the score. The method incorporates the accuracy and recall of the model to determine a single value, the harmonic mean. Precision is the amount of significance the measurement has with the outcome. Precision can be defined as the number of true positives divided by the number of true positive and false positive. Accuracy is the most critical metric used to measure the efficiency (trueness) of the classifier model. The algorithm calculates the ratio of correctly defined photos vs the overall number of photographs in the dataset. Sensitivity and specificity scores help assess the accuracy of performance. A receiver operating characteristic curve (ROC curve) is a graph of a classification model that shows how well a classification performs to decrease classification thresholds. A ROC curve maps True and False Positive against True Positive at differing degrees of classification. The Receiver Operational Characteristic (ROC) study shows a reliable calculation of COVID-19 classification. The region under the ROC curve indicates the sensitivity of the measurement; it is a curve drawn as the relationship between 1-specificity and sensitivity, showing the characteristic representation of the classification. The field under the curve is a partially independent vector. The plot of True Positive Rate versus False Positive Rate is the receiver operating characteristics (ROC) curve. By calculating the degree of separability between different groups, the ROC reflects the model's investigative competence. It explores how our CNN can discriminate between normal and COVID-19.

The area under the receiver operating characteristic curve (AUC) is used to assess the model's performance; the area is the entire two-dimensional area under the receiver operating characteristic curve from (0,0) to (1,1). The greater superior region under the curve (AUC) is the model in which various groups are divided. Ideally, the AUC should be 1.0, its value is 0, and 0.5 is when the model is equal to random guessing

Table 6 Comparative analysis of the classification results of the 19 models during experiment 1

Experiment -1	Metric -1	Metric-2	Metric-3
Image nets	F1 Score	AUC	Accuracy
VGG-16	0.79330	0.79375	0.82350
VGG-19	0.83333	0.83482	0.77240
MobileNet	0.90000	0.90000	0.91000
InceptionResNetV2	0.83000	0.84000	0.90000
InceptionV3	0.53000	0.50000	0.52100
ResNet-101	0.79000	0.79330	0.86000
ResNet-101V2	0.9012	0.9147	0.9167
ResNet50V2	0.82410	0.82660	0.83000
Xception	0.79333	0.79370	0.73790
SqueezeNet	0.52000	0.53000	0.53000
DarkNet-53	0.53450	0.54230	0.53240
SqueezeNet + DarkNet-53 + MobileNetV2 + Xception + ShuffleNet	0.89650	0.90120	0.87450
EfficentNetB7	0.51230	0.52330	0.53000
DCGAN	0.82145	0.82569	0.90000
LSTMCNN	0.84123	0.82365	0.87456
U-Net(Glaucoma)	0.83645	0.82654	0.88975
Proposed Model	0.98746	0.98654	0.99655
COVID research paper	0.87645	0.88875	0.86457
NASNet Large	0.87457	0.88570	0.98746
DarkNet-53 + MobileNetV2 + Resnet-101 + NASNet Large + Xception + GoogLeNet	0.84563	0.87163	0.86347

Table 7 Comparative analysis of the classification results of the best performing 9 models during experiments 2 and 3

Experiment-2	Metric-1	Metric-2	Metric-3	Experiment-3	Metric-1	Metric-2	Metric-3
Image nets	F1	AUC	Accuracy	Image nets	F1	AUC	Accuracy
VGG-19	0.98756	0.96587	0.99634	InceptionResNetV2	0.96359	0.97584	0.96359
InceptionResNetV2	0.96548	0.96387	0.96471	ResNet-101	0.99785	0.97458	0.98745
Xception	0.96348	0.94127	0.96547	ResNet50V2	0.96547	0.95412	0.98745
DCGAN	0.93647	0.94226	0.96785	DCGAN	0.97465	0.96548	0.97845
LSTMCNN	0.96355	0.98564	0.97459	LSTMCNN	0.96359	0.97855	0.96359
Proposed Model	0.92940	0.96245	0.99245	U-Net(Glaucoma)	0.93235	0.96235	0.96984
COVID research paper	0.97459	0.96326	0.96458	Proposed Model	0.95487	0.95055	0.96358
NASNet Large	0.97847	0.98636	0.97414	COVID research paper	0.97423	0.98746	0.98632
ResNet-101V2	0.8865	0.8659	0.8737	ResNet-101V2	0.8963	0.9147	0.9213
DarkNet-53 + MobileNetV2 + Resnet-101 + NASNet Large + Xception + GoogLeNet + Resnet-101 + NASNet Large + Xception + GoogLeNet	0.99483	0.93584	0.98742	DarkNet-53 + MobileNetV2	0.97478	0.99789	0.97779

Table 8 Comparative analysis of the classification results of the best performing 9 models during experiments 11 and 13

Experiment-11	Metric-1	Metric-2	Metric-3	Experiment-13	Metric-1	Metric-2	Metric-3
Image nets	F1	AUC	Accuracy	Image nets	F1	AUC	Accuracy
ResNet50V2	0.80426	0.84035	0.89550	MobileNet	0.98763	0.97856	0.98565
Xception	0.84660	0.86301	0.89206	InceptionResNetV2	0.96547	0.95635	0.98615
DCGAN	0.93642	0.93348	0.93476	ResNet50V2	0.98236	0.96584	0.99655
LSTMCNN	0.97457	0.97895	0.97456	LSTMCNN	0.97845	0.98746	0.97855
U-Net(Glaucoma)	0.96805	0.98069	0.98660	U-Net(Glaucoma)	0.89233	0.98698	0.98524
Proposed Model	0.98630	0.98604	0.98936	Proposed Model	0.89975	0.98555	0.92126
Covid research paper	0.97456	0.95479	0.96470	Covid research paper	0.96325	0.98464	0.97456
ResNet-101V2	0.8924	0.9124	0.9010	ResNet-101V2	0.8947	0.9214	0.9007
NASNet Large	0.96548	0.94756	0.94752	NASNet Large	0.94365	0.96412	0.97413
DarkNet-53 + MobileNetV2 + Resnet-101 + NASNet Large + Xception + GoogLeNet	0.95686	0.97845	0.98875	DarkNet-53 + MobileNetV2 + Resnet-101 + NASNet Large + Xception + GoogLeNet	0.97878	0.98747	0.99785

Table 9 Comparative analysis of the classification results of the best performing 9 models during experiments 14 and 15

Experiment-14	Metric-1	Metric-2	Metric-3	Experiment-15	Metric-1	Metric-2	Metric-3
Image nets	F1	AUC	Accuracy	Image nets	F1	AUC	Accuracy
VGG-19	0.89641	0.84563	0.94567	ResNet-101	0.94563	0.93654	0.96413
MobileNet	0.95675	0.95641	0.96875	ResNet50V2	0.92146	0.96285	0.96421
DCGAN	0.98745	0.93647	0.97456	DCGAN	0.94569	0.93215	0.97459
LSTMCNN	0.97855	0.97465	0.96578	LSTMCNN	0.94742	0.97456	0.96458
U-Net(Glaucoma)	0.98523	0.96785	0.96875	U-Net(Glaucoma)	0.99455	0.99214	0.98669
Proposed Model	0.97854	0.96874	0.98756	Proposed Model	0.93298	0.94783	0.98547
ResNet-101V2	0.8938	0.8997	0.9045	ResNet-101V2	0.8932	0.8897	0.8740
COVID research paper	0.98631	0.96346	0.97852	COVID research paper	0.93648	0.97413	0.98746
NASNet Large	0.96875	0.98215	0.99785	NASNet Large	0.96335	0.94758	0.96354
DarkNet-53 + MobileNetV2 + Resnet-101 + NASNet Large + Xception + GoogLeNet	0.98634	0.98632	0.94578	DarkNet-53 + MobileNetV2 + Resnet-101 + NASNet Large + Xception + GoogLeNet	0.97459	0.98875	0.97879

Table10 Comparative analysis of the classification results of the best performing 7 models during experiment 16

Experiment-16	Metric-1	Metric-2	Metric-3
Image nets	F1	AUC	Accuracy
DenseNet53 + Mobile-Netv2 + ResNet-101 + NASNetLarge + Xception + GoogLeNet	0.9156	0.9063	0.9074
LSTMCNN	0.8765	0.91368	0.8936
ResNet-101V2	0.9111	0.8993	0.9054
U-Net(Glaucoma)	0.8623	0.8836	0.8845
COVID research paper	0.9012	0.90014	0.8856
NasNetLarge	0.8745	0.91141	0.8763
Self-Made Model	0.9123	0.91256	0.8934

Table 11 Average-wise performance comparison of all the models

Model NAME	All experiments average F1	All experiments average AUC	All experiments average accuracy
VGG-16	0.920733	0.888575	0.914426
VGG-19	0.915139	0.903742	0.921005
MobileNet	0.911087	0.919878	0.925677
InceptionResNetV2	0.904686	0.913151	0.920659
InceptionV3	0.896025	0.885667	0.882325
ResNet-101	0.899679	0.900224	0.899411
ResNet50V2	0.880875	0.910234	0.919958
ResNet-101V2	0.8971	0.9036	0.8989
Xception	0.917826	0.902268	0.91461
SqueezeNet	0.521689	0.530245	0.51925
DarkNet-53	0.529491	0.54358	0.532157
SqueezeNet + DarkNet-53 + MobileNetV2 + Xception + ShuffleNet	0.897339	0.902605	0.894416
EfficentNetB7	0.513148	0.532155	0.532913
DCGAN	0.949983	0.940607	0.957902
LSTMCNN	0.966804	0.964544	0.964641
U-Net(Glaucoma)	0.962745	0.959091	0.950293
Proposed Model	0.945833	0.960826	0.946373
COVID research paper	0.961163	0.957247	0.962947
NASNetLarge	0.959736	0.957844	0.954607
DarkNet-53 + MobileNetV2 + Resnet-101 + NASNetLarge + Xception + GoogLeNet	0.968263	0.95837	0.965129

in the worst case. This is an overall indicator of success across every conceivable degree of identity.

$$Sensitivity = \frac{True\ positives}{True\ Positives\ +\ False\ Negatives} \quad (1)$$

$$Specificity = \frac{True\ negative}{True\ Negatives\ +\ False\ positives} \quad (2)$$

$$Recall = \frac{True\ positives}{True\ Positives\ +\ False\ Negatives} \quad (3)$$

$$Precision = \frac{True\ positives}{True\ Positives\ +\ False\ Positives} \quad (4)$$

Accuracy

$$= \frac{True\ positives\ +\ True\ Negatives}{True\ Positives\ +\ False\ Positives\ +\ False\ Negatives\ +\ True\ Negatives} \quad (5)$$

$$F1 - Score = 2 * \frac{(Precision*Recall)}{(Precision\ +\ Recall)} \quad (6)$$

Table 12 Average-wise best performing 5 models on three efficiency measuring parameters

Ranking of Models	On the basis of Accuracy	On the basis of AUC	On the basis of F1-SCORE
First	DarkNet-53 + Mobile-NetV2 + Resnet-101 + NASNet-Large + Xception + GoogLeNet	LSTMCNN	DarkNet-53 + Mobile-NetV2 + Resnet-101 + NAS-Net Large + Xception + Goog-LeNet
Second	LSTMCNN	Proposed Model	LSTMCNN
Third	COVID Research Paper	U-Net(Glaucoma)	U-Net(Glaucoma)
Fourth	U-Net(Glaucoma)	DarkNet-53 + Mobile-NetV2 + Resnet-101 + NASNet Large + Xception + GoogLeNet	COVID Research Paper
Fifth	NASNetLarge	NASNetLarge	NASNetLarge
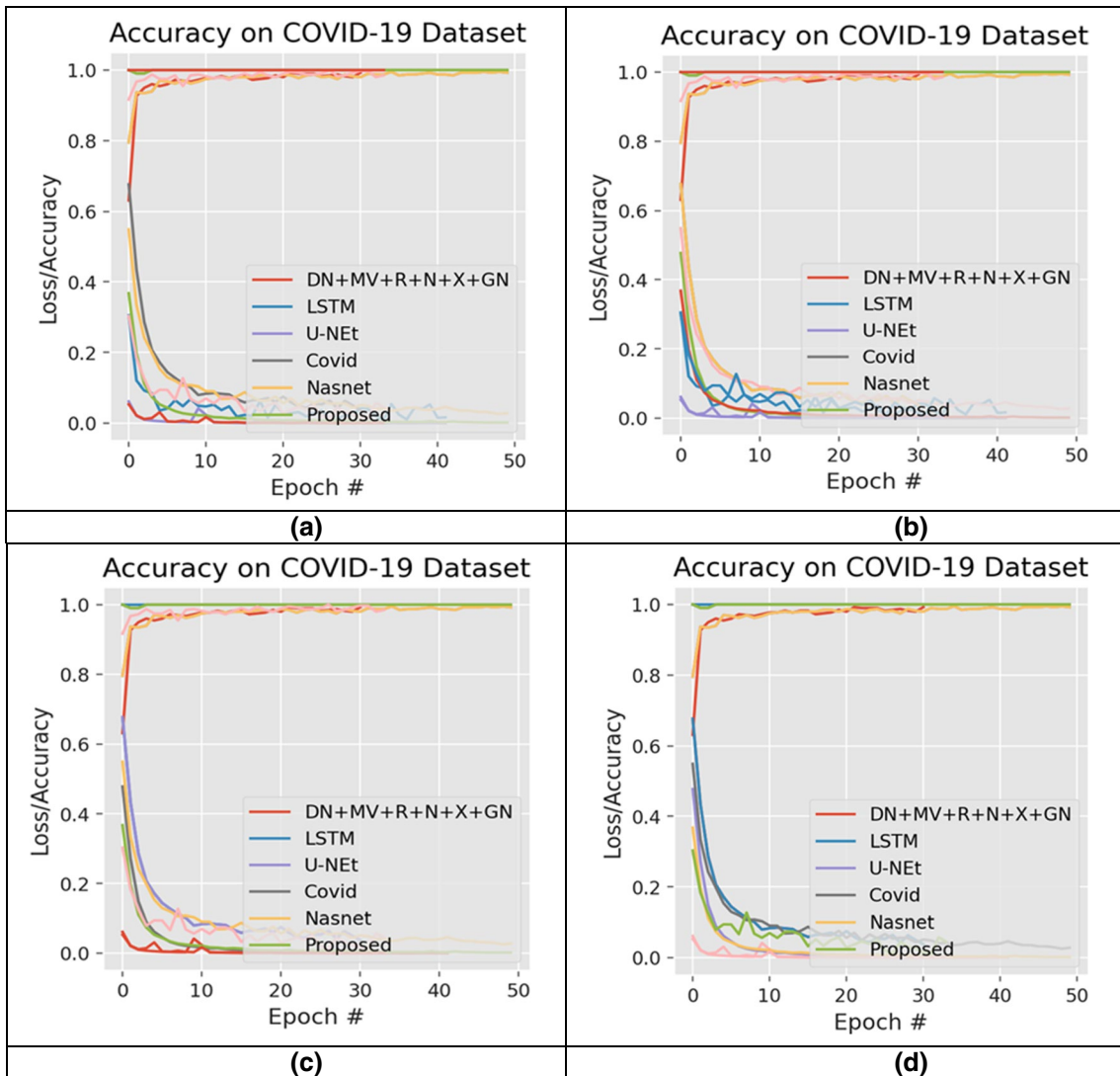
Fig. 13 Generated Training loss, Validation loss, Training accuracy and Validation accuracy curves of 5 models during 4 experiments
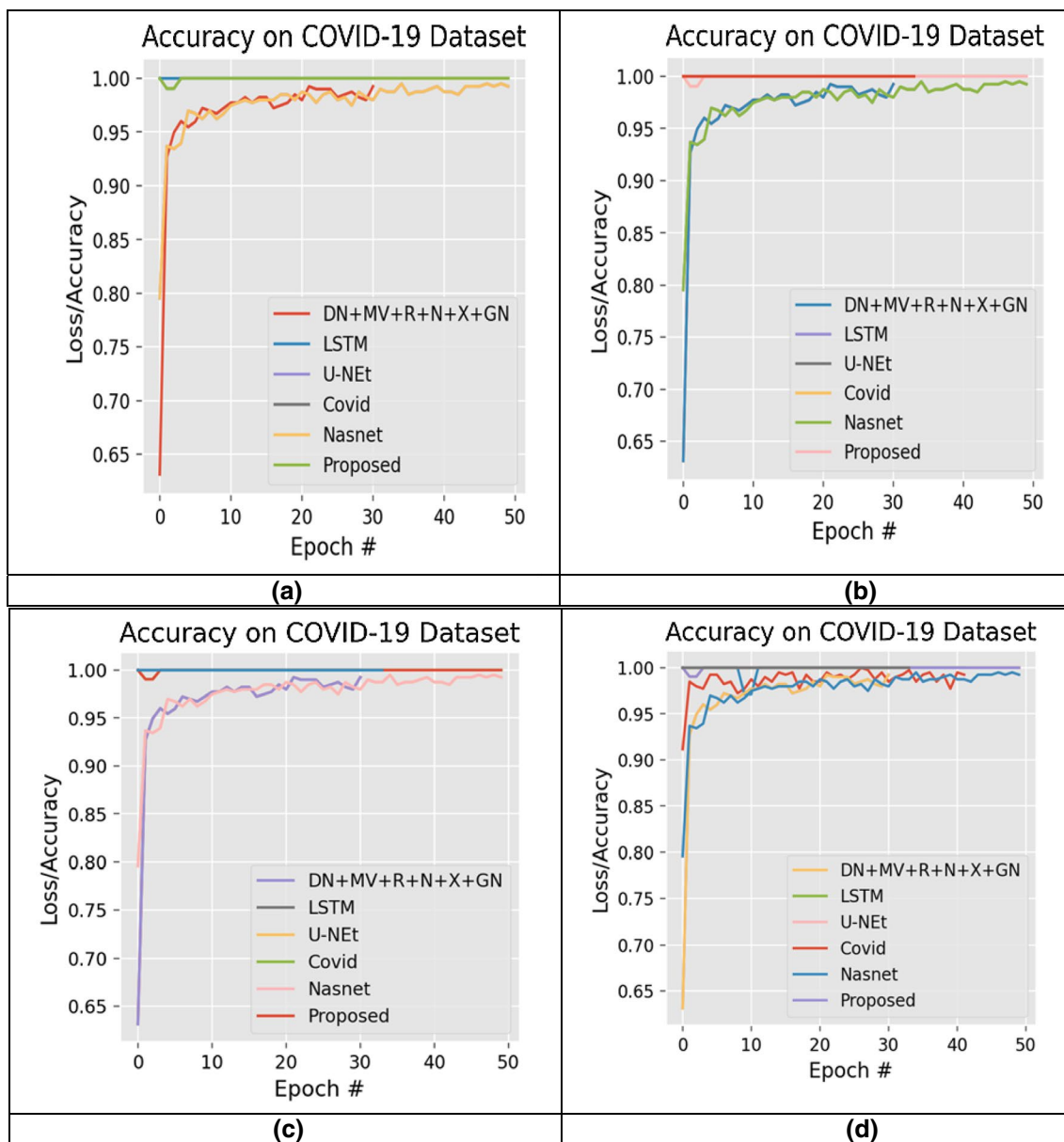
Fig. 14 Generated Testing accuracy curves of 5 models during 4 experiments

The result of confusion matrixes for binary classification problem in detecting COVID-19 positive are shown in Figs. (16, 17, 18). In order to show the produced result during each experiment, Tables 6, 7, 8, 9, 10, 11, 12 are compiled. During the first experiment, Table 6 shows the performance of all 20 selected models; we presented the performance of the best 09 models (top 50 percent) in experiments 2,3,11,13,14,15 and16 (Refer Tables 7, 8, 9, 10) due to space constraint. After that, a cumulative average table is generated where the effects of each model are summed and presented on individual performance metrics during all 16 experiments (Table 11). Combined Table 12 is extracted

from this Average Table, which highlights the final best five models for each metric. Bold numbers highlight the best score produced, corresponding to each metric, in particular experiment, during each table (Refer Tables 6, 7, 8, 9, 10, 11, 12).

Four graphs, where each one displays Validation_loss, Training_loss, Validation_accuracy and Training_accuracy, for four experiments (11,13,14 and 15) performed on 6 models ((DarkNet-53 + MobileNetV2 + Resnet-101 + NAS-Net Large + Xception + GoogLeNet), LSTMCNN, U-Net(Glaucoma),COVID research paper, NASNetLarge and Self-Made Proposed Model) are demonstrated using
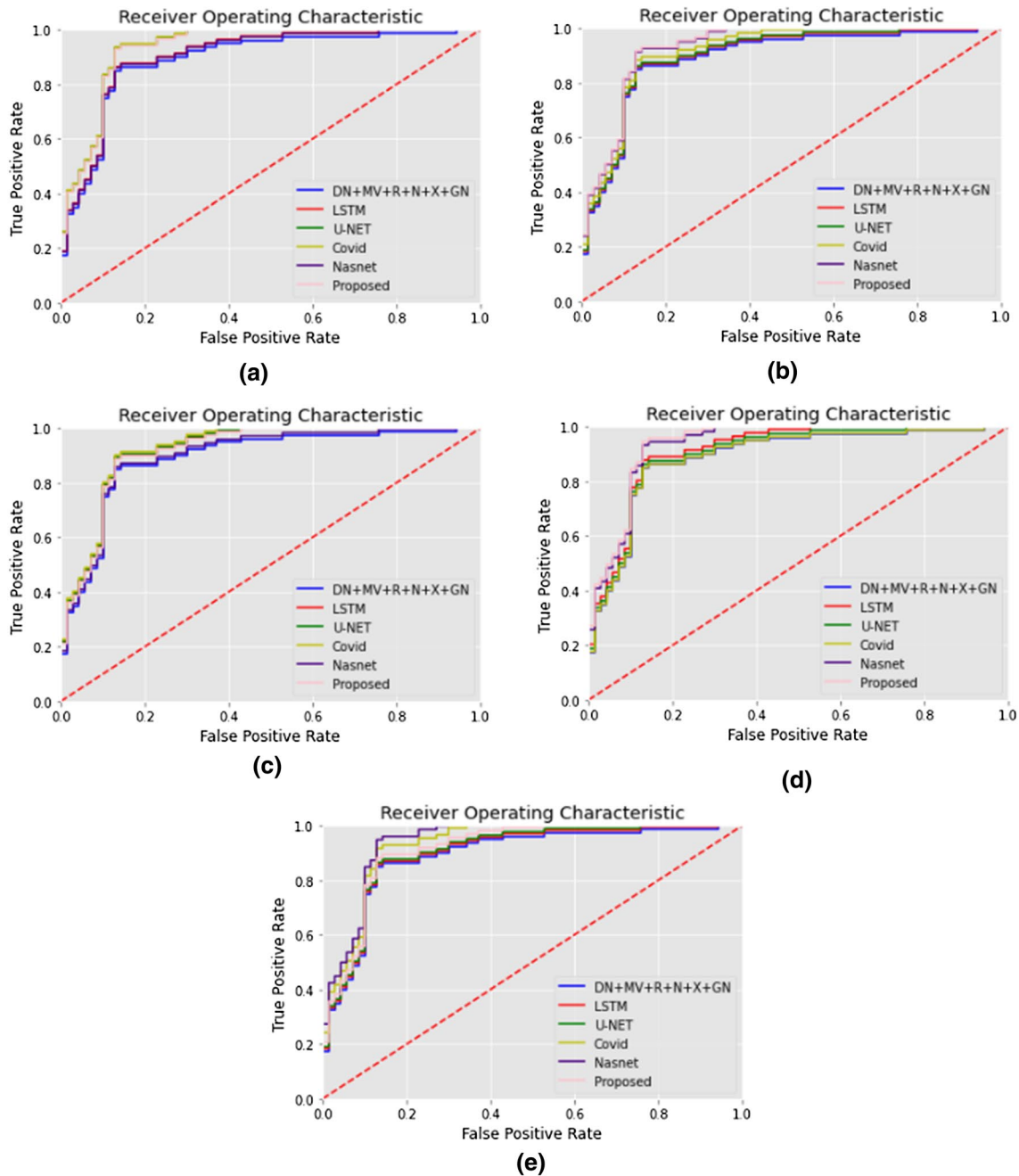
Fig. 15 Generated ROC curves of the 6 models during 5 experiments

Fig. 13. As we increase epochs in the validation phase, the training and testing losses become almost equal which shows that the model is fit adequately and well. Regarding for train and validation accuracy, we found that training and validation accuracy is almost equal, shows that there is no overfitting and the accuracies are showing an upward trend which shows the model is performing well with each epoch.

Four graphs, where each one displays testing_accuracy, for four experiments (11,13,14 and 15) performed on 6

models ((DarkNet-53 + MobileNetV2 + Resnet-101 + NAS-NetLarge + Xception + GoogLeNet),LSTMCNN,U-Net(Glaucoma), COVID research paper, NASNet Large and Self-Made Proposed Model) are demonstrated using Fig. 14.
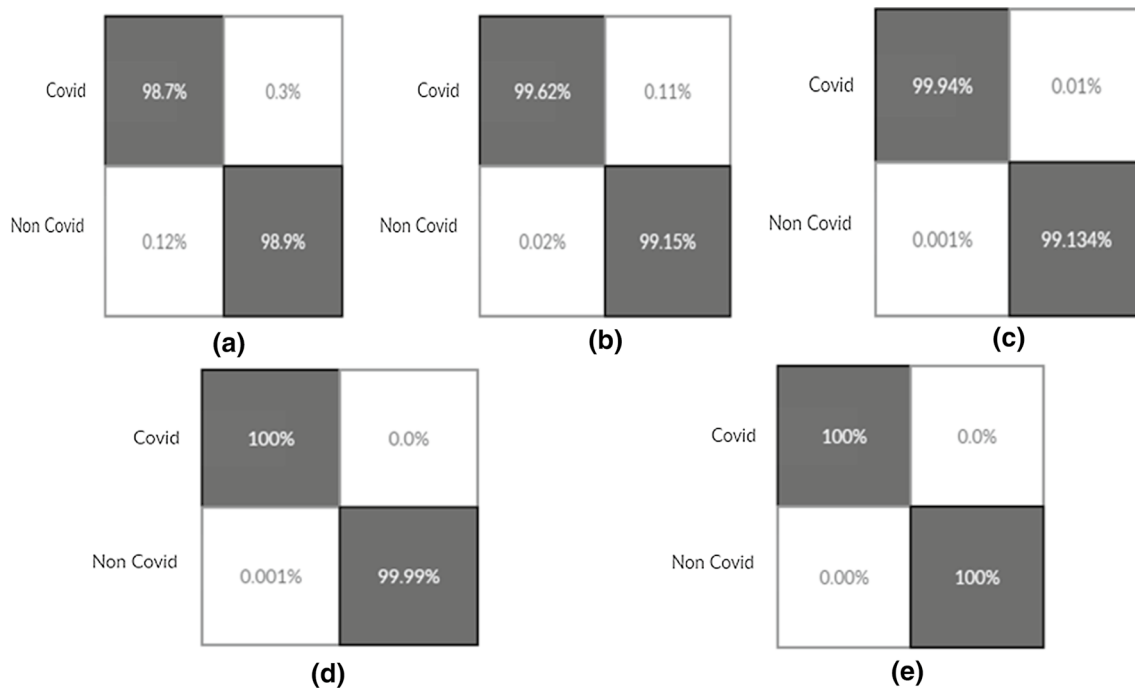
Five ROC graphs(Fig. 15), where each one displays AUC, for five experiments (11,12,13,14 and 15) performed on 6 models (DarkNet-53 + MobileNetV2 + Resnet-101 + NAS-NetLarge + Xception + GoogLeNet,LSTMCNN,U-Net(Glaucoma),COVID research paper, NASNet Large and

Fig. 16 Confusion Matrixes for Proposed Model build from scratch (Experiment 14(Fig-a), 15(Fig-b), 16(Fig-c), 17(Fig-d), 18(Fig-e))
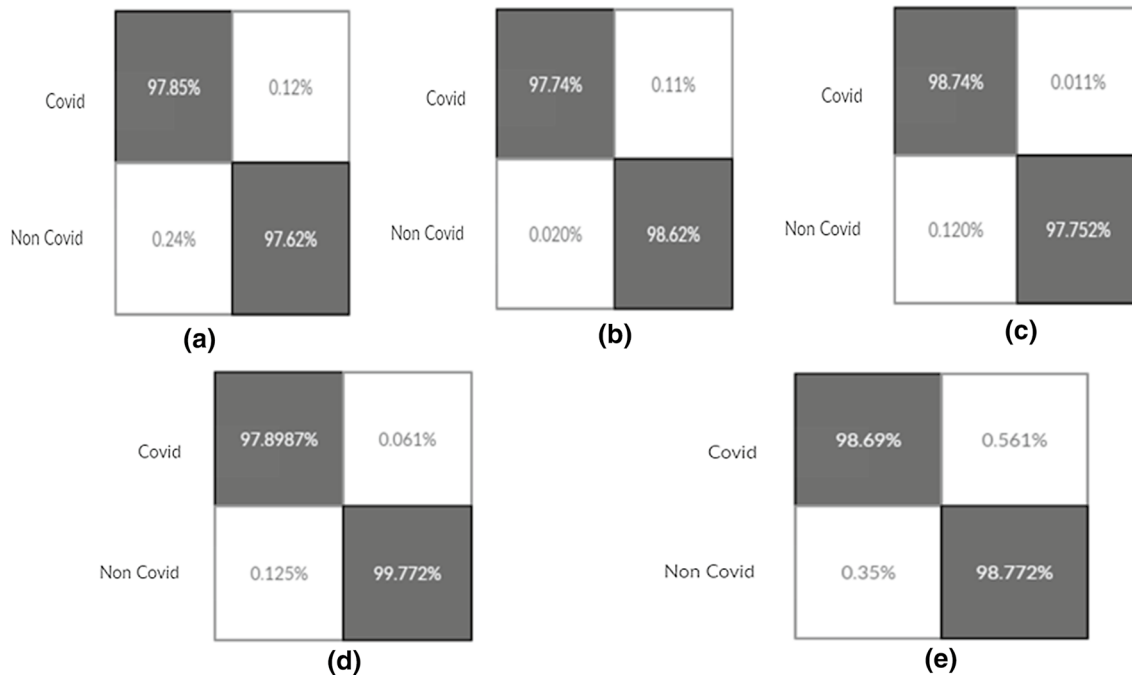

Fig. 17 Confusion Matrixes for LSTMCNN Model (Experiment 14(Fig-a), 15(Fig-b), 16(Fig-c), 17(Fig-d), 18(Fig-e))

Self-Made Proposed Model) were generated to analyze the classification effectiveness.

Comparison in terms of confusion matrixes is illustrated in Figs. 16, 17, 18.

Based on intensive experiments on a considerably large dataset, the acquired outcomes and in-depth data analysis, we find out that our proposed hybrid LSTMCNN model and our proposed ensemble deep transfer learning model (**DarkNet-53 + MobileNetV2 + Resnet-101 + NASNet-**
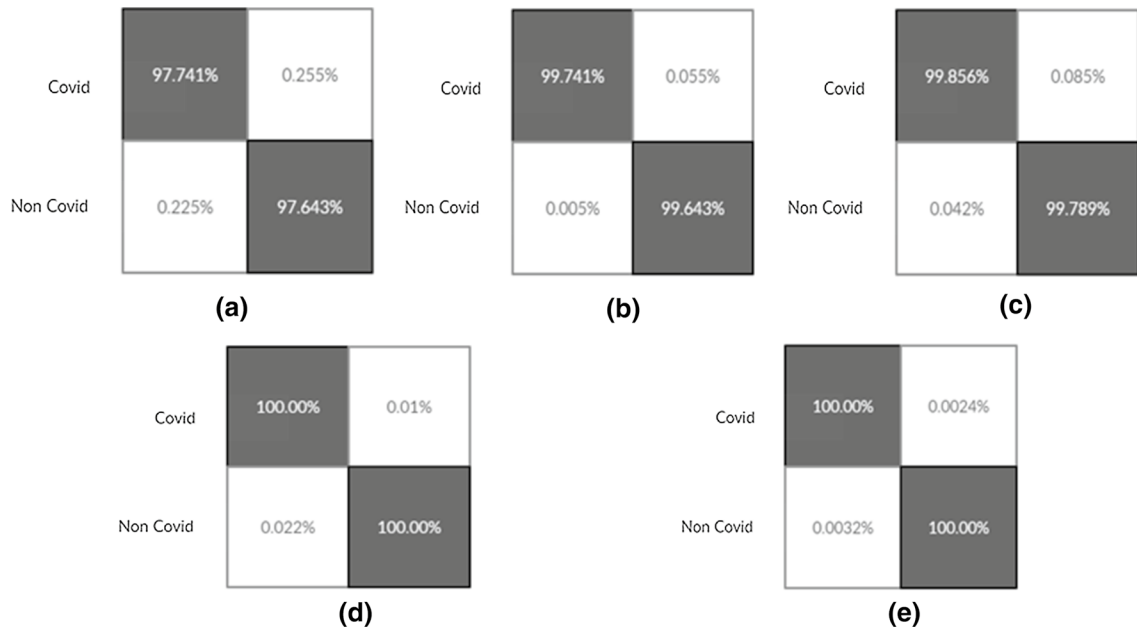
Fig. 18 Confusion Matrixes for Proposed Ensemble Model (Experiment 14(Fig-a), 15(Fig-b), 16(Fig-c), 17(Fig-d), 18(Fig-e))

Fig. 19 (**a**) Healthy Ima (**b**) COVID-19 Image

(**a**) Healthy Image   (**b**) COVID-19 Image

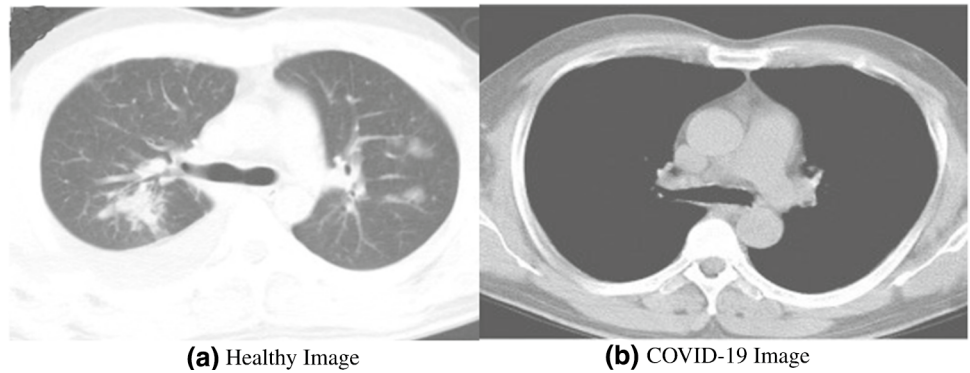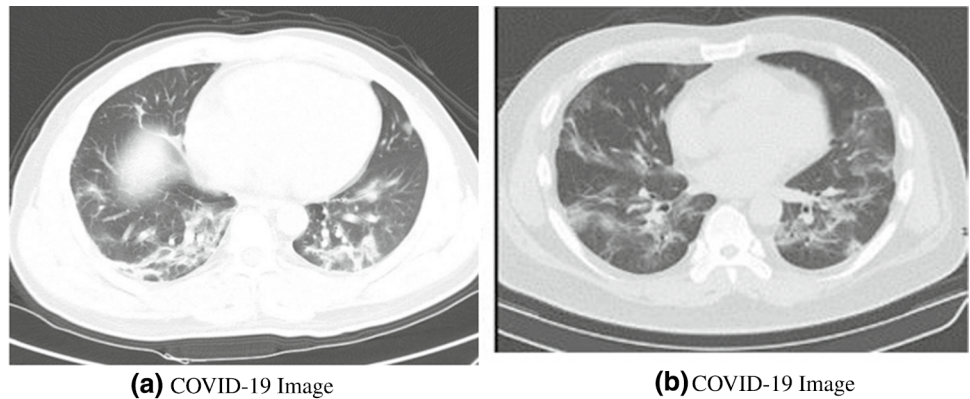**Fig. 20** (**a**) COVID-19 Image (**b**) COVID-19 Image

(**a**) COVID-19 Image   (**b**) COVID-19 Image

Large + Xception + GoogLeNet) performed auspiciously in COVID-19 detection and produced better performance metrics such as accuracy and F1-score than current state-of-the-art methods. Our last proposed model (Fig. 7) yields outstanding average-wise performance for AUC, but average-wise results for F1 and accuracy were less

promising; however, if we observe some of the experiment, for example Experiment 1, the response was overwhelming.

Our reproducible research reveals that patients with no apparent CT abnormalities could be distinguished from COVID-19 positive patients. While COVID-19 characteristic such as bilateral involvement, peripheral dispersion, mixed ground-glass opacity or naked eye consolidation or vascular thickening has occurred, subtle COVID-19 changes have occurred. Various filters recorded differences in the semantic features in the digital deep learning network. While other profound learning algorithms were proposed for the COVID-19 screening compared to other studies, this study had advantages using X-ray images and CT pictures.

3.3 Success and Failure Cases

In this subsection, classification of the given COVID-19 images is focused with the help of four scenarios. Figure 19 shows two images from the Healthy class. Case represented by Fig. 19a. is a successful because the image is normal and our proposed system also predict as normal. Case represented by Fig. 19b. is a failure case because the image represents COVID-19 and our proposed system predict as normal. Figure 20 depicts two scenarios with the help of two images of COVID-19 and suspicious class. Figure 20a presents a success case and 20-b presents failure case.

4 Discussion on the Results and Comparison with Other State-of-the-Art Methods

When all the selected studies are compared with the present study, our datasets are comparatively more immense in terms of the number of COVID-19 positive images used for classification. According to the shortlisted studies, the three more extensive datasets are found to be used in [27] with 453 COVID-19 dataset images, [35] with 580 images and 1000 COVID-19 images in [34]. Still, COVID-19 images count scale is far greater than that of 580 and 453 images in our many experiments. Many previous studies have justified that the number of datasets should be the highest for training and testing is applied in this work. We say that we are the first to undertake such a systematic experiment where 20 CNN models have been used for 16 experiments on different combinations of 4 chosen datasets to provide a comprehensive analysis of artificial intelligence's duty to recognize this infection. Besides, on all 16 experiments, we have summed the output of all the shortlisted image networks, showing the corresponding average values in Table 11. For all performance appraisal metrics, which shows the best models, we have complied Table 12. We conclude that our selected models have demonstrated auspicious efficiency by achieving greater than 99% maximum accuracy and more than 96%

average accuracy. Similarly, executing a maximum AUC greater than 0.995 and an average performance greater than 0.96, we have achieved fantastic performance. Finally, our models have produced the best F1 score of more than 0.99 again, focusing on F1 score, the average performance is also excellent with a value of 0.9682. Considering all the model's average performance over all the experiments, we infer from Tables 11, 12 that LSTM generates the highest average of 0.9645 AUC. In contrast, the highest generated standalone value of 16 experiments is greater than 0.99.

When we tried to compare our performance with other shortlisted studies, we realize that a fair comparison cannot be made because of the difference in the chosen datasets, performance metrics and validation. However, in a comparatively large dataset, we have justified the use of ours. The scale of the dataset that other researchers have chosen is slightly smaller. For instance, in the study [25], the subject model achieved 88.10% accuracy on 284 images (142 COVID images). When the performance comparison is made between ours and [24], we find that in the case of [24], the number of tests, the number of models chosen for research and the size of the dataset is smaller, the accuracy shown by the model is 95.80% which is lower than that of average performance and highest performance.

In [29], practitioners suggested a deep neural network-based DarkNet inspired model with seventeen convolution layers, with LeakyReLU, which was tested on a total of 125 COVID images and 1125. The accuracy generated (87.08%), and the F1-score (0.8737) is either on average or without averaging less than ours. With nine studies published in arXiv and medRxiv, the authors [29] have indeed justified their models' success. However, with only recently published premium studies [2, 19–33], we have made a sincere attempt to explain our work. In-depth analysis shows that if the F1-score is treated as an efficiency parameter, Table 13 shows that our best model will perform higher than that proposed in [29] on an average classification basis and standalone basis.

However, in the case of precision, we do well in binary classification on a standalone basis and satisfactorily on an average basis. Researchers [20] tested 219 COVID-19 images using their proposed model, consisting of a ResNet23 model and a self-designed model. The F1 score and the model's measured accuracies are 83.9% and 86.7%, respectively, both of which are well behind ours. Besides, the size of the data set chosen for testing is not too large, and, also, the performance of the proposed model is not compared to any other pre-existing models. Practitioners, [21], by addressing various advantages over it, justified the usability of X-ray images against CT image. The Inception (Xception)-based model implemented by automated deep transfer learning is proposed in the study using the same dataset [29] for preparation, testing and validation.

Table 13 Comparison of the results with state-of-the-are CNN methods

Reference	Number of COVID-19 images	Method	Accuracy (in %)	F1-Score	AUC	Split method Or Cross-Validation
Goel et al. [34]	1000	Optimized CNN	97.78	0.9525	–	Split
Jain et al. [35]	580	Xception	97.97	0.86	–	–
Das et al.[21]	127	Deep CNN and Xception-based self-crafted model	97.40	0.9696	–	Split
Ardakani et al. [23]	108 patients CT slices	ResNet-101 and Xception	99R.51		0.994	Split
Singh et al. [24]	132	Modified XceptionNet	95.80	0.9588	–	Split
Ozturk et al. [29]	127	DarkNet inspired model	98.08	0.9651	–	Cross-Validation
Panwar et al. [25]	192	VGG-16 inspired nCOVnet	88.10	–	–	Split
Abraham et al.[27]	453	Squeezenet + DarkNet-53 + MobileNetV2 + Xception + ShuffleNet	91.16	0.914	0.963	Cross-Validation
Abbas et al.[36]	105	DeTraC method with ResNet	93.10	–	–	Split
Zebin et al. [37]	202	EfficientNetB0	96.80	–	–	Cross-Validation
Punn et al. [38]	108	NASNet Large	98.00	–	0.99	Cross-Validation
Song et al.[7]	98 patients Chest CT images	Customized GAN Model	–	–	0.972	Split
Shalbaf et al.[57]	349 CT images	Ensemble model of 5 deep transfer learning architecture	85.00	0.852	–	Split
Autee et al.[58]	868 Chest X-ray images	StackNet-DenVIS	95.07	0.955 (Weighted Average)	–	Cross-Validation
Apostolopoulos et al. [59]	224 Chest X-ray images	VGG-19 and MobileNet v2	96.78	–	–	Cross-Validation
Gianchandani et al.[5]	401 Chest X-ray images	Ensemble model of 2 deep transfer learning architecture	96.15	0.961	–	Split
Singh et al.[40]	2373 Chest CT scanned images	Ensemble model of 3 deep transfer learning architecture	98.83	0.9830	0.9828	Split
Jaiswal et al.[41]	1262 chest CT scan dataset	Modified DenseNet201	96.25	0.9629	–	Split
Our Proposed	Maximum upto 1250	LSTMCNN (Standalone Performance)	Upto 98.66	Upto 0.9974	Upto 0.9856	Cross-Validation
Our Proposed	Maximum upto 1250	LSTMCNN (Average-wise Performance)	96.46	0.9668	0.9645	Cross-Validation
Our Proposed	Maximum upto 1250	Hybrid/multiCNN (Standalone Performance)	Upto 99.78	Upto 0.9977	Upto 0.9978	Cross-Validation
Our Proposed	Maximum upto 1250	Hybrid/multiCNN (Average-wise Performance)	96.51	0.9682	0.9583	Cross-Validation

The efficiency of the ten pre-existing versions was compared with proposed one. Compared to all, the proposed model reached the highest accuracy of 97.40%, which is almost equal to ours on average and less than our highest achievement. Similarly, focusing on the proposed model's F1 score, 0.9696 score was produced, which also follows the same course, which in average cases is roughly comparable to ours and more petite than our highest achieved. In a study

[22] focusing on the diagnosis of COVID-19 infection, practitioners suggested a predictive model from the 18 laboratory findings of 600 patients. This study was a unique study on COVID-19 diagnosis and maybe the first of its kind. The experimental findings show that 86.66%, 0.9189 and 0.6250 respectively, were the highest produced accuracy, F1-score and AUC values when identifying patients with infection. Two approach tests, tenfold cross-validation and train-test split was conducted. LSTM performs best in the previous strategy, while one CNNLSTM combination performs best with 92.30% accuracy, 0.9300 F1-score and 0.900 AUC in the latter. Researchers [23] have experimented with a private dataset for clinical screening of COVID-19 infection using 10 pre-existing transfer-learned structured convolutionary networks. A single experiment on a moderate size dataset (510 COVID-19 images) was performed during the study. Researchers propose ResNet-101 and Xception as the best performing model that can play an adjuvant instrument in the radiology departments of medical colleges/hospitals due to this experimental research. These models obtained the highest AUC (0.994) and accuracy (99.51%), however, the number of tests, the size of combined datasets, and the number and form of selected models are more significant.

The authors of [25] proposed a transfer-learned CNN model based on VGG-16, called the nCOVnet model, for rapid diagnosis of infection from patients' X-ray images. The authors used 127 positive patients with COVID-19 as a dataset to validate nCOVnet performance using a single experiment only. Also, we note that the efficiency of the proposed nCOVnet is not compared to any other current/novel model/technique. The analysis does not explicitly present the accomplished AUC and F1-score. In the meantime, we can claim that our number of experiments, the size of combined datasets and the number and form of models shortlisted again exceeds them.

Two deep learning (DL) models, one for discrimination and another for localization, were presented in the subsequent study by Wang et al. [26]. The first model extracts lung characteristics from chest X-rays radiographs and the second locates and assigns known X-ray radiographs to the left, right or bipulmonary lungs. A total of 3500 images containing only 225 images of COVID-19 positive patients were selected. The first model, which corresponds to our target, provided 98.71% accuracy, somewhat lower than our best model(s). For research on 453 COVID-19 X-ray images, Abraham et al. [27] used 10 pre-trained CNN models and 15 combinations of pre-trained CNNs. Only two experiments were conducted. The authors conclude on the results obtained that their hybrid model (comprising 5 pre-trained models) performs best with 0.963 AUC, 0.914 F1-score and 91.15% accuracy. In our case, the average accuracy (of sixteen experiments) and standalone (non-averaged) accuracy are also higher. In standalone (non-averaging) F1 and AUC,

a similar trend is also observed. The second test's dataset size consists of 71 infectious images and 7, non-COVID images.

The authors of [19] presented research to screen for COVID-19 infection from private CT images of 618 patients. Using a 3D deep learning model to be classified into COVID-19, Influenza-A viral pneumonia, and irrelevant to infection groups and their corresponding confidence scores, segmentation is initially carried out to classify infection regions employing a local-attention categorization model. The system provided a total accuracy rate of 86.7%. In their empirical study [30], practitioners suggested a deep neural network based on Xception CNN called CoroNet for automatic 3-class classification (COVID, Pneumonia and normal) with 95% accuracy. The experiment was conducted on 284 chest radiograph images of COVID-19 and approximately 1250 images in total. Binary classification has also been performed by authors where the model generates a 99% and 0.985 accuracy and F1 score, respectively. Binary classification has also been carried out; our highest results are also in the same line as that produced in this report. Besides, the authors have not compared their proposed model to other existing models, and the number of models tested and the size of COVID-19 images used in experiments is also smaller than that of our research. Researchers from the following study [31] present a deep SqueezeNet-based, COVIDiagnosis-Net, Bayesian optimization-based method to diagnose the disease from X-ray images. The experimentation was carried out on 5310 images that contain only 66 images of COVID-19. Overall, the technique yields 0.9826 precision and 0.9825 F1-score. Due to the proposed system's very high delivered accuracy, no other model is compared to the proposed one. The count of COVID-19 images is too limited, and it is expected that the model's output should also be validated on a large dataset.

A convolutionary layer-based CNN model is built in the subsequent study by Nour et al. [2] that also includes the Bayesian optimization algorithm for the detection of COVID-19 infection. The proposed method yields 98.97% accuracy on a limited size of 219 public COVID-19 images, but the total number of images is 2905. Our research has implemented this CNN model (without a Bayesian optimization algorithm) to assess the output of our datasets and compare the results achieved with other topic models. In this study by Brunesa et al. [32], investigators applied model pre-trained VGG-16 imagenet to build their model on the top of it to screen COVID-19 from X-ray images. 3520 images were selected in which the size of COVID-19 positive patient's images count is 250 only. The model achieves overall 97% accuracy. Similarly, like others, the performance of the proposed model was not compared with any pre-existing imagenets. In the following study by Toraman et al.[28], a novel ANN, convolutional CapsNet was

proposed, which achieved accuracy and F1-score of 97.24% (with data augmentation),both for binary class classification. 231 COVID-19 were used to validate the performance of the proposed one; only one model was presented. VGG-19-based model was offered in the experimental research by Panwar et al. [33] where 285 images of COVID-19 patients were shortlisted for performance assessment. Three experiments were performed to finally achieve the best accuracy of 95.61% and 0.97 F1-Score.

Approximately 850 COVID-19 positive patient X-ray images were chosen by Goel et al. [34] from the publicly accessible repositories for the work. The best accuracy 97.78% and the best F1-score of 0.9525 was offered; for both of these performance assessment metrics our best performing model(s) generated best scores is better than these ones. Jain et al.[35] compared three models, Inception V3, Xception and ResNeXt, in the following study just published; authors shortlisted the public repository for about 575 COVID-19 images. Authors of [35] have raised a cause of concern stating that the high accuracy obtained may be due to overfitting influence. In this context, we want to assure that there is a very rare chance of this in present empirical work. Data augmentation was used to correct the imbalance dataset and avoid overfitting during training. The Xception model was the best performer, with the highest accuracy and F1 score of 97.97% and 0.95, respectively. Researchers suggested a deep CNN (DeTraC) that decomposes, transfers and composes for the classification of COVID-19 chest X-ray images in the subsequent empirical analysis by Abbas et al. [36]. Just 105 COVID-19 images were in the database selected for function, 70:30 split ratio training–testing was preserved. For the transfer learning stage in DeTraC, five dissimilar ImageNet pre-trained CNN models were used. The highest accuracy of 97.35 percent was achieved by combining VGG-19 with DeTraC. Practitioners Zebin et al. [37] used VGG-16, ResNet50 and EfficientNetB0 for classification and 202 shortlisted COVID-19 images were split into training–testing ratios of 80:20. In the case of EfficientNetB0, the best accuracy of 96.8% was achieved. The investigators engaged five pre-trained CNN models in this final analysis by Punn et al. [38]; authors shortlisted 153 COVID-19 photos. NASNetLarge was the best model that achieves an accuracy of 98% and an AUC of 0.99 for two classifications. Table 13 is compiled, which indicates that the acquired performance parameters for our two proposed CNN's are superior to those of other state-of-the-art approaches.

Based on the above intensity comparison, we discover that our and other effective deep learning-based models can detect the COVID-19 infection's biomarkers, such as asymmetric peripheral ground-glass opacities without pleural effusions, multifocal patchy consolidation regions, and nodular opacities. Our best performing deep learning model(s) can easily sense bilateral, lower lobe, and lesion localization (areas on chest X-ray images) involvement.

We are also skeptical about the comparison, as they used different datasets, comparison techniques, and performance metrics. However, the number and scale of datasets, the number of experiments conducted, and the number and variety of image networks used made the findings evident and valuable to the medical community. The present study presents a computer-assisted diagnostic method using X-ray and CT images and delivering results like a radiologist.

Our proposed model(s) accuracy is optimistic and superior to other shortlisted competitive models; the others have less classification accuracy; this may be since the other architectures have a larger and more complicated structure and more training parameters than our proposed ones. Our proposed architecture(s) provides better generalization capabilities than the other shortlisted models. With the aid of generated results, a deep learning framework can help detect infection subjects using X-ray/CT images by providing a fast, cheap and accurate solution for the diagnosis of the disease. The study showed that a promising model(s) for diagnosing and classifying infections is suggested as the LSTMCNN and ensemble model (consisting of 6 networks). These automated models can identify complex patterns from medical images with an expert radiologist's equivalent ability. The model does not entail extensive costs and can be used in radiology departments as the best supporting method during image analysis.

5 Discussion on Practical Applicability and Outlook

Early diagnosis of coronavirus disease 2019 (COVID-19) is crucial for disease treatment and control. A rise in patients, insufficient government planning, and an increase in funding (people and services) add to the crisis. The situation is more difficult because the RT-PCR test kits (having lower sensitivity) are not readily available in certain countries. The patients have not been diagnosed as expected (suspected or proven) and have not been placed in an isolated area, resulting in hotspots (potential infection source). The detrimental impact is profound in all industries like engineering, tours and tourism, vehicles, culture, and the education market. A report (https://www.dentons.com/en/insights/alerts/2020/march/11/COVID-19-and-its-impact-on-the-global-economy) predicts the global loss of 1.1 trillion dollars. To date, the exact solution to this deadly disease is not found, and even the just retired first citizen of the USA got also infected from it.

It is a common challenge to differentiate and reliably diagnose COVID-19 from other diseases. We proposed a system in which a fully automated end-to-end chest X-ray

and CT interpretation was evaluated using well-known and designed deep-learning-based approaches. The system extracts what they need on their own, and there is no need for human intervention. When training is done on large datasets, network weights are trained and fine-tuned on small datasets. The data is standardized, and preprocessing steps are carried out first, and then models produce outputs by differentiating the input images from unsafe ones.

As seen in the previous section of this paper, we are the front runners to perform systematic experiments on four datasets and their comparative review to justify the robust and best model(s), out of 20, for COVID-19 outbreak prediction. This will satisfy the immediate need for human society to explain its practical applicability. Anyone can see the results within 5 s from our trained models with high accuracy and discriminative ability; it can also play an assistant's role during X-ray/CT imaging in radiology departments of hospitals / medical colleges. To check the outcomes, all the standard performance evaluation indicators such as Accuracy, AUC and F1-score are validated. Our hybrid and ensemble deep learning model offer a unique benefit in this study. The user can pick the deep learning model's best output to identify each diseased case separately.

Our research demonstrated that deep learning-based techniques helped radiologists to identify the COVID-19 infection accurately. The best results were found using our LSTMCNN model and our proposed ensemble deep transfer learning model. We have used these two robust deep learning-based models to characterize and detect the infection; they can recognize chest X-ray/CT images abnormalities. The outputs of these auspiciously performing model(s) can be used as a second opinion and can help physicians make the best decision and accurate diagnosis by effectively detecting abnormalities. These models capture radiographic features from low-level image data and can recognize complex patterns that radiologists may overlook. These models reduce wait times for diagnosis. Using these models, hospitals can identify and isolate the patients early, preventing the virus from spreading in the community, ultimately saving the life and money. Besides, it reduces the exposure of nurses and medical staff to the outbreak. We can use this proposed system to guide isolation in a real-time setting for patients at risk for COVID-19. Preliminary analysis of the samples can help in determining whether quarantine measures are appropriate. Our best performing, fully automated, inexpensive, rapid and accurate binary classifier models can be employed in health care centres situated at remote locations to analyze X-ray and CT images to resolve the issues like the absence of qualified radiologists and wait for the testing kits to arrive.

6 Conclusion

In this paper, a hybrid model (LSTMCNN) and ensemble model using deep CNN are proposed for COVID-19 detection from chest X-ray images and CT images. Intense experimentation is performed in this work, where 16 experiments are performed using 20 models. The proposed models are tested over three well-known datasets and one private dataset to prove their versatility. Extensive experimental results reveal that the proposed models achieved incredible accuracy for binary classification and outperformed the other shortlisted competitive COVID-19 classification models. Health practitioners and researchers may leverage these models to improvise this infection's prediction in this worldwide ongoing pandemic situation. Although two of our models (LSTMCNN and ensemble deep transfer learning model) produced an encouraging performance in a considerably more extensive dataset, it has few limitations worth mentioning. The method is limited to 2 class classifications, COVID-19 and Non-COVID, of the X-ray and CT images. Testing in multi-class classification (like COVID-19, normal and pneumonia) is not performed in this work. Moreover, the segmentation of the infected region is not performed. In this study, it is believed that patients have a single disease if it exists.

In future, more patient data can be leveraged that can support better feature extraction capability of the proposed model(s). We can also extend this work by adding factors like risk and survival prediction of the potential patients that can be useful in framing a stronger healthcare framework. In the future, we will analyze how the hyperparameters were tuned in with available algorithms and how feature extraction approaches can be applied on X-rays/CT images for this infection identification.

Declarations

Conflict of interest On behalf of all authors, the corresponding author states that there is no conflict of interest.

References

1. https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200810-covid-19-sitrep-203.pdf?sfvrsn=aa050308_4
2. Nour, M.; Cömert, Z.; Polat, K.: A novel medical diagnosis model for COVID-19 infection detection based on deep features and Bayesian optimization. Appl. Soft Comput. **97**, 106580 (2020)
3. Karar, M.E.; Hemdan, E.E.D.; Shouman, M.A.: Cascaded deep learning classifiers for computer-aided diagnosis of COVID-19 and pneumonia diseases in X-ray scans. Complex Intell. Syst. **7**(1), 235–247 (2021)
4. Shankar, K.; Perumal, E.: 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**(3), 1277–1293 (2021)

5. Gianchandani, N.; Jaiswal, A.; Singh, D.; Kumar, V.; Kaur, M.: Rapid COVID-19 diagnosis using ensemble deep transfer learning models from chest radiographic images. J. Ambient. Intell. Human. Comput. (2020). https://doi.org/10.1007/s12652-020-02669-6

6. Xie, X.; Zhong, Z.; Zhao, W.; Zheng, C.; Wang, F.; Liu, J.: Chest CT for typical coronavirus disease 2019 (COVID-19) pneumonia: relationship to negative RT-PCR testing. Radiology **296**(2), E41–E45 (2020)

7. Song, J.; Wang, H.; Liu, Y.; Wu, W.; Dai, G.; Wu, Z.; Deng, K.: End-to-end automatic differentiation of the coronavirus disease 2019 (COVID-19) from viral pneumonia based on chest CT. Eur. J. Nucl. Med. Mol. Imaging **47**(11), 2516–2524 (2020)

8. Nihashi, T.; Ishigaki, T.; Satake, H.; Ito, S.; Kaii, O.; Mori, Y.; Ohashi, M.: Monitoring of fatigue in radiologists during prolonged image interpretation using fNIRS. Jpn. J. Radiol. **37**(6), 437–448 (2019)

9. Taylor-Phillips, S.; Stinton, C.: Fatigue in radiology: a fertile area for future research. Br. J. Radiol. **92**(1099), 20190043 (2019)

10. Ardila, D.; Kiraly, A.P.; Bharadwaj, S.; Choi, B.; Reicher, J.J.; Peng, L.; Naidich, D.P.: End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography. Nat. Med. **25**(6), 954–961 (2019)

11. Kanne, J.P.; Little, B.P.; Chung, J.H.; Elicker, B.M.; Ketai, L.H.: Essentials for radiologists on COVID-19: an update—radiology scientific expert panel (2020)

12. Xie, X.; Zhong, Z.; Zhao, W.; Zheng, C.; Wang, F.; Liu, J.: Chest CT for typical coronavirus disease 2019 (COVID-19) pneumonia: relationship to negative RT-PCR testing. Radiology **296**(2), E41–E45 (2020)

13. Liu, X.; Guo, S.; Yang, B.; Ma, S.; Zhang, H.; Li, J.; Fu, Y.: Automatic organ segmentation for CT scans based on super-pixel and convolutional neural networks. J. Digit. Imaging **31**(5), 748–760 (2018)

14. Gharbi, M.; Chen, J.; Barron, J.T.; Hasinoff, S.W.; Durand, F.: Deep bilateral learning for real-time image enhancement. ACM Trans. Graph. (TOG) **36**(4), 1–12 (2017)

15. Litjens, G.; Kooi, T.; Bejnordi, B.E.; Setio, A.A.A.; Ciompi, F.; Ghafoorian, M.; Sánchez, C.I.: A survey on deep learning in medical image analysis. Med. Image Anal. **42**, 60–88 (2017)

16. Ker, J.; Wang, L.; Rao, J.; Lim, T.: Deep learning applications in medical image analysis. Ieee Access **6**, 9375–9389 (2017)

17. Lodwick, G.S.: Computer-aided diagnosis in radiology: a research plan. Invest. Radiol. **1**(1), 72–80 (1966)

18. Doi, K.: Computer-aided diagnosis in medical imaging: historical review, current status and future potential. Comput. Med. Imaging Graph. **31**(4–5), 198–211 (2007)

19. Xu, X.; Jiang, X.; Ma, C.; Du, P.; Li, X.; Lv, S.; Yu, L.; Ni, Q.; Chen Y.; Su. J.; Lang, G.: A deep learning system to screen novel coronavirus disease 2019 pneumonia. Engineering **6**(10), 1122–1129 (2020)

20. Butt, C.; Gill, J.; Chun, D.; Babu, B.A.: Deep learning system to screen coronavirus disease 2019 pneumonia. Appl. Intell. 1 (2020)

21. Das, N.N.; Kumar, N.; Kaur, M.; Kumar, V.; Singh, D.: Automated deep transfer learning-based approach for detection of COVID-19 infection in chest X-rays. IRBM (2020). https://doi.org/10.1016/j.irbm.2020.07.001

22. Alakus, T.B.; Turkoglu, I.: Comparison of deep learning approaches to predict COVID-19 infection. Chaos, Solitons Fractals **140**, 110120 (2020)

23. Ardakani, A.A.; Kanafi, A.R.; Acharya, U.R.; Khadem, N.; Mohammadi, A.: Application of deep learning technique to manage COVID-19 in routine clinical practice using CT images:

24. results of 10 convolutional neural networks. Comput. Biol. Med. **121**, 103795 (2020)

24. Singh, K.K.; Siddhartha, M.; Singh, A.: Diagnosis of coronavirus disease (COVID-19) from chest X-ray images using modified XceptionNet. Roman. J. Inform. Sci. Technol. **23**, S91-105 (2020)

25. Panwar, H.; Gupta, P.K.; Siddiqui, M.K.; Morales-Menendez, R.; Singh, V.: Application of deep learning for fast detection of COVID-19 in X-Rays using nCOVnet. Chaos, Solitons Fractals **138**, 109944 (2020)

26. Wang, Z.; Xiao, Y.; Li, Y.; Zhang, J.; Lu, F.; Hou, M.; Liu, X.: Automatically discriminating and localizing COVID-19 from community-acquired pneumonia on chest X-rays. Pattern Recognit. **110**, 107613 (2021)

27. Abraham, B.; Nair, M.S.: Computer-aided detection of COVID-19 from X-ray images using multi-CNN and Bayesnet classifier. Biocybern. Biomed. Eng. **40**(4), 1436–1445 (2020)

28. Toraman, S.; Alakus, T.B.; Turkoglu, I.: Convolutional capsnet: a novel artificial neural network approach to detect COVID-19 disease from X-ray images using capsule networks. Chaos, Solitons Fractals **140**, 110122 (2020)

29. Ozturk, T.; Talo, M.; Yildirim, E.A.; Baloglu, U.B.; Yildirim, O.; Acharya, U.R.: Automated detection of COVID-19 cases using deep neural networks with X-ray images. Comput. Biol. Med. **121**, 103792 (2020)

30. Khan, A.I.; Shah, J.L.; Bhat, M.M.: CoroNet: a deep neural network for detection and diagnosis of COVID-19 from chest x-ray images. Comput. Methods Programs Biomed. **196**, 105581 (2020)

31. Ucar, F.; Korkmaz, D.: COVIDiagnosis-Net: Deep Bayes-SqueezeNet based diagnosis of the coronavirus disease 2019 (COVID-19) from X-ray images. Med. Hypotheses. **140**, 109761 (2020)

32. Brunese, L.; Mercaldo, F.; Reginelli, A.; Santone, A.: Explainable deep learning for pulmonary disease and coronavirus COVID-19 detection from X-rays. Comput. Methods Programs Biomed. **196**, 105608 (2020)

33. Panwar, H.; Gupta, P.K.; Siddiqui, M.K.; Morales-Menendez, R.; Bhardwaj, P.; Singh, V.: A deep learning and grad-CAM based color visualization approach for fast detection of COVID-19 cases using chest X-ray and CT-Scan images. Chaos, Solitons Fractals **140**, 110190 (2020)

34. Goel, T.; Murugan, R.; Mirjalili, S.; Chakrabartty, D.K.: Opt-CoNet: an optimized convolutional neural network for an automatic diagnosis of COVID-19. Appl. Intell. **51**(3), 1351–1366 (2021)

35. Jain, R.; Gupta, M.; Taneja, S.; Hemanth, D.J.: Deep learning based detection and analysis of COVID-19 on chest X-ray images. Appl. Intell. **51**(3), 1690–1700 (2021)

36. Abbas, A.; Abdelsamea, M.M.; Gaber, M.M.: Classification of COVID-19 in chest X-ray images using DeTraC deep convolutional neural network. Appl. Intell. **51**(2), 854–864 (2021)

37. Zebin, T.; Rezvy, S.: COVID-19 detection and disease progression visualization: Deep learning on chest X-rays for classification and coarse localization. Appl. Intell. **51**(2), 1010–1021 (2021)

38. Punn, N.S.; Agarwal, S.: Automated diagnosis of COVID-19 with limited posteroanterior chest X-ray images using fine-tuned deep neural networks. Appl. Intell. **51**(5), 2689–2702 (2021)

39. Kaur, M.; Kumar, V.; Yadav, V.; Singh, D.; Kumar, N.; Das, N.N.: Metaheuristic-based deep COVID-19 screening model from chest X-ray images. J. Healthc. Eng. **2021**, 8829829 (2021). https://doi.org/10.1155/2021/8829829

40. Singh, D.; Kumar, V.; Kaur, M.: Densely connected convolutional networks-based COVID-19 screening model. Appl. Intell. **51**(5), 3044–3051 (2021)

41. Jaiswal, A.; Gianchandani, N.; Singh, D.; Kumar, V.; Kaur, M.: Classification of the COVID-19 infected patients using

DenseNet201 based deep transfer learning. J. Biomol. Struct. Dyn. (2020). https://doi.org/10.1080/07391102.2020.1788642

42. Tiwari, S.; Jain, A.: Convolutional capsule network for COVID-19 detection using radiography images. Int. J. Imaging Syst. Technol. **31**(2), 525–539 (2021)

43. Zhang, Y.D.; Satapathy, S.C.; Liu, S.; Li, G.R.: A five-layer deep convolutional neural network with stochastic pooling for chest CT-based COVID-19 diagnosis. Mach. Vis. Appl. **32**(1), 1–13 (2021)

44. Zhang, Y.D.; Satapathy, S.C.; Zhang, X.; et al.: COVID-19 diagnosis via densenet and optimization of transfer learning setting. Cogn Comput (2021). https://doi.org/10.1007/s12559-020-09776-8

45. Li, L.; Xu, M.; Liu, H.; Li, Y.; Wang, X.; Jiang, L.; Wang, N.: A large-scale database and a CNN model for attention-based glaucoma detection. IEEE Trans. Med. Imaging **39**(2), 413–424 (2019)

46. Simonyan, K.; Zisserman, A.: Very deep convolutional networks for large-scale image recognition (2014). arXiv:1409.1556

47. Howard, A.G.; Zhu, M.; Chen, B.; Kalenichenko, D.; Wang, W.; Weyand, T.; Andreetto, M.; Adam, H.: Mobilenets: Efficient convolutional neural networks for mobile vision applications (2017). arXiv:1704.04861

48. Szegedy, C.; Ioffe, S.; Vanhoucke, V.; Alemi, A.A.: Inception-v4, inception-resnet and the impact of residual connections on learning. In: Thirty-first AAAI Conference on Artificial Intelligence (2017)

49. He, K.; Zhang, X.; Ren, S.; Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778 (2016)

50. Osawa, K.; Tsuji, Y.; Ueno, Y.; Naruse, A.; Yokota, R.; Matsuoka, S.: Second-order optimization method for large mini-batch: training resnet-50 on imagenet in 35 epochs (2018). arXiv:1811.12019

51. Chollet, F.: Xception: deep learning with depthwise separable convolutions. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1251–1258 (2017)

52. Iandola, F.N.; Han, S.; Moskewicz, M.W.; Ashraf, K.; Dally, W.J.; Keutzer, K.: SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and < 0.5 MB model size (2016). arXiv:1602.07360

53. Redmon, J.; Farhadi, A.: Yolov3: An incremental improvement (2018). arXiv:1804.02767

54. Tan, M.; Le, Q.: Efficientnet: rethinking model scaling for convolutional neural networks. In: International Conference on Machine Learning, pp. 6105–6114. PMLR (2019)

55. Radford, A.; Metz, L.; Chintala, S.: Unsupervised representation learning with deep convolutional generative adversarial networks (2015). arXiv:1511.06434

56. Zoph, B.; Vasudevan, V.; Shlens, J.; Le, Q.V.: Learning transferable architectures for scalable image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 8697–8710 (2018)

57. Shalbaf, A.; Vafaeezadeh, M.: Automated detection of COVID-19 using ensemble of transfer learning with deep convolutional neural network based on CT scans. Int. J. Comput. Assist. Radiol. Surg. **16**(1), 115–123 (2021)

58. Autee, P.; Bagwe, S.; Shah, V.; Srivastava, K.: StackNet-DenVIS: a multi-layer perceptron stacked ensembling approach for COVID-19 detection using X-ray images. Phys. Eng. Sci. Med. **43**(4), 1399–1414 (2020)

59. Apostolopoulos, I.D.; Mpesiana, T.A.: Covid-19: automatic detection from x-ray images utilizing transfer learning with convolutional neural networks. Phys. Eng. Sci. Med. **43**(2), 635–640 (2020)

60. Hochreiter, S.; Schmidhuber, J.: Long short-term memory. Neural Comput. **9**(8), 1735–1780 (1997)

61. Kumar, J.L.M.; Rashid, M.; Musa, R.M.; Razman, M.A.M.; Sulaiman, N.; Jailani, R.; Majeed, A.P.A.: The classification of EEG-based wink signals: a CWT-transfer learning pipeline. ICT Express (2021)

62. Mehmood, A.; Iqbal, M.; Mehmood, Z.; et al.: Prediction of heart disease using deep convolutional neural networks. Arab. J. Sci. Eng. **46**, 3409–3422 (2021). https://doi.org/10.1007/s13369-020-05105-1

63. Nanda, N.; Kakkar, P.; Nagpal, S.: Computer-aided segmentation of liver lesions in CT scans using cascaded convolutional neural networks and genetically optimised classifier. Arab. J. Sci. Eng. **44**, 4049–4062 (2019). https://doi.org/10.1007/s13369-019-03735-8